sensors

Volume 1

State-of-the-Art Sensors Technology in Spain 2017

Edited by
Gonzalo Pajares Martinsanz

MDPI

State-of-the-Art Sensors Technology in Spain 2017

Volume 1

Special Issue Editor

Gonzalo Pajares Martinsanz

MDPI

Special Issue Editor
Gonzalo Pajares Martinsanz
Department Software Engineering and Artificial Intelligence
University Complutense of Madrid
Spain

This edition is a reprint of the Special Issue published online in the open access journal *Sensors* (ISSN 1424-8220) from 2017–2018 (available at: http://www.mdpi.com/journal/sensors/special_issues/SA_sensors_technology_spain_2017).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, F.M.; Lastname, F.M. Article title. *Journal Name* **Year**, *Article number*, page range.

Table of Contents

About the Special Issue Editor

Gonzalo Pajares Martinsanz received the Ph.D. degree in Physics from Distance University from Spain in 1995, discussing a thesis on stereovision. Since 1988 he worked at Indra in critical real-time software development. He also was working at Indra Space and INTA in advanced image processing for Remote Sensing. He joined the University Complutense of Madrid in 1995 on the Faculty of Informatics (Computer Science) at the Department of Software Engineering and Artificial intelligence. His current research interests include computer and machine visual perception, artificial intelligence, decision making, robotics and simulation with a lot of publications, including several books, on these topics. He is director of the ISCAR Research Group. He is Associated Editor in the indexed journal of Remote Sensing and serves as a member of the Editorial Board in the following indexed journals: Sensors, EURASIP Journal of Image and Video processing, Pattern Analysis and Applications. He is the Editor-in-Chief of Journal of Imaging.

Preface to "State-of-the-Art Sensors Technology in Spain 2017"

Since 2009 six special issues has been published on sensors and technologies in Spain, where researchers present their successful progress. Forty high quality works have been collected and reproduced on this book, demonstrating significant achievements.

They are self-contained works addressing different sensor-based technologies, procedures and applications on several areas with a wide range of devices and useful developments. Readers will also find an excellent source of resources when necessary in the development of his research, teaching or industrial activity.

Although the book is focused on sensors and technologies in Spain, it describes worldwide advanced developments and references on the covered topics useful on the contexts addressed. Some works have been or come from international collaborations.

Our society is demanding new technologies for data acquisition, processing and transmission for immediate actions or knowledge, with important contributions to the welfare or actuations when required.

The international scientific and industrial communities worldwide are also indirect beneficiary. Indeed, the book provides insights and solutions for the different problems covered. Also, it lays the foundation for future advances toward new challenges. In this regard, new sensors contribute to the solution of existing problems, but also conversely, where the need to resolve certain problems demands the development of new technologies or procedures. This book is in this way.

We are grateful to all the persons involved in the edition of this book. Without the invaluable contribution of authors together with the excellent help of reviewers, this book would not have seen the light. More than 200 authors have contributed to this book.

Thanks to Sensors journal and the team of people involved in the edition and production of this book for the support and encouragement.

Gonzalo Pajares Martinsanz
Special Issue Editor

The Use of Phononic Crystals to Design Piezoelectric Power Transducers

Silvia Ronda [1], José Luis Aragón [2], Elvira Iglesias [1] and Francisco Montero de Espinosa [1,*]

[1] ITEFI-CSIC, Serrano 144, 28006 Madrid, Spain; silvia.ronda@csic.es (S.R.); eiglesias.jml@gmail.com (E.I)
[2] Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México, Apartado Postal 1-1010, Querétaro 76230, Mexico; aragon@fata.unam.mx
* Correspondence: francisco.montero@csic.es; Tel.: +34-620-068-406

Abstract: It was recently proposed that the lateral resonances around the working resonance band of ultrasonic piezoelectric sandwich transducers can be stopped by a periodic array of circular holes drilled along the main propagation direction (a phononic crystal). In this work, the performance of different transducer designs made with this procedure is tested using laser vibrometry, electric impedance tests and finite element models (FEM). It is shown that in terms of mechanical vibration amplitude and acoustic efficiency, the best design for physiotherapy applications is when both, the piezoceramic and an aluminum capsule are phononic structures. The procedure described here can be applied to the design of power ultrasonic devices, physiotherapy transducers and other external medical power ultrasound applications where piston-like vibration in a narrow band is required.

1. Introduction

Ultrasonic piezoelectric power transducers normally work in the length mode and are made of two sections: a piezoceramic resonator, as the active part, and either one or two different metallic sections bonded or fixed by a screw (pre-stressed transducers). From these, the Langevin type [1] is the most common design, where the transducer length is usually half the wavelength (λ) at the resonance frequency. A slightly different design is that of the power transducers used for physiotherapy, which are made with two sections: the piezoceramic resonator and a metallic capsule [2]. Piezoceramics and metals with low mechanical losses are used to optimize energy delivery at the working frequency.

Langevin type transducers are mainly used when a clean length resonance with a piston-like mechanical displacement at the emitting transducer surface is required. To achieve this, a length-to-radial aspect ratio higher than 2 should be used to avoid any coupling between length and radial resonance modes and to optimize the transducer electromechanical coupling coefficient [1].

In this work, we are interested in transducers for physiotherapy, where the typical design consists of a piezoceramic disc resonator and an attached metal case that will be in contact with the human body. The simplest device can be a piezoceramic disc bonded to a metallic disc. Most of these transducers operate at discrete frequencies in the range 1–3 MHz, thus the desired length-to-radial aspect ratio to produce no interference between the thickness and the radial modes cannot be achieved since the required transducer acoustic aperture is bigger than the transducer length. The transducer aperture, and in consequence the transducer diameter, is typically in the range between 10 mm and 30 mm, and the thickness should be lower than 5 mm. These acoustic constraints imply a low length-to-radial aspect ratio, thus radial modes always interfere with the main length mode. In this case the transducer emitting surface is far from the ideal piston-like behavior [2].

Although, as it has already been mentioned, this piston-like behavior can be achieved by adequately controlling the geometry of the resonators, it would be desirable to find a way to achieve this regardless the geometry and dimensions of the resonator. Piezocomposite structures have been proposed since the 80s to improve the electromechanical efficiency of the piezoelectric resonators by decreasing the specific acoustic impedance of the piezoceramics [3]. The most successful piezocomposite design consists of dicing a piezoceramic resonator along its thickness, forming a square pattern with grooves filled with a polymer. By controlling the width and pitch of the dice, lateral resonances in the planar direction are strongly attenuated, the surface vibrates in-phase and larger electromechanical coefficients are attained. Also, triangular-cut piezocomposites have been designed for unimodal operation [4]. Piezocomposites are the best option for ultrasonic imaging applications where low acoustic power is used and large frequency bandwidth is required. However, they are not useful for ultrasonic high power applications at narrow frequency band, which is our main interest in this work. The basic requirement of a piezocomposite design is to fix the spacing of the bars away from one half the wavelength.

Recently we proposed the application of the principles behind photonic and phononic materials to attenuate or even stop the lateral resonances, producing in this way a piezoceramic resonator adequate for operation in the thickness mode with an in-phase vibration surface [5]. Phononic materials prevent propagation of waves in certain frequency ranges by making use of the fundamental properties of waves, such as scattering and interference, which can produce band gaps, that is, a range of frequencies within which waves cannot propagate through the structure [6–11]. The dispersion relation of the radial waves coming from the lateral contour of a transducer can be modified by making a periodic array of holes along the longitudinal main vibration direction such that these radial waves cannot propagate. Thus, contrary to the usual requirement of piezocomposite design, in this case the strict Bragg condition is required, so the diffracting elements must be at half a wavelength. It should be said that in the field of piezoelectric resonators, phononic crystals where the periodic array consists of piezoelectric elements embedded on a substrate were first studied [12–14]. A different approach consisting of a piezoelectric material with a periodic repetition of hollow cylinders was used to study surface acoustic waves in these piezoelectric phononic crystals [15].

Here, we use the phononic crystal design proposed in [15] and compare the performance of different transducer designs, looking for the most adequate for physiotherapy applications, that is, in terms of mechanical vibration amplitude and acoustic efficiency. We study transducers consisting of a periodic array of cylindrical voids drilled along the main vibrating direction in either one or both the transducer component sections: the piezoceramic and the metallic (aluminum) capsule. We tested four designs and concluded that that the best design for physiotherapy applications is obtained when both, the piezoceramic and an aluminum capsule are phononic structures. For clinical applications the design where only the piezoceramic is a phononic structure could also be adequate because the emitting face (capsule) has no holes.

2. Materials and Methods

The materials used to prepare the samples were PZT-4 piezoceramic as active element and aluminum as passive acoustic line. Four different transducer designs were built. Model 1 consists of a piezoceramic PZT-4 disc with thickness t_c = 2 mm and radius r_c = 15 mm bonded to an aluminum disc with thickness t_{Al} = 3 mm and the same radius. With these dimensions, and the thickness condition, the transducer should vibrate at 1 MHz. In Model 2 both transducer sections have the same dimensions, but the commercial piezoceramic is transformed into a phononic structure by drilling a square lattice of holes in the piezoceramic disc, with a lattice parameter a = 2 mm and holes of 1.3 mm diameter [2]. Model 3 is the same as Model 1, but with a square lattice of holes drilled only in the aluminum disc (not in the piezoceramic); the lattice parameter and filling factor used are the same as in Model 2. Finally, Model 4 consists of a transducer where both, the piezoceramic and the aluminum discs are phononic structures with the same design as in Model 2. Geometries of the four models can be found

in Figure 1, where the *z* axis is the piezoceramic polarization direction. Throughout this paper, the *z* axis will be the out-of-plane vibration direction.

These models were studied numerically and experimentally. For the numerical study, the piezoelectric devices were simulated first using finite elements with COMSOL Multiphysics® 4.4, using the PZT-4 and the aluminum material properties included in the COMSOL library. The calculated magnitudes were the module of the input electrical impedance, the phase of the piezoceramic and the out-of-plane vibration mechanical amplitude of the emission face (in this case, the bottom face of the aluminum).

Experimental measurements of the input electrical impedance were carried out with a 4294A Impedance Analyzer (Agilent, Santa Clara, CA, USA) and the out-of-plane mechanical amplitudes were recorded using laser vibrometry. The vibrometry experimental setup consisted of an OFV 50000 single-point vibrometer (Polytec, Waldwronn, Germany) with VD06 and DD900 decoders with a OFV 505 head, a set of UT 100-75 linear stages (Micro-Controle, Newport Co., Irvine, CA, USA) with a Micro-Controler TL78 main frame, a 3000 Wave Surfer oscilloscope (Teledyne LeCroy, Heidelberg, Germany,) and homemade software to control the stages and the signal acquisition.

3. Results

In a phononic crystal consisting of a square array of air holes, the lattice parameter of the perforations must be around half the wavelength of the wave to be diffracted and the band-gap width is proportional to the so-called filling factor $\pi(r^2/a^2)$, where *a* is the lattice parameter and *r* is the radius of the perforation [5]. When a transducer is made of different materials as it is the case of the transducers studied here, the optimal lattice parameter is different for each transducer section. Aluminum has propagation velocities at different propagation modes (bulk or plate), around 1.5 times the corresponding to PZT. That means that both the lattice parameter and the hole diameter should be 1.5 times bigger for the same frequency. In Ref. [5] a PZT-5A piezoelectric disc with 20 mm of diameter and 2 mm of thickness was studied. There, it was shown that with a lattice parameter *a* = 2 mm and holes of 0.7 mm radius (38% of filling factor), clear indications of the filtering of radial modes at the resonance frequency existed. With these lattice parameters, the band gap in aluminum should be centered at a frequency around 1.5 MHz. Nevertheless, if we use these results, where a filtering band close to 700 kHz around the thickness resonance was measured, we expected a good degree of filtering of the radial modes in the aluminum plate around 1 MHz.

3.1. FEM Modelling

To compare with experimental results, a frequency range from 300 kHz to 1.2 MHz was used for all calculations. In the four models, the bottom piezoelectric disc face was grounded and the upper face was excited with an electric potential of 1V. A mechanical damping was introduced, as isotropic structural factor, of $1 + 2 \times 10^{-2}i$. Given the symmetry of the perforated disc, it is enough to consider only a quarter of the disc, provided that symmetry conditions are applied. The typically sampling size (the finite element mesh size) to solve a wave propagation problem should be at least five elements per wavelength. In the calculated models, bulk and plate waves (mainly symmetric Lamb modes) propagate. Velocities of cylindrical waves coming from the contour of the PZT and aluminum plates with 2 mm and 3 mm of thickness are around 4000 ms^{-1}, which means a wavelength of 4mm at 1 MHz. Therefore, finite element mesh of about $h = 1.5 \times 10^{-4}$ m in the drilled domain and $h = 5 \times 10^{-4}$ m in the remaining domain was used. The mesh was even refined at the boundaries.

Model 1 (Figure 1a) where neither the piezoceramic or the aluminum section have perforations, is the typical commercial design of a 1 MHz physiotherapy transducer. Figure 2a shows its simulated module input electrical impedance and its phase, calculated using a frequency step of 4.66×10^3 Hz (about 193 points of the used frequency range). The first minimum of the impedance module is at 500 kHz and the second minimum is at 1 MHz, the first and second thickness resonances related to half and one wavelength, respectively.
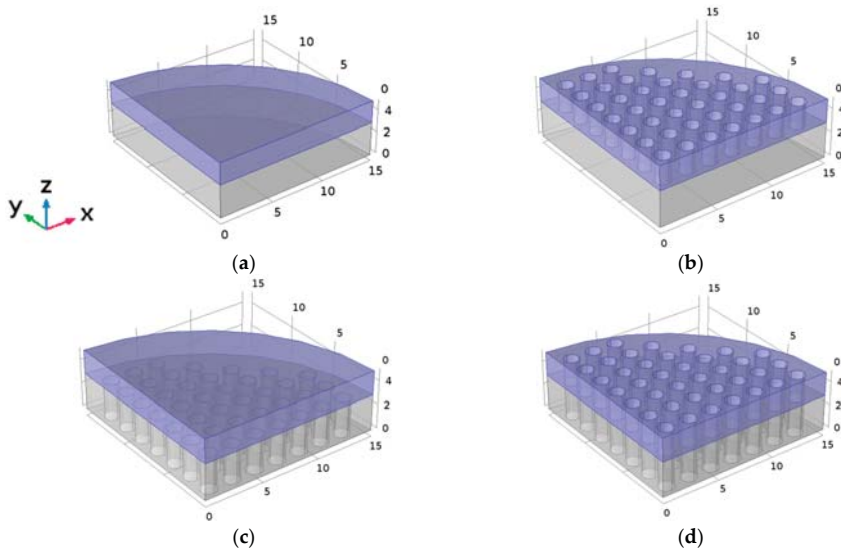
Figure 1. Geometry of the piezoelectric transducer used for the simulations. The scale is in mm; the piezoceramic PZT-4 disc is in blue and the aluminum plate in grey. (**a**) Model 1 and reference mechanical axis; (**b**) Model 2; (**c**) Model 3 and (**d**) Model 4.

These two main thickness resonances are obviously coupled with all the radial modes of the entire cylindrical structure. The plate modes excited at the cylindrical contour interfere with the thickness mode, thus modifying the amplitude and phase of the out-of-plane displacement of both the PZT and the aluminum transducer faces. The signature of a combination of Bessel functions are superimposed to the plane wave in-phase displacement. The z component of the cylindrical waves is not so strong as to create complete counter-phase vibration. The consequence of this mixture of thickness and radial vibration modes is that the mechanical displacement amplitude pattern of the main circular surfaces of the transducer at thickness resonance (997 kHz) becomes far from the thickness in-phase assumption, showing the signature of radial modes both in the amplitude and in the phase, as shown in Figures 3a and 4a. Consequently, the acoustic field emitted into a medium, such as the human body, is not related to the acoustic diffraction piston-like assumption and the physiotherapists, even aware of the circular diffraction theory, will have no idea about the real field distribution emitted into the body [2].

Figure 2b shows the input electrical impedance of Model 2. In this model, only the piezoelectric disc is a phononic structure. As it can be seen, the signature of the radial modes is less evident; the more one approaches the center of the phononic crystal frequency band (1 MHz), the cleaner this signature is. Moreover, as reported in [5], the resonance frequency and the quality factor Q decrease. The mechanical displacement amplitude pattern of the aluminum surface at the resonance frequency (951 kHz) is shown in Figures 3b and 4b. It is evident that the influence of the radial modes is lower than in the case of the device without phononic crystal. The out-of-plane vibration still has radial symmetry, but it is less concentrated and in-phase (Figure 3b). The phononic crystal structure of the piezoceramic was capable of improving the phase vibration, even at the aluminum transducer surface.

Figure 2c shows the input electrical impedance of Model 3, where only the aluminum disc is a phononic structure. The filtering of the radial modes is not as evident as in Model 2 since the design of the phononic crystal is not well fitted to the aluminum plate, as explained at the beginning of this Section. The vibration is mainly in phase and its amplitude is more regular along the surface. As expected, circular modes are not present (Figure 3c). Despite that the aluminum phononic crystal is not centered at 1 MHz, Model 3 seems to be effective in some degree. The resonance frequency is

lower than the measured in Model 1 because the aluminum section suffers the same softening effect reported in [5].
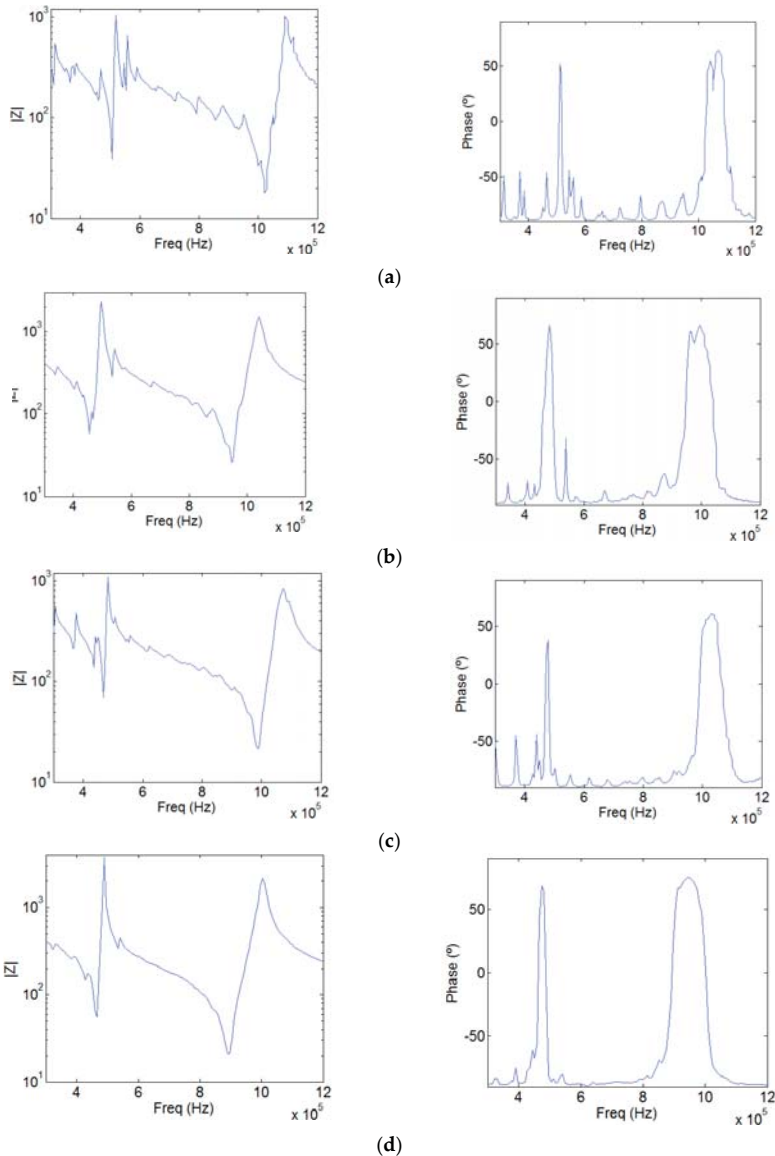
Figure 2. Calculated input electrical impedance. (**a**) Model 1 impedance module in logarithmic scale and phase. The same for Model 2 (**b**), Model 3 (**c**) and Model 4 (**d**).

Finally, concerning Model 4, where both the piezoceramic and the aluminum discs were drilled, in Figure 2d we observe that the filtering is quite clear and effective in all the frequency range studied. The softening effect is now stronger, lowering the frequency up to 890 kHz. Figures 3d and 4d show that, apart from the transducer contour, all the surface vibrates in phase and with a quite regular

amplitude distribution. Local maxima are located at the inter-hole space. No radial modes are observed, neither in the input electrical impedance or in the vibration pattern. Figure 5 shows the maximun out of plane vibration amplitude at the maximum elongation of the aluminum surface in the *yz* section ($x = 0$) for each transducer model at their corresponding resonance frequency (997 kHz, 951 kHz, 988 kHz and 894 kHz, respectively). All of them were excited with the same electrical amplitude signal (1 V).
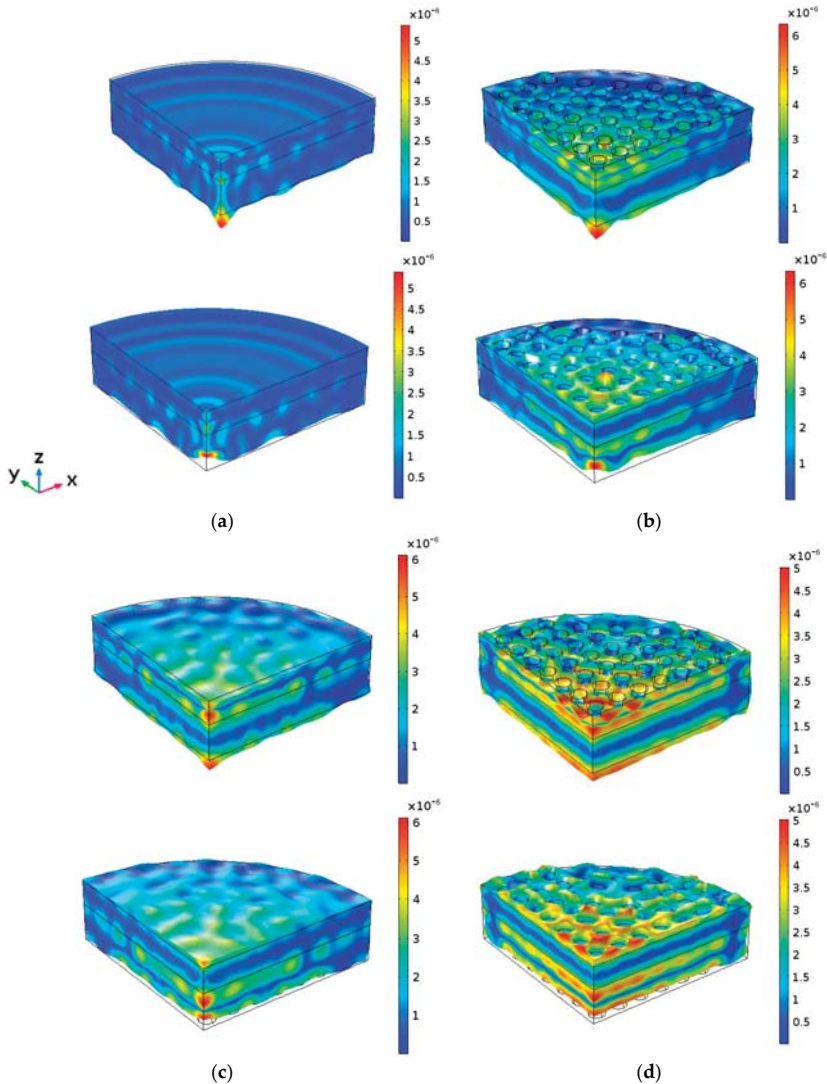
Figure 3. *Z* direction vibration amplitude calculated results for the four models at their respective main thickness resonance (one wavelength resonance). (**a**) Maximum elongation time position of the aluminum face (**up**) and minimum elongation time position (**bottom**) of Model 1. The same for Model 2 (**b**), Model 3 (**c**) and Model 4 (**d**). The excitation voltage of the piezoceramic disc in all the models was 1V. The bar scale is in mm. The frequency for each model was 997 kHz, 951 kHz, 988 kHz and 894 kHz, respectively.
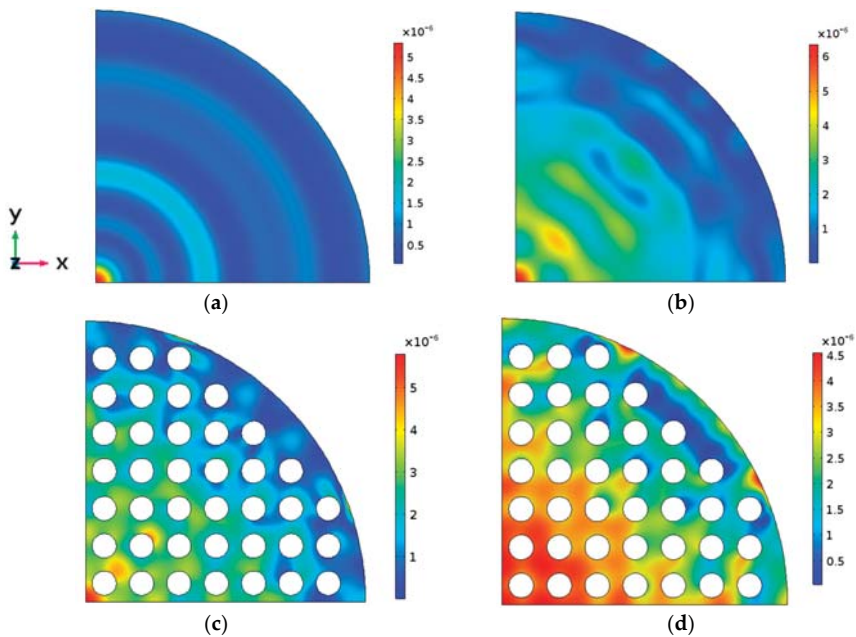
Figure 4. The z direction vibration amplitude calculated at the aluminum surface for the four models at their respective main thickness resonance (one wavelength resonance). (**a**) Model 1, (**b**) Model 2, (**c**) Model 3 and (**d**) Model 4. The excitation voltage of the piezoceramic disc in all the models was 1V. The bar scale is in mm. The frequency for each model was 997 kHz, 951 kHz, 988 kHz and 894 kHz, respectively.

To compare the transduction efficiency of the different models, the average of the maximum value of the vibration along the z axis of all the calculated points in the aluminum face was calculated as a function of the vibration frequency. Figure 6 shows the average, without taking into account the effective radiating surface of each model. As observed, the highest value of each model is attained at the corresponding resonance frequency. The bandwidth is similar. Model 4 displays the highest average mechanical displacement. The softening of the resonance structure is correlated with the increase of the mechanical displacements as expected (see [5]). If the actual emitting surface is considered (Model 3 and Model 4 has a surface with holes), if the average is multiplied by the surface, then these results are modified. From Figure 7 we see that this average amplitude vibration is the same as the obtained from Models 1, 2 and 4. Since Model 4 is the only one with a homogenous phase vibration, we conclude that it is the most efficient.

Summarizing. Model 4 is the best option in terms of mechanical vibration amplitude and acoustic efficiency. Nevertheless, Model 2 is more practical for clinical applications because the emitting face has no holes. Model 2 has the same average mechanical displacement that Model 4 but the vibration distribution is less uniform in amplitude (Figures 3–5).

Since the standard lowest frequency used in physiotherapy is 1 MHz, we now consider two more models (Model 5 and 6), similar to Models 2 and 4, but where the piezoceramic disc has 1.5 mm of thickness (Figure 8a). The KLM monodimensional model was used to choose this piezoceramic thickness to get a transducer with a frequency close to the standard 1 MHz.
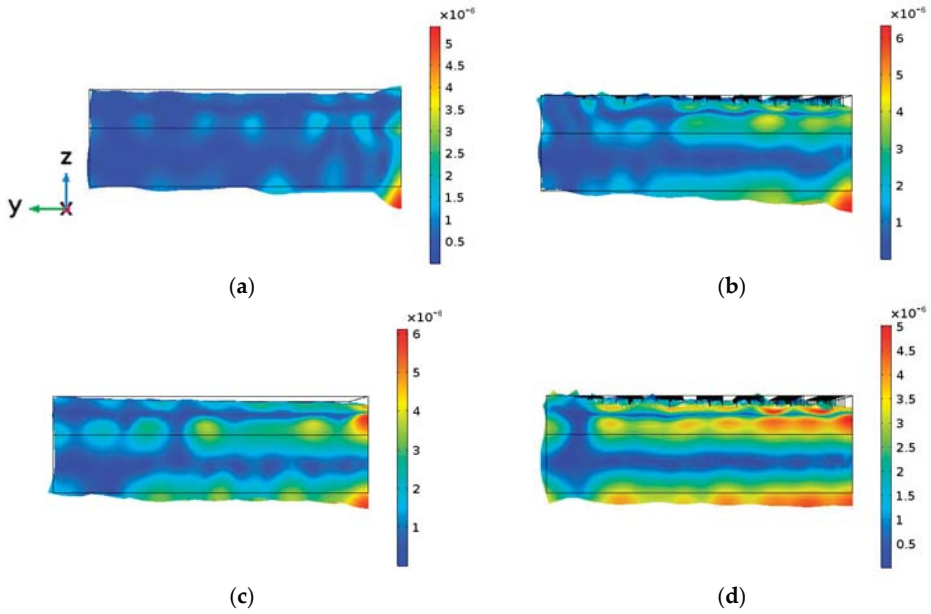
Figure 5. Maximum out-of-plane vibration amplitude at maximum elongation of the aluminum surface in the yz section, $x = 0$, of the transducer for 1 V excitation of the piezoceramic. (**a**) Model 1, (**b**) Model 2, (**c**) Model 3 and (**d**) Model 4. The bar scale is in mm. The frequency for each model was 997 kHz, 951 kHz, 988 kHz and 894 kHz, respectively.
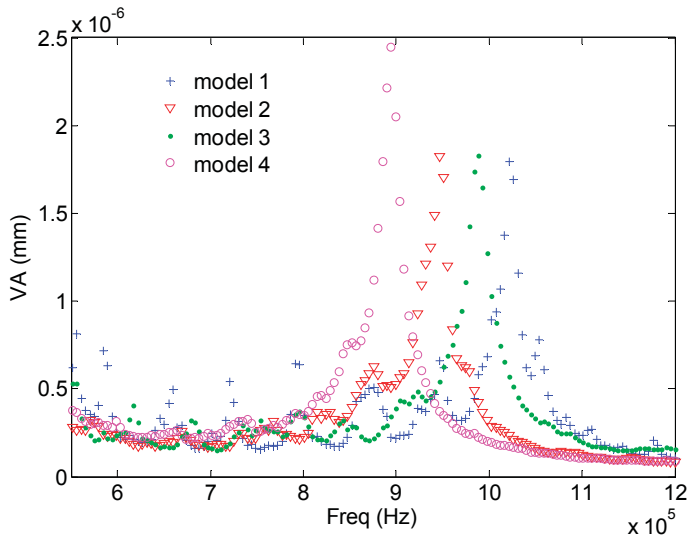
Figure 6. Average of the maximum value of the vibration along the *z* axis of all the calculated points in the aluminum face (VA) as a function of the vibration frequency.
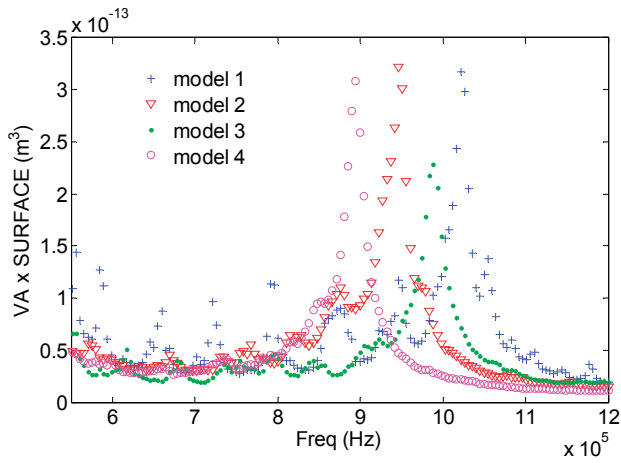
Figure 7. Average of the maximum value of the vibration along the *z* axis of all the calculated points in the aluminum face (VA × surface) as a function of the vibration frequency, multiplied by the real surface of each mode.

Figure 8 shows the module and phase of the input electrical impedance as a function of frequency, the out-of-plane component of the mechanical displacement at the aluminum surface for the resonance frequency and the value of the maximum out-of-plane vibration amplitude at maximum elongation of the aluminum surface in the yz section, for Models 5 and 6. The corresponding resonance frequencies are 1.082 MHz and 1.042 MHz, respectively. We observe that behaviour of these models is quite similar to those of Models 2 and 4.

Figure 8. *Cont.*

(**b**)

(**c**)

(**d**)
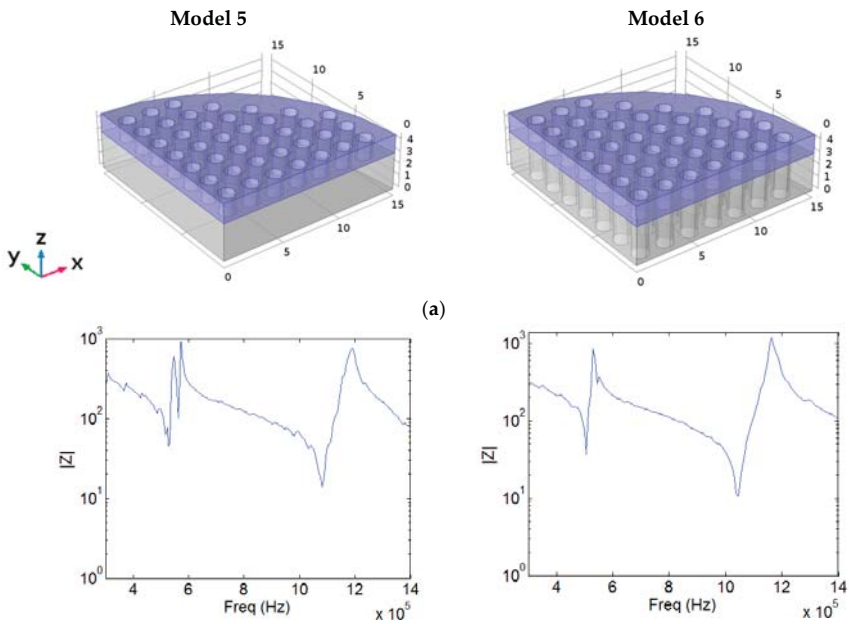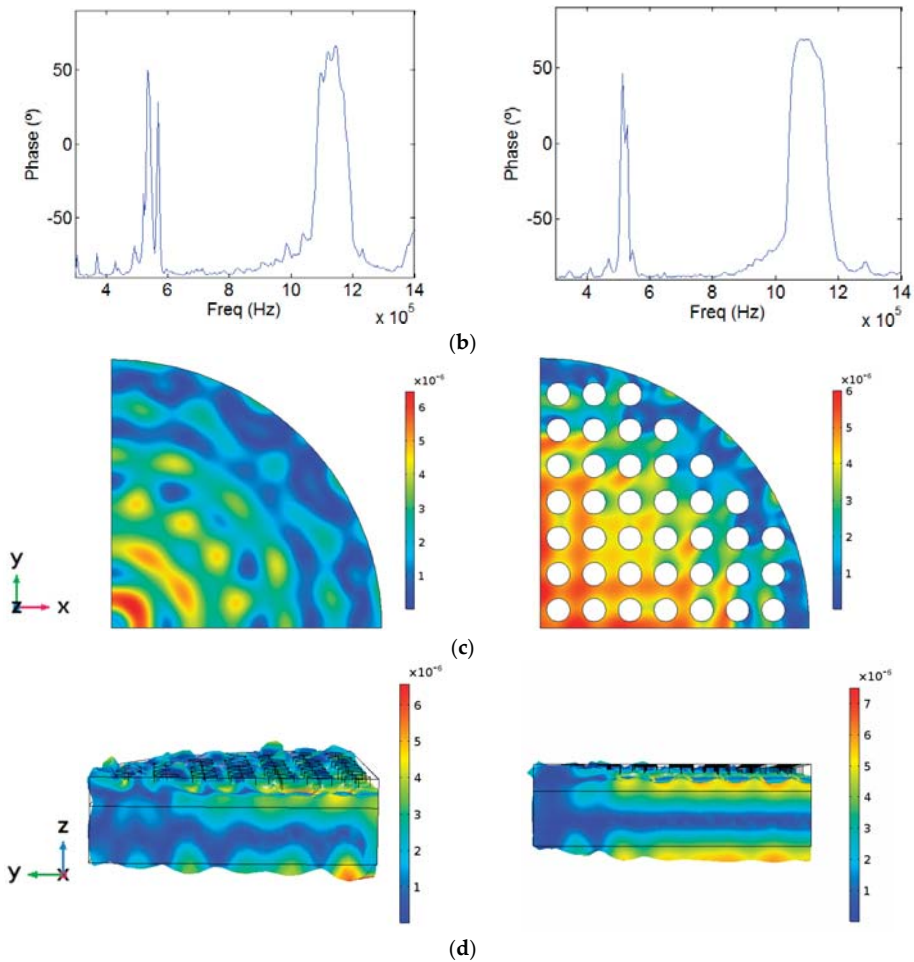
Figure 8. Simulations for Model 5 (left column) and Model 6 (right column) for a 1V excitation of the piezoceramic disc. Scales are in mm: (**a**) Model Setup, (**b**) Module of the input electrical impedance and phase, (**c**) Value of the out of plane component of the mechanical displacement at the aluminum surface for the resonance frequency (1.082 MHz for Model 5 and 1.042 MHz for Model 6) and (**d**) Value of the maximum out-of-plane vibration amplitude in the *yz* section of the transducers for the resonance frequency.

3.2. Experimental Results

Ferroperm Pz26 hard piezoceramics were used to build Models 1, 3, 5 and 6. The aluminum discs and the drilling of the piezoceramics were made at our workshop. The piezoceramic drilling process was made with an automatic programmable drilling system. After several trials, a set of drilling parameters including the maximum number of holes for each drill were defined and finally used to produce the discs with holes. The piezoceramic and the aluminum sections were bonded with Araldite D & HY935.

Models 1 was first tested to compare with the numerical results. Figure 9 shows the set of electrical and mechanical outputs of Model 1. Figure 9a shows the electrical impedance module and phase

and Figure 9b shows the aluminum surface out-of-plane mechanical displacement measured by laser vibrometry in linear arbitrary units. Figure 9c shows the maximum amplitude of the mechanical displacement along the radius.
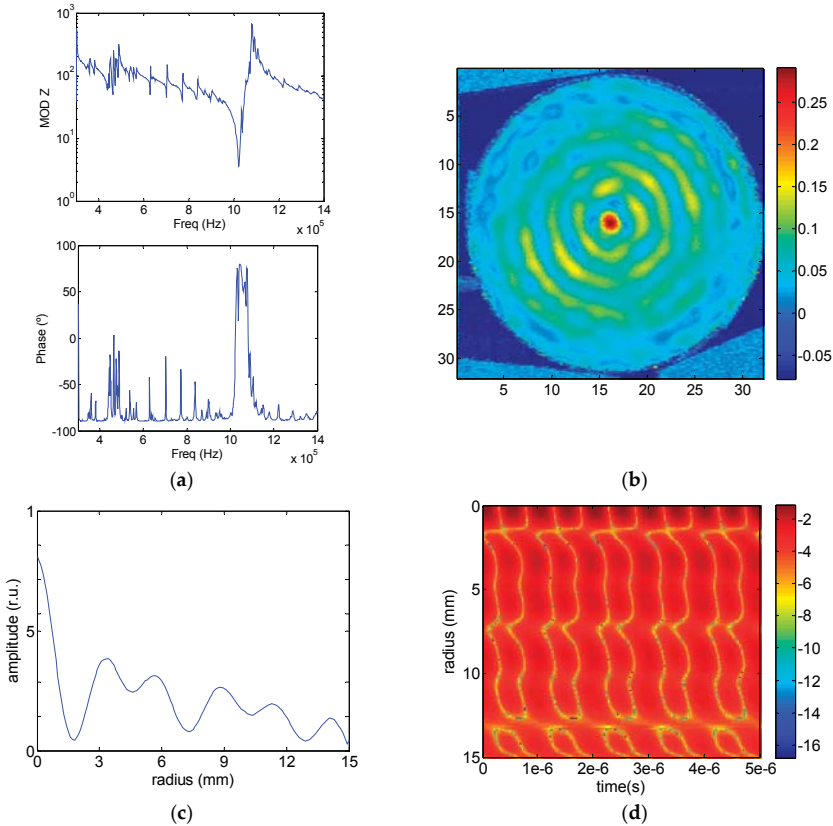
Figure 9. Experimental results with Model 1, (**a**) Module and phase of the electrical impedance; (**b**) Aluminum surface out-of-plane mechanical displacement in linear arbitrary units; (**c**) Mechanical displacement amplitude along the radius; (**d**) Relative mechanical displacement amplitude as a function of time along the radius in logarithmic scale.

As it can be seen, the agreement with the simulated model is quite good (see Figures 2a, 3a, 4a and 5). As the vibrometer spot is scaned from the centre to the border the maximum mechanical amplitude, the wave phase along the radius and the recorded vibration signal as a function of time, were measured and are shown in Figure 9d. The phase variation can be observed by following the amplitude variation for a given time. Horizontal axis correspond to a time of 5 μs. Bar scale gives the amplitude of the signal in relative units logarithmic scale. The figure shows that the phase of the mechanical displacemet varies along the radius as a consequence of the strong coupling between the thickness mode and the thickness component of the radial modes.

Now Model 3 was tested to observe the filtering produced by the aluminum section with drilled holes. Figure 10 shows the same tests as those shown in Figure 9 for Model 1. As predicted by the numerical results, the electrical and vibration measurements show a clear filtering of the radial modes. Interestingly, the phase change observed in Figure 10d agrees with the one observed in Figure 3c.

The surface pattern of Figure 10b displays the same square shadow, rotated in the 110 crystalographic direction, which is also observed in Figure 4c. Finally, Models 5 and 6 were fabricated and tested. Since the goal is to get power transducers with a vibration pattern close to the piston-like assumption, the Model 6, with a complete phononic structure, is the best model as was also predicted by the FEM model. The variation of the mechanical amplitude distribution along the radius is less than 25% along the 75% of the entire transducer emitting face (Figure 11d) with an in-phase vibration behavior (Figure 11f). Model 5 shows a 40% variation of the mechanical amplitude vibration along the radius in the 50% of the emitting surface with in–phase vibration (Figure 11c,e). As in the case of the Model 2, the Model 5 is also a possibility if a metal plate without holes is required for technical reasons.
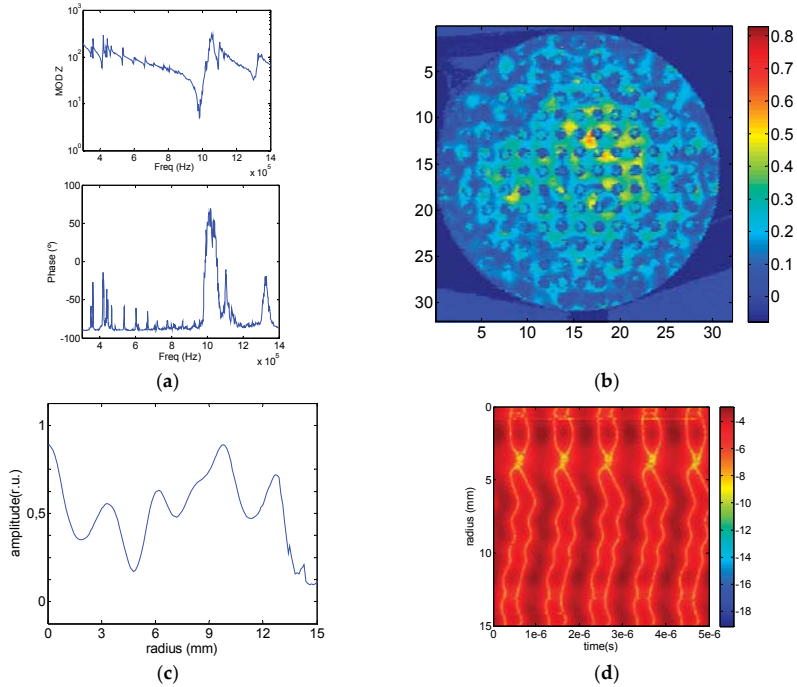
Figure 10. Experimental results for Model 3, (**a**) Module and phase of the electrical impedance; (**b**) Aluminum surface out-of plane mechanical displacement in linear arbitrary units; (**c**) Mechanical maximum displacement amplitude along the radius; (**d**) Relative mechanical displacement amplitude as a function of time along the radius in logarithmic scale.
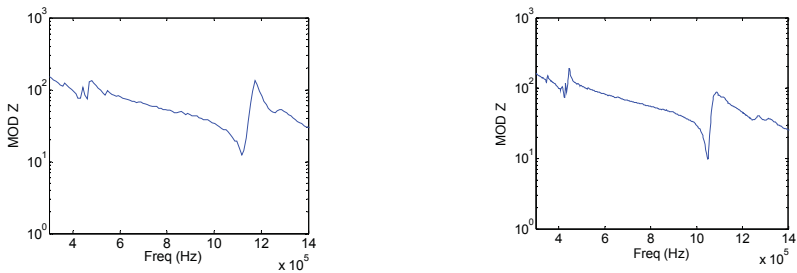
Figure 11. *Cont.*
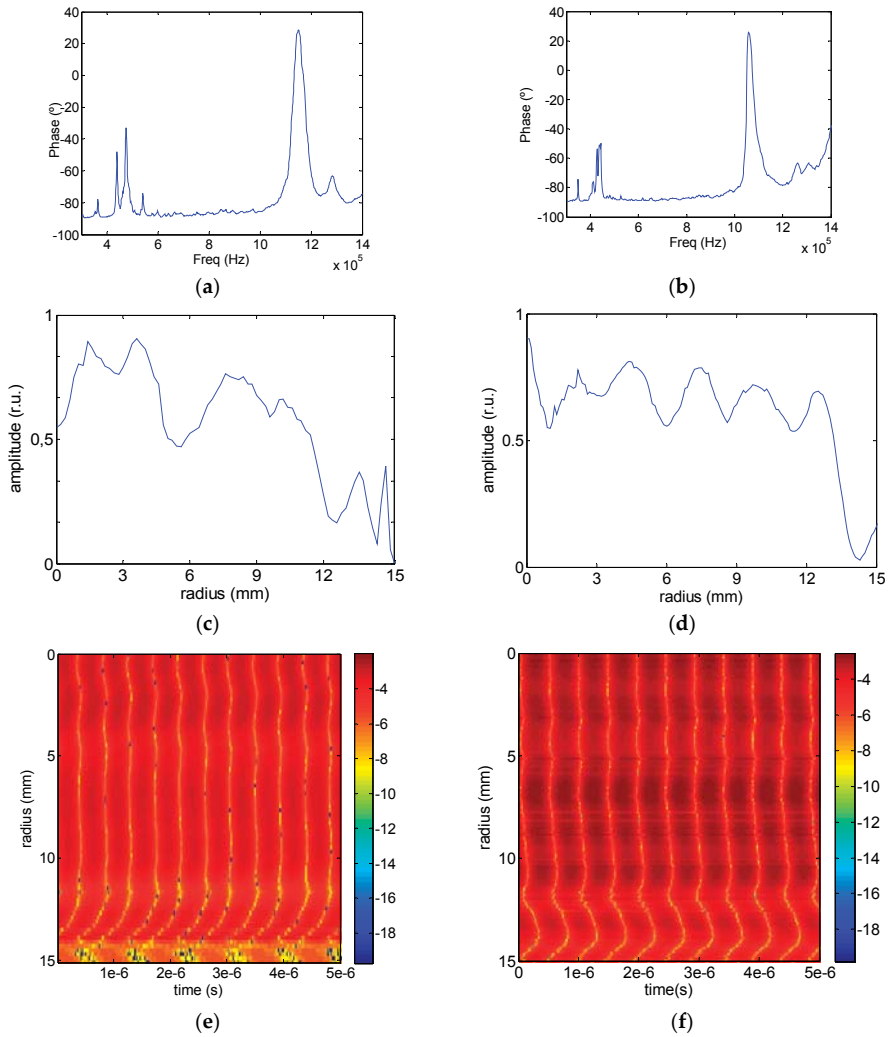
Figure 11. Experimental tests of Models 5 and 6; (**a**) Module and phase of the electrical impedance of Model 5; (**b**) Module and phase of the electrical impedance of Model 6; (**c**) Amplitude of the mechanical displacement along the radius of Model 5; (**d**) Amplitudeof the mechanical displacement along the radius of Model 6; (**e**) Relative mechanical displacement amplitude as a function of time, also along the radius (logarithmic scale) in Model 5; and (**f**) Relative mechanical displacement amplitude as a function of time, also along the radius (logarithmic scale) in Model 6.

Finally, in Figure 12 the calculated maximum z mechanical displacement of the Model 6 is compared with the experimentally measured maximum out-of-plane mechanical displacement. The vibration distribution agrees quite well. The bar scales show that the relative values also agree.
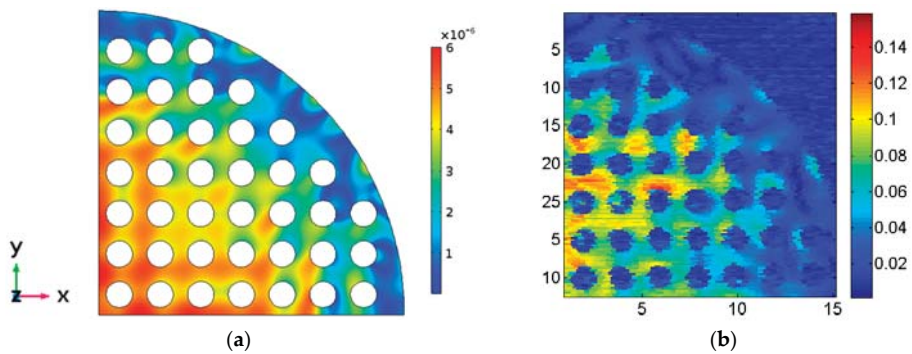
Figure 12. Model 6 out-of-plane mechanical displacement for the resonance frequency around 1 MHz. (**a**) FEM model, (**b**) experimental measurement in linear arbitrary units.

4. Conclusions

Typical power ultrasonic transducer model with two sections, piezoceramic and metal, has been modified following the phononic crystal working principle. This is achieved by transforming the transducer into a phononic crystal by drilling a lattice of holes spaced by an adequate distance calculated to stop the radial modes of the resonator, that is, the resonances arising from the cylindrical contour and those coming from the holes themselves are stopped. The effect of the array of holes in one of the sections or in both sections has been studied with FEM models and tested with real transducers. Electrical and vibrational experimental tests have also been made. The effect produced by performing a lattice of holes either in one of the transducer sections (PZT and aluminum plates) or in both has been studied. Although cancelling the radial modes in the case of PZT is more effective because the design is optimum for this case, the same crystal lattice and filling ratio are used for both transducer sections. With only one section drilled with a phononic structure, a clear filtering of the radial modes is observed, but the effect is maximum when the two transducer sections are drilled. We conclude that in terms of mechanical vibration amplitude and acoustic efficiency, the most efficient design is obtained when both, the piezoceramic and an aluminum capsule, are phononic structures. However, for clinical applications the design where only the piezoceramic is a phononic structure could also be a solution because the emitting face (capsule) has no holes. The method for designing piezoelectric power transducer and testing procedure using here can be applicable to fabricate more efficient physiotherapy ultrasonic transducers with piston-like emitting surfaces.

Author Contributions: F. Montero de Espinosa conceived and designed the experiments; S. Ronda, E. Iglesias and F. Montero de Espinosa performed the experiments; S. Ronda and J.L. Aragón made the FEM models. J.L. Aragón and E. Iglesias analyzed the FEM data; S. Ronda, J.L. Aragón and F. Montero de Espinosa wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Neppiras, E.A. *The Pre-Stressed Piezoelectric Sandwich Transducer*; Ultrasonic International: London, UK, 1973; pp. 295–302.
2. Iglesias, E.; Chinchurtreta, F.; De Frutos, J.; Montero de Espinosa, F. Tailoring the efficiency of ultrasonic transducers for physiotherapy. *Ferroelectrics* **2016**, *499*, 83–89. [CrossRef]
3. Smith, W.A.; Auld, B.A. Modeling 1–3 composite piezoelectrics: thickness-mode oscillations. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **1991**, *38*, 40–47. [CrossRef] [PubMed]

4. Dziewierz, J.; Ramadas, S.N.; Gachagan, A.; O'Leary, R.L.; Hayward, G. A 2D Ultrasonic array design incorporating hexagonal-shaped elements and triangular-cut piezocomposite. In Proceedings of the 2009 IEEE International Ultrasonics Symposium, Rome, Italy, 20–23 September 2009; pp. 422–425.

5. Aragón, J.L.; Quintero-Torres, R.; Domínguez-Juárez, J.L.; Iglesias, E.; Ronda, S.; Montero de Espinosa, F. Planar modes free piezoelectric resonators using a phononic crystal with holes. *Ultrasonics* **2016**, *71*, 177–182. [CrossRef] [PubMed]

6. Maldovan, N. Sound and heat revolutions in phononics. *Nature* **2013**, *503*, 209–217. [CrossRef] [PubMed]

7. Sigalas, M.N.; Economou, E.N. Elastic and acoustic wave band structure. *J. Sound. Vib.* **1992**, *158*, 377–382. [CrossRef]

8. Kushwaha, M.S.; Halevi, P.; Dobrzynski, L.; Djafari-Rouhani, B. Acoustic band structure of periodic elastic composites. *Phys. Rev. Lett.* **1993**, *71*, 2022. [CrossRef] [PubMed]

9. Montero de Espinosa, F.; Jiménez, E.; Torres, M. Ultrasonic band-gap in a periodic two-dimensional composite. *Phys. Rev. Lett.* **1998**, *80*, 1208. [CrossRef]

10. Sánchez-Pérez, J.V.; Caballero, D.; Martínez-Sala, R.; Rubio, C.; Sánchez-Dehesa, J.; Meseguer, F.; Llinares, J.; Gálvez, F. Sound attenuation by a two-dimensional array of rigid cylinder. *Phys. Rev. Lett.* **1998**, *80*, 5325.

11. Vasseur, J.O.; Deymier, P.A.; Frantziskonis, G.; Hong, G.; Djafari-Rouhani, B.; Dobrzynski, L. Experimental evidence for the existence of absolute band gaps in two-dimensional periodic composite media. *J. Phys. Condens. Matter* **1998**, *10*, 6051–6064. [CrossRef]

12. Wilm, M.; Ballandras, S.; Laude, V.; Pastureaud, T. A full 3D plane-wave-expansion model for 1–3 piezoelectric composite structures. *J. Acoust. Soc. Am.* **2002**, *112*, 943–952. [CrossRef] [PubMed]

13. Wang, Y.Z.; Li, F.M.; Huang, W.H.; Wang, Y.S. Effects of inclusion shapes on the band gaps in two-dimensional piezoelectric phononic crystals. *J. Phys. Condens. Matter* **2007**, *19*, 496204. [CrossRef]

14. Wang, Y.Z.; Li, F.M.; Kishimoto, K.; Wang, Y.S.; Huang, W.H. Wave band gaps in three-dimensional periodic piezoelectric structures. *Mech. Res. Commun.* **2009**, *36*, 461–468. [CrossRef]

15. Laude, V.; Wilm, M.; Benchabane, S.; Khelif, A. Full band gap for surface acoustic waves in a piezoelectric phononic crystal. *Phys. Rev. E* **2005**, *71*, 036607. [CrossRef] [PubMed]

Multisensory System for the Detection and Localization of Peripheral Subcutaneous Veins

Roemi Fernández * and Manuel Armada

Centre for Automation and Robotics (CAR) CSIC-UPM, Carretera de Campo Real, Km. 0,200. La Poveda, Arganda del Rey, 28500 Madrid, Spain; manuel.armada@csic.es

* Correspondence: roemi.fernandez@car.upm-csic.es; Tel.: +34-918-711-900

Abstract: This paper proposes a multisensory system for the detection and localization of peripheral subcutaneous veins, as a first step for achieving automatic robotic insertion of catheters in the near future. The multisensory system is based on the combination of a SWIR (Short-Wave Infrared) camera, a TOF (Time-Of-Flight) camera and a NIR (Near Infrared) lighting source. The associated algorithm consists of two main parts: one devoted to the features extraction from the SWIR image, and another envisaged for the registration of the range data provided by the TOF camera, with the SWIR image and the results of the peripheral veins detection. In this way, the detected subcutaneous veins are mapped onto the 3D reconstructed surface, providing a full representation of the region of interest for the automatic catheter insertion. Several experimental tests were carried out in order to evaluate the capabilities of the presented approach. Preliminary results demonstrate the feasibility of the proposed design and highlight the potential benefits of the solution.

Keywords: SWIR camera; TOF camera; peripheral vein detection; subcutaneous vein localization; automatic catheter insertion

1. Introduction

It is estimated that about 90–95% of patients in hospitals receive some kind of intravenous therapy. In comparison with other methods of administration, the intravenous technique is the fastest means of providing fluids and medications throughout the human body. That is the reason why it is widely used to transfuse blood and blood products, provide parenteral nutrition, correct dehydration and electrolyte imbalances, deliver medication and chemotherapeutic agents, and provide avenues for dialysis/apheresis, hemodynamic monitoring, and diagnostic testing [1].

Within intravenous therapy, the peripheral venous catheter is the most frequently used vascular access device. It is applied to most emergency room and surgical patients, and before some radiological imaging techniques using radiocontrast, resulting in more than a billion peripheral intravenous catheters being used per year worldwide [2]. The procedure implies inserting a catheter into the vein by a needle, which is subsequently removed while the small tube of the cannula remains in. The process involves then a very demanding motor coordination, requiring proper training and significant experience, especially when dealing with children and the elderly. Even so, errors are very common, and often, nursing staff has to try several times before to place the needle successfully, which causes pain and distress to the patient, and frustration to the clinicians. Replacement of catheters and rotation of site is recommended every 72 to 96 h in order to reduce the risk of phlebitis and bloodstream infection [3–5], which means repeating the process several times in case of prolonged stays in the hospital. In addition, the insertion of peripheral intravenous catheters exposes health care workers to certain occupational risks, such as a needlestick injury and exposure to blood, which can lead to bloodborne infections [6,7].

In the last decade, several robotic systems have been integrated successfully in the medical field, offering objective and measurable advantages in comparison with traditional procedures, mainly due to the high accuracy and repeatability of their actions [8,9]. These features also make robotic systems stand out as an alternative to provide faster, safer and less painful insertion of intravenous catheters [10–12]. One fundamental step to attain such systems is the effective detection and localization of target subcutaneous veins.

The current state of the art related to automatic detection of veins is predominantly focused on the recognition of finger-vein patterns that can be used as biometric signatures in personal identity authentication systems [13]. For instance, in [14], the repeated tracking of dark lines on an image of a finger captured under infrared light is proposed for finger-vein pattern extraction. In [15], the same authors present an improved method based on the calculation of local maximum curvatures in cross-sectional profiles of the image. In [16], a mean curvature method, which uses geometrical properties of the intensity field to find alley-like structures with negative mean curvatures is proposed for the same purpose. In [17], authors present a finger vein extraction method using gradient normalization and principal curvature calculation. In [18], a finger-vein pattern identification system based on PCA for image pre-processing and feature extraction using LDA is described. Finger-vein image enhancement is proposed in [19] by using a fuzzy-based fusion method with Gabor and Retinex filtering, whereas in [20], authors propose a new finger-vein capturing device that ensures accurate finger positioning to reduce misalignment when veins images are captured. All these methods have in common the fact that they are based on an image of a finger illuminated with an infrared light and acquired by a CCD camera.

Vein identification for automated intravenous drug delivery was proposed in [21]. The system consisted of a web-camera, a near-infrared light and a Digital Single Lens Reflex (DSLR) camera with an external filter to block visible light. The authors concluded that better results are achieved for the images captured with the DSLR and poor results for images obtained with the web-camera can be attributed to the low contrast quality of these images.

Multispectral imaging systems and NIR spectroscopy have also been considered by several authors. Subcutaneous vein detection using multispectral imaging is proposed in [22]. In this case, visible and NIR images acquired by a multispectral imager are used in a normalized subtraction algorithm for improving contrast, and consequently the vein detection performance. In [23], near-infrared spectroscopy is proposed for an education-focused mobile medical application devised to help to improve the decision-making skills of healthcare students in venipuncture. There are also three commercial devices approved by the FDA that use NIR spectroscopy to facilitate peripheral intravenous catheter insertion: the VeinViewer [24] (Christie Medical Holdings, Memphis, TN, USA), the AccuVein [25] (AccuVein LLC, Cold Spring Harbor, New York, NY, USA) and the VascuLuminator [26] (De Koningh Medical Systems, Arnhem, The Netherlands). All these solutions are conceived to improve visualization of subcutaneous veins instead of detecting them automatically.

There is scarce literature related to 3D localization of subcutaneous veins. In [11,12,27], authors propose active optical triangulation for range data acquisition and parametric surface modelling to store the 3D shape of the patient arm. Active optical triangulation is achieved by combining a camera and a laser stripe line generator. Worthy of mention is also the research presented in [28]. The proposed system provides augmented vein structures that are back-projected and superimposed on the skin surface of the hand for assisting doctors in locating the injection veins. The system consists of two industrial cameras, a color micro projector, NIR light sources, a support structure with multi-degree of freedom, and an underpan. Veins are segmented by a multiple-feature clustering method. Vein structures captured by the two cameras are matched and reconstructed based on epipolar constraint and homographic property. The skin surface is reconstructed by active structured light with special encoding values. Results show that the system effectively provides augmented display and visualization of subcutaneous veins.

This paper presents an automatic system that combines a SWIR (Short-Wave Infrared) camera, a TOF (Time-Of-Flight) camera, a NIR (Near Infrared) lighting system and the associated algorithm for

the detection and localization of peripheral subcutaneous veins. The solution is intended to be used for the future automation of peripheral intravenous catheter insertion.

The rest of the paper is organized as follows: Section 2 describes the design and implementation of the proposed multisensory system and the associated algorithm. Section 3 presents the results obtained from the different experimental tests that have been carried out. Section 4 discusses the main results of this work and finally, Section 5 summarizes the major conclusions.

2. Materials and Methods

This section describes the multisensory system that has been designed for the data acquisition and explains the processing algorithm that has been implemented for the automatic detection and localization of peripheral subcutaneous veins.

2.1. Multisensory System Description

The proposed multisensory system is based on the combination of a GoldEye P-032 SWIR camera (Allied Vision, Stradtroda, Germany), a SwissRanger SR-400011 TOF 3D camera (Mesa Imaging, Zürich, Switzerland), and a NIR light source that consists of 96 LEDs of 940 nm distributed in two linear arrays. The GoldEye P-032 SWIR camera has a high spectral response from 900 to 1700 nm thanks to its indium gallium arsenide (InGaAs) sensor and features a maximum frame rate of 30 fps at its full resolution of 636×508 pixels with 14-bit A/D conversion. The TOF camera provides a depth map and amplitude image at the resolution of 176×144 pixels with 16 bit floating-point precision, as well as x, y and z coordinates to each pixel in the depth map. IR light can penetrate human tissues to about 3 to 5 mm subcutaneous depth before losing coherence and directionality to diffusion due to the low optical absorption of human skin and muscles in the NIR region of the electromagnetic spectrum [29,30]. On the contrary, blood is a strong absorber of NIR radiation, increasing the contrast between the subcutaneous veins and the surrounding tissues in NIR images [31,32]. Thus, the SWIR camera enables the acquisition of the required data for the detection of areas of interest that could belong to peripheral subcutaneous veins, whereas the TOF camera supplies simultaneously fast acquisition of accurate distances and intensity images of targets, enabling their localization in the coordinate space. Intrinsic and extrinsic calibration parameters of both cameras were estimated by using the Matlab camera calibrator app (http://www.mathworks.com/products/matlab/). A distance measurement calibration was also carried out in Matlab for the TOF camera by following the method proposed in [33].

Figure 1 shows the layout of the different elements that make up the proposed system. Note that the linear arrays of lEDs are placed side by side with the SWIR camera, so the SWIR camera captures the light reflected by the patient arm.
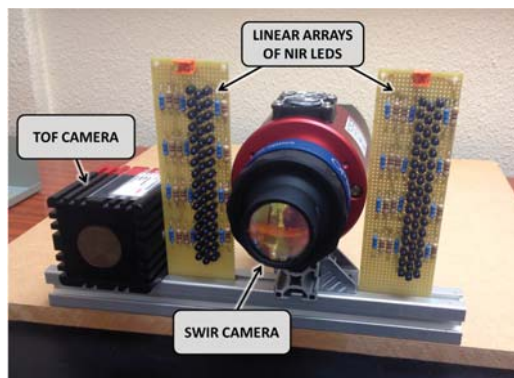
Figure 1. Close-up view of the proposed multisensory system.

In addition, the custom-made multisensory rig that integrates the SWIR camera, the TOF camera and the two linear arrays of NIR LEDs is mounted in a Bosch frame, in such a way that the image planes of the cameras are parallel to the table where subjects place their hands or arms for the automatic detection and localization of peripheral subcutaneous veins.

2.2. Algorithm for Automatic Detection and Localization of Peripheral Subcutaneous Veins

Figure 2 shows a block diagram of the proposed algorithm, which consists of two main parts: one devoted to the features extraction from the SWIR image, and another envisaged for the registration of the range data provided by the TOF camera, with the SWIR image and the results of the peripheral veins detection.
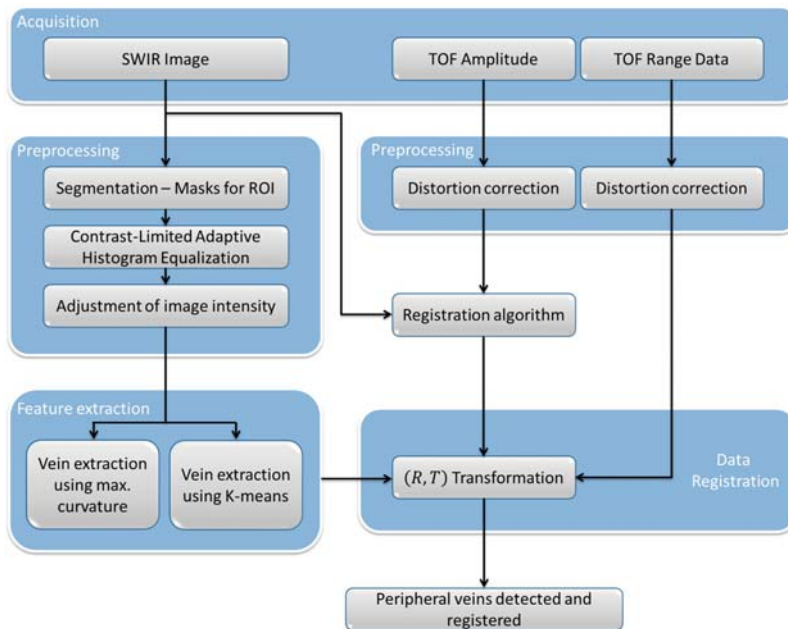
Figure 2. Algorithm for automatic detection and localization of peripheral subcutaneous veins.

For features extraction, the first step involves the segmentation of the acquired SWIR image into two regions, the background and the foreground. The foreground represents the part of the patient's body where the detection of the peripheral veins is going to be carried out, whereas the background represents the rest of the image pixels that are not required for further processing. Otsu's method [34] is then utilized for choosing a global threshold that minimizes the intraclass variance of the background and the foreground pixels. With the attained threshold, a binarization of the original SWIR image is conducted, followed by a dilatation of the background region. The aim of this dilatation is discarding those pixels where the transition from background to foreground region takes place, and which can produce detection errors in the subsequent steps. In fact, the dilatation process is performed twice by using two square structuring elements with two different width values, 10 and 15 pixels, respectively. Thus, two masks are obtained and applied to the original SWIR images. In this way, once the vein extraction is accomplished, a logical AND will be applied to both images, eliminating false detections due to the effects produced by the edges during the segmentation process. Thus, the obtained masks ensure that only the region of interest of the original SWIR image is considered for features extraction, increasing the algorithm performance.

Next, contrast-limited adaptive histogram equalization [35] is applied to the masked images for compensating non-uniform lighting conditions, followed by an adjustment of the image intensity values. This last adjustment consists on mapping the intensity values to new values such that 1% of data is saturated at low and high intensities of the SWIR image. In this way, the contrast between the veins and the surrounding tissues on the SWIR image is enhanced. After these preprocessing steps, features extraction is accomplished by two different techniques: the maximum curvature method and the k-means clustering.

The maximum curvature method presented in [15], which is one of the better finger-vein extraction methods [13], is based on the fact that a vein appears like a dent with high curvature in the cross-sectional profile, and consequently, calculates the local maxima of each cross-sectional profile (in four directions: horizontal, vertical and two diagonal directions).

On the other hand, the k-means clustering used in this application partitions the image pixels into three clusters: background, surrounding tissues and subcutaneous veins. From these three clusters, only the subcutaneous veins cluster is considered as final solution. In this case, the centroids of clusters used to characterize the data are determined by minimizing the sum of squad errors given by:

$$J_K = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - m_k)^2 \tag{1}$$

where $(x_1, \ldots, x_n) = X$ is the data matrix, $m_k = \sum_{i \in C_k} (x_i / n_k)$ is the centroid of the cluster c_k and n_k is the number of points in C_k [36].

Once the peripheral subcutaneous veins have been detected on the SWIR image, second part of the algorithm addresses the registration of the detection results with the range data provided by the TOF camera, so that they share a common reference frame. Before registration, radial distortion of the TOF data is corrected. The relationship between the distortion coordinate system and the imaging coordinate system is given by:

$$x_d = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \, y_d = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \tag{2}$$

where (x_d, y_d) is the distorted coordinate, (x, y) is the normalized imaging coordinate, k_1, k_2, k_3 are the radial distortion coefficients of the lens, and $r^2 = x^2 + y^2$. Normalized image coordinates are calculated from pixel coordinates by translating to the optical center and dividing by the focal length in pixels. Thus, x and y are dimensionless. Radial distortion coefficients are determined with the intrinsic and extrinsic parameters of the camera during the calibration process. This calibration process is carried out once, by taking pictures of different orientations of a planar checkboard considered as a metric reference for the system. The position of all the square corners in a set of 50 pictures was analyzed in order to obtain the camera parameters.

Then, as SWIR image and TOF data are acquired with cameras that exhibit a different pixel array and a different field of view, the random sample consensus (RANSAC) algorithm [37] is adopted for registering the acquired data, in such a way that a direct correspondence between the pixels of the different images is obtained. As relative positions of SWIR and TOF camera remain fixed on the designed set-up, the RANSAC algorithm is used only once and offline, for finding the rotation and the translation (R, T) that enable the mapping of the TOF data into the reference frame of the processed SWIR image, being R a 2×2 matrix and T a 2×1 vector. For that, N pairs of control points' correspondences between frames F_1 and F_2 are selected, where F_1 and F_2 correspond to TOF and SWIR frames respectively. In this particular case, five pairs of control points were selected manually in 20 different scenes, resulting in a total of 100 pairs of control points ($N = 100$). The control points are represented by 2D coordinates $\left(X_1^i, X_2^i\right)$ in their respective reference systems. RANSAC samples the solution space of (R, T) and estimates its fitness by counting the number of inliers, f_0:

$$f_0(F_1, F_2, R, T) = \sum_i^N L\left(X_1^i, X_2^i, R, T\right) \tag{3}$$

where:

$$L\left(X_1^i, X_2^i, R, T\right) = \begin{cases} 1, & e = \|RX_1^i + T - X_2^i\| < \epsilon \\ 0, & otherwise \end{cases} \tag{4}$$

and ϵ is the threshold beneath which a features match $\left(X_1^i, X_2^i\right)$ is determined to be an inlier. RANSAC chooses the transform with the largest number of inlier matches [38,39]. The resulting (R, T) is then utilized online for matching the range data provided by the TOF camera with the processed SWIR data that contains the resulting peripheral veins detection.

3. Results

In order to validate the proposed multisensory system and the associated algorithm for the detection and localization of peripheral veins, several experimental tests have been carried out. Figures 3 and 4 show the data acquired by the multisensory system for one of the experimental tests performed to evaluate the detection of the peripheral subcutaneous veins on the front of the hand.

(**a**)   (**b**)

Figure 3. Test 1—Acquired images. (**a**) SWIR image; (**b**) TOF amplitude image.
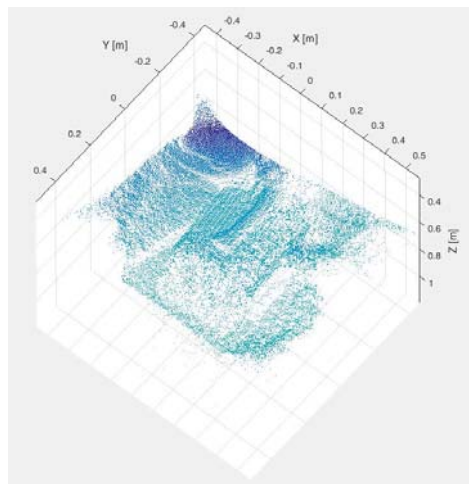
Figure 4. Test 1—Raw point cloud provided by the TOF camera.

Figure 3 presents the SWIR image and the TOF amplitude image, whereas Figure 4 displays the raw point cloud provided by the TOF camera.

The vein extraction results achieved after applying the proposed algorithms to the acquired SWIR image are presented on Figure 5. Left-hand side image displays the result from the adapted maximum curvature method, while right-hand side image shows result obtained with the method based on the k-means clustering. Red color is utilized to visualize pixels identified or classified as vein.

Once the vein extraction is completed, TOF data are registered in order to locate the veins spatially. Figure 6a shows the result from registering the original SWIR image with the TOF data (termed SWIR-D visualization, as it combines the SWIR information with the estimated depth for each pixel), whereas Figure 6b displays the 3D mapping of the veins detected with the method based on the k-means clustering after applying the proposed registration algorithm.

(**a**) (**b**)
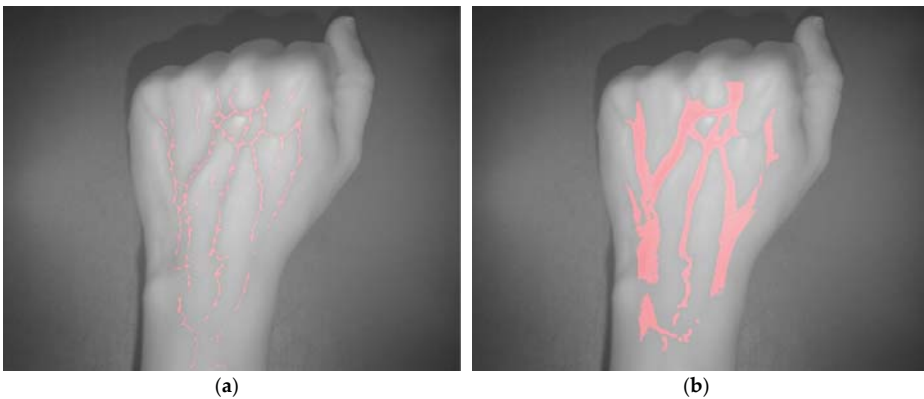
Figure 5. Test 1—Vein extraction. (**a**) Results from applying the proposed preprocessing and the adapted maximum curvature method; (**b**) Results from applying the method based on the k-means clustering.

(**a**) (**b**)
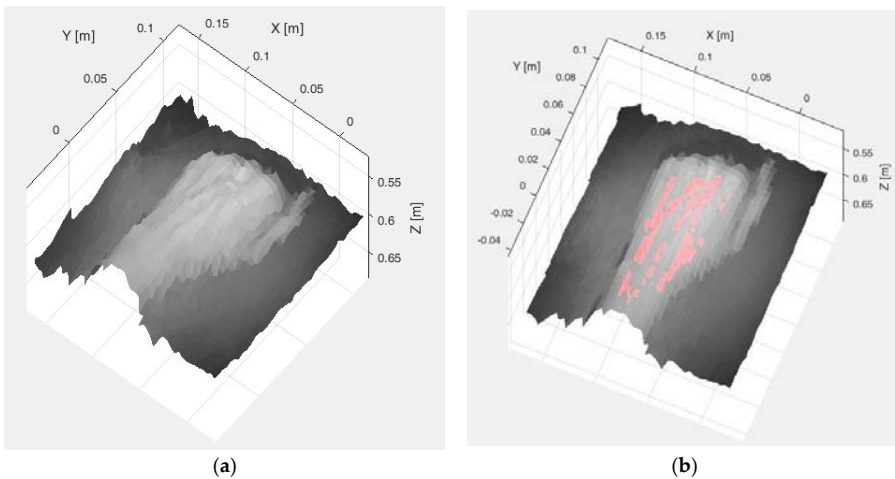
Figure 6. Test 1. (**a**) SWIR-D visualization; (**b**) Registered SWIR image with vein detection results.

Figures 7 and 8 display the dataset acquired with the proposed multisensory system for a second test scene. In this case, it is desired to evaluate the detection of the subcutaneous veins on the wrist anterior view. The dataset includes the SWIR image (Figure 7a), the TOF amplitude image (Figure 7b) and the raw point cloud provided by the TOF camera (Figure 8).

(**a**) (**b**)

Figure 7. Test 2 - Acquired images. (**a**) SWIR image; (**b**) TOF amplitude image.
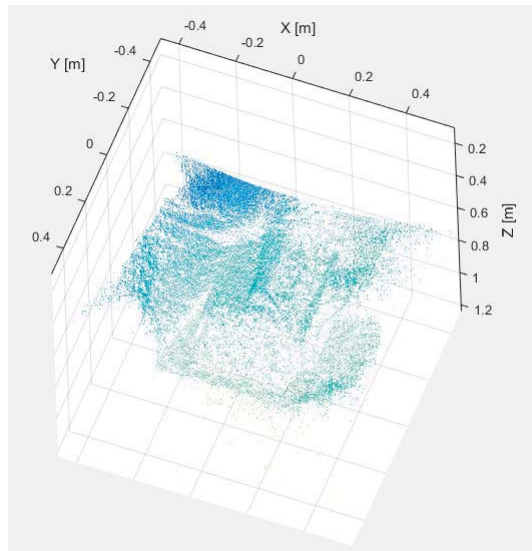
Figure 8. Test 2—Raw point cloud provided by the TOF camera.

Figure 9a,b shows the results obtained after applying the adapted maximum curvature method and the method based on the *k*-mean clustering, respectively. Next, Figure 10a displays the obtained result after registering the TOF data with the original SWIR image, and Figure 10b illustrates the SWIR-D visualization obtained after applying the proposed registration algorithm. In this SWIR-D visualization, the subcutaneous veins detected on the SWIR image with the method based on the k-means clustering are mapped onto the 3D reconstructed surface. Lastly, the third experiment is intended to evaluate the detection and localization of subcutaneous veins on the anterior view of the forearm. Figures 11 and 12 display the dataset acquired with the proposed multisensory system,

including a SWIR image (Figure 11a), a TOF amplitude image (Figure 11b) and a raw point cloud provided by the TOF camera (Figure 12).

(**a**)   (**b**)
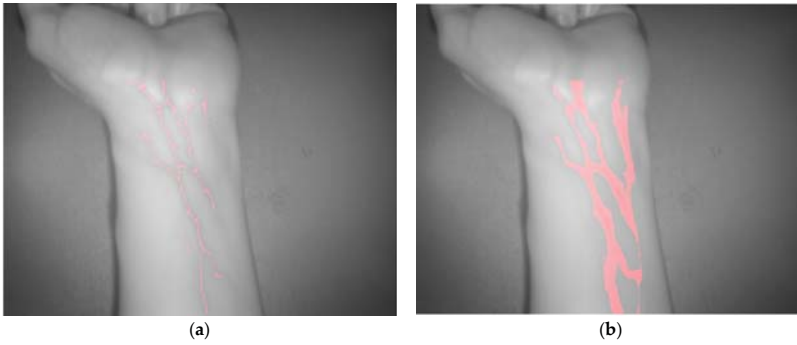
Figure 9. Test 2—Vein extraction. (**a**) Results from applying the proposed preprocessing and the adapted maximum curvature method; (**b**) Results from applying the method based on the k-means clustering.

(**a**)   (**b**)
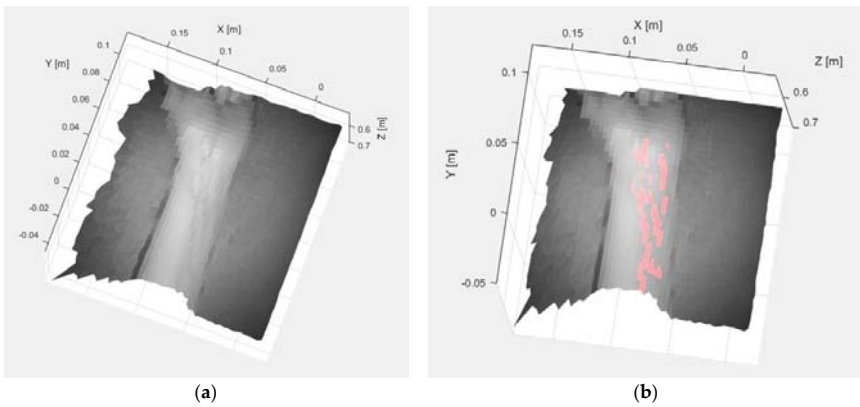
Figure 10. Test 2. (**a**) SWIR-D visualization; (**b**) Registered SWIR image with vein detection results.

(**a**)   (**b**)

Figure 11. Test 3—Acquired images. (**a**) SWIR image; (**b**) TOF amplitude image.
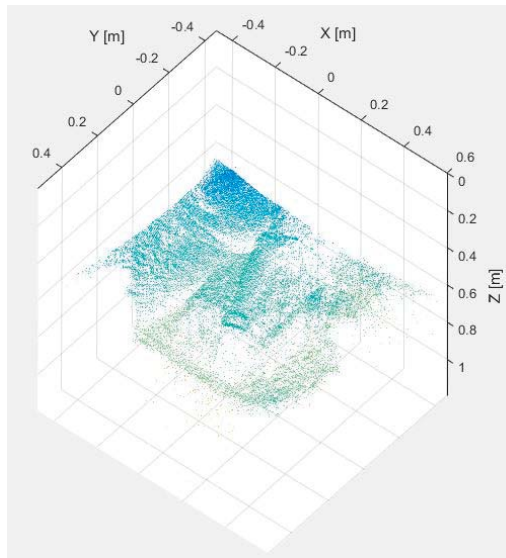
Figure 12. Test 3—Raw point cloud provided by the TOF camera.

Figure 13a,b shows the vein detection results obtained with the adapted maximum curvature method and the method based on the k-means clustering, respectively. Finally, Figure 14a displays the obtained result after registering the TOF data with the original SWIR image, while Figure 14b illustrates the mapping of the veins detected with the method based on the k-means clustering onto the 3D reconstructed surface, after applying the proposed registration algorithm. Note that in the three presented experiments, only the registration results obtained with the detection based on the k-means clustering are displayed. This is due to the fact that the method based on the adapted maximum curvature provides a low number of image pixels identified as vein, which demerits the registration results. On the contrary, the detection method based on the k-means clustering provides a higher a proportion of image pixels that are correctly identified as vein, and consequently, fits better as input for the subsequent registration procedure, improving performance results.

(**a**)   (**b**)
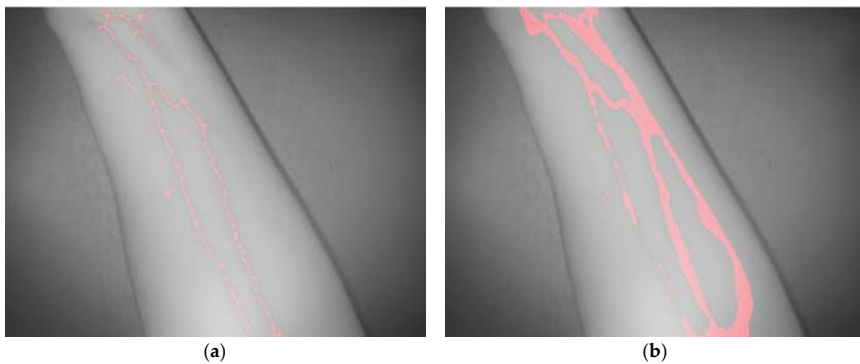
Figure 13. Test 3—Vein extraction. (**a**) Results from applying the proposed preprocessing and the adapted maximum curvature method; (**b**) Results from applying the method based on the k-means clustering.

(**a**)   (**b**)
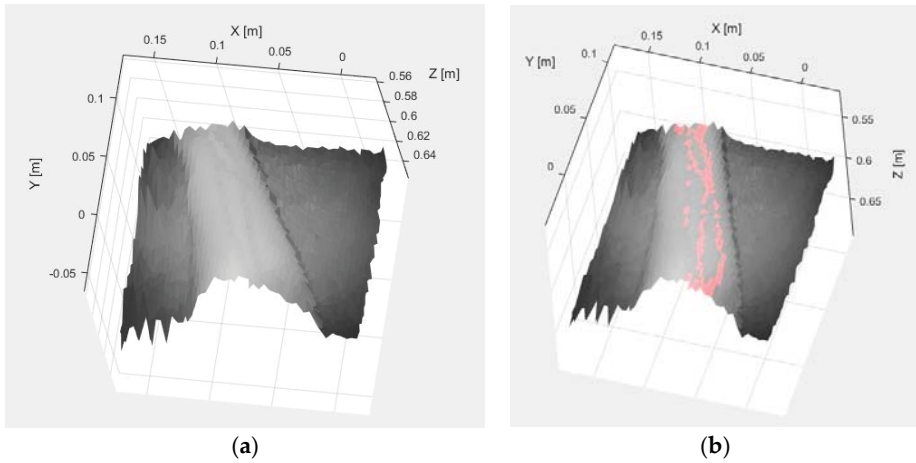
Figure 14. Test 3. (**a**) SWIR-D visualization; (**b**) Registered SWIR image with vein detection results.

In order to evaluate quantitatively the performance of the proposed approach, ground truth data was carefully collected and produced for the three test scenes presented previously. This process included the manual labelling of the pixels that visually appear to belong to the veins on the SWIR images and the manual measurement of the Cartesian coordinates of ten control points located on the test scenarios. These coordinates were measured with respect to the reference frame located at the intersection of the optical axis with the front face of the TOF camera.

Figure 15 shows the labelled images for the SWIR images presented on Figures 3a, 7a and 11a. Note that the same masks applied during the segmentation process were utilized after the manual labelling of the SWIR images, in order to get the same regions of interest used during the features extraction. These images were then utilized as ground truth data in the pixel-level comparison carried out with the detection results obtained with the proposed methods. Detection performance is then evaluated in terms of true positive rate, accuracy and total error rate. The true positive vein detection rate, which is a proportion of the pixels that are correctly identified as vein, is defined as:

$$\mathrm{TP} = \frac{\text{number of pixels correctly identified as vein}}{\text{total number of pixels identified as vein}} \cdot 100\% \tag{5}$$

Accuracy is the overall correctness of the detection algorithm and it is calculated as:

$$\mathrm{Accuracy} = \frac{\text{sum of correctly identified pixels}}{\text{total number of pixels}} \cdot 100\% \tag{6}$$

Lastly, the total error rate is given by:

$$\mathrm{Error\ rate} = \frac{\text{sum of incorrectly identified pixels}}{\text{total number of pixels}} \cdot 100\% \tag{7}$$

Performance evaluation results are gathered in Table 1.

On the other hand, after evaluating the data registered from the TOF camera with respect to the collected ground truth, we obtained that the position errors measured for the defined control points go from 2.1 to 3.5 mm in the *x*-axis, from 1.8 to 4.6 mm in the *y*-axis and from 0.1 to 9.9 mm in the *z*-axis, with a mean error of 2.5 mm in the *x*-axis, 2.9 mm in the *y*-axis and 3.1 mm in the *z*-axis. Note that the *x*-axis coincides roughly with the transverse axis of the hand or forehand, the *y*-axis

with the longitudinal axis, and the *z*-axis with the optical axis of the TOF camera. Table 2 summarizes these results.

Table 1. Detection performance evaluation.

	Test	TP	Accuracy	Error Rate
1	Method based on adapted max. curvature	21.9%	97.0%	3.0%
	Method based on k-means	86.7%	96.5%	3.5%
2	Method based on adapted max. curvature	16.0%	97.3%	2.7%
	Method based on k-means	72.3%	97.0%	3.0%
3	Method based on adapted max. curvature	22.0%	97.1%	2.9%
	Method based on k-means	80.9%	96.1%	3.9%

(a)   (b)

(c)
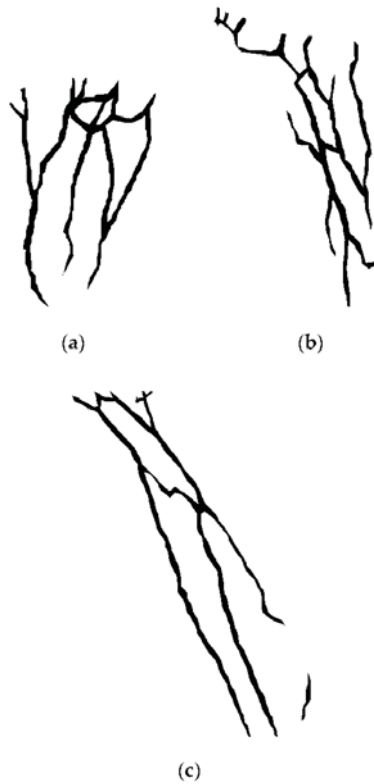
Figure 15. Labelled images. (**a**) Ground truth data for test 1; (**b**) Ground truth data for test 2; (**c**) Ground truth data for test 3.

Table 2. Position errors from the 3D registered data.

Axis	Minimum Error	Maximum Error	Mean Absolute Error
x	2.1 mm	3.5 mm	2.5 mm
y	1.8 mm	4.6 mm	2.9 mm
z	0.1 mm	9.9 mm	3.1 mm

Lastly, the response time of the proposed algorithm was been evaluated on a 2.60 GHz Intel® Core™ i5-4210M CPU with 8 GB of RAM memory. The average time for vein extraction using the adapted maximum curvature method was 1.2 s and 588 ms for the method based on the k-means clustering, whereas the average time for the registration procedure was 140 ms. That is, the total time for automatic detection and localization of peripheral subcutaneous veins with the proposed system is between 728 ms and 1.34 s.

4. Discussion

Performance evaluation results show that detection method based on the adapted maximum curvature exhibits a slightly better performance in terms of accuracy and total error rate than the method based on the k-means. This is due to the intrinsic nature of the algorithm, which provides less pixels identified as vein, but with a very high level of correctness. On the other hand, the method based on the k-means presents a much higher true positive rate detection, while keeping the accuracy and the total error rate quite close to the values provided by the adapted maximum curvature method. That means that a higher number of pixels are correctly classified as vein, and although incorrect detections also increase, they do not demerit the algorithm accuracy. With a higher number of correctly identified pixels, it is easier to determine some characteristics of veins, such as their width, which can help in selecting the best target vein for the automatic insertion of catheters. In addition, as during the mapping that takes place in the registration procedure some true positive detection points can be lost, it is better to have a high positive detection rate that guarantees proper registration of the pixels identified as vein.

Therefore, both methods attain a high level of correctness in identifying the image pixels that belong to the peripheral subcutaneous veins, but the method based on the k-means clustering fits better as input for the subsequent registration procedure, improving the overall performance result. It is also important to remark that ground truth labelling of images that were used for the performance evaluation was done manually, and this process is not 100% free from mistakes. Consequently, labelling errors can also contribute to shorten the performance values.

Regarding performance results related to the TOF data registration procedure, it is relevant to note that the TOF camera is characterized by suffering from flying pixels and noise. In addition, the registration algorithm is dealing with a correspondence between images of 144×176 pixels from the TOF camera and images of 636×508 pixels from the SWIR camera. Moreover, manual measurement of distances for ground truth data is not exempt from errors, which could explain the appearance of some isolated maximum errors, far from the mean values. Thus, the mean position errors obtained during the experimental test are quite acceptable, but should be further improved in the near future for increasing the reliability of the stated application.

On the other hand, the response time of the proposed algorithm satisfactorily fulfils the requirements for real-time applications. This coupled with its affordable cost and its reduced size and compactness makes it suitable for clinical routine use. Therefore, performance evaluation results highlight the feasibility and the potential benefits of the proposed solution.

5. Conclusions

This paper proposed a multisensory approach for the detection and localization of peripheral subcutaneous veins as a first step for achieving a guidance system that can be used in the future for automatic catheter insertion with the help of a robotic system. The solution includes a SWIR camera for acquiring reflectance measurements in the NIR region that are used for detecting the image pixels that belongs to the subcutaneous veins, a TOF camera that provides fast acquisition of Cartesian coordinates for enabling the localization of the target veins and a NIR lighting source for improving the contrast between the subcutaneous veins and the surrounding tissues.

The algorithm designed for the proposed multisensory system includes the vein features extraction from the SWIR image and the registration of the detection results with the data provided by the TOF camera, in such a way that range data can be associated to the image pixels identified as veins. For vein features extraction, two methods were proposed and evaluated experimentally, one based on the adaptation of the maximum curvature method and another based on the *k*-means clustering, both combined with several preprocessing steps for improving detection performance. Although both methods exhibit satisfactory detection accuracy, the method based on the k-means clustering fits better for the posterior registration process, given its higher TP rate.

Preliminary experimental results demonstrate the feasibility of the proposed design and highlight the potential benefits of the solution. However, future work should be directed to enhance the localization performance in order to achieve a more reliable application.

Author Contributions: The work presented here was carried out in collaboration between the authors. Roemi Fernández designed the study and wrote the manuscript. Roemi Fernández and Manuel Armada conceived and designed the experiments. Roemi Fernández designed and implemented the proposed algorithm. Roemi Fernández and Manuel Armada performed the experiments for data acquisition. Roemi Fernández processed and analyzed the data and drew the main conclusions. Manuel Armada contributed to the review of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jones, A. Dressings for the management of catheter sites: A review. *J. Assoc. Vasc. Access* **2004**, *9*, 26–33. [CrossRef]
2. Alexandrou, E.; Ray-Barruel, G.; Carr, P.J.; Frost, S.; Inwood, S.; Higgins, N.; Lin, F.; Alberto, L.; Mermel, L.; Rickard, C.M. International prevalence of the use of peripheral intravenous catheters. *J. Hosp. Med.* **2015**, *10*, 530–533. [CrossRef] [PubMed]
3. O'Grady, N.P.; Alexander, M.; Burns, L.A.; Dellinger, E.P.; Garland, J.; Heard, S.O.; Lipsett, P.A.; Masur, H.; Mermel, L.A.; Pearson, M.L.; et al. *Guidelines for the Prevention of Intravascular Catheter-Related Infections*; Center for Disease Control and Prevention: Atlanta, GA, USA, 2011.
4. Bernatchez, S.F. Care of peripheral venous catheter sites: Advantages of transparent film dressings over tape and gauze. *J. Assoc. Vasc. Access* **2014**, *19*, 256–261. [CrossRef]
5. Pujol, M.; Hornero, A.; Saballs, M.; Argerich, M.J.; Verdaguer, R.; Cisnal, M.; Peña, C.; Ariza, J.; Gudiol, F. Clinical epidemiology and outcomes of peripheral venous catheter-related bloodstream infections at a university-affiliated hospital. *J. Hosp. Infect.* **2007**, *67*, 22–29. [CrossRef] [PubMed]
6. Richardson, D.; Kaufman, L. Reducing blood exposure risks and costs associated with spivc insertion. *Nurs. Manag.* **2011**, *42*, 31–34. [CrossRef] [PubMed]
7. Strauss, K.W.; Onia, R.; Van Zundert, A.A.J. Peripheral intravenous catheter use in Europe: Towards the use of safety devices. *Acta Anaesthesiol. Scand.* **2008**, *52*, 798–804. [CrossRef] [PubMed]
8. Díaz, C.E.; Fernández, R.; Armada, M.; García, F. A research review on clinical needs, technical requirements, and normativity in the design of surgical robots. *Int. J. Med. Robot. Comput. Assist. Surg.* **2017**. [CrossRef] [PubMed]
9. Díaz, C.E.; Fernández, R.; Armada, M.; García Gutiérrez, F.D.J. State of the art in robots used in minimally invasive surgeries. Natural orifice transluminal surgery (NOTES) as a particular case. *Ind. Robot Int. J.* **2015**, *42*, 508–532. [CrossRef]
10. Saltarén Pazmiño, R.J.; Aracil Santonja, R.; Puglisi, L.J.; Poletti, G.A.; Rey Portoles, G.; Cabrera Parra, W. Robot de estructura cinemática híbrida para el guiado de la inserción de agujas, catéteres y elementos quirúrgicos para procedimientos de cirugía mínimamente invasiva. Spanish Patent ES-2392059 B2, 20 December 2011.

11. Paquit, V.; Price, J.R.; Seulin, R.; Mériaudeau, F.; Farahi, R.H.; Tobin, K.W.; Ferrell, T.L. Near-infrared imaging and structured light ranging for automatic catheter insertion. In *Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display 2006*; Cleary, K.R., Robert, L., Galloway, J., Eds.; SPIE: San Diego, CA, USA, 2006; Volume 6141.

12. Paquit, V.; Price, J.R.; Mériaudeau, F.; Tobin, K.W.; Ferrell, T.L. Combining near-infrared illuminants to optimize venous imaging. In *Medical Imaging 2007: Visualization and Image-Guided Procedures*; Cleary, K.R., Miga, M.I., Eds.; SPIE: San Diego, CA, USA, 2007; Volume 6509.

13. Ton, B.T.; Veldhuis, R.N.J. A high quality finger vascular pattern dataset collected using a custom designed capturing device. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–5.

14. Miura, N.; Nagasaka, A.; Miyatake, T. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach. Vis. Appl.* **2004**, *15*, 194–203. [CrossRef]

15. Miura, N.; Nagasaka, A.; Miyatake, T. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE Trans. Inf. Syst.* **2007**, *E90-D*, 1185–1194. [CrossRef]

16. Song, W.; Kim, T.; Kim, H.C.; Choi, J.H.; Kong, H.-J.; Lee, S.-R. A finger-vein verification system using mean curvature. *Pattern Recognit. Lett.* **2011**, *32*, 1541–1547. [CrossRef]

17. Choi, J.H.; Song, W.; Kim, T.; Lee, S.-R.; Kim, H.C. Finger vein extraction using gradient normalization and principal curvature. *Proc. SPIE* **2009**, *7251*, 725111.

18. Wu, J.-D.; Liu, C.-T. Finger-vein pattern identification using svm and neural network technique. *Expert Syst. Appl.* **2011**, *38*, 14284–14289. [CrossRef]

19. Shin, K.; Park, Y.; Nguyen, D.; Park, K. Finger-vein image enhancement using a fuzzy-based fusion method with gabor and retinex filtering. *Sensors* **2014**, *14*, 3095–3129. [CrossRef] [PubMed]

20. Pham, T.; Park, Y.; Nguyen, D.; Kwon, S.; Park, K. Nonintrusive finger-vein recognition system using NIR image sensor and accuracy analyses according to various factors. *Sensors* **2015**, *15*, 16866–16894. [CrossRef] [PubMed]

21. Deepa, P.; Mohanavelu, K.; Sundersheshu, B.S.; Padaki, V.C. Vein identification and localization for automated intravenous drug delivery system. In *Wireless Networks and Computational Intelligence, Proceedings of the 6th International Conference on Information Processing (ICIP 2012), Bangalore, India, 10–12 August 2012*; Venugopal, K.R., Patnaik, L.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 270–281.

22. Wang, F.; Behrooz, A.; Morris, M.; Adibi, A. High-contrast subcutaneous vein detection and localization using multispectral imaging. *J. Biomed. Opt.* **2013**, *18*, 050504. [CrossRef] [PubMed]

23. Juric, S.; Flis, V.; Debevc, M.; Holzinger, A.; Zalik, B. Towards a low-cost mobile subcutaneous vein detection solution using near-infrared spectroscopy. *Sci. World J.* **2014**, *2014*, 365902. [CrossRef] [PubMed]

24. Phipps, K.; Modic, A.; O'Riordan, M.A.; Walsh, M. A randomized trial of the vein viewer versus standard technique for placement of peripherally inserted central catheters (PICCs) in neonates. *J. Perinatol.* **2012**, *32*, 498–501. [CrossRef] [PubMed]

25. Kaddoum, R.N.; Anghelescu, D.L.; Parish, M.E.; Wright, B.B.; Trujillo, L.; Wu, J.; Wu, Y.; Burgoyne, L.L. A randomized controlled trial comparing the accuvein AV300 device to standard insertion technique for intravenous cannulation of anesthetized children. *Pediatr. Anesth.* **2012**, *22*, 884–889. [CrossRef] [PubMed]

26. Cuper, N.J.; Klaessens, J.H.G.; Jaspers, J.E.N.; de Roode, R.; Noordmans, H.J.; de Graaff, J.C.; Verdaasdonk, R.M. The use of near-infrared light for safe and effective visualization of subsurface blood vessels to facilitate blood withdrawal in children. *Med. Eng. Phys.* **2013**, *35*, 433–440. [CrossRef] [PubMed]

27. Meriaudeau, F.; Paquit, V.; Walter, N.; Price, J.; Tobin, K. 3D and multispectral imaging for subcutaneous veins detection. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 2857–2860.

28. Ai, D.; Yang, J.; Fan, J.; Zhao, Y.; Song, X.; Shen, J.; Shao, L.; Wang, Y. Augmented reality based real-time subcutaneous vein imaging system. *Biomed. Opt. Express* **2016**, *7*, 2565–2585. [CrossRef] [PubMed]

29. Anderson, R.R.; Parrish, J.A. The optics of human skin. *J. Investig. Dermatol.* **1981**, *77*, 13–19. [CrossRef] [PubMed]

30. Tuchin, V.V.; Utz, S.R.; Yaroslavsky, I.V. Tissue optics, light distribution, and spectroscopy. *Opt. Eng.* **1994**, *33*, 3178–3188. [CrossRef]

31. Sfareni, R.; Boffi, A.; Quaresima, V.; Ferrari, M. Near infrared absorption spectra of human deoxy- and oxyhaemoglobin in the temperature range 20–40 °C. *Biochim. Biophys. Acta BBA Protein Struct. Mol. Enzymol.* **1997**, *1340*, 165–169. [CrossRef]

32. Wray, S.; Cope, M.; Delpy, D.T.; Wyatt, J.S.; Reynolds, E.O.R. Characterization of the near infrared absorption spectra of cytochrome AA3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochim. Biophys. Acta BBA Bioenerg.* **1988**, *933*, 184–192. [CrossRef]

33. Chiabrando, F.; Chiabrando, R.; Piatti, D.; Rinaudo, F. Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. *Sensors* **2009**, *9*, 10080–10096. [CrossRef] [PubMed]

34. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

35. Zuiderveld, K. Contrast limited adaptive histogram equalization. In *Graphics Gems IV*; Heckbert, P.S., Ed.; Academic Press Professional, Inc.: San Diego, CA, USA, 1994; pp. 474–485.

36. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [CrossRef]

37. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

38. Du, H.; Henry, P.; Ren, A.X.; Cheng, A.M.; Goldman, A.D.B.; Seitz, A.S.M.; Fox, D. Interactive 3D modeling of indoor environments with a consumer depth camera. In Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp'11), Beijing, China, 17–21 September 2011; ACM: New York, NY, USA, 2011; pp. 75–84.

39. Fernández, R.; Salinas, C.; Montes, H.; Sarria, J. Multisensory system for fruit harvesting robots. Experimental testing in natural scenarios and with different kinds of crops. *Sensors* **2014**, *14*, 23885–23904. [CrossRef] [PubMed]

Robot-Beacon Distributed Range-Only SLAM for Resource-Constrained Operation

Arturo Torres-González, Jose Ramiro Martínez-de Dios * and Anibal Ollero

Robotics Vision and Control Group, University of Sevilla, Escuela Superior de Ingenieros,
c/Camino de los Descubrimientos s/n, 41092 Seville, Spain; arturotorres@us.es (A.T.-G.); aollero@us.es (A.O.)
* Correspondence: jdedios@us.es; Tel.: +34-954-487-357; Fax: +34-954-487-340

Abstract: This work deals with robot-sensor network cooperation where sensor nodes (beacons) are used as landmarks for Range-Only (RO) Simultaneous Localization and Mapping (SLAM). Most existing RO-SLAM techniques consider beacons as passive devices disregarding the sensing, computational and communication capabilities with which they are actually endowed. SLAM is a resource-demanding task. Besides the technological constraints of the robot and beacons, many applications impose further resource consumption limitations. This paper presents a scalable distributed RO-SLAM scheme for resource-constrained operation. It is capable of exploiting robot-beacon cooperation in order to improve SLAM accuracy while meeting a given resource consumption bound expressed as the maximum number of measurements that are integrated in SLAM per iteration. The proposed scheme combines a Sparse Extended Information Filter (SEIF) SLAM method, in which each beacon gathers and integrates robot-beacon and inter-beacon measurements, and a distributed information-driven measurement allocation tool that dynamically selects the measurements that are integrated in SLAM, balancing uncertainty improvement and resource consumption. The scheme adopts a robot-beacon distributed approach in which each beacon participates in the selection, gathering and integration in SLAM of robot-beacon and inter-beacon measurements, resulting in significant estimation accuracies, resource-consumption efficiency and scalability. It has been integrated in an octorotor Unmanned Aerial System (UAS) and evaluated in 3D SLAM outdoor experiments. The experimental results obtained show its performance and robustness and evidence its advantages over existing methods.

Keywords: robot-sensor network cooperation; SLAM; sensor networks

1. Introduction

This paper deals with Range-Only (RO) Simultaneous Localization and Mapping (SLAM) in which the robot uses range measurements to build a map of an unknown environment and to self-localize in that map. RO-SLAM does not require line-of-sight operation and is independent of lighting conditions, whereas visual SLAM is sensitive to them and is not suitable in the presence of dense dust or smoke [1]. Using radio beacons as landmarks naturally solves the data association problem, typical of visual SLAM. Besides, RO-SLAM operates with shorter measurement and state vectors than visual SLAM, resulting in significantly lower computational burden.

Our envisioned application is aerial robot autonomous navigation for maintenance and repairing in industrial plants. We are interested in RO-SLAM schemes where the robot uses nodes from a sensor network as landmarks. Consider a GPS-denied scenario where a large number of sensor nodes (beacons) have been deployed at unknown static locations. For instance, they have been placed at random locations for real-time monitoring an accident, or they are used for monitoring an industrial facility, and their exact location was not registered during deployment. This is not a

constraint: sensor nodes are already used for monitoring and maintenance in many industrial plants. Each node periodically gathers and filters measurements and transmits them to a monitoring station. We assume that each node can measure the distance to other nodes. This is not a limitation, in fact most Commercial Off-The-Shelf (COTS) nodes can measure the Radio Signal Strength of Incoming packets (RSSI) and estimate the range to the emitting node [2]. Mapping and robot localization is essential for UAS navigation in the above GPS-denied environments. Besides, knowing the location of beacons enables robot-sensor network cooperative missions of interest in these scenarios such as sensor node transportation and deployment [3,4], collection of data from sensors [5,6] or node remote powering [7]. The potential capabilities of robot-sensor network collaboration have recently originated significant interest in RO-SLAM methods that use sensor nodes as beacons [8–11]. However, most existing techniques consider beacons as passive landmarks disregarding the computational, communication and sensing capabilities with which they are actually endowed.

The performance of RO-SLAM methods improves as the number of measurements integrated in SLAM increases. For instance, integrating beacon-beacon (inter-beacon) measurements improves map and robot estimation accuracy [12]. However, using more measurements requires spending more resources such as the energy, bandwidth and computational power required to gather, transmit and integrate the measurements. SLAM is a resource demanding task. In the above scenarios the robot and beacons perform many other tasks apart from SLAM, all of them sharing the available resources. Besides, the available resources can be, and usually are, constrained by technological or application limitations such as the available energy and computational capacity of beacons or the bandwidth of the sensing channel. Resource consumption reduction has attracted high interest in SLAM, but very few resource-constrained RO-SLAM methods have been reported.

This paper presents a RO-SLAM scheme that is capable of exploiting robot-beacon cooperation in order to improve SLAM accuracy and efficiency while meeting a given resource consumption bound. In this work, the resource consumption bound is expressed in terms of the maximum number of measurements that can be integrated in SLAM per iteration. The sensing channel capacity used, the beacon energy consumed or the computational capacity employed, among others, are proportional to the number of measurements that are gathered and integrated in SLAM. The proposed scheme can meet static and dynamic bounds, e.g., determined by an online resource allocation tool, enabling high flexibility, which can be of interest in many cases. Our scheme employs a distributed Sparse Extended Information Filter (SEIF) SLAM method, in which each beacon gathers and integrates in the SLAM update stage robot-beacon and inter-beacon measurements. Like SEIF algorithm, our scheme can be executed in constant-time. Our scheme also comprises a distributed tool that uses a gain-cost analysis to dynamically select the most informative measurements to be integrated in SLAM.

The proposed scheme has a robot-beacon distributed approach were beacons actively participate in measurement selection, gathering and integration in SLAM. It has the following main properties: (1) adaptive resource-constrained operation, since it dynamically adapts to satisfy the given resource consumption bound; (2) accuracy, since it integrates inter-beacon measurements, significantly improving map and robot localization accuracies and speeding up beacon initialization; (3) efficiency, since it gathers and integrates the most informative measurements and; (4) scalability, since all the involved tasks are executed in a distributed manner. This paper presents the scheme, evaluates and compares its performance with existing methods in experiments performed with an octorotor UAS.

The proposed scheme employs the distributed SEIF SLAM that we sketched in [13]. In this paper, it is enhanced and combined with a distributed measurement selection tool that enables resource-constrained operation. The main improvements in this paper over [13] are:

- development of a distributed robot-beacon tool that selects the most informative measurements that are integrated in SLAM fulfilling the resource consumption bound;
- harmonious integration of the distributed SEIF SLAM and the measurement selection tool. The resulting scheme outperforms that in [13], as described in Section 6;
- extension to 3D SLAM, integration and experimentation of the scheme with an octorotor UAS;

- new experimental performance evaluation and comparison with existing methods;
- new subsection with experimental robustness evaluation;
- extension and more detailed related work. Furthermore, the paper has been restructured and all sections have been completed and rewritten for clarity.

The paper is organized as follows. Section 2 summarizes the related work. Section 3 presents the problem formulation. Sections 4 and 5 briefly describe the operation of the robot and beacons. Section 6 presents its evaluation, comparison and robustness analysis in 3D SLAM experiments. The conclusions are in Section 7.

2. Related Work

As it is known, in Simultaneous Localization and Mapping (SLAM) a robot builds a map of an unknown environment while simultaneously keeping track of its location within that map. RO-SLAM methods solve the SLAM problem using only range measurements. Recursive Bayesian Filters (RBFs) provide a solid statistical basis for estimation integrating measurements in presence of noise. A good number of RO-SLAM methods have been developed in the last years. Most of them employ approaches based on the Extended Kalman Filter (EKF) SLAM or on Fast-SLAM. These filters have been combined with different initialization tools in order to deal with the partial observability of range measurements. Particle Filters (PFs) [14] involve significant computational burden but provide accurate results. Probability grids [12,15] have also been used but are computationally inefficient to implement in 3D. Trilateration [9] is the simplest approach but it can lead to high initialization errors, which can originate inconsistent estimations. Gaussian mixtures [16,17] provide a multi-hypothesis representation that allows integrating measurements from the beginning in undelayed SLAM methods. Dual to Kalman Filters, Information Filters represent the state by its information vector ζ and its information matrix Ω. The Sparse Extended Information Filter (SEIF) [18] for online SLAM maintains a sparse representation of Ω. SEIF is naturally efficient in the update stage and the sparseness of Ω enables using efficient algorithms for motion update and state recovery. SEIF can be executed in constant time regardless of the map size.

Most RO-SLAM techniques employ only robot-beacon range measurements. Work [12] was the first in using also measurements between beacons (inter-beacon). Integrating inter-beacon measurements improves map and robot estimation accuracy and speeds up beacon initialization. Despite its advantages few methods employ them. Integrating more measurements involves higher consumption of energy, bandwidth and processing time to gather, transmit and integrate them in SLAM. Work [19] integrates inter-beacon measurements adopting decentralized methods to deal with the increase in the consumption of resources.

Efficiency in the use of resources has received high interest in SLAM. Some methods use decision making tools based on information metrics to balance efficiency and accuracy. Work [20] developed an approach that selects highly informative loop-closure links to simplify the Pose SLAM information form and reduce its computational cost. They also achieve a compact information matrix reducing the number of measurements by operating in open loop for long periods. This open loop operation requires good odometry estimations, which are not always available in all scenarios and applications. Work [21] presented a method for estimating the quality of visual measurements in single-camera SLAM maximizing the mutual information between measurements and states in order to select the best camera measurements. Despite its potential advantages, very few SLAM techniques, and even less RO-SLAM, for resource-constrained operation have been reported.

In work [22] we proposed a RO-SLAM scheme that improves efficiency by choosing between two measurement gathering modes. Mode selection is performed using a simple centralized tool based on heuristics of robot and map estimation uncertainty: It does not consider resource consumption requirements, hence it cannot adapt to given resource consumption bounds. Besides, mode decision is fully centralized at the robot, lacking scalability. The scheme proposed in this paper is capable of satisfying a given resource consumption bound. It uses a distributed measurement selection tool that

prioritizes measurements analyzing cost and expected uncertainty improvement. Besides, the scheme is distributed and beacons actively participate in measurement selection, gathering and integration in SLAM.

2.1. Range Only SEIF SLAM in a Nutshell

This section briefly summarizes the SEIF SLAM algorithm as an introduction to the proposed scheme. Most expressions have been omitted. For notation simplicity, time subindex t has also been omitted. Refer to [23] for further details.

RO-SLAM methods commonly adopt a state vector $\vec{x}$ comprised of the robot position (x_r) and the location of all the beacons in the map ($x_1, x_2, \ldots, x_N$). SEIF is based on Extended Information Filter (EIF). Duals to Kalman Filters, Information Filters (IFs) employ the canonical representation of multivariate Gaussian distributions. The standard representation is based on the mean vector μ and the covariance matrix Σ, whereas the canonical representation is based on the information vector $\xi = \Sigma^{-1}\mu$ and the information matrix $\Omega = \Sigma^{-1}$.

$$\Omega = \begin{pmatrix} \Omega_{x_r,x_r} & \Omega_{x_r,x_1} & \cdots & \Omega_{x_r,x_N} \\ \Omega_{x_1,x_r} & \Omega_{x_1,x_1} & \cdots & \Omega_{x_1,x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{x_N,x_r} & \Omega_{x_N,x_1} & \cdots & \Omega_{x_N,x_N} \end{pmatrix} \tag{1}$$

The information matrix Ω is symmetric and positive-semidefinite. Each off-diagonal entry of Ω, called the link [18], represents the relation between two elements in $\vec{x}$. At any time in the SEIF SLAM operation some of the off-diagonal elements of Ω are zero meaning lack of information between the involved elements; some of them have high values, called strong links, meaning high amount of information; and a number of them have values close to zero, called weak links. Weak links have lower impact on the estimation of $\vec{x}$ than strong links but both involve similar computational burden. SEIF operates as EIF but maintains a sparse representation of Ω by keeping the number of active beacons bounded by a threshold. At each step the discovered beacons are classified in active and passive. Active beacons are those with non-zero links to the robot. Every time the number of active beacons is above the bound, the sparsification step is performed deactivating the beacons with the weakest links.

EIF measurement update modifies only the entries of Ω corresponding to the elements involved in the measurement. Factorizing Ω allows efficient update stages regardless of the map size. SEIF inherits this efficiency. Furthermore, by bounding the number of active beacons, SEIF significantly reduces the computational burden in the prediction stage. For linearizing the prediction and observation models it is necessary to recover the state estimate μ from the predicted $\bar{\Omega}$ and $\bar{\xi}$. SEIF recovers only the useful part of $\vec{x}$, robot and active beacons. Of course, enforcing sparseness in Ω involves an approximation error in the estimations obtained by SEIF. Work [18] suggests using sparsification bounds in the range [4,5,7–11] to balance between accuracy degradation and burden reductions.

2.2. Integration of Range Measurements

Range measurements have the problem of partial observability. To cope this it SEIF should be combined with an auxiliary beacon initialization mechanism. Our scheme uses PFs for beacon initialization due to their high flexibility. Each beacon can be in the "initialization phase" or in the "state vector phase". In the "initialization phase" each beacon executes its own PF. When the PF converges, the beacon initializes its state vector: The "state vector phase" starts.

The observation model used for range measurement $z_{r,i}$ taken by the robot to beacon b_i is:

$$h_{r,i}(\mu) = \sqrt{\delta_x^2 + \delta_y^2 + \delta_z^2}, \tag{2}$$

where $\delta_x = \mu_x - \mu_x^i$, $\delta_y = \mu_y - \mu_y^i$ and $\delta_z = \mu_z - \mu_z^i$. ($\mu_x, \mu_y, \mu_z$) and ($\mu_x^i, \mu_y^i, \mu_z^i$) are respectively the estimated positions of the robot and of beacon b_i. $h_{r,i}$ is nonlinear. Its Jacobian is:

$$H_{r,i} = \frac{\partial h_{r,i}}{\partial \mu} = \begin{bmatrix} \frac{\delta_x}{h_{r,i}} & \frac{\delta_y}{h_{r,i}} & \frac{\delta_z}{h_{r,i}} & \cdots & \frac{-\delta_x}{h_{r,i}} & \frac{-\delta_y}{h_{r,i}} & \frac{-\delta_z}{h_{r,i}} & \cdots \end{bmatrix} \tag{3}$$

All the entries of $H_{r,i}$ that are not shown in (3) are zero. Only the entries of $H_{r,i}$ corresponding to the robot and beacon b_i are not zero. The proposed method also integrates inter-beacon measurements, such as $z_{i,j}$ gathered by beacon b_i to b_j. The adopted model $h_{i,j}$ is an extension of that in (2) taking $\delta_x = \mu_x^i - \mu_x^j$, $\delta_y = \mu_y^i - \mu_y^j$ and $\delta_z = \mu_z^i - \mu_z^j$. The entries of its Jacobian $H_{i,j}$ are zero except for those corresponding to the beacons involved in the measurement:

$$H_{i,j} = \begin{bmatrix} 0 & \cdots & \frac{\delta_x}{h_{i,j}} & \frac{\delta_y}{h_{i,j}} & \frac{\delta_z}{h_{i,j}} & \cdots & \frac{-\delta_x}{h_{i,j}} & \frac{-\delta_y}{h_{i,j}} & \frac{-\delta_z}{h_{i,j}} & \cdots \end{bmatrix} \tag{4}$$

3. Problem Formulation

Consider a GPS-denied scenario where a number of sensor nodes endowed with sensing, computational and communication capabilities have been deployed at static locations. We assume that the location of the nodes is not known. For instance, they have been deployed at unknown locations or their exact location has not been registered during deployment. Each node is assumed equipped with a range sensor and can measure the distance to the robot or to other nodes within its sensing region. Range measurements are assumed affected by statistically independent Gaussian noise. We are interested in RO-SLAM techniques that use sensor nodes as radio beacons (landmarks), to online estimate the locations of sensor nodes and of the robot. The SLAM method can exploit the capabilities of the beacons deployed but, at the same time, should make efficient use of the available resources and should comply with resource consumption constraints.

In conventional RO-SLAM techniques the robot gathers range measurements to the beacons within its sensing region and integrates these measurements in the update stage, see Figure 1a. In the adopted scheme measurement gathering and integration in SLAM is distributed among the beacons. The robot computes the SEIF SLAM prediction stage and transmits the predicted state to the beacons within its sensing region. Each beacon receiving the message: (1) gathers range measurements to the robot and other beacons (inter-beacon measurements), see Figure 1b; (2) integrates these measurements and computes its contribution to the update stage and; (3) transmits its contribution to the robot. Finally, the update stage of SLAM expressed in the information form is additive [23] and the robot reconstructs the updated state by simply adding the contributions it received. The proposed scheme naturally exploits robot-beacon cooperation to solve online SLAM: Accurately, since it integrates inter-beacon measurements; efficiently, distributing measurement gathering and integration in SLAM, and hence sharing burden and energy consumption; and in a scalable manner.

Efficiency in the use of resources is very important in robot-sensor network cooperation. In most cases the SLAM algorithm is executed simultaneously with other tasks, all of them sharing the available resources. Furthermore, radio beacons gathering range measurements such as RSSI, Time of Arrival (ToA) or Differential ToA, make use of some kind of communication, which requires using a channel with a certain (constrained) capacity. In fact the capacity of the sensing channel is one of the most relevant constraints in settings with a high number and density of deployed beacons. In our problem resource consumption can be expressed in terms of the maximum number of measurements that are gathered and integrated in SLAM at each iteration. The use of the sensing channel, the consumption of beacons energy and of beacons computational capacity are proportional to the number of measurements that are gathered and integrated in SLAM. Hence, bounding the number of measurements constrains the consumption of the main resources involved.
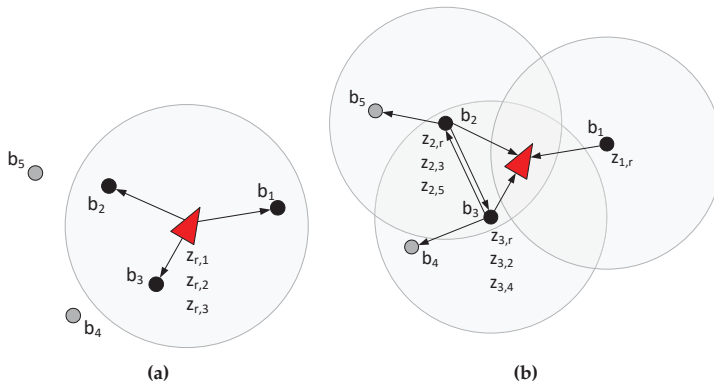
Figure 1. Measurement gathering in conventional SLAM methods (**a**) and in the proposed distributed SLAM scheme (**b**). Gray circles represent the sensing regions of the robot (**a**) and beacons b_1, b_2 and b_3 (**b**).

A diagram of the proposed scheme with its main modules is shown in Figure 2. The scheme combines (1) a distributed SEIF SLAM method in which beacons gather and integrate in SLAM robot-beacon and inter-beacon measurements; and (2) a distributed information-driven measurement selection tool that dynamically selects the measurements that are integrated in SLAM in order to improve performance while fulfilling the bound in the total number of measurements. Both components are executed in a distributed manner by the robot and by the beacons. Each beacon maintains a local version of the SLAM state whereas the global state is maintained only by the robot. The message interchange and the operation of the robot and beacons is summarized in Figure 3.

Methods that select the measurements that best reduce the uncertainty in the SLAM global state are necessarily centralized and have to deal with the information matrix of the global state. These methods incur in high computational burden with large maps and scale badly with the map size. The proposed scheme approximates this centralized measurement selection by a robot-beacon distributed tool that preserves the constant time execution and scalability. In Section 6.2 it is experimentally shown that the adopted tool is almost as accurate as the centralized measurement selection, adifference in map and robot RMS errors lower than 3%.

The distributed measurement selection tool is performed in two steps: *Measurement distribution* and *measurement allocation*. In *measurement distribution* the robot dynamically decides the number of measurements that are assigned to each beacon within the robot sensing region using expectations of uncertainty reductions. In *measurement allocation*, each beacon b_i decides the actual measurements that it will gather and integrate in SLAM analyzing the cost and expected uncertainty improvement of integrating each measurement.

The bound in the number of measurements that can be gathered in each iteration is NM_{max}. As example where NM_{max} has a high value is shown in Figure 1b: Each beacon within the robot sensing region gathers and integrates a measurement to each beacon within its sensing region. An example with a low NM_{max} in shown in Figure 4. b_1 and b_2 are assigned with only one measurement and gather $z_{1,r}$ and $z_{2,r}$. b_3 is assigned with two measurements and besides $z_{3,r}$, it gathers $z_{3,2}$, the measurement that achieves the best expected uncertainty improvement-cost trade-off.
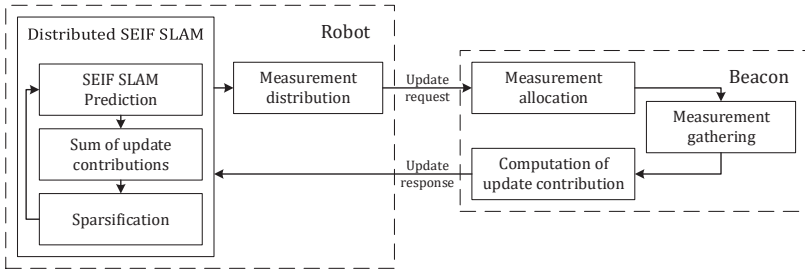
Figure 2. Diagram of the proposed resource-constrained RO-SLAM scheme.
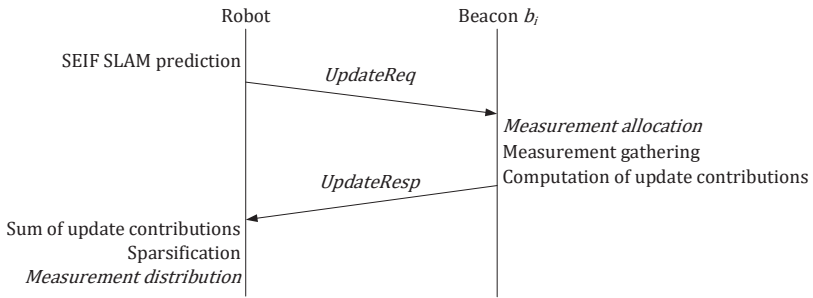
Figure 3. Operation and message interchange in the proposed scheme.
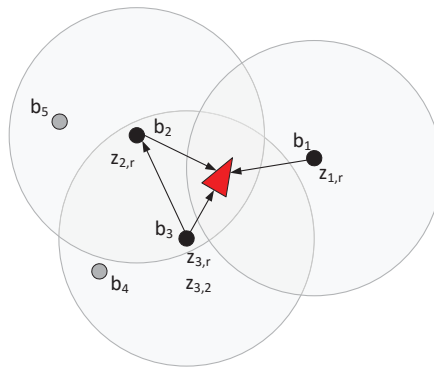
Figure 4. Measurement gathering with a low value of NM_{max}; b_1 gathers $z_{1,r}$; b_2 gathers $z_{2,r}$; b_3 gathers $z_{3,r}$ and $z_{3,2}$.

4. Operation of the Robot

The operation of the robot can be decomposed in four main tasks: (1) computation of the SEIF SLAM prediction stage; (2) reconstruction of the updated state using the contributions received by the robot; (3) computation of the sparsification step; and (4) *measurement distribution*. For brevity, most SEIF equations have been omitted. Refer to [23] for further details.

The robot operation is as follows, see Algorithm 1. First, the robot computes the SEIF SLAM prediction. Static beacons and nonlinear robot kinematic model are assumed. The robot

Jacobian is computed at each time, which requires recovering the state. Our scheme uses the efficient algorithm described in [18] for motion update and state recovery. This algorithm computes the predicted information vector $\bar{\xi}_t$ and information matrix $\bar{\Omega}_t$ and recovers the predicted μ_t.

Algorithm 1: Summary of the operation of the robot.

Require: $\xi_{t-1}, \Omega_{t-1}, NM_{max}, LM$
1: SEIF motion update and state recovery
2: Create and broadcast *UpdateReq* message
3: Receive *UpdateResp* messages
4: Extract ξ_i, Ω_i and ui_i from *UpdateResp* messages
5: Compute ξ_t and Ω_t as in (6)–(7)
6: SEIF Sparsification
7: *Measurement distribution.* Create *LM*
8: **return** ξ_t, Ω_t, LM

As the robot moves beacons go in and out of the robot sensing region. The robot maintains BS_r, a list with the beacons that are currently within its sensing region. At each time t the robot broadcasts an *UpdateReq* message that includes μ_t and *LM*, a list created by the robot at $t-1$ that contains the number of measurements that have been assigned to each beacon $b_i \in BS_r$. Transmitting the whole μ_t in the *UpdateReq* message is not suitable in cases with large maps. Only the elements in μ_t required for each beacon are transmitted. Let ev_i be the vector with the estimates required by beacon b_i to compute its contribution to the update stage:

$$ev_i = [\mu_r \quad \mu_i \quad \mu_j]^T, \tag{5}$$

where μ_r is the estimation of the robot current location, μ_i is the estimation of the location of beacon b_i and μ_j represents the estimations of the location of every beacon b_j within the sensing region of b_i.

When b_i receives the *UpdateReq* message, it performs as described in Section 5 and transmits to the robot an *UpdateResp* message with $\xi_{i,t}$ and $\Omega_{i,t}$, its contribution to the SLAM update stage. The robot reconstructs the updated state, ξ_t and Ω_t, using $\bar{\xi}_t$ and $\bar{\Omega}_t$ and the update contributions it received:

$$\xi_t = \bar{\xi}_t + \sum_i F_i^T \xi_{i,t}, \tag{6}$$

$$\Omega_t = \bar{\Omega}_t + \sum_i F_i^T \Omega_{i,t} F_i, \tag{7}$$

where F_i is the projection matrix that implements the operations necessary to allocate $\xi_{i,t}$ and $\Omega_{i,t}$ at the suitable entries in ξ_t and Ω_t.

Next, the robot checks if Ω_t satisfies the SEIF sparsification bound. If not, the beacons with the weakest links are deactivated as described in [18].

The final step performed by the robot is to distribute the number of measurements NM_{max} between $b_i \in BS_r$. NM_{max} is considered an input to our scheme. It can be static or dynamic, computed by an online resource allocation tool, for instance analyzing the capacity of the channel using link quality estimators, see e.g., [24]. Measurement distribution is performed proportionally to $IG_{i,t}$, the usefulness of the measurements from beacon b_i to reduce the uncertainty of the SLAM state. b_i has impact on the SLAM state only if its update contribution reaches the robot, i.e., if b_i receives the *UpdateReq* message sent by the robot and if the robot receives the *UpdateResp* message with the update contribution from b_i. These two events are statistically independent. Taking $p_{r,i}$ as the Packet Reception Rate (PRR) from the robot to b_i and assuming symmetric PRRs, $IG_{i,t}$ can be estimated as:

$$IG_{i,t} = p_{r,i}^2 \, ui_{i,t}, \tag{8}$$

where $ui_{i,t}$ estimates the capability of the measurements gathered by b_i to reduce the uncertainty in the SLAM state. Transmission errors in sensor networks are not infrequent. This criterion naturally assigns more measurements not only to the most informative beacons, but also to those with better communication with the robot.

Each b_i computes its own $ui_{i,t}$, described in Section 5, and transmits it to the robot in an *UpdateResp* message. The robot can measure $p_{r,i}$ to each $b_i \in BS_r$ by simply analyzing message transmission success. Next, the robot allocates the NM_{max} measurements among beacons $b_i \in BS_r$ proportionally to $IG_{i,t}$ and creates LM, the list with the number of measurements assigned to each beacon $b_i \in BS_r$.

5. Operation of Beacons

The operation of beacons is summarized in Algorithm 2. Once beacon b_i has received the *UpdateReq* message it performs as follows: (1) executes *measurement allocation* and selects the most informative measurements; (2) gathers and integrates in SLAM the selected measurements and; (3) transmits to the robot its update contribution in an *UpdateResp* message.

Algorithm 2: Summary of the operation of beacon b_i

1: Receive *UpdateReq* message. Extract ev_i and LM_i
2: *Measurement allocation.* Compute $J_{i,j}$ and ui_i
3: Gather the LM_i measurements with the highest $J_{i,j}$. Create MS_i
4: **if** (b_i is at "state vector phase") **then**
5:    Compute $\xi_{i,t}$ and $\Omega_{i,t}$ with MS_i as in (15)–(16)
6: **else**
7:    Use MS_i to update the PF of b_i
8: **end if**
9: Create *UpdateResp* message and transmit it to the robot

5.1. Measurement Allocation

UpdateReq messages include LM. Once beacon b_i has received an *UpdateReq* message, it extracts LM_i, the number of measurements it was assigned with. Let BS_i be the set of beacons b_j within the sensing region of b_i. In *measurement allocation* each beacon b_i selects which measurements $z_{i,j}, b_j \in BS_i$ it should gather and integrate in SLAM. We adopt a common approach in information-driven measurement selection and formulate the problem as the greedy optimization of a utility function that establishes a trade-off between information gain and resource consumption:

$$J_{i,j} = r_{i,j} - \alpha c_{i,j} \tag{9}$$

$c_{i,j}$ is the cost of the resources consumed in gathering and integrating measurement $z_{i,j}$. In sensor networks energy is maybe the most constrained resource. We take $c_{i,j}$ as the energy consumed by b_i in gathering and integrating $z_{i,j}$. $c_{i,j}$ could be different for each beacon, e.g., depending on the remaining energy in its batteries. For simplicity, $c_{i,j}$ was assumed the same for all measurements. The reward $r_{i,j} = ui_{i,j}$ is the expected SLAM uncertainty improvement resulting after integrating $z_{i,j}$. α is a weighting factor that balances the cost the reward. Its effects will be evaluated in Section 6.3. This cost-reward approach has been used in the literature, like in [25], where they used it to make decisions on a robot exploration problem. In our work, α was determined experimentally.

The reward $r_{i,j} = ui_{i,j}$ is determined as follows. In our distributed scheme each beacon maintains its own local state. $\overline{\Omega}_{i,t}$ is the predicted information matrix for time t of the local state of b_i. $\overline{\Omega}_{i,t}$ was computed by b_i at $t-1$. It is easy to notice that the updated information matrix of the local state of b_i that would result after integrating measurement $z_{i,j}$ is:

$$\Omega'_{i,t} = \overline{\Omega}_{i,t} + \Omega_{i,j,t} \tag{10}$$

where $\Omega_{i,j,t}$ is the expected contribution of $z_{i,j}$ and is computed as follows:

$$\Omega_{i,j,t} = H_{i,j,t}^T R^{-1} H_{i,j,t}, \tag{11}$$

where R is the covariance matrix of the measurement noise and $H_{i,j,t}$ is the Jacobian of the observation model of measurement $z_{i,j}$ computed with ev_i, just received from the robot in the *UpdateReq* message.

On the other hand, in case of not integrating $z_{i,j}$, the updated information matrix for b_i would be $\Omega_{i,t}^n = \overline{\Omega}_{i,t}$. The uncertainty improvement $ui_{i,j}$ is the difference of the uncertainty in $\Omega'_{i,t}$ and in $\Omega_{i,t}^n$. Entropy is maybe the most widely-used metric for the uncertainty in a probability distribution. It is adopted in our scheme. Entropy can be used to measure the uncertainty of beacons in the "state vector phase" and also of beacons in the "initialization phase", giving the same treatment to both cases. If beacon b_i is in the "state vector phase", its state follows a Gaussian probability distribution and its entropy can be computed using an exact expression. In this case $ui_{i,j}$ is as follows:

$$ui_{i,j} = \frac{1}{2} \log \left(\frac{|\overline{\Omega}_{i,t}|}{|\overline{\Omega}_{i,t} + \Omega_{i,j,t}|} \right) \tag{12}$$

If b_i is in the "initialization phase", i.e., its PF has not converged, its probability distribution is approximated by the set of PF particles. In this case, there is not an exact expression and each b_i computes its $ui_{i,j}$ using the approximate calculation described in [26].

It should be noticed that if b_j is in the "initialization phase", it is still not in the state vector of its neighbor b_i. Hence, $z_{i,j}$ is not useful to update the local map of b_i, either if b_i is in the "initialization phase" or in the "state vector phase". Hence, the uncertainty improvement $ui_{i,j}$ is taken as zero.

Long-term optimization of $J_{i,j}$ involves high computational burden and bad scalability. We adopted a simple but efficient greedy approach: At each time b_i selects the LM_i beacons $b_j \in BS_i$ that achieve the highest value in $J_{i,j}$. Of course, measurements with negative gain-cost utility, $J_{i,j} < 0$, are not selected.

Each beacon b_i receiving the *UpdateReq* message also computes ui_i, which will be used by the robot in *measurement distribution*. ui_i estimates how good it is to assign measurements to b_i, i.e., the expected improvement in the uncertainty of the local state of b_i if b_i integrates one measurement to each beacon $b_j \in BS_i$. Similarly, the updated information matrix for b_i that would result after integrating one measurement to each beacon $b_j \in BS_i$ is:

$$\Omega''_{i,t} = \overline{\Omega}_{i,t} + \sum_{j \in BS_i} \Omega_{i,j,t} \tag{13}$$

As above, if no measurement is integrated, the updated information matrix is $\Omega_{i,t-1}$. Thus, ui_i is computed as follows:

$$ui_i = \frac{1}{2} \log \left(\frac{|\overline{\Omega}_{i,t}|}{|\overline{\Omega}_{i,t} + \sum_{j \in BS_i} \Omega_{i,j,t}|} \right) \tag{14}$$

5.2. Integration of Measurements

At this step, beacon b_i has already gathered one measurement to the robot and to each of the beacons selected in *measurement allocation*. Let MS_i be the set of gathered measurements. The next step is to integrate them. If beacon b_i is in the "initialization phase", it updates its PF with the measurements in MS_i. If b_i is in the "state vector phase", it integrates them in its local state and computes its update contribution as follows:

$$\xi_{i,t} = \sum_{j \in MS_i} H_{i,j,t}^T R^{-1} [z_{i,j} - h_{i,j}(ev_i) + H_{i,j,t} ev_i], \tag{15}$$

$$\Omega_{i,t} = \sum_{j \in MS_i} H_{i,j,t}^T R^{-1} H_{i,j,t}, \tag{16}$$

where $h_{i,j}(ev_i)$ and $H_{i,j,t}$ are respectively the predictions and Jacobians for each measurement in MS_i, either robot-beacon or inter-beacon measurement. Finally, b_i transmits an *UpdateResp* message to the robot with its contribution to the SEIF update ($\xi_{i,t}$ and $\Omega_{i,t}$) and to *measurement distribution* (ui_i).

6. Experiments

The validation of the proposed scheme was performed in sets of 3D SLAM outdoor experiments with AMUSE UAS, an octorotor developed in the UE-FP7 ARCAS project for maintenance and repairing of industrial facilities [27], see Figure 5. Maintenance of industrial facilities is currently performed using sensor nodes that gather measurements for process monitoring and anomaly detection. In these complex scenarios GPS is often unavailable or has bad quality. UAS are suitable tools for confirming and eventually repairing the anomalies detected but they require accurate localization. The proposed resource-constrained RO-SLAM scheme is very interesting in this problem. Besides the typical technological constraints of UAS and beacons, in these scenarios there are often a high number of sensors and wireless devices involving significant bandwidth limitations. Besides, the energy consumed by nodes (beacons) is constrained in order to avoid frequent battery replacements.

Figure 5. AMUSE UAS flying during one experiment at the School of Engineering of Seville.

A total of 24 beacons were deployed at random locations and different heights in a 20 × 20 m scenario, beacons are marked in Figure 5. Each beacon was comprised of a *RaspberryPi* running Linux connected through USB to a *Nanotron nanoPAN 5375* [28] Time-of-Flight (ToF) range sensor and to a WiFi USB adapter, all powered by an external battery, see Figure 6b. The performance of *Nanotron* sensors in outdoors was characterized experimentally. Their range error can be modeled as a Gaussian PDF with zero mean and a standard deviation of $\sigma_m = 0.6$ m, see Figure 6c. The resolution of these sensors is 0.01 m. It should be noticed that measurements from two or more different beacons can always be distinguished because the measurements are tagged with a unique identifier for each beacon. Each beacon run an independent ROS (Robots Operating System) node. The ROS node implements the beacon algorithm, gathers range measurements with the *Nanotron* ToF range sensor using an ad-hoc developed ROS driver and communicates with the other beacons using WiFi. One beacon was mounted on the landing skid of AMUSE, see Figure 6a. In the experiments the proposed SLAM scheme was executed at a rate of 1 Hz, one iteration per second. The robot beacon transmitted *UpdateReq* messages at a rate of 1 Hz. In these experiments the main resource constraint was imposed by the

sensing channel bandwidth. $NM_{max} = 80$ was the maximum number of range measurements per iteration that we could gather without interference in the experiments.
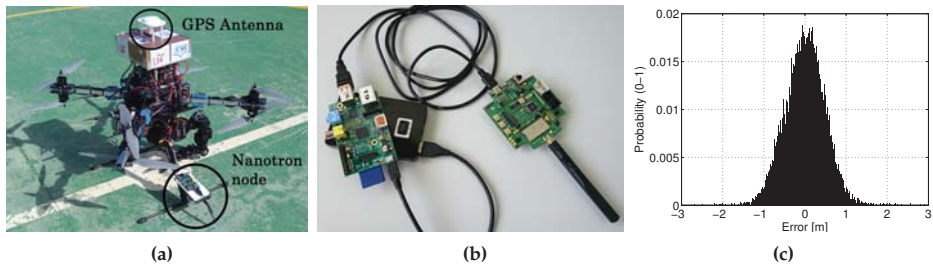
Figure 6. (**a**) Picture of AMUSE octorotor during one experiment; (**b**) Picture of a beacon comprised of a *Nanotron* ToF range sensor connected through USB to a *RaspberryPi* module powered by a battery; (**c**) Experimental outdoor characterization of *Nanotron nanoPAN 5375* sensors.

In these experiments AMUSE was in manual flight. The objective was not to use the proposed scheme for real-time navigation. AMUSE is equipped with a *Novatel OEM6* GPS unit with 2 cm accuracy. We used GPS only as ground truth for accuracy assessment. SLAM provides the generated map and robot location in a local coordinate frame. To compare with the ground-truth, an affine transform is performed on the final beacon locations, re-aligning the local solution into the same global coordinate frame. UAS odometry obtained from Inertial Measurement Units is often too noisy to be used in SLAM. Like most works in 3D SLAM, e.g., [29], we opted for employing some beacons, 5 in the experiments performed, as anchors for correcting the UAS localization.

The auxiliary PFs for beacon initialization were set with 500 particles. Each beacon executes its own PF, initializes the PF when it receives the first measurement and decides that it has converged when the maximum eigenvalue of the covariance matrix is lower than 0.4. The eigenvalues of the covariance matrix are directly proportional to the variance along the corresponding eigenvectors. Thus, the maximum eigenvalue is a measure of the volume (or at least the largest axis) of the confidence ellipsoid of the distribution. Using the eigenvalues puts the convergence condition on each axis. Our scheme has only two parameters: NM_{max} and α. We used $NM_{max} = 80$ and $\alpha = 7.5$ in all the experiments. For simplicity in these experiments the cost of measurement $z_{i,j}$ in $J_{i,j}$ was taken constant and equal to the energy consumed by beacons when taking one measurement $c_{i,j} = 6.6$ mJ.

6.1. Validation

The result of the proposed scheme in one experiment in XY (left) and 3D (right) views is shown in Figure 7. Blue lines and stars represent respectively the ground truth UAS trajectory and beacon locations. Red lines and triangles are the resulting estimates. The total number of measurements integrated at each iteration along the experiment is shown in Figure 8a. At the beginning all beacons gathered all the measurements they were assigned with: In total $NM_{max} = 80$ between all the beacons. As the experiment advanced the beacon local states had lower and lower uncertainty and inter-beacon measurements became less and less informative. From $t = 108$ s on, some measurements achieved $J_{i,j} < 0$; their reward was lower than the cost, and were not gathered anymore. In average the number of measurements per iteration in this experiment was 61, lower than NM_{max}. The proposed scheme ensures the given NM_{max} bound avoiding reward-cost inefficient measurements.

(a)

(b)
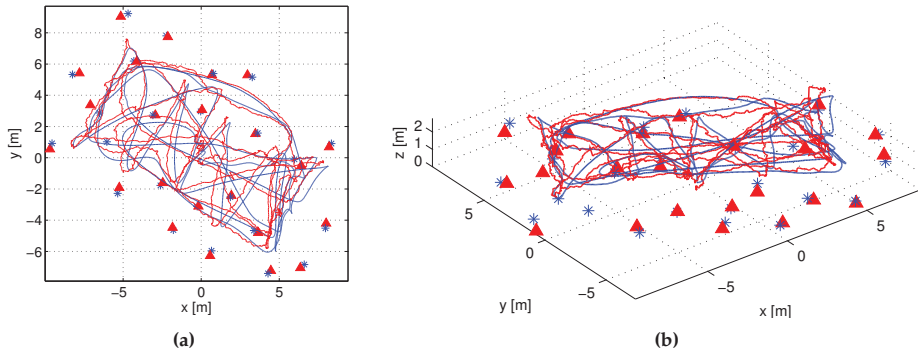
Figure 7. Results of the proposed scheme in a 3D SLAM experiment with AMUSE UAS: XY (**a**) and 3D views (**b**).
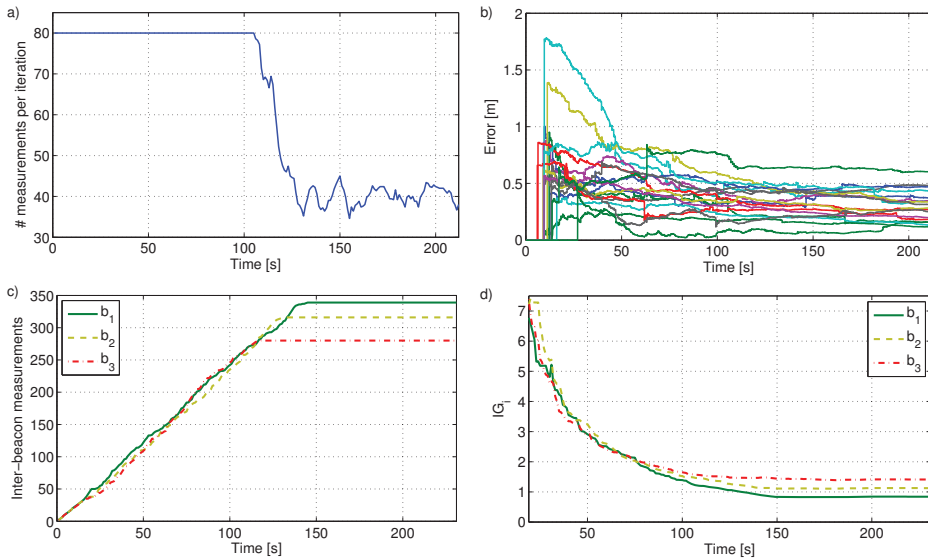
Figure 8. (**a**) Number of measurements integrated at each iteration along the experiment; (**b**) Evolution of beacon localization errors; (**c**) Number of measurements gathered by three beacons along the experiment; (**d**) Values of IG_i along the experiment for the beacons in (**c**).

The evolution of beacon localization errors along the experiment is shown in Figure 8b. The drawing for each beacon starts when its PF converged. The majority of the PFs converged between $t = 6$ s and $t = 16$ s, shortly after the start of the experiment. The UAS localization errors in the three coordinates are shown in Figure 9. The red dashed lines represent the 3σ bounds showing the consistency of the estimations.
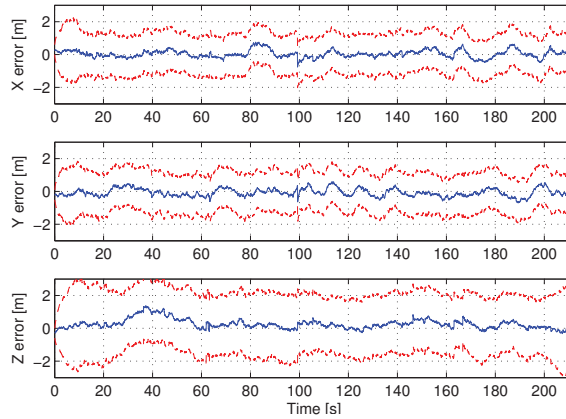
Figure 9. Evolution of UAS localization error in the experiment.

The cumulative number of inter-beacon measurements gathered by three beacons along the experiment is shown in Figure 8c. Similar curves were obtained for all beacons. The shape of each curve is a ramp with almost constant slope until the beacon stops gathering measurements. This evolution is useful to analyze the performance of *measurement distribution* and *measurement allocation*. At the beginning each beacon gathers all the measurements it is assigned with. *Measurement distribution* assigns measurements to b_i proportionally to IG_i. More measurements are assigned to beacons with higher uncertainty. Hence, *measurement distribution* naturally balances the values of IG_i of all the beacons. Figure 8d shows that the three beacons represented in Figure 8c have similar evolution in IG_i. This can be observed for all beacons. As a result all beacons are assigned with a similar number of measurements, resulting in similar slopes in Figure 8c. Beacon b_i gathers and integrates $z_{i,j}$ as long as $J_{i,j} > 0$. The uncertainty of b_i will be lower as it integrates more measurements. After a while, the measurements gathered by b_i will not satisfy $J_{i,j} > 0$ and it will stop taking measurements, the lope becomes zero in Figure 8c. Each beacon has its own different situation (number of neighbors, time of PF convergence, etc.) hence, they will reach zero-slope at different times.

The proposed scheme can dynamically adapt to different values of NM_{max}. The previous experiment was repeated simulating that during interval $t \in [90, 105]$ the number of measurements was bounded by $NM_{max} = 30$, see Figure 10. In this case the robot RMS error was 0.516 m, very similar to that with $NM_{max} = 80$ along the entire experiment, which was 0.51 m. The difference in the map error was even smaller. The scheme selects the most informative measurements reducing the impact of changes in NM_{max}. The only effect is that beacons stop gathering inter-beacon measurements, reaching $J_{i,j} < 0$, later in the experiment. They keep gathering measurements and at the end of the experiment the number of measurements integrated are the same in both cases.
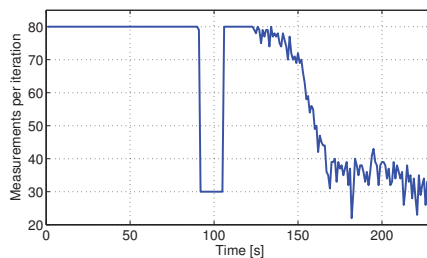
Figure 10. Experiment in Figure 8 taking $NM_{max} = 30$ during $t \in [90, 105]$.

6.2. Performance Comparison

The proposed scheme was evaluated and compared with other methods in 20 sets of real experiments with different beacon settings and UAS trajectories. Method *M1* is a conventional SEIF SLAM scheme that integrates only robot-beacon measurements. Method *M2* is the distributed SEIF SLAM reported in [13]. It integrates robot-beacon and all inter-beacon measurements. Method *M3* is *M2* combined with a tool that selects the NM_{max} measurements that best improve the uncertainty in the global state: This tool is necessarily centralized at the robot. In the proposed distributed scheme each beacon selects the best measurements to improve its local uncertainty. Comparing with method *M3* allows evaluating how far our distributed measurement selection is from the centralized selection. The data from the sets of experiments was logged and the four methods were executed offline with the same parameters. Their performance is compared in Table 1, which analyzes robot and map RMS errors, convergence times of auxiliary PFs, number of measurements actually integrated per iteration, average energy consumed by beacons and average robot CPU time consumed evaluated in percentage w.r.t. that of *M1*. Recall that the number of measurements integrated per iteration is proportional to the energy consumed by beacons (shown in the table) and to the beacon computational time required for measurement integration (not shown in the table).

Table 1. Comparison of the proposed scheme versus methods *M1*, *M2* and *M3*.

	M1	*M2*	*M3*	**Proposed**
Map RMS error (m)	0.49	0.33	0.34	0.34
Robot RMS error (m)	0.59	0.49	0.50	0.51
PF convergence times (s)	25.2	5.4	5.6	5.7
# of measurements/iteration	33.2	206.9	80	61.7
Beacon energy consumption (J)	43.7	272.2	105.2	81.1
Robot CPU time (% of *M1*)	100	65.6	265.5	58.6

M1 does not integrate inter-beacon measurements and hence had the poorest errors and PF convergence times. *M2* integrates inter-beacon measurements, which significantly reduces PF convergence times, 78%, and map and robot RMS errors, 32% and 16%, respectively, over *M1*. On the other hand, *M2* gathered and integrated 523% more measurements, which largely increased beacon energy consumption. *M2* distributes computation between the robot and beacons and hence reduces the robot CPU times over *M1*. *M3* is *M2* combined with a centralized tool that selects the NM_{max} measurements that best improve the uncertainty of the global state. *M3* integrated 80 measurements per iteration, 61% lower than *M2*, and achieved similar RMS errors (difference <3%) and PF convergence times (<4%). However, it required much larger robot CPU times: It uses the information matrix of the global state to select the most informative measurements and computing determinants has $O(n^3)$ complexity.

The proposed scheme obtained similar RMS errors, PF convergence times and robot CPU times to *M2* [13] requiring 70% less measurements and 70% lower beacon energy consumption. Besides, our scheme achieved similar RMS errors and PF convergence times to *M3* but required 23% less measurements (and beacon energy consumption). Each beacon uses the information matrix of its local state for measurement selection, requiring 78% lower robot CPU burden than *M3*. Besides, in our scheme each beacon maintains a local version of its map, which can be useful in some cases. Once beacons have built their local map, they can transmit it to any robot, which can immediately recover the full map applying map-joining techniques.

6.3. Discussion

In the following we discuss on the scalability of the proposed scheme and analyze the impact of transmission errors and of the parameters of the method: NM_{max} and α.

The proposed method preserves the scalability of the distributed SEIF presented in [13]. *Measurement distribution* involves only the beacons within the robot sensing region, whereas *measurement allocation*, performed by each beacon b_i, involves only the beacons within the sensing region of b_i. Thus, the computational complexity depend on the beacon density, not on the map size. Beacons are used as landmarks in RO-SLAM: Highly inhomogeneous local beacon densities are not suitable in RO-SLAM. It is often more interesting if beacons are deployed in densities with some homogeneity, leading to constant time execution regardless of the map size.

NM_{max} is taken as an input to our scheme. The performance of the proposed scheme with different values of NM_{max} is summarized in Table 2. The measurements from all the experiments were logged and the proposed scheme was offline executed with NM_{max}. The integration of measurements is critical for PF convergence and low values of NM_{max} decelerate PF convergence. On the other hand, the estimation accuracy was only slightly affected, which is attributed to its capability to select informative measurements. With $NM_{max} = 80$ the average number of measurements actually integrated in SLAM was similar to that with $NM_{max} = 60$. The explanation is the role of α. In the experiments all measurements are assumed to have the same cost. Thus, α acts as a threshold since measurement $z_{i,j}$ is gathered only if $J_{i,j} = r_{i,j} - \alpha c_{i,j} > 0$.

Table 2. Average performance of the proposed scheme with different values of NM_{max}.

	$NM_{max} = 40$	$NM_{max} = 60$	$NM_{max} = 80$
Map RMS error (m)	0.35	0.346	0.34
Robot RMS error (m)	0.52	0.51	0.51
PF convergence times (s)	15.8	9.5	5.7
# of measurements/iteration	40	56.5	61.7

The performance of the proposed scheme with different values of α is summarized in Table 3. α allows setting our scheme to prevent integrating measurements that are not very informative. With $\alpha = 1.5$, almost all measurements satisfy $J_{i,j} > 0$ and the number of measurements integrated in SLAM is almost NM_{max}. With $\alpha = 15$, many measurements do not satisfy $J_{i,j} > 0$ soon in the experiments and in average only 49.3 measurements were integrated per iteration. These were the two extremes in the range of α, an intermediate value $\alpha = 7.5$ was used. Despite the difference in the number of measurements, the value of α affects accuracy very slightly as shown in Table 3.

Table 3. Average performance of the proposed scheme with different values of α.

	$\alpha = 1.5$	$\alpha = 7.5$	$\alpha = 15$
Map RMS error (m)	0.34	0.34	0.37
Robot RMS error (m)	0.51	0.51	0.52
PF convergence times (s)	5.7	5.7	5.9
# of measurements/iteration	78.9	61.7	49.3

The proposed distributed scheme needs communication between the robot and beacons. In this sense the transmission errors in sensor networks cannot be ignored. Its performance assuming different PRR levels is summarized in Table 4. Our method explicitly considers PRR in the estimation of IG_i and assigns more measurements to the beacons that have better link quality with the robot. As expected, it exhibits good robustness to PRR. Even with PRR = 40%, transmission error rate of 60%, its performance is very slightly perturbed.

Table 4. Average performance of the proposed scheme with different PRR levels.

	PRR = 40	PRR = 60	PRR = 80	PRR = 100
Map RMS error (m)	0.4	0.37	0.35	0.35
Robot RMS error (m)	0.57	0.53	0.52	0.51
PF convergence times (s)	9.6	7.1	6.4	5.7
# of measurements/iteration	44.8	51.4	57.1	61.7

7. Conclusions

RO-SLAM has some characteristics that make it more suitable than visual SLAM in a wide range of robot-sensor node cooperative missions. Most RO-SLAM methods consider beacons as passive landmarks and do not exploit the capabilities with which beacons are actually endowed. The accuracy of RO-SLAM estimations improves with the number and variety of measurements that are integrated in SLAM. However, using more measurements requires consuming more resources to gather, transmit and integrate the measurements in SLAM, which often contrasts with existing technological or application constraints.

This paper presents a scalable robot-beacon distributed RO-SLAM scheme for resource-constrained operation. The objective is to improve SLAM performance while meeting a given resource consumption bound expressed as the maximum number of measurements that can be integrated in SLAM per iteration. In our problem, the number of measurements is a good metric for resource consumption since it directly impacts the sensing channel capacity used, the beacon energy consumed and the computational capacity employed, among others.

The proposed scheme efficiently combines a distributed SEIF SLAM method that integrates robot-beacon and inter-beacon measurements, together with a distributed information-driven tool that selects the measurements to be integrated in SLAM balancing uncertainty improvement and resource consumption. The scheme has a robot-beacon distributed approach where beacons actively participate in measurement selection, gathering and integration in SLAM. Our scheme ensures resource-constrained operation with static or dynamic bounds, showing significant flexibility. It achieves higher accuracy and lower beacon initialization times than conventional SLAM methods. Besides, it can be executed in constant time regardless of the map size.

Its performance was validated and compared with existing methods in sets of 3D SLAM experiments. Robustness analysis confirmed its stable and predictable performance against transmission errors and different values of its parameters.

Author Contributions: A. Torres-González and J.R. Martínez-de Dios contributed equally in the research that leads to this paper. A. Torres-González developed, implemented the algorithms and performed the real experiments and wrote some parts of the manuscript. J.R. Martínez-de Dios helped in the design of the algorithms and elaborated the major part of the manuscript. A. Ollero provided suggestions and corrections during the preparation of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

Σ_t	Covariance matrix of the SLAM global state at time t
μ_t	Mean of the SLAM global state at time t
Ω_t	Updated information matrix of the SLAM global state at time t
ξ_t	Updated Information vector of the SLAM global state at time t
$\bar{\Omega}_t, \bar{\xi}_t$	Predicted information matrix and predicted information vector of the SLAM global state for time t
$\xi_{i,t}, \Omega_{i,t}$	Update contribution of beacon b_i to ξ_t
$z_{r,i}, z_{i,j}$	Measurement gathered by the robot to beacon b_i. Measurement gathered by beacon b_i to b_j
$h_{r,i}, h_{i,j}$	Observation models for robot-beacon and inter-beacon measurements
$H_{r,i}, H_{i,j}$	Jacobians of the observation models for robot-beacon and inter-beacon measurements
BS_r, BS_i	Sets of the beacons that are currently within the sensing region of the robot and beacon b_i, respectively
LM	List with the number of measurements assigned to each beacon in BS_r in *measurement distribution*
MS_i	Set of measurements gathered by beacon b_i
NM_{max}	Maximum number of measurements that can be gathered and integrated per SLAM iteration
$J_{i,j}$	Utility function for measurement $z_{i,j}$
$r_{i,j}, c_{i,j}$	Reward and cost for measurement $z_{i,j}$
α	Weighting factor between reward and cost

References

1. Brunner, C.J.; Peynot, T.; Vidal-Calleja, T.A.; Underwood, J.P. Selective Combination of Visual and Thermal Imaging for Resilient Localization in Adverse Conditions: Day and Night, Smoke and Fire. *J. Field Robot.* **2013**, *30*, 641–666.

2. Banatre, M.; Marron, P.; Ollero, A.; Wolisz, A. *Cooperating Embedded Systems and Wireless Sensor Networks*; John Wiley & Sons: Hoboken, NJ, USA, 2008.

3. Corke, P.; Hrabar, S.; Peterson, R.; Rus, D.; Saripalli, S.; Sukhatme, G. Deployment and Connectivity Repair of a Sensor Net with a Flying Robot. In *Experimental Robotics IX*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 21; pp. 333–343.

4. Maza, I.; Caballero, F.; Capitan, J.; Martinez-de Dios, J.R.; Ollero, A. A distributed architecture for a robotic platform with aerial sensor transportation and self-deployment capabilities. *J. Field Robot.* **2011**, *28*, 303–328.

5. Martinez-de Dios, J.; Lferd, K.; de San Bernabe, A.; Nunez, G.; Torres-Gonzalez, A.; Ollero, A. Cooperation Between UAS and Wireless Sensor Networks for Efficient Data Collection in Large Environments. *J. Intell. Robot. Syst.* **2013**, *70*, 491–508.

6. Cobano, J. A.; Martínez-de Dios, J.R.; Conde, R.; Sánchez-Matamoros, J.M.; Ollero, A. Data retrieving from heterogeneous wireless sensor network nodes using UAVs. *J. Intell. Robot. Syst.* **2010**, *60*, 133–151.

7. Todd, M.; Mascarenas, D.; Flynn, E.; Rosing, T.; Lee, B.; Musiani, D.; Dasgupta, S.; Kpotufe, S.; Hsu, D.; Gupta, R.; et al. A different approach to sensor networking for SHM: Remote powering and interrogation with unmanned aerial vehicles. In Proceedings of the 6th International Workshop on Structural Health Monitoring (Citeseer), Stanford, CA, USA, 11–13 September 2007.

8. Challa, S.; Leipold, F.; Deshpande, S.; Liu, M. Simultaneous Localization and Mapping in Wireless Sensor Networks. In Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference, Melbourne, Australia, 5–8 December 2005, pp. 81–87.

9. Menegatti, E.; Zanella, A.; Zilli, S.; Zorzi, F.; Pagello, E. Range-only SLAM with a mobile robot and a Wireless Sensor Networks. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, 12–17 May 2009; pp. 8–14.

10. Sun, D.; Kleiner, A.; Wendt, T. Multi-robot Range-Only SLAM by Active Sensor Nodes for Urban Search and Rescue. In *RoboCup 2008: Robot Soccer World Cup XII*; Iocchi, L., Matsubara, H., Weitzenfeld, A., Zhou, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5399; pp. 318–330.

11. Nogueira, M.; Sousa, J.; Pereira, F. Cooperative Autonomous Underwater Vehicle localization. In Proceedings of the IEEE Oceans, Sydney, Australia, 24–27 May 2010; pp. 1–9.

12. Djugash, J.; Singh, S.; Kantor, G.; Zhang, W. Range-only SLAM for robots operating cooperatively with sensor networks. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Orlando, FL, USA, 15–19 May 2006; pp. 2078–2084.

13. Torres-González, A.; Martinez-de Dios, J.; Ollero, A. Efficient Robot-Sensor Network Distributed SEIF Range-Only SLAM. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2014), Hong Kong, China, 31 May–7 June 2014; pp. 1319–1326.

14. Menegatti, E.; Danieletto, M.; Mina, M.; Pretto, A.; Bardella, A.; Zanconato, S.; Zanuttigh, P.; Zanella, A. Autonomous discovery, localization and recognition of smart objects through WSN and image features. In Proceedings of the IEEE GLOBECOM 2010; Miami, FL, USA, 6–10 December 2010; pp. 1653–1657.

15. Olson, E.; Leonard, J.; Teller, S. Robust range-only beacon localization. In Proceedings of the IEEE/OES Autonomous Underwater Vehicles, Sebasco, ME, USA, 17–18 June 2004; pp. 66–75.

16. Blanco, J.L.; Fernandez-Madrigal, J.A.; Gonzalez, J. Efficient probabilistic Range-Only SLAM. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'08, Nice, France, 22–26 September 2008; pp. 1017–1022.

17. Torres-González, A.; Martinez-de Dios, J.; Ollero, A. Integrating Internode Measurements in Sum of Gaussians Range Only SLAM. In *ROBOT2013: First Iberian Robotics Conference*; Springer International Publishing: Basel, Switzerland, 2014; pp. 473–487.

18. Thrun, S.; Liu, Y.; Koller, D.; Ng, A.Y.; Ghahramani, Z.; Durrant-Whyte, H. Simultaneous Localization and Mapping with Sparse Extended Information Filters. *Int. J. Robot. Res.* **2004**, *23*, 693–716.

19. Djugash, J.; Singh, S.; Grocholsky, B. Decentralized mapping of robot-aided sensor networks. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2008), Pasadena, CA, USA, 19–23 May 2008; pp. 583–589.

20. Ila, V.; Porta, J.; Andrade-Cetto, J. Information-Based Compact Pose SLAM. *IEEE Trans. Robot.* **2010**, *26*, 78–93.

21. Vidal-Calleja, T.A.; Sanfeliu, A.; Andrade-Cetto, J. Action selection for single-camera SLAM. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2010**, *40*, 1567–1581.

22. Torres-González, A.; Martinez-de Dios, J.R.; Ollero, A. An Adaptive Scheme for Robot Localization and Mapping with Dynamically Configurable Inter-Beacon Range Measurements. *Sensors* **2014**, *14*, 7684–7710.

23. Thrun, S.; Burgard, W.; Fox, D. *Probabilistic Robotics*, 3rd ed.; The MIT Press: Cambridge, UK, 2005.

24. Baccour, N.; Koubâa, A.; Mottola, L.; Zúñiga, M.A.; Youssef, H.; Boano, C.A.; Alves, M. Radio link quality estimation in wireless sensor networks: A survey. *ACM Trans. Sens. Netw. (TOSN)* **2012**, *8*, 34.

25. Stachniss, C.; Grisetti, G.; Burgard, W. Information Gain-based Exploration Using Rao-Blackwellized Particle Filters. *Robot. Sci. Syst.* **2005**, *2*, 65–72.

26. Boers, Y.; Driessen, H.; Bagchi, A.; Mandal, P. Particle filter based entropy. In Proceedings of the IEEE Conference on Information Fusion (FUSION 2010), Edinburgh, UK, 26–29 July 2010.

27. Heredia, G.; Jiménez-Cano, A.; Sánchez, M.; Llorente, D.; Vega, V.; Braga, J.; Acosta, J.; Ollero, A. Control of a Multirotor Outdoor Aerial Manipulator. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), Chicago, IL, USA, 14–18 September 2014.

28. Nanotron nanoPAN 5375. Available online: http://nanotron.com/EN/PR_ic_modules.php (accessed on 6 April 2017).

29. Torres-González, A; Martinez-de Dios, JR; Jiménez-Cano, A; Ollero, A An Efficient Fast-Mapping SLAM Method for UAS Applications Using Only Range Measurements. *Unmanned Syst.* **2016**, *4*, 155–165.

Influence of Wind Speed on RGB-D Images in Tree Plantations

Dionisio Andújar [1,*], José Dorado [2], José María Bengochea-Guevara [1], Jesús Conesa-Muñoz [1], César Fernández-Quintanilla [2] and Ángela Ribeiro [1]

[1]   Centre for Automation and Robotics, Spanish National Research Council, CSIC-UPM, Argandadel Rey, 28500 Madrid, Spain; jose.bengochea@car.upm-csic.es (J.M.B,-G.); jesus.conesa@csic.es (J.C.-M.); angela.ribeiro@csic.es (Á.R.)

[2]   Institute of Agricultural Sciences, Spanish National Research Council, CSIC, 28006 Madrid, Spain; jose.dorado@ica.csic.es (J.D.); cesar@ica.csic.es (C.F.-Q.)

*   Correspondence: dionisioandujar@hotmail.com; Tel.: +34-91-745-2500

Abstract: Weather conditions can affect sensors' readings when sampling outdoors. Although sensors are usually set up covering a wide range of conditions, their operational range must be established. In recent years, depth cameras have been shown as a promising tool for plant phenotyping and other related uses. However, the use of these devices is still challenged by prevailing field conditions. Although the influence of lighting conditions on the performance of these cameras has already been established, the effect of wind is still unknown. This study establishes the associated errors when modeling some tree characteristics at different wind speeds. A system using a Kinect v2 sensor and a custom software was tested from null wind speed up to 10 m·s^{-1}. Two tree species with contrasting architecture, poplars and plums, were used as model plants. The results showed different responses depending on tree species and wind speed. Estimations of Leaf Area (LA) and tree volume were generally more consistent at high wind speeds in plum trees. Poplars were particularly affected by wind speeds higher than 5 m·s^{-1}. On the contrary, height measurements were more consistent for poplars than for plum trees. These results show that the use of depth cameras for tree characterization must take into consideration wind conditions in the field. In general, 5 m·s^{-1} (18 km·h^{-1}) could be established as a conservative limit for good estimations.

Keywords: RGB-D images; Kinect sensor limits; depth information; wind speed; woody crops

---

1. Introduction

Plant reconstruction by non-destructive methods is of high value for decision making processes [1]. The use of sensors for plant characterization may lead to a better knowledge of the processes involved in plant development all over throughout the life cycle and may improve the decisions taken for plant production, contributing to create new protocols to enhance the profitability of crops [2].

Current techniques for plant characterization vary from manual to fully automatic, using a great variety of imaging to non-imaging technologies [3]. Plant characterization using machine vision by portable imaging and analysis software is the most investigated technique [4]. RGB cameras have been widely used for phenology monitoring [5], plant geometric characterization [6], nitrogen application [7], yield monitoring [8] and weed/crop discrimination [9]. These cameras can acquire images with a high resolution and at a low cost. However, their limited capacity to provide spectral and structural information is a deterrent to their usage in plant reconstruction. In addition, under outdoor conditions, the variable and uncontrolled illumination and the presence of shadows may represent a serious problem [10].

Although various systems can be used to estimate plant parameters, the third dimension is usually missing. Distance sensors, which measure the distance from the sensor to an object, allow one to create 3D models, increasing the effectiveness in plant description [11]. There are two main types of distance sensors available: ultrasonic and light detection and ranging (LiDAR). Both of them can be used for plant height and volume estimation. The ultrasonic sensors measure the reflected echo that the sensor transmits as sound waves of high frequency toward an object. They have been used for tree canopy characterization [12] and for discrimination between plant types (grasses and broad-leaved) [13,14]. Distance measurements may be improved by using a terrestrial laser scanner (TLS). This is a simple device that is able to provide 2D or 3D plant models by displacing the sensor along the row and storing the relative position of the sensor. The sensor measures the distance to the impacted object with a laser beam by time-of-flight method or phase-shift measurement. Llorens et al. [15] used a dual methodology of LiDAR and ultrasonic sensors in the characterization of tree canopies, concluding that LiDAR assessments were the most accurate. Andújar et al. [16] obtained promising results using a TLS to characterize the profile of poplar trees. Various other systems have also been explored for plant characterization in precision agriculture, including radar systems [17], hemispherical photography [18], stereo-vision [19], and magnetic resonance or X-ray visualization [20].

In the last few years, a new geometric reconstruction concept based on the use of depth cameras has arisen [21]. The Kinect device, originally designed by Microsoft to track body position and movement for video-games and digital entertainment, has also been used for plant characterization in agriculture. Paulus et al. [22] compared two low-cost 3D imaging systems (a Kinect device and a David laser scanning system) with a high precision laser scanner to estimate volumetric shape of sugar beet leaves and taproots and of wheat ears. They concluded that the two low-cost sensors could replace the expensive laser scanner in some of the plant phenotyping scenarios. Chéné et al. [23] developed an algorithm to segment depth images of plant obtained from top view. They showed the potential of this sensor for volume estimation, leaf curvature and morphology, or the use for pathogens detection. Similarly, Andújar et al. [24] proposed the use of depth cameras for determining the weed volume and defining herbicide spraying date [25]. Wang and Li [26] used volume estimation from Kinect models for onions, showing that volume and fruit density were directly related with Kinect measurements.

The use of sensors located on field platforms may allow monitoring plant growth under natural environments. However, these sensors may be influenced by different environmental conditions present throughout the growing season. Recent studies have showed the influence of light on sensors measurements [27]. Although daytime light contamination has limited the potential use of Kinect v1 (projected light pattern and triangulation) sensors in open spaces [28], the new version (Kinect v2) uses a completely different measurement principle (time-of-flight) that should overcome, at least partially, this problem [29–31]. Other parameters such as wind can modify the captured information related to the plant structure. When a plant is moving due to wind conditions sensor readings can differ along the time. Decision making requires standardized conditions when the information is captured. Since wind speed cannot be controlled outdoors, the influence of its effect should be quantified. The overall objective of this work was to assess the effect of wind speed on measurements recorded by a Kinect v2 sensor related to plant height, volume and leaf area, as well as to establish the limits and errors in these parameters measured in cultivated plants and associated with different wind speeds.

2. Materials and Methods

2.1. Hardware, Software and Other Equipment

Microsoft Kinect is a RGB-D camera that uses the time-of-flight method for the calculation of the camera distance to an object. The Kinect one is composed by an RGB camera of 1080 p, a depth camera, an infrared (IR) camera and an array of microphones. Kinect v2 has a self-adaptation of the exposure time of the RGB image. Thus, by automatically adapting this parameter the sensor obtains brighter images. However, this limits the number of frames that are captured in a time interval to

a minimum frame rate of 15 fps. The time-of-flight system modulates a camera light source with a square wave. It uses phase detection to measure the time it takes light to travel from the light source to the object and back to the sensor, estimating distance from the results. The system calculates distance from the speed of light in air by estimating the received light phase at each pixel with knowledge of the modulation frequency. The RGB camera creates raw color images with a 1920×1080 resolution. Kinect only provides an array of pixel RGB values that are converted into a proper Windows Presentation Foundation (WPF) representation. The IR camera can take a clear view into the dark. It has a resolution of 512×424 pixels. The sensor allows tracking of IR reflective objects while filtering out IR lights. The Kinect v2 has a wider field of view in the IR being able of acquiring depth information at 70 degrees horizontally and 60 degrees vertically. This effect allows to capture objects closer to the camera and still in its field of view. The camera is effective at distances from 0.5 m to 4.5 m. The use of reconstruction tools allows the creation of bigger models by displacing the sensor and using custom software based on the Iterative Closest Point (ICP) algorithm [32].

The acquisition process was based on an Intel desktop computer with Windows 8 supported by Kinect SDK, which helped to acquire data by classes, functions and structures that also provided options to combining more than one sensor. The SDK provides the necessary drivers for the sensor, and some sample functions that were implemented for the measurements combined with some OpenCV functions [33]. The RGB, depth and infrared images use classes and functions to the camera acquisition process and, if applicable, also coordinate the images acquired by different sensors. The software creates a point cloud by detecting the overlapping areas in sequential frames taken during the acquisition process by assessing the relative position of the sensor for each frame. The overlapped areas allow the creation of the final 3D model and removal of outliers in the mesh.

The assessment of the influence of wind speed was conducted by placing the sensor at 1 m height and at 1 m of distance from the tree. The sensor was located on a tripod and turned from $45°$ to $-45°$, allowing a full view of the tree. The time to complete the acquisition was 10 s from the top $(45°)$ to the ground view $(-45°)$. The system was supplied with electric power by a field vehicle that allows field measurement and also the storage of the additional devices needed during the process (Figure 1). A set of measurements were conducted at different wind speeds: 0, 2.5, 5, 7.5, and 10 m·s^{-1}. Measurements were performed on a sunny cloudless day without natural wind, to avoid the influence of meteorological parameters other than the purpose of this study, i.e., wind speed in a fix direction. In order to create the wind draft a hand-held blower-vacuum (Stihl^{TM} model SH 86, Stihl Inc., Virginia Beach, VA, USA) equipped with a plastic tube was used to artificially create the windy conditions. The wind speed was assessed with a portable anemometer (Testo model 410-1, Testo, Lenzkirch, Germany) at each measurement with a precision of ± 0.2 m·s^{-1}. Consequently, under these artificial but controlled conditions we assessed the isolate effect of wind on the 3D models created with a RGB-D camera.
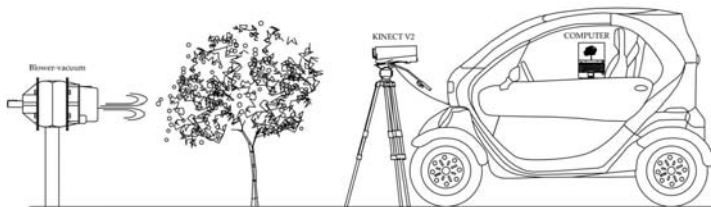
Figure 1. Schematic design of the portable system electrically powered at 220 V by an electric car.

2.2. Site Study and Field Measurements

Field experiments were conducted on tree plantations located at the experimental farm "La Poveda" (Arganda del Rey, Madrid, Central Spain) during May 2016 and April 2017. Measurements

were taken at direct sunlight in a sunny day (approx. 40,000 lux) and without any shading structure. Two tree species with contrasting shape and plant structure were considered: poplar (*Populus* spp.) and plum (*Prunus domestica* L.).These two species were selected for their agronomic importance in the study area, which were naturally available on the experimental farm. The first study was conducted on poplar grown as short rotation coppice, with 1-year old trees, scanning a total of fourteen trees at different wind speeds. The distance between trees was 0.5 m. Trees were selected randomly within the field, but including a representative sample of the population (within 95% confidence interval of mean plant height; data not shown). The adjacent trees were separated with solid wooden sheets in order to avoid interferences from neighboring tree leaves in the model. At sampling time, poplar heights ranged from 0.4 m to 0.6 m. A second study was conducted on a 2-year old plum orchard following the same procedures described previously. Fourteen trees were randomly selected covering the variability in height present within the field.

After the Kinect measurements, the actual height of every tree was manually measured. Thereafter, trees were cut for determination biomass and Leaf Area (LA). Biomass values were determined by drying the samples (stems plus leaves) of the ten trees in an oven at 78 °C during 48 h and weighting the dry biomass. The LA was calculated from the set of images by placing all the leaves on a white surface. As a reference surface, a standard 100 cm^2 black square was also placed in the image in order to calculate, by correlation, the leaf area of each sample (Figure 2). The images were obtained with a Canon EOS 7D (Canon, Tokyo, Japan) camera fitted with a 50 mm lens. The RGB images were transformed to binary images. A linear combination of the RGB planes with coefficients (r = − 0.884, g = 1.262, b = − 0.311) was performed. The applied coefficients were obtained by a genetic algorithm optimization process [34] that proved to perform better than Excess Green coefficients (ExG = 2G-R-B) [35]. In the resulting grey level image, the green objects (plants) appear bright, in contrast to objects with a different color which appear dark. Then, the Otsu's thresholding method [36] was applied to separate the objects pixel-wise into foreground (plants) and background in a binary image. From this image, the values marked as black denoted leaf area and those pixels off the range, which corresponded to white, denoted the background.

(**a**) (**b**)
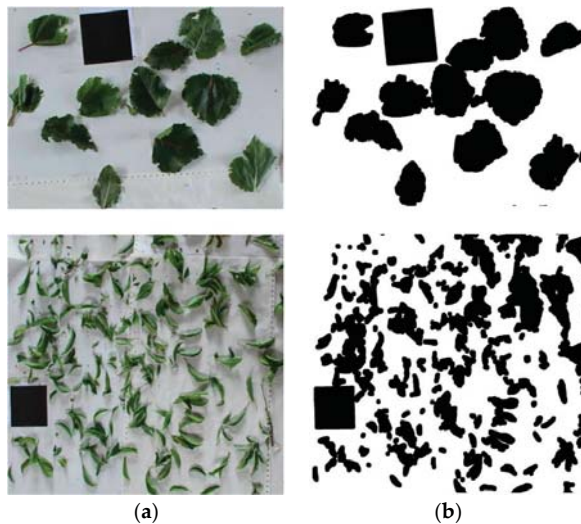
Figure 2. RGB images (**a**) used to quantify the leaf area, after their transformation to binary images (**b**) and subsequent application of the Otsu's thresholding method. Upper side corresponds to a poplar sample. Down side corresponds to a plum-tree sample. A 100 cm^2 black square was included in each image as reference area.

2.3. Data Processing

The raw data were recorded using Kinect Studio in video mode during a whole period including every wind speed. Thereafter, the file was divided in sections of time corresponding to each wind speed. The meshes were then reconstructed using Kinect fusion algorithm (Microsoft, Redmond, WA, USA). The parameters selected corresponded to Volume Max Integration Weight = 300, Volume Voxels per meter = 128 and Volume Voxels Resolution = 384. From the processed mesh, the desired parameters were calculated by offline processing using the open software Meshlab® (University of Pisa, Italy) (Figure 3). The processing tools incorporated into the software use filters to remove duplicated points, unreferenced vertices, and null faces. The software allows the cleaning, smoothing, visualization and processing of the acquired data readings and plots them as new meshes. The point cloud was processed in two steps [21,37].

(a) Data outliers and noise in the point cloud were filtered out. The filter identifies the points with no connection (1 cm out of the grid and removes them. (b) Tree volumes were calculated from the mesh by computation of polyhedral Mass properties [37]. The algorithm locates the body's center of mass and computes the moment of inertia about various axes in a dynamic simulation in rigid bodies considering a uniform density. The algorithm is based on a reduction of the volume integrals, minimizing the errors that result from poorly conditioned alignment of polyhedral faces. The volume integrals of a polyhedron are simultaneously computed in a single walk over the boundary of the polyhedron. The algorithm computed the normal, weighting the product over the k-nearest neighbors. The obtained isosurface is approximated from the normal field, whose gradient is related to the indicator function that describes the isosurface. Then, the mesh volume was obtained.
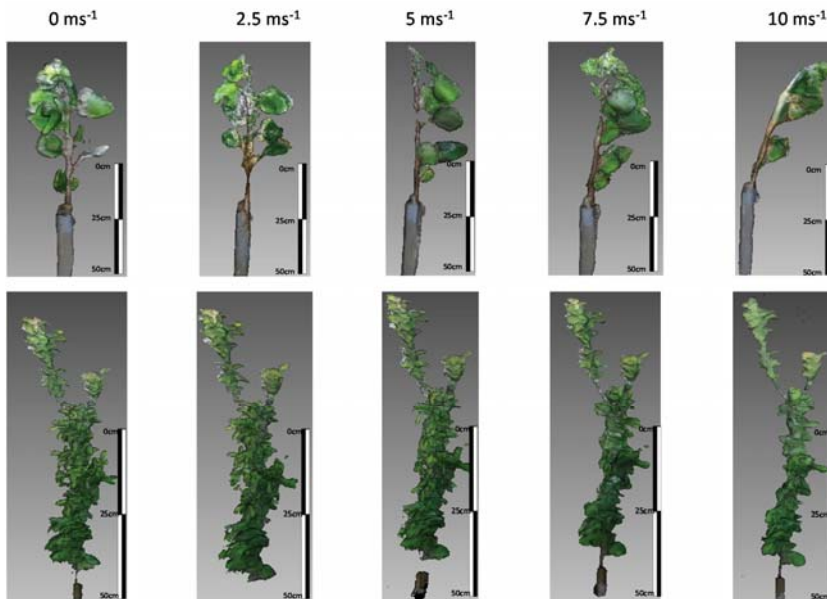
Figure 3. Example of poplar (figures on top) and plum (figures at the bottom) tree models created at different wind speeds, from 0 to 10 m·s^{-1}.

Information extracted from models was analyzed and compared with ground truth data. From the model, parameters regarding plant volume, maximum height and surface were extracted. These values were correlated with actual parameters of height, dry biomass and LA. Pearson's correlation coefficients

were used in the evaluation on simple linear relationships between the data from models and ground truth data. The analyses were repeated for both kinds of trees. The correlation analysis provided the initial information for the scatter plots comparing the cited parameters. The scatter diagrams provided the functional relationship existing between actual parameter and the models at different wind speeds.

3. Results and Discussion

In general, variability in Kinect measurements, noise and invalid pixels increased as wind speed increased (Figures 4–6). This was expected since the wind moved the trees during data acquisition, changing their position. At null and low wind speeds, the actual height, LA, and biomass showed good correlations with those estimated from the models. However, correlations decreased as wind speed increased.
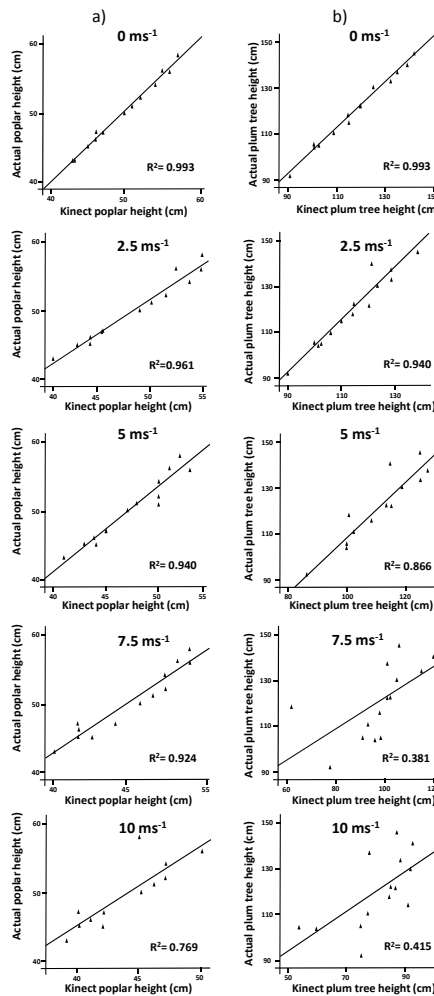
Figure 4. Regression analysis comparing actual height vs. model height for (**a**) poplar trees and (**b**) plum trees at wind speeds ranging from 0 to 10 m·s^{-1}.
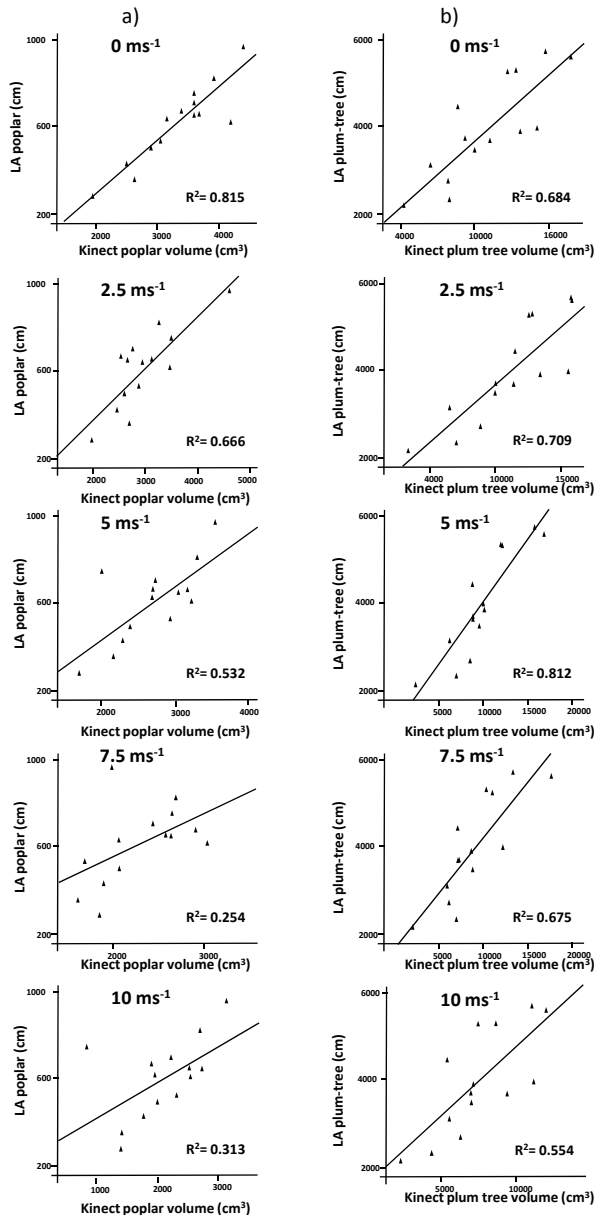
Figure 5. Regression analysis comparing Leaf Area (LA) vs. tree volume for (**a**) poplar trees and (**b**) plum trees at wind speeds ranging from 0 to 10 m·s^{-1}.

Actual poplar height ranged from 43 to 58 cm (average: 50.6 cm). These values were strongly correlated (r = 0.996) with those measured by the Kinect v2 under no wind conditions. When wind speed increased to 2.5, 5 and 7.5 m·s^{-1}, correlations decreased slightly (r = 0.980, r = 0.970 and r = 0.961, respectively). In contrast, the mean absolute error (MAE) increased from 0.21 in the absence of wind to 1.21 for 2.5 and 5 m·s^{-1}, and to 1.14 for 7.5 m·s^{-1}. Similarly, the root mean square errors (RMSEs)

increased with the wind speed (2.21, 2.14, 2.00 and 5.50 cm for 2.5, 5, 7.5 and 10 m·s^{-1}, respectively) relative to the windless scenario (0.21 cm). Nevertheless, actual and estimated measurements of poplar height were still significantly correlated (Figure 4a). In general, although Kinect measurements tended to slightly underestimate plant height when wind speed increased in a fixed direction, we can conclude that this parameter could be satisfactorily estimated by this method up to 7.5 m·s^{-1}.
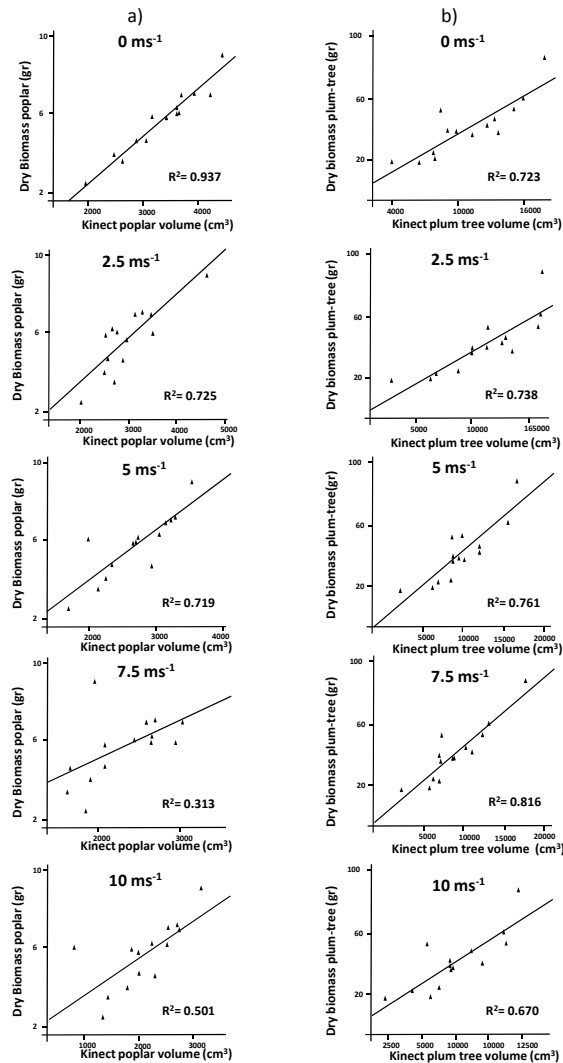
Figure 6. Regression analysis comparing dry biomass (g) vs. tree volume for (**a**) poplar trees and (**b**) plum trees at wind speeds ranging from 0 to 10 m·s^{-1}.

In the case of plum trees, a good agreement was observed between actual and model estimated tree height up to 5 m·s^{-1} wind speed. Actual tree height ranged from 91 to 142 cm (average: 118 cm). The Kinect v2 measurements achieved correlations of r = 0.996, r = 0.970 and r = 0.930 for wind speeds of 0, 2.5 and 5 m·s^{-1}, respectively. No significant correlations were obtained with higher wind

speeds. MAE resulted in values of 2.14, 3.71 and 4.21 cm in the absence of wind and for the two lower wind speeds. When speeds increased to 7.5 and 10 m·s^{-1}, MAE increased up to 12.21 and 17.07 cm. Regression analysis showed good fittings till wind speeds of 5 m·s^{-1} (Figure 4b). Apparently, plum trees were more affected by high wind speeds than poplar trees. The model showed good estimations until 5 m·s^{-1}. In general, plum-trees compared with poplar trees were easier affected by higher wind speeds.

The calculation of maximum height using Kinect v2 sensors may have various sources of errors. Wind moves the top of the trunk and the small leaves located in that area affecting the reconstruction of this parameter. This measurement could be improved by analyzing data without filters. However, under these conditions, noise and invalid pixels increase and volume calculation led to worse predictions. The errors associated with the Kinect sensor have been assessed for the previous [38] and the new version [39]. The first version of Kinect was unable to detect branches smaller than 6 mm width. The results were tested for distances of 0.5 to 1.25 m with similar results. In the case of Kinect v2, the range of measurement is higher and error tended to decrease. Under field and laboratory conditions and no influence of other external factors, absolute errors lower than 12 mm were obtained when cotton plants were measured [39]. The accuracy of the system depends not only on the sensor; the target object highly influences the accuracy of measurements. Azzary et al. [21] compared manual and Kinect v1 measurements of plant basal diameter and plant height, concluding that the underestimation or overestimation depends on the plant species and the characteristic of the plant to measure. In our study, models for large leaves were more accurate than those with small leaves. In general, solid objects create more accurate models than complex shapes such as trees. In a study conducted to estimate the dimensions of commercial onions, a 2 mm error in diameter was obtained [26]. The study compared the models obtained with Kinect v1 sensor and RGB cameras. Although both sensors underestimated onion dimensions, depth images created more accurate models than those obtained with RGB images. In a similar study, Andújar et al. [40] showed an error higher than 1 cm when cauliflower dimensions were measured with the first version of Kinect. Those errors could be reduced with Kinect v2, which have showed lower error. However they depend of several parameters such as object reflection, distance object-sensor, foreground, etc. [41]. Measuring the shapes of trees and other plants is an important challenge due to their complex shape.

The parameters extracted from 3D models allowed to characterize leaf area, photosynthetic active area, leaf orientation, dynamics of plant growth, etc. These parameters are directly related with phenotyping processes.

In poplar trees, LA was highly correlated with volume estimation in models created at 0 and 2.5 m·s^{-1} and correlation was still significant at 95% at 5 m·s^{-1}. Higher wind speed resulted in not significant correlations. The regressions between poplar volume and LA obtained from the RGB images indicate good agreements at wind speeds below 2.5 m·s^{-1} (Figure 5a). Above this speed wind impedes a good estimation of LA. Since, LA is a parameter directly related with plant growth and productivity and it is a good indicator for nutrient application, the use of sensors in precision applications in poplars should be avoided when wind conditions exceed 2.5 m·s^{-1}. In plum trees, model measurements improved significantly compared to poplar trees. Correlations were significant under all wind conditions. However, regression results suggest a decrease in the accuracy when wind speed exceeds 5 m·s^{-1} (Figure 5b). The differences in shape between poplar and plum trees created different patterns in the models. Plum trees were more elongated and internal parts of the trees were less occluded, improving LA estimation under windy conditions.

The actual biomass and the volume estimated using the Kinect v2 were consistent at different wind speeds for the two tree species studied (Figure 6). Poplar trees showed correlations of r = 0.968, r = 0.851 and r = 0.848 for wind speeds of 0, 2.5 and 5 m·s^{-1}, respectively. Regression analysis showed similar results. Tree biomass was not properly estimated when wind speeds exceeded 5 m·s^{-1} (Figure 6a). In plum trees, the relationship between actual biomass and the estimated volume was significant at 99% at all wind speeds. However, dispersion of values was greater at high wind speeds (Figure 6b).

Volume quantification using Kinect sensors has been tested previously [32]. These studies concluded that plant volume can be properly estimated although with some errors. However, when working with solid objects, such as onions or cauliflowers, estimated values were more similar to real objects [26,40]. When working with whole plants, many points of the cloud impact on internal parts, introducing some errors [2]. In order to reduce these errors, algorithms more accurate than those used for characterization of solid plant products are required.

Structural parameters extracted from the models generated by depth cameras such as Kinect must take into consideration the natural conditions present on the field. The used sensor should work properly outdoors when the scene has low ambient IR light. The Time-of-Flight distance measurement method is based on measuring the difference between two accumulators in the sensor, each one containing a portion of the returning IR light. However, when sunlight radiation is high the captured radiation on accumulators can be low and this limits the data output. In addition the accuracy depends on the angle of incidence of sunlight. However, several papers have been published evaluating the effect of sunlight in outdoors conditions [42]. Thus the study took constant values of light, time and angle of measurements coinciding with a sunny day at midday. Deformations caused by wind are different from those caused by illumination and they are more difficult to correct by software. In our study, we set up the Kinect Fusion algorithm for a good plant reconstruction at different wind speeds and we extracted values from the created model. The practical relevance of the information generated at high wind speeds is conditioned by the fact that, to avoid the drift of agrochemicals, spraying pesticides or liquid nutrients must be done when wind is almost null. This study confirms the need of considering every weather conditions, including wind speed, when monitoring vegetation outdoors.

4. Conclusions

This study has proved that a constant one-direction wind can influence the acquisition of visual depth information. This conclusion is supported by studies conducted with two different tree species (poplars and plum trees) at various wind speeds (up to 10 m·s^{-1}). Higher wind speed resulted in higher variability in Kinect v2 height and volume measurement due to tree movement. In general, this sensor was unable to obtain accurate 3D models when wind speed exceeded a 5 m·s^{-1} threshold. Although this threshold was similar for both tree species, individual cases should be studied when performing an outdoor study.

Branch flexibility, leave shape and size affect the fidelity of the 3D models. New algorithms processing maximum height while reducing invalid pixels are required in order to estimate plant volume and height from a single-shot model developed at high wind speeds.

Author Contributions: José Dorado and César Fernández-Quintanilla helped to design the experiments and provided material to the experiment; Ángela Ribeiro contributed by providing material to the experiment, analyzing data and revising the document; José María Bengochea-Guevara and Jesús Conesa-Muñoz helped during the experiment, Dionisio Andújar analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Duan, T.; Chapman, S.C.; Holland, E.; Rebtzke, G.J.; Guo, Y.; Zheng, B. Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *J. Exp. Bot.* **2016**, *67*, 4523–4534. [CrossRef] [PubMed]
2. Rosell-Polo, J.R.; Auat, F.; Gregorio, E.; Andújar, D.; Puigdomenech, J.M.; Escola, A. Advances in Structured Light Sensors Applications in Precision Agriculture and Livestock Farming. *Adv. Agron.* **2015**, *133*, 71–112.
3. Li, L.; Zhang, Q.; Huang, D. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* **2014**, *14*, 20078–20111. [CrossRef] [PubMed]

4.   Barbedo, A.; Garcia, J. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* **2016**, *144*, 52–60. [CrossRef]

5.   Anderson, H.B.; Nilsen, L.; Tommervik, H.; Karlsen, S.R.; Nagai, S.; Cooper, E.J. Using Ordinary Digital Cameras in Place of Near-Infrared Sensors to Derive Vegetation Indices for Phenology Studies of High Arctic Vegetation. *Remote Sens.* **2017**, *8*, 847. [CrossRef]

6.   Brandão, Z.; Zonta, J. Hemispherical photography to estimate biophysical variables of cotton. *Revista Brasileira de Engenharia Agrícola e Ambiental* **2016**, *20*, 789–794. [CrossRef]

7.   Cui, R.X.; Liu, Y.D.; Fu, J.D. Estimation of Winter Wheat Leaf Nitrogen Accumulation using Machine Learning Algorithm and Visible Spectral. *Spectrosc. Spect. Anal.* **2016**, *36*, 1837–1842.

8.   Schirrmann, M.; Giebel, A.; Gleiniger, F.; Pflanz, M.; Lentschke, J. Monitoring Agronomic Parameters of Winter Wheat Crops with Low-Cost UAV Imagery. *Remote Sens.* **2016**, *8*, 706. [CrossRef]

9.   Herrera, P.J.; Dorado, J.; Ribeiro, Á. A novel approach for weed type classification based on shape descriptors and a fuzzy decision-making method. *Sensors* **2014**, *14*, 15304–15324. [CrossRef] [PubMed]

10.   Campos, Y.; Sossa, H.; Pajares, G. Spatio-temporal analysis for obstacle detection in agricultural videos. *Appl. Soft Comput.* **2016**, *45*, 86–97. [CrossRef]

11.   Vázquez-Arellano, M.; Griepentrog, H.W.; Reiser, D.; Paraforos, D.S. 3-D Imaging Systems for Agricultural Applications—A Review. *Sensors* **2016**, *16*, 618. [CrossRef] [PubMed]

12.   Li, H.Z.; Zhai, C.Y.; Weckler, P.; Wang, N.; Yang, S.; Zhang, B.A. Canopy Density Model for Planar Orchard Target Detection Based on Ultrasonic Sensors. *Sensors* **2017**, *17*, 31. [CrossRef] [PubMed]

13.   Andújar, D.; Escola, A.; Dorado, J.; Fernandez-Quintanilla, C. Weed discrimination using ultrasonic sensors. *Weed Res.* **2011**, *51*, 543–547. [CrossRef]

14.   Andújar, D.; Weis, M.; Gerhards, R. An ultrasonic system for weed detection in cereal crops. *Sensors* **2012**, *12*, 17343–17357. [CrossRef] [PubMed]

15.   Llorens, J.; Gil, E.; Llop, J.; Escolà, A. Ultrasonic and LIDAR sensors for electronic canopy characterization in vineyards: Advances to improve pesticide application methods. *Sensors* **2011**, *11*, 2177–2194. [CrossRef] [PubMed]

16.   Andújar, D.; Escolà, A.; Rosell-Polo, J.R.; Sanz, R.; Fernández-Quintanilla, C.; Ribeiro, A.; Dorado, J. A LiDAR-based system to assess poplar biomass. *Gsunde Planz* **2016**, *68*, 155–162. [CrossRef]

17.   Dutta, D.; Wang, K.X.; Lee, E.; Goodwell, A.; Woo, D.K.; Wagner, D.; Kumar, P. Characterizing Vegetation Canopy Structure Using Airborne Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1160–1178. [CrossRef]

18.   Wei, S.S.; Fang, H.L. Estimation of canopy clumping index from MISR and MODIS sensors using the normalized difference hotspot and darkspot (NDHD) method: The influence of BRDF models and solar zenith angle. *Remote Sens. Environ.* **2016**, *187*, 476–491. [CrossRef]

19.   Gibbs, J.A.; Pound, M.; French, A.P.; Wells, D.M.; Murchie, E.; Pridmore, T. Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct. Plant Biol.* **2016**, *44*, 62–75. [CrossRef]

20.   Khakimov, B.; Engelsen, S.B. High-throughput cereal metabolomics: Current analytical technologies, challenges and perspectives. *J. Cereal Sci.* **2014**, *59*, 393–418. [CrossRef]

21.   Azzari, G.; Goulden, M.L.; Rusu, R.B. Rapid Characterization of Vegetation Structure with a Microsoft Kinect Sensor. *Sensors* **2013**, *13*, 2384–2398. [CrossRef] [PubMed]

22.   Paulus, S.; Behmann, J.; Mahlein, A.K.; Plümer, L.; Kuhlmann, H. Low-cost 3D systems: Suitable tools for plant phenotyping. *Sensors* **2014**, *14*, 3001–3018. [CrossRef] [PubMed]

23.   Chéné, Y.; Rousseau, D.; Lucidarme, P.; Bertheloot, J.; Caffier, V.; Morel, P.; Belin, E.; Chapeau-Blondeau, F. On the use of depth camera for 3D phenotyping of entire plants. *Comput. Electron. Agric.* **2012**, *82*, 122–127. [CrossRef]

24.   Andújar, D.; Fernández-Quintanilla, C.; Dorado, J.; Ribeiro, A. An Approach to the Use of Depth Cameras for Weed Volume Estimation. *Sensors* **2016**, *16*, 972. [CrossRef] [PubMed]

25.   Andújar, D.; Fernández-Quintanilla, C.; Dorado, J. Matching the Best Viewing Angle in Depth Cameras for Biomass Estimation Based on Poplar Seedling Geometry. *Sensors* **2015**, *15*, 12999–13011. [CrossRef] [PubMed]

26.   Wang, W.; Li, C. Size estimation of sweet onions using consumer-grade RGB-depth sensor. *J. Food Eng.* **2014**, *142*, 153–162. [CrossRef]

27.   Hernandez-Aceituno, J.; Arnay, R.; Toledo, J.; Acosta, L. Using Kinect on an Autonomous Vehicle for Outdoors Obstacle Detection. *IEEE Sens. J.* **2016**, *16*, 3603–3610.

28. El-Iaithy, R.A.; Huang, J.; Yeh, M. Study on the use of Microsoft Kinect for robotics applications. In Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium, Myrtle Beach, SC, USA, 23–26 April 2012; pp. 1280–1288.

29. Butkiewicz, T. Low-cost coastal mapping using Kinect v2 time-of-flight cameras. In Proceedings of the Oceans, St. John's, NL, Canada, 14–19 September 2014; pp. 1–9.

30. Fankhauser, P.; Bloesch, M.; Rodriguez, D.; Kaestner, R.; Hutter, M.; Siegwart, R. Kinect v2 for mobile robot navigation: Evaluation and modeling. In Proceedings of the 2015 International Conference on Advanced Robotics (ICAR), Istambul, Turkey, 27–31 July 2015; pp. 388–394.

31. Lachat, E.; Macher, H.; Mittet, M.A.; Landes, T.; Grussenmeyer, P. First Experiences with Kinect v2 Sensor for Close Range 3D Modelling. In Proceedings of the 2015 International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Avila, Spain; 2015; pp. 93–100.

32. Niessner, M.; Zollhöfer, M.; Izadi, S.; Stamminger, M. Real-time 3D reconstruction at scale using voxel hashing. *ACM Trans. Graph. (TOG)* **2013**, *32*, 169. [CrossRef]

33. The Open Source Computer Vision (Open CV). Available online: http://opencv.org (accessed on 19 December 2016).

34. Burgos-Artizzu, X.P.; Ribeiro, A.; Tellaeche, A.; Pajares, G.; Fernández-Quintanilla, C. Analysis of natural images processing for the extraction of agricultural elements. *Image Vis. Comput.* **2010**, *28*, 138–149. [CrossRef]

35. Woebbecke, D.M.; Meyer, G.E.; Von Bargen, K.; Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* **2015**, *38*, 259–269. [CrossRef]

36. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

37. Mirtich, B. Fast and Accurate Computation of Polyhedral Mass Properties. *J. Graph. Tools* **1996**, *1*, 31–50. [CrossRef]

38. Nock, C.A.; Taugourdeau, O.; Delagrange, S.; Messier, C. Assessing the potential of low-cost 3D cameras for the rapid measurement of plant woody structure. *Sensors* **2013**, *13*, 16216–16233. [CrossRef] [PubMed]

39. Jiang, Y.; Li, C.; Paterson, A.H.; Roberts, J.S. An imaging-based system for high-throughput phenotyping of cotton plants under field conditions. In Proceedings of the 2016 ASABE Annual International Meeting, Orlando, FL, USA, 17–20 June 2016.

40. Andújar, D.; Ribeiro, R.; Fernández-Quintanilla, C.; Dorado, J. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Comput. Electron. Agric.* **2016**, *122*, 67–73. [CrossRef]

41. Steward, J.; Lichti, D.; Chow, J.; Ferber, R.; Osis, S. Performance Assessment and Calibration of the Kinect 2.0 Time-of-Flight Range Camera for Use in Motion Capture Applications. In Proceedings of the Fig Working Week 2015, Sofia, Bulgaria, 17–21 May 2015.

42. Pagliari, D.; Pinto, L. Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. *Sensors* **2015**, *15*, 27569–27589. [CrossRef] [PubMed]

Development of a Spectrophotometric System to Detect White Striping Physiopathy in Whole Chicken Carcasses

Maria Victoria Traffano-Schiffo [1], Marta Castro-Giraldez [1], Ricardo J. Colom [2] and Pedro J. Fito [1,*]

[1] Instituto Universitario de Ingeniería de Alimentos para el Desarrollo, Universitat Politecnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain; matrasc@upv.es (M.V.T.-S.); marcasgi@upv.es (M.C.-G.)

[2] Instituto de Instrumentación para Imagen Molecular, Universitat Politecnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain; rcolom@eln.upv.es

* Correspondence: pedfisu@tal.upv.es; Tel.: +34-63-699-8785; Fax: +34-96-387-9832

Abstract: Due to the high intensification of poultry production in recent years, white chicken breast striping is one of the most frequently seen myopathies. The aim of this research was to develop a spectrophotometry-based sensor to detect white striping physiopathy in chicken breast meat in whole chicken carcasses with skin. Experiments were carried out using normal and white striping breasts. In order to understand the mechanism involved in this physiopathy, the different tissues that conform each breast were analyzed. Permittivity in radiofrequency (40 Hz to 1 MHz) was measured using two different sensors; a sensor with two flat plates to analyze the whole breast with skin (NB or WSB), and a two needles with blunt-ended sensor to analyze the different surface tissues of the skinless breast. In the microwave range (500 MHz to 20 GHz), permittivity was measured as just was described for the two needles with blunt-ended sensor. Moreover, fatty acids composition was determined by calorimetry techniques from −40 °C to 50 °C at 5 °C/min after previously freeze-drying the samples, and pH, microstructure by Cryo-SEM and binocular loupe structure were also analyzed. The results showed that the white striping physiopathy consists of the partial breakdown of the pectoral muscle causing an increase in fatty acids, reducing the quality of the meat. It was possible to detect white striping physiopathy in chicken carcasses with skin using spectrophotometry of radiofrequency spectra.

---

1. Introduction

In recent decades, the overall consumption of chicken and turkey meat has increased considerably and it is expected that in the coming years, chicken will become the most produced type of meat in the world. It has been estimated that in 2020, global chicken meat production will be around 122.5 million tons [1]. The two main reasons that are driving the success of poultry meat are the low price and the healthy nutritional profile compared to pork and beef.

As a result of increasing demand, there has been an intensification in poultry production, where modern hybrid broilers show a pectoral development greater than 20% of the body weight [2]. This production intensification has increased the incidence of abnormalities in the pectoralis muscle [3–5]. The most common malformations are deep pectoral myopathy [6,7] and white striping [8,9].

White striping disease is a serious, emerging issue characterized by the appearance of white stripes (WS) parallel to the muscle fiber on the surface of the *pectoralis major* muscle [10]. Some authors

have classified the incidence into three categories, according to the intensity and thickness of the WS: NORMAL when the breast is not affected, MODERATE when the thickness of the WS is less than 1 mm and SEVERE when the WS covers most of the surface and the thickness is greater than 1 mm [11,12].

The presence of WS on the surface of chicken breasts affects the visual appearance of the product and decreases the degree of consumer acceptance [11]. However, the negative visual impact is not the only problem; white striping also affects the chemical composition: protein content decreases as the degree of affection increases, while there is an opposite trend in fat content, thus the presence of this hypertrophy decreases the nutritional value of the meat [13].

For this reason, white striping breasts (WSB) should not be commercialized as normal breast (NB), and should be used in the formulation of meat products such as sausages and nuggets, however, when the chicken is sold whole, the presence of the skin means that it is not possible to see the effects of this disorder. Considering this issue, it is necessary to find a reliable, efficient industrial system to detect the presence of white striping; for this reason, non-invasive sensors based on spectrophotometry at range of radiofrequency and microwaves could be a good tool to meet this challenge.

Dielectric properties expressed as permittivity can be explained as vector, polar or complex numbers. As a complex number, permittivity (ε) is composed of two terms, the dielectric constant ε' and the loss factor ε'', which are the real and imaginary terms of permittivity, respectively. The dielectric constant is related to the tissue's ability to absorb and store electric energy, and the loss factor is related to the dissipation of the electric energy into other energies such as thermal or mechanical energies.

In the radiofrequency and microwave range, the interaction of the photon flux with biological tissue produces three main dispersions: α, β, and γ [14]. In particular, α-dispersion (from a few Hz to a few kHz), also called the counterion effect, is induced by the orientation of mobile charges in a dielectric medium [15]. β-Dispersion (from kHz to tens of MHz) is related to the orientation of fixed charges in macromolecules such as proteins [16]. At higher frequencies of the radiofrequency range, β-dispersion results due to the surface tension charges, this phenomenon is called the Maxwell-Wagner effect. Finally, in the microwave range, γ-dispersion occurs at GHz frequencies and is due to the orientation and induction of dipolar molecules such as water [17–19]. Another important effect in the microwave range is ionic conductivity. It only affects the loss factor, as it produces a repulsion of charged molecules, transforming electric energy into other types [20].

The usefulness of spectrophotometry at low frequencies in the food industry has been demonstrated as a monitoring technique in a wide range of the permittivity spectra: radiofrequency, microwave and infrared (from Hz to THz). Talens et al. [20] developed a dielectric isotherm technique capable of predicting the water activity in dried orange peel using the ε' at 20 GHz (microwave range); Shang, Guo, and Nelson [21] identified the apple varieties in the radiofrequency range. Castro-Giráldez et al. [22] demonstrated the usefulness of two dielectric ageing indexes at different frequencies (140 Hz, 500 Hz and 300 kHz) to determine pork meat ageing. Also, Trabelsi and Roelvink [23] and Damez et al. [24] demonstrated that dielectric spectroscopy is able to predict chicken and beef ageing using microwave and radiofrequency ranges, respectively. Finally, Cuibus et al. [25] and Traffano-Schiffo et al. [26] demonstrated that infrared is a good non-destructive technique to monitor freezing and drying processes.

The aim of this research was to develop a sensor based in spectrophotometry to detect white striping physiopathy in chicken breast meat in whole chicken carcass with skin.

2. Experimental

2.1. Raw Material

The experiments were carried out using boneless broiler breasts (*pectoralis major*) obtained from the SADA Group slaughterhouse located in Rafelbunyol, Valencia, Spain. After slaughter, male broilers of 42 d were bled, plucked, tempered in a cooling tunnel at 4 °C for 3 h and finally dismembered. Breasts were classified by the quality experts of the slaughterhouse plant, according to

the appearance of severe WS (WSB) or NB. The selected breasts (including the skin) were transported using an isothermal bag with ice blocks to the laboratory of the Institute of Food Engineering for Development (IuIAD) at the Polytechnic University of Valencia (UPV). The samples were measured at 12 h post-mortem and maintained at 2 °C during the experimental procedure.

2.2. Experimental Procedure

Experiments were carried out using 40 chicken breasts: 20 were classified as NB and 20 as WSB. Prior to the dielectric measures, pH analyses of the samples were carried out using a punch pH-meter S-20 SevenEasy™ (Mettler Toledo, Barcelona, Spain) to characterize the samples according to their quality. Permittivity of the samples was measured in the surface of the breasts (ventral side) in radiofrequency and microwave ranges (non-destructive measurements). In radiofrequency range, firstly, a sensor of two flat plates with circular surfaces was used to measure the permittivity of whole breast with skin (NB or WSB). After that, a two needles with blunt-ended sensor was used to measure the permittivity of each tissue of the skinless breast (muscle and adipose tissues of NB and WSB, and white stripes of WSB). In the microwave range, permittivity was measured as described above for the two needles with blunt-ended sensor. Subsequently, with the aim of analyzing the state of the fatty acids, a differential scanning calorimetry (DSC) study was performed by triplicate for samples of the ventral portion of NB and WSB. Finally, a macro and microstructural study of NB and WSB was performed using a binocular loupe and a low-temperature scanning electron microscope, respectively.

2.3. Structural Analysis

2.3.1. Low-Temperature Scanning Electron Microscopy (Cryo-SEM)

The microstructure of NB and WSB was analyzed by Cryo-SEM. A Cryostage CT-1500C unit (Oxford Instruments, Witney, UK), coupled to a JSM-5410 scanning electron microscope (Jeol, Tokyo, Japan), was used. The sample was immersed in N_2 slush (-210 °C) and then quickly transferred to the Cryostage at 1 kPa where sample fracture took place. Sublimation (etching) was carried out at -95 °C. The final point was determined by direct observation under the microscope, working at 5 kV. Then, once again in the Cryostage unit, the sample was coated with gold under vacuum (0.2 kPa) applied for 3 min, with an ionization current of 2 mA. The observation in the scanning electron microscope was carried out at 20 kV, at a working distance of 15 mm and temperature ≤ -130 °C.

2.3.2. Optical Measurements

Optical measurements were made with a Leica MZ APO™ binocular loupe (Leica Microsystems, Wetzlar, Germany) with low magnification ($8\times$ to $80\times$) using incident light illumination (light reflected off the surface of the sample). It uses two separate optical paths with two objectives and two eyepieces to provide slightly different viewing angles for the left and right eyes. In this way it permits a three-dimensional visualization of the sample.

2.4. Permittivity Measurements

2.4.1. Radiofrequency Range

As explained above, two different sensors were used to carry out this experimental work (see Figure 1). One of them consists of two flat plates with circular surfaces sensor (Figure 1a) and it was used to measure the whole breast with skin (NB or WSB) (this configuration has high penetration depth). The other sensor consists of two needles with blunt-ended (Figure 1b) and it was used to measure the different tissues of surface breast (this configuration has low penetration depth): muscle and adipose tissues in NB and WSB, and WS in WSB. In case of the two needles sensor, its size was small enough to touch the specific tissue, and its penetration depth was low enough just to include the specific tissue. In WS and in adipose tissue, the globular conformation of the fat adipocytes,

similar in the three axes, does not affect to the displacement of photon flux. However, in case of muscle, the mobility of liquid phase in the direction of the fibers changes the displacement of photon flux if it is compared with the perpendicular direction. Therefore, it was measured in perpendicular direction of fibers. It is important to highlight that the muscle tissue of WSB is referred to the muscle between the WS.

Both sensors were developed by The Institute of Food Engineering for Development (IuIAD) and The Institute for Molecular Imaging Technologies (I3M), both at the Polytechnic University of Valencia. The sensors were connected to an 4694A impedance analyzer (Agilent, Santa Clara, CA, USA). Permittivity were estimated using Equations (1)–(3). The frequency range measured was from 40 Hz to 1 MHz. Calibration of the equipment was performed in open (air) and short-circuit.

The signal obtained by the Agilent analyser is the impedance Z, and taking into account that the impedance as a complex number is $\overline{Z} = R + jX$, where the real part of the impedance is the resistance R and the imaginary part is the reactance X. The parameters R and X were transformed in ε', ε'' as follows:

$$\varepsilon' = \frac{-X}{(R^2 + X^2)} \frac{1}{2\pi C_0} \tag{1}$$

$$\varepsilon'' = \frac{R}{R^2 + X^2} \frac{1}{2\pi f C_0} \tag{2}$$

$$C_0 = \frac{\varepsilon_0 \, S}{d} \tag{3}$$

where f is the frequency (Hz), C_0 is the capacitance in the vacuum (F), S is the surface of the electrodes (m^2), ε_0 is the vacuum permittivity (F/m) and d is the separation between the electrodes with differential tension ($V_H - V_L$) (m).
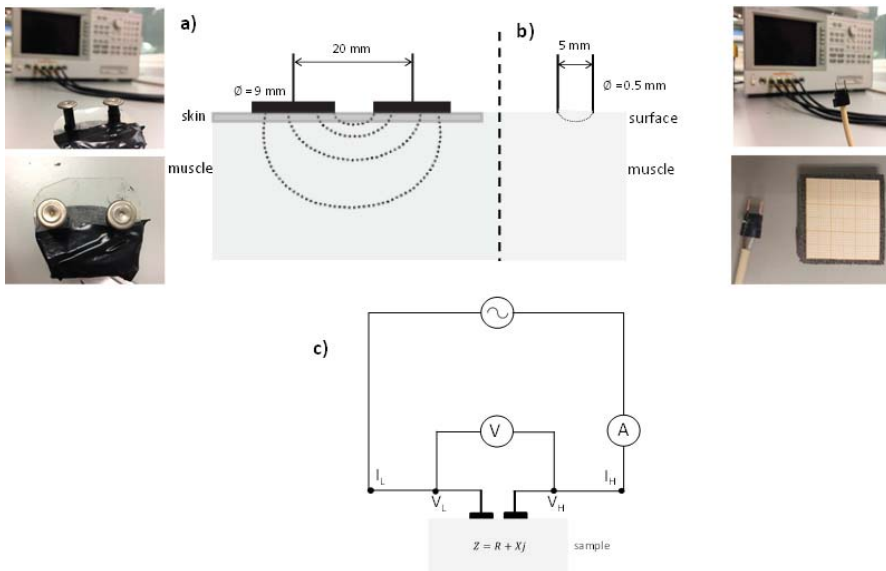
Figure 1. Permittivity measurement in radiofrequency, detail of each sensor: (**a**) two flat plates sensor to measure the whole breast with skin; (**b**) two needles with blunt-ended sensor to determine the permittivity in surface of the breast and (**c**) electric circuit of both sensors.

2.4.2. Microwave Range

Permittivity of the samples in the microwave range was measured with an Agilent 85070E open-ended coaxial probe connected to an Agilent E8362B Vector Network Analyzer. The system was calibrated using three different types of loads: open (air), short-circuit and 4 °C Milli®-Q water. Once the calibration was carried out, 4 °C Milli®-Q water was measured again to check calibration suitability. All determinations were made from 500 MHz to 20 GHz. All permittivity measurements were performed in triplicate.

2.5. Differential Scanning Calorimetry (DSC)

The thermal transitions of the fatty acids was performed according to the method proposed by Benedito et al. [27] using a Mettler Toledo DSC 1 differential scanning calorimeter (Mettler Toledo, Barcelona, Spain) provided with a full range temperature sensor FRS5. The calibration of the equipment was performed with the automatic calibration function FlexCal supplied by the manufacturers. Samples taken from the surface of tissue (around 22–25 mg) were introduced into aluminium pans (Mettler Toledo, ME-00026763). Water was removed by lyophilisation for 48 h using a lyophiliser Telstar Lyoalfa-6 (Telstar, Dewsbury, U.K.) as it can interfere with the fat melting in the DSC curve. The pans were hermetically sealed and an empty one was used as the reference sample. Before obtaining the DSC curves, the samples were tempered at 60 °C for 5 min, cooled to −40 °C at 10 °C/min and held at −40 °C for 5 min. The DSC curves were obtained by heating the sample from −40 °C to 50 °C at a rate of 5 °C/min.

2.6. Statistical Analysis

The statistical analysis was carried out with the Statgraphics Centurion XVI software (Statgraphics, Warrenton, VA, USA). One-Way ANOVA analyses were made in order to find statistically significant differences between the parameters studied for the different samples. The logistic model of modified Gompertz was fitted by using nonlinear regression.
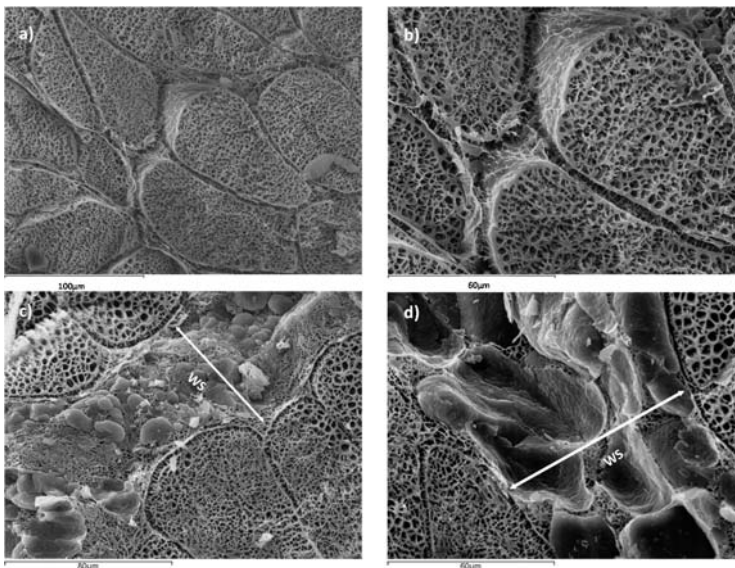
Figure 2. Micrographs of normal breast (NB) (**a**) 500×; (**b**) 1000× and breast affected by white striping (WSB), with the detail of the white stripe (WS) (**c**) 750× and (**d**) 1000× at 12 h post-mortem.

3. Results and Discussion

All the samples presented a pH value between 5.7 and 6.1, which were classified as NORMAL according to the Zhang and Barbut [28] classification. In Figure 2, a microstructural analysis of NB and WSB samples is shown. Figure 2a,b correspond to NB, where muscle tissue with correct packaging of the myofibrils is observed. However, in Figure 2c,d, it is possible to observe the deposition of adipocytes, conforming a new adipose tissue surrounded by muscle tissue, defined as white stripe (WS). This deposition of adipocytes is produced in areas which have suffered muscle breakdown. Other authors suggested that there is also an accumulation of collagen in the WS [13,29].

Figure 3 shows a detail of WS in chicken breast, where a muscular rupture, from the surface breast to the internal muscle, can be appreciated, induced by forces perpendicular to the fissure, and filled with fat tissue, in order to maintain the breast integrity. Thus, the accelerated muscular growth induced by pectoral hypertrophy has caused a partial muscular rupture and the deposition of adipose tissue in this area.
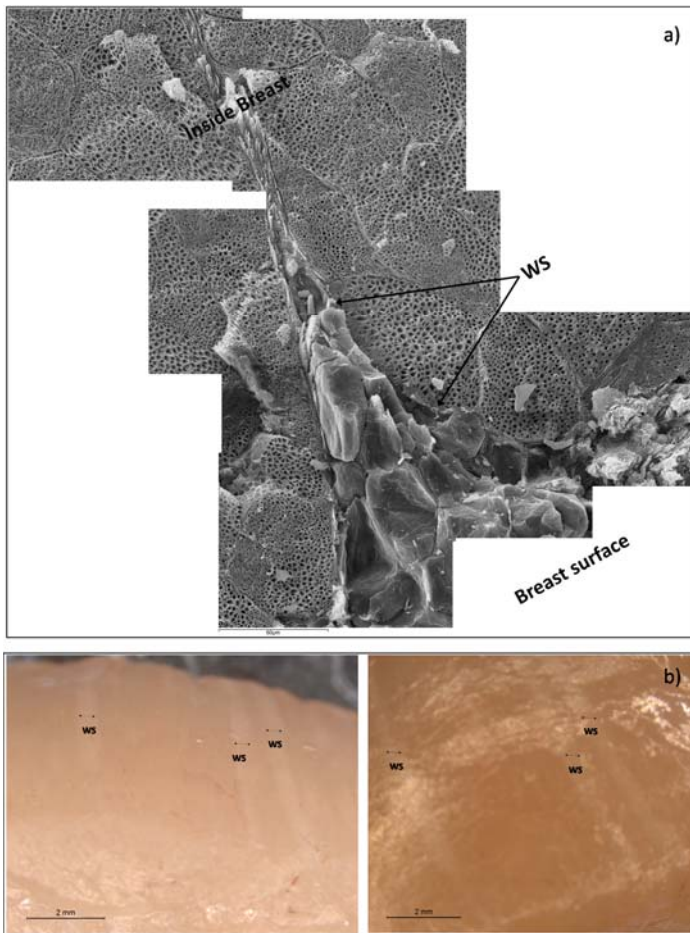
Figure 3. (**a**) Micrographs showing the detail of a white stripe (WS) (1000×); (**b**) binocular microscope images of WS in surface breast samples (8×).

Adipose tissue maintains the cohesion between muscle fibers and the elasticity of muscle tissue and therefore its activity of contraction-relaxation, conferring to the muscle the ability to transmit mechanical tension.

In addition, it is possible to appreciate that tissue breakdown is generated from the outside to the inside, so the quantity of adipose tissue is higher at the surface of the breast than the interior. Major muscle breakdown is in the external surface because the maximum tension is caused by the pectoral activity of poultry during flapping. This phenomenon produces a *pectoralis major* expansion, being maximum in the surface (reaching in this area the breakdown tension level).

Compared to NB, the WSB contains higher quantities of fatty acids, which cause the diminution of the protein content [30]. In particular, collagen increases due to hypertrophy as it is necessary to maintain the muscle structure, however, sarcoplasmic and myofibrillar proteins decrease [13].

The fatty acids of WSB and NB, in the surface, were measured by differential scanning calorimetry. The thermograms obtained show four transitions for the WSB (Figure 4a). According to the melting temperatures, it was possible to relate each transition to the different groups of fatty acids [31,32], where polyunsaturated fatty acids (PUFA) show a melting point in the range between $-17\,^\circ$C and $-5\,^\circ$C; monounsaturated fatty acids (MUFA) between $-5\,^\circ$C and $27\,^\circ$C and saturated fatty acids (SFA) between $27\,^\circ$C and $40\,^\circ$C.
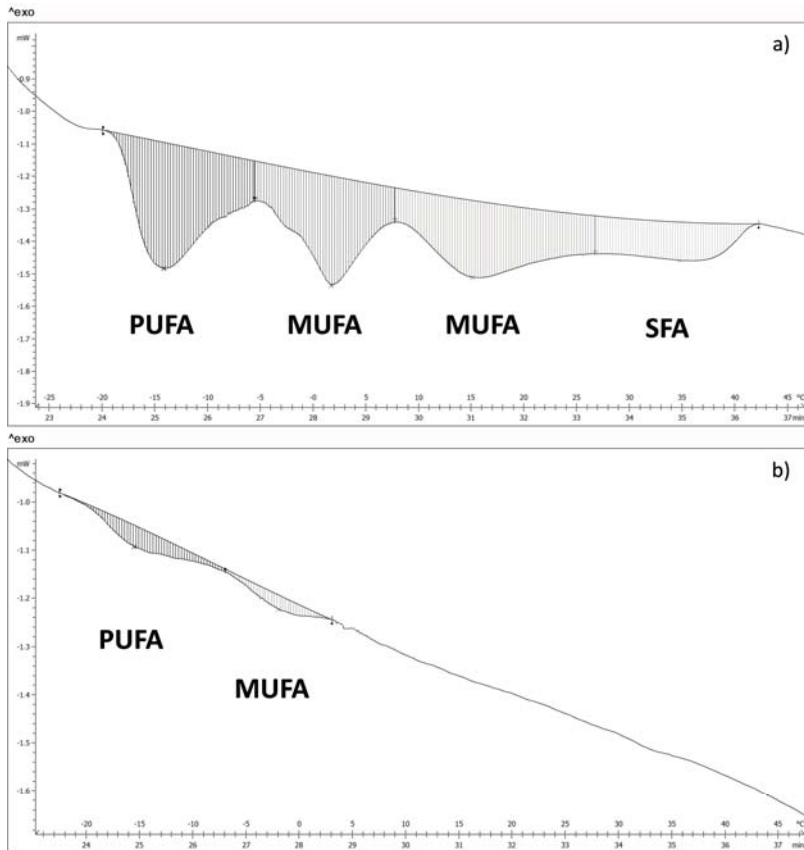
Figure 4. DSC thermogram of (**a**) white striping breast and (**b**) normal breast.

Figure 4b shows a thermogram corresponding to a NB, where it is possible to appreciate that there are only two transition peaks associated with PUFA and MUFA, which suggests that there is a big difference in fatty acids content between NB and WSB. Energy values and the temperature of the different transitions for both groups of samples are shown in Tables 1 and 2, respectively.

Table 1. Energy (J/g) and temperatures (°C) values of the fatty acids transitions of the white striping samples. Where: T_0 corresponds to the initial transition temperature, T_p peak transition temperature and T_f the final transition temperature.

	1st Transition	2nd Transition	3rd Transition	4th Transition
E (J/g)	3.3 ± 1.3	2.9 ± 1.3	3.8 ± 1.9	1.3 ± 0.5
T_0 (°C)	17 ± 2	4.6 ± 1.7	5 ± 3	27 ± 6
T_P (°C)	12.6 ± 1.6	0.9 ± 1.8	13.68 ± 0.96	35.0 ± 1.9
T_f (°C)	5 ± 2	4 ± 3	27 ± 5	39.9 ± 1.4

Table 2. Energy (J/g) and temperatures (°C) values of the fatty acids transitions of the normal samples. Where: T_0 corresponds to the initial transition temperature, T_p peak transition temperature and T_f the final transition temperature.

	1st Transition	2nd Transition	3rd Transition	4th Transition
E (J/g)	0.57 ± 0.04	0.24 ± 0.06	-	-
T_0 (°C)	20 ± 2	7.3 ± 0.2	-	-
T_P (°C)	15.1 ± 0.4	1.5 ± 0.6	-	-
T_f (°C)	8.1 ± 1.5	3.5 ± 0.6	-	-

According to Kuttappan et al. [31] and Knothe and Dunn [32], the majority of fatty acids per groups are PUFA: linoleic acid (18:2n-6), MUFA: palmitoleic (16:1c) and oleic acids (18:1c), and finally SFA: palmitic acid (16:0). The transition temperatures of these fatty acids correspond to the temperatures 1st, 2nd, 3rd and 4th, respectively, from Tables 1 and 2.

Mass fraction of fatty acids of NB and WSB (Table 3) were obtained from the transition energies and the latent heat of fusion of the major fatty acids in each transition [32] according to Equation (4):

$$x_{fa} = \frac{E}{\Delta H^f} \tag{4}$$

Being x_{fa} the mass fraction of fatty acids (kg/kg), E the transition energy of specific fatty acids group (J/g) and the ΔH^f the latent heat of fusion of specific fatty acid (J/g).

As can be appreciated in the table, WSB show a much higher total fatty acid content than NB.

Table 3. Mass fractions of fatty acids performed by DSC in normal (NB) and white striping (WSB) samples.

Fatty Acid	Chain	ΔH^f * (J/g)	NB x_{fa} (kg/kg)	WSB x_{fa} (kg/kg)
SFA				
Palmitic Acid	16:0	208.2	0	0.0050 ± 0.0013
MUFA				
Oleic acid	18:1c9	75.5	0	0.059 ± 0.019
Palmitoleic acid	16:1c	189.6	0.0013 ± 0.0003	0.013± 0.005
PUFA				
Linoleic acid	18:2n-6	139	0.0041 ± 0.0003	0.0235 ± 0.0091
TOTAL Fatty Acids			0.005367 ± 0.000007	0.101 ± 0.013

* Data obtained from [33,34].

The fatty acid profile also differs between NB and WSB. WSB showed approximately 5% SFA, 23% PUFA and a 72% MUFA (expressed in relation to total fatty acids); while samples NB showed 76% PUFA, 24% MUFA and no presence of SFA (expressed in relation to total fatty acids). The change

in the fatty acid profile of WSB is very important, mainly due to the presence of saturated fatty acids, principally palmitic acid. Numerous studies have shown that the consumption of saturated fat increases blood cholesterol levels, especially the LDL fraction [35–37]. Furthermore, the presence of WS produces, in tissue surface, an increase in total fatty acid content from 0.54% to 10.1%, so the chicken breast can no longer be considered a totally lean meat.

Permittivity was measured with different sensors. A small sensor with two needles with blunt-ended was used at radiofrequency and a coaxial probe was used in the microwave range in order to characterize the dielectric properties of the different parts of breast: muscle and adipose tissues, in NB and WSB, and WS in WSB.

One of the main problems to fit the full spectrum of radiofrequency and microwave is the appearance of three dispersions on a very large frequency range with sigmoidal shape. The Debye model is a physic model that explains the electric behaviors in this specific frequency range. This model uses different parameters to define each relaxation [38], but it is tedious to handle and difficult to fit with statistic tools. Others authors have used different math models trying to approach the Debye model [39,40]. A powerful sigmoidal model used in biological systems is the Gompertz model [41].

The dispersions shown in the dielectric constant spectra, α, β and γ, are similar to the sigmoidal models aforementioned. Therefore, dielectric constant (ε') was modelled adjusting the experimental data using an own adaptation of the modified Gompertz model (Equation (5)) in order to obtain information on each of the dispersions:

$$l\varepsilon'(\omega) = l\varepsilon'_\infty + \sum_{n=1}^{3} \frac{\Delta l'\varepsilon_n}{1 + e^{((l\omega^2 - l\omega_\tau^2)*\alpha_n)}} \tag{5}$$

where $l\varepsilon'$ represents the decimal logarithm of the dielectric constant, $l\varepsilon'_\infty$ the logarithm of the dielectric constant at high frequencies, $l\omega$ represents the decimal logarithm of the angular velocity (obtained from the frequency), $\Delta l\varepsilon'_n$ ($\Delta l\varepsilon'_n = \log \varepsilon'_n - \log \varepsilon'_{n-1}$) the magnitude of the dispersion, $l\omega_t$ the logarithm of the angular velocity at relaxation time for each dispersion n, and α_n are the dispersion slopes.

Figure 5 shows an example of the different dispersions that have been modelled. In the figure the parameters of the Gompertz model are also represented.
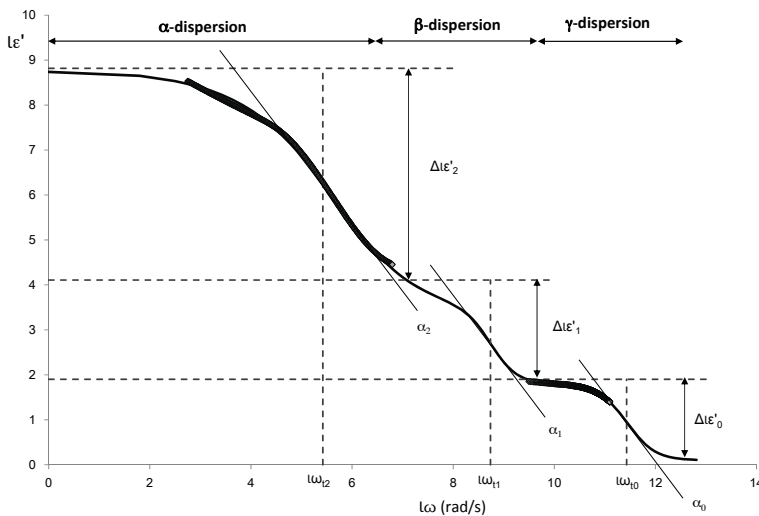
Figure 5. The adapted modified Gompertz model application, where the black line corresponds to the values of the mathematical model and the black diamonds are the measured data of normal breast.

From the Gompertz parameters, it is possible to determine the relaxation frequencies and dielectric constants of each relaxation (Equations (6)–(9)):

$$\varepsilon'_\alpha = 10^{(l\varepsilon'_\infty + \Delta l\varepsilon'_\gamma + \Delta l\varepsilon'_\beta + \frac{\Delta l\varepsilon'_\alpha}{2})} \tag{6}$$

$$\varepsilon'_\beta = 10^{(l\varepsilon'_\infty + \Delta l\varepsilon'_\gamma + \frac{\Delta l\varepsilon'_\beta}{2})} \tag{7}$$

$$\varepsilon'_\gamma = 10^{(l\varepsilon'_\infty + \frac{\Delta l\varepsilon'_\gamma}{2})} \tag{8}$$

$$f_i = 10^{\frac{l\omega_{\tau i}}{2\cdot\pi}} \tag{9}$$

Being i for Equation (9) each dispersion (α, β and γ). Figure 5 shows an example of radiofrequency and microwave spectra for each tissue. From the Gompertz adjustment and using Equations (6)–(9), it was possible to obtain the relaxation dielectric constants and the relaxation frequencies for each dispersion (Tables 4 and 5).

Table 4. Relaxation dielectric constant at each dispersion (α, β and γ) of the different tissues of normal breast (NB) and white striping breast (WSB). Data obtained by the two needles with blunt-ended sensor.

Sample	ε'								
	α (10^6)			β (10^2)			γ		
Muscle tissue WSB	22	$\pm$	6 [a]	25	$\pm$	3 [a]	17.3	$\pm$	0.2 [a]
White stripe WSB	13	$\pm$	6 [b]	11	$\pm$	5 [b]	15.5	$\pm$	0.9 [b]
Adipose tissue WSB, NB	6	$\pm$	2 [c]	4	$\pm$	1 [c]	15.0	$\pm$	0.9 [b]
Muscle tissue NB	18	$\pm$	4 [a]	21	$\pm$	5 [a]	17.15	$\pm$	0.14 [a]

Different letters ([a–c]) indicate significant difference between values in each column with $p < 0.05$.

Table 5. Relaxation frequency at each dispersion (α, β and γ) of the different tissues of normal breast (NB) and white striping breast (WSB), obtained by the two needles with blunt-ended sensor.

Sample	f								
	α (kHz)			β (MHz)			γ (GHz)		
Muscle tissue WSB	6	$\pm$	1 [b]	8	$\pm$	4 [b]	45	$\pm$	3 [a]
White stripe WSB	7	$\pm$	4 [b]	10	$\pm$	1 [b]	38	$\pm$	1 [b]
Adipose tissue WSB, NB	6	$\pm$	4 [b]	110	$\pm$	7 [a]	34	$\pm$	2 [c]
Muscle tissue NB	18	$\pm$	4 [a]	13	$\pm$	5 [b]	43	$\pm$	1 [a]

Different letters ([a–c]) indicate significant difference between values in each column with $p < 0.05$.

As Figure 6 shows, three relaxations per tissue can be appreciated. In α-dispersion, the relaxation dielectric constant of muscles tissues (WSB and NB) shows a significant higher value ($p < 0.05$) than the fatty tissues (WS and adipose). The main mobile charges in muscle tissue are Ca^{2+}, K^+, Na^+ and Mg^{2+}, which have different functions, for example as the second messenger of ATP signaling; these electrolytes are solved in liquid phase, and their mobility is high. However, in fatty tissues, the electrolytes suffer the high attraction of surface tension of fatty globules, thus electrolytes maintain their orientation ability although this ability is reduced. Moreover, the white stripe sample shows a significantly higher value ($p < 0.05$) than the adipose tissue because white stripe has an accumulation of electrolytes in the interface between the partially breakdown muscle tissue and the fat globules deposited in this fissure (Table 4), where the MUFA predominance increases the adsorption of electrolytes. In order to understand the nature and interaction of the electrolytes responsible of α dispersion, the relaxation frequencies were analyzed (Table 5). The fatty tissues show the electrolytes in the same state (spin orientation); nevertheless, in case of muscle tissues, the muscle in WSB shows a similar state of electrolytes as fatty

tissues. This could be due to the effect of the partial breakdown of muscle in WSB, where the nature of the electrolytes associated with the interface between fat globules and muscle tissue is different from the electrolytes in the liquid phase. This is shown clearly in the very significant difference ($p < 0.001$) of the relaxation frequency (state of electrolytes) between the muscle tissue in NB (muscle fibers without fat interactions) and the muscle in WSB.
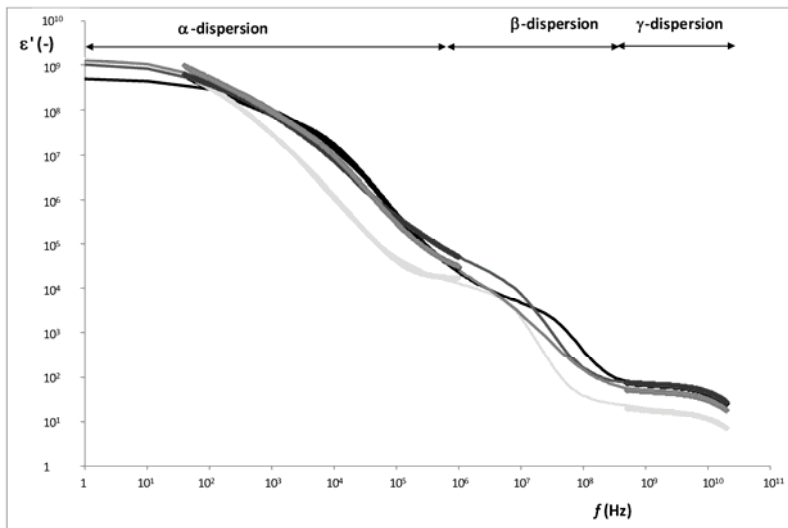
Figure 6. Dielectric constant spectra in radiofrequency and microwave ranges of the different tissues of white striping breast (WSB) and normal breast (NB). Where, the lines correspond to Gompertz model and the diamonds to the experimental data for each tissue.

With regard to β-relaxation, the dielectric constant of muscles tissues (WSB and NB) shows a significant higher value ($p < 0.05$) than the fatty tissues (WS and adipose), being also significant ($p < 0.05$) the difference between WS and adipose tissue.

The β-relaxation phenomena are explained by the orientation of fixed charges of the dielectric media. In muscle tissue, the structural proteins have active sites, fixed charges with orientation capacity; moreover, the main of these charges with orientation capacity are the charges involved in the actin-myosin complex [42]. This phenomenon is produced in the low MHz range as Table 5 shows, where muscles tissues have 8 ± 4 and 13 ± 5 MHz, for WSB and NB, respectively, without significant differences. Nevertheless, in the high MHz range another phenomenon occurs which is associated to the surface tension, this phenomenon is called the Maxwell-Wagner effect [15,43], and produces less absorption of energy than the effect of the active sites of proteins. In case of adipose tissue, as the relaxation frequency shows (110 ± 7 MHz), the high surface tension of the fat globules is the main contributor to β-relaxation, being the relaxation dielectric constant the lowest. In the case of WS, the breakdown of the muscle tissue and the formation of the white stripes by fat globule deposition produces both effects: the breakdown of the structural proteins of muscle tissue produces a partial degradation of myosin-actin complex and new active sites with orientation capacity, and the effect of the orientation of surface charges of fat globules. Comparing the dielectric constant of muscle tissue and adipose tissue (Table 4), it is possible to observe that the orientation capacity of active protein sites produces more absorption of energy than the capacity of orientation of surface charges of the fat globules. Taking into account that the WS is composed by broken muscle tissue and fat globules, then the β-relaxation of WS should be affected by the stronger effect, which is the orientation

of active sites of proteins. This theory is strengthened by observing the β-relaxation frequency in range of the effect of protein charges (10 ± 1 MHz) and an intermediate dielectric constant value between muscles tissues and adipose tissue. The partial denaturalization of myosin-actin complex, the protein breakdown and the reduction of the electric field space occupied by proteins (Figure 3) produces a reduction of the absorption energy capacity in the WS with regard to the muscles tissues, as Table 4 shows.

Finally, regarding the γ-dispersion, the dielectric constant of the muscle tissues is higher than that of the adipose tissues (Table 4). Orientation of the dipolar molecules occurs at these frequencies, the main dipolar molecule in animal tissue is water; the adipose tissue shows less interaction with water molecules than muscle tissue due to its hydrophobic character. Water molecules in adipose tissue interact with adsorbed electrolytes of the fat, while in muscle tissue the water can be in liquid phase and also adsorbed to the tissue. Therefore the relaxation frequency is larger in muscle tissue with regard to the fat tissue since water has a higher mobility in liquid phase than adsorbed in the solid matrix (Table 5).

The permittivity of NB and WSB samples were also measured using a sensor with two flat circular plates. This kind of sensor with a raised surface was used in order to increase the penetration and to measure the permittivity at radiofrequency through the chicken skin.

Table 6 shows the average values of the dielectric constant of α and β dispersions after adjusting the data with the adapted modified Gompertz model (Equation (5)). The ANOVA performed on the WS samples compared to the normal samples reveals that the differences in the dielectric constant are significant ($p < 0.05$). As was explained in the analysis of the different tissues, in α and β dispersion, the muscle tissue produces the major effect in the dielectric constant. The measurements made in the whole breast (including the skin) follows the same behaviors, where the NB (Table 6) behaves as the muscle of the NB (Table 4) and the WSB (Table 6) behaves as a mixture between muscle of the WSB and WS (Table 4). Therefore, using sensors in the radio frequency range, it is possible to segregate chickens with white striping physiopathy before dismembering, carrying out the measurements on whole chickens.

Table 6. Values of dielectric constant in α and β dispersions of white striping breast (WSB) and normal breast (NB), obtained by the sensor with two flat plates with circular surfaces.

Sample	ε'					
	$\alpha\ (10^6)$			$\beta\ (10^2)$		
WSB	3.1	±	0.8 [b]	21	±	5 [b]
NB	4.2	±	0.3 [a]	31	±	3 [a]

Different letters ([a]–[b]) indicate significant difference between values in each line with $p < 0.05$.

4. Conclusions

It was possible to conclude that white striping physiopathy consists of the partial rupture of the pectoral muscle, possibly associated with the overgrowth process of hypertrophic breeds; the organism solves this rupture of the tissue by depositing fat globules in the break area in order to maintain the functionality of contraction-relaxation of the muscular system. The fat content and the relationship between types of fatty acids vary when white striping occurs. The change in fat content is remarkable, rising from 0.5% to 10.1%, exhibiting a higher content of MUFA and SFA fatty acids; thus, WSB cannot be considered to be a healthy food with low saturated fats. Finally, it was possible to detect white striping physiopathy in chicken carcasses with skin using sensors with flat circular surfaces in radiofrequency spectra.

References

1. Best, P. *Worldwide Poultry Meat Production, Consumption Forecasts*; Watt: Rockford, IL, USA, 2011.
2. Havenstein, G.B.; Ferket, P.R.; Qureshi, M.A. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult. Sci.* **2003**, *82*, 1500–1508. [CrossRef] [PubMed]
3. Soglia, F.; Laghi, L.; Canonico, L.; Cavani, C.; Petracci, M. Functional property issues in broiler breast meat related to emerging muscle abnormalities. *Food Res. Int.* **2016**, *89*, 1071–1076. [CrossRef]
4. Petracci, M.; Mudalal, S.; Soglia, F.; Cavani, C. Meat quality in fast-growing broiler chickens. *Worlds Poult. Sci. J.* **2015**, *71*, 363–374. [CrossRef]
5. Dransfield, E.; Sosnicki, A.A. Relationship between muscle growth and poultry meat quality. *Poult. Sci.* **1999**, *78*, 743–746. [CrossRef] [PubMed]
6. Kijowski, J.; Konstańczak, M. Deep pectoral myopathy in broiler chickens. *Bull. Vet. Inst. Pulawy* **2009**, *53*, 487–491. [CrossRef]
7. Lien, R.J.; Bilgili, S.F.; Hess, J.B.; Joiner, K.S. Induction of deep pectoral myopathy in broiler chickens via encouraged wing flapping. *J. Appl. Poult. Res.* **2012**, *21*, 556–562. [CrossRef]
8. Bailey, R.A.; Watson, K.A.; Bilgili, S.F.; Avendano, S. The genetic basis of pectoralis major myopathies in modern broiler chicken lines. *Poult. Sci.* **2015**, *94*, 2870–2879. [CrossRef] [PubMed]
9. Russo, E.; Drigo, M.; Longoni, C.; Pezzotti, R.; Fasoli, P.; Recordati, C. Evaluation of white striping prevalence and predisposing factors in broilers at slaughter. *Poult. Sci.* **2015**, *94*, 1843–1848. [CrossRef] [PubMed]
10. Petracci, M.; Mudalal, S.; Bonfiglio, A.; Cavani, C. Occurrence of white striping under commercial conditions and its impact on breast meat quality in broiler chickens. *Poult. Sci.* **2013**, *92*, 1670–1675. [CrossRef] [PubMed]
11. Kuttappan, V.A.; Lee, Y.S.; Erf, G.F.; Meullenet, J.F.; McKee, S.R.; Owens, C.M. Consumer acceptance of visual appearance of broiler breast meat with varying degrees of white striping. *Poult. Sci.* **2012**, *91*, 1240–1247. [CrossRef] [PubMed]
12. Lorenzi, M.; Mudalal, S.; Cavani, C.; Petracci, M. Incidence of white striping under commercial conditions in medium and heavy broiler chickens in Italy. *J. Appl. Poult. Res.* **2014**, *23*, 754–758. [CrossRef]
13. Petracci, M.; Mudalal, S.; Babini, E.; Cavani, C. Effect of white striping on chemical composition and nutritional value of chicken breast meat. *Ital. J. Anim. Sci.* **2014**, *13*, 179–183. [CrossRef]
14. Schwan, H.P. Electrical properties of tissues and cell suspensions. *Adv. Biol. Med. Phys.* **1957**, *5*, 147–209. [PubMed]
15. Kuang, W.; Nelson, S.O. Low-Frequency dielectric properties of biological tissues: A review with some new insights. *Trans. ASAE-Am. Soc. Agric. Eng.* **1998**, *41*, 173–184. [CrossRef]
16. Wolf, M.; Gulich, R.; Lunkenheimer, P.; Loidl, A. Relaxation dynamics of a protein solution investigated by dielectric spectroscopy. *Biochim. Biophys. Acta Proteins Proteom.* **2012**, *1824*, 723–730. [CrossRef] [PubMed]
17. Metaxas, A.C.; Meredith, R.J. *Industrial Microwave Heating*; IEE Power Engineering Series 4; Institution of Electrical Engineers. Peter Peregrinus Ltd.: London, UK, 1983.
18. Castro-Giráldez, M.; Fito, P.J.; Dalla Rosa, M.; Fito, P. Application of microwaves dielectric spectroscopy for controlling osmotic dehydration of kiwifruit (Actinidia deliciosa cv Hayward). *Innov. Food Sci. Emerg. Technol.* **2011**, *12*, 623–627. [CrossRef]
19. Traffano-Schiffo, M.V.; Castro-Giraldez, M.; Colom, R.J.; Fito, P.J. Study of the application of dielectric spectroscopy to predict the water activity of meat during drying process. *J. Food Eng.* **2015**, *166*, 285–290. [CrossRef]
20. Talens, C.; Castro-Giraldez, M.; Fito, P.J. Study of the effect of microwave power coupled with hot air drying on orange peel by dielectric spectroscopy. *LWT-Food Sci. Technol.* **2016**, *66*, 622–628. [CrossRef]
21. Shang, L.; Guo, W.; Nelson, S.O. Apple variety identification based on dielectric spectra and chemometric methods. *Food Anal. Methods* **2015**, *8*, 1042–1052. [CrossRef]

22. Castro-Giráldez, M.; Botella, P.; Toldrá, F.; Fito, P. Low-frequency dielectric spectrum to determine pork meat quality. *Innov. Food Sci. Emerg. Technol.* **2010**, *11*, 376–386. [CrossRef]

23. Trabelsi, S.; Roelvink, J. Investigating the influence of aging on radiofrequency dielectric properties of chicken meat. *J. Microw. Power Electromagn. Energy* **2014**, *48*, 215–220. [CrossRef]

24. Damez, J.L.; Clerjon, S.; Abouelkaram, S.; Lepetit, J. Beef meat electrical impedance spectroscopy and anisotropy sensing for non-invasive early assessment of meat ageing. *J. Food Eng.* **2008**, *85*, 116–122. [CrossRef]

25. Cuibus, L.; Castro-Giráldez, M.; Fito, P.J.; Fabbri, A. Application of infrared thermography and dielectric spectroscopy for controlling freezing process of raw potato. *Innov. Food Sci. Emerg. Technol.* **2014**, *24*, 80–87. [CrossRef]

26. Traffano-Schiffo, M.V.; Castro-Giráldez, M.; Fito, P.J.; Balaguer, N. Thermodynamic model of meat drying by infrarred thermography. *J. Food Eng.* **2014**, *128*, 103–110. [CrossRef]

27. Benedito, J.; Carcel, J.A.; Rossello, C.; Mulet, A. Composition assessment of raw meat mixtures using ultrasonics. *Meat Sci.* **2001**, *57*, 365–370. [CrossRef]

28. Zhang, L.; Barbut, S. Rheological characteristics of fresh and frozen PSE, normal and DFD chicken breast meat. *Br. Poult. Sci.* **2005**, *46*, 687–693. [CrossRef] [PubMed]

29. Kuttappan, V.A.; Shivaprasad, H.L.; Shaw, D.P.; Valentine, B.A.; Hargis, B.M.; Clark, F.D.; Owens, C.M. Pathological changes associated with white striping in broiler breast muscles. *Poult. Sci.* **2013**, *92*, 331–338. [CrossRef] [PubMed]

30. Mudalal, S.; Babini, E.; Cavani, C.; Petracci, M. Quantity and functionality of protein fractions in chicken breast fillets affected by white striping. *Poult. Sci.* **2014**, *93*, 2108–2116. [CrossRef] [PubMed]

31. Knothe, G.; Dunn, R.O. A comprehensive evaluation of the melting points of fatty acids and esters determined by differential scanning calorimetry. *J. Am. Oil Chem. Soc.* **2009**, *86*, 843–856. [CrossRef]

32. Kuttappan, V.A.; Brewer, V.B.; Apple, J.K.; Waldroup, P.W.; Owens, C.M. Influence of growth rate on the occurrence of white striping in broiler breast fillets. *Poult. Sci.* **2012**, *91*, 2677–2685. [CrossRef] [PubMed]

33. Uehara, H.; Suganuma, T.; Negishi, S.; Uda, Y.; Furukawa, Y.; Ueno, S.; Sato, K. Physical properties of two isomers of conjugated linoleic acid. *J. Am. Oil Chem. Soc.* **2008**, *85*, 29–36. [CrossRef]

34. Cedeño, F.O.; Prieto, M.M.; Espina, A.; García, J.R. Measurements of temperature and melting heat of some pure fatty acids and their binary and ternary mixtures by differential scanning calorimetry. *Thermochim. Acta* **2001**, *369*, 39–50. [CrossRef]

35. Lichtenstein, A.H.; Ausman, L.M.; Jalbert, S.M.; Schaefer, E.J. Effects of different forms of dietary hydrogenated fats on serum lipoprotein cholesterol levels. *New Engl. J. Med.* **1999**, *340*, 1933–1940. [CrossRef] [PubMed]

36. Mauger, J.F.; Lichtenstein, A.H.; Ausman, L.M.; Jalbert, S.M.; Jauhiainen, M.; Ehnholm, C.; Lamarche, B. Effect of different forms of dietary hydrogenated fats on LDL particle size. *Am. J. Clin. Nutr.* **2003**, *78*, 370–375. [PubMed]

37. Socarrás Suárez, M.M.; Bolet Astoviza, M. Alimentación saludable y nutrición en las enfermedades cardiovasculares. *Rev. Cuba Investig. Bioméd.* **2010**, *29*, 353–363.

38. Kuang, W.; Nelson, S.O. Low-Frequency dielectric dispersion from ion permeability of membranes. *J. Colloid Interface Sci.* **1997**, *193*, 242–249. [CrossRef] [PubMed]

39. Hurt, W.D. Multiterm Debye dispersion relations for permittivity of muscle. *IEEE Trans. Biomed. Eng.* **1985**, *32*, 60–64. [CrossRef] [PubMed]

40. Gabriel, S.; Lau, R.W.; Gabriel, C. The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues. *Phys. Med. Biol.* **1996**, *41*, 2271. [CrossRef] [PubMed]

41. El-Gohary, A.; Alshamrani, A.; Al-Otaibi, A.N. The generalized Gompertz distribution. *Appl. Math. Model.* **2013**, *37*, 13–24. [CrossRef]

42. Rayment, I.; Holden, H.M.; Whittaker, M.; Yohn, C.B.; Lorenz, M.; Holmes, K.C.; Milligan, R.A. Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **1993**, *261*, 58–65. [CrossRef] [PubMed]

43. Traffano-Schiffo, M.V.; Balaguer, N.; Castro-Giráldez, M.; Fito, P.J. Emerging technologies in juice processing. In *Juice Processing: Quality, Safety and Value-Added Opportunities*; Ibartz, A., Falguera, V., Eds.; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2014; pp. 197–212.

Article

Radon Mitigation Approach in a Laboratory Measurement Room

Patricia Blanco-Rodríguez [1,‡], Luis Alfonso Fernández-Serantes [2,‡], Alberto Otero-Pazos [2,‡], José Luis Calvo-Rolle [2,*,†,‡] and Francisco Javier de Cos Juez [3,‡]

[1]   Department of Energy, University of Oviedo, Oviedo 33004, Spain; blancopatricia@uniovi.es
[2]   Department of Industrial Engineering, Universidade da Coruña, Ferrol 15405, Spain; luis.alfonso.fernandez.serantes@udc.es (L.A.F.-S.); alberto.otero.pazos@udc.es (A.O.-P.)
[3]   Department of Mines Exploitation and Prospecting, University of Oviedo, Oviedo 33004, Spain; fjcos@uniovi.es
*   Correspondence: jlcalvo@udc.es; Tel.: +34-88-101-3117
†   Current address: Department of Industrial Engineering, Universidade da Coruña, 15405 Coruña, Spain.
‡   These authors contributed equally to this work.

Abstract: Radon gas is the second leading cause of lung cancer, causing thousands of deaths annually. It can be a problem for people or animals in houses, workplaces, schools or any building. Therefore, its mitigation has become essential to avoid health problems and to prevent radon from interfering in radioactive measurements. This study describes the implementation of radon mitigation systems at a radioactivity laboratory in order to reduce interferences in the different works carried out. A large set of radon concentration samples is obtained from measurements at the laboratory. While several mitigation methods were taken into account, the final applied solution is explained in detail, obtaining thus very good results by reducing the radon concentration by 76%.

Keywords: radon; mitigation; positive ventilation; sump system; sealing; radon reduction

---

1. Introduction

Radon is a gas produced by the natural radioactive decay of U^{238}, a very active chemical element capable of associating with many others and moving with them [1]. Uranium is present in small proportions (ppm) in almost all rocks and soils of the planet [2]. Galicia Autonomy (Spain) is an area prone to radon gas due to the common presence of granites, a type of rock and soil that has great uranium concentration [3,4].

Radon is a naturally produced inert radioactive gas. It is odorless, colorless and tasteless [5]. It is easily released from the soil into the air, where it disintegrates forming various short-lived products known as radon progeny [6]. In the decay mode of radon gas, alpha particles are emitted, which adhere to aerosols, dust and other airborne particles [7]. As a consequence, when people breathe, radon progeny is deposited on epithelial cells lining the airways where alpha particles can damage DNA and, therefore, cause lung cancer [8] and thousands of deaths all over the world each year [9]. According to the United States Environmental Protection Agency (EPA), radon gas is the second leading cause of lung cancer, after smoking, causing 21 thousand deaths per year in the United States [10].

Due to its accumulation, radon can be a severe issue in houses, workplaces, schools or any building where people or animals are located. It comes from the soil basement where buildings are located and it enters them through cracks in the floors, construction joints, gaps in suspended floors, cavities inside walls or water supply.

Normally, radon gas is not produced by common building materials [9] and, as a general rule, outdoor radon concentration is very low. On average, it varies between 5 and 15 Bq/m^3 [7]. Conversely, indoor concentrations are higher and reach maximums at places such as mines, caves and treatment water plants, among others [7].

Like in other fields [11,12], Spanish legislation establishes that the average concentration of radon at workplaces should be less than 600 Bq/m^3 (300 Bq/m^3 for workplaces with high permanence of workers). Over this value (>600 Bq/m^3), a mitigation control system with periodical tests should be installed to keep radon levels low [13].

The equipment used for environmental radioactivity measurements is very sensitive and needs specific conditions, especially regarding temperature, pressure and relative humidity. Consequently, it is necessary to achieve optimal operating conditions and, therefore, there should be no presence of perturbation or any radioactive contamination such as radon gas concentration in the place where the equipment is located. Significant radon gas levels will involve an undesirable increase in the variability of measurements. Of course, this should not occur if very precise measures are required.

There are some studies about radon mitigation in the literature, mainly in households; however, there are no cases of laboratory measurement rooms. This research shows a specific study of this last case, where an initial analysis and monitoring of radon gas concentration was made with common equipment. Then, consequent mitigation solutions were checked and implanted for the required reduction, providing specific data during operation for the adopted solutions. It should also be taken into consideration that the laboratory location is in a geographical area with high radon gas concentrations due to the soil nature, which significantly complicates the main goal.

The use of a high precision sensor under a strictly controlled scenario enables the validation of a certain set of assumptions neither verified nor published so far.

Different factors play an important role in decreasing radon gas concentration when the number of air changes per hour in a room is above four. This is of particular importance in laboratories equipped with environmental radioactivity measurement devices, each one of them having its own sensor. These specific measurement rooms are key to controlling environmental radioactivity, a vital issue both at national and European level.

The EU Member States have a collective responsibility to ensure continuous monitoring of radioactivity in the air, water and soil, in accordance with Articles 35 and 36 of the Euratom Treaty concerning monitoring of radioactivity in the environment. Additional EU legislation sets out detailed criteria for the quality of water intended for human consumption, and the maximum permitted levels of radioactive contamination of foodstuffs (Directive 98/83/EC and Regulation 733/2008/EC respectively).

Reliability of the environmental radioactivity equipment measurements largely depends on the mitigation methods adopted to reduce radon gas concentration. As a consequence, in a critical area such as environmental radioactivity monitoring, it is essential not only to include a contrast methodology but also to establish a clear configuration of the measurement room and its equipment.

In order to reduce radon gas concentration at workplaces, several mitigation solutions can be implemented, such as:

- Sealing: trying to seal all obvious cracks, gaps and holes in floors or walls to prevent radon from entering the building. However, in most cases, reduction is not effective [14,15].
- Positive ventilation: installing small fans to blow filtered fresh air into the building. The fundamental principle is to increase the pressure inside the building in relation with outside pressure, so radon gas cannot flow inside [4,14,15].
- Sump systems: the purpose is to discharge the air from inside the building when the concentration reaches a set point value. Usually, the extraction is forced to improve the performance and to reduce the response time [15–17].
- Underfloor ventilation: installing vents at the foundations to provide natural or force ventilation underfloor to suck or blow radon [4,15].

The proposed solutions in the present work are based on positive ventilation and their main aim is to avoid interferences with the measurement equipment at the radioactivity laboratory. This study is focused on the analysis of different steps that have been carried out in the mentioned case. For each step, a change has been introduced and several measurements have been done. Finally, a comparison was made, taking into account all recorded cases.

This paper begins with a description of the case study, where both the installation and all the different components used to decrease radon concentration are detailed. It follows with the carried out process that analyses the approach, step by step. Then, the results are shown and finally the conclusions are drawn.

2. Materials and Methodology

In this section, the specific radiation monitoring sensor as well as the followed methodology are described.

2.1. Materials

The sensor employed to measure radon gas activity is the AlphaGUARD PQ 2000PRO model from SAPHYMO [18].

AlphaGUARD PQ 2000PRO is a portable device used for instantaneous or continuous radon concentration measurements, both for short- or long-term tests in buildings and outdoors.

This sensor, which is insensitive to vibrations as well as humidity, has a pulse-counting ionization chamber (alpha spectroscopy) and allows radon gas levels to be measured in a wide range, between 2 and $2,000,000$ Bq/m^3. Operating in diffusion mode, the user can select measuring cycles from 10 min to 60 min, while operating in flow-through mode the cycle time varies between 1 and 10 min.

Among its many advantages, AlphaGUARD PQ 2000PRO performs fast response with high detection efficiency (1 cpm at 20 Bq/m^3), and enables maintenance-free operation with a stable calibration. The instrument calibration error amounts to 3%, plus uncertainty of the primary standard.

In addition to the radon gas concentration, AlphaGUARD PQ 2000PRO is equipped with different integrated sensors that enable other essential parameters in radon mitigation to be measured and stored, such as ambient temperature, atmospheric pressure and relative humidity. The internal sensor included to measure ambient temperature, from -10 °C to 60 °C, is a precision monolithic integrated circuit. The device also has a laser-trimmed silicon bridge transducer in order to determine atmospheric air pressure, from 800 mbar to 1050 mbar. Finally, the integrated sensor that allows to measure relative humidity, from 0% rH to 99% rH, is an hydrophilic polymer film on hybrid.

A specific software called DataEXPERT, included together with the AlphaGUARD Radon Monitor in the basic configuration sold by company Saphymo GmbH, was used for storing and analyzing these data.

The main technical characteristics of the AlphaGUARD PQ 2000PRO sensor are provided below (Table 1).

Table 1. AlphaGUARD PQ 200PRO characteristics.

Physical Characteristics	
Measurement range (Rn-222) -Lower limit -Upper limit	 2 (Bq/m^3) $2,000,000$ (Bq/m^3)
Foldback protection	>10,000,000 (Bq/m^3) verified
Type of radon detector	Ionization chamber ($HV = 750VDC$)
Operation mode	3D alpha spectroscopy and current mode
Detector volume -Total -Active	 0.62 liter 0.56 liter
Type of Rn-FP filter (detector entry window)	Glass fiber filter (retension coefficient > 99,9 %)
Detector filling mechanisms	Design optimized for fast passive diffusion (10/60 min cycle) Flow mode (1/10 min cycle) Rn/Tn mode (10 min cycle)
Detector signal acquisition	Fast digital signal sampling network, using three separate ADC-channels
Calibration error (Rn-222)	±3%
Detector efficiency	1 cpm at 20 (Bq/m^3)
Background signal due to detector contamination	<1 (Bq/m^3)
Measurement cycle time -Diffusion mode -Flow-through mode	 10 min or 60 min (user selectable) 1 min or 10 min (user selectable)
Storage capacity	3 days at 1 min measuring cycles 1 month at 10 min measuring cycles 6 months at 60 min measuring cycles
Internal sensors -Ambient temperature -Atmospheric air pressure -Relative air humidity	 Precision monolithic integrated circuit (−15 ºC–60 ºC) Laser-trimmed silicon bridge transducer (800 mbar–1050 mbar) Hydrophilic polymer film on hybrid (0% ∼ 99%rH)
Environmental Characteristics	
System operating range -Operating temperature -Air pressure -Relative humidity	 −10 ºC–50 ºC 700 mbar–1150 mbar 0% ∼ 95%
Mechanical Characteristics and Display	
Dimensions (excluding handle)	120 mm × 315 mm × 175 mm (H × W × D)
Weight	4.5 kg
Electrical Characteristics	
Power supply	Battery or external mains (120 V–230 V)
Battery	Rechargable, allows > 10 days autonomous operation, >30 days using external booster
Interfaces	
DataEXPERT software for data collection, management and professional analysis	

2.2. Methodology

The AlphaGUARD PQ 2000PRO detector is based on a design-optimized pulse ionization chamber. In regular operation, the measuring gas enters diffusion mode through a large-surface glass fiber filter into the ionization chamber. In addition, the glass fiber filter protects the interior of the chamber and thus prevents contamination of dusty particles. While the radon progeny products do not enter the ionization chamber, only the gaseous Radon-222 reaches the chamber passing through the glass fiber filter.

For our purpose, the AlphaGUARD sensor operates in diffusion mode measuring radon gas concentration every 10 min. With the obtained values, the equipment calculates the average as in Equation (1).

$$Average = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad (1)$$

It also calculates the uncertainty as the standard deviation using the Equation (2).

$$Uncertainty = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}} \qquad (2)$$

3. Case Study

The Environmental Radioactive Laboratory of the University of A Coruña (Spain) carries out different works related to radioactive measurements and the analysis of the obtained data. These tasks are carried out in a specific room (measurement room), where all equipment used in this study is located. The case study is described in detail in the following subsections.

3.1. Environmental Radioactive Laboratory

The proposal has been implemented at the *"Laboratorio de Radiactividad Ambiental"* (LRA) of the University of A Coruña, located at the Faculty of Engineering in Ferrol (Galicia Autonomy). The laboratory carries out different studies on the radiological characterization of environmental samples on the north coast of Galicia or with a specific agreement with the Nuclear Safety Council, on a program of Environmental Radiation Monitoring, for example.

The characterization measurements of samples are done in a small room, where a concentration of radon gas has been detected. This laboratory measurement room was chosen because it is a small room, closed and with little ventilation, an appropriate place where radon gas concentrations can be high.

The radon level was first measured in the laboratory measurement room. Because radon gas is present in detectable concentrations in this specific laboratory, it was chosen to be the site to study different mitigation methods. Some of these methods were implemented with the consequent reduction of radon level.

3.2. Laboratory Measurement Room

The LRA measurement room is a small room of 20×60 m^2 where all radioactive measurements are done and where the used equipment is placed. The room is located inside the LRA, on the second floor of the building. It is west-oriented and its dimensions are, approximately, 3.20 m high, 5.80 m long and 3.55 m wide. Figure 1 shows the overall appearance of the room.

The building structure and the floor are made of reinforced concrete with the walls made of coated and painted bricks. Due to the heavy weight of the equipment, it was necessary to install steel sheets on the floor in order to distribute the weight uniformly. To get inside this room, there is a door that connects with the chemistry laboratory and the office. Moreover, there is a window situated on the west wall where the different mitigation systems are installed.
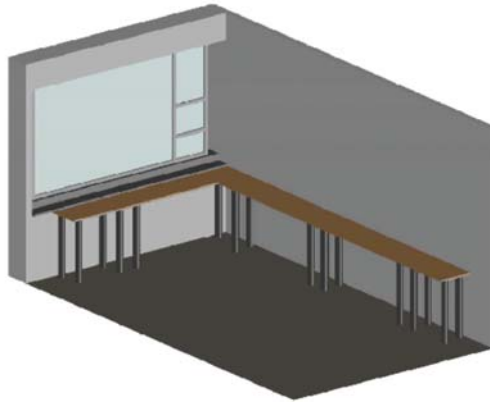
Figure 1. Measurement room.

4. Radon Mitigation Solution Approach

With the aim of reducing radon gas concentration and avoiding perturbations in the equipment, several mitigation methods have been taken into account. The implemented solution had to consider building limitations and restrictions. On the other hand, due to the fact that the measurement room is independent of the office, the applied solution does not disturb the work environment.

The measurement room is located on the second floor and, as a result, some mitigation methods such as underfloor ventilation are impossible to carry out. In addition, sealing and sump systems cannot be implemented because of the building age and because there are different entrances to the room. In any case, radon gas would have entered the measurement room through the cracks.

Instead, positive ventilation could be a good solution because of its viability and low-cost implementation and, due to the reasons described above, this is the chosen solution to mitigate radon.

4.1. Positive Ventilation Solution

Positive ventilation consists of fans that blow fresh and filtered air into the measurement room, by pressurizing to prevent radon from entering and diluting it [14,15]. To implement the selected solution, two fans and an air conditioning system were installed. Different steps were followed to carried out this option.

The fans were installed in order to pressurize and dilute radon gas. Installation of two fans was found to be necessary to provide enough ventilation.

Besides, an air conditioning system was installed to reduce air humidity, keep the temperature around the value of 20 Celsius degrees and to weatherize the measurement room. Moreover, this system blows filtered air into the measurement room, contributing to radon mitigation and its stability throughout the year. Air conditioning was necessary to ensure constant environmental conditions.

These systems also benefit the installation because they reduce condensation.

They were installed at the unique window of the room, so the works carried out for their installation were minimal. Figure 2 shows the measurement room with the installed systems.

The components used to decrease radon concentration are the following:

- Fans: The implemented fans are the HV-230 model from S&P, whose characteristics are:

    - Rotating speed of 1250 revolutions per second.
    - Extraction caudal at high speed of 600 m^3/h and at low speed of 450 m^3/h.
    - Impulsion caudal of 330 m^3/h.

The power consumed by the fans amounts to 34 watts each, and the sourced current is 0.15 amperes.

- Air conditioning: The air conditioning system is the PCA-RP50xB7KA model from Mitsubishi Electric, whose characteristics are:

  – For cold, 5 kW and 4300 kCal/h.
  – For hot, 5.5 kW and 4730 kCal/h.
  – Sound level, for low speed, around 32/40 dB.
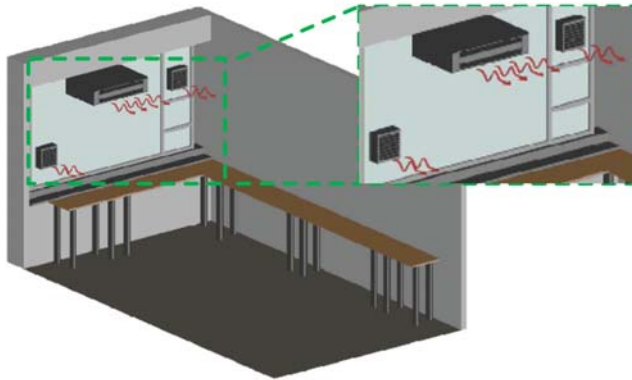
Figure 2. Measurement room with all systems.

4.2. Steps Followed until the Final Solution is Reached

In this subsection, the steps carried out to obtain the desired results are explained:

1. The starting point was the installation of one fan to pressurize the measurement room, see Figure 3. Radon concentration was measured some time later, once its level had achieved a steady state, but the level did not decrease enough with this action.
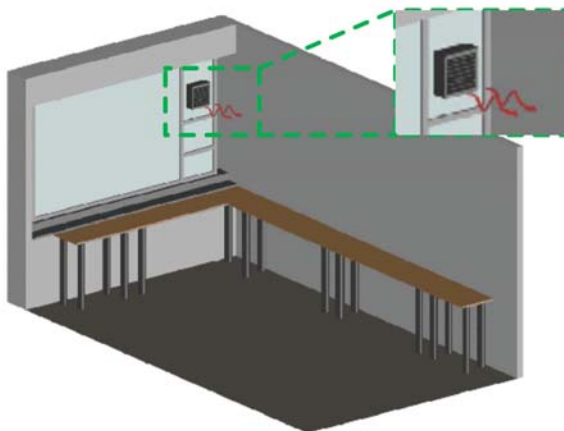
Figure 3. Measurement room with one fan.

2. As a result, a second fan was installed to increase the pressure inside the measurement room, see Figure 4. After several measurements, it was noticed that the introduced air was sufficient to reduce radon concentrations, thus avoiding interferences in the measurements.
3. Once the fan systems were implemented, air conditioning was installed, see Figure 2. Furthermore, it blows air inside the measurement room and weatherized it at a constant condition at all times. The air conditioning helps to have stable measurements and, therefore, minimize the uncertainty due to the temperature control of the room and the reduction in humidity.
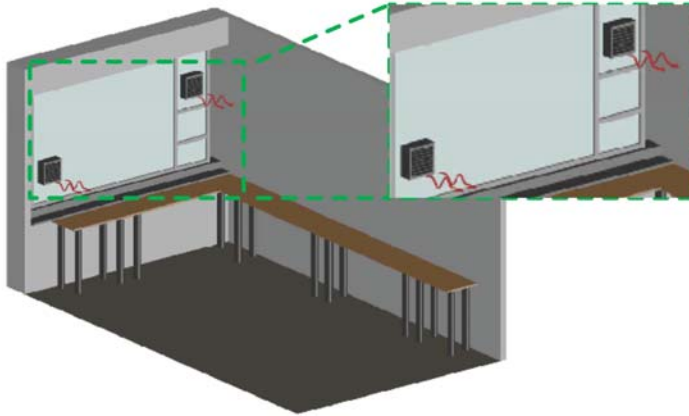
Figure 4. Measurement room with two fans.

4. In the end, all systems—fans and air conditioning—were working together with the aim to decrease radon concentration and to have repeatable conditions for the measurements. In addition, different analyses were done periodically to keep control of radon levels in order to know how they vary throughout the year.

In this particular case, fans work in impulsion mode and the air conditioning works keeping the temperature around the value of 20 °C and controlling air humidity.

5. Results

In this section, the radon concentration measurements for the different cases mentioned above are shown. At first, the radon concentration was measured on different days to quantify its level. Table 2 shows the average concentration for the period from April 25th to November 23th 2007, without any mitigation solution.

Table 2. Concentration before applying any solution.

Radon-222 Activity	Concentration (Bq/m^3)
Average	79
Uncertainty	37
Maximum concentration	158
Number of measurements	212

Then, from November 30, 2007, to January 25, 2008, one fan was installed and measurements were carried out. The next period of measuring was from January 25 to December 1, 2008. During this period, a second fan was installed and several measurements were taken with both fans operating.

Table 3 shows the concentrations during these periods. The analysis of the concentration while the mitigation solutions were implemented revealed how it decreased to the values shown in Table 3.

Table 3. Concentration after installation of fans.

Radon-222 Activity	One Fan (Bq/m^3)	Two Fans (Bq/m^3)	One/Two Fans (Bq/m^3)
Average	29	19	21
Uncertainty	18	10	13
Maximum concentration	64	53	64
Number of measurements	56	278	334

After the installation of the two fans, an air conditioning system was installed, achieving the results shown in Table 4. The period when this system (both fans and air conditioning) is on is from 1 December 2008 to nowadays. At this moment, the value of radon concentration is very low. Finally, Table 5 shows the percentage concentration reduction summary by using the different components of the mitigation solution. Remark, in general terms, it is necessary to apply a minimum overpressure of 5 mbar into the measurement room. Nowadays, both fans are always in operation to achieve this value.

Table 4. Concentration after air conditioning installation.

Radon-222 Activity	Concentration (Bq/m^3)
Average	19
Uncertainty	9
Maximum concentration	40
Number of measurements	1174

Table 5. Concentration reduction after mitigation systems installation.

Radon-222 Activity	Mitigation Solution	Compared with	Reduction %
Average	One fan	Any method	62.8
Uncertainty	One fan	Any method	50.3
Average	Two fans	One fan	35.6
Uncertainty	Two fans	One fan	43.7
Average	Two fans	Any method	76.0
Uncertainty	Two fans	Any method	72.0
Average	One or two fans	Any method	73.8
Uncertainty	One or two fans	Any method	65.8
Average	Two fans and air conditioning	Any method	76.0
Uncertainty	Two fans and air conditioning	Any method	75.7

While Figure 5 shows the measurement room with all the installed systems, two fans and the air conditioning system, Figure 6 shows a summary of the reduction of the radon gas with the different systems, showing how it is reduced when the methods are implemented.

Figure 5. Picture of the measurement room with all systems.

Figure 6. Evolution of radon gas measurement with the different systems.

6. Conclusions

With the radon mitigation solution approach explained in this paper, very good results have been obtained. As it is shown in the results section, after the installation of different removing systems, the
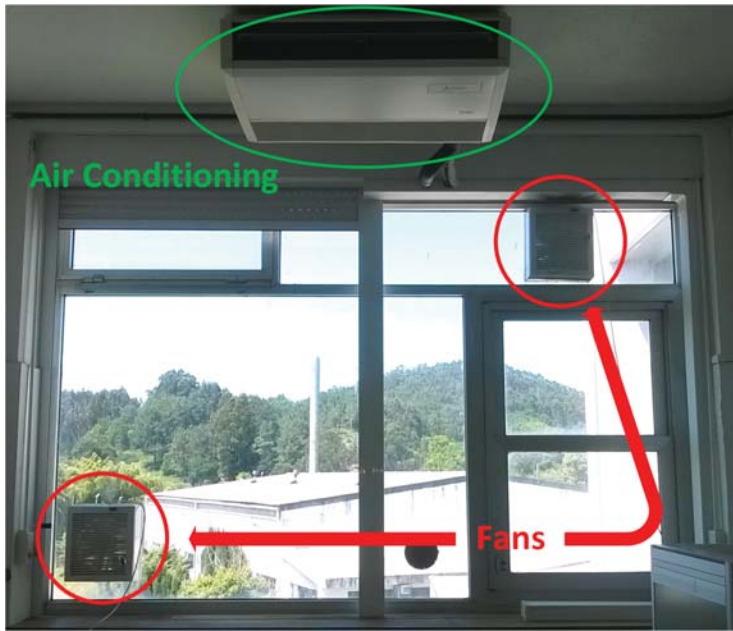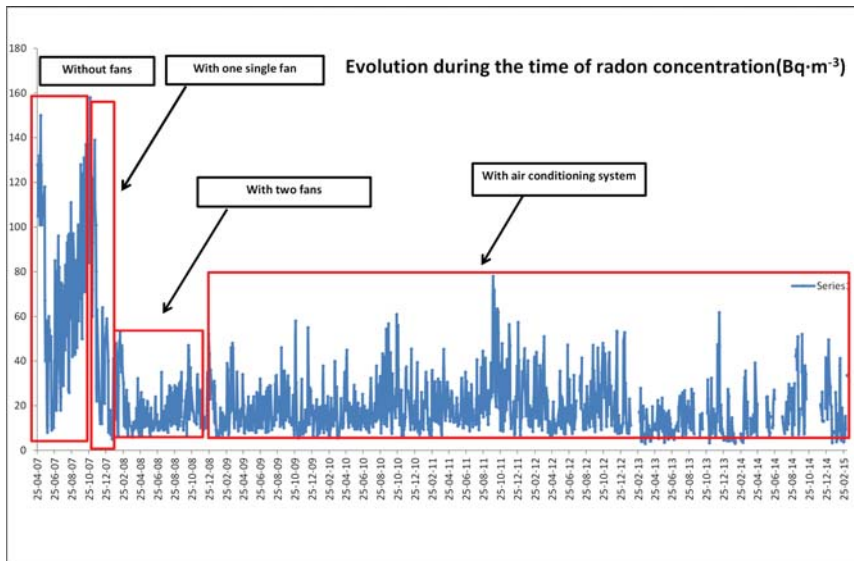
radon concentration decreased in a very significant way, up to the value of 76%. As a consequence, the implemented solution avoids interferences in the laboratory measurements.

Positive ventilation achieved good results by pressurizing the measurement room and avoiding radon entrance. For this purpose, two fans were installed with the aim of reducing the radon concentration to a low value. In addition, air conditioning contributed slightly to radon mitigation and allowed a very good repeatable condition performance for the measurements.

Author Contributions: Luis Alfonso Fernández-Serantes, Patricia Blanco-Rodríguez, Alberto Otero-Pazos, José Luis Calvo-Rolle and Francisco Javier de Cos Juez conceived the study. Luis Alfonso Fernández-Serantes, Patricia Blanco-Rodríguez and Alberto Otero-Pazos carried out the experiments and defined the measurements. José Luis Calvo-Rolle and Francisco Javier de Cos Juez interpreted the results and drafted the manuscript. Patricia Blanco-Rodríguez, Alberto Otero-Pazos and José Luis Calvo-Rolle supervised the experimental data analysis; they also contributed to the critical revision and improvement of the paper. All of the authors have approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sahu, P.; Mishra, D.P.; Panigrahi, D.C.; Jha, V.; Patnaik, R.L. Radon emanation from low-grade uranium ore. *J. Environ. Radioact.* **2013**, *126*, 104–114.

2. Llerena, J.; Cortina, D.; Durán, I.; Sorribas, R. 222Rn concentration in public secondary schools in Galicia (Spain). *J. Environ. Radioact.* **2010**, *101*, 931–936.

3. Barros-Dios, J.; Ruano-Ravina, A.; Gastelu-Iturri, J.; Figueiras, A. Factors underlying residential radon concentration: Results from Galicia, Spain. *Environ. Res.* **2007**, *103*, 185–190.

4. Welsh, P. Trials of radon remedies in a UK test house: An introduction. *Environ. Int.* **1996**, *22* (Suppl. 1), 1059–1067.

5. Lamonaca, F.; Nastro, V.; Nastro, A.; Grimaldi, D. Monitoring of indoor radon pollution. *Measurement* **2014**, *47*, 228 – 233.

6. Veleva, B.; Valkov, N.; Batchvarova, E.; Kolarova, M. Variation of short-lived beta radionuclide (radon progeny) concentrations and the mixing processes in the atmospheric boundary layer. *J. Environ. Radioact.* **2010**, *101*, 538–543.

7. World Health Organization. El radón y el Cáncer. Descriptive Note nº 291, 2009. Available online: http://www.who.int/mediacentre/factsheets/fs291/es/ (accessed on 9 May 2017).

8. Radon, C.; Sciences, C.; Studies, D.; Council, N. *Health Effects of Exposure to Radon: BEIR VI*; BEIR (Series), National Academies Press: Washington, DC, USA, 1999.

9. Nsibande, C. *Radon (RN) Measurement in Buildings*; LAP Lambert Acad. Publ.: Mannheim, Germany, 2011.

10. Environmental Protection Agency. EPA's Assessment of Risks from Radon in Homes, EPA 402-R-03-003., 2013. Available online: https://www.epa.gov/sites/production/files/2015-05/documents/402-r-03-003.pdf (accessed on 9 May 2017).

11. Sánchez, A.S.; Iglesias-Rodríguez, F.J.; Fernández, P.R.; de Cos Juez, F. Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *Int. J. Ind. Ergon.* **2016**, *52*, 92–99.

12. Sánchez, A.S.; Fernández, P.R.; Lasheras, F.S.; de Cos Juez, F.J.; Nieto, P.G. Prediction of work-related accidents according to working conditions using support vector machines. *Appl. Math. Comput.* **2011**, *218*, 3539–3552.

13. Instruction, IS-33, 21st of December 2011, from Nuclear Security Council, on Radiological Criteria for Protection Against Exposure to Natural Radiation, 2012. Available online: https://www.csn.es/images/stories/publicaciones/unitarias/normativa/is-33.pdf (accessed on 9 May 2017).

14. Korhonen, P.; Kokotti, H.; Kalliokoski, P. Survey and mitigation of occupational exposure of radon in workplaces. *Build. Environ.* **2000**, *35*, 555–562.

15. Scivyer, C. *Radon in the Workplace: A Guide for Building Owners and Managers*; Building Research Establishment: Garston, UK, 2011.

16. Maringer, F.; Akis, M.; Kaineder, H.; Kindl, P.; Kralik, C.; Lettner, H.; Lueginger, S.; Nadschläger, E.; Ringer, W.; Rolle, R.; et al. Results and conclusions of the Austrian radon mitigation project 'SARAH'. *Sci. Total Environ.* **2001**, *272*, 159–167.

17. Vázquez, B.F.; Adán, M.O.; Poncela, L.S.Q.; Fernandez, C.S.; Merino, I.F. Experimental study of effectiveness of four radon mitigation solutions, based on underground depressurization, tested in prototype housing built in a high radon area in Spain. *J. Environ. Radioact.* **2011**, *102*, 378–385.

18. Alphaguard-Professional Radon Monitor. Available online: http://www.saphymo.com/ (accessed on 22 March 2017).

sensors

MDPI

Ionizing Radiation Measurement Solution in a Hospital Environment

Antonio-Javier Garcia-Sanchez [1],*, Enrique Angel Garcia Angosto [2],
Pedro Antonio Moreno Riquelme [1], Alfredo Serna Berna [3] and David Ramos-Amores [3]

[1] Department of Information and Communication Technologies, Universidad Politécnica de
    Cartagena (UPCT), Campus Muralla del Mar, E-30202 Cartagena, Spain; pamr0@alu.upct.es
[2] Services, General Electric Healthcare España S.L.U., 28023 Madrid, Spain; enrique.garciaangosto@ge.com
[3] Department of Radiophysics, Hospital General Universitario Santa Lucía, C/Mezquita, s/n, Paraje Los
    Arcos, E-30202 Cartagena, Spain; alfredo.serna@carm.es (A.S.B.); david.ramos@carm.es (D.R.-A.)
* Correspondence: antoniojavier.garcia@upct.es; Tel.: +34-968-326-538

Abstract: Ionizing radiation is one of the main risks affecting healthcare workers and patients worldwide. Special attention has to be paid to medical staff in the vicinity of radiological equipment or patients undergoing radioisotope procedures. To measure radiation values, traditional area meters are strategically placed in hospitals and personal dosimeters are worn by workers. However, important drawbacks inherent to these systems in terms of cost, detection precision, real time data processing, flexibility, and so on, have been detected and carefully detailed. To overcome these inconveniences, a low cost, open-source, portable radiation measurement system is proposed. The goal is to deploy devices integrating a commercial Geiger-Muller (GM) detector to capture radiation doses in real time and to wirelessly dispatch them to a remote database where the radiation values are stored. Medical staff will be able to check the accumulated doses first hand, as well as other statistics related to radiation by means of a smartphone application. Finally, the device is certified by an accredited calibration center, to later validate the entire system in a hospital environment.

---

1. Introduction

The exposure of healthcare workers to ionizing radiation is a relevant concern due to the health risks involved [1]. Governments and authorities are working to foster ionizing radiation safety. In this way, recent directives (e.g., Council Directive 2013/59/Euratom) [2] have set basic safety standards to protect the health of individuals from dose exposures. Under these directives, healthcare workers in charge of radiology exposed to this type of radiation need to be monitored on a daily basis.

The most effective way of protecting medical staff who regularly work in controlled areas is the continuous measurement of the accumulated radiation level in their bodies. To this end, we have performed an exhaustive analysis of the current commercial radiation monitoring instruments, and divided them into two main groups: personal dosimeters and area survey meters. Personal dosimeters are small devices which healthcare workers wear on the body part most exposed to radiation. Thermo-luminescence based solutions are the most widely used technology where, after a period (usually a few weeks), the device is sent to a specialized laboratory to analyze the received doses. Therefore, the measurement is off-line, which restricts its usefulness. On the other hand, area dosimeters measure the radiation doses in real time, with most of them lacking Internet connections or database storage. These large devices are only placed in certain areas of the hospital (for instance, in the Nuclear Medicine department).

To overcome these drawbacks, a system has been designed, developed, verified and checked as a hybrid solution, valid as both a personal dosimeter for specific zones and an area dose meter. The entire system is based on open-source technology, integrating commercial off-the-shelf (COTS) electronic components including Internet connections via 3G, Ethernet or WiFi. The system functionality is the following one: a device collects the doses employing a commercial GM tube and dispatches them to a remote database hosted in a server. It is responsible for storing all the information received for each healthcare worker, immediately calculating accumulated, average and deviation doses, and providing a highly flexible solution. Also, medical staff are aware of these values in real time, thanks to a smartphone application. This code, as well as the remaining software developed for this work, is available in [3].

The device was verified at the National Dosimetry Centre (CND) in Valencia, Spain under reference laboratory conditions. The complete solution was tested at the Sta. Lucia University Hospital, located in Cartagena (southeastern Spain), using two devices. The first one was placed in a corridor next to the PET department and the second one at the control area of the Computer Tomography (CT) area. The results were compared to calibrated electronic dosimeters, demonstrating the correct procedure carried out by the system's measurement.

The rest of this paper is organized as follows: Section 2 introduces the technological background related to this research, where dosimetry concepts and a commercial dosimeter classification are thoroughly described; Then, the following Section 3 presents our proposal detailing both its hardware and software architecture, thus, the benefits of the solution offered in respect to the remaining commercial dosimeters are also highlighted. Section 4 deals with the procedure for verifying the device in hospital environments; Section 5 validates our system at Sta. Lucia University Hospital; results are analyzed and discussed, and finally, Section 6 summarizes the main conclusions.

2. Technological Background

2.1. Ionizing Radiation

Ionizing radiation is that with enough energy to ionize matter, by removing electrons from their states joined to their corresponding atom. It may be produced by radioactive substances, which emit such radiation naturally, or by artificial generators such as X-ray generators and particle accelerators.

The ionizing radiation sources naturally occurring in the earth's surface can be classified as alpha, beta, gamma or X-ray particles. Their features are explained in the following paragraphs:

- Alpha particles (α): They are usually generated in nuclear reactions or radioactive disintegration. Their capacity of penetration is low. They quickly lose their kinetic energy in the atmosphere, because they interact strongly with other molecules due to their great mass and electrical charge, generating a significant number of ions per centimeter of covered length. In general, they cannot pass through the thickness of multiple sheets of paper. However, if an alpha emitter enters the body it can be very harmful.
- Beta particles (β): They have greater penetration than alpha particles. They can be stopped by a few centimeters of wood or a thin sheet of metal. However, as with alpha particles, they can cause serious damage to the human body.
- Gamma particles (γ): Gamma radiation (γ) is a type of electromagnetic radiation [4]. Gamma rays can cause serious damage to the nucleus of cells, so they are used to sterilize medical equipment and food. Gamma rays easily pass through the skin and other organic substances; therefore, they could cause serious damage to internal human organs. X-rays are included under this category. They are also photons, but with a lower penetration capacity than gamma rays.

2.2. Recommended Criteria to Assign Dosimeters for Healthcare Workers

- The Spanish Health Protection regulation against Ionizing Radiation (Royal Decree 783/2001 (BOE 26/07/2001) defines exposed healthcare workers as those who, due to the circumstances

in which their work is carried out, are habitually/occasionally subject to the risk of exposure to ionizing radiation that may involve doses higher than the limits recommended for members of the public. Exposed healthcare workers are further classified into two categories, A and B:

1. Category A includes those who can receive an effective dose greater than 6 mSv (miliSievert) per official year, or a dose greater than 3/10 of the equivalent dose limit for eye lenses (45 mSv/year), skin and limbs (150 mSv/year). In Category A, exposed workers must use a personal dosimeter.
2. Category B includes exposed workers who rarely receive effective doses in excess of 6 mSv per official year, or 3/10 of the equivalent dose limit for eye lenses, skin and limbs. These workers are not necessarily required to wear personal dosimeters. It is sufficient for them to be controlled by a dosimetry monitoring system that guarantees that the doses received are compatible with their level B classification.
3. Non-exposed workers, such as members of the public health system, do not require any dosimetry control.

Table 1 summarizes the effective dose levels together with the equivalent doses in eye lenses, skin and limbs used in this staff classification.

Table 1. Dose level ranged by exposition category.

Healthcare Professional Category	Effective Dose	Eye Lense Dose	Extremity Dose
Exposed worker A classification	>6 mSv/year	>45 mSv/year	>150 mSv/year
Exposed worker B classification	<6 mSv/year	<45 mSv/year	<150 mSv/year
Non-exposed workers	<1 mSv/year	<15 mSv/year	<50 mSv/year

A minimum number of people working under ionizing radiation conditions in the healthcare environment are classified as Category A exposed workers, and would be required to be controlled by individual dosimetry. These workers are the following:

* Those who have to stay close to the radiation beam in interventional radiology and hemodynamics.
* Healthcare workers preparing and administering radioactive doses in nuclear medicine.
* Healthcare workers who are involved in the preparation of treatment and care of patients in radioactive metabolic therapy.
* Healthcare workers performing these functions in brachytherapy.

Table 2 shows the percentages of workers with accumulated annual doses exceeding the annual dose limit for members of the public (1 mSv per year, established by the General Radiation Protection Manual, CSN/SEPR/SEFM Forum, June 2003), classified by category and work type.

Table 2. Percentage of healthcare workers exceeding the annual dose limit according to their job type.

Job Type	Doctor	Medical Assistant	Operator	Other
Outpatient clinic Radiology	0.0%	0.40%	0.0%	0.0%
Hospital Radiology	2.7%	1.40%	0.5%	0.4%
Vascular Radiology	24.1%	5.90%	2.9%	1.5%
Radiotherapy	0.1%	2.60%	0.4%	0.7%
Nuclear Medicine	1.7%	50.70%	36.6%	13.0%
Interventional	2.2%	0.40%	0.0%	0.1%
Others	0.6%	0.30%	0.0%	0.0%

2.3. Dosimetry Classification

2.3.1. Personal Dosimetry

Dosimetry estimates the doses received by a given person. Depending on the body area you wish to estimate, personal dosimetry is classified into:

1. Dosimetry of the whole organism: It is implemented by a lapel dosimeter, which must be placed in a position that is representative of the most exposed part of the body surface. In those cases where a lead apron is required, the dosimeter should be placed under the apron.
2. Dosimetry of the upper extremities: It is implemented by wearing wrist or ring dosimeters. Recommended only in those cases where the doses to the hands can be notably higher than those of the whole organism, due to their proximity to the radiation beam.
3. Abdomen dosimetry: It is implemented by lapel dosimeters. Only used on the abdomen of exposed pregnant workers. Where a lead apron is required, the dosimeter should be placed under the apron.

There are a lot of personal dosimeters on the market. The most common ones are described below:

*   Thermo-luminescence dosimeter (TLD) [5].

It incorporates anodized aluminum foil with four thermo-luminescent detectors. These detectors are usually made of lithium fluoride activated with magnesium or calcium fluoride activated, in turn, with manganese.

The detectors store the energy received from ionizing radiation. In order to know the amount of radiation received by the dosimeter, it is necessary to heat it to a temperature of 300 °C, thus releasing the stored energy in the form of light. The amount of light emitted is proportional to the radiation dose received by each detector.

The lapel dosimeter is designed for the measurement of deep personal equivalent doses denoted as Hp (10) and superficial doses called Hp (0.07) [6], as well as for energy discrimination from incident radiation.

The main advantages are its low cost, easy handling, sensitivity, and that it does not depend only on environmental conditions. Furthermore, it is reusable: Once the dosimeter receives the radiation dose during a period of time, it can be employed again. However, its main drawback is related to that the radiation cannot be observed in real time, which negatively impacts on its applicability.

*   Film dosimeter [7].

It contains a small piece of radiographic film, placed between two metal filters (usually aluminum or copper) inside a plastic casing, in order to protect the film from light exposure. Some film dosimeters have two emulsions, one for low doses and the other for high dose measurements. These two emulsions can be on separate film substrates or on either side of a single substrate.

When the film is irradiated, an image of the protective box is projected onto the film. In this way, the amount of radiation received is proportional to the optical density of the film being exposed, compared to the optical density of previously calibrated films.

The main advantage is its functionality. Film can provide information on the exposure conditions such as the direction of the incident radiation or suspected contamination. The main drawback is that the reading is estimated, not immediate, and not reusable. It also depends on external conditions, such as humidity.

*   Radiophotoluminescent Glass Dosimeter (RPLGD) [8].

It consists of a housing that incorporates a radiophotoluminescent (RPL) glass in the center. The RPL crystal material is activated silver with crystallized phosphate glass (P_2O_5). When this

material is exposed to radiation, stable luminescent centers are created in the silver ions, denoted as Ag° and Ag^{2+}. These ions emit light when stimulated. To produce this stimulation, the crystal is irradiated with ultraviolet light, measuring the light emitted with a photo multiplier tube. So, the intensity of the light emitted will be proportional to the radiation dose received.

The main advantage is its reusability. In the reading procedure, the stored dose is not eliminated. Therefore, it allows us to accumulate the radiation dose in the very long term. In addition, it is not sensitive to ambient temperature, which favors the measurement precision. The main drawback is that it doesn't measure data in real time and it doesn't export data.

- OSL (Optically stimulated luminescence) dosimeter [9].

These dosimeters incorporate a sheet of carbon-activated aluminum oxide, located between filters, to obtain qualitative information on the conditions during exposure. To know the amount of radiation dose received, the aluminum oxide is stimulated at specific frequencies by means of a laser. In this way, the intensity of the light emitted will be proportional to the dose of radiation received.

As in previous cases, its main advantage is its reusability. It can be reused a few number of times without losing sensitivity. On the other hand, commercial devices lack remote connectivity and data storage.

- Electronic Personal dosimeter [10].

It uses electronic sensors and signal processing, and shows the radiation dose normally received in microSievert (µSv). This dosimeter either has a miniature Geiger-Müller tube or is equipped with silicone sensors.

The main advantage is that it can continuously display the accumulated dose on the screen, so it is often used as an alternative in emergency cases in which it is necessary to know a specific amount of radiation within a short period of time. Additionally, it can incorporate alarms if a previously programmed accumulated radiation dose is exceeded during a certain period of time. It works with batteries. The main drawbacks are the cost and the lack of connectivity.

- Pen/pocket dosimeter [11].

It is a small ionization chamber, cylindrical (approximately 2 cm^3) in shape, filled with air and equipped with a central electrode. It is shaped like a ballpoint pen and contains an ionization chamber with a quartz fiber electrode that functions as an electroscope. By means of an optical arrangement, the response of the radiation action can be observed.

The main advantage of this equipment is that the reading of this dosimeter is straightforward and very easy to use. The main drawback is the fact that it has a poor range of use and low sensitivity compared to other dosimeter systems. It requires daily radiation monitoring and daily charging. Occasionally, it can even be discharged when hit. Therefore, for some years now, its use has been reduced significantly.

Finally, Table 3 summarizes the radiation ranges for each personal dosimeter type listed above.

Table 3. Radiation range for personal dosimeter.

Dosimeter	Range
Film and TLD dosimeters	100 µSv–10 Sv
OSL and RPL dosimeters	10 µSv–10 Sv
Self-reading pocket dosimeters	50 µSv–0.2 Sv
Electronic Personal dosimeter	0.1 µSv–10 Sv

2.3.2. Non-Personal Dosimetry

Non-Personal Dosimeters are devices that estimate the received dose in a particular place [12]. They can be classified as:

1. Area dosimetry: It is performed by means of lapel dosimeters. It is reserved for the estimation of doses in places or areas where workers are exposed under category A.
2. Dosimetry of rotating workplaces: It is performed using lapel and/or wrist dosimeters. It is focused on estimating doses in those workplaces where an activity is occasionally performed, and category B workers could be exposed.
3. Research dosimetry: It is carried out using lapel and/or wrist dosimeters. It is requested by a Radiation Protection Service to measure doses in all those situations not covered in the previous cases.

The following paragraphs will describe the main types of area dosimeters used in hospital environments today. To this respect, they can be divided into two main groups: (i) gas detectors and (ii) solid-state detectors.

*   Gas detectors

Gas detectors are based on the direct gathering of the ionization phenomenon caused by a particle when passing through a noble gas, enclosed between two electrodes subjected to a potential difference. This type of gas is used because the dose rate that can be monitored should be as high as possible. In this group, there are three main types of detectors: (1) ionization chambers; (2) proportional counters and (3) Geiger-Müller counters. Each one accomplishes specific functions, implying different shapes and sizes.

1. Ionization chamber [13]. It is the most straightforward gas detector, considered as a plane-parallel capacitor, in which the region between planes is completely filled with a gas, preferably with air. The electric field in this region prevents the recombination of the ions with electrons, which involves electrons moving towards the positive electrode, while positively charged ions move to the negative electrode. The applied voltage sets the velocity at which electrons and ions move to the chamber electrodes. For instance, for a typical voltage value of about 100 V, ions move at speeds of 1 m/s. Under these conditions, ions take up to 0.01 s to get through a 1 cm thick chamber (note that electrons are more mobile than ions and they will move 1000 times faster). An inconvenience is that these temporal values are excessively long for the time dealt with in radiation detection (e.g., a Computer Tomography).
2. Proportional counter [14]. It consists of a metal case filled with a noble gas such as argon or xenon, with a very fine thread crossing the center. The wire (anode) is set at a high potential difference in relation to the box, so that an electric field constantly passes through the gas. The process is the following: when an ionizing particle, such as an electron, goes through the gas, it releases electrons from its atoms (it ionizes the atoms), leaving behind a positively charged ion and a free electron. The free electrons generated in the gas are accelerated to the anode and their amount is proportional to the initial energy of the particle or X-ray. Unlike simple ionization chambers, the potential difference used in proportional counters is much greater, so the electrons accelerated towards the anode have enough energy to produce secondary but proportional ionizations (hence the name of the device), causing an electronic "cascade". The electric current or voltage generated in the anode can then be measured and digitized and, as mentioned above, the voltage or charge is proportional to the energy of the particle or X-ray incident. To be able to observe individual pulses, we must increase the applied voltage (which means exceeding 1000 V, its main inconvenience). To this respect, the largest electric field is capable of accelerating electrons so that they can generate secondary ionizations. Accelerated secondary electrons produce new ionizations, eventually creating an avalanche or cascade of ionizations.

3. Geiger-Müller counter [15]: If the electric field increases further, the Geiger-Müller region is reached. In this case, secondary avalanches are generated anywhere in the tube, caused by photons emitted by atoms stimulated in the original avalanche. These photons move relatively far from the original avalanche location and, in a short time, the entire tube is involved in the process. Counters based on this principle are known as Geiger-Müller meters. Since the entire tube is involved in each initial avalanche, there is no information on the energy of the original radiation (all incident radiation produces identical output pulses valued around 1 V). So, Geiger-Müller counters are employed as pulse counters. However, in these counters there is a serious problem. During their movement into the tube, ions can be accelerated, reaching the anode with enough energy to release electrons and starting the process again (due to the nature of the multiple avalanches). To overcome this inconvenience, a second type of gas, called quenching gas, is added. This gas is made up of complex organic molecules such as ethanol (while the primary gas is usually made up of simple molecules, such as argon). A typical mixture of gases will be: 90% argon and 10% ethanol. The molecular nature of this added gas prevents the appearance of continuous avalanches. Geiger-Müller sensors allow the obtaining of accurate radiation doses with a remarkably low cost, and its integration into any electronic system to manage the information and dispatch it to remote places is a straightforward process. These are the main reasons why the Geiger-Müller counter has been adopted as part of our proposal.

- Solid State Detectors

The solid state detector, also known as Semiconductor Radiation Detector, is a radiation detector in which a semiconducting material (silicon or germanium) constitutes the detecting medium. Solid detectors have higher densities that provide reasonable absorption probabilities for a normal detector size. Two types will be described: (1) scintillating counters and (2) semiconductor detectors:

1. Scintillator counters [16]. The scintillator counter works as follows:

- Incident radiation interacts with the atoms and molecules of the material, stimulating them.
- Stimulated states are de-energized by emitting visible (or nearly visible) fluorescent light.
- The light reaches a photosensitive surface by pulling out photoelectrons.
- The electrons accelerate and multiply in number to form an electrical pulse.

There is a wide variety of available scintillators and photomultiplier tubes, depending on the type of application. The properties to be considered in the selection of the material include the fraction of incident energy that appears as light, efficiency (the probability of radiation being absorbed), response time and energy resolution.

2. Semiconductor detectors [17]

Semiconductor solid materials (germanium and silicon) are alternatives to scintillators for building radiation detectors. When a type-p semiconducting material is in contact with a type-n semiconducting material, the electrons of the type-n semiconductor can diffuse through the junction in the type-p semiconductor and combine with the vacant ones. In the vicinity of the p-n junction, the charge conductors are neutralised, creating a region called the depletion zone. The diffusion of electrons from the n-type region leaves behind ionized, fixed donor states, while in the p-type region there are fixed acceptor states charged negatively. Thus, an electric field is created, which ultimately makes diffusion phenomenon advances impossible. A typical p-n junction of a diode is formed.

If any radiation penetrates into the depletion zone and creates an electron-hole pair, the result is very similar to that of an ionization chamber. The electrons flow in one direction and the vacancies/holes in the other. The final number of collected electrons creates an electronic pulse whose amplitude is proportional to the energy of the radiation.

Concerning the manufacturing of these detectors, it is possible to initiate them from a p-type semiconductor in which lithium atoms are diffused. The n-type layer created by producing detectors such as Ge (Li) or Si (Li) is in the order of 1 mm thick, which is easily penetrable by medium-energy gamma rays (the range of a photon of 100 kiloelectronvolt in Ge is about 4 mm and in Si about 2 cm). However, in the case of charged particles the range is much smaller (e.g., for 1 megaelectronvolt electron the range is 1 mm in Si and Ge; for 5 megaelectronvolt alpha particles, the range is only 0.02 mm in both) and there is a 1 mm thick layer to prevent the particles from accessing the depletion zone.

The main advantages are that the time required to collect the load of large detectors is in the range of 10–100 ns, depending on the shape of the detector (flat or coaxial) and the point of entry of the radiation with respect to the electrodes. This time is less than that obtained in an ionization chamber, since here the path followed by the created loads is reduced by several orders of magnitude. Another clear advantage of these types of solid detectors is related to energy. Less energy is required to create an electron-hole pair (~3.6 eV/pair in Si at 300 °K), thus, excellent energy resolution is obtained. An inconvenience is the lack of connectivity in commercial devices.

2.4. Discussion

As for personal dosimetry, thermo-luminescent dosimeters are the most widely used in Spanish hospitals. However, as previously mentioned, their main drawback is the lack of real time measurements. Each personal dosimeter will be transferred at least once a month to a specialized center (in Spain, the National Dosimetry Centre—CND) which will heat it to up to 300 °C. The amount of light thus emitted is proportional to the radiation dose received.

During this period of time, in which the CND estimates the level of radiation received by the dosimeter, health personnel will have another dosimeter made available to measure the radiation of the following month. So, a monthly radiation level is obtained and there is no possibility of knowing specific periods of time with higher doses during this time because the radiation dose is accumulated.

As for area dosimetry, most of the devices used in a hospital environment lack any connection to remote systems in order to receive the dose. Usually, the only way to get the data is to be in front of the dosimeter and to check the dose.

On the other hand, there is an important concern regarding data storage. These types of devices store the average radiation values in specific temporal periods. Data are released after each period to restrict data storage. To carry out a later evaluation of emitted radiation, it will not be possible to access radiation values outside of the current period.

In addition, if users want to store the data directly on a PC, it is necessary to have it in the same room as the area dosimeter, which in some cases, can be a complex or impractical task, for instance, in a hospital corridor or in a computer tomography room (cases of study proposed in this work).

In summary, the major drawbacks detected in the current instrumentation for ionizing radiation measurement are described in the following points:

1. Limited storage. There are devices that can store the radiation captured internally. However, due to the limited memory capacity of these devices, it is not possible to have a large amount of stored data.
2. Scarce Flexibility. Most devices do not allow the collection of data in any other way than that established by the manufacturer. In the particular case of personal thermo-luminescent dosimeters, where the stored data can be accessed, radiation collection must be stopped and be replaced by another device while accessing the data.
3. High cost. Those alternatives that have some storage capacity and offer results in accurate, real time radiation doses (especially for area dosimeters) have a clear impact on the cost. Therefore, these solutions are addressed to public entities or big, private companies.
4. Power supply. A large number of dosimeters described in this section are supplied by batteries. In the case of failing batteries, the received dose is not computed and a health risk could result.

5. Manufacturer dependent. A priori, any modification of the device firmware to adapt it to new conditions of the application/environment is not possible.

3. Proposal

3.1. General Overview

The radiation measurement proposal is a step forward, which solves diverse problems presented by current radiation systems. Figure 1 shows the main system components where the entire system integrating wireless communications, database, mobile application and measurement device have been highlighted.
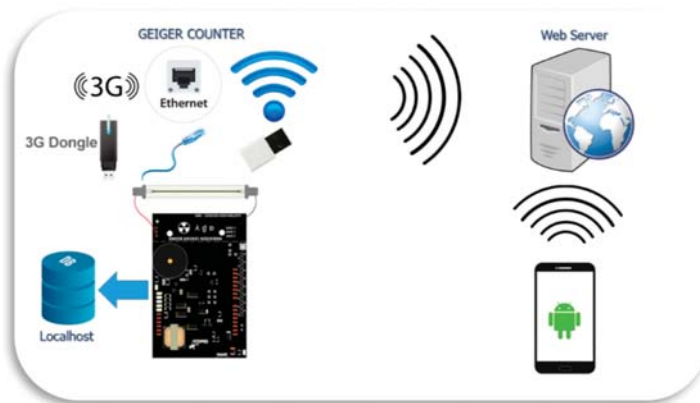
Figure 1. System components.

As a general approach, the radiation is measured with a commercial Geiger-Müller tube [18], receiving two values: (i) the effective dose-rate value (μSv/h) and (ii) the counts per minute (CPM). These values are stored in two redundant databases: one hosted on a remote web server and the second in a database implemented in a Raspberry Pi [19] (localhost), which in addition, provides memory storage, processing and connectivity. Note that this commercial detector can be replaced by another; without thereby the system functionality lessen. Therefore, it is not an objective of this work to characterize and study this sensor, tackling these concerns in future works. For instance, it should be interesting to model the GM directivity, to later implement it in the Raspberry. Furthermore, in the case of using detectors such as the solid state ones (because the application requires it), environmental parameters (temperature, pressure, humidity, etc.) must be considered to adjust its operation. As in previous case, the detector adjustment will be developed in the corresponding software module of the Raspberry. Under these premises, to have of a COST-based system facilitates the design of specific solutions in accordance with application or service requirements.

The set of radiation measurement, capture, and storage procedure is performed in periods of one second, i.e., the sensor measures radiation doses and counts per minute (both values form the data) at one-second intervals and stores both values in the Raspberry Pi and web server. In this way, it provides a versatile solution for the storage problem that current systems are facing today.

The data uploading to the database of the web server is carried out through 3G, Ethernet or WiFi technology, which ensures flexibility and reliability in the face of any contingency in the transmission of information (e.g., falling 3G network, WiFi router failure, etc.). Furthermore, the user has access to a smartphone application, which is able to show information in real time, or statistics of radiation values computed in remote web server: Daily measurements, averages, deviations and so on.

The user could access data through the web server, as well as being able to export them to other formats, for instance, a spreadsheet (e.g., Excel) to analyze the data in more detail. In addition, access to the data does not stop the operation of the device, which is the main drawback of some of the current systems.

The system could be used either as a personal dosimeter and/or as an area dose meter. Regarding area dosimeters, it is obvious, since its functioning is the same as Geiger counters. To use it as a personal dosimeter, healthcare staff have to log in and register in the smartphone application. So, for each registered user, the received radiation is stored in the corresponding databases, to later consult instantaneous, accumulated, mean radiation values of dose. In this regard, it provides a real-time solution in personal dosimetry, avoiding long waiting periods in order to know the received radiation data, as occurs in many of the current systems (e.g., thermo-luminescent dosimeters) which are connectable from any place through Internet access.

The Geiger-Müller detector is powered by the Raspberry Pi, so this device just has to be connected to the power supply, thus avoiding the energy limitations of other electronic dosimeters. Once connected to the power supply, the first step that the system performs is to connect to the remote database and start receiving radiation values.

Among all the aforementioned technologies, the system has a Geiger-Müller tube, therefore a Geiger counter. A sensor of this type usually triggers pulses, which is not an energy value, so it would not be useful for knowing the dose of radiation received by health personnel. Users could get the radiation value multiplying CPM by a conversion factor (K):

$$CPM \times K \; (\mu Sv/h) \tag{1}$$

This conversion factor is achieved by means of a thorough calibration process by the manufacturer. In our case, this factor is valued in 0.008120 [18], thus the radiation received is derived in units of micro Sieverts per hour. The app shows Counts Per Minute (CPM), and Radiation Value ($\mu Sv/h$) at the same time. This is a new feature that commercial Geiger counters do not provide today.

It is important to note that the system is able to estimate the radiation value in real time and send this information to a database. This system contributes to monitoring radiation in two ways: (i) by acquiring/collecting radiation values in real-time and (ii) in the ability to store each radiation value labeling it with its corresponding date/time along with the user receiving the dose. Figure 2 shows how the proposed system works.

Nowadays, there are some alternatives, including GM detectors, which show radiation in real time, and in which data are also stored. To accomplish these functions, in [20], the device only ensures connectivity to a PC via USB, while the system proposed in [21] sets a proprietary system up to send data to a PC. Therefore, comparing with our proposal, the lack of remote communications incurs in a clear disadvantage. Furthermore, an additional advantage to our system, contrasted with similar systems such as the one mentioned above, is its low cost. Currently, this type of system is valued at thousands of Euros. In this sense, the work [22] tests a medical environment through a proprietary and costly solution, although in this case, remote communications are considered. We offer an alternative capable of improving current services at a much lower cost (the prototype proposed here does not exceed 200 Euros).
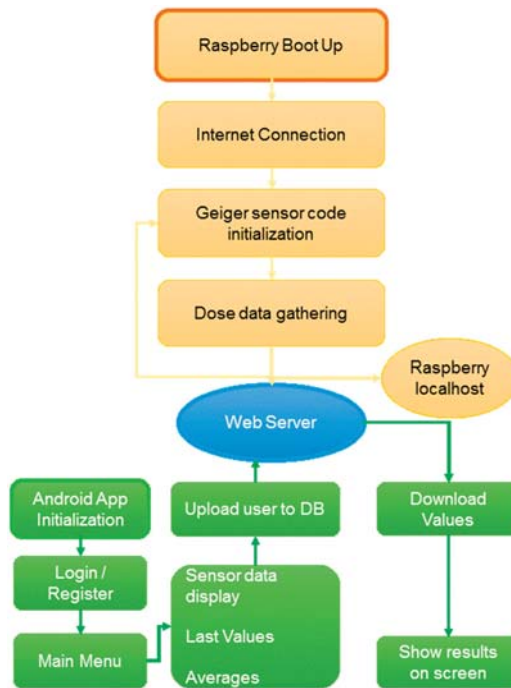
Figure 2. System's block diagram.

Different low-cost dosimeter-based solutions can be found in [23,24]. In [23], a dosimeter prototype captures the dose rate at 5 Hz through a CMOS camera. However, this proposal lacks communication support which prevents, for instance, statistics calculations. The work [24] leads to security concerns in power nuclear plant communications when data include radiation values. To face this, authors propose a solution based on XBee module communication, a well-known Wireless Sensor Network (WSN) device. Authors carefully explain aspects such as modulation, packet format or firmware, and how this device takes actions when malicious attacks occur; although they never evaluate its operation under the proposed nuclear environment.

The work [25] presented a low-cost, open-source and remote communication solution. In this paper, authors analyzed the variation of the energy in relation to the dose rate in a commercial and personal X-rays detector denoted as POKEGA. It is conceived to monitor the radiation received by members working in interventional radiology. The POKEGA solution was proposed in other works [26,27]; it is composed of a photodiode detector capturing the radiation, and a smartphone responsible for dispatching the dose (together with, for instance, GPS position, time, date, etc.) to Google docs, to later visualize the emplacement where each dose was collected. Both, photodiode and smartphone are connected by a cable. In comparison with our system providing reliability and robustness to the communication thanks to diverse wireless interfaces, POKEGA depends on the smartphone and its telephony connection. Furthermore, our system stores the monitored radiation both in a remote server and in the local device (Raspberry).

In summary, our system could be used as a personal and area dosimeter. It provides the advantages of displaying data in real time and storing it, either at an area level or for a specific user. In addition, the data reading procedure is more suitable and quicker than in the case of current personal dosimeters.

These features are also ensured by its connectivity (USB, Ethernet, WiFi or 3G). So, this heterogeneous connectivity allows us to know the amount of received doses outside of the measurement location. It is notable that its connectivity, together with the ability to connect it to electrical current, guarantees robustness and reliability. Finally, the device has been verified by the CND simulating hospital conditions and validated in a real hospital environment, giving accuracy to our proposal as a real alternative.

3.2. Advantages of Our Proposal in Respect to Current α, β, γ Radiation Meters

The main advantages of our proposed system in respect to other proposals are the following:

1. Low cost. In comparison to other systems, the features and functions offered by our system are more competitive in terms of cost.
2. Double functionality. The device can be used for both personal (in specific places as, for instance, the CT room) and area dosimetry.
3. Real time doses. This system is capable of displaying the dose values in real time either as a personal or area dosimeter.
4. High storage capacity. In contrast to other current systems, our system has a large storage capacity, due to the database usage.
5. Statistics. Several statistics are calculated immediately by the remote server, available for users at any time.
6. Connection alternatives. The device is able to connect to remote databases through a set of interfaces (USB, Ethernet, WiFi and 3G).
7. Data recovery. If the device loses the connection to the remote database, all radiation values are stored in the device until the connection is restored.
8. Flexibility and modulation. The system is built using general-purpose and open-source technologies, which allows software/hardware modifications.
9. Easy access. Users can check the data everywhere, due to the fact that the interface is a smartphone app that is connected to the remote database.
10. Reliability. Different connection alternatives, replicated databases, and continuous energy supply contribute to a robust and reliable system.
11. Verification. The device has been verified in CND, under standard calibration protocols.
12. Easy handling and installation. Once the power is turned on, the device is automatically set to capture radiation, store it on the system and connect to the remote web service.
13. Authentication. Security tasks will be implemented in a future work; however, the user must register to access dose values.
14. Validation. This proposal has been validated at St. Lucia University Hospital in different environments, which will be further described in Section 5.

Finally, Table 4 summarizes a comparison among different current radiation meters. This comparison is extended to include our proposal.

Table 4. Comparison among different dosimeters including our system.

Device	Real Time	Reusable	Low-Cost	Internet	Database	Reliability	Smartphone App	Open Source
TLD Dosimeter	No	Yes	Yes	No	No	Yes	No	No
Film Dosimeter	No	No	N/A	No	No	Yes	No	No
RPLGD	No	Yes	N/A	No	No	Yes	No	No
OSL Dosimeter	No	Yes	N/A	No	No	Yes	No	No
Personal Elec. Dosimeter	Yes	Yes	No	No	No	Yes	No	No
Pocket Dosimeter	Yes	Yes	Yes	No	No	No	No	No
Ionization Chamber	Yes	Yes	No	No	N/A	Yes	No	No
Proportional Counter	Yes	Yes	No	No	N/A	Yes	No	No
Geiger-Müller Counter	Yes	Yes	No	No	Yes	Yes	No	No
Scintillator Counter	Yes	Yes	No	No	N/A	Yes	No	No
Semiconductors detector	Yes	Yes	N/A	No	N/A	Yes	No	No
Instadose-Mirion	Yes	Yes	N/A	No	Yes	N/A	Yes	No
Dosicard-Canberra	Yes	Yes	N/A	No	Yes	N/A	No	No
PROPOSED SYSTEM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

3.3. In Detail: Description HW/SW of the System

In the following sections, the system proposed here will be discussed in depth from a twofold perspective: (i) hardware components used, their connection/integration and (ii) software implemented.

3.3.1. Hardware Description

The device hardware is built as the integration of different printed circuit boards. These include:

1. A Raspberry Pi.
2. The radiation sensor board.
3. The Bridge board that connects the Geiger counter to the Raspberry Pi, denoted as Arduino Shield Connection Bridge.

Figure 3 shows the Raspberry Pi and Geiger Sensor, while Figure 4 depicts the entire system integrated and connected.

Figure 3. Raspberry Pi and Geiger Sensor (**left**) connected to the Bridge board and the Geiger Sensor board (**right**).

Figure 4. End device.

In the following paragraphs, we describe the main components of the system in detail.

- Raspberry Pi [19]. It is a low-cost embedded board developed by the Raspberry Pi Foundation in the UK. Its size is approximately that of a credit card and encapsulates a Broadcom BCM2835 chip with an ARM processor up to 1 GHz speed. In our work, the Raspberry Pi B++ model has been selected for its low cost and ability to carry out all the functions of our system. Note that a MicroSD card including the operating system (Raspbian), is inserted in the back of the Raspberry, which will be explained later.

  The following components of the Raspberry Pi are employed and configured in our development:

  — A/V connector power boards to operate (5 V).
  — USB ports used to connect the 3G dongle or wireless adapter. In addition, if the Raspberry needs to be programmed, they will be used to connect the mouse and keyboard.
  — A General Purpose Input/Output (GPIO) port consisting of 40 pins allows the Raspberry Pi to communicate with external elements. In particular, GPIO pins are managed and handled directly through code from the intermediate Bridge board.
  — An Ethernet port joins our system to the Internet.
  — An HDMI port connects the Raspberry to a monitor, for configuration tasks.

  The following functions are carried out using the Raspberry:

  1. Executing the code corresponding to the Geiger counter in order to display/visualize the dose values.
  2. Connecting to the Internet via Ethernet, WiFi adapter or 3G dongle through USB.
  3. Enabling the automatic execution of the Geiger counter code and automatic internet connection.
  4. Running a localhost, including its corresponding database, which allows us to have a backup database.
  5. Ensuring that collected doses are uploaded to the database on the remote web server every second.

- Raspberry Pi to Arduino Shield Connection Bridge [28]. Since the connection between the Geiger Sensor Board and Raspberry is not possible, an intermediate board or bridge is required. This board is called Raspberry Pi to Arduino Shield Connection Bridge, but for the sake of brevity, we will denote it a bridge board.

The goal of this bridge board is to facilitate the integration of modules designed for Arduino on Raspberry Pi. This allows us to attach analog and digital sensors using the same pins for Arduino, but with the power and capacity of a Raspberry Pi. So, on the back of this main board, a bridge board will be connected through the GPIO pins. The board including the Geiger counter will be connected to the pins I/O, UART, SPI and I2C, belonging to the bridge board.

- Geiger Sensor Board [18]. This board is responsible for measuring radiation values using a Geiger-Müller tube. The radiation board is composed of two parts: the supply side and the signal side. The supply side is responsible for ensuring the high voltage required by the Geiger-Müller tube (400 V–800 V). The signal side adapts the output analog pulses from the GM tube to digital data understandable by the Raspberry Pi.

To achieve high voltage in the GM tube, an oscillator connected to a voltage multiplier consisting of diodes, transistors, resistors and capacitors is employed. This electronic circuit guarantees a potential difference of 500 V. In the particular case of requiring voltage values above 500 V to feed the tube, they will be obtained from zener diodes connected in series.

- 3G dongle. When connecting to the Internet via 3G, it is necessary to put a 3G dongle into the interface USB of the Raspberry Pi. Although there are several models compatible with Raspberry Pi, the selected dongle [29] together with a prepaid card have allowed the use of the device anywhere in the hospital environment. In addition, in the specific case of not having access to WiFi or Ethernet in the hospital under study, or when the system was verified at CND, the use of the 3G dongle has been essential to dispatch the captured radiation data to the remote web server.
- WiFi Adapter. This component is placed in a USB port, providing access to the Internet thanks to the WiFi hospital infrastructure. It should be noted that our device ensures an Internet connection (and therefore, the dose information for each user in the remote web server) by selecting of one these technologies: WiFi, 3G and Ethernet.

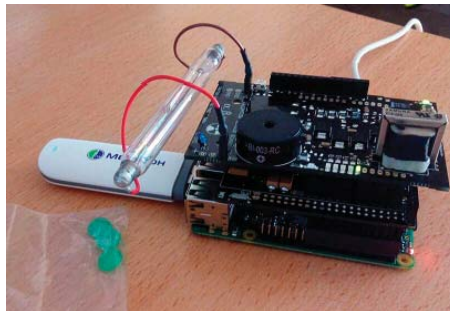Figure 5 shows the complete system with 3G USB dongle connected.

Figure 5. Device with 3G dongle attached.

Figure 6 depicts the complete system inside the protection cover.

Figure 6. Cover case including device.

3.3.2. Software Description

1. Operating System. The operating system (OS) for Raspberry Pi is denoted as Raspbian. In particular, Raspbian Jessie was the version installed, because it offers multiple functionality and documentation in terms of capturing data and communication issues.
2. Geiger sensor code. This code, implemented in C language, quantifies the counts per minute and equivalent radiation measured by the GM detector on the board. Raspberry Pi executes the programmed code obtaining the dose data from the Geiger detector through its GPIO pins. Since the sensor uses the same connections as Arduino, it is necessary to include the arduPi library codes provided by the manufacturer.

Figure 7 shows a flowchart of the way a Geiger operates (and its associated code).
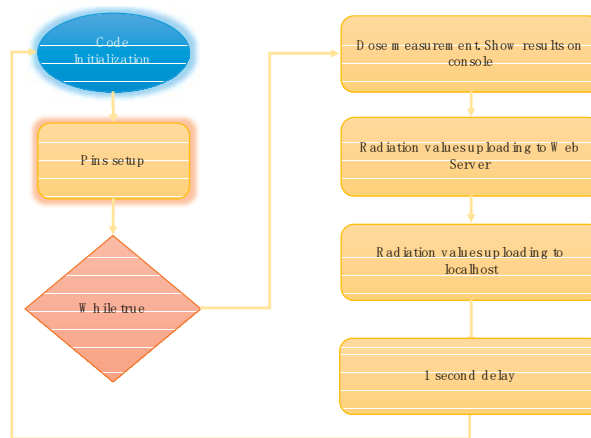
Figure 7. System general diagram.

Every time the code is initiated, the application remains waiting for a stimulation to occur in the GM tube, increasing the number of counts. Then, the application enters into an infinite loop, in which the radiation values are continuously captured in periods of one second. In reference to the previous figure, each of the blocks is discussed:

- Radiation measurement: display by console terminal

As mentioned above, the app shows counts per minute (CPM) and radiation dose (μSv/h). Firstly, it is necessary to calculate the number of counts per minute and to later obtain the equivalent radiation value related to the CPM calculated, multiplying the CPM by the conversion factor provided by the manufacturer. The number of counts corresponding to the GM tube stimulation is stored in a specific variable. The temporal interval lasts one second. In order to obtain equivalent CPMs, it will be necessary to multiply the number of counts per second by 60 (countPerMinute = 60 $*$ count).

Finally, before the radiation values are stored in the local and remote databases, the number of counts is reset (variable is set to zero), so when the loop is restarted, a new count value is obtained.

- CPM and radiation value upload to remote web server

Once CPM and radiation values have been obtained, the next step is to upload these data to the MySQL remote web server database. To this end, it is necessary to use a method which connects to the server by IP, user name and his/her password. Then, the programmer/developer selects the database

which connects to it. Note that each device will have its own database on the web server. Thus, each device is independent of the other ones.

On the other hand, to upload the CPM and radiation values, a SQL query is implemented. In this case, programmers dispatch an INSERT type query, including countPerMinute (CPM) and radiationValue (equivalent radiation) values to the remote database. Finally, once these values have been stored, the connection is closed.

The results displayed by the smartphone application will be the data (together with their statistics) stored in this database installed in the remote web server.

- CPM and radiation value storage in localhost

The dose values are also stored in the localhost database.

Any error detected in the remote web server will be mitigated by the localhost. That is, those radiation values which were lost in the period of non-operation of the web server will be recuperated by the localhost. This function is carried out by a method which shares the same code as the previous one, except for the IP address.

- End of loop

Finally, once data has been uploaded to the web server and stored in the localhost, there is a one second delay in the following set of measurements. This delay implies an appropriate time interval, from the time the radiation level is measured by the Geiger detector, to the time the Android application displays this value on the screen. If the delay decreases to values inferior to one second, the GM detector would be able to measure more accurately. However, there would not be enough time to upload the data to the web server and download it onto the app. It is notable that other systems are capable of showing radiation levels, but taking more time, demonstrating that this system is more accurate in adjusting the intervals to radiation samples.

Finally, an automatic code startup has been implemented. To avoid the manual execution of the code in Raspberry Pi (resulting in inefficient application in hospital environments), a script has been programmed to automatically start the code every time the Raspberry Pi is connected to a power supply or is booted. To this end, as Raspberry Pi is a Unix system, the script will be created within the */etc./init.d* directory, in which the start and stop commands are specified. Additionally, the different interfaces to access Internet (WiFi, 3G and Ethernet) will also connect automatically, without requiring support from specialist personnel or developers.

3. Servers. A server is required to store the dose data collected by the device. The one selected for both the remote web server and the localhost is Apache [30], which is a free-service and open-source HTTP server with the same features as other high-cost competitors. The Apache web server is responsible for storing the GM-captured radiation data and managing them in an Android application. In this case, the Android application accesses this server, both to log in or register a new user, and to display the radiation measurements/statistics. The remote web server requires Internet access, so, it must have a domain and therefore, a public IP address. In a certain way, both the device and the Android application connect to the server through this public IP. Under this premise, note that a web server allows us to access data anywhere. Using the Android application or a PC, we check the radiation values at any time and place, even if the detector fails or turns off. This also allows us to control the login and registration of users, avoiding, for instance, that two different users employ the same user name. The Apache localhost is configured as a backup service, lacks an Internet connection, and data are recovered via USB.

4. Database. They are responsible for storing both radiation data from the GM detector and information from user registers. PhpMyAdmin [31] has been selected as the tool to generate database, both in the local server and the remote web server. PhpMyAdmin written in *PHP* ensures the database management and handles the MySQL facilitates. For instance, the user can

create/delete databases, generate/delete/modify tables, delete/edit/add fields, execute any SQL statement, handle keys in fields, manage privileges, and export data in various formats.

Keeping in mind that the devices are independent from each other, labeling each device with its own unique identification number means that each device has its own database. The result is a more intuitive and appropriate management of different devices deployed in a hospital, thus facilitating the data recovery and the data security.

The fields composing the databases are, as well as the CPM and radiation value, the following: (i) an identifier for each measurement uploaded; (ii) a timestamp of the captured dose, both coming from the device; and (iii) registered users, including user name, password and email data, dispatched by the smartphone application. Regarding the latter, the goal is for each hospital to have its own databases of users, which would be independent of the others.

5. Android application. The application developed for smartphones is in charge of displaying all the radiation results captured by devices. This application has been implemented in Android because, among other factors, it is an open source. Furthermore, Android makes the simulation and testing of different versions of our application on different devices easier, without the requirement of them having been previously developed in emulators. The development of the app is divided into two main parts: The appearance, focused on the graphic interfaces and the specific functions of each of them.

Figure 8 presents a flowchart including the overall functionality of the app.
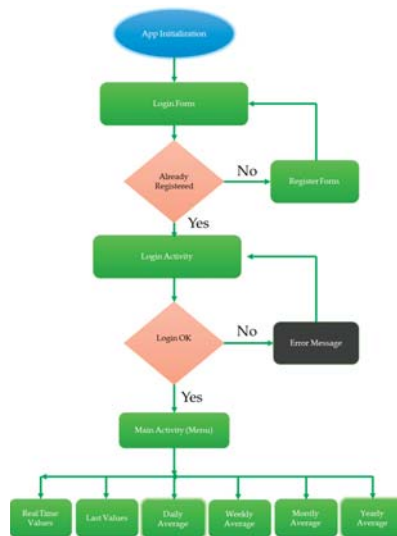
Figure 8. App flow-chart.

The application links the database through SQL queries. These communications are asynchronous due to the fact that the server runs diverse processes, so this issue will stop the main application if it is coded in a synchronous way. To implement asynchronous tasks, the *AsyncTask* class [32] is employed. This communicates with the process of the main thread, only performing operations and showing the results in the main thread.

An advantage of our implementation is the reduction of the number of database accesses. So, the application is more efficient and dynamic, decreasing the probability of connection errors. To this end, the *Intent* class [33] is in charge of sending the user name and device number, among

different activities, without requiring remote connection to the database. Therefore, SQL queries are only demanded when users want to obtain some type of results, such as radiation statistics. Finally, in the following subsections, the appearance and function of each of the parts in which the app has been divided will be detailed, according to the general diagram shown in Figure 8.

App Initialization

Each time the app is started, an image, like in Figure 9, presenting a login form, appears.

Figure 9. Login screen.

As shown, this form inputs the user name, password and device number. First of all, the user name linked with the password is required. Those values are stored in the database.

The device number is the identifier assigned to each device. Each hospital has a specific number of catalogued devices, allowing the user to select the device which they want to receive radiation information from. This selection is manually accomplished when logging in, and depends on the area where the staff is working. Therefore, by filling in the device number field, the app connects to the database corresponding to that device. This is possible because, as discussed above, each healthcare worker is exposed to radiation in an enclosed hospital area, and it will not be necessary to modify the device number every time.

On the other hand, if the user doesn't input data in the form fields or a device number is not one of those assigned to the hospital, an error message will be shown, alerting the user. The *Inputvalidation* class is in charge of these functions. However, if the proccess is correct, the *Login* button is enabled and the user name and password are checked against the database through the *CheckUser* class. In the affirmative case, the main view of the application is depicted in Figure 10.

Figure 10. Login successful.

All views and styles have been programmed in XML. To achieve this, the official Android API has been used. The code corresponding to Start session is denoted as "LogicActivity". The flowchart of this activity is presented in Figure 11.
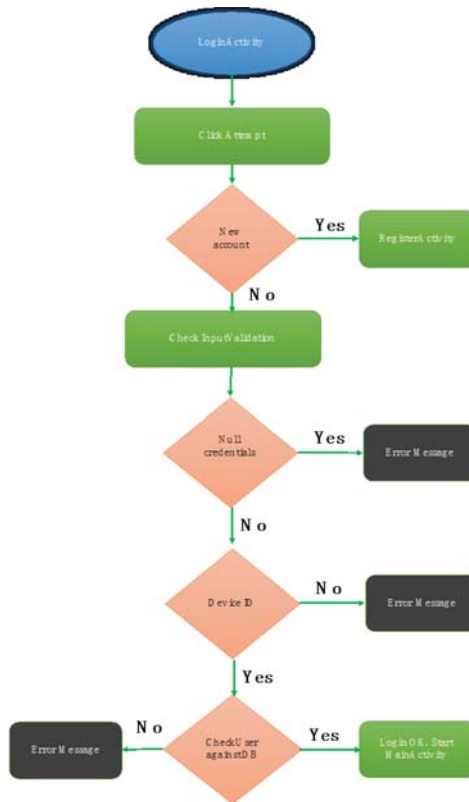
Figure 11. Start procedure flow-chart.

User Register

Each time the user clicks on the screen "No account yet? Create one", located in "LoginActivity", the activity called "RegisterActivity" will start. This activity shows a registration form, as shown in Figure 12.

Figure 12. User registration.

As we can see, this form asks for the user name, e-mail address, password and password confirmation of the user. As in "LoginActivity", the login form is accessed again by clicking on the sentence "Already a member? Login", shown in the following Figure 13.

Figure 13. App login screen.

"RegisterActivity" contains the same features as "LoginActivity". The *InputValidation* class executes a function in charge of ensuring that if the user introduces empty fields in the form or if the *password* field does not match the *confirm password* field, an error message will alert the user. The class *CheckUserRegis*, extended to the class *AsyncTask*, checks if the application has a user name which already exists, thus comparing the one introduced in the application with the user's database. Finally, if a user name doesn't match a previous one introduced in the database, it means that this name is available and the new user's data is stored in the database. The class responsible for this function is called *AddUser*. The following flowchart summarizes how this activity works (Figure 14).
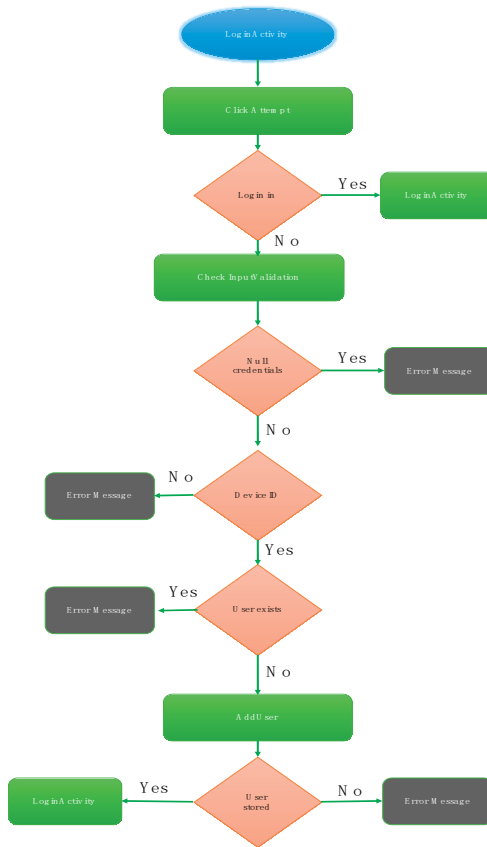
Figure 14. User register flow-chart.

Main Menu

The appearance of the main menu corresponds to the Technical University of Cartagena (UPCT) logo as well as a dropdown menu which lets us select the action we want the device to furnish is shown in Figure 15.
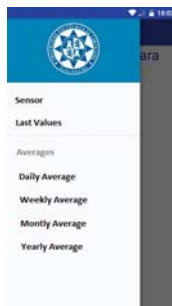
Figure 15. App statistical options.

Each menu item is associated with a different activity. A particular case is the "Sensor" action, which shows GM detector (sensor) results in real time. The *MainActivity* allows a user who requires any stadistical action, for example, to download the monthly average radiation, to finish this action, and again access real time dose values in their smartphone. The goal is to provide an extra function to the main menu, which goes beyond an intermediary between activities. Finally, the function corresponding to the main activity is "MainActivity", and is reflected in the following flowchart (Figure 16).
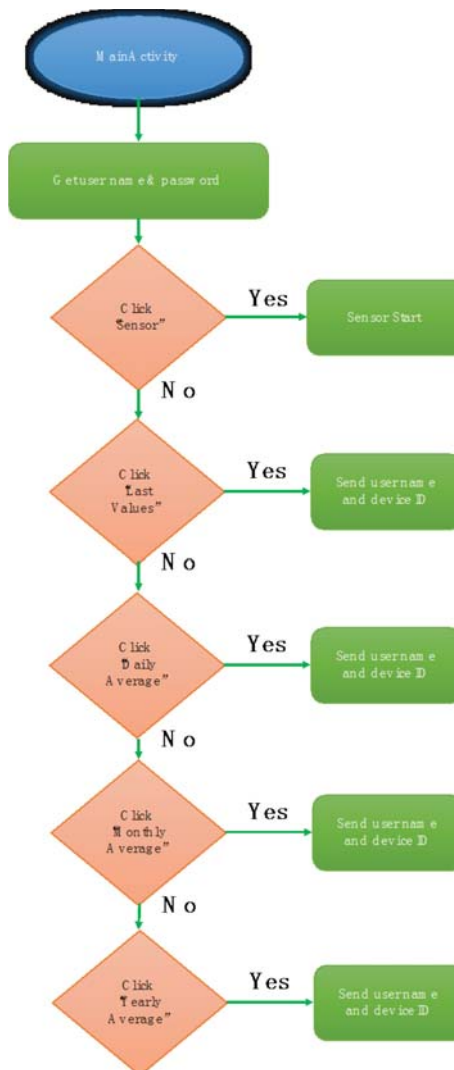
Figure 16. Main menu method flow-chart.

As shown in Figure 16, depending on the item of the dropdown menu which the user clicks, one action or another starts, and then a different activity begins. As previously illustrated in the

"Intent" description, before starting a new activity, the user name and device number are obtained from "LoginActivity".

Sensor (GM Detector)

This activity is implemented in the "MainActivity", sharing appearance and functionality with the dropdown main menu. The appearance depends its state. If the "Sensor" icon has not been clicked by the user to start collecting and displaying values in real time, the following Figure 17 depicts the graphic user interface which appears. However, if the user clicks the "Sensor" icon, the window is Figure 18.

Figure 17. Main screen.

Figure 18. Dose values screen shot.

Values belong to counts per minute (CPM) and their corresponding radiation value, together with the user who has downloaded them, as shown in this view (Figure 18). These values are updated each second, collected by the GM detector. If the device turns off, the last value captured by the sensor is displayed. Note that the appearance changes (each second) once the first SQL query arrives to the database. The *Query* class directs this issue.

Showing Last Values

This activity addresses the last counts per minute (and thus, radiation values) and the timestamp registered by the user device as well. The activity in charge of this function is Main2Activity. However, the class which leads to the connection to the database is Last class. This furnishes the required operations to obtain last dose values by means of SQL queries as shown in Figure 19.

Figure 19. Last values screen shot.

Statistics

The appearance and functionality for the daily, weekly, monthly and annual average is the same. Only the SQL queries vary, taking into consideration the time period to compute. The statistic view shows the user name, the device number and the average value that is required to download from the remote database. The function is similar to the previous Last Values view. However, they are very different in the SQL query message. In detail, the activity addressed to the daily average is *Main4Activity* and the attachment to the database is carried out by the *diarilyaverage* class. Regarding the weekly average, it is accomplished by *Main5Activity* and the *weeklyaverage* class. The monthly average is run by *Main3Activity* and the *Middle* class. Finally, for the annual average, *Main6Activity*, together with the *MediaAnnual* class exhibit the annual statistics. These codes, together with the complete software here proposed, are available in [3].

4. Verification of the Device

To guarantee the reliability and traceability of radiation measurements, it is necessary to verify the appropriate operation of the device in hospital environments. The verification procedure was carried out in the Laboratory of Ionizing Radiation Metrology at the National Dosimetry Center (CND). The CND is the largest Personal Dosimetry Service in Spain, and one of the largest in Europe in terms of monthly reading volume. It has been authorized for this purpose by the Nuclear Safety Council. Likewise, the National Dosimetry Center has been authorized to act as a Radiological Protection Unit in radio diagnostic installations.

The CND has an Ionizing Radiation Metrology Laboratory, available to the public healthcare sector and private initiatives. The calibration laboratory was certified by the National Accreditation Agency (ENAC) for calibrations in ionizing radiation and radioactivity. The certification is also recognized by the European Organization for Accreditation of Laboratories (EA). The Laboratory performs calibration of instruments for measuring ionizing radiation, as well as irradiation of personal dosimeters (DTL, film dosimeters, etc.), at protection and diagnostic levels.

4.1. Verification Procedure

The verification performed on the system corresponds to the PT-03 procedure, which is characterized in the certification of instruments for measuring ionizing radiation under X-rays.

To furnish the procedure, the device was set-up in the Test Room without any external casing, taking the center of the GM tube as a reference point. In this room, the device was put in a vertical position, orienting the GM tube towards the X-ray source, so that the tube was parallel to the anode-cathode direction.

In order to obtain the best measurement quality, the reference point was matched to the center of the radiation field. This was accurately marked by a laser, so the device was displaced to that point, as can be seen in the sequence of images in Figure 20.

Figure 20. Device verification sequence.

The distance between the X-ray source and the reference point is 5 m. The appropriate radiation field diameter is 19 cm, capturing the radiation dose in the GM tube. The radiation qualities employed in the verification process are defined as N-80 and N-300 of the Narrow Spectrum ISO Series. The characteristics of these radiation qualities are broken down in Table 5, below. Note that *Kerma* is the sum of the kinetic energies of all charged particles set in motion by the radiation; its magnitude in the international system is the Gray, equivalent to a 1 Joule per Kilogram. On the other hand, the so-called half-value layer (HVL) is the layer thickness of copper reducing the intensity of radiation due to absorption and scattering phenomena by half (first half-value layer) and by four times (second half-value layer), respectively. As it will be observed in the next subsection, one of the test was carried out in the PET surroundings. In this device, the energy peak is 511 keV around, value that is out of the range of the N-300 quality. However, in usual medical tests where the radionuclide is inside the human body, the continuous spectrum generated is lower than the maximum peak. This fact is even greater in our case since we measure doses in the PET surroundings. Therefore, the N-300 quality is, a priori, appropriate to validate our device under these conditions. This also is in accordance with the CND standards. On the other hand, N-80 is the spectrum recommended for X-rays which usually varies from 20 to 125 keV; N-80 is selected as a central energy value.

Table 5. Main features for N-80 and N-300 Qualities.

Quality Code	Voltage (Kv)	E_{avg} (keV)	1st HVL (mm Cu)	2nd HVL (mm Cu)	Kerma (Air) (µGy/min) Minimum	Kerma (Air) (µGy/min) Maximum
N-80	80	65	0.05776	0.619	1.9	1000
N-300	300	250	6.28	6.29	2.9	450

Thus, the verification process performed is summarized in Table 6.

Table 6. Verification procedure for N-80 and N-300 Qualities.

Quality Code	System Scale	Kerma Air Rate (μGy/h)	Radiation Time
N-80	100	110–130	360
N-80	100	180–210	360
N-300	100	350–400	180

Table 6 exhibits the three tests performed on our device. They were defined by the CND according to the environment where our system will be deployed. Additionally, a thorough study should be carried out to enhance the GM measurement accuracy. For instance, in a future, a correction factor which reflects the difference between the response of the commercial GM in relation to the dose rate could be computed, to later program it in the Raspberry.

As regards the tests, firstly, an N-80 was implemented with a radiation range from 110 to 130 μGy/h, at an interval of 360 s. The second test was carried out with similar characteristics to the previous one, but increasing the radiation level from 180 to 210 μGy/h. Finally, an N-300 test was furnished with a radiation range from 350 to 400 μGy/h for an interval of 180 s.

In each test, the radiation values captured/sensed by the device were written down in Table 7. Later, these values were compared to the radiation values measured by the certified laboratory instrumentation at the same time interval. Thus, the relationship between the values captured by the device and those collected by the laboratory instrumentation results in an N_H verification factor. This factor must be multiplied by the radiation values measured by the device to obtain the true radiation values.

4.2. Verification Results

The verification results are shown Table 7 below.

Table 7. Verification results for our proposal.

Quality Code	System Scale	Kerma Air Rate (μGy/h)	Radiation Time	N_H
N-80	100	110–130	360	3.37 ± 0.20
N-80	100	180–210	360	5.28 ± 0.58
N-300	100	350–400	180	10.2 ± 1.0

Concerning the first N-80 test, the verification factor N_H is 3.37 with a deviation of ± 0.20. The second N-80 test results in a N_H factor of 5.28 with a deviation of ± 0.58. Finally, for the last N-300 test, the verification factor N_H is 10.2 with a deviation of ± 1. Observing these results, we notice that as the radiation level increases, the radiation factor also grows considerably, having to multiply the radiation value of our device by 10.2 in the case of N-300 quality.

However, a last correction factor (which has been denoted as N_F), motivated by the sampling period to which the GM tube is subject, must be considered. This factor adjusts our sampling times (1 s) to the one defined by the manufacturer (10-s blocks). In other words, by dividing the sampling period indicated by the manufacturer by 10, the sampling period under study is obtained, and then the CPMs are calculated. Therefore, the final verification equation is as follows:

$$\text{Real Radiation Measurement} = \text{Captured Radiation Level} *(N_H/N_F), \tag{2}$$

where NF = 10.

This means that for the N-300 radiation quality, the standard deviation obtained in the verification process at the CND is the only one that computes.

5. Performance Evaluation in the Hospital

The main contribution of this work is to offer a step forward in radiation measurement systems and in particular, those affecting healthcare personnel. As a final issue in the development of the proposal, it was tested in a hospital environment under real conditions.

Two devices were configured to be tested at St. Lucia University Hospital in Cartagena. A device was placed in the corridor adjacent to the Nuclear Medicine department (near the Positron Emission Tomography room, Figure 21), specifically, in the same place as the current area dosimeter. The device continuously measured the ambient radiation for 7.5 h. Its main purpose was to capture the radiation emitted by patients who walked by and stayed in this area.

The second device was located in the CT area, specifically, in the room where the healthcare staff control the procedures (Figure 22). As in the previous case, this device measured environmental radiation. However, the radiation received by the healthcare staff was collected, measured, stored and computed in our databases and smartphone application.

An additional dosimeter was placed into our device cases for comparison purposes after the data collection. Those dosimeters were two small GM dosimeters from the Radiation Protection Department at Sta. Lucia University Hospital. The goal was to contrast the accumulated radiation in µSv/h measured by our devices with that obtained by Sta. Lucia dosimeters.

Figure 21. Device located in Positron Emission Tomography (PET) corridor.

Figure 22. Device placed next to the CT room.

Result Discussion

The radiation values obtained by the devices can be observed in Figures 23 and 24. Figure 23 shows doses captured in the corridor, while Figure 24 depicts results in the room next to CT. We will discuss the entire results, dividing them into two subsections depending on the placement of the devices.
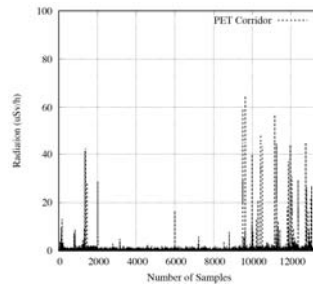
Figure 23. PET corridor results.
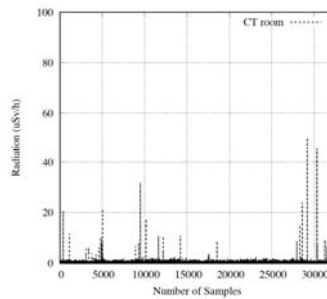
Figure 24. Results in room next to CT room.

Device Located in the Corridor

Before starting the tests, it was verified that the environmental radiation value captured by the device was similar to the one measured by the area dosimeter belonging to St. Lucia University Hospital. In detail, the latter obtained a dose range from 0.005 μSv/h to 0.5 μSv/h in a period of ten minutes. During the same interval, our device collected a radiation range from 0.0044 μSv/h to 0.4872 μSv/h.

As shown in Figure 23, we can observe that the radiation values increase to a maximum value (peak) and then they decrease gradually. As expected, a patient in treatment increments the activity of the device. Therefore, when values grow, it indicates that the patient is approaching the device. The maximum radiation value points out that the patient is closest to the device. Finally, as the patient moves away, the radiation values decrease.

Results were compared to the portable GM device at St. Lucia University Hospital. Our platform obtained an accumulative radiation of 2.46 μSv/h, compared to the 2.12 μSv/h of the hospital device during the time period evaluated. All in all, the viability and operation of our system as an area dosimeter is demonstrated, and validates the design and development proposed. Furthermore, these results can be displayed in real time, thanks to the app, or later analyzed, by dispatching data from the database (e.g., PC).

Device Located in the Room Next to CT System

Once installed and configured, the device was checked to observe if it captured radiation values when patients were scanned in real time. This task was carried out by two healthcare staff members, downloading and using the smartphone application, simultaneously. The result was that both received an accumulative dose of 0.45 μSv/h in their working day (eight hours).

Figure 24 shows the radiation values received by personnel during an 18 h period. As can be observed, the number of samples indicating a radiation level above the background value is lower

than in the previous case. Therefore, the time dedicated to a complete CT scan is shorter than the radioactive pharmacological treatment received by a patient.

On the other hand, as in the corridor device, data obtained was compared to the portable GM device at St. Lucia University Hospital. Our platform obtained an accumulative radiation of 1.86 μSv/h in comparison with the value of 2.09 μSv/h measured by the hospital device during the time period evaluated, validating our system as a personnel dosimeter for specific areas. As in the previous case, these results can be monitored in real time or later, by transferring data from the database.

In both scenarios, it is noteworthy to mention that medical devices barely interfere in the communications of our devices due to the distances among them. This means that packet losses motivated by interferences are scarce, ensuring that data dispatched each second by the device are in turn received by the mobile application. Additionally, medium access mechanisms such as CSMA-CA (Carrier sense multiple access with collision avoidance) in WiFi support packet collision; the device transceiver senses the physical medium and, only when it is free of interferences, sends the packet including the collected dose.

6. Conclusions

In this paper, we propose the design and implementation of a low-cost, open-source, portable, COTS system to collect ionizing radiation (gamma particles) in real time, deriving (i) the average radiation in different temporal periods and (ii) the accumulated radiation received by medical personnel. The solution is offered by a device that contains a commercial Geiger-Muller detector and all the electronic boards required (hardware and software) to dispatch the dose values to a remote database. Healthcare staff will obtain information of the radiation received in their bodies through a mobile application, developed in Android, and connected to the database.

To validate the complete system, a twofold procedure was adopted. Firstly, it was verified in the CND Calibration Laboratory, an authorized and accredited center for this type of calibrations. Secondly, a test-bed was set up at Sta. Lucia University Hospital, deploying two devices in two potential radiation areas and comparing measurements with other validated dosimeters placed next to our devices. Results show that this solution is able to accurately measure the radiation doses received by two heathcare workers in real time, and generate statistics for later studying and analyzing. Finally, the system proposed appropriately operates as a general purpose area dosimeter.

Author Contributions: A.-J.G.-S. conceived the idea. E.A.G.A. and P.A.M.R. looked into the related literature. A.-J.G.-S. proposed the scientific methodology. A.-J.G.-S. and P.A.M.R. contributed to the HW/SW definition. E.A.G.A., A.S.B. and D.R.-A. improved technical issues. All authors participated in the elaboration of the manuscript and its revision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Richardson, D.B.; Cardis, E. Risk of cancer from occupational exposure to ionising radiation: retrospective cohort study of workers in France, the United Kingdom, and the United States (INWORKS). *BMJ* **2015**, *351*. [CrossRef] [PubMed]
2. Directive 2013/59/Euratom. Available online: https://ec.europa.eu/energy/sites/ener/files/documents/CELEX-32013L0059-EN-TXT.pdf (accessed on 17 December 2017).
3. Solution Code. Available online: https://www.dropbox.com/sh/38g9yr34loyeb4h/AADk0q_KMaN7FYKICt0sCkgga?dl=0 (accessed on 17 December 2017).
4. Gilmore, G.R. *Practical Gamma-Ray Spectrometry*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2008. [CrossRef]
5. TLD Dosimeter for RE-2000 Readers. Available online: http://nusim.com/wp-content/uploads/2012/07/20996042_TLDdosimeter_EN_A.pdf (accessed on 17 December 2017).

6. Noriega Sánchez, K.; Ruiz Manzano, P. Dosimetría de Cristalino de GESTISA y Su Aplicación a Personal. 4° Congreso Conjunto Sociedad Española de Física Médica (SEFM) y la Sociedad Española de Protección Radiológica (SEPR), Valencia, 2015. Available online: http://www.crystallineproject.eu/wp-content/uploads/2015/06/SISTEMA-DOSIMETRIA-GESTISA-Y-APLICACION-AMBITO-SANITARIO-pdf.pdf (accessed on 17 December 2017). (In Spain)

7. European Nuclear Society. Film Dosimeter. Available online: https://www.euronuclear.org/info/encyclopedia/f/filmdosimeter.htm (accessed on 17 December 2017).

8. LLNL Environmental RPL Dosimeter. Available online: https://www.orau.org/PTP/collection/dosimeters/ron.htm? (accessed on 17 December 2017).

9. Dosimetría BeOSL. Available online: http://www.radmedical.com.mx/?page_id=2 (accessed on 17 December 2017).

10. Electronic Dosimeter DMC2000. Available online: https://www.dosimetry.com/images/pdf/LIT4319_EDs_combined.pdf (accessed on 17 December 2017).

11. Pen Dosimeters. Available online: http://seintl.com/radiationalert/pen_dosimeters.html? (accessed on 17 December 2017).

12. Glenn, F.K. *Radiation Detection and Measurement*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2000.

13. Ionization Chamber. Available online: https://www.orau.org/PTP/collection/ionchamber/introionizationchamberr.htm (accessed on 17 December 2017).

14. Proportional Counter. Available online: https://www.orau.org/PTP/collection/proportional%20counters/introprops.htm (accessed on 17 December 2017).

15. Geiger Muller Counter. Available online: https://www.cpp.edu/~pbsiegel/bio431/texnotes/chapter4.pdf? (accessed on 17 December 2017).

16. Scintillation Detectors. Available online: http://www-physics.lbl.gov/~spieler/physics_198_notes/PDF/III-Scint.pdf? (accessed on 17 December 2017).

17. Helmuth Spieler. Semiconductor Detector. 1998. Available online: http://www-physics.lbl.gov/~spieler/SLAC_Lectures/PDF/Sem-Det-I.pdf (accessed on 17 December 2017).

18. Geiger Counter-Radiation Sensor Board for Arduino and Raspberry Pi. Available online: https://www.cooking-hacks.com/documentation/tutorials/geiger-counter-radiation-sensor-board-arduino-raspberry-pi-tutorial (accessed on 17 December 2017).

19. Raspberry Pi. Available online: https://www.raspberrypi.org/ (accessed on 17 December 2017).

20. Instadose 1. Available online: https://www.mirion.com/products/instadose-dosimetry-services/ (accessed on 17 December 2017).

21. Dosicard. Available online: http://www.canberra.com/products/hp_radioprotection/dosimetry.asp (accessed on 17 December 2017).

22. Poudel, S.; Weir, L.; Dowling, D.; Medich, D.C. Changes in Occupational Radiation Exposures after Incorporation of a Real-time Dosimetry System in the Interventional Radiology Suite. *Health Phys.* **2016**, *111*, S166–S171. [CrossRef] [PubMed]

23. Servoli, L.; Bissi, L.; Fabiani, S.; Magalotti, D.; Placidi, P.; Scorzoni, A.; Calandra, A.; Cicioni, R.; Chiocchini, S.; Dipilato, A.C.; et al. Personnel real time dosimetry in interventional radiology. *Phys. Med.* **2016**, *32*, 1724–1730. [CrossRef] [PubMed]

24. Nuclear Open-Source Radiation Monitoring System. Available online: https://www.blackhat.com/docs/us-17/wednesday/us-17-Santamarta-Go-Nuclear-Breaking%20Radition-Monitoring-Devices-wp.pdf (accessed on 17 December 2017).

25. Terasaki, K.; Toshioh, F.; Murazaki, H.; Kuramoto, T.; Umezu, Y.; Ishigaki, Y.; Matsumoto, Y. Evaluation of basic characteristics of a semiconductor detector for personal radiation dose monitoring. *Radiol. Phys. Technol.* **2017**, *10*, 189–194. [CrossRef] [PubMed]

26. Ishigaki, Y.; Matsumoto, Y.; Ichimiya, R.; Tanaka, K. Ultra-low-cost radiation monitoring system utilizing smartphone-connected sensors developed with internet community. In Proceedings of the IEEE Sensors Conference, Taipei, Taiwan, 28–31 October 2012; pp. 652–655.

27. Ishigaki, Y.; Matsumoto, Y.; Ichimiya, R.; Tanaka, K. Development of mobile radiation monitoring system utilizing smartphone and its field tests in Fukushima. *IEEE Sens. J.* **2013**, *13*, 3520–3526. [CrossRef]

28. Raspberry Pi to Arduino Shields Connection Bridge. Available online: https://www.cooking-hacks.com/raspberry-pi-to-arduino-shield-connection-bridge (accessed on 17 December 2017).

29. Huawei E173. Available online: https://www.amazon.es/Huawei-E173-M%C3%B3dem-7-2Mbps-blanco/dp/B008BQEFYM (accessed on 17 December 2017).
30. Apache. Available online: https://www.apache.org/ (accessed on 17 December 2017).
31. PHPMyAdmin. Available online: https://www.phpmyadmin.net/ (accessed on 17 December 2017).
32. Async Task—Android. Available online: https://developer.android.com/reference/android/os/AsyncTask.html (accessed on 17 December 2017).
33. Intent Class—Android. Available online: https://developer.android.com/reference/android/content/Intent.html (accessed on 17 December 2017).

Article

Measurement of Vibrations in Two Tower-Typed Assistant Personal Robot Implementations with and without a Passive Suspension System

Javier Moreno [1], Eduard Clotet [1], Marcel Tresanchez [1], Dani Martínez [1], Jordi Casanovas [2] and Jordi Palacín [1,*]

[1] Department of Computer Science and Industrial Engineering, University of Lleida, 25001 Lleida, Spain; jmoreno@diei.udl.cat (J.M.); eclotet@diei.udl.cat (E.C.); mtresanchez@diei.udl.cat (M.T.); dmartinez@diei.udl.cat (D.M.)

[2] Department of Chemistry, University of Lleida, 25001 Lleida, Spain; jcasanovas@quimica.udl.cat

* Correspondence: palacin@diei.udl.cat; Tel.: +34-973-702-724

Abstract: This paper presents the vibration pattern measurement of two tower-typed holonomic mobile robot prototypes: one based on a rigid mechanical structure, and the other including a passive suspension system. Specific to the tower-typed mobile robots is that the vibrations that originate in the lower part of the structure are transmitted and amplified to the higher areas of the tower, causing an unpleasant visual effect and mechanical stress. This paper assesses the use of a suspension system aimed at minimizing the generation and propagation of vibrations in the upper part of the tower-typed holonomic robots. The two robots analyzed were equipped with onboard accelerometers to register the acceleration over the X, Y, and Z axes in different locations and at different velocities. In all the experiments, the amplitude of the vibrations showed a typical Gaussian pattern which has been modeled with the value of the standard deviation. The results have shown that the measured vibrations in the head of the mobile robots, including a passive suspension system, were reduced by a factor of 16.

1. Introduction

The use of mobile robots is gaining popularity as an alternative way to offer reliable and more efficient solutions to problems faced not only in industries [1–3] but also in domestic environments [4]. Providing appropriate assistance and care to elderly people has become one of the main concerns in developed countries. This problem has already been highlighted by organizations such as the United Nations [5] and the World Health Organization [6], whose studies on the evolution of the world population postulate that the percentage of elderly people (aged 60 or more) will rise from 12% to 21% during the next 35 years, as a result of the clear increase of human life expectancy. In this context, the use of robots in domestic environments is a way of offering proper assistance to people of advanced years due to the evolution of robotics and electronic health systems.

The motion system of assistive robots must be able to maneuver in unstructured and limited spaces without interfering with the inhabitant's lifestyle. This is a disadvantage for mobile robots equipped with conventional two-independent driving wheels, as their mobility is restricted to one directional movement due to the motion system only providing two degrees of freedom (DOF). As a result, mobile robots equipped with classical steering mobility systems are unable to perform

lateral displacements. To overcome these limitations, it is essential to develop omnidirectional mobile robots (OMR), where the term omnidirectional describes the ability of a motion system to change the direction of movement without having to perform any intermediate rotation. As the Assistant Personal Robot (APR) [7] is equipped with three omnidirectional wheels shifted 120° (kiwi drive), it offers a 3-DOF motion system over a two dimensional plane [8]. This allows the robot to move in any angle or direction [9], as well as perform arbitrary movements in arbitrary directions without changing its orientation.

The primary components involved in creating an OMR are the omnidirectional wheels. There are several types of omnidirectional wheels which are based on the same operating principle, providing traction in the direction normal to the motor axis while using the inner passive rollers placed along the periphery of the main wheel which can, as well, slide in the direction of the motor axis.

Figure 1 shows different types of omnidirectional wheels and their traces. Figure 1a shows a wheel (universal wheel) designed with multiple passive rollers (or inner passive wheels) where the axes are at a tangent with the main wheel circumference. This construction cannot avoid the existence of discontinuities in the trace; therefore, it provides irregular contact with the surface due to the presence of gaps between the successive rollers producing vibrations in the robot. Some solutions are proposed to minimize those effects in these types of wheels, such as reducing the size of the gap between the passive rollers [10]. Figure 1b shows the Mecanum [11,12], invented in 1973 by Bengt Ilon, an engineer working for the Swedish company Macanum AB. This wheel is based on the use of overlapping rollers providing continuous contact between the wheel and the ground. The design and adjustment of this type of wheel's parameters allow for a remarkable decrease of the vibrations [13]. These wheels are often positioned in pairs over the same axle but in opposite orientations to a four-wheel structure. The drawback of these wheels is the generation of horizontal vibrations due to the parasite torque generated as a result of the contact points moving along a parallel line to the wheel shaft. Figure 1c shows the double wheel concept based on the use of two overlapping parallel wheels. In this case, the contact between the assembled wheel and the ground is continuous; however, this design still generates a significant horizontal vibration originated by the gaps between the rotating inner wheels [14]. Finally, in the design shown in Figure 1d the contact points are aligned in order to reduce the horizontal vibration while using alternating passive rollers of different sizes and shapes in order to minimize the gap between them, thus causing a slight vertical vibration. As with Macanum wheels, there are jobs where vibrations are analyzed according to the design parameters [15].
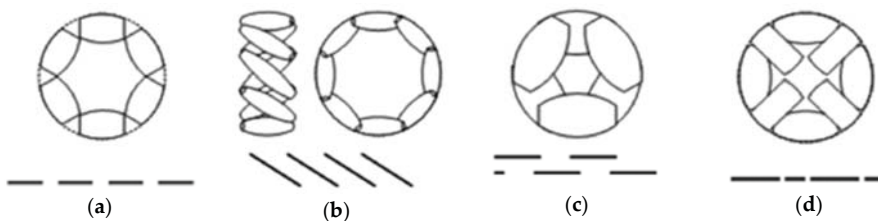
Figure 1. Types of omnidirectional wheels and their traces: (**a**) multiple passive rollers (or inner passive wheels) in which the axes are positioned tangent to the main wheel circumference; (**b**) with the rollers arranged in an overlapping way where the contact between the wheels and the ground is continuous; (**c**) based on two overlapping parallel wheels; (**d**) based on using alternating passive rollers with different size and shape.

The common point of the omnidirectional wheels (Figure 1) are the vibrations generated due to the inner trace wheel discontinuities. In the case of the tower-typed mobile robots, these vibrations can cause visible oscillations at the top of the mobile robot. For example, the proposal of [16] analyzes this problem in the case of two-wheeled mobile robots, testing a vibration minimization technique based on using different acceleration and deceleration velocity profiles along with the effects of using soft and

hard wheels instead of using a dedicated suspension system. The occurrence of vibrations is frequently used for reconnaissance of the terrain [17–20], or even power generation [21], but is often a problem for the safety of the components and for their correct operation and comfort [22–26]. The vibrations appearing at the head of the mobile robot cause mechanical stress, cracks [27], and also an undesired visual effect during the displacements.

This paper proposes the measurement of the vibrations originating in two tower-typed APR mobile robot prototypes: APR-1 and APR-2. The second prototype has been improved by including a suspension system mounted at the holonomic base, which is equipped with three omnidirectional wheels based on the alternation of passive rollers (previously exemplified in Figure 1d). The vibration patterns of the two mobile robots have been obtained by measuring the accelerations at one wheel, chassis, and at the top of the structure.

2. Materials and Methods

The materials used in this paper are two prototypes of the mobile robot APR. The mechanical structure of the first APR-1 prototype was described in [7] and the second improved APR-2 prototype was described in [28]. This paper is focused on the measurement of the vibration patterns found in both prototypes by using multiple linear three-axis accelerometer devices in order to gather the vibration patterns from different locations on the mobile robots and evaluate the improvements achieved.

2.1. Mobile Robot Prototypes

The experimental part of this paper has been carried out using two APR prototypes (Figure 2); both devices are tower-typed omnidirectional mobile robots with three omnidirectional wheels shifted 120°. The APR-1 [7] was the first APR prototype implemented (Figure 2a) and was built on a rigid structure with a motion system, with the wheels and a tower directly attached. The APR-2 was the second improved prototype implementation (Figure 2b) and it includes a passive suspension system implemented at the base of the robot.

| (a) | (b) |

Figure 2. Images of the two mobile robot prototypes analyzed; (**a**) APR-1, (**b**) APR-2.

Humans are the inspiration for the physical design of the APR, which resembles them; it is designed to navigate through environments planned to ease their mobility and it is also capable of moving its head and arms. The mobile robot design lacks, on purpose, sharp edges or projecting parts for safety reasons and to facilitate its use in domestic unstructured environments. The structure of the APR is also designed to maintain its reliability as a multi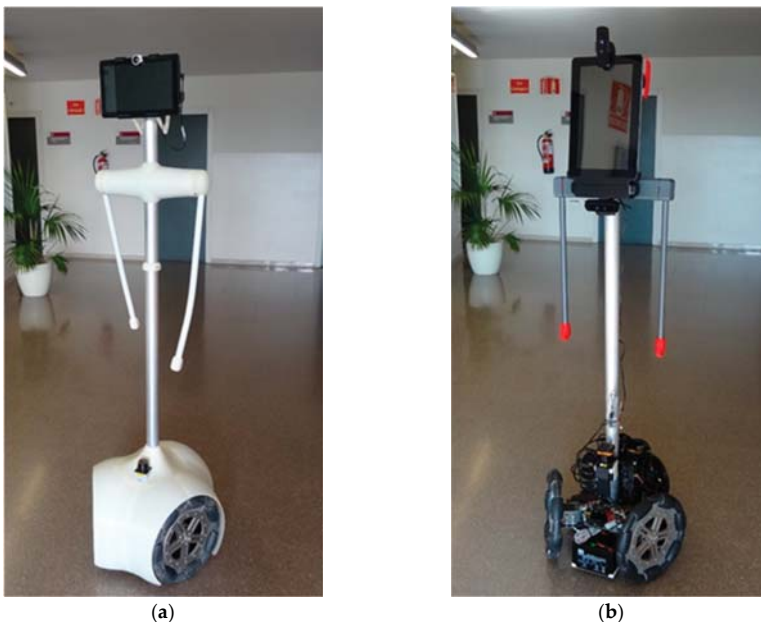purpose mobile platform; therefore, the main body of the robot is a single thin aluminum tube, on which various measuring devices can be easily clamped on.

The inner mechanical structure of the APR was made with a combination of stainless steel and aluminum parts to guarantee its durability and resistance. All of the heavy elements are located at the base of the robot, lowering its center of mass almost to ground level, and also reducing vibrations at the top of the robot. In both prototypes, the structure of the APR is divided in two parts, the holonomic motion system and the thin body or tower.

The mobile robot has a circular base section which supports the holonomic motion system composed of three omnidirectional wheels, the batteries, the Light Detection and Ranging (LIDAR) device, and the main electronic boards. The inner design of the base is covered with a low-cost 3D printed thermoplastic polymer, Acrylonitrile Butadiene Styrene (ABS), which provides flexible protection that will contribute to absorbing part of the energy of impacts in the case of collision. The circular design minimizes the probability of getting accidentally hooked on furniture objects such as mats, curtains, or clothing, and plays a crucial role when operating in tight indoor spaces, simplifying the mobile robot tele-control when passing through doorways, small corridors, or complicated environments. This three-wheeled robot has three independent geared DC motors attached to the omnidirectional wheels that provide 3 DOF to the motion system. Each wheel has the same distance, R, from its center to the center of the mobile robot. The APR has a triangular contact area with the ground due to the three wheels used in the motion system.

The mobile robot has a thin body and an upper part mainly designed for human interaction. The head of the mobile robot has a multi-touch panoramic screen and two arms each with one degree of freedom in order to move them forward and backward. The chest and shoulders of the APR are located approximately at 1.3 m height, which is slightly lower than the average for human shoulders. The shoulders of the APR enclose two DC geared motors which are connected to two soft arms with a 35 cm separation in between. The arms are 55 cm long for esthetical reasons and can be used as a support by elderly people when walking or can be used for basic gesture interaction. The arms are periodically moved during a forward displacement in order to replicate the natural movements performed by humans when walking. Table 1 summarizes the main physical features of the two mobile robots used in this paper.

Table 1. Physical specifications of the two APR prototypes.

	Height (cm)	Width (cm)	Weight (kg)	Suspension	Screen (Inches)
APR-1	164	48	34.5	No	7
APR-2	168	54	38.3	Yes	13

Figure 3 shows the structural diagram of the two mobile robot prototypes used in this paper. The chassis in the design of the APR-2 prototype (Figure 3b) improves the wheels and tower joints. The improved joints are based on a damped arm and a joint between the tower and the chassis, allowing the tower to balance and improve its stability as the pivot point is positioned above the tower's center of gravity. Figure 4 shows the CAD design of the suspension system. Finally, Figure 5 shows an overview of the implemented suspension system (Figure 5a), and a close-up image of a single arm of the base (Figure 5b). The red parts of the APR-2 prototype are made of a flexible 3D printing filament. This prototype also includes two sets of spring-based connections; the first one is located between the motor's and batteries' structures; and the second one is found between the mobile robot batteries' and the central tower's structures.
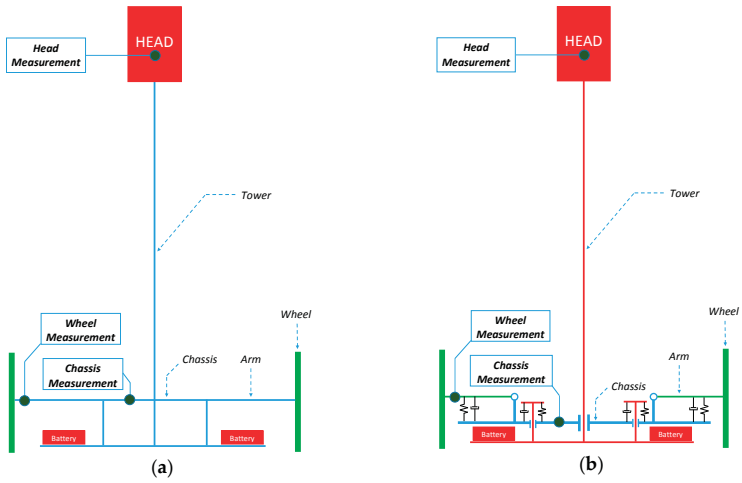
Figure 3. Structural diagram of the (**a**) APR-1 prototype and the (**b**) APR-2 prototype. The dark green dots depict the location of the measurement points: wheel, chassis, and head.
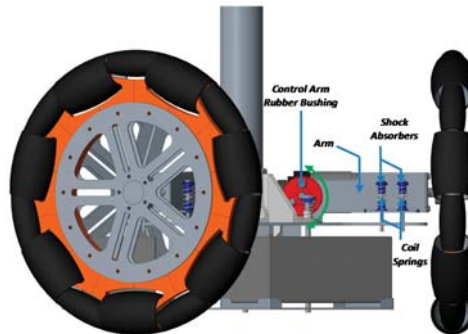
Figure 4. CAD design of the passive suspension system of the APR-2.
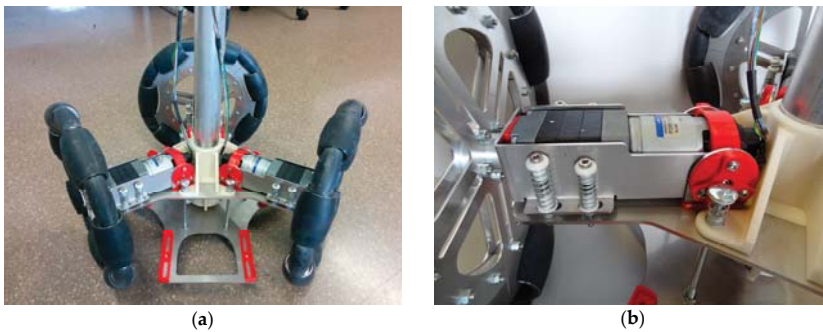
Figure 5. Image of the passive suspension system of the APR-2; (**a**) overview, (**b**) arm detail.

2.2. Profile Wheel Vibration

The most common holonomic motion system for robots is based on the use of three or four omnidirectional wheels. The operational function of the different types of omnidirectional wheels is based on providing traction in the natural direction towards the motor axis, and using numerous passive rollers (where the axes are at the tangent of the wheel circumference) that can slide in the direction of the motor axis. This implementation allows the wheel to spin as well as perform perpendicular displacements from the wheel's forward trajectory, allowing the robot to move in any direction.

The common motion system implemented in the APR is based on the use of three omnidirectional wheels shifted 120° and is composed of passive rollers placed along the periphery of the main wheel. Figure 6 shows the design of one omnidirectional wheel based on the use of alternating passive rollers with different sizes and shapes in order to minimize the gap between the rollers. Table 2 summarizes the main mechanical features of the wheels. In this case, the theoretical contact points between the wheel and the ground are aligned, but the gap between rollers and some manufacturing defects cause horizontal and vertical vibrations.
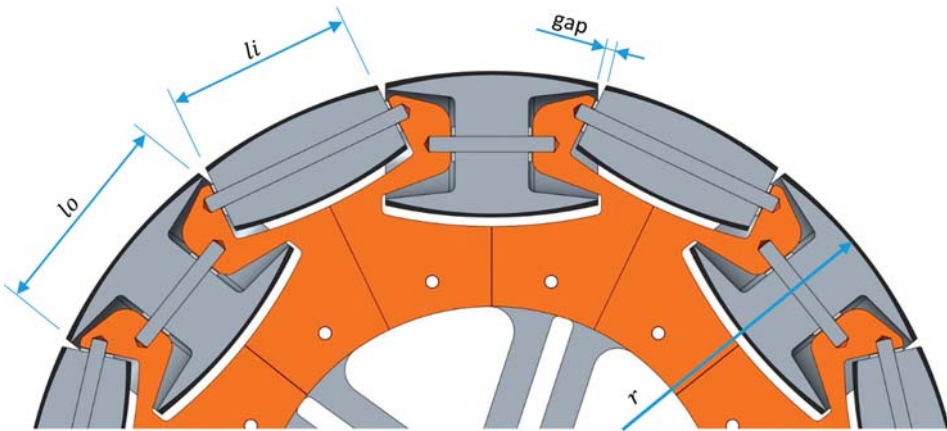
Figure 6. CAD section showing the alternate use of the two passive roller types and the gap between them.

Table 2. Main mechanical features of the omnidirectional wheels.

Diameter (cm)	Width (cm)	Outer Rollers Specifications	Inner Rollers Specifications	Gap (mm)	Maximum Vertical Displacement [1] (mm)
300	44.48	7 units $\varnothing_{max}$ 44.48 mm Length 67.5 mm	7 units $\varnothing_{max}$ 26.52 mm Length 60.5 mm	2.82	0.0066

[1] Maximum distance of the typical shape of the vertical displacement profile.

Equation (1) describes the analytical expression used to obtain the *gap* or distance between the rollers depending on the mechanical wheel features:

$$gap = 2 \cdot r \cdot \sin \left(\frac{2 \cdot \pi \cdot r - ni \left(2 \cdot \sin^{-1} \left(\frac{li}{2 \cdot r} \right) \cdot r \right) - no \left(2 \cdot \sin^{-1} \left(\frac{lo}{2 \cdot r} \right) \cdot r \right)}{2 \cdot r} \right), \tag{1}$$

where *ni* and *no* are the number of the inside and outside passive rollers, *li* and *lo* are the length of the inside and outside passive rollers, and *r* is the radius of the wheel.

Equations (2) and (3) describe the analytical expression used to obtain the vertical displacement profile, *d*, of the center of the wheel when the gap between the passive rollers are in contact with the ground.

$$d(\alpha) = \frac{\left| r \cdot \left(\sin\left(\frac{3 \cdot \pi}{2} - \sin^{-1}\left(\frac{gap}{2 \cdot r} \right) \right) + \frac{\cos\left(\frac{3 \cdot \pi}{2} - \sin^{-1}\left(\frac{gap}{2 \cdot r} \right) \right)}{\tan\left(\left(\frac{3 \cdot \pi}{2} - \sin^{-1}\left(\frac{gap}{2 \cdot r} \right) \right) + \alpha \right)} \right) \right|}{\sqrt{\left(\frac{-1}{\tan\left(\left(\frac{3 \cdot \pi}{2} - \sin^{-1}\left(\frac{gap}{2 \cdot r} \right) \right) + \alpha \right)} \right)^2 + (-1)^2}} \rightarrow \alpha = 0 : to : \sin^{-1}\left(\frac{gap}{2 \cdot r} \right), \qquad (2)$$

$$d_{max} = r - \sqrt{r^2 - \left(\frac{gap}{2} \right)^2} \qquad (3)$$

where α is the angle that forms the *gap*.

Figure 7 shows the vertical displacement profile's typical shape, *d*, that appears in the center of the omnidirectional wheel due to the gap between the rollers. This displacement profile has a sharp transition that causes strong vertical vibrations. The maximum theoretical amplitude or displacement is 0.006 mm but in the practice this value is also affected by the weight of the mobile robot that crushes asymmetrically the soft cover of the passive rollers.
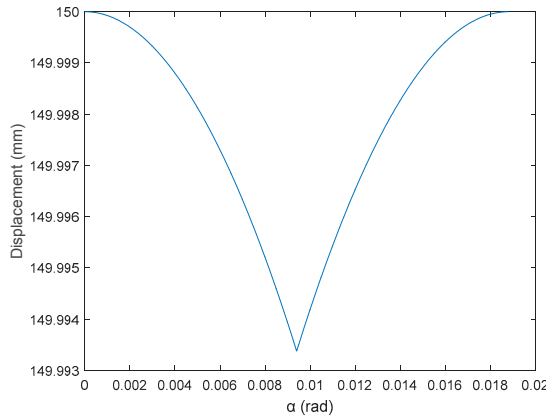
Figure 7. Profile of the vertical displacement of the wheel's center found across the gap.

2.3. Vibration Measurement

The vibrations measurement is carried out by a LIS3DSH accelerometer manufactured by STMicroelectronics (Rennes, Switzerland). The LIS3DSH is a low-power high-performance three-axis linear accelerometer of compact size ($3 \times 3 \times 1$ mm), the dynamic measurement range is selectable between ± 2 to ± 16 g, the sensitivity is between 0.06 to 0.73 mg/digit depending on the range, and it has a maximum sampling rate of 1.6 kHz.

The complete data acquisition system is based on the STM32F4-Discovery board from STMicroelectronics that includes on the same board the LIS3DSH accelerometer sensor and a STM32F407VGT6 32-bit high-performance microcontroller based on the ARM Cortex™-M4 processor developed by ARM (Cambridge, UK), and which is also manufactured by STMicroelectronics (Figure 8). This acquisition system has an USB 2.0 on the go (OTG) interface for connectivity, a reset pushbutton, and a user configurable pushbutton used for data or experiment triggering.

Figure 8. Data acquisition system based on the STM32F4-Discovery board.

The microcontroller uses an internal timer to read the raw accelerometer data using a Serial Port Interface (SPI). The maximum sampling rate is 1.6 kHz and the raw accelerometer data is stored in the microcontroller's RAM and then transferred as a text file to a USB flash-disk memory connected to the USB 2.0 OTG interface. The microcontroller is able to store up to 12,424 samples of the three-axis accelerometer, which means 7.765 s at the maximum sampling rate supported by the accelerometer.

The measurements have been carried out by configuring the accelerometers with a dynamic range of ± 4 g and a sensitivity of 0.12 mg/digit. This range has been obtained by a trial and error procedure in order to avoid any saturation in the raw data obtained from the accelerometers. The standard deviations of the sensor for the X, Y, and Z axes at zero speed were 0.0132 g, 0.0181 g, and 0.0144 g, respectively.

3. Measurements

The vibration measurement has been performed by attaching the accelerometers to one wheel, to the chassis, and to the top of the mobile robot (see Figures 3 and 9 for reference). The accelerometers' axes orientation is the same as the reference axes plotted in Figure 9, where the Z axis corresponds to the normal vector, the X axis to the longitudinal vector, and the Y axis corresponds to the side vector.

| (a) | (b) |
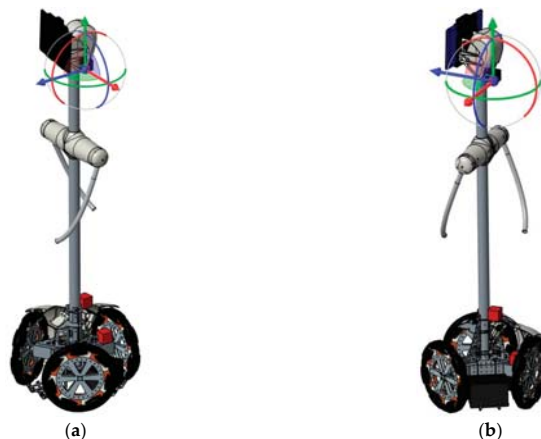
Figure 9. Mobile robot APR-1 with a red box representing the location of the accelerometers: front (**a**) and side (**b**) view. The longitudinal, normal, and side axes of the measurements are depicted with a blue, green, and red arrow, respectively.

Tables 3–5 show the raw data obtained by the accelerometers located on both robots along the longitudinal, side, and normal axes, respectively. The maximum speed of the two mobile robots is

100 cm/s. The operational speed range is in an array from 19 and 66 cm/s hence the data provided in the tables depicts these two extreme velocity cases. The gathered data from the APR-1 (without suspension) is shown in blue; the gathered data from the APR-2 (with passive suspension) is shown in orange.

Tables 3 and 4 show the vibrations' dynamic evolution measured in both the wheels and chassis of the two mobile robots. The dynamic evolution is similar even though the APR-2 prototype (with a passive suspension) reduces the overall vibrations' amplitude in the normal axis, hence a reduction in the transmission of the vibrations to the upper part of the mobile robot is expected. Tables 3 and 4 also show the presence of some acceleration peaks in the APR-2 prototype caused by small irregularities on the ground, although these peaks are not directly transmitted to the head of the mobile robots (see Table 5).

Table 3. Vibration representation measured in the wheel of the APR-1 and APR-2 prototypes.
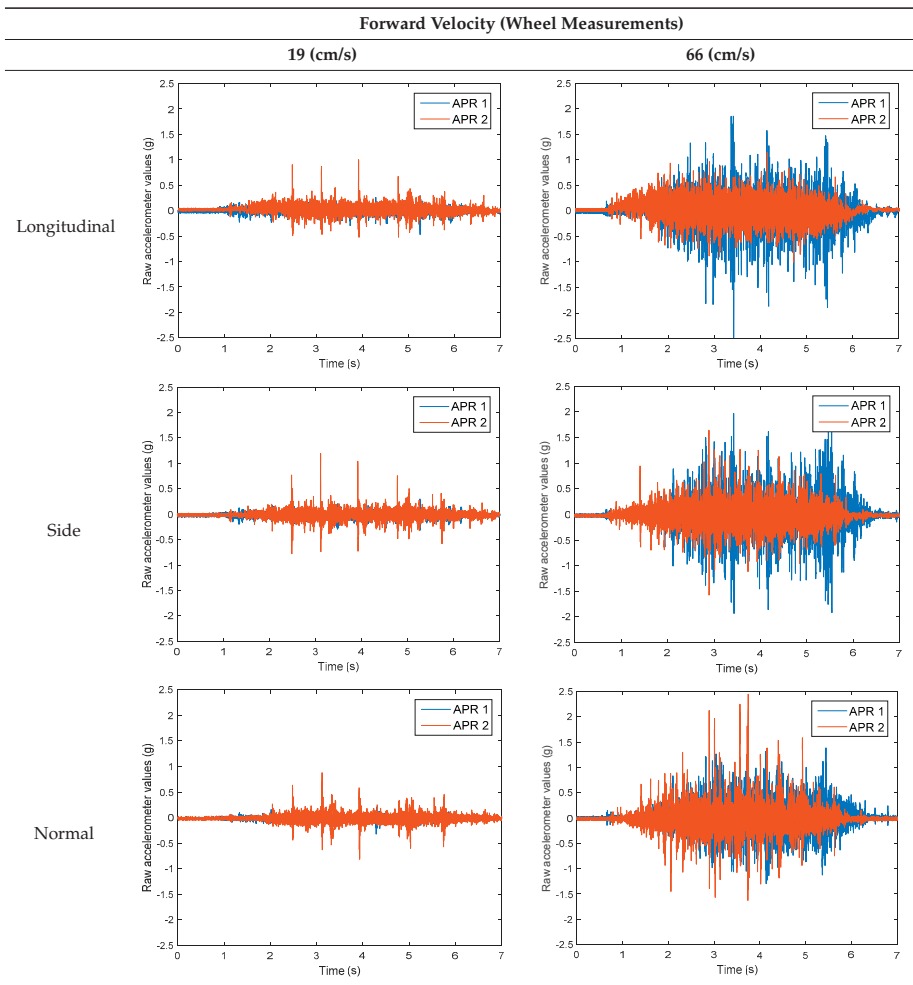
Table 4. Vibration representation measured in the chassis of the APR-1 and APR-2 prototypes.
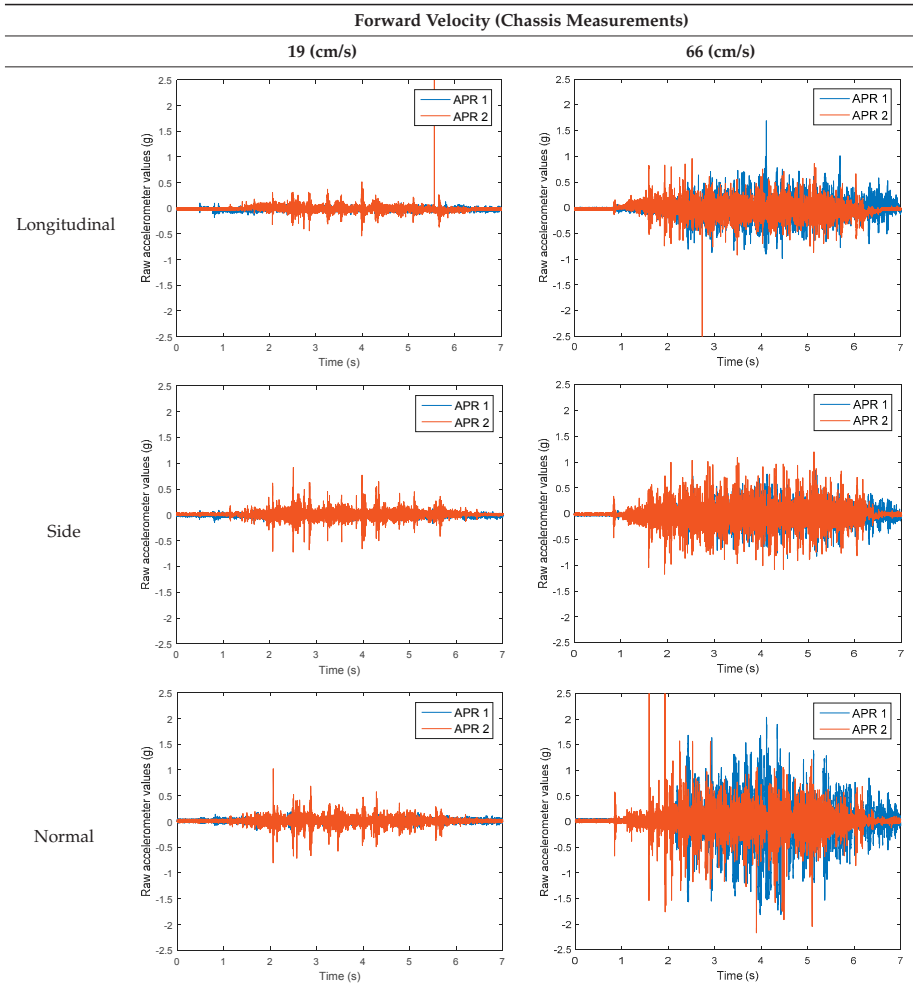
Table 5 shows the vibrations' dynamic evolution measured at the head of the two mobile robots at two velocities. In this case, the passive suspension system has reduced the overall amplitude and peaks of the vibrations measured in all axes. The maximum amplitude of the vibrations in the APR-1 prototype (without suspension) was raised from 0.45 g for a forward velocity of 19 cm/s to 3.97 g for a forward velocity of 66 cm/s (~8.8 times). The vibrations' maximum amplitude measured in the APR-2 prototype (with passive suspension) was raised from 0.17 to 0.40 g in the same velocity range (~2.3 times). Therefore, the vibrations' maximum amplitude in the head of the APR-2 prototype was lower than the minimum vibrations measured in the APR-1 prototype. Additionally, the measurements over the longitudinal axis also showed a smooth undamped effect caused by the passive suspension when accelerating and decelerating the APR-2 prototype.

Table 5. Vibration representation measured at the head of the APR-1 and APR-2 prototypes.
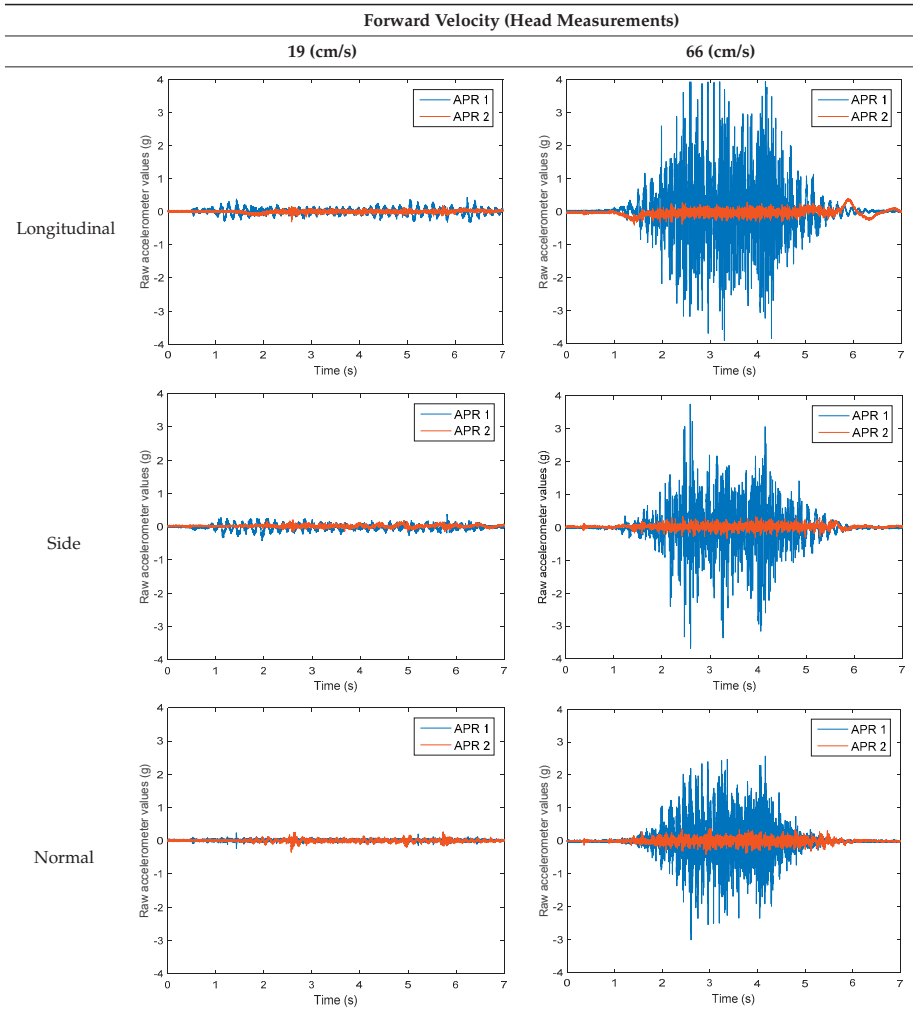
The raw data measured during the displacements has been manually divided in four different states: repose, initial acceleration, constant speed, and deceleration (Figure 10). The data gathered by the accelerometers while the robot is going forward at a constant speed (reference speed) has been selected to create a histogram of the accelerations' amplitude measured during the displacement. As an example, Figure 11 shows the histogram of the side acceleration measured at the head of APR-1 when moving forward at a reference velocity of 55 cm/s. Figure 11 shows a typical Gaussian distribution with a standard deviation of 0.80 g. The value of the standard deviation will be used in this paper to compare the overall vibrations measured in both mobile robots.
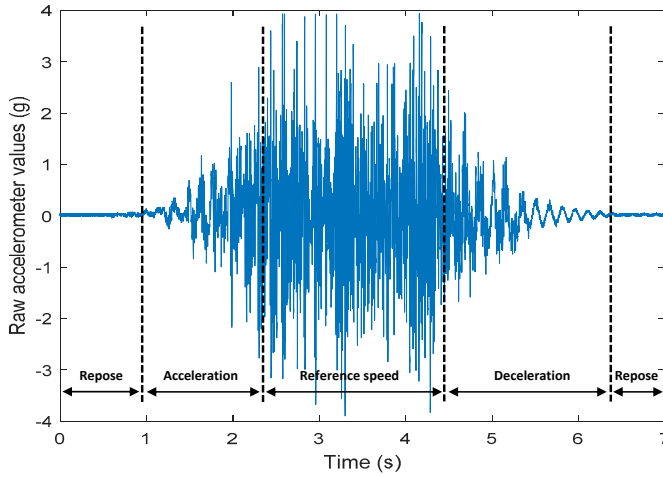
Figure 10. Raw accelerometer data obtained at the head of the APR-1 prototype when the forward speed was fixed at 66 cm/s.
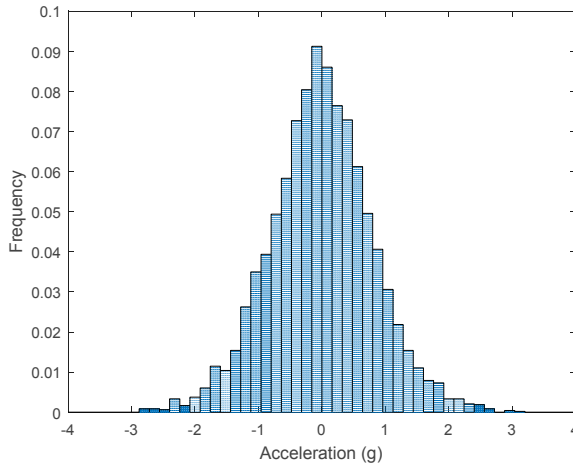
Figure 11. Histogram of the side accelerations measured at the head of the APR-1 prototype when moving at a constant velocity of 55 cm/s during a forward displacement.

For the sake of comparison, Tables 6–8 summarize the vibration pattern histograms obtained from each mobile robot when moving forward at two different constant velocities (23 and 66 cm/s). These histograms also represent a typical normal distribution. Note that, in some cases, the Gaussian distribution is not zero centered due to small misalignments between the accelerometer axes and gravity.

Table 6. Vibration pattern histograms measured in the wheel of the two APR prototypes.
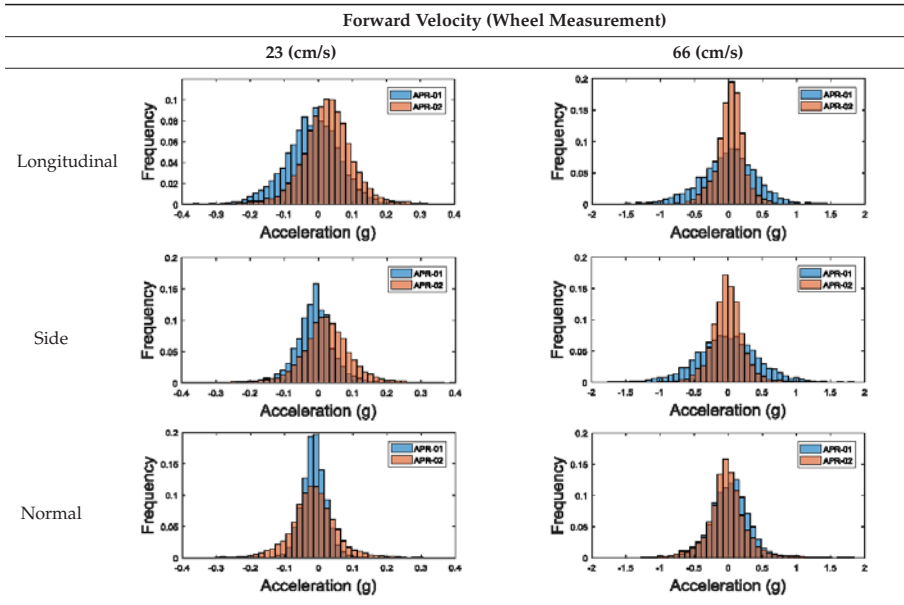
Table 7. Vibration pattern histograms measured in the chassis of the two APR prototypes.
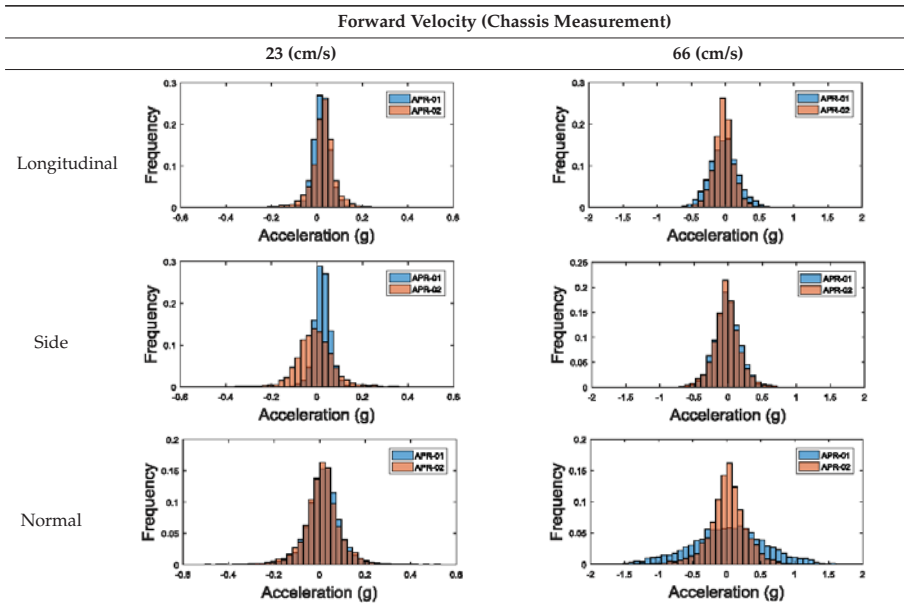
Table 8. Vibration pattern histogram measured in the head of the two APR prototypes.
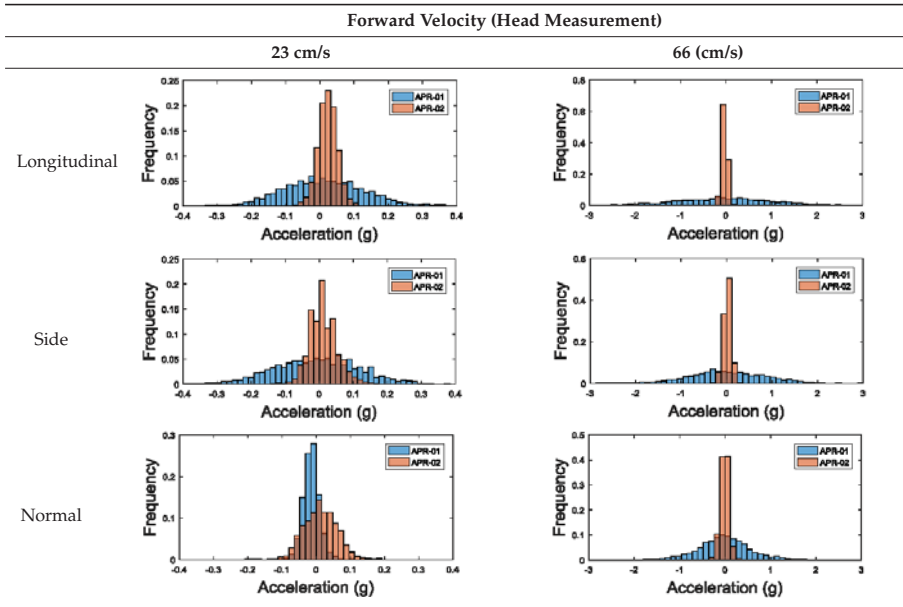
Table 9 compares the evolution of the standard vibration deviation measured at different velocities in the wheel, chassis, and head of the APR-1 and APR-2 mobile robot prototypes. The results show that the vibrations on the APR-1 (without suspension) have a clear tendency to increase proportionally with the velocity. Alternatively, the vibrations measured at the head of the APR-2 (with passive suspension) are almost constant with a standard deviation of 0.80 g at the maximum speed. This is because in the first prototype the wheels' vibrations are directly transmitted and amplified due to the tower-typed structure, reaching a maximum acceleration peak of almost 4 g at the head when moving forward at a constant speed of 55.2 cm/s. The measurements also show that the longitudinal and side measures at the chassis of the APR-1 are similar while there is a considerable variation between them and the normal axis; this is mainly caused by the edge of the rollers hitting the ground during their transition, shaking the structure along the normal axis.

The acceleration data gathered at the head of the APR-2 shows that the suspension system allows the central structure to balance itself in a controlled way when the robot is moving at a constant velocity. The measures obtained at the wheel of the APR-2 prototype show that the addition of flexible shock absorbers and spring based suspensions are an effective method to minimize and stabilize the wheels' vibrations which are almost constant and not transmitted to the tower. The rigid APR-1 prototype cannot absorb the vibrations originating in the wheels which are directly transmitted to the chassis and the tower of the mobile robot.

Finally, Figures 12 and 13 show the frequency spectra of the accelerations measured over the longitudinal axis at the head of the two mobile robot prototypes in the case of moving forward at a constant speed of 66 cm/s (Table 9—Longitudinal and Head data). Figure 12 shows that in the case of the APR-1 prototype (without suspension), the highest peak frequencies are in a range from 10 to 300 Hz, while Figure 13 shows that in the case of the APR-2 prototype (with passive suspension), these frequencies have overall smaller amplitudes. These spectra differences go along with the results shown in Table 9 based on the computation of the standard deviation of the vibrations measured in the mobile robots.

Table 9. Comparison of the vibrations' standard deviation measured in the APR-1 and APR-2 prototypes.

	Standard Deviation of the Acceleration Measured at the:		
	Wheel	**Chassis**	**Head**
Longitudinal			
Side			
Normal			
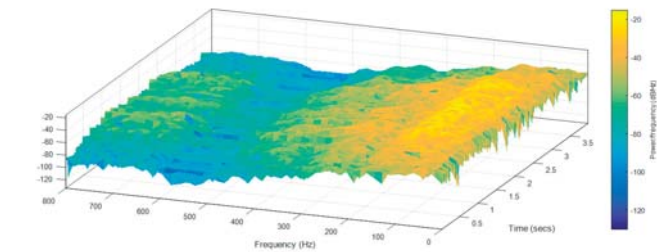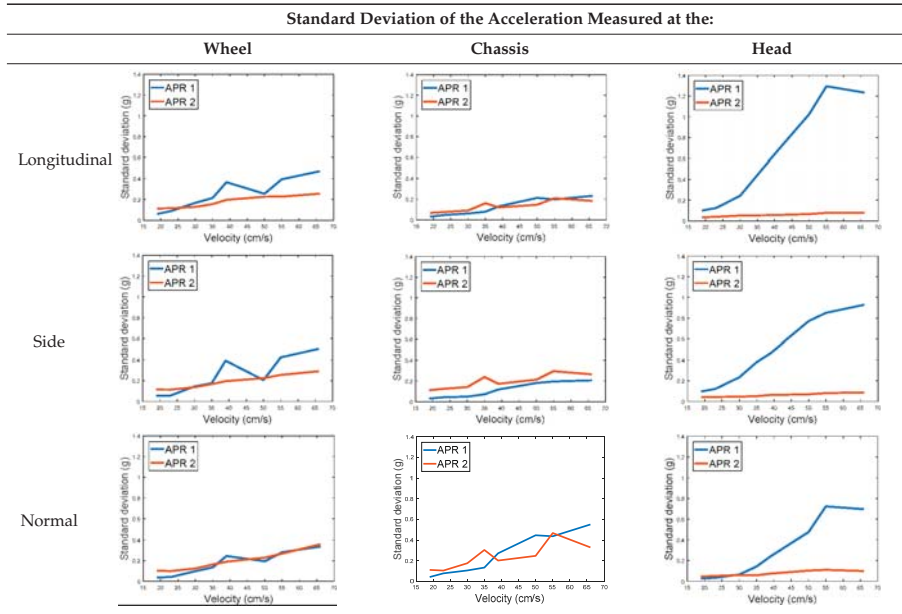
Figure 12. Frequency spectra of the accelerations measured over the longitudinal axis at the head of the APR-1 prototype (without suspension).
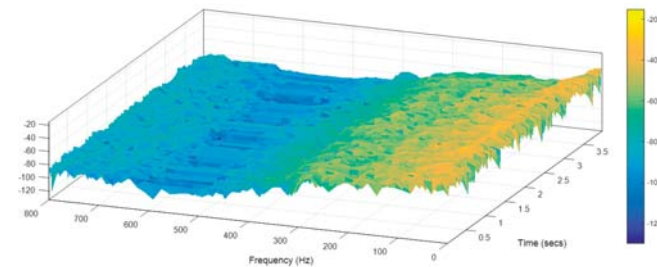
Figure 13. Frequency spectra of the accelerations measured over the longitudinal axis at the head of the APR-2 prototype (with passive suspension).

4. Conclusions

This paper presents the vibration pattern measurements obtained from two tower-typed mobile robots based on a holonomic motion system. The mobile robots are both configured as Assistant Personal Robots (APR). The first implemented prototype (APR-1) includes a rigid structure whereas the second implemented prototype (APR-2) includes a passive suspension system. The design of the APR-2 prototype includes flexible joints between the wheels and the chassis and between the chassis and the tower, and also a damped arm between the chassis and the tower. This design improves the stability as the pivot point is positioned above the center of gravity of the tower. The vibrations have been measured in one wheel, the chassis, and the head of the tower-type mobile robots. In both cases the omnidirectional wheels' gap is the main vibration source. The results have shown that the vibration amplitude measured on both mobile robots when going forward at a constant speed have a typical Gaussian distribution that can be modelled with the value of the standard deviation, and similar results are shown for the vibrations measured in the wheels and in the chassis of the two mobile robots, as they use the same wheel design. Similarly, the results have shown a large reduction of the vibration standard deviation measured in the head of the mobile robot without suspension (1.3 g) and with the passive suspension system (0.08 g). These results represent a huge vibration reduction (1/16) and a huge reduction of associated problems, such as the electronic devices' mechanical stress located at the head of the mobile robots.

Author Contributions: J.M., E.C., M.T., and D.M. designed the holonomic motion system and also designed and performed the experiments; J.C. and J.P. supervised the experiments; J.M., E.C., and J.P. analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fish, S. UGVs in future combat systems. Proceedings of SPIE 5422 Unmanned Ground Vehicle Technology VI, Orlando, FL, USA, 13–15 April 2004; pp. 288–291.
2. Blitch, J.G. Artificial intelligence technologies for robot assisted urban search and rescue. *Expert Syst. Appl.* **1996**, *11*, 109–124. [CrossRef]
3. Martinez, D.; Teixidó, M.; Font, D.; Moreno, J.; Tresanchez, M.; Marco, S.; Palacín, J. Ambient Intelligence Application Based on Environmental Measurements Performed with an Assistant Mobile Robot. *Sensors* **2014**, *14*, 6045–6055. [CrossRef] [PubMed]
4. Wada, M.; Asada, H.H. Design and control of a variable footprint mechanism for holonomic omnidirectional vehicles and its application to wheelchairs. *IEEE Trans. Robotic. Autom.* **1999**, *15*, 978–989. [CrossRef]
5. United Nations. World Population Ageing 2013. Available online: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf (accessed on 15 December 2015).
6. World Health Organization. World Report on Ageing and Health, 2015. Available online: http://apps.who.int/iris/bitstream/10665/186463/1/9789240694811_eng.pdf (accessed on 18 February 2016).
7. Clotet, E.; Martínez, D.; Moreno, J.; Tresanchez, M.; Palacín, J. Assistant Personal Robot (APR): Conception and Application of a Tele-Operated Assisted Living Robot. *Sensors* **2016**, *16*, 610. [CrossRef] [PubMed]
8. Muir, P.F.; Neuman, C.P. *Kinematic Modeling of Wheeled Mobile Robots*; Technical Report, CMU-RETR-8612; Carnegie Mellon University: Pittsburgh, PA, USA, 1986.
9. West, M.; Asada, H. Design of Ball Wheel Mechanisms for Omnidirectional Vehicles with Full Mobility and Invariant Kinematics. *J. Mech. Des.* **1997**, *119*, 153–161. [CrossRef]
10. Blumrich, J. Omnidirectional Wheel. U.S. Patent No. 3,789,947; Patent and Trademark Office, Washington, DC, USA, 5 February 1974.

11. Ilon, B.E. Wheels for a Course Stable Selfpropelling Vehicle Movable in any Desired Direction on the Ground or Some Other Base. U.S. Patent No. 3,876,255; Patent and Trademark Office, Washington, DC, USA, 8 April 1975.

12. Nagatani, K.; Tachibana, S.; Sofne, M.; Tanaka, Y. Improvement of odometry for omnidirectional vehicle using optical flow information. In Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, Takamatsu, Japan, 31 October–5 November 2000; Volume 1, pp. 468–473.

13. Bae, J.J.; Kang, N. Design Optimization of a Mecanum Wheel to Reduce Vertical Vibrations by the Consideration of Equivalent Stiffness. *Shock Vib.* **2016**, *2016*, 8. [CrossRef]

14. Bradbury, H.M. Omni-Directional Transport Device. U.S. Patent No. 4,223,753; Patent and Trademark Office, Washington, DC, USA, 23 September 1980.

15. Park, Y.K.; Lee, P.; Choi, J.K.; Byun, K.S. Analysis of factors related to vertical vibration of continuous alternate wheels for omnidirectional mobile robots. *Intell. Serv. Robot.* **2016**, *9*, 207–216. [CrossRef]

16. Jo, G.J.; Ryoo, Y.J. Vibration Minimization of Tower Typed 2-Wheeled Mobile Robot Using Acceleration and Deceleration Velocity Profile Method. In *Soft Computing in Advanced Robotics*; Springer International Publishing: Cham, Switzerland, 2014; pp. 45–53.

17. Brooks, C.A.; Iagnemma, K. Vibration-based terrain classification for planetary exploration rovers. *IEEE Trans. Robot.* **2005**, *21*, 1185–1191. [CrossRef]

18. Brooks, C.; Iagnemma, K.; Dubowsky, S. Vibration-based terrain analysis for mobile robots. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 3415–3420.

19. Weiss, C.; Frohlich, H.; Zell, A. Vibration-based terrain classification using support vector machines. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 4429–4434.

20. Vicente, A.; Liu, J.; Yang, G.Z. Surface classification based on vibration on Omni-wheel mobile base. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 916–921.

21. Zhang, Y.; Zheng, R.; Shimono, K.; Kaizuka, T.; Nakano, K. Effectiveness testing of a piezoelectric energy harvester for an automobile wheel using stochastic resonance. *Sensors* **2016**, *16*, 1727. [CrossRef] [PubMed]

22. De Pauw, B.; Lamberti, A.; Ertveldt, J.; Rezayat, A.; van Tichelen, K.; Vanlanduit, S.; Berghmans, F. Vibration Monitoring Using Fiber Optic Sensors in a Lead-Bismuth Eutectic Cooled Nuclear Fuel Assembly. *Sensors* **2016**, *16*, 571. [CrossRef] [PubMed]

23. Xu, X.; Chen, S. Field balancing and harmonic vibration suppression in rigid AMB-rotor systems with rotor imbalances and sensor runout. *Sensors* **2015**, *15*, 21876–21897. [CrossRef] [PubMed]

24. Zhan, D.; Yu, L.; Xiao, J.; Chen, T. Multi-camera and structured-light vision system (MSVS) for dynamic high-accuracy 3D measurements of railway tunnels. *Sensors* **2015**, *15*, 8664–8684. [CrossRef] [PubMed]

25. Moron, C.; Garcia, A.; Ferrandez, D. New System for Measuring Impact Vibration on Floor Decking Sheets. *Sensors* **2014**, *15*, 635–641. [CrossRef] [PubMed]

26. Gomez-Gil, J.; Gomez-Gil, F.J.; Martin-de-Leon, R. The influence of tractor-seat height above the ground on lateral vibrations. *Sensors* **2014**, *14*, 19713–19730. [CrossRef] [PubMed]

27. Gómez, M.J.; Castejón, C.; García, J.C. Automatic condition monitoring system for crack detection in rotating machinery. *Reliab. Eng. Syst. Saf.* **2016**, *152*, 239–247. [CrossRef]

28. Moreno, J.; Clotet, E.; Lupiañez, R.; Tresanchez, M.; Martínez, D.; Pallejà, T.; Casanovas, J.; Palacín, J. Design, Implementation and Validation of the Three-Wheel Holonomic Motion System of the Assistant Personal Robot (APR). *Sensors* **2016**, *16*, 1658. [CrossRef] [PubMed]

Enhanced Flexibility and Reusability through State Machine-Based Architectures for Multisensor Intelligent Robotics

Héctor Herrero [1],*, Jose Luis Outón [1], Mildred Puerto [1], Damien Sallé [1] and Karmele López de Ipiña [2]

[1] Tecnalia Research and Innovation, Industry and Transport Division, San Sebastián 20009, Spain; joseluis.outon@tecnalia.com (J.L.O.); mildred.puerto@tecnalia.com (M.P.); damien.salle@tecnalia.com (D.S.)

[2] Department of Systems Engineering and Automation, Universidad del País Vasco/Euskal Herriko Unibertsitatea, EleKin Research Group, San Sebastián 20009, Spain; karmele.ipina@ehu.eus

* Correspondence: hector.herrero@tecnalia.com

Abstract: This paper presents a state machine-based architecture, which enhances the flexibility and reusability of industrial robots, more concretely dual-arm multisensor robots. The proposed architecture, in addition to allowing absolute control of the execution, eases the programming of new applications by increasing the reusability of the developed modules. Through an easy-to-use graphical user interface, operators are able to create, modify, reuse and maintain industrial processes, increasing the flexibility of the cell. Moreover, the proposed approach is applied in a real use case in order to demonstrate its capabilities and feasibility in industrial environments. A comparative analysis is presented for evaluating the presented approach versus traditional robot programming techniques.

Keywords: intelligent robotics; flexibility; reusability; multisensor; state machine; software architecture; computer vision

1. Introduction

An analysis [1] of the current situation in manufacturing plants highlights three major trends:

- An ever-increasing customization of products and short lifecycle, which require an increase in the flexibility of the production means (one unique system must handle all of the product diversity and operations) [2,3]. Robots fit perfect into this topic due to their versatility; robot programs can adapt to the customizations of the products.
- A large variation in production volumes, which requires an increase in the reconfigurability of production (one system for one product/task within recombinable production lines) [2,4]. Robotic mobile platforms play an important role in this trend; easy to move robots are necessary in some production chains where production volumes change frequently.
- Limited access to skilled operators due to an aging workforce, changes in education and an ever-faster technology development. This requires new solutions to assist operators and provide collaborative work environments [5]. Collaborative robotics are being developed for this topic.

The research addressed in this paper focuses on the first trend: the need for highly flexible and intelligent robotic systems. Despite the large effort in the research community, large companies, as well as small and medium enterprises (SME) still do not have appropriate software tools and solutions to react rapidly with economic viability for an interesting return of investment for the automation of their processes. The direct consequence is that production operations are mostly performed manually, with

high operation costs that endanger those companies with respect to lower wage countries. This research is thus oriented toward developing and providing a software ecosystem that allows for a rapid and efficient programming of production processes, providing the required flexibility and permitting an effective integration of auxiliary sensors and artificial vision systems. Even if this approach is generic and applicable to industrial manipulators, this paper will be focused on dual-arm multisensor robotic operations.

The dual-arm robots provide more dexterity, in addition to the advantage that they can be used in the existing workstations. Due to these arguments, the dual-arm robot deployment is growing year by year, not only in large multinationals, but also in SMEs. Sector experts [2,6] affirm investments for robot deployment are amortized in 1–2 years; however, this information cannot be extrapolated to all cases. However, applications with short production batches, environments prone to many changes and processes that need human-robot collaboration or special environment supervision do not comply with this trend. Dual-arm robots are being introduced in such contexts. The growth of dual-arm systems [7] is resulting in many efforts made by robotic researchers to manage them. Programming, coordinating and supervising bi-manual robots is a need that is increasingly being demanded by the community; even more with the rise of collaborative robots, which have to integrate different sensors for cell supervising and monitoring [8,9]. In this scenario, the need for actuation when external signals are received becomes essential, e.g., a person enters the workspace of the robot, and the robot must stop its movement and adapt its behavior.

In this paper, we present an approach to alleviate the challenges that can be identified for dual-arm robotic programming. The presented framework eases the deployment of industrial applications and allows managing the execution control, increasing the reliability and traceability of the system (Section 2). To ease the deployment of this kind of application, we present how the framework can integrate skill-based programming. For understanding the advantages, the assembly operation of an aeronautical part is detailed. Moreover, an evaluation of the architecture is presented (Section 3). Finally, we present the discussion, conclusions and future work (Sections 4 and 5).

2. Materials and Methods

2.1. State Machine-Based Execution Coordination for Dual-Arm Robots

Traditional robot programming is still not very flexible; thus, the dual-arm programming suffers the same problems. In the industry, smaller and smaller batches are ordered, and as a consequence, the costs of reprogramming the robots grow. Even though there are usually different parts, the process is very similar, e.g., assembling parts with different types of screws. In this case, the assembly operation is the same; only the screw size, type or position is changing. Those tasks can be modeled; the key is to be able to subdivide a task (screw operation) into smaller operations (robot movement, end-effector actuation, etc.). Then, re-using these tasks can be made parametrizing correctly the corresponding suboperations without needing to reprogram the whole task. Grouping the robot basic movements (primitives) according to tasks or skills is an alternative that many authors have followed [10–14].

One of the most relevant issues in dual-arm robotic programming, especially for industrial applications, is the lack of powerful and easy to use graphical user interfaces [15]. An easy to configure graphical user interface (GUI), which allows the previously-mentioned skill-based programming, will enable operators to program and maintain the industrial processes. This, in addition to the workers feeling a part of the automation process, will also contribute to reduce the costs of the robotic systems' deployment.

Regarding the execution control, state machines can address dual-arm challenges. These tools are commonly used for general-purpose processes, and in particular, they have been extensively adopted by the robotics community. In this aspect, the work made by different authors combining finite state machines with knowledge and skills is very relevant [16–18]. State machines are an easy way for describing behaviors and for modeling how components react to different external and internal

stimuli [19,20]. In this area, there are different implementation alternatives, e.g., there are many projects using Orocos rFSM [21]. rFSM is a small and powerful state-chart implementation designed for coordinating complex systems, such as robots. SMACH [22] is another implementation of state machines. It can be defined as a task-level architecture for rapidly creating complex robot behavior. In this work, SMACH has been selected for implementing the state machine. One of the reasons is because SMACH can be used under the ROS (Robot Operating System) [23,24], which is a flexible framework for writing robot software. ROS is a collection of tools, libraries and conventions [25] that aims to simplify the task of creating complex and robust robot behavior across a wide variety of robotic platforms [26]. As a complementary element to the execution control, multi-agent systems can be useful for decision making in coordination and synchronization tasks [27,28].

2.1.1. Proposed Architecture

As illustrated in Figure 1, the proposed state machine interconnects the application development framework (graphical user interface) and the robotic lower level control system. The presented state machine is composed of different states where each state corresponds to one of the basic operations that the robot can execute. Basic operations are considered the functions or commands that by themselves are able to achieve a goal, e.g., Cartesian point to point interpolation. It can be understood as a robot API (application programming interface). Following the program provided by the user, the active state triggers its corresponding state to execute the necessary functions.
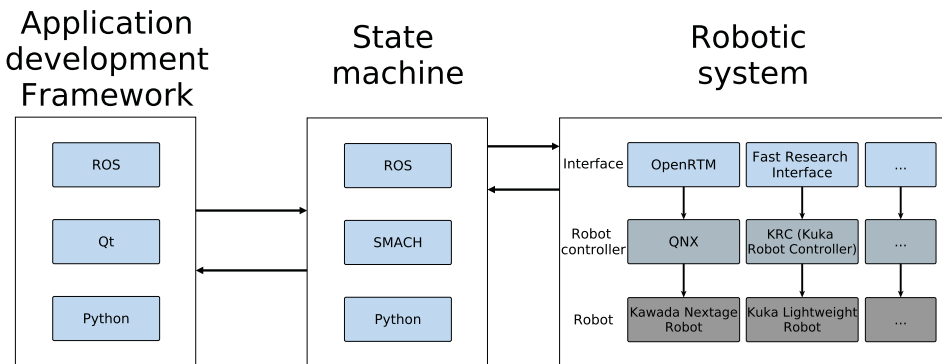
Figure 1. Proposed overall architecture. The figure shows how it is divided into three levels.

In this research, all of the prototypes are being tested and validated in a dual-arm robot, specifically in a Kawada Nextage Open Robot (Figure 2). This robot has humanoid aspects, with two arms of 6 degrees of freedom (DOF) attached to a rotatory torso; it is equipped with a 2-DOF head, which incorporates the stereo vision system. In conclusion, it is a 15-DOF robot managed by a single controller. In order to obtain more precision, other stereo vision systems have been added to each wrist.

As is detailed in Section 3.1, the applications are composed of tasks, and these in turn are composed of primitives (or previously-mentioned basic operations), which are translated to states. On the one hand, the execution engine triggers state changes at the low level. On the other hand, in the case of the Kawada Nextage Open robot, the states are connected to the robotic system through an OpenRTM bridge [29]. Even so, it should not be forgotten that ROS allows hardware independence, and changing the bridge properly, another robotic system can be used (for example, Orocos or the Fast Research Interface [30] to interface a Kuka LWR with the proposed architecture).

This combination of the application development framework and a low level state machine allows us to considerably improve the flexibility and hardiness, make the programming easier, achieve hardware independence and environment control, resulting in a more industry-oriented solution.

Figure 2. Nextage Open Robot where all developments are being tested.

2.1.2. Core Description

One of the first requirements that was identified was introspection, which is a tool able to provide the current execution state continuously, allowing us to manage possible errors and improving the recovery from them. In Figures 3 and 4, the proposed architecture is outlined. The proposed architecture consists of two state machines, one per arm, with some common states. These common states are used when a synchronization between the arms is required, i.e., when both arms of the robot have to move at the same time. This combination of two state machines related by some common states combines the advantages of having individual machines for processes that do not need dual-arm cooperation, with the robustness that allows centralized states for dual-arm requiring processes. The use of the SMACH/ROS combination provides some tools that are very useful for introspection. SMACH uses ROS messages for publishing, besides other information, the current state; thus, any module of the software can be checked easily.
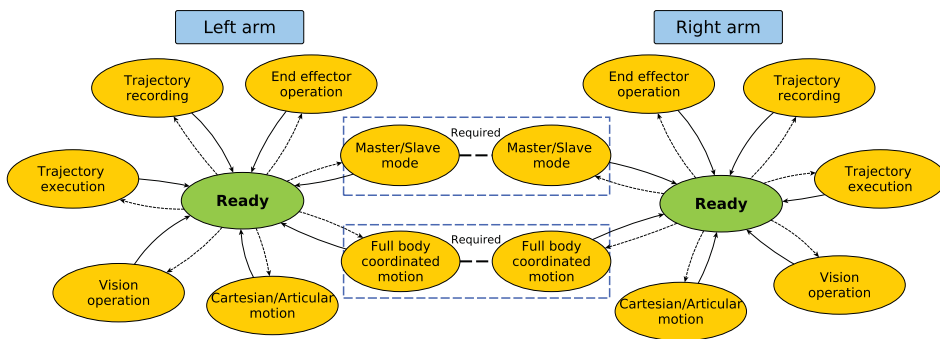
Figure 3. Proposed state machine-based architecture. The figure represents an overview of the architecture.

When the application is launched, the system starts from a ready state and keeps changing to different states that can be seen as available abilities or capacities of the robot. Note that some states have not been included in order to simplify the diagram. These states are pause/stop, error handling and finish. The proposed work in this paper allows either human or sensor-based supervision of the environment and permits canceling or adapting plans according to sensor values and perception system information. When an error occurs, e.g., in a trajectory execution, the system is able to cancel

the current operation in order to handle the error and return to a safe position (if possible) or enter an alarm state that requires operator intervention.
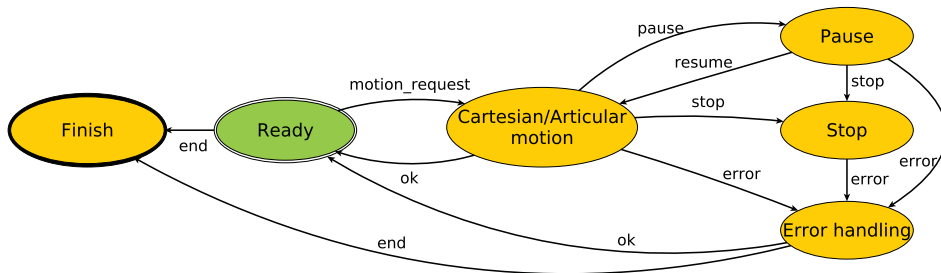
Figure 4. Proposed state machine-based architecture in detail. The figure shows existing states and transitions.

2.1.3. Description of the Developed States

Each state has been implemented as a module that generally is independent from the core. Only a few modules have been defined as fundamentals. These special modules are articular/Cartesian, full body coordinated motion and trajectory execution. All available modules for this version are shown in Figure 3. It should be emphasized that according to the requirements of the different applications, the available states can be updated by incorporating new capabilities or removing others that will not be used.

Table 1. Summary of the main elements of the state machine.

State	Description
Ready	The state machine is ready for receiving new instructions. This state is waiting until the execution engine sends a new request.
Cartesian articular motion	Manages the robot movements both in the Cartesian space and the articular space. If the movement cannot be executed correctly, there is an error handling state to manage it.
Full body coordinated motion	Allows controlling both arms in coordination. Two arms must be in this state to start coordinated motion. Sending the values of the 15 joints of the robot is necessary.
Record trajectory	Allows recording trajectories with a trajectory planner or teaching by demonstration. These trajectories are stored in a database for future use.
Trajectory execution	Executes trajectories, provided by a trajectory planner or previously stored in a database.
End-effector operation	Manages end-effector operations; depending on the end effector, different operations can be made, e.g., gripper open/close, deburring tool activate/deactivate, screwing operation, etc.
Vision operation	Manages different computer vision operations. This includes picture acquisition, processing and reference frame transformation, among others. As the robotic system has multiple vision systems, this state is responsible for managing them depending on the operation that will be executed.
Master/slave mode	Puts robot in bi-manual coordinated manipulation mode; one arm actuates as the master and the other one as the slave. Consists of planning a trajectory for the master arm and then computing this trajectory with an offset for the slave arm.

In order to understand the proposed architecture, Table 1 summarizes the different states and their utility. Besides, in Table 2, a summary of the signal and transitions is presented. Each state may contain a more or less complex structure according to its purpose. On the one hand, for example, the vision operation state only contains the calls to different vision functions. On the other hand, the articular/Cartesian motion state is highly general, i.e., this state contains all of the required code to manage motions both in Cartesian and articular spaces. For a state transitioning, different events are handle; these events can be thrown out by the safety supervision system or by any module.

Table 2. Summary of the signals and transitions of the state machine.

State	Signal	Transition to
Ready	motion_request	Cartesian/articular motion
	vision_request	Vision operation
	end_effector_request	End effector operation
	...	...
	end	Finish
Cartesian Articular motion	ok	Ready
	pause	Pause
	stop	Stop
	error	Error handling
Pause	resume	Cartesian/articular motion
	stop	Stop
	error	Error handling
Stop	error	Error handling
Error handling	ok	Ready
	end	Finish

2.2. Flexible Application Development

The proposed architecture in this paper not only refers to the state machine-based execution manager, but contains everything necessary for deploying different robotic applications. One of the key advantages of the proposed work is that different applications reuse the common structure of the framework.

2.2.1. Software Structure of the Framework

In order to ease the maintainability and assure software quality, the developed framework is organized into different packages. In this way, following the ROS philosophy, each package must fulfil minimum quality criteria.

The simplest application is composed by at least the following three packages: execution engine, core functions and application functions. Figure 5 illustrates these packages (three columns) and the relation between them. As can be seen, the execution engine creates (instantiates) the state machines. Each state machine has an instance of an application function (RivetInstallation, AntenaAssembly, etc.). Application functions inherit from core functions all of the attributes and methods, which allows using the robot basic operations (Section 2.1.3), enhancing and particularizing them for applying into specific industrial applications. In this way, all applications are composed by core functions (basic operations) and application functions, which are a combination of the previous ones. These function libraries basically configure the requests for the state machine filling required parameters. This organization also allows having specific graphical user interfaces for each project (rivet_installation_gui) and a common one for basic robot guiding or teaching (dashboard).
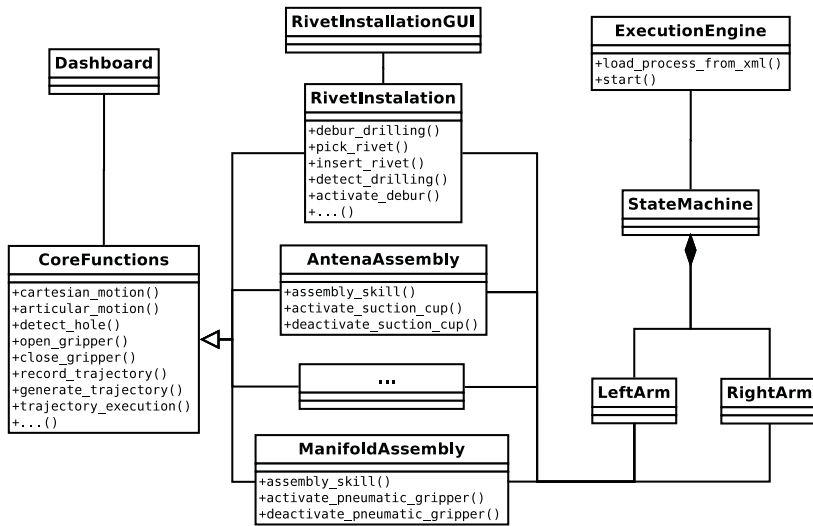
Figure 5. Software structure of the framework.

2.2.2. Execution Engine

The execution engine creates two threads, one per arm; these threads will contain instances of the proposed state machine. The execution engine will continue its execution managing the request of operations, i.e., the execution engine is responsible for orchestrating the application flow.

At this point, it is important to think about the change of paradigm for executing robotic applications. As has been explained here, there are three "independent" threads. As the proposed architecture is running under ROS, the state machine threads are actually ROS nodes and basically act like threads with their own parametrization and independent behavior. The execution engine communicates with these nodes via ROS messages, which contain robot commands with the necessary parametrization; in this way, each node receives commands to execute and starts triggering the state machine to the convenient state. When the operation is finished, the state machine returns to the ready state. The heart of the matter remains in how these messages are generated and managed (Section 2.2.3).

The consistency of the execution is guaranteed by the deterministic operation of the state machine. Each node will not receive the next operation until necessary synchronization requirements are met, i.e., until the execution engine can assure that state machines are in the ready state. In Section 3.1, a real use case is presented explaining how the operations are executed maintaining the coordination of both arms.

2.2.3. Application to Executable XML

As mentioned above, applications are stored in XML files, with the particularity that each group of the robot (left arm, right arm and torso/head) has its own instructions. This is because each state machine needs to execute operations both synchronously and asynchronously: in some cases, a process requires both arms of the robot at the same time, e.g., a big part that needs two arms for a correct handling; in other cases, some process can require the use of both arms, but not at the same time. XML files contain, in addition to the operations, the necessary flags and synchronization tools to assure this coordination. In this paper, for the presented use case, the simplest instruction for coordination is used: a wait instruction. This allows one arm to wait until the other arm finishes its ongoing operation.

Generating a simple application (as can be seen in Figure 6) can be performed writing each XML file by hand; however, when the application and complexity grow, it is difficult to maintain the correct perspective and timeline, leading to errors. To address this, a simple graphical interface can be used. The presented GUI in Figure 7 obtains a list of available functions from core functions and application function packages (introduced in Section 2.2.1). For creating new applications, the user has to add functions and parametrize them. With the help of the graphical interface many programming errors are avoided, especially for the synchronization of both arms, allowing a global vision of the execution flow. In Figure 7, at the right frame, the application program is represented; the displayed example is for rivet installation process. When the application is ready, an XML file is created, containing the list of commands that each arm has to execute. The wait function represents the simplest synchronization mechanism, because in those time lapses, the left arm has to wait until the right arm finishes; therefore, in the generated XML file, this will be translated as the wait synchronization operation.

```xml
<operation id = '1'>
    <left_arm>
        <function name = 'debur_drilling'>
            <params>
                <param value = 'drill_1'></param>
            </params>
        </function>
    </left_arm>
    <torso_head>
        <function name = 'wait'></function>
    </torso_head>
    <right_arm>
        <function name = 'pick_rivet'>
            <params>
                <param value = 'rivet_1'></param>
                <param value = 'ref_R10'></param>
            </params>
        </function>
    </right_arm>
</operation>
<operation id = '2'>
    <left_arm>
```

Figure 6. Application program fragment.

Figure 7. Simple GUI for new application development.

3. Results

3.1. Validation in a Real Use Case

As Tecnalia [31] is in direct contact with companies in different industrial sectors, these developments have been tested in several scenarios with different requirements. One of the most relevant use cases is for the aeronautics sector; Tecnalia and Airbus Operations (Puerto Real facilities, Spain) have been working together for several years developing pilot cells for a dual-arm robot (see LIAA [32] (the EU's FP7 program) for flexible assembling operations). The first steps toward the technology transfer for industry validation of this architecture are currently in process in the Rapid Reconfiguration of Flexible Production Systems (ReCaM) (this research has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 680759) [33] project (the EU's Horizon 2020 program). The relation between a technological center (Tecnalia), a robotic system integrator (DGH [34]) and the end user (CESA [35]) is a key issue in ReCaM, where the aim to demonstrate a set of integrated tools for the rapid reconfiguration of flexible production systems, particularly the assembly of aeronautical actuators, is addressed.

As had been mentioned in previous work [36], one of the most relevant tasks in the aerostructure assembly is the rivet installation operation. In this paper, the progress made on the automation of the riveting installation is presented; in the current prototype a deburring operation has been added, because this prepares the surface of the drilling perimeter for the correct rivet installation. This operation is performed with an integrated deburring tool in one of the grippers of the robot. The other gripper is prepared for taking and introducing rivets into drilled holes. This demonstration takes advantage of the dual-arm capabilities. Furthermore, for robot perception, a stereo vision system has been incorporated for the precise hole detection; using incorporated stereo cameras on the arms, production tolerances (0.2 mm) can be achieved [37]. In the same way that the vision system is used, different kinds of sensors can be integrated adding the corresponding state to the state machine.

Summarizing, the current demonstrator is composed of the following steps:

a. Detect and debur the drilled hole with the left arm (Figure 8).
b. Pick and extract the rivet from a tray with the right arm (Figure 9a).
c. Insert the rivet into the detected hole with the right arm (Figure 9b).

Figure 8. A drilled hole is deburred after detecting its position by vision.

Figure 9. (**a**) The right arm of the robot is taking a rivet from a tray; (**b**) after taking the rivet, it is introduced in the previously-detected drilled hole.

Note that Steps a and b can be performed at the same time, because the rivet extraction operation take more time than the deburring operation.

If these operations are viewed as skills, the deburring skill, pick rivet skill and rivet inserting skill are obtained. Figure 10 shows how skills are decomposed into primitives. The organization into skills eases the composition of new programs, because the parametrization is perceptibly easier. This parametrization contains the key features that vary between different skill executions. The way to determine the parameters is as follows: the system programmer starts by selecting the references or elements that change for different scenarios. For example, in the case of deburring and insertion, the references of the holes and rivets to be inserted must be parametrized. If that would not be enough, the parameters that allow one to configure the differences between scenarios would be added. Once these skills have been validated, abstracting from primitives is possible. In the case of the deburring operation, only the theoretical position must be provided, taking into account that this information can be extracted from the CAD model of the piece.

Calibration or the referencing process of the cell is beyond the scope of this work (even if it will be addressed in future work); nevertheless, it can be easily summarized in three steps: at first, the positions of the drilled holes are obtained, referenced to the origin of the CAD model. After that, using an accurate tool center point (TCP), three known points of the real piece are touched; the easiest way is usually touching one corner and their adjacent edges with the TCP. With these points, the position and orientation of the piece can be estimated. Finally, using the obtained theoretical position of the piece in the robot frame and the position of the hole in the piece frame, a frame transformation can be done to obtain the approximate position of the piece drillings. Of course, this approximate position must be corrected using artificial vision to achieve the required 0.2 mm of precision.

Returning to the proposed architecture, once the skills are decomposed, the resulting primitives are the ones that are executed by the state machine. Each state is processing the primitive callbacks and handling errors if they take place. Thus, the error handling is simpler, and it is managed specifically in each state or module. Taking one of the operations that are being analyzed, the sequence of the machine state is shown in Figure 11.

The execution engine sends to the state machines the request for the next operation, based on the information that is stored in the application XML (see Figure 6). The state machine changes from one state to another, completing the requested operations. As can be seen in Figure 12, some operations of the task of installing one rivet can be performed using both arms of the robot at the same time,

improving the cycle time. After these coordinated operations, an exclusive movement of the right arm is performed; at this moment, the left arm is waiting until the right arm finishes the installation of the rivet. The whole process of rivet installation is composed by the repetition of this block of skills. In order to demonstrate the adaptability of the presented framework, the CESA [35] use case is presented.
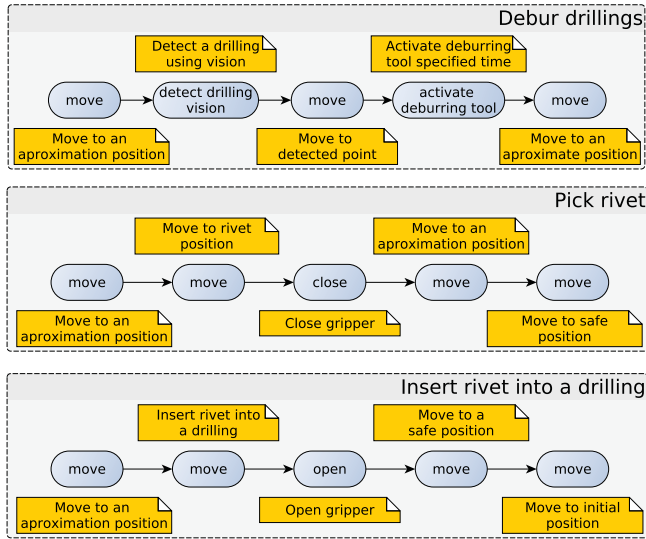
Figure 10. Install rivet process organized into skills. Skills are composed by primitives.
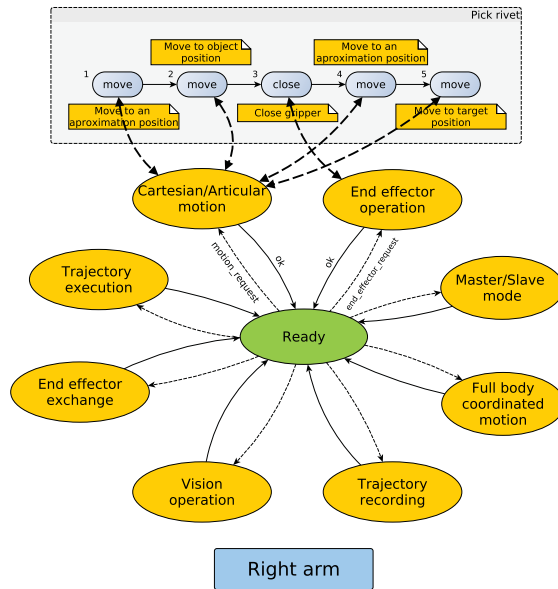
Figure 11. Debur drilling skill mapping into the state machine.
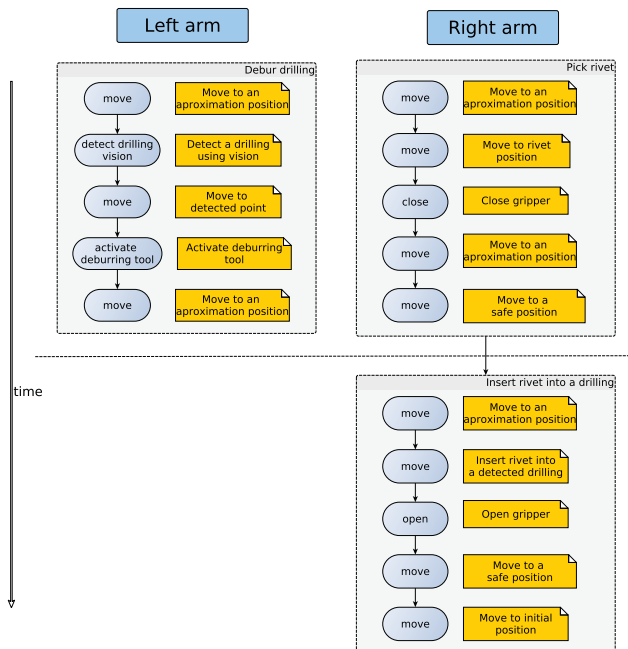
Figure 12. Coordination between both arms' timeline.

As has been mentioned above, Tecnalia is working on different projects with assembling operations in the aeronautical sector. This use case is being developed under the ReCaM project [33], one of the principal topics of which is the development of assembly capabilities for robots. In the ReCaM project, the starting point will be the product requirement description, which is first matched against the resource capabilities existing on the current system layout. If no matches are found, the system needs to be reconfigured. New resources can be searched from the resource catalogs. This matchmaking and search is allowed by the OWL-based capability model [38], which is used to describe the resource capabilities in a formal, computer- and human-interpretable manner. The capability matchmaking approach is presented in [39]. Once the system has been re-configured (or found suitable as such), the actual operations need to be programmed and executed. For this programming, the skill-based approach by Tecnalia is utilized. Basically, the required steps are the following: pick and assembly various elements (valves, springs, caps, etc.) into a manifold. All of the elements are stored in a kit, which can be referenced and located by artificial vision.

In this demonstrator, the information extracted from CAD models (an offline process that is not in the scope of this work) plays an important role. This information is modeled into different XML files: fixture information and element information. The fixture information XML file contains the position and orientation of the relevant points of the fixture; these points are marked as targets for pick and place operations. The element information XML file contains the grasp position, the necessary gripper for grasping and the assembly point in the model, i.e., the point that is necessary to align with the fixture relevant point. This skill is able to perform the steps listed above to complete the assembly of different elements into the manifold, only taking into account the information provided in the XML files. Figure 13 shows a detailed example of how the assembly skill is parametrized using the provided information.
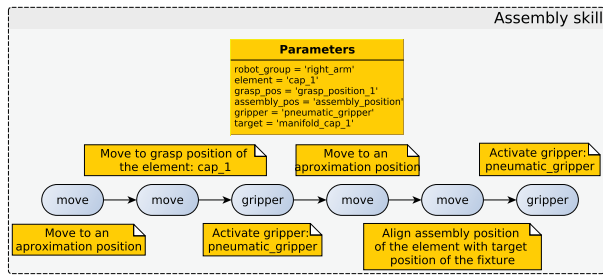
Figure 13. Assembly skill configuration for one cap of the manifold.

As can be seen, different applications can be modeled following the same schema; the parameters that appear in the skill configuration are codified in an XML file (as has been presented for the previous use case in Figure 6). This XML is completely compatible with the state machine and execution engine (Sections 2.1.1 and 2.2.2, respectively). The system capacity to adapt to changes in the environment provides advantages. For instance, if there is variation in the position of parts (elements) or in the number of parts to process, the high-level program can be adjusted through minor changes (e.g., reprocess the CAD model for updating positions and adding more blocks of a particular skill). No changes in the low-level program are needed. As a consequence, an increase of system flexibility has been achieved.

3.2. Evaluation

In the last few years, several methods for evaluating software architectures have been defined: scenario-based (SAAM, architecture tradeoff analysis method (ATAM), ALMA, etc.) [40–45], mathematical model-based (reliability analysis, performance analysis) [46] and metrics-based software architecture evaluation methods (QuADAI) [47]. In order to evaluate the advantages of the proposed architecture, based on the previously-mentioned methods, a simplified approach of the architecture tradeoff analysis method (ATAM) has been selected to perform a comparative analysis [40,48,49]. This method is widely used by the research community for architecture evaluation [50–52]. When the architecture is evaluated, depending on the requirements, different qualities must be analyzed. ATAM concentrates on evaluating suitability; therefore, the selection of the appropriate qualities has a remarkable relevance.

As has been mentioned above, ATAM is a scenario-based method; that is why different scenarios have been selected in order to compare different desirable qualities. On the one hand, the creation of a new application from the beginning scenario has been chosen. New application deployment implies working environment definition, relevant position acquisition, fixture calibration, robot process programming, simulating, testing and adjusting. On the other hand, another common scenario is proposed, adapting an existing application to new product references (the required process would be the same, but could change the number of operations or the dimensions of the elements).

The proposed architecture has been compared with different ways of addressing the automation of an industrial process [53,54]. The traditional and most commonly-used method is online programming, i.e., teach by demonstration (moving the robot with the teach pendant), replicating the process and acquiring required way points. In other cases, the use of offline programming software can be found. This approach is composed by the following steps: the generation of the 3D scene, tag creation, trajectory planning, process planning, post-processing simulation and calibration [54]. As can be seen, the proposed framework in this article is very similar to an offline programming process, though with some improvements.

In order to evaluate different approaches, a set of desirable qualities have been analyzed:

Ease of use: To deal with the first scenario, differences between online programming and other alternatives are evident. A new application deployment requires stopping the production for fixture calibrations, way point acquisition, process replication, simulations and adjustments. These tasks require a high expertise in robotics and programming. With offline alternatives, the process can be offline almost entirely; only calibration and final adjustments require stopping the production. Generally, offline programming software is very complex and also requires highly trained staff. The cost of these technicians (plus license costs) could not be affordable for SMEs. The proposed framework provides a set of ease-to-configure primitives and skills, which reduces the training costs.

Adaptability: This quality impacts the second scenario. Modifying an existing process using online programming is very time consuming; new position acquisition moving the robot implies stopping the production. For offline programming, changes can be made without stopping the production, although depending on the nature of the changes, this could imply repeating many tasks in order to adapt the application. In the case of the proposed approach, the process is similar to offline programming, though with the particularity that the developed skills are programmed keeping in mind possible changes. For example, in the case of deburring and riveting holes (Section 3.1), possible changes in the hole positions and rivet size are anticipated, so the skill takes the information of the hole position and size from a processed CAD file. Then, the skill adapts its behavior, configuring the target position and gripper aperture with respect to the obtained information. The same idea is applied in the assembly operation; the developed skills can adapt to usual changes in this kind of process: changes in assembly points' positions, changes in parts' size, etc.

Reliability: The presented approach provides an implicit supervision tool: the state machine allows knowing the current status of the execution. Besides, the modular error handling permits an individualized response for the different types of errors. Traditional robot programming techniques require ad hoc error handling in each critical part of the program.

Subsetability: This is the ability to support the production of a subset of the system [54]. This concept could be important in different ways. For the commercial side, the possibility of having different optional modules (states or even skills) is an advantage. In the case of requiring incremental developments, the possibility to deliver simple prototypes that are enhanced with new modules and abilities is interesting. For the end user, having only the functionalities that are required could reduce the training time and increase the ease of use. Subsetability quality does not exist for traditional online programming, and for offline programming, software usually is used only for the commercial aspect.

Performance: Both online programming and offline programming have the best performance, because these methods do not add any layer of software in the execution time, i.e., when the configuration or set-up phase concluded, only a robot specific code is executed in the controller. In the proposed framework, the XML program is parsed for executing existing skills, which are composed by primitives that execute directly in the robot controller. This, combined with the overhead from the state machine, results in greater demands on processing resources. Even so, the executed process and robot movements are the same for all alternatives, so these differences in performance do not affect the overall operation.

Table 3 summarizes the strengths and weakness of different robot programming approaches. Online programming is the simplest approach, which only has the performance as the clear advantage. The proposed approach can be seen as an enhanced offline programming method; both have in common many insights, in spite of the fact that through the skill programming and state machine-based architecture, the ease of use, adaptability and subsetability have been improved. Thanks to the developed skills, many of assembly applications that are composed by pick and place operations can be easily modeled and resolved by the presented framework. This proposal is a step forward in the generalization of this kind of problem. These improvements have a performance drawback, but taking into account the advantages, the trade-off is acceptable.

Table 3. Strengths and weakness of different robot programming approaches.

Quality	Online Programming	Offline Programming	State Machine and Skill Based Programming Framework
Ease of use	−	+	++
Adaptability	−	+	++
Reliability	−	+−	+
Subsetability	−	+	++
Performance	++	++	−

Based on the obtained conclusions in Table 3, the representation of the claimed improvements has been done. The more important qualities that have been improved are the ease of use and the adaptability. These improvements are translated directly into the reduction of the development time. Despite that the required time for the programming of different automation processes can vary widely, one of the most usual operations has been selected: pick and place. If the CESA use case has been taken as the reference (Section 3), in the following lines, an analysis of the required time for programming the assembling operations can be found.

Figure 14 shows how the online programming development time grows linearly according to the number of operations that must be programmed. Each operation requires moving the robot manually and storing waypoints. Regarding offline programming, an initial overrun can be perceived, mainly due to the required time for cell referencing, i.e., the transition between simulation and reality. After that, successive operations require less time than manual teaching. Concerning skill-based programming, higher initial overrun is necessary, due to the required cell referencing and the additional information, which complements the skills (grasp positions, assembly positions, gripper information, etc). When this information is modeled, the successive instantiation of assembly skills is faster; only drag and drop and simple parametrization are required. In conclusion, it can be perceived how when more than five operations are required, the skill-based programming offers better performance.

Figure 15 presents the required development time for adjusting an existing process, i.e., when something has moved or another reference of the product requires position adjustments. As before, online programming will require repeating all of the process, teaching new waypoints and assuring no collisions. Regarding offline programming and skill-based programming, in this case, they behave in a similar way: one the one hand, an initial cell referencing is necessary, and on the other hand, as the program is already created, only parameter modifications are required. Of course, the necessary changes are different in both methodologies, but the same required time has been estimated.

Finally, Figure 16 shows the required time if the robot of the process is changed to a different one. Taking a process composed by 10 operations, for an online programming approach, this is a completely new process. Using an offline solution, in the best case, the program sequence can be reused. However, it must be noted that a revision of all of the waypoints must be done. In the case of skill-based programming, the developed skill does not require a revision in terms of programming or parametrization because the problem to resolve is the same. For the proposed framework, this scenario is taken as another process adjustment, requiring the same time as in the previous case.
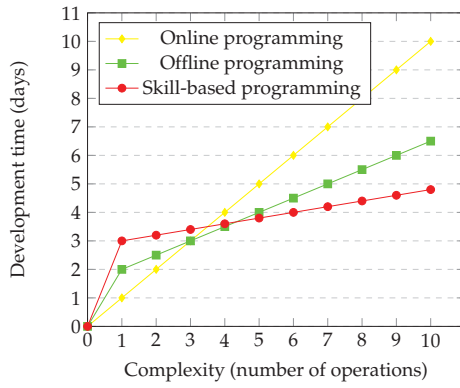
Figure 14. Comparison of the process development time according to its complexity.
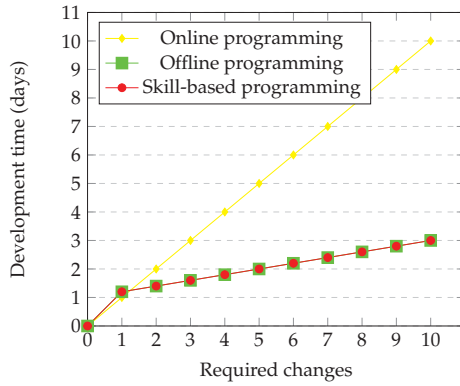
Figure 15. Comparison of the process development time according to the number of adjustments in element positions.
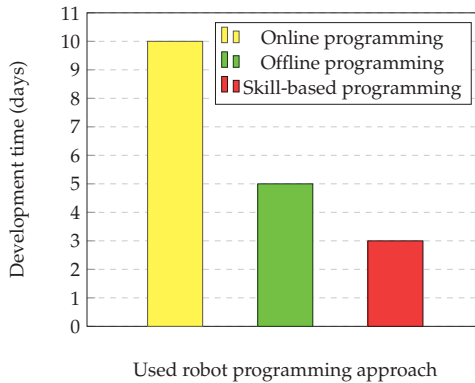
Figure 16. Comparison of the process development time when the robot provider is changed.

4. Discussion

As has been analyzed in the previous section, the presented approach in this article offers greater flexibility and reusability (adaptability) than traditional frameworks. On the one hand, the flexibility of this approach is demonstrated by the fact that the same skills can be used to perform different processes although they suffer from certain variations, e.g., variations in the rivet models, variations in the drilled holes' number or positions, etc. This assertion is supported by the work that the authors have made in different applications [55–58]: another deburring process was performed using very similar skills; the antenna assembling skill was presented; workspace monitoring and vision operations for hole detection and 3D CAD matching were integrated as skills; and finally, the interaction between the skills and the state machine was presented. On the other hand, new applications can be generated graphically (Section 2.2.3), reducing the required expertise and increasing the ease of use. When the user adds a skill to the execution flow, all required parameters must be filled. In this way, a succession of blocks, which composes the application, is generated. The developed GUI allows exporting sections or entire applications into XML files in order to increase the re-usability.

One of the foreseen advantages of the present approach is that the state machine architecture can be enhanced with different modules (states) that could be useful in completely different processes. In the proposed scenario, the states are related to the robot primitives, i.e., robot movements controlled in velocity in the Cartesian space. Nevertheless, the proposed primitives can be combined with nonlinear controllers, such as predictive control [59], neural networks or fuzzy approaches [60,61], needed in other industrial processes with high uncertainty in the model like chemical processes (i.e., petrochemical plants). The skills approach could provide additional information and actuation; basic functionality could operate the aperture or closure of valves, and a complex implementation could cover other acting elements. This is an idea explored in the TOP-REF project [62].

Regarding reliability and robustness that the state machine provides, it permits users to abstract from the specifics of dual-arm robotic programming. The proposed framework eases the coordination of both arms with the help of a simple GUI (Figure 7). Besides, a complete traceability of the program status combined with modular error handling increases the overall reliability compared with traditional online and offline software.

One of the drawbacks of the presented approach is the performance. The entire ROS ecosystem added to the state machine requires a powerful computer, but taking into account the cost of a computer in relation to automation project costs, this is not a relevant issue. Another relevant topic is that the proposed architecture is hardware agnostic; the developed skills are not using robot-specific functions; however, when primitives are executed, ROS interfaces are used. ROS is compatible with a large number of robots [26], though for an industrial environment, ROS-Industrial [63] is more adequate. ROS-Industrial appears with the support of a large research community and robot manufacturers. Their goal is to provide reliable and robust ROS packages. The list of supported industrial robots [64] is growing day by day. This can be a disadvantage compared with available offline programming software, e.g., Delmia, which offers a huge database of robots.

In the industrial world, presenting a framework mostly composed of open source modules always causes a discussion. Even so, as has been mentioned in Section 1, nowadays, more flexibility and novel solutions are demanded, and open source initiatives like ROS are responding to these requirements of the industry.

5. Conclusions and Future Work

To improve the control and coordination of anthropomorphic multisensor robots, state machine-based architectures have been introduced. This approach allows us to increase the robustness and reliability of the whole system. The proposed architecture is designed to act as a basis for easier programming methodologies. Thanks to the presented graphical user interface, new applications can be generated without the need to be an expert in robotics. With the proper training, the operator will be able to create, adapt and maintain industrial processes.

Besides these advantages, the reusability has been noticeably increased. By employing the software architecture that has been presented, completely different applications can leverage well-tested modules and functions used in previous developments. At present, the same architecture is being used in different pilot stations with different types of robots and requirements; in these pilot stations, this technology is under intense tests for validating the usability, robustness and feasibility.

The proposed architecture has been compared with traditional approaches in order to analyze and highlight the strengths and weakness. ATAM has been selected in order to evaluate the qualities that have notable relevancy: ease of use, adaptability, reliability, subsetability and performance. The required development time for accomplishing assembly operations has been compared. The results of the evaluation reveal that the framework improves almost all of the mentioned qualities; the exception is the performance in terms of computational cost, which is inevitably increased by the additional software layers introduced. The next step to follow in the future will be performing a wider test bench for evaluating and comparing the performance of the robot operation with other alternatives, i.e., online and offline programming and other programming frameworks. In this evaluation, users with different levels of training could be requested. Additionally, some stress tests will be applied for assuring the stability of the system.

In future work, we will further investigate how to integrate different skill formalisms into the proposed architecture, especially for the ease of the automatic creation of new skills. The database of skills proposed in the LIAA project is another topic that will be reviewed in order to integrate more skills in the architecture. Additionally, this architecture will be integrated with the reconfigurable and flexible production system under development at ReCaM project. The tools provided by this framework will enable auto-programming and self-adjusting to the required task by utilizing parametric capabilities in the CESA use case.

Regarding the state machine-based architecture, if the proposed approach is used, the industrial processes that can benefit from dual-arm robots are more controlled, and this allows an easier and faster deployment of new applications. In the future, the focus will be set on the coordinated manipulation of the arms with the intention of easing this kind of task. Besides, the integration of a multi-agent system for decision making in coordination and synchronization tasks is being considered.

Author Contributions: All the authors have collaborated in the research: H.H. carried out the development and wrote the paper; H.H., M.P. and D.S. conceived and designed the experiments; J.L.O. and H.H. performed the experiments; D.S., K.L.I. and M.P. supervised the work and improved the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, 21–23 September 2005; Volume 1, pp. 886–893.
2. Duguay, C.R.; Landry, S.; Pasin, F. From mass production to flexible/agile production. *Int. J. Oper. Prod. Manag.* **1997**, *17*, 1183–1195.
3. Hu, S.J. Evolving paradigms of manufacturing: From mass production to mass customization and personalization. *Procedia CIRP* **2013**, *7*, 3–8.
4. Wang, W.; Koren, Y. Scalability planning for reconfigurable manufacturing systems. *J. Manuf. Syst.* **2012**, *31*, 83–91.
5. Tao, F.; Cheng, Y.; Zhang, L.; Nee, A. Advanced manufacturing systems: Socialization characteristics and trends. *J. Intell. Manuf.* **2015**, *28*, 1–16.
6. Haslarn, C. The end of mass production? *Econ. Soc.* **1987**, *16*, 405–439.

7. Smith, C.; Karayiannidis, Y.; Nalpantidis, L.; Gratal, X.; Qi, P.; Dimarogonas, D.V.; Kragic, D. Dual arm manipulation—A survey. *Robot. Auton. Syst.* **2012**, *60*, 1340–1353.

8. Xia, L.; Chen, C.C.; Aggarwal, J.K. Human detection using depth information by kinect. In Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Colorado Springs, CO, USA, 20–25 June 2011; pp. 15–22.

9. Blumrosen, G.; Miron, Y.; Intrator, N.; Plotnik, M. A Real-time kinect signature-based patient home monitoring system. *Sensors* **2016**, *16*, 1965.

10. Sen, S.; Sherrick, G.; Ruiken, D.; Grupen, R.A. Hierarchical Skills and Skill-based Representation. In Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-11), San Francisco, CA, USA, 7–8 August 2011.

11. Thomas, U.; Hirzinger, G.; Rumpe, B.; Schulze, C.; Wortmann, A. A new skill based robot programming language using UML/P Statecharts. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 6–10 May 2013.

12. Zhou, J.; Ding, X.; Qing, Y.Y. Automatic planning and coordinated control for redundant dual-arm space robot system. *Ind. Robot Int. J.* **2011**, *38*, 27–37.

13. Andersen, R.H.; Solund, T.; Hallam, J. Definition and Initial Case-Based Evaluation of Hardware-Independent Robot Skills for Industrial Robotic Co-Workers. In Proceedings of the 41st International Symposium on Robotics (ISR/Robotik 2014), Munich, Germany, 2–3 June 2014.

14. Vanthienen, D.; De Laet, T.; Decré, W.; Smits, R.; Klotzbücher, M.; Buys, K.; Bellens, S.; Gherardi, L.; Bruyninckx, H.; De Schutter, J. iTaSC as a unified framework for task specification, control, and coordination, demonstrated on the PR2. In Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011.

15. Poppa, F.; Zimmer, U. RobotUI-A software architecture for modular robotics user interface frameworks. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura-Algarve, Portugal, 7–11 October 2012.

16. Björkelund, A.; Bruyninckx, H.; Malec, J.; Nilsson, K.; Nugues, P. Knowledge for intelligent industrial robots. In Proceedings of the AAAI Spring Symposium: Designing Intelligent Robots, Stanford, CA, USA, 26–28 March 2012.

17. Huckaby, J.; Vassos, S.; Christensen, H.I. Planning with a task modeling framework in manufacturing robotics. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013.

18. Stenmark, M.; Malec, J. A helping hand: Industrial robotics, knowledge and user-oriented services. In Proceedings of the 2013 IEEE/RSJ International Conferenceon Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013.

19. Alonso, D.; Vicente-Chicote, C.; Pastor, J.A.; Alvarez, B. Stateml : From graphical state machine models to thread-safe ada code. In *Reliable Software Technologies—Ada-Europe 2008*; Springer: Berlin, Germany, 2008; pp. 158–170.

20. Armentia, A.; Gangoiti, U.; Priego, R.; Estévez, E.; Marcos, M. Flexibility support for homecare applications based on models and multi-agent technology. *Sensors* **2015**, *15*, 31939–31964.

21. Klotzbuecher, M. rFSM. Available online: https://github.com/orocos/rFSM/tree/master/doc (accessed on 1 June 2016).

22. Bohren, J. Package Summary. Available online: http://wiki.ros.org/smach. (accessed on 1 June 2016).

23. Quigley, M.; Conley, K.; Gerkey, B.P.; Faust, J.; Foote, T.; Leibs, J.; Wheeler, R.; Ng, A.Y. ROS: An open-source Robot Operating System. Available online: http://www.willowgarage.com/sites/default/files/icraoss09-ROS.pdf (accessed on 24 May 2017).

24. ROS. Available online: http://www.ros.org/ (accessed on 1 June 2016).

25. ROS. Core Components. Available online: http://www.ros.org/core-components/ (accessed on 1 June 2016).

26. ROS. Robots. Available online: http://wiki.ros.org/Robots (accessed on 1 January 2017).

27. Badawy, R.; Yassine, A.; Heßler, A.; Hirsch, B.; Albayrak, S. A novel multi-agent system utilizing quantum-inspired evolution for demand side management in the future smart grid. *Integr. Comput.-Aided Eng.* **2013**, *20*, 127–141.

28. Pinto, T.; Praca, I.; Vale, Z.; Morais, H.; Sousa, T.M. Strategic bidding in electricity markets: An agent-based simulator with game theory for scenario analysis. *Integr. Comput.-Aided Eng.* **2013**, *20*, 335–346.

29. OpenRTM. Available online: http://openrtm.org/ (accessed on 1 June 2016).
30. Fast Research Interface Library. Available online: http://cs.stanford.edu/people/tkr/fri/html/ (accessed on 1 June 2016).
31. Tecnalia. Available online: http://www.tecnalia.com/en/ (accessed on 1 February 2017).
32. LIAA. Available online: http://www.project-leanautomation.eu/ (accessed on 1 June 2016).
33. ReCaM. Available online: http://recam-project.eu/ (accessed on 1 February 2017).
34. DGH. Available online: http://www.grupodgh.es/en/ (accessed on 1 February 2017).
35. DGH. Available online: http://www.cesa.aero/en/ (accessed on 1 February 2017).
36. Herrero, H.; Outón, J.L.; Esnaola, U.; Sallé, D.; de Ipiña, K.L. State machine based architecture to increase flexibility of dual-arm robot programming. In *Bioinspired Computation in Artificial Systems*; Springer: Berlin, Germany, 2015; pp. 98–106.
37. Herrero, H.; Esnaola, U.; Sallé, D. TECNALIA HIRO Performing Aeronautics Assembly–Deburring and riveting–Showcased at BIEMH2014. Available online: https://www.youtube.com/watch?v=pvxlqyJtPNo (accessed on 1 April 2017).
38. Järvenpää, E.; Siltala, N.; Lanz, M. Formal resource and capability descriptions supporting rapid reconfiguration of assembly systems. In Proceedings of the 12th Conference on Automation Science and Engineering, and International Symposium on Assembly and Manufacturing, Fort Worth, TX, USA, 21–22 August 2016.
39. Järvenpää, E.; Siltala, N.; Hylli, O.; Lanz, M. Capability matchmaking procedure to support rapid configuration and re-configuration of production systems. 2017. Unpublished.
40. Babar, M.A.; Zhu, L.; Jeffery, R. A framework for classifying and comparing software architecture evaluation methods. In Proceedings of the Software Engineering Conference, Melbourne, Australia, 13–16 June 2004.
41. Dobrica, L.; Niemela, E. A survey on software architecture analysis methods. *IEEE Trans. Softw. Eng.* **2002**, *28*, 638–653.
42. Ionita, M.T.; Hammer, D.K.; Obbink, H. Scenario-based software architecture evaluation methods: An overview. In Proceedings of the International Conference on Software Engineering (ICSE/SARA) Orlando, FL, USA, 19–25 May 2002.
43. Kazman, R.; Klein, M.; Clements, P. *ATAM: Method for Architecture Evaluation*; Technical Report, DTIC Document; Software Engineering Institute: Pittsburgh, PA, USA, 2000.
44. Gonzalez-Huerta, J.; Insfran, E.; Abrahão, S. Models in software architecture derivation and evaluation: Challenges and opportunities. In Proceedings of the International Conference on Model-Driven Engineering and Software Development, Lisbon, Portugal, 7–9 January 2014.
45. Babar, M.A.; Gorton, I. Comparison of scenario-based software architecture evaluation methods. In Proceedings of the 11th Asia-Pacific Software Engineering Conference, Busan, South Korea, 30 November–3 December 2004.
46. Cheung, L.; Roshandel, R.; Medvidovic, N.; Golubchik, L. Early prediction of software component reliability. In Proceedings of the 30th International Conference on Software Engineering, Leipzig, Germany, 10–18 May 2008.
47. Gonzalez-Huerta, J.; Insfran, E.; Abrahão, S.; Scanniello, G. Validating a model-driven software architecture evaluation and improvement method: A family of experiments. *Inf. Softw. Technol.* **2015**, *57*, 405–429.
48. Kazman, R.; Klein, M.; Clements, P. *Evaluating Software Architectures-Methods and Case Studies*; Addison-Wesley Professional: Boston, MA, USA, 2001.
49. Ringert, J.O.; Rumpe, B.; Wortmann, A. A Case Study on Model-Based Development of Robotic Systems using MontiArc with Embedded Automata. *arXiv* **2014**, arXiv:1408.5692.
50. Giorgini, P.; Kolp, M.; Mylopoulos, J. Multi-agent and software architectures: A comparative case study. In Proceedings of the International Workshop on Agent-Oriented Software Engineering, Bologna, Italy, 15 July 2002.
51. Bravo, J.; Villarreal, V.; Hervás, R.; Urzaiz, G. Using a communication model to collect measurement data through mobile devices. *Sensors* **2012**, *12*, 9253–9272.
52. Aman, W.; Snekkenes, E. EDAS: An evaluation prototype for autonomic event-driven adaptive security in the internet of things. *Future Internet* **2015**, *7*, 225–256.
53. Biggs, G.; MacDonald, B. A survey of robot programming systems. In Proceedings of the Australasian Conference on Robotics And Automation, Brisbane, Australia, 1–3 December 2003.

54. Pan, Z.; Polden, J.; Larkin, N.; Van Duin, S.; Norrish, J. Recent progress on programming methods for industrial robots. *Robot. Comput.-Integr. Manuf.* **2012**, *28*, 87–94.

55. Herrero, H.; Outon, J.L.; Esnaola, U.; Salle, D.; Lopez de Ipina, K. Development and evaluation of a Skill Based Architecture for applied industrial robotics. In Proceedings of the 2015 4th International Work Conference on Bioinspired Intelligence (IWOBI), San Sebastian, Spain, 10–12 June 2015.

56. Herrero, H.; García, F.; Esnaola, U.; Sallé, D. 2015 TECNALIA NextageOpen—Dual-Arm Robot for Aeronautics Pilot Station. Available online: https://www.youtube.com/watch?v=x-eJ66jM1Rk (accessed on 1 April 2017).

57. Herrero, H.; Moughlbay, A.A.; Outón, J.L.; Sallé, D.; de Ipiña, K.L. Skill based robot programming: Assembly, vision and Workspace Monitoring skill interaction. *Neurocomputing* **2017**, doi:10.1016/j.neucom.2016.09.133.

58. Herrero, H.; Pacheco, R.; Alberdi, N.; Rumayor, M.; Salle, D.; Lopez de Ipiña, K. Skills for vision-based applications in robotics application to aeronautics assembly pilot station. In Proceedings of the 2015-International Conference on Computer as a Tool (EUROCON), Salamanca, Spain, 8–11 September 2015.

59. Wang, T.; Gao, H.; Qiu, J. A Combined Fault-Tolerant and Predictive Control for Network-Based Industrial Processes. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2529–2536.

60. Wang, T.; Zhang, Y.; Qiu, J.; Gao, H. Adaptive fuzzy backstepping control for a class of nonlinear systems with sampled and delayed measurements. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 302–312.

61. Wang, T.; Qiu, J.; Gao, H.; Wang, C. Network-Based Fuzzy Control for Nonlinear Industrial Processes With Predictive Compensation Strategy. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, doi:10.1109/TSMC.2016.2616904.

62. TOPREF. Available online: http://toprefproject.eu/ (accessed on 1 April 2017).

63. ROS. Available online: http://rosindustrial.org/about/description/ (accessed on 1 January 2017).

64. ROS. Supported Hardware. Available online: http://wiki.ros.org/Industrial/supported_hardware (accessed on 1 January 2017).

Torsional Ultrasound Sensor Optimization for Soft Tissue Characterization

Juan Melchor *, Rafael Muñoz and Guillermo Rus

Department of Structural Mechanics, University of Granada; ETS Ingenieros de Caminos, Severo Ochoa s/n, 18071 Granada, Spain; rmb@ugr.es (R.M.); grus@ugr.es (G.R.)

* Correspondence: jmelchor@ugr.es; Tel.: +34-958-240-037

Abstract: Torsion mechanical waves have the capability to characterize shear stiffness moduli of soft tissue. Under this hypothesis, a computational methodology is proposed to design and optimize a piezoelectrics-based transmitter and receiver to generate and measure the response of torsional ultrasonic waves. The procedure employed is divided into two steps: (i) a finite element method (FEM) is developed to obtain a transmitted and received waveform as well as a resonance frequency of a previous geometry validated with a semi-analytical simplified model and (ii) a probabilistic optimality criteria of the design based on inverse problem from the estimation of robust probability of detection (RPOD) to maximize the detection of the pathology defined in terms of changes of shear stiffness. This study collects different options of design in two separated models, in transmission and contact, respectively. The main contribution of this work describes a framework to establish such as forward, inverse and optimization procedures to choose a set of appropriate parameters of a transducer. This methodological framework may be generalizable for other different applications.

Keywords: torsional ultrasound; probability of detection; soft tissue mechanics; finite element method; optimization; inverse problem

1. Introduction

Torsional ultrasonic waves can be used to characterize the biomechanics of soft tissue. The design process of a torsional ultrasonic sensor has as a key point the use of piezoelectric materials [1,2]. Piezoelectric materials (PZT) are an inorganic compound that changes shape when an electric field is induced. They are actually used in the design of sensors in several applications, such as music microphones and instruments, precision positioning, cancellation of noise, motors, and ultrasonic devices in our study.

This design problem provides two steps, the understanding of propagation of torsional waves and the optimization of a transducer with the capacity to transmit and receive them. Additionally, some applications require large displacements, such as improvement of micropositioning in robotics, CD drivers, or control of trailing edge flaps of helicopters. For the rest, these transducers are able to execute large circumferential displacements, it is necessary for smart structures, and they can be adapted for a range of applications on a wide variety of requirements [3,4].

Several optimization criteria are available for ultrasonic transducers depending on specific parameters being considered. For example, the use of central frequency and insertion loss (a factor associated with frequencies that disappear through a filter). Furthermore, new variables as the case of central frequency or the energy have been shown to be the most appropriate. Other new features have been introduced that describe the waveform and amplitude spectra improving the inputs of the optimization formulation, widening the range of applications as it is shown in this work [5,6].

The combined use of simulation based on finite elements analysis (FEA) and optimization methods leads to a suitable way of designing ultrasonic sensors. This process needs use of a global

cost function that measures the maximization/minimization criterion and is calculated using the model. Several design alternatives can be obtained using different parameters to optimize the cost function [7,8].

The use of FEA together with the multi-objective decision facilitates an efficient design where conflicts among several criteria such as acoustic effects, impedance, vibration of the sensors are optimized, in addition to the design time [9,10].

Recently, sensors which can be used for similar applications have been designed. Bolt-Langevin type transducers (BLT) are an example, composed of piezoelectric disks with a pair of elastic bodies mounted at the ends which generate mechanical oscillations with a specific design frequency. They mainly use longitudinal waves in the thickness direction. Additionally, there are some works taking into account the torsional vibrations generated by the elastic part [11–14]. The main application of this type of sensors was the generation of uniform micro-droplets [15], as well as in acoustic levitation (NFAL), in non-destructive testing, and biomedical engineering [16,17].

Numerical analysis may be useful to consider other designs, setting the tolerance of each parameter related to sensor efficiency within a particular range [18,19].

Torsional waves has also been used as guided waves in nondestructive testing of pipes [20,21], in liquids inserting a bar in the fluid where waves are propagated to measure density [22,23], liquid level [24], temperature [24], or viscosity [23].

One of the main applications of ultrasonic transducers in medicine is elastography with ultrasonic longitudinal waves, where there exists a special interest in efficiency enhancements of sensor desings [25–28]. Recently, a new field of investigation has been opened introducing the concept of transversal waves to measure the shear modulus of soft tissue, due to the importance of correlating it with many disorders [29,30].

Accordingly, the motivation of this study began with the aim at designing an optimized ultrasonic sensor focus on clinical applications to obtain high levels of sensitivity in mechanical identification of soft tissue. A part of the computational design of the transducer was described in detail [31]. In parallel to this, since the first design study, some reference works have emerged as another shear ultrasonic sensor based on induced resonances [32] providing a new screening method for breast cancer from wave viscoelasticity imaging [33]. The last approaches in this line of research have improved developing the generation of remote adaptive torsional shear waves (ATSW) using an octagonal phased array to improve displacements and reducing the dispersion of shear wave speeds. The ATSW method also reveal that it is possible to estimate the viscoelasticity of biological tissues when small biases in lesion appear [34,35]. Torsional waves has also been used to characterized mechanical properties of tissues like liver [36] or vocal fold [37,38].

The transducer is going to be adapted at facing preterm birth assessment in direct contact with cervix, since the survival and morbidity carries a burden in the health care system [39]. The prevalence of preterm birth is between 8.1–12.7% of all births in European countries [40].

The concept of shear waves to evaluate cervix stiffness to birth prediction has been considered with another physical principle, Supersonic Shear Imaging (SSI), without separation between P and S waves and high energy levels, since shear waves are generated using the nonlinear acoustic radiation force [29,41]. On the other hand, direct contact can be avoided.

This paper describes the complete view of the methodology that has been used to design the torsinal wave transducer for soft tissue evaluation in direct contact, dividing it in two phases. The first one studies only the torsional propagation in tissue and how the excitation conditions are optimized using an inverse problem. It is based on a finite element model (FEM) of the tissue by FEAP open software. The second phase optimize the transducer design adding the FEM of the transducer and using a inverse problem that maximize, what the authors call the robust probability of detection (RPOD) [31], as a measure of the sensibility of the transducer. The probability of detection (POD) [42] of the transducer as a function of a model parameter measures the probability of the transducer to detect a specific small change in that parameter. The RPOD concept is introduced to deal with more

than one parameter, conservatively selecting the worse POD of the parameters as the representative POD of the transducer. This definition of RPOD will be used in an inversion procedure to optimize the parameters of the transducer that maximize the RPOD.

2. Methods

The optimization of the proposed design of the torsional wave transducer has been undertaken in two phases:

- Phase 1: Modelization of the torsional wave propagation in tissue, including a sensitivity test on the parameters of the excitation wave.
- Phase 2: Joint modelization of tissue-transducer interaction. Optimization of the transducer under maximum POD (Probability Of Detection) criterion.

The purpose of phase 1 is to test the capacity of the torsional wave model and the inverse method to identify changes in mechanical properties of a multilayered tissue, independently on the transducer that is used. Additionally, the POD estimator is checked as the tool to measure sensitivity. Phase 2 incorporates the modelization of the proposed transducer design and uses the POD estimator to optimize the transducer design maximizing sensitivity measure. Both phases use a similar (not identical) inverse problem design. The common parts of the inverse problem and POD are firstly presented, and afterwards the phases.

In this work, POD is the algorithm to measure sensitivity and may be defined as the probability of the transducer and the reconstruction algorithm to detect changes on the mechanical properties of the tissue, given the presence of uncertainties, mainly, signal noise. A multilayered tissue will be supposed, four layers in phase 1 (Figure 1) just to test the multilayer capacity, and two layers in phase 2 (dermic and connective, Figure 2). The mechanical parameters are Young Modulus and Reynolds attenuation in phase 1 for each layer, and shear moduli for both layers of phase 2.
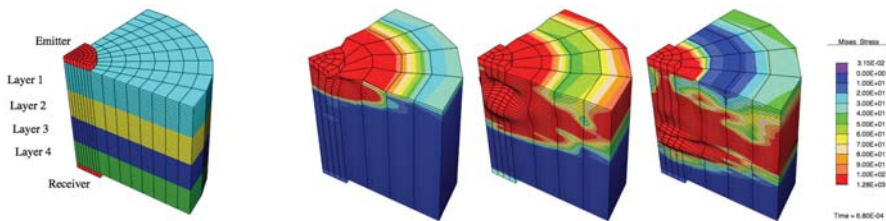
Figure 1. Left: geometry, layers, transducers and mesh. Following pictures: three instants of torsional propagation *t*= 280, 460 and 680 μs, using a simpler mesh.
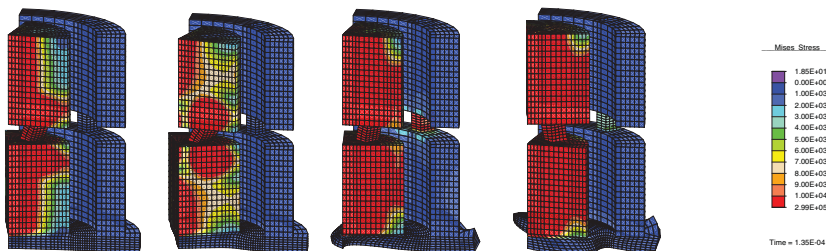
Figure 2. Torsional transducer at instants *t* = 9, 18, 117, 135 μs. Tissue is the soil.

2.1. Robust Probability of Detection

A plausible pathology may be inferred when enough changes in the mechanical properties of the tissue layers are detected. Let **P** be a specific parametrization for POD study, consisting of a vector with the mechanical properties under study. For instance for phase 2 shear moduli *G*, $\mathbf{P} = (G_c \; G_d)$ (*c* connective and *d* dermic). A pathological change implies that the healthy state presents values of reference *r* on the parameters, and pathology would exist in case that enough variation is observed $\mathbf{p} = (\Delta G_c \; \Delta G_d)$. If P_k is a generic element of the array **P**, $p_k = \Delta P_k = P_k - P_k^r$, being N_p the size of the array.

The mathematical definition of the probability of detection (POD) is presented as the probability that the modifications on the received signal (SIGNAL) given the emergence of pathology surpasses the signal noise level (NOISE),

$$\text{POD} = P\left(\frac{|\,\text{SIGNAL}\,|^{\,2}}{|\,\text{NOISE}\,|^{\,2}} > 1 \right) \tag{1}$$

The variables that determines the POD estimation for a specific location are the severity of the pathology (**p**) and the noise level σ. The following deduction will be shown for one single parameter p_k, considering that this calculation will have to be repeated N_p times, one for each parameter of **p**.

As signal and noise changes depend on several parameters p_k, a robust POD (RPOD), is defined so that the RPOD is the POD of the less favorable case, the POD of the parameter with worst POD.

$$\text{RPOD} = \min_k \text{POD}(p_k) \qquad\qquad \text{POD}(p_k) = P\left(\frac{|\text{SIGNAL}(p_k)|^2}{|\text{NOISE}(p_k)|^2} > 1 \right) \tag{2}$$

The linear nature of the physical models that will be used, assumes a linear relation between measurements and mechanical properties. The measurement dependency with noise will also be suppose to be linear. Under these two assumptions of linearity, the measurements may be developed as a Taylor expansion of first order, centered at the healthy case and the absence of noise.

$$\psi_i(p_k, \sigma) = \psi_i(0,0) + \underbrace{p_k \frac{\partial \psi_i}{\partial p_k}(0,0)}_{\text{SIGNAL}} + \underbrace{\sigma \frac{\partial \psi_i}{\partial \sigma}(0,0)}_{\text{NOISE}} + hot \tag{3}$$

being $i = 1, ..., N$ the receiving points. $\psi_i(0,0)$ are the measurements in the absence of pathology and noise at point *i*, the second term the linear variation of the measurement given the only presence of the pathology, which has been labeled as SIGNAL, and the third term the linear variation due to the presence of noise and healthy conditions, labeled NOISE.

The second term of the sum may be approximately expressed as a finite difference, but introducing a very small degradation $p_{k0} \to 0$ instead of a pure healthy state to assure that the FEM computational implementation of the models captures the perturbations produced at small p_k, and compute $\psi_{i,p_k}(p_{k0}, 0) \approx \psi_{i,p_k}(0,0)$,

$$\frac{\partial \psi_i}{\partial p_k}(p_{k0}, 0) = \psi_{i,p_k}(p_{k0}, 0) = \frac{\psi_i(p_{k0} + \Delta p_k, 0) - \psi_i(p_{k0}, 0)}{\Delta p_k}$$

For the RPOD calculation, the FEM algorithm must be run $N_p + 1$ times. The first one, to calculate the signal applying the mechanical properties of reference (representing no degradation), $\psi_i(p_{k0}, 0)$, and N_p additional signals altering each parameter of the parametrization, $\psi_{i,p_k}(p_{k0} + \Delta p_k, 0)$.

The noise term of the *Taylor* series on Equation (3) is proposed to be,

$$\frac{\partial \psi_i}{\partial \sigma} = \xi_i \text{RMS}(\psi_i^{\text{FEM}}) = \xi_i \text{RMS} \qquad\qquad \text{RMS}(g) = \sqrt{\frac{1}{M} \sum_{j=0}^{M-1} g(t_j)^2} \tag{4}$$

with ξ_i a random variable (noise generator) applied over the root mean square (RMS) of the simulated signal. (RMS) defined for any sampled function g in time domain $g(t_j)$ at M samples.

Equations (3) and (4), and the relationship $|Y_i|^2 = \frac{1}{m} \sum_{i=1}^{m} Y_i^2$, can be applied on Equation (1),

$$\text{POD} = P\left(\frac{p_k^2 \frac{1}{N} \sum_{i=1}^{N} (\psi_{i,p_k}(0,0))^2}{\sigma^2 \text{RMS}^2 \frac{1}{N} \sum_{i=1}^{N} \xi_i^2} > 1\right) = P\left(p_k^2 > \frac{\text{RMS}^2 \sigma^2 \sum_{i=1}^{N} \xi_i^2}{S_k}\right) \tag{5}$$

being,

$$S_k = \sum_{i=1}^{N} (\psi_{i,p_k}(0,0))^2 \tag{6}$$

The POD may be reinterpreted in terms of parameter noise, so that it is the probability of a specific stochastic mechanical property p_k^2, having the following cumulative probability density function,

$$\text{POD} = F\left(\frac{\text{RMS}^2 \sigma^2 \sum_{i=1}^{N} \xi_i^2}{S_k}\right) \tag{7}$$

which represents the noise distribution function in terms of the parameter p_k (in this case, a mechanical property of the material). That is, how the noise distribution function of the signal is translated through the direct model into a random variation of the parameter values. In order to detect a variation of the mechanical properties originated by a pathological zone, this variation must exceed this parameter-noise level, a fact that defines the POD.

Following techniques of the theory of error propagation and Monte Carlo sampling, the noise of the signals at receiving locations can be assumed to behave as a normal distribution [42], and consequently, the squared sum of the noise ξ_i follows a *chi-square* distribution, given that $\sum_{i=1}^{N} \xi_i^2 \longrightarrow \chi_N^2$ (e.g., [43]). Its parameter is the number N of degrees of freedom, corresponding to the number of receiving locations. This *Chi-square* distribution may be approximately assumed as a normal $\mathcal{N}$ distribution in the case of $N > 10$, with $N - 2/3$ the mean and $\sqrt{2N}$ the standard deviation, $\chi^2(N) \approx \mathcal{N}(N - 2/3, \sqrt{2N})$. In (7) it leads to,

$$p_k^2 \longrightarrow \mathcal{N}\left[\frac{\text{RMS}^2 \sigma^2 (N - 2/3)}{S_k}, \frac{\text{RMS}^2 \sigma^2 \sqrt{2N}}{S_k}\right] \tag{8}$$

expression that is only valid for noise with Gaussian distribution.

As this probability density function f posses a cumulative probability $F(x) = \int_{-\infty}^{x} f(y) dy$, and being its inverse $x = G(F(x))$, the pathology to noise ratio p_k/σ can be deduced from (8) for a fixed and desired POD value as,

$$\frac{p_k}{\sigma} = \sqrt{\frac{\text{RMS}^2 (N - 2/3)}{S_k}\left(1 + G[\text{POD}] \frac{\sqrt{2N}}{N - 2/3}\right)} \tag{9}$$

This is the final formula, where a specific value of POD can be fixed, to obtain the relationship of the alteration of the mechanical property p_k against noise, around the mechanical property's value of reference.

2.2. Inverse Problem

Two different IP approaches has been used. In phase 1 to adjust the mechanical properties of the tissue that minimize the difference between simulated and experimental signals (misfit function), and in phase 2 to optimize the transducer parameters that maximize the RPOD.

Although the target is different, both approaches present similarities. On one hand, Genetic Algorithms (GA) are used, mainly because they are global search algorithms. It implies that the search

for the optimal parameters is made by reasonably exploring the whole parameter space, so that falling in local maxima or minima is avoided (see [42,44] for more detailed information on the algoritm that is used). On the other hand, the cost functional will serve the optimization criterion. If f is the function to optimize (misfit in phase 1 and RPOD in phase 2), a modified version f^l has been reported to improve convergence of the GA. [45],

$$f^l = \log{(f + \epsilon)} \qquad (10)$$

being ϵ a very small value ensuring the logarithm's existence.

2.3. Phase 1: Modelization of Torsional Wave Propagation in Tissue

Phase 1 aims at checking the suitability of the propagation model and the inverse problem for optimization purposes with torsional waves. Additionally, the POD estimator is implemented and tested. Table 1 shows the steps followed in this phase.

In order to show the capability of torsional waves to characterize changes in mechanical properties of soft tissue, a direct contact emitter-receiver configuration on a layered tissue with parallel interfaces is hypothesized (see Figure 1). Torsional waves are generated and propagated through the normal direction to the layer interfaces and are collected on the opposite face of the tissue. Each layer is assumed to be isotropic. A total tissue of 40 mm with four layers of the same thickness will be assumed in this study.

Several important points are remarked as a summary,

- A linear elastic, attenuating and multilayered physical model, solved by finite elements, was used to simulate the torsional wave propagation.
- An inverse problem is proposed to characterize the mechanical properties of the tissue and detect pathology.
- The inverse problem is applied to several sets of excitation parameters (geometry and emitted waveforms), to see the capability of the method to select that with best detection.
- The use of the semi-analytic POD estimator with the selected excitation, to measure its capability of detection.
- The calculation process is outlined in Figure 3.

Step 1 defines the problem and the four sets of excitation parameters. The adopted ranges of variation of the parameters are a frequency (f) range of $(2, 20)$ kHz, bandwidth (b) of $(0.5f, f)$ and a transducer radius (r) range of $(10, 20)$ mm.

Within the mentioned ranges, four arbitrary candidates are shown in Table 2.
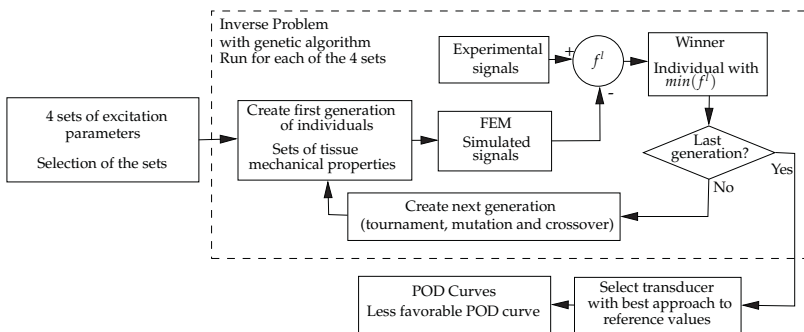
Figure 3. Calculation method for phase 1.

Table 1. Methodology of phase 1.

Step	Target	Outcome	Tools or Inputs
1. Problem configuration	Problem geometry, boundary conditions, configuration, sets of excitation parameters	Problem definition. 4 sets of excitation parameters to test	
2. Forward physical model selection	Simulate propagation with the excitation parameters	Physical model: differential equations	Outcome from step 1: geometry, materials, etc.
3. Finite element model (FEM)	Computational implementation of the physical model	Computational code of the model	Outcomes from steps 1 and 2.
4. Discretization study and FEM test	Convergence study Balance computational burden-time Assure accuracy of simulations	Spatial element Time interval FEM checked	Forward model (FEM) Geometrical parameters and tissue mechanical properties L-S wave speed check
5. Inverse problem with genetic algorithm (GA)	Design and implementation Evaluate GA's convergence - Quality of identification - Convergence speed Apply IP to 4 excitation parameter sets	Observed right behaviour and identification on the mechanical properties of the tissue Winner: excitation parameter set with best identification	FEM Cost functional for GA optimization: misfit function Synthetic signals
6. POD evaluation of the winner	POD estimator evaluation Checking coherency on results	Graphics of POD on modifications of mechanical parameters of tissue No perceived anomalies	Forward model Winner parameter set POD estimator

Table 2. Sets of excitation parameters under test.

Transducer	f (kHz)	b (kHz)	r (mm)
Design 1	6.32	6.32	10
Design 2	20	20	10
Design 3	2	1	10
Design 4	6.32	3.16	20

Following with steps 2 and 3, the torsional propagation process is described using a 3D physical model consisting of standard linear elastic equations with Rayleigh attenuation. Equilibrium, constitutive and kinematic equations are,

$$\sigma_{ji,j} + b_i = \rho \ddot{u}_i + R\rho \dot{u}_i \qquad \begin{cases} \sigma_{ij} = c_{ijkl}\varepsilon_{kl} \\ \varepsilon_{kl} = \frac{1}{2}(u_{k,l} + u_{l,k}) \end{cases} \tag{11}$$

being σ_{ij} the stress tensor , ε_{ij} the strain tensor, u_i the displacement vector, b_i volumetric forces, ρ material's density, R the Rayleigh damping coefficient, c_{ijkl} the fourth-order stiffness tensor of material properties.

These equations are applied over every layer of the tissue considering continuity on displacement and stress through the interfaces. Every layer is supposed homogeneous and isotropic. In the case hereby treated, four layers of soft tissue with different mechanical properties are assumed for algorithmic testing purposes. The first and last layers are assumed to be the same tissue features, as it use to happen in many experiences, for instance, with epithelial tissue in emission and reception. In such conditions, the parameters **P** of the model are the mechanical properties of every layer, in this case, their Young moduli E and Rayleigh attenuation R, which can be presented in the vector $\mathbf{P} = (E_4 R_4 E_3 R_3 E_2 R_2 E_1 R_1)$, being in this case $N_p = 8$, the number of parameters. But a simpler parametrization will be used to improve IP convergence using $N_p = 4$ parameters, $\mathbf{p} = (E_3 R_3 E_2 R_2)$, assuming the rest of the parameters constant, as indicated in Table 3.

Table 3. Parameters of the layered soft tissue. Poisson's ratio depends on E.

Material	Young Modulus E (MPa)	Poisson Ratio ν	Density ρ (kg/m³)	Thickness a (mm)	Attenuation R ($\frac{\text{dB}}{\text{MHz}\cdot\text{mm}}$)
Layer 1	20	0.48	1070	10	1836.07
Layer 2	$E_2 \in [2 \sim 15]$	ν_2	920	10	$R_2 \in [1 \sim 5]$
Layer 3	$E_3 \in [15 \sim 50]$	ν_3	1070	10	$R_3 \in [1 \sim 5]$
Layer 4	20	0.48	1070	10	1836.07

The forward problem is performed using a 3D FEM model implemented in FEAP package [46]. Solid hexaedric elements with 8 nodes are used in 26 blocks to generate a structured mesh (Figure 1), so that remeshing perturbation is avoided on sensitivity analysis. The material's diameter being simulated is 100 mm. Other FEM parameters: total time 3000 µs, incremental time 20 µs and mesh extension (radius) 50 mm.

Summarizing, the model system is defined by the specimen materials (steel and four layers of tissue, Table 3), geometry and the boundary conditions: at the cylindrical axis, r = 0 mm (being r the radial direction), null displacements at the points; energy absorbing face at the external face at r = 100 mm to avoid reflections; and free displacements at the rest of the mesh points. The input is introduced as an applied displacement with the shape of a spike pulse.

Two criteria have been tested to verify the validity of the FEM model. First, a mesh convergence test has been carried out, introducing a trade-off between computational cost and numerical precision. Second, a verification of compressional (p) and shear (s) wave speeds have been compared to the time of arrival of the first wavefront.

In the second criteria, the speed of compressional c_p and shear c_s waves are estimated, in relationship to the bulk modulus K of all layers (and consequently to Young's modulus E) by,

$$K = \frac{E}{3(1 - 2\nu)} = \rho \left(c_p^2 - \frac{4}{3} c_s^2 \right) \tag{12}$$

$$c_p = \sqrt{\frac{E(1 - \nu)}{\rho(1 + \nu)(1 - 2\nu)}} = \sqrt{\frac{K + \frac{4}{3}G}{\rho}} = \sqrt{\frac{M}{\rho}} \qquad c_s = \sqrt{\frac{G}{\rho}} \tag{13}$$

In particular, for $\rho = 1070$ kg/m³, $E = 20$ MPa and $\nu = 0.3$, $t_p = 0.6$ µs and $t_s = 1.1$ µs is the time used for compression and shear waves to propagate with speeds $c_s = 0.05 \sim 0.5$ km/s and $c_p = 0.94 \sim 1$ km/s. For $\nu = 0.49$, the corresponding times are $t_p = 0.25$ µs and $t_s = 1.16$ µs. If $\rho = 1000$ kg/m³ and $c_p = 1.5$ km/s, $M = 2.25$ GPa and estimation of total time is $t = 1.46$ µs for maximum time of propagation of waves across the soft tissue.

For the inverse problem calculation, shown in Figure 3, let ψ^0 be the samples of the received signals in the case of healthy tissue, ψ^x the samples of the received signal of the test with the soft tissue on its unknown current state, and ψ the samples of the simulated signal on reception. The discrepancy against the healthy state may initially be described by,

$$\Phi = \frac{\psi - \psi^0}{\text{RMS}(\psi^0)} \tag{14}$$

The misfit or discrepancy between simulated and experimental signals is defined as the residual $\gamma = \Phi^x - \Phi$. The definition of the fitness function f is generated as a quadratic form from the residual vector γ of size M,

$$f = \frac{1}{2}|\gamma|^2 = \frac{1}{2}\frac{1}{M}\sum_{j=1}^{M}\gamma_j^2 \tag{15}$$

Synthetic signals are created as experimental signals. Four received signals are calculated using the FEM model, one for each of the four set of excitation parameters. In order to transform these

signals, avoiding exact matching between experiment and simulation, random Gaussian white noise is added (same function for all signals).

Finally, for POD estimation of the best excitation conditions, four curves will be obtained, one for each parameter of the vector **p**. The less favorable is considered.

2.4. Phase 2: Tissue-Transducer Modelization, Optimization with RPOD

The conceptual transducer design to be optimized is shown in Figures 2 and 4. A central inner disk with four piezoelectric emitters electrically joined, create a circumferential shear displacement of the disk, which is transmitted to a thin double-layered tissue (dermic and connective, Figure 2, bottom area), and from it to the external circular crown where four piezoelectric elements captures the shear movement generating one signal. The time of flight is measured to infer shear moduli of the tissue layers.

The shear moduli of the layers $\mathbf{P} = (G_c \, G_d)$ (c connective, d dermic) vary when pathologic changes occur, $p = (\Delta G_c \, \Delta G_d)$, $\Delta G_k = G_k - G_k^r$, $k = c, d$, and this change must be greater than noise level.
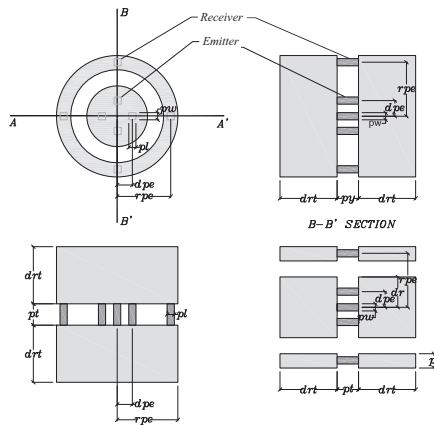
Figure 4. Transducer geometry. Piezoelectric elements in dark gray. Emitters at the inner disk.

Table 4 summarizes the methodology of this phase.

The forward model (Step 1) includes the coupled electrical-mechanical piezoelectric effect, under the following constitutive equations,

$$\mathbf{T} = \mathbf{C}_E \cdot \mathbf{S} + \mathbf{e}^T \cdot \mathbf{E} \tag{16}$$

$$\mathbf{D} = \mathbf{e} \cdot \mathbf{S} - \varepsilon_\mathbf{S} \cdot \mathbf{E} \tag{17}$$

being **T** the stress tensor, **S** strain tensor, **E** electric field, **D** charge-density displacement, $\mathbf{C}_E$ piezoelectric stiffness matrix, **e** piezoelectric coupling coefficient matrix, $\mathbf{e}^T$ its transpose and ε_S permittivity coefficient matrix.

Equilibrium and kinematic relationships are as follows,

$$\nabla \cdot \mathbf{D} = 0; \quad \mathbf{E} = -\nabla \phi \tag{18}$$

$$\nabla^S \cdot \mathbf{T} = 0; \quad \mathbf{S} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T) \tag{19}$$

with ϕ the electric potential and $\mathbf{u} = (u_1 u_2 u_3)$ the vector field of displacements.

<div align="center">**Table 4.** Methodology of phase 2.</div>

Step	Target	Results	Tools or Inputs
1. Selection of the forward model	Simulate generation, propagation, and reception of waves	Physical model: differential equations	Initial transducer design: geometry, materials, etc.
2. Finite element model (FEM)	Computational implementation of the physical model	Computational code of the model	Physical model Boundary conditions(geometry and properties of conceptual design, etc.)
3. Discretization analysis	Convergence enhancement Balance computational burden-time	Spatial element Time interval	FEM Initial geometry and mechanical properties
4. Validation of FEM results	Check accuracy on simulations	Test discrepancies using simplified models	Approximated analytic model of torsional waves for comparison [31].
5. Sensitivity test on mechanical properties of tissue and parameters of the transducer geometry [31]	To know the response to transducer geometry and tissue propagation Select materials Reduce P-wave generation and propagation Maximize amplitude of received torsional wave New check of the FEM	Material influence Select a transducer geometry among tested vanishing P-wave Ranges of mechanical properties in tissue with low P-wave propagation Ranges of P-wave existence	FEM Materials to test Variation ranges of the transducer parameters (geometry)
6. Inverse problem as a new test of the forward problem	New check of the FEM Quality of identification of tissue mechanical properties	Valid identification of mechanical properties (G_c G_d) of the tissue (Figure 5a)	FEM Cost function for GA optimization: Misfit function
7. Evaluation of RPOD for best sensor in step 5 [31]	Test RPOD estimator Evaluate POD for three values of tissue shear modulus of reference (0.3, 3 and 30 kPa) Testing coherency on results	POD graphics (Figure 6) on a range of values of (G_c^r, G_d^r) Select the initial parameters for optimization No apparent anomalies	FEM POD estimator
8. Transducer optimization with best RPOD criterion	Find transducer design (geometry) with maximum RPOD Evaluate improvement in RPOD	Select the optimal geometric design with the worst response to mechanical properties Quantification of improvement	FEM RPOD estimator GA design with RPOD as function to maximize

(a)

(b)
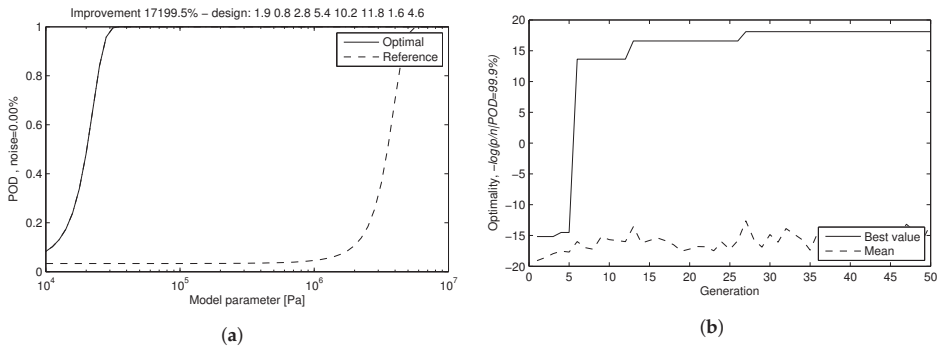
Figure 5. Result of the RPOD optimization process in phase 2. (**a**) RPOD functions of the optimized transducer design and the reference design; (**b**) evolution through generations; best design captured in 27th generation.
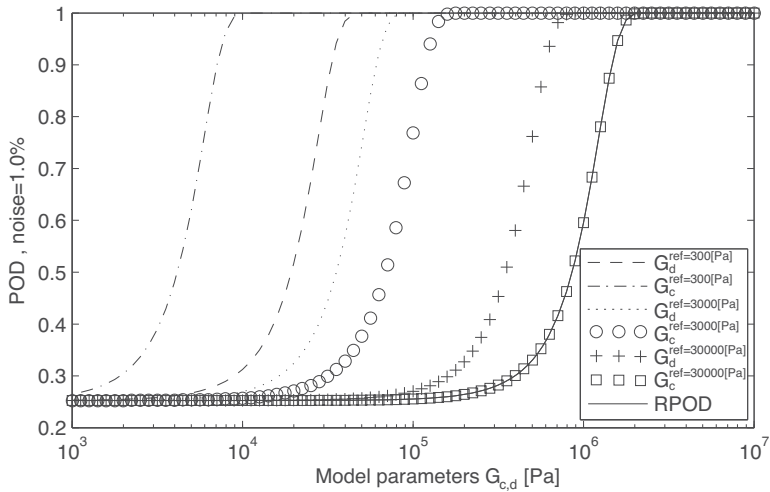
Figure 6. RPOD evaluation (Step 7) of phase 2, for different shear modulus values of reference.

FEAP software is used to solve the FEM algorithm. The mesh and the stress distribution at different instants can be seen in Figure 2 (see [31] for more detail).

The transducer parameters represents and an 8 dimensional space $(pw, pl, pt, dpe, dr, rpe, rw, drt)$ as depicted in Figure 4. A sensitivity analysis of this parameters determines their range of variation that will be used in next steps. Additionally, another sensitivity analysis of the parameters $(G_d^r G_c^r)$ on the POD is performed for values 0.3, 3 and 30 kPa, which is shown in Figure 6.

The RPOD optimization runs an inverse problem that searches for the optimal transducer parameters that maximize the RPOD. The RPOD calculation considers a 2-dimensional space of tissue mechanical parameters of reference $(G_d^r G_c^r)$ to explore a wide range of possible tissues.

Figure 7 shows the schematics of the GA, being Equation (9) the optimization criterion. A generation of $N_d = 20$ specific individuals (transducers designs) from the 8D space are firstly selected. For each individual d, the parameters $[S_{kj}]_d$ (Equation (6)) are calculated for $N_j = 10$ samples of the 2D space $(G_d^r G_c^r)$, selected using a Montecarlo procedure, obtaining $2xN_j$ values of S_{kj}. Then, the values of $[\frac{p_{kj}}{\sigma}]_d$ (see Equation (9)) are calculated for a fixed desired level of 99.9% POD. The minimum value $[\frac{p}{\sigma}]_d$ is selected, thus implementing the Robust POD idea (RPOD), where any other values of reference $(G_d^r G_c^r)$ for the individual d will have a better POD. So, among all individuals of the generation, $d = 1...N_d$, that with maximum RPOD is selected (higher value $[\frac{p}{\sigma}]$). A new generation is formed combining the features of the transducers from the first generation with operations of tournament, mutation, and cross-over [44] and the process is repeated until 50th generation has been calculated. The overall winner, is the transducer design with optimized capacity of detection.
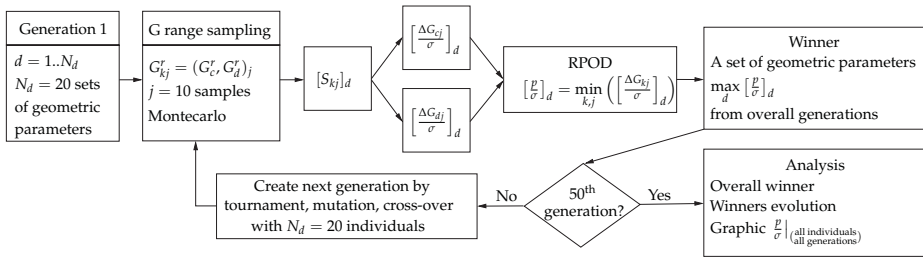
Figure 7. Optimization process maximizing RPOD with the genetic algorithm.

3. Results

Beginning with phase 1, Table 5, shows the optimized mechanical properties that corresponds to the minimized misfit function(discrepancy) between experimental (synthetic) signals and simulated signals, for each of the 4 sets of excitation parameters. Design 3 presents the better minimization of the fitness function, since is the one that best approaches the reference values of the mechanical properties.

Table 5. Adjusted parameter for each set of excitation conditions including the relative error against reference values. Thus, Number 3 is the optimal design.

IP Results	E_3 (MPa)	E_2 (MPa)	R_3 (10^{-3})	R_2 (10^{-3})	f (kHz)	b (kHz)	rs (mm)
Reference Values	27.3861	5.4772	2.2361	2.2361			
Design 1	30.4330 (11%)	5.3381 (3%)	1.6567 (26%)	3.3893 (52%)	6.32	6.32	10
Design 2	25.2412 (8%)	5.9172 (8%)	1.7507 (22%)	3.2839 (47%)	20	20	10
Design 3	29.1313 (6%)	5.1269 (6%)	2.4648 (10%)	1.7845 (20%)	2	1	10
Design 4	30.6323 (12%)	5.0484 (8%)	2.5155 (12%)	1.7478 (22%)	6.32	3.16	20

Figure 8a shows the POD estimation of the selected Design Number 3 for increasing degraded values for a fixed noise level 10%. There are four curves of cumulative probability, one for each parameter p_k. Initially, the less favourable curve would be the one with higher displacement to the right, but taking relative variations instead of absolute values, it results in E_2, so it is conservatively taken as the one which determines the probability of detection. The convergence evolution through generations for Design number 1 is shown in Figure 8b.

(a)

(b)
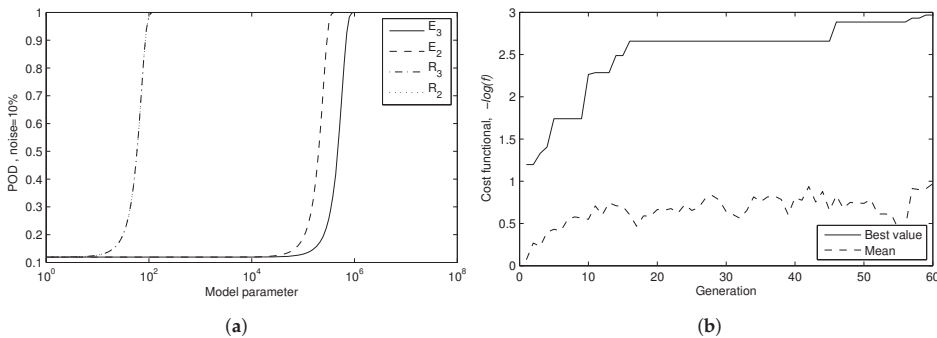
Figure 8. Results of phase 1. (**a**) POD dependency on mechanical properties; (**b**) convergence evolution of Design 1.

The POD curves show that the properties of the second layer can be detected to a level of $\Delta E_2 = 0.5$ MPa (9.12% of the reference value of the property) and $\Delta R_2 = 100$ (4.47% respectively) with

a probability above 99%, under the simulated noise. Similarly, properties of third layer $\Delta E_3 = 1$ MPa (3.65% of the reference value) and $\Delta R_3 = 100$ (4.47%) are detectable with probability above 99%. These figures are compatible with the detectability based on the global minima of the fitness function (with results in Table 5, Design 3). In a conservative position, the POD curve of the selected design in the presented situation can be determined by that of the parameter of E_2, that is the less favorable curve.

The excitation of Design 3 has been found to provide the higher sensitivity to pathology and, therefore, accurate identifiability, where a frequency range between 0.2 and 2 MHz delivers stable results.

The second phase presents the following results: (a) the optimized transducer with the best geometry (winner), shown in Table 6, that maximizes RPOD, applied on the less favorable values of shear moduli of the reference; (b) the evolution of the convergence process through the generations, in Figure 5b; (c) the sampled function of eight dimensions $\frac{p_k}{\sigma}(pw, pl, pt, dpe, dr, rpe, rw, drt)$ used as the criterion decision in the maximization process.

Figure 5a shows the RPOD function of the winner transducer and the initial of the reference. The optimization presents a improvement of 17,199.5% (172-times better) in sensibility (in RPOD), detecting lower changes in the mechanical properties.

The configuration of the optimized transducer produces an excitation with a central frequency of 28 kHz.

Table 6. Optimized transducer, along with assumed ranges and the initial values of reference.

| Design Parameters (mm)	Range	Initial Value	Optimal Value	Label
Piezoelectric Length	(0.5, 2)	1	0.8	pl
Piezoelectric Width	(0.75, 2)	1	1.9	pw
Piezoelectric Thickness	(0.4, 4)	2	2.8	pt
Disc Radius	(1.75, 5.75)	4.25	5.1	dr
Disc Piezoelectric Eccentricity	(1.5, 3.5)	2.5	2.7	dpe
Ring Width	(1.5, 2.5)	2	1.6	rw
Ring Piezoelectric Eccentricity	(5.75, 8.5)	7.5	5.9	rpe
Disc-Ring Thickness	(3, 13)	8	4.6	drt

4. Discussion and Conclusions

A strategy to optimize the design parameters of a torsional shear ultrasonic transducer, based on the inverse problem, has been developed. The main points of the methodology are a forward procedure based on FEM and analytical simplified validation with a sensibility analysis, and an inverse procedure to maximize the transducer sensitivity using a POD estimator which is developed in a robust manner to improve outcomes. Previously, in the first phase a similar methodology is performed to study just the propagation of torsional shear waves across the soft tissue. The parameters of the excitation waves and dimensions has been adjusted looking for the maximum reduction of undesired P wave emissions. The results obtained in both cases are the design parameters and biomechanical clear identification of the mechanical properties of the soft tissue layers within the expected ranges of variation.

This general methodology of a transducer optimization based on maximizing the probability of detection, together with the use of torsional waves, can open an additional line of research to obtain transducers for soft tissue evaluation at low levels of acoustic energy.

In a second step, the desired geometry of the transducer has been manufactured and currently facing adjustments of its functional performance. In the following work, this torsional ultrasonic design will be experimentally validated and improved characterizing mechanical properties in quasi-fluids and soft tissues. The experimental setup will be performed using tissue mimicking phantoms proving a sensor sensitivity study (in terms of angles of incidence and pressures) assessing the robustness of a new proposed elastography technique based on torsional quantitative principles. Three rheological models

will be fitted with the experimental data beside a static independent testing method. The results of the previous rheological tests will be compared with the transducer reconstruction of the biomechanical parameters from the propagated torsional wave.

The transducer manufacturing is presenting the following challenges. The material of the ring, the steel, has been changed to polylactic acid (PLA) to obtain a better amplitude of the received wave, the same occurs with the emitter disk that has been replaced by a small motor with a PLA disk in the contact surface. The whole analytical modelization is still under investigation due to the complexity to make the problem tractable, although this is not affecting the functionality of the transducer. On the contrary, minimization of mechanical and electronic cross-talk effects between emission and reception are being faced before proceeding to the experimental validation.

Safer tests based on ultrasonic techniques than those using ionizing radiations are being investigated and some good results have been arisen, like those of Yonetsu [47] using quantitative sonography, showing a capacity to characterize tumors and differentiate their benign and malignant nature. As changes in consistency of tumors alters their mechanics, mechanical waves could represent an impacting method of diagnosis because of its direct nature of detection.

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-8220/17/6/1402/s1.

Author Contributions: Guillermo Rus and Juan Melchor conceived of and designed the experiments and simulations. Juan Melchor performed the experiments and simulations. Guillermo Rus, Juan Melchor and Rafael Muñoz analyzed the data. Rafael Muñoz and Juan Melchor contributed analysis tools. Rafael Muñoz and Juan Melchor wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wells, P.N.; Liang, H.D. Medical ultrasound: Imaging of soft tissue strain and elasticity. *J. R. Soc. Interface* **2011**, *8*, 1521–1549.

2. Wu, Z.; Taylor, L.S.; Rubens, D.J.; Parker, K.J. Sonoelastographic imaging of interference patterns for estimation of the shear velocity of homogeneous biomaterials. *Phys. Med. Biol.* **2004**, *49*, 911.

3. Glazounov, A.; Zhang, Q.; Kim, C. Torsional actuator based on mechanically amplified shear piezoelectric response. *Sens. Actuators A Phys.* **2000**, *79*, 22–30.

4. Harkness, P.; Lucas, M.; Cardoni, A. Coupling and degenerating modes in longitudinal–torsional step horns. *Ultrasonics* **2012**, *52*, 980–988.

5. Szabo, T. *Diagnostic Ultrasound Imaging: Inside Out*; Academic Press: Oxford, UK, 2004.

6. Thijssen, J.; Verhoef, W.; Cloostermans, M. Optimization of ultrasonic transducers. *Ultrasonics* **1985**, *23*, 41–46.

7. Heikkola, E.; Laitinen, M. Model-based optimization of ultrasonic transducers. *Ultrason. Sonochemistry* **2005**, *12*, 53–57.

8. Schröder, A.; Rautenberg, J.; Henning, B. Evaluation of cost functions for FEA based transducer optimization. *Phys. Procedia* **2010**, *3*, 1003–1009.

9. Heikkola, E.; Miettinen, K.; Nieminen, P. Multiobjective optimization of an ultrasonic transducer using NIMBUS. *Ultrasonics* **2006**, *44*, 368–380.

10. Fu, B.; Hemsel, T.; Wallaschek, J. Piezoelectric transducer design via multiobjective optimization. *Ultrasonics* **2006**, *44*, e747–e752.

11. Kim, J.; Kwon, O. Vibration characteristics of piezoelectric torsional transducers. *J. Sound Vib.* **2003**, *264*, 453–473.

12. Kaneko, S.; Nomoto, S.; Yamamori, H.; Ohya, K. Load characteristics of a bolted Langevin torsional transducer. *Ultrasonics* **1996**, *34*, 239–241.

13. Lin, S. Study on the Langevin piezoelectric ceramic ultrasonic transducer of longitudinal–flexural composite vibrational mode. *Ultrasonics* **2006**, *44*, 109–114.

14. Aoyagi, M.; Suzuki, T.; Tomikawa, Y. Characteristics of a bolt-clamped torsional vibrator using shear-mode piezo-ceramics sandwiched in the axial direction. *Ultrasonics* **1996**, *34*, 219–222.

15. Harada, T.; Ishikawa, N.; Kanda, T.; Suzumori, K.; Yamada, Y.; Sotowa, K. Droplet generation using a torsional Langevin-type transducer and a micropore plate. *Sens. Actuators A Phys.* **2009**, *155*, 168–174.

16. Li, J.; Liu, P.; Ding, H.; Cao, W. Modeling characterization and optimization design for PZT transducer used in Near Field Acoustic Levitation. *Sens. Actuators A Phys.* **2011**, *171*, 260–265.

17. Lin, H.; Fang, W. A rib-reinforced micro torsional mirror driven by electrostatic torque generators. *Sens. Actuators A Phys.* **2003**, *105*, 1–9.

18. Kwun, H.; Jolly, W.; Light, G.; Wheeler, E. Effects of variations in design parameters of ultrasonic transducers on performance characteristics. *Ultrasonics* **1988**, *26*, 65–72.

19. Wang, F.; Zhao, X.; Zhang, D.; Wu, Y. Development of novel ultrasonic transducers for microelectronics packaging. *J. Mater. Process. Technol.* **2009**, *209*, 1291–1301.

20. Murav'evaa, O.V.; Len'kovb, S.V.; Murashova, S.A. Torsional waves excited by electromagnetic acoustic transducers during guided-wave acoustic inspection of pipelines. *Acoust. Phys.* **2016**, *62*, 117–124.

21. Kwun, H.; Teller, C.M. Magnetostrictive generation and detection of longitudinal, torsional, and flexural waves in a steel rod. *J. Acoust. Soc. Am.* **1994**, *96*, 1202–1204.

22. Fan, Z.; Lowe, M.J.S.; Castaings, M.; Bacon, C. Torsional waves propagation along a waveguide of arbitrary cross section immersed in a perfect fluid. *J. Acoust. Soc. Am.* **2008**, *124*, 2002–2010.

23. Kim, J.O.; Bau, H.H. Instrument for simultaneous measurement of density and viscosity. *Rev. Sci. Instrum.* **1989**, *60*, 1111–1115.

24. Spratt, W.K.; Vetelino, J.F.; Lynnworth, L.C. Torsional Ultrasonic Waveguide Sensor. *IEEE Int. Ultrason. Symp.* **2010**, 702–706.

25. Ophir, J.; Alam, S.; Garra, B.; Kallel, F.; Konofagou, E.; Krouskop, T.; Varghese, T. Elastography: Ultrasonic estimation and imaging of the elastic properties of tissues. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **1999**, *213*, 203–233.

26. Van Kervel, S.; Thijssen, J. A calculation scheme for the optimum design of ultrasonic transducers. *Ultrasonics* **1983**, *21*, 134–140.

27. Sunny, Y.; Bawiec, C.; Nguyen, A.; Samuels, J.; Weingarten, M.; Zubkov, L.; Lewin, P. Optimization of un-tethered, low voltage, 20–100 kHz flexural transducers for biomedical ultrasonics applications. *Ultrasonics* **2012**, *52*, 943–948.

28. Rus, G.; Bochud, N.; Melchor, J.; Alaminos, M.; Campos, A.; Linde, B.B.; Pączkowski, J.; Ponikwicki, N. Dispersive model selection and reconstruction for tissue culture ultrasonic monitoring. *AIP Conf. Proc.* **2012**, *1433*, 375–378, doi:10.1063/1.3703208.

29. Peralta, L.; Molina, F.; Melchor, J.; Gómez, L.; Massó, P.; Florido, J.; Rus, G. Transient elastography to assess the cervical ripening during pregnancy: A preliminary study. *Ultraschall Med.-Eur. J. Ultrasound* **2015**, doi:10.1055/s-0035-1553325.

30. Rodríguez, J.M.M. Mechanics of Nonlinear Ultrasound in Tissue. Ph.D. Thesis, Universidad de Granada, Granada, Spain, 2016.

31. Melchor, J.; Rus, G. Torsional ultrasonic transducer computational design optimization. *Ultrasonics* **2014**, *54*, 1950–1962.

32. Hadj Henni, A.; Schmitt, C.; Trop, I.; Cloutier, G. Shear wave induced resonance elastography of spherical masses with polarized torsional waves. *Appl. Phys. Lett.* **2012**, *100*, 133702.

33. Ekeom, D.; Henni, A.H.; Cloutier, G. Design of a phased array for the generation of adaptive radiation force along a path surrounding a breast lesion for dynamic ultrasound elastography imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2013**, *60*, 552–561.

34. Ouared, A.; Montagnon, E.; Cloutier, G. Generation of remote adaptive torsional shear waves with an octagonal phased array to enhance displacements and reduce variability of shear wave speeds: Comparison with quasi-plane shear wavefronts. *Phys. Med. Biol.* **2015**, *60*, 8161.

35. Ouared, A.; Kazemirad, S.; Montagnon, E.; Cloutier, G. Ultrasound viscoelasticity assessment using an adaptive torsional shear wave propagation method. *Med. Phys.* **2016**, *43*, 1603–1614.

36. Mattei, G.; Ahluwalia, A. Sample, testing and analysis variables affecting liver mechanical properties: A review. *Acta Biomater.* **2016**, *45*, 60–71.

37. Li, L.; Stiadle, J.M.; Lau, H.K.; Zerdoum, A.B.; Jia, X.; Thibeault, S.L.; Kiick, K.L. Tissue engineering-based therapeutic strategies for vocal fold repair and regeneration. *Biomaterials* **2016**, *108*, 91–110.

38. Jiao, T.; Farran, A.; Jia, X.; Clifton, R.J. High frequency measurements of viscoelastic properties of hydrogels for vocal fold regeneration. *Exp. Mech.* **2009**, *49*, 235–246.

39. Moster, D.; Lie, R.T.; Markestad, T. Long-term medical and social consequences of preterm birth. *N. Engl. J. Med.* **2008**, *17*, 262–273.

40. Blondel, B.; Macfarlane, A.; Gissler, M.; Breart, G.; Zeitlin, J. Preterm birth and multiple pregnancy in European countries participating in the PERISTAT project. *BJOG* **2006**, *113*, 528–535.

41. Bercoff, J.; Tanter, M.; Fink, M. Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2004**, *51*, 396–409.

42. Rus, G.; Lee, S.Y.; Chang, S.Y.; Wooh, S.C. Optimized damage detection of steel plates from noisy impact test. *Int. J. Numer. Methods Eng.* **2006**, *68*, 707–727.

43. Rade, L.; Westergren, B. *Mathematics Handbook for Science and Engineering*; Springer: Berlin/Heidelberg, Germany, 1999.

44. Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, USA, 1989.

45. Gallego, R.; Rus, G. Identification of cracks and cavities using the topological sensitivity boundary integral equation. *Comput. Mech.* **2004**, *33*, 154–163.

46. Taylor, R.L.; Govindjee S. *FEAP Manual*; University of Berkeley: Berkeley, CA, USA, 2003.

47. Yonetsu, K.; Ohki, M.; Kumazawa, S.; Eida, S.; Sumi, M.; Nakamura, T. Parotid tumors: Differentiation of benign and malignant tumors with quantitative sonographic analyses. *Ultrasound Med. Biol.* **2004**, *30*, 567–574.

Article

A Novel Active Imaging Model to Design Visual Systems: A Case of Inspection System for Specular Surfaces

Jorge Azorin-Lopez *, Andres Fuster-Guillo, Marcelo Saval-Calvo, Higinio Mora-Mora and Juan Manuel Garcia-Chamizo

Department of Computer Technology, University of Alicante, Carretera San Vicente s/n, San Vicente del Raspeig, Alicante 03690, Spain; fuster@dtic.ua.es (A.F.-G.); msaval@dtic.ua.es (M.S.-C.); hmora@ua.es (H.M.-M.); juanma@dtic.ua.es (J.M.G.-C.)
* Correspondence: jazorin@dtic.ua.es; Tel.: +34-96-590-3681

Abstract: The use of visual information is a very well known input from different kinds of sensors. However, most of the perception problems are individually modeled and tackled. It is necessary to provide a general imaging model that allows us to parametrize different input systems as well as their problems and possible solutions. In this paper, we present an active vision model considering the imaging system as a whole (including camera, lighting system, object to be perceived) in order to propose solutions to automated visual systems that present problems that we perceive. As a concrete case study, we instantiate the model in a real application and still challenging problem: automated visual inspection. It is one of the most used quality control systems to detect defects on manufactured objects. However, it presents problems for specular products. We model these perception problems taking into account environmental conditions and camera parameters that allow a system to properly perceive the specific object characteristics to determine defects on surfaces. The validation of the model has been carried out using simulations providing an efficient way to perform a large set of tests (different environment conditions and camera parameters) as a previous step of experimentation in real manufacturing environments, which more complex in terms of instrumentation and more expensive. Results prove the success of the model application adjusting scale, viewpoint and lighting conditions to detect structural and color defects on specular surfaces.

Keywords: visual inspection; specular surfaces; structured lighting; quality control

1. Introduction

Scientific models and theories aimed at explaining the behaviour of specular reflections are sufficiently consolidated. However, the automatic processing, using artificial vision techniques, of scenes where there are specular surfaces, has problems that have not been solved yet. Vision systems designed to deal with specular objects must cope with the optical difficulties associated with this type of material [1]. The surfaces have a high reflection coefficient which causes undesired reflections and shine, concealing, in some cases, chromatic, morphological and topographical information about the object.

Most artificial vision techniques ignore specular reflections and focus on the diffuse component of the interaction of light with objects. Thus, these techniques, which are conceived to deal with Lambertian surfaces, produce wrong results when other surfaces are in the visual scene [2,3]. The solutions put forward in the literature adopt opposite ways of approaching the problem: developing methods to detect specularities in images, in order to take advantage of existing artificial vision techniques; and designing new vision techniques that explicitly deal with these surfaces.

The methods designed for detecting specularities in the scene differ depending on the vision level in which they are approached. They take advantage of certain phenomena and characteristics of light to separate the contribution of diffuse and specular reflection: spectral distribution of reflections [4–7], polarization [8–11], analysis of the behaviour of specularities in several images [12–16], and combinations of them [17–19]. They are used either to avoid or to remove specularities in the original image. The solutions in the last case are at the low level of vision systems because reflections and shine are considered as noise to be removed. They offer a filtered image to be processed at higher levels.

Designing new vision techniques allows specular reflections to be considered as a peculiar characteristic of the surface, which allows better performance of the vision process [20–26]. They use mechanisms of active vision that, in some cases, are modifications or refinements of classic techniques. They are generally focused on the extraction of the object shape [27–38].

The techniques developed to deal with specular surfaces, either for detecting specularities or for extracting the object shape, must fulfill certain requirements, making them only viable in specific applications: for example, requirements related to specific electromagnetic characteristics of objects (i.e., specific methods for metallic or dielectric objects), having previous knowledge of the geometry of the scene or of the surface reflectance, etc.

Specifically, automated visual inspection is the most used tool for testing products in industry due to their ease of use and its cost [39–42]. However, there are very few vision systems that deal with the surface inspection of specular products. They share the same aforementioned drawback of the generic vision system for specular objects: they are very specific as they assume specific constraints. The lighting of the scene and the acquisition equipment are determinant factors in the proposed solutions since they help the detection of defects in images [43–47].

Techniques that use structured lighting appear in few systems cited in the literature. They are considered to be the most reliable and suitable for inspecting the 3D shape of products [36,48–50]. Moreover, they have advantage over other techniques including laser, time of flight or LIDAR because the same sensor is used to determine colour information instead of acquiring colour and shape information in an independent manner. Techniques differ in the way they acquire the lighting pattern by means of projecting on the surface [51–53] or focusing the system on the reflection [36,48,54,55] (assuming the object as part of optical system where lighting patterns are projected); the equipment used to generate patterns [56–58] (screens, projectors, etc.); and, in the method used to codify patterns [59–63].

Proposed solutions try to satisfy specific requirements by means of adapting to the application domain and to specific constraints of the products. A system designed to satisfy the quality control of a product generally cannot be applied to another system. As a consequence, the contribution of this paper is to provide an active vision model able to explain the problem of inspecting specular surfaces and able to help in designing vision systems for this purpose. Moreover, the paper proposes a method based on the model that is able to minimize the negative effects of specular surfaces in visual inspection and able to take advantage of specular reflections as a peculiar characteristic. The method is focused on controlling the acquisition conditions (e.g., lighting angles, viewpoints, chromaticity and other lighting characteristics, etc.) to maximize the likelihood of detecting defects. Particular characteristics of the inspection problem make control of the acquisition conditions possible.

In order to validate the model the use of simulation is proposed as an inspection system design methodology that could be systematically applied, studying conditions in which the inspection has to be carried out and designing solutions in a flexible way. Virtual inspection makes use of the virtual manufacturing technology to model and simulate the inspection process, and the physical and mechanical properties of the inspection equipment, to provide an environment for studying the inspection methodologies [64]. Simulation based on virtual imaging enables the rendering of realistic images, providing a very efficient way to perform tests compared to the numerous attempts of manual experiments [65]. The introduction of simulation provides a flexible and low-cost method (compared

with experimentation in the laboratory) of testing original hypotheses and the benefits that can be drawn from this research.

The rest of this paper is organized as follows. Section 2 describes the active vision model. The method based on the model for inspecting specular surfaces is developed in Section 3. Sections 5 and 6 evaluate the proposed method by controlling image acquisition conditions. Finally, Section 7 concludes the paper.

2. Active Vision Model to Deal with Specular Surfaces

We are interested in modeling the automatic process of artificial vision in order to provide solutions to the problem of inspecting specular surfaces. First of all, a model describing the image formation and the variables that take part in this process is presented.

An image I is defined as a two-dimensional representation provided by F. Let F a function that models a visual acquisition system, *VAS*. It includes all equipment and scene configuration to capture an image: lighting, positions, viewpoints, cameras, etc. Let ρ a vector made up of *scene magnitudes* that contribute to the formation of I Equation (1). Each vector ρ is an element of a representation space P related to optical magnitudes of the visual perception phenomenon.

$$I(x,y) := F(\rho), \ \rho = (\rho_1, \rho_2, \rho_3, ..., \rho_n) \in P \tag{1}$$

The components ρ_i of the vector of *scene magnitudes* are measurable physical values involved in the process of image formation. They could be, in practice: scale, viewpoint, light intensity, frequency, saturation, etc. Each component could be modeled as a function depending on three inputs Equation (2): the subject of interest, m, in the scene (e.g., the object to be inspected), the environment, e, in which the subject is placed and, finally, the camera, c, that captures images from the scene.

$$\rho_i = \rho_i(m, e, c) \tag{2}$$

The contribution of each element (m, e and c) can be expressed as three vectors made up of magnitudes: μ_i related to the object Equation (3), ϵ_i related to the environment Equation (4) and γ_i related to the camera Equation (5). Intensity and wavelength of light sources, medium of transmission, relative position between scene and vision device, are examples of environment variables ϵ_i. Regarding the camera contribution γ_i, the variables are related to the sensor characteristics, optical and electronic elements: zoom, focus, diaphragm, size of the sensor, signal converters, etc. Finally, the reflectance, colour, shape, topography of the object are examples of object variables μ_i. The values of each vector establish elements of the set M for the object, E for the environment and Γ for the camera. In Diagram 1 an outline of the magnitudes can be found.

$$M = \{m_0, m_1, m_2, ...\}, m_i = (\mu_1, \mu_2, ..., \mu_m) \tag{3}$$

$$E = \{e_0, e_1, e_2, ...\}, e_i = (\epsilon_1, \epsilon_2, ..., \epsilon_n) \tag{4}$$

$$\Gamma = \{c_0, c_1, c_2, ...\}, c_i = (\gamma_1, \gamma_2, ..., \gamma_l) \tag{5}$$

To define our model, it is worth to remember the sensitivity as an important static characteristic of a sensor. The sensitivity is the slope of the calibration curve (see Figure 1). First, we can define the calibration curve as the function that maps a physical scene magnitude and its representation in the image space. Depending on the camera parameters the calibration curve will have a different function for different scene magnitudes. For instance, a camera with large depth of field will have a smoother calibration curve for intensity measurement (along a larger set of scene magnitudes for intensity, the camera will be able to distinguish between them) than a camera with short depth of field which will have a very abrupt calibration curve for the same scene magnitude (only a small set of intensity values will be distinguishable). With this function, sensitivity indicates the detectable output

change for the minimum change of the measured magnitude. As a naive example, the detectable output could be the ability to perceive two different color that are actually different in the real scene. If the color change is too small, the sensor would return the same value of intensity. In our case, the detectable output in F for the minimum change of the *scene magnitudes*:

$$sensitivity = \frac{\Delta F(\rho)}{\Delta \rho} \tag{6}$$
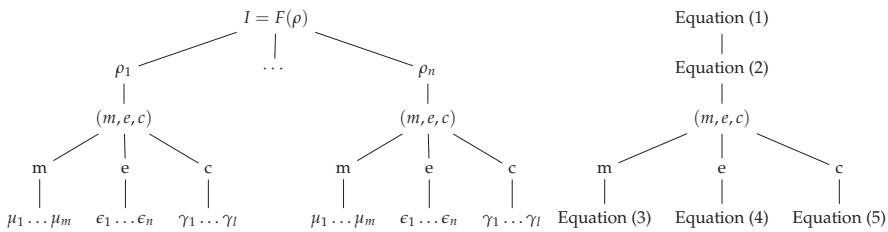
Diagram 1: Levels of magnitudes involved in the formation of the image. The image function as the most abstract level on the top, then each scene magnitude, the three elements that composes ρ, and the individual measurements of each element on the bottom of the tree.

Usually, the camera parameters are calibrated to a set of values γ_i so that the sensitivity is optimized for all variables of ρ simultaneously with respect to a single metric (not necessarily maximized for each variable ρ_i). For example, a camera could be calibrated using a specific zoom, focus and sensor size optimizing the sensitivity of the system to perceive the colour of a subject in a wide range of object distances and viewpoints but the system is not optimized separately for distances or viewpoints. However, a given camera c has maximum sensitivity for a value of each ρ_i. For convenience, we are going to define the *tuning point* as the corresponding point ρ_s in the *scene magnitudes* space P for each camera of the set Γ in which the sensitivity of the *VAS* is the optimum (see Figure 1) . The sensitivity decreases in general for values of ρ differently from ρ_s.
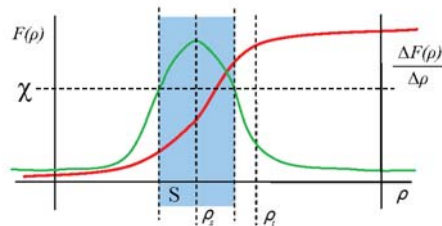
Figure 1. Calibration curve (**red**) and sensitivity (**green**) for the visual acquisition system (*VAS*) including *tuning* ρ_s and *working point* ρ_t.

In the same way, since the *VAS* depends on the contribution of the environment e, for each value of the set E, the corresponding point ρ_t in the *scene magnitudes* will be named as the *working point*. The detected output of this point is related to the sensitivity curve of the *VAS* for each of the magnitudes ρ_i because it restricts the limits where the system can work. In average conditions or in simple approaches to vision problems, its effect is usually considered to be negligible because perception takes place close to ρ_s. This simplification is unacceptable in the case of adverse conditions as it occurs, for example, in dark environments, remote objects from the camera, or in the presence of specular objects that limit the capability of the *VAS* to perceive the scene.

Acquisition of the scene and representation on a plane carried out by *F* Equation (1) cause situations, related to vision in adverse conditions, where it is not possible to distinguish between the different scene magnitude vectors *ρ* that contribute to an image. The capacity to discern elements of *P* is related to the measurement in the image *I* of the magnitudes *m*, *e*, and/or *c*, that contribute to each of the components ρ_i. A specific application could be interested in knowing the intensity lighting of a scene from the environment *e*, or the focal length from the camera *c*, for calibration purposes. However, generally, the contribution of the object *m* to the image, and, therefore, its magnitudes μ_i, is the aim of the measurement. For example, as we can see in Figure 2, a *VAS* used to perceive color (μ_2) of objects is able to distinguish the different objects in the scene. However, if it is used to perceive shape (μ_1) using the same camera (including all variables γ_i), neither objects m_8 and m_9 nor m_{10} and m_{11} could be distinguished each other (as we mentioned above, a given camera has maximum sensitivity for a specific value of each ρ_i). Hence, capacity to discern magnitudes that contribute to the vector *ρ* can be delimited to distinguish elements of set *M* in the image (e.g., the colour of a region of an object, or the shape of a surface, or both, etc.). It is important to understand that the scene magnitudes are continuous, not discrete. The representation in the Figure 2 shows a set of magnitudes that are distinguishable in the space of the image $I = F(\rho)$, but it is an example of the certain magnitudes in the continuous space of magnitudes *P*.

An object could be distinguishable from another one in the *VAS* when they are distinguishable in the measurement performed in *F* (i.e., in the vertical axis of Figures 1 and 2A). Let Ω_{m_i} be the set of objects that can be distinguished for an specific object m_i, and let *χ* the minimum difference perceptible by the system (sensitivity), then Ω_{m_i} could be established as:

$$\Omega_{m_i} = \left\{ m_j \in M : (\exists \chi > 0)\left[\, |F(^{m_j}\rho) - F(^{m_i}\rho)| \geq \chi\right] \right\} \tag{7}$$

Figure 2B, shows the measurement performed in *F* analysed from the point of view of the object space *M* considering just colour and shape. Objects close to m_1, in the yellow area, are not distinguishable, for that object in this example. However, they are distinguishable for the object m_2 (light green area). Following this example, $m_{10} \in \Omega_{m_9}$ but $m_{10} \notin \Omega_{m_{11}}$, if the measurement performed in *F* takes into account both colour and shape.

A *VAS* has to deal with different situations that are a consequence of the subsets of *P* delimited by the problem to be solved (e.g., delimited by objects or by their characteristics to be analysed, or by environments, or by the cameras, etc.). A minimum value of sensitivity *χ* can be established in which any object is distinguishable from another considered in the acquisition (see Figure 1). The values of the vector of magnitudes *ρ* in which sensitivity is higher than a threshold *χ* conform the subset *S* of *P* defined by Equation (8).

$$S = \left\{ \rho_i : (\exists \chi > 0)(\forall m_k, \forall m_j \in M)\left[\frac{|F(\rho_i) - F(\rho_{i-1})|}{\Delta \rho} \geq \chi \rightarrow m_j \in \Omega_{m_k} \right] \right\} \tag{8}$$

There are no perception difficulties for situations of the subset $S \subset P$. They involve values of environment and camera, in addition to characteristics of objects, that make up magnitudes *ρ* of the set *S* Equation (8). In other words, it is possible to distinguish the images of two different objects (m_i and m_j in Equation (9)) from the acquisition performed by *F* if the *VAS* is working on points of *S* ($^{m_i}\rho$ denote a scene magnitude vector whose component *m* is the object m_i).

$$(\exists^{m_i}\rho, ^{m_j}\rho \in S)(\forall m_i, m_j \in M)\left[(m_i \neq m_j) \rightarrow |F(^{m_i}\rho) - F(^{m_j}\rho)| > 0\right] \tag{9}$$
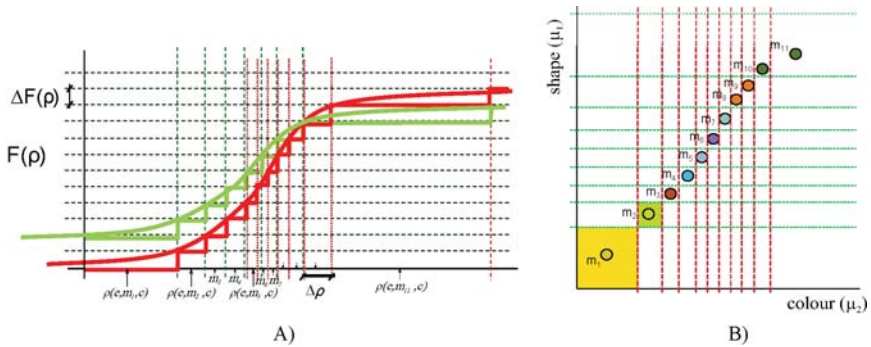
Figure 2. (**A**) Example of calibration curves for a *VAS* aimed to measure shape (**green**) and colour (**red**); (**B**) Measurement performed in *F* transformed to the object space *M* considering shape μ_1 and colour μ_2. In this example, the colour of the objects m_{10} and m_{11} can be distinguishable but their shape cannot.

Vision systems working on *scene magnitudes* of the complementary of *S*, S^c, ($P \setminus S$) present the aforementioned adverse conditions to distinguish characteristics of objects from images. Hence, solutions have to be provided to achieve distinct images from different objects that the camera perceives as the same one. These solutions should be able to compensate the low sensitivity in S^c (sensitivity less than χ in Figure 1). Among the three variables that provide values to the *scene magnitudes*, object is a constant due to them being the subject of interest. However, environment conditions and camera characteristics could be modified to set up *scene magnitudes* ρ of the subset *S*. Thus, the *VAS* is going to be able to have different images from different objects in the scene in order to distinguish them Equation (9). For this purpose, we propose two complementary alternatives (Diagram 2):

- *System Calibration* (calibrating the system). This alternative tries to minimize the distance between the *working point* ρ_t and the *tuning point* ρ_s.
- *Measurement Enhancement* (conditioning the measurement). Enhance the target measurements or parameters. This alternative can be considered as conditioning the measurement performed by the *VAS*.

System Calibration consists of shifting one of the points so that the *working point* be an element of the set *S*. The goal could be generating a new image of the object m_i by a transformation Y_S Equation (10) in which the *working point* ($\rho_k^{m_i}$) be close enough to the *tuning point*. Do not confuse calibrating the system with calibrating the camera. The alternative presented here could be calibrating the sensor as well as changing other parameters of environment or object of interest. Hence, to carry out this alternative it is necessary to adjust the environment to shift the *working point*, for example moving the object closer to the camera, adjusting lighting conditions, etc. (Figure 3 shows an outline of this process). On the other hand, in order to shift the *tuning point* (see Figure 4), the camera could be recalibrated or new acquisition equipment can be used (traditionally this is done by replacing the camera with a more suitable one). In this case, subset *S* of *P* changes for the new camera from S_o to S_n.

$$(\exists Y_S)(\exists \rho_k^{m_i} \in S)(\forall \rho^{m_i} \in P)\left[F(\rho_k^{m_i}) = Y_S\left(F(\rho^{m_i})\right)\right] \qquad (10)$$

Solution alternatives

System Calibration   *Measurement Enhancement*

Y_S (Equation (10)) Y_C (Equation (12))

Moving working point (Figure 3)   Moving tuning point (Figure 4)   Amplify perception   Increasing differences of input magnitudes (Figure 5)

Adjust lights, · · ·   *Calibration, · · ·*   *Signal amplification, · · ·*   *Light pattern, · · ·*
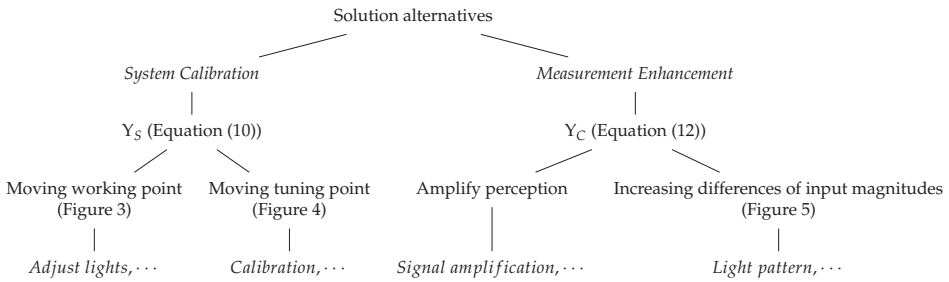
Diagram 2: Diagram with the different alternatives to improve the system perception.

For the second alternative, *Measurement Enhancement* aims to directly influence the sensitivity curve of the acquisition system. In a nutshell, it tries to somehow highlight or enhance the parameters ρ which want to be perceived. It can be carried out by means of two new alternatives. First, the output signal of the perception system can be amplified. That is, the classic conception of measurement system amplification at the signal conditioning step. Limitations of this technique are related to increasing the amplitude of the signal in ranges of minimal sensitivity (both minimum and maximum of the range of the sensor because the contribution of the object in the output signal is insignificant, the signal-to-noise ratio is very low). In this way, acquisition system improvements are limited because they are only applied in ranges of intermediate sensitivity.
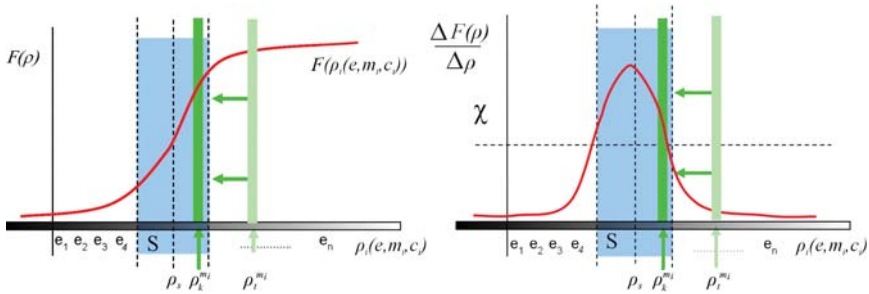
Figure 3. Outline for a *System Calibration* Y_S shifting the *working point* to *S*. The initial *working point* $\rho_l^{m_i}$ (**light green**) has been moved to $\rho_k^{m_i}$ (**dark green**) after transformation is applied. The new *working point* $\rho_k^{m_i}$ is close enough to the *tuning point* ρ_s.

The other alternative of enhancing the target measurement is increasing the differences of the values of the input magnitudes. This is to operate with large differences (η) of the input (ρ) to increase the differences of the output F for different objects (m_i and m_j) until the differences are perceptible at the output ($m_j \in \Omega_{m_i}$). Figure 5 schematically shows this concept. Elements used as input magnitudes of the set P to get large differences of the input make up the subset A in Equation (11). The goal is to reduce the number of possibilities that the *VAS* has to deal with; for example, restricting camera positions, viewpoints or lighting characteristics (in this paper, a lighting pattern has been used to inspect specular surfaces).

$$A = \left\{ \; \rho_k^{m_i} : (\exists \eta > 0)(\forall \rho_l^{m_j} \in P)\left[\left[(m_i \neq m_k) \wedge \left| \rho_k^{m_i} - \rho_l^{m_j} \right| \geq \eta \right] \to m_j \in \Omega_{m_i} \right] \; \right\} \qquad (11)$$

We model the transformation able to increase the differences of the values of the input magnitudes as Y_C:

$$(\exists Y_C)(\exists \rho_k^{m_i} \in A)(\forall \rho^{m_i} \in P)\left[F(\rho_k^{m_i}) = Y_S\left(F\left(\rho^{m_i}\right)\right)\right] \tag{12}$$
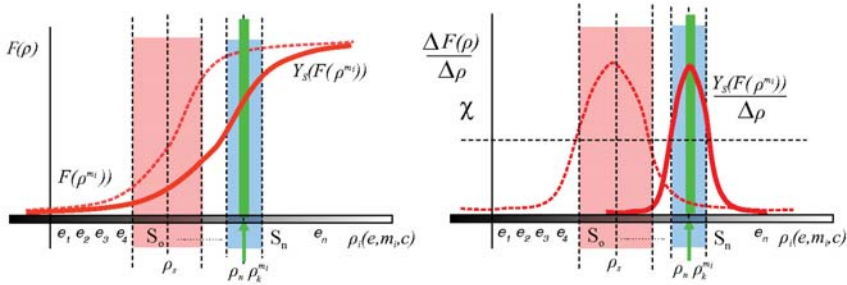
Figure 4. Outline for a transformation Y_S shifting the *tuning point* to the *working point* $\rho_k^{m_i}$. Dotted line curve and light red area represents the old calibration and S_o respectively. Red line curve and blue area represent the new calibration curve and the new S_n. The old *tuning point* ρ_s has been consequently moved to the new one ρ_n after transformation is applied.

The techniques are not exclusive and can be used together for designing vision system in which images of different objects can be distinguished. An example of the use could be a system which needs to perceive two colliding objects separately through a sensor with a fisheye lens. First, a Y_S transformation by means of calibrating the camera to reduce the distortion could be applied. After, Y_C could mean to colorize the objects to enlarge the perceived difference between them.
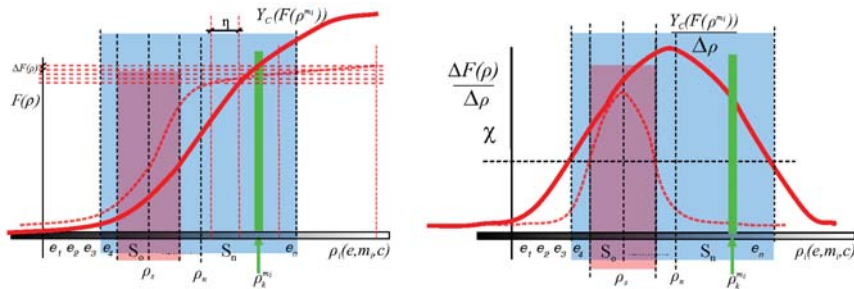
Figure 5. Outline for a transformation Y_C able to increase differences of input magnitudes values until the change is perceptible in the output F. Dotted line curve and light red area represents the old calibration and S_o respectively of the *VAS*. Red line curve and blue area represent the new calibration curve and the new S_n. The vertical dotted red lines represent values of the set A that allow perceptible changes in the output (horizontal dotted red lines).

3. Method for Inspecting Specular Surfaces

In this section, we are going to use the previous model to specify a method for inspecting specular surfaces. The objective of an automated visual inspection system, *AVI*, aims to determine if a product differs from the manufacturer's specifications. This implies the *AVI* has to measure the object magnitudes in the scene in order to compare them with values of the magnitudes established in the design step of the product (e.g., reflectance, colour, shape, topography).

Two sets of objects are considered for modelling the *AVI*: M_P and M_I. The first one is composed of objects that are made up by magnitudes μ_i defined in the manufacturing specifications. The set

M_I is composed by objects to be inspected looking for any deviation from objects of the set M_P. These deviations cause, depending on the magnitude, different defects: morphological, chromatic, topographic defects, etc. The union of M_P and M_I is the subset $M_S = \{M_P \cup M_I\} \subset M$ Equation (3) of the possible objects that the inspection system must consider.

The inspection goal could be modelled as in Equation (13). The *AVI* has to decide whether an object m_i of the set M_I could be distinguished (9) from an object m_j of the set M_P if any of the magnitudes $\mu_i^{m_j}$ differ in some value η from the original $\mu_i^{m_k}$. The object m_j is considered the object model of m_i and contains the manufacturing specifications.

$$(\exists i)(\exists \eta > 0)(\forall m_j \in M_I)(\forall m_k \in M_P)\left[\left|\mu_i^{m_j} - \mu_i^{m_k}\right| \geq \eta \rightarrow m_j \in \Omega(m_k)\right] \tag{13}$$

In order for the deviations (η) of the object magnitudes $\mu_i^{m_j}$ from $\mu_i^{m_k}$ to be detected in the image Equation (13), contributions from the environment and the camera to ρ must allow a suitable sensitivity to the *AVI*.

Since it was previously shown that the sensitivity of the *VAS* (*AVI* in this specific case) is optimized simultaneously for all variables of ρ, *AVI* sensitivity is not necessarily maximized for each variable ρ_i. It is not maximized for each object m and each of its variables μ_i. Moreover, generally an *AVI* is designed for performing a measurement of a subset of the object magnitudes μ_i in the image (i.e., colour, shape). Therefore, the sensitivity of the *AVI* can be very low for some of the magnitudes μ_i to be measured. That is, it is possible that the perception be suitable for measuring the surface colour or shape but it could not be suitable for both together. In other words, the intended measurement determines the perception capacity of the *AVI*. Calibration parameters or environment conditions must be adjusted to adequately perceive magnitudes μ_i. This process requires great knowledge of the problem and accuracy for the solution.

For specular surfaces, the difficulty to perceive in *scene magnitudes* of S^c is a consequence of the surface reflectance. The environment conditions produce an effect in the perception of the object being more important than other types of surfaces (e.g., Lambertian surfaces). For example, if it is considered that the calibration parameters γ_i are the same for two different images, the difficulty to perceive with different environment conditions is given by:

- The mirror itself: a given camera can be confused so that it cannot distinguish between the environment and the object. The spatial modulation of the environment contribution creates the illusion of the objects in a scene.
- The lighting of the environment can cause shine on the surface and confuse the two images with different objects. For example, the image formed by a grey surface with a high reflection coefficient illuminated by white lights can be confused with the image of a lighter surface object.

The specular reflection causes the *working point* ρ_t be easily located at the limit of the range of the *VAS*, at the maximum value of ρ that can be measured. Sensitivity is very low for the perception of objects in these conditions because specularity saturates the camera sensor.

The viable solution for compensating the lack of sensitivity produced in the ranges of the *scene magnitudes* ρ (in order to obtain an improved image) is related to the *System Calibration* that performs the perception system (see Section 2). It is necessary to increase the differences in the values of the input magnitudes, Δ_ρ, to raise the differences of the output magnitude until they can be measured (Equation (12)).

In other words, using the resolution (another important sensor characteristic that indicates the smallest change in the magnitude being measured that the sensor can detect, e.g., the smallest feature size of an object or the smallest change in colour that the *VAS* can distinguish), since the resolution in those ranges is very low, it is necessary to force large differences at the input. Thus, it is necessary to work using a subset of *scene magnitudes* $A \subset P$ in order the perception will be suitable in the points around the *tuning point* and it will be facilitated in the ranges of sensor saturation. Differences in F will

enable the perception among objects using the set S_A in Equation (14). The elements do not necessarily correspond to those of subset S Equation (8).

$$S_A = \left\{ \; \rho_i \in A : (\exists \chi > 0)(\forall m_i \in M_P)(\forall m_j \in M_I) \left[\frac{\Delta F(\rho_i)}{\Delta \rho} \geq \chi \rightarrow m_j \in \Omega_{m_i} \right] \; \right\} \qquad (14)$$

In addition, if the *Measurement Enhancement* is not sufficient to discern the defects in the inspection, the distance between the *working point* ρ_t and the *tuning point* ρ_s must be minimized using the Equation (10). In this case, minimization is performed on the input magnitudes of the set A. Therefore, the transformation Y_S operates with the values of the input magnitudes S_A Equation (14). In consequence, Equation (15) models the proposed solution for inspecting specular surfaces combining the two proposed transformations Y_S and Y_C according to the objects to be inspected and the possible defects to be detected (see Figure 6).

$$
\begin{aligned}
&(\exists Y_C)(\exists Y_S)(\exists \chi, \eta > 0)(\exists \rho_k^{m_i}, \rho_l^{m_l} \in A \subset P) \\
&(\exists \rho_o^{m_i}, \rho_q^{m_j} \in S_A)(\forall \rho^{m_i}, \rho^{m_j} \in P) \\
&F(\rho_k^{m_i}) = Y_C(F(\rho^{m_i})), F(\rho_l^{m_j}) = Y_C(F(\rho^{m_j})) \\
&F(\rho_o^{m_i}) = Y_S(F(\rho_k^{m_i})), F(\rho_q^{m_i}) = Y_S(F(\rho_l^{m_j})) \\
&(m_i \neq m_j) \wedge \left| \rho_k^{m_i} - \rho_l^{m_j} \right| \geq \eta \rightarrow \left| F(\rho_o^{m_i}) - F(\rho_q^{m_j}) \right| \geq \chi
\end{aligned}
\qquad (15)
$$
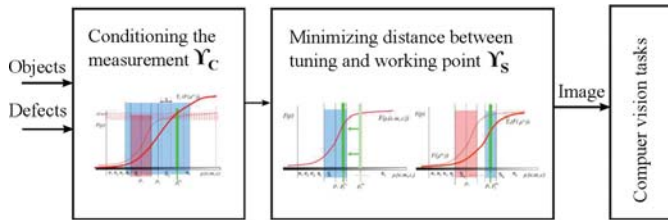
Figure 6. Method for inspecting specular surfaces. According to objects and defects, transformations Y_S and Y_C are selected to provide an image in which defects can be detected using computer vision techniques.

4. AVI Method Controlling Environmental Parameters

As it was previously shown, increasing the differences of the input magnitude values Δ_ρ can be performed by means of the control of the environmental conditions or of the camera parameters. These are two input parameters of the components ρ_i Equation (2) that contribute to the formation of the image I Equation (1) and are variables that can be affected by the system. The third input parameter, the object, is considered as a constant because it is the object to be inspected.

It is known that the camera, as a photoelectric transducer, provides a measurement related to the scene radiance (see Figure 7). It is a function of environment and object magnitudes. The contribution of interest to the radiance is the radiance coming from the inspected object. This radiance, L_R, is related to the object reflectance, f_R [66] and the irradiance E incident on the surface according to Equation (16).

$$L_R = f_R E i \qquad (16)$$

Reflectance f_R contributes the necessary information about the behaviour of the light interacting surface of the object. Irradiance, the second factor of Equation (16), is related to the environment variables, in fact, to be precise, to the electromagnetic radiation that reaches the surface of the object. These magnitudes affect the contribution of the object in the camera.

Controlling the luminous energy of the environment, a function of the lighting sources and the transmission modulations, is a way of working directly with input magnitudes of the system. Without

considering any other perception characteristics, irradiance is the key. Structuring the energy that reaches the object is a method of affecting the environment or the object. It enables areas of the surface of the object to be isolated, the contrast in the camera to be increased or decreased, etc. Thus, this variable allow us to force large differences at the input to perceive changes in the output image.

Figure 7. Parameters involved in the scene radiance L_R: object reflectance (f_R) and irradiance (E) incident on the surface.

Measurement Enhancement is the task of the transformation Y_C Equation (15). In this paper, we propose a instance Y_Φ of Y_C focused on lighting conditions of the environment. 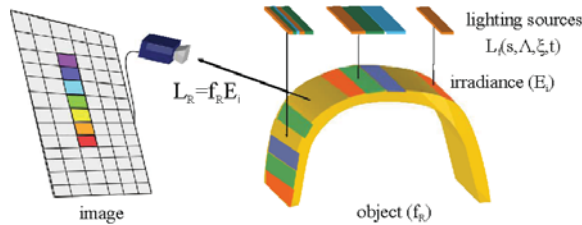The transformation Y_Φ structures the lighting in order to establish regions on the object radiated by different spectral powers Φ. Areas of different radiances are formed in this way. The regions can be formed by means of spatial modulation, forming a grid. Also, a sequence of lightings using temporal modulation or a combination of both can be used. The regions contribute with different radiances to the input of the camera establishing independent areas in the image. The increasing of the differences at the input of the system is produced in space or time domain Equation (12).

In this paper, the transformation Y_Φ is carried out by means of spatial modulation (space domain). The different areas in the image are formed with characteristics proportional to the irradiance E. The object irradiance E actively affects the perception process. The environment parameters are established so that great lighting gradients on the object are formed. The projection of the radiance L, which arrives at the optics, forms an image with regions. The projection of the radiance L, which arrives at the optics, forms an image with regions. The gradient of the image is a function of the one generated on the object. The greater the gradient of the pattern, the greater the gradient formed on the photodetectors. Then, a larger difference among adjacent photodetectors is obtained. Therefore, each photodetector has a spectral power associated in an instant (that is the function of irradiance and object characteristics).

Structuring the lighting enables the projection of a pattern on the surface object. This pattern is deformed by the object characteristics. It may be considered that the irradiance E is modulated by the object. Then, any other object modulates the generated pattern in a different way, and, therefore, modulates the spectral power, which is received in the space or the time by each photodetector, in a different way (in terms of inspection systems, any object with defects will modulate the generated pattern in a different way than the same object without defects). Controlling the input values of the perception system, for each photodetector, enables a reduction in the elements of the set of *scene magnitudes* that the system has to deal with. In addition, the pattern has to be configured so that the differences of the output are perceptible according to the magnitudes of the object (shape, colour, topology, etc.) to be perceived.

Control of energy that reaches the surface of the object is needed in order to design a certain pattern. The task depends on the number of environment lights, the spatial distribution of lighting, the wavelengths that conform each of the sources, the time, the modulations of transmission, etc. Moreover, the pattern of spectral power Φ could be different in order to inspect a specific magnitude μ_i of the object m. Then to this purpose, the transformation Y_Φ will determine the spectral power $\Phi \in P$ as a function of $\mu_i \in U$ and $m \in M$:

$$\Upsilon_\Phi : M \times U \to P \tag{17}$$

For practical considerations, it is interesting that the function of energy establishing the spectral power Φ is established in terms of the field radiance L_f by considering four parameters: $s \in S$, $\Delta \in D$, $\xi \in X$ and $t \in R$ (see Figure 8).

$$L_f : S \times D \times X \times R \to P \tag{18}$$

Hence, the transformation Υ_Φ could be defined by:

$$\Upsilon_\Phi(m, \mu_i) := L_f(s, \Delta, \xi, t) \tag{19}$$

The regions of lighting *ROL* on the surface of the object can be determined from the parameter *s*. It is a function that establishes the morphology of the regions of the pattern formed on the surface. Parameter Δ determines the set of lighting characteristics that radiates each of the established regions *ROL*. The function ξ determines the spatial configuration of energy reached by the object. The task of the function is to distribute the lighting characteristics of the set Δ over each of the regions determined by *s*. Finally, the function depends on time *t*. If the structured lighting is temporal, it is necessary to generate a sequence of patterns. Moreover, for practical reasons, we define *ROG* as the regions of lighting on the source. In the same way as *ROL*, the *ROG* is the morphology of the regions of the pattern conformed on the lighting source, in this case (see Figure 8).
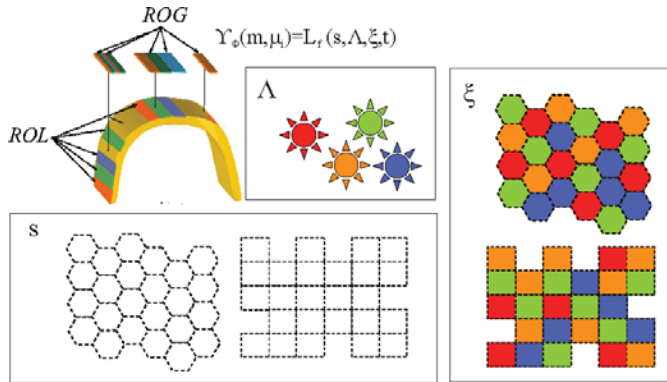
Figure 8. Parameters *s*, Δ, ξ of the transformation Υ_Φ.

5. Experiments

In this section, experiments performed by simulations to validate the model to deal with specular surfaces are presented. Specifically, the objective is to prove whether the transformation Υ_Φ is able to compensate for the lack of sensitivity that takes place in certain values of the magnitudes of the scene. In other words, probing whether the transformation *Measurement Enhancement*, by actively controlling the lightning pattern, is able to detect object characteristics (e.g., defects in visual inspection) that are not detected in other conditions. The experiments are mainly based on the scale of perception and the point of view as magnitudes of the scene.

We propose the use of simulation based on virtual imaging as an efficient way to perform a large battery of tests (different environment conditions and camera parameters) as a previous step of experimentation in real manufacturing environments, more complex in terms of instrumentation, and thus more expensive. Then, in this section, experiments performed by simulations to validate the model to deal with specular surfaces are presented.

5.1. Experimental Setup

An extensive explanation of the experimental setup is shown. It integrates the simulator, the subject of interest, and the environmental parameters that allow to design the lighting patterns.

5.1.1. Simulator

A simulator has been implemented considering the model presented in Section 2. It is able to synthesize images by $F(\rho)$ Equation (1). The components of ρ_i Equation (2) of the *scene magnitudes* (scale, viewpoint, light intensity, etc.) are modelled by the characteristics of the subject of interest, m, the environment, e, and, finally, the camera, c (see Diagram 1).

According to the characteristics μ_i of the subject of interest m, the simulator is basically interested in the reflection characteristics of the surfaces. The light reflection on an object depends on the atomic nature of the surface material and on the scattering and dispersion that takes place when the light contacts it. Consequently, in order to determine the reflected flow densities, the spatial configuration of the surface μ_S and its electromagnetic properties, which determine the type of material, are taken into account. These properties provide the refraction index μ_n of the surface, which represents the macroscopic optical properties required.

Regarding the environment e, the problem of specifying it is limited to the specific application of inspection. Therefore, in the process of transmitting the light signal from generating sources to the image, the different lighting sources are considered together with the modulations that the signal experiences, with the exception of those referring to the object to be inspected and those of the acquisition system. As environment variables, a distinction will be made between those related to lighting sources (that make up the fundamental element in the environment magnitudes) and the relative positions of the different elements involved in the scene. In order to carry out the study of local light reflection, the Bidirectional Reflectance Distribution Function (BRDF) presented by Cook and Torrance [67] has been used because different studies show its great capacity to adapt to reflectance extracted from objects [68].

Finally, variables related to the characteristics of the perception system include both the functions that define the optics and those related to the sensor itself. The aim of this research is to study the influence of environmental conditioning on the perception of the object. Then, controlled experimentation aimed to measure the radiance coming from the subject of interest is performed. Therefore, an ideal perception system is considered (images without noise, constant sensitivity for all spectral powers, constant diaphragm, etc.). It influences the radiometric measurement only with geometric variables like the focal distance and the size of the sensor. The other variables have been set to be constant.

5.1.2. Subject of Interest

Plane objects are considered in the experiments. They permit a simple control of the angle formed by the elements involved in the scene on all regions of the object. Control is necessary due to that fact that one of the aims of experimentation is to carry out a thorough test of the viewpoint.

The set M_P of the objects defined in the manufacturing specifications is made up of planes of 12 mm $\times$ 12 mm. The function μ_S establishes the points in a coordinate system that is local to the object independently of microscopic shape. The surface roughness is calculated by means of the Beckmann distribution model [69] by coherence with the BRDF model used. The root mean square is assigned to 0.1 (rms, or m in [67]). According to electromagnetic characteristics, there are two types of plane: dielectric and metallic. Dielectric planes have a refraction index μ_n of 1.6 and metallic ones (base on chromium) have a refraction index μ_n of 2.8 and an extinction index μ_{n_i} of 3.2 using the Fresnel equations. Also, other variables of the Cook-Torrance model such as a specularity coefficient (s in [67]) of 0.75 and a constant diffuse component (R_s) associated to a RGB of (0.6,0.6,0.6) have been considered.

The set M_P to be inspected is made up of objects with the same characteristics of M_P but including defects. Three types of defects have been distributed over these planes (see Figure 9): two changes in topography (a 0.6 mm-diameter crack or crater in Figure 9a,b and a change in colour (an area of 0.6 mm $\times$ 0.6 mm in Figure 9c). In this last case, a constant diffuse component R_s associated to a RGB of (0.4,0.4,0.4) on a surface measuring 0.6 mm $\times$ 0.6 mm is established. In total, for each plane in M_P, 3 planes (1 per type of defect) with 25 defects have been considered in M_P. Table 1 summarizes the data used in the experiments.
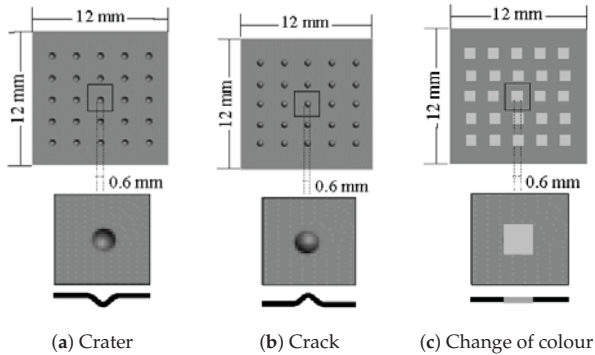

(**a**) Crater   (**b**) Crack   (**c**) Change of colour

Figure 9. Defects considered in the experiments.

5.1.3. Environment

Environment characteristics play an important role in validating the transformation Y_Φ Equation (19). Specifically, the goal is to study the transformation that configures the lighting of the scene to establish regions on the object surface radiated by different spectral powers by means of spatial modulation. Therefore, we have considered different spatial configurations of the energy that reaches the object surface according to different gradients by considering different s, Δ and ξ.

Since the objects are planes, the regions of lighting *ROL* formed by s are established on a plane. In the experiments, the parameters are established from the lighting source domain using the *ROG* (see Figure 8). The set of lights conform a grid so that the whole lighting extension can be defined for different conditions. Different areas of the regions of grid *ROG* are considered 0.1 mm $\times$ 0.1 mm, 0.2 mm $\times$ 0.2 mm, 0.3 mm $\times$ 0.3 mm and 0.6 mm $\times$ 0.6 mm to determine the influence of the size of the regions inside the areas of defects. Then, the ratios between the size of the region and the size of the defect in one of the dimensions (0.6 mm) are: 6 (0.6/0.1), 3 (0.6/0.2), 2 (0.6/0.3) and 1 (0.6/0.6).

The set Δ Equation (20) is made up of lighting sources using the same characteristics (polarization, power, etc.) except the wavelengths, δ_i, to conform different spectral powers, Φ_δ. The wavelengths used are from the visible electromagnetic spectrum. Also, lighting sources only radiate for specific wavelengths (monochromatic lights).

$$\Delta = \left\{ \Phi_{\delta_0}, \Phi_{\delta_1}, \Phi_{\delta_2}, ... \right\} \tag{20}$$

Finally, the function ξ determines the spatial configuration of energy emitted by the lighting. The lighting characteristics of the set Δ are distributed over each of the regions *ROL* determined by s establishing different gradients: spatial and amplitude. In the experiments, for practical considerations, the *ROG* composed of a squared grid has been used (Figure 10).

In this paper, four different configurations of lighting are considered: two for spatial gradients (ξ_x and ξ_{xy}) and two for amplitude gradients (ξ_L and ξ_I). Regarding the former two, the function ξ establishes spectral powers of Δ Equation (20) into two different spatial distributions. Specifically,

in the experiments the spatial gradient established by ξ is organized in one direction, ξ_x, and in two directions, ξ_{xy}, (see Figure 11). Let $ROG(x,y)$ be the region of column x and row y of the lighting grid and let N_x, N_y be the column and row of neighbouring regions of the grid. A function ξ will be defined as ξ_x if it only sets up different lighting characteristics in adjacent positions of an axis of the grid (Figure 10a) and the same ones in adjacent positions of the other axis of the grid. Then, any region in the grid ROG is assigned an element of the set Δ such that:

$$\xi_x\big(ROG(x,y),\Delta\big) = \Phi_{\delta_i} \in \Delta : \xi\Big(N_y\big(ROG(x,y)\big),\Delta\Big) = \Phi_{\delta_i} \wedge \xi\Big(N_x\big(ROG(x,y)\big),\Delta\Big) \neq \Phi_{\delta_i} \quad (21)$$

For the second spatial gradient, a function ξ will be defined as ξ_{xy} if sets up different lighting characteristics in all adjacent positions of a region of the grid (Figure 10b). Any region in ROG is assigned to an element of the set Δ Equation (20) such that:

$$\xi_{xy}\big(ROG(x,y),\Delta\big) = \Phi_{\delta_i} \in \Delta : \xi\Big(N_y\big(ROG(x,y)\big),\Delta\Big) \neq \Phi_{\delta_i} \wedge \xi\Big(N_x\big(ROG(x,y)\big),\Delta\Big) \neq \Phi_{\delta_i} \quad (22)$$

(**a**) ξ_x (**b**) ξ_{xy}
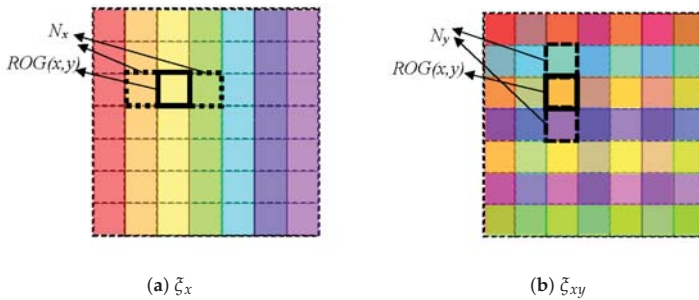
Figure 10. Spatial distribution considered for function ξ in the experiments.

Regarding the amplitude gradients, two configurations are also considered for ξ: *Linear*, ξ_L, and *Interlaced*, ξ_I. A function ξ will be defined as ξ_L if sets up lighting characteristics with close wavelengths in adjacent positions N of the regions of the lighting grid ROG (see Figure 11a) whereas the *Interlaced*, ξ_I, configuration maximize the differences among wavelengths in these positions (see Figure 11b). Then, in case of ξ_L, any region in ROG is assigned to an element of the set Δ Equation (20) such that:

$$\xi_L\big(ROG(x,y),\Delta\big) = \underset{\Phi_{\delta_i}}{\operatorname{argmin}}\Big(\xi\big(N(ROG(x,y)),\Delta\big) - \Phi_{\delta_i}\Big) \text{ subject to } \sum \xi(N(ROG(x,y)),\Delta) - \Phi_{\delta_i} > 0 \quad (23)$$

In case of any function that is considered as an *Interlaced* function, ξ_I, the ROG is assigned such that:

$$\xi_I\big(ROG(x,y),\Delta\big) = \underset{\Phi_{\delta_i}}{\operatorname{argmax}}\Big(\xi\big(N(ROG(x,y)),\Delta\big) - \Phi_{\delta_i}\Big) \quad (24)$$

(**a**) Linear ξ_L (**b**) Interlaced. ξ_I

Figure 11. Amplitude distribution considered for function ξ in the experiments.

Different transformations Y_Φ are formed using combinations of functions ξ accomplishing different spatial and amplitude gradients: *Linear X* ($\xi_{L,x}$), *Interlaced X* ($\xi_{I,x}$), *Linear XY* ($\xi_{L,xy}$), and *Interlaced XY* ($\xi_{I,xy}$) to carry out the experiments (see Table 2).

The function *Linear X* ($\xi_{L,x}$) is made up by using a function meeting a spatial gradient in one axis of the grid Equation (21) and an amplitude gradient formed by *Linear* configuration Equation (23). This function uses an ordered sequence Δ^L of the set Δ. Let Φ_{δ_i} be an element of Δ that radiates energy with the wavelenght δ_i. Let δ^i and $\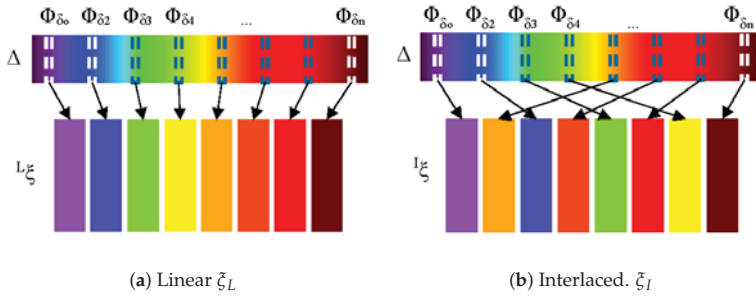delta^e$ be the minimum and maximum values and let δ^n be the number of wavelengths considered in the sequence, then Δ^L is:

$$\Delta^L = \left[\Phi_{\delta_0}, \Phi_{\delta_1}, \Phi_{\delta_2}, ..., \Phi_{\delta_{n-1}} \right] : n \geq 2, \delta^e \geq \delta^i, \delta_i = \delta^i + mod(i, \delta^n) \left(\frac{\delta^e - \delta^i}{\delta^{n-1}} \right) \tag{25}$$

An element of the sequence Δ^L is assigned to the region (x,y) of the grid lighting *ROG* by the function ξ^x:

$$\xi^x(ROG(x,y), \Delta^L) = \Delta^L\big(mod(x, n)\big) \tag{26}$$

The differences of wavelengths among neighbouring regions are constant in one of the axes of the grid.
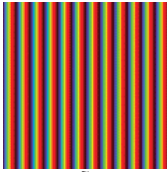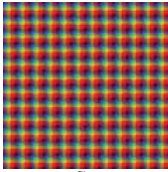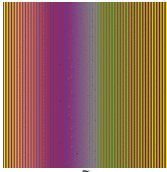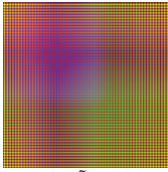
The function *Interlaced X*($\xi_{I,xy}$) is made up by a function that accomplishes a spatial gradient in one axis of the grid Equation (21) and an amplitude gradient formed by the '*Interlaced*' configuration Equation (24). In this paper, this function uses an ordered sequence Δ^E of the set Δ in which a maximum difference of the wavelengths between adjacent positions is established. Let Δ_t and Δ_b be the top half and bottom half of Δ^L Equation (25).

$$\begin{aligned} \Delta_t &= \left[\Phi_{\delta_0}, \Phi_{\delta_1}, ..., \Phi_{\delta_i - 1} \right], \\ \Delta_b &= \left[\Phi_{\delta_i}, \Phi_{\delta_{i+1}}, ..., \Phi_{\delta_{n-1}} \right] : \\ i &= \lceil n/2 \rceil, \Delta_t \cup \Delta_b = \Delta^L \end{aligned} \tag{27}$$

Table 1. Characterists of the objects to be generated in the experiments.

Characteristic	Metallic	Dielectric
Surface roughness (rms)	0.1	
Refraction index	2.8	1.6
Extinction index	3.2	0
Specularity coefficient	0.75	
Diffuse component (RGB)	(0.6,0.6,0.6)	
Diffuse component defect (RGB)	(0.4,0.4,0.4)	
Object size (mm)	12×12	
Defect size (mm)	0.6 diameter or side	

Table 2. Lighting configuration using combinations of the spatial and amplitude gradients.

	Spatial Gradient	
	X (ξ_x)	**XY (ξ_{xy})**
Amplitude gradient Linear (ξ_L)	$\tilde{\xi}_{L,x}$	$\tilde{\xi}_{L,xy}$
Interlazed (ξ_I)	$\tilde{\xi}_{I,x}$	$\tilde{\xi}_{I,xy}$

Then, the sequence Δ^E combines Δ_t and Δ_b in an interlaced manner:

$$\Delta^E = [\delta_0, \delta_1, ..., \delta_{n-1}] :$$
$$\delta_i = \left\{ \begin{array}{ll} \Delta_t(\lfloor i/2 \rfloor) & \text{if } (i \bmod 2) = 0 \\ \Delta_b(\lfloor i/2 \rfloor) & \text{if } (i \bmod 2) = 1 \end{array} \right] \tag{28}$$

An element of the sequence Δ^E is assigned to the region (x,y) of the grid by the function ξ^x Equation (26).

The function '*Linear XY*' accomplish a spatial gradient in two axes of the grid Equation (22) and an amplitude gradient formed by the '*Linear*' configuration Equation (23). This function uses the ordered sequence Δ^L Equation (25) of the set Δ. An element of the sequence Δ^L is assigned to the region (x,y) of the lighting grid by the function ξ^{xy} :

$$\xi^{xy}(ROG(x,y), \Delta) = \frac{\Delta(x \bmod n) + \Delta(y \bmod n)}{2} \tag{29}$$

Finally, the function '*Interlaced XY*' is defined as meeting a spatial gradient in two axes of the grid Equation (22) and an amplitude gradient formed by the *Interlaced* configuration Equation (24). This function uses the ordered sequence Δ^E Equation (28) of the set Δ. An element of the sequence Δ^L is assigned to the region (x,y) of the grid by the function ξ^{xy} Equation (29).

Also, a reference lighting configuration is defined. It permits the comparison of the improvement produced by enhancing the target parameters to measure. In this case, all the regions of the grid have the same characteristics. The set Δ is built with one element (monochromatic or polychromatic lights).

$$\Delta = \delta, \ \xi(ROG(x,y)) = \delta \tag{30}$$

The parameters used in the transformations Y_Φ are summarized below (see Table 3). The lighting covers the whole surface (12 mm × 12 mm). The areas of *ROL*s generated by *s* are 0.1 mm × 12 mm, 0.2 mm × 12 mm, 0.3 mm × 12 mm and 0.6 mm × 12 mm for the functions '*Linear X*' and *Interlaced X*. The '*Linear XY*' and '*Interlaced XY*' functions from *ROL* have areas of 0.1 mm × 0.1 mm, 0.2 mm × 0.2 mm, 0.3 mm × 0.3 mm and 0.6 mm × 0.6 mm. According to lighting characteristics Δ, '*Linear X*' and '*Linear XY*' use the sequence Δ^L Equation (25) with 10 elements for all cases whereas *Interlaced X*

and '*Interlaced XY*' use the sequence Δ^E Equation (28) with 120, 60, 40 and 20 wavelengths uniformly distributed from the visible electromagnetic spectrum (380 nm to 780 nm). Finally, the function ξ^{xy} is used for '*Linear XY*' and '*Interlaced XY*'. It establishes (120 × 120), (60 × 60), (40 × 40) and (20 × 20) *ROL*s on the plane. '*Linear X*' and *Interlaced X* are made up using the function ξ^x establishing 120, 60, 40 and 20 *ROL*s according to the areas of the regions considered. Figure 12 shows the different lighting patterns used in the experiments.

Table 3. Parameters for the transformations Y_Φ used in experiments.

Characteristics	Linear X	Linear XY	Interlazed X	Interlazed XY
s: *ROL* areas (mm)	0.1 × 12 0.2 × 12 0.3 × 12 0.6 × 12	0.1 × 0.1 0.2 × 0.2 0.3 × 0.3 0.6 × 0.6	0.1 × 12 0.2 × 12 0.3 × 12 0.6 × 12	0.1 × 0.1 0.2 × 0.2 0.3 × 0.3 0.6 × 0.6
s: *ROG* areas (mm)		0.1 × 0.1 0.2 × 0.2 0.3 × 0.3 0.6 × 0.6		
Δ (# wavelenghts)	10		120 60 40 20	

Figure 12. Samples of lighting patterns generated by the transformation v_Φ used in the experimentation.

5.2. Performance Results

In order to obtain the performance results, the components ρ_i Equation (2) considered are scale ρ_E, angle ρ_θ and intensity lighting ρ_I. They are the most influential scene variables in the image formation and in the characteristics of the visual inspection systems.

The system is tuned to a set of scale values ρ_E: 1 pixel/mm, 2 pixels/mm, 5 pixels/mm, 10 pixels/mm and 15 pixels/mm. The set of the component angle ρ_θ, as the angle formed by the surface normal and the camera normal, is formed by the sequence from 0° to 90° with an increase of 10° (10 different angles in total). Finally, the intensity lighting ρ_I is the basic parameter measured by the camera and is related to the solution proposed for increasing the perception capacity of the system. A large set of camera, environment and object variables take part in the formation of this magnitude. The goal of the tests is to study the intensity lighting according to environment values related to the proposed transformations Υ_Φ and not to influence any other variables. However, the angle formed by the surface normal and the lighting plane normal is very important in the design of inspection systems. It affects the intensity lighting ρ_I. Therefore, it is considered as this parameter. The values contemplated in the tests are from 0° to 90° with differences of 10° (10 different angles in total).

In order to measure the effectiveness of the transformation Υ_Φ to inspect, the performance is calculated as the number of different pixels between the image of an object without defects and the image of the same object with defects. Specifically, this difference divided by the estimated number of pixels containing the defects, for a specific resolution, measures the success rate. The calculation of this rate is performed in all the possible combinations of the variables ρ_i contemplated previously. A total of 68,000 images have been synthesized. That is, 17,000 images from objects without defects M_P and 51,000 (17,000 × 3 types of defects) images from inspection objects M_I with 3 types of defects (2 in topography and 1 in colour). The 17,000 images are the images of 17 lighting configurations (4 Υ_Φ functions of environment condition using 4 different lighting areas defined by *s* and a reference lighting configuration Equation (30)) conditioning the measure of 2 objects (dielectric and metallic) for 500 values of vector ρ (5 ρ_E, 10 ρ_θ and 10 ρ_I).

The experimental results of the average success rates for scale ρ_E, angle ρ_θ, lighting ρ_I and the influence of the size of lighting *s* regions are detailed in the next sections.

5.2.1. Scale of Perception

The study of the perception scale makes discerning the influence of the size of the defects in the image possible.

Figure 13a shows the success rates according to the dielectric object. The function '*Interlaced XY*' offers the best success rates whereas the '*Linear X*' offers the minimum of the transformations Υ_Φ. The average difference between the best function and the reference is 4.33%. The data from functions ('*Linear XY*' and '*Interlaced X*') that uses only one of the maximum proposed gradients, scale or amplitude, is very similar. The average difference is only 0.16%.

The differences in success rates according to metallic object are more noticeable (see Figure 13a). The function '*Interlaced XY*' shows similar success rates to the case of dielectric material. The values are practically independent of the type of material. However, the success rates of the reference lighting significantly decrease to values between 46% and 57.9%. This is an average difference of 5.85% lower than the success rate of the dielectric case. Therefore, the improvement of the capacity of perception of the system is better using the alternative *Measurement Enhancement*. It differs more than 10% using the best function (*Interlaced XY*). Also, the results show an increase in sensitivity using spatial distributions organized in two directions (ξ^{xy}).

It is interesting to consider the shape of the curve in Figure 13a. The graph represents the success rate of the system as a function of the scale. This is pixels per millimetre and not pixels per defect. The defects used in the tests have a maximum size of 0.6 mm for one of their three dimensions. Therefore, the scale values ρ_E correspond to 0.6, 1.2, 3, 6 and 9 pixels per dimension of the defect. For the minimum scale (1 pixel/mm), a crack or crater defect is projected into 0.28 pixels^2 and a chromatic defect is projected into 0.36 pixels^2. This corresponds to an area of 7.065 and 9 pixels^2 respectively in the image of a defective plane (25 defects, see Figure 9). Then, the differences of 1 pixel in the image suppose that success rates vary 11.11% or 14.15%. The ratio of success rate to pixel is very large. Conditions for the next scale (2 pixels/mm) are similar. In this case, differences of 1 pixel in the

image means the detection or not of a defect. The topographic defects are projected into an area of 1.13 pixels^2 and the chromatic ones into an area of 1.44 pixels^2. Differences of 1 pixel for the rest of the scale values mean a lesser impact on the detection.
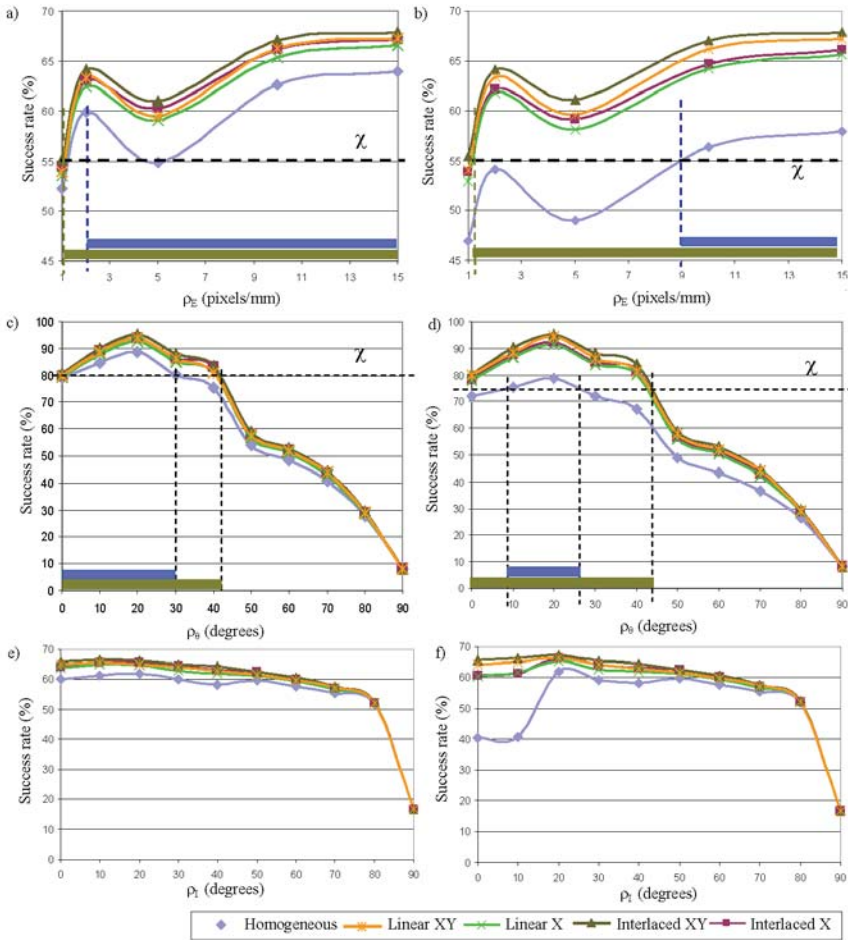
Figure 13. Success rates for the dielectric (**a**,**c**,**e**) and metallic (**b**,**d**,**f**) material according to scale (**a**,**b**), angle (**c**,**d**) and intensity lighting (**e**,**f**).

In short, the average success rates show higher scale ρ_E, more capacity of perception of the defects both for dielectric and metallic materials, avoiding short scales (1 pixel/mm and 2 pixels/mm) due to the characteristics mentioned in the previous paragraph.

Comparing the functions Y_Φ and the reference function, according to the capacity of perception, the data shows that a greater adjustment of the scale is required using the latter lighting. For example, if the minimum threshold χ for determining the defect in the dielectric material (this parameter will depend on the type of application) is 55%. Using the reference lighting, approximately more than 2 pixels/mm are necessary to perceive the defects whereas with '*Interlaced XY*' lighting only 1 pixel/mm is needed. In the case of metals, greater adjustment of the system is required. For the same threshold χ, more than 9 pixels/mm would be necessary using reference lighting and only 2 pixel/mm

for each of the functions Y_Φ proposed. The best lighting, '*Interlaced XY*', makes a system tune to only 1 pixel/mm to detect the possible defects.

5.2.2. Angle of Perception

The success rates of the different transformations Y_Φ according to the perception angle ρ_θ are shown in Figure 13c,d, for the dielectric and metallic objects respectively.

The study of the average of the success rates shows that the best conditions to perceive according to ρ_θ are in the interval $[0°, 40°]$ for all lighting configurations. The function '*Linear X*' offers the worst results of the transformations Y_Φ even though their differences are minimum, an average of 2.85%. The functions '*Linear XY*' and '*Interlaced X*' present a similar behaviour. The average difference between them is 0.77% and it reaches a maximum of 1.12% for an angle of 20°.

The best results are provided by the function '*Interlaced XY*'. The improvement in the capacity of perception reaches the maximum differences, an average difference of 11.4%, in the interval $[10°, 40°]$ compared to using the reference lighting. According to the material type in this interval, the differences are 6.84% and 16.05% for the dielectric and metallic case respectively. This is due to the fact that the success rates significantly decrease from inspecting dielectrics materials to metallic one, an average of 8.99% in the interval. Also, the differences in the success rates between the transformations Y_Φ that light the metallic objects are more noticeable.

An angle of 20° formed by the surface normal and the camera normal sets the maximum success rate. Using the function '*Interlaced XY*' for this angle, similar results are obtained according to the type of material: 95.05% for dielectric materials and 95.19% for metallic one. Furthermore, for the metallic object in this angle, the difference between the '*Interlaced XY*' and the lighting reference is 16.53%.

The success rate and the differences between the transformations and reference lighting decrease from 40° to reach 8.2% at 90°. In the latter case, it is only possible to detect convex defects.

Comparing the functions Y_Φ and the reference function, according to the capacity of perception, the data shows that the increase in the number of the magnitudes of a scene which the system is able to perceive is more significant than in the study of the scale. For example, if the minimum threshold χ for determining the defect in the dielectric material is 80% using the reference lighting, the angle ρ_θ must be less than 30° to perceive the defects. In this case, using the '*Interlaced XY*' lighting, the angle can increase by up to approximately 43°. Hence, the increase is more than 10°. Assuming a threshold of 75%, the angle formed by the surface normal and the camera normal can be established in the interval $[10°–25°]$ for determining the defect in the metallic material using the reference lighting. Using any transformation Y_Φ, the angle tuning can be established in the interval $[0°–45°]$. Then, the increase is about 30°.

5.2.3. Intensity Lighting

Regarding the intensity lighting ρ_I as the angle formed by surface normal and lighting plane normal, the success rates of the different lighting configurations are shown in the Figure 13e,f, for the dielectric and metallic objects respectively. The function '*Interlaced XY*' provides the maximum success rate and the greatest increase in the capacity of perception of the system.

The study of the inspection capacity considering the material type shows the limited sensitivity for perceiving metallic objects using angles ρ_I close to 0° and the reference lighting. The characteristics of the lighting sources and of the metallic reflection cause sensor saturation in the angle interval $[0°, 10°]$. The decreasing of the success rate, compared to the dielectric material, is considerably marked with values close to 20% (concretely 19.72% for 0° and 20.77% for 10°). Also, transformations using minimum spatial differences show differences of more than 3% between metallic and dielectric in that interval: '*Interlaced X*' 3.30% (0°) and 4.36% (10°), '*Linear X*' 3.47% (0°) and 3.30% (10°). However, if the transformations with spatial distributions in two dimensions are used, sensor saturation is not important. The differences are 0.33% (0°) and 0.45% (10°) for the function '*Interlaced XY*' and 0.73%

(0°) and 0.37% (0°) for 'Linear XY'. The success rate is independent of the material and the differences are not significant for the rest of the lighting angles.

The success rate shows a low gradient for the intensity lighting ρ_I according to lighting angle. The magnitude has an almost constant behaviour in the interval [0°, 60°]. The standard deviation is lower than 2 for the transformations Y_Φ whereas the value is 4.21 for the reference lighting for all angles due to the behaviour in the initial values of the metallic plane (the standard deviation for the dielectric plane is 1.44).

5.2.4. Size of Lighting Regions

The influence of the size of the lighting regions ROL defined by s on the capacity of perception according to the scale is indicated on the left of Figure 14 and according to the angle on the right of the Figure 14. The size of ROL is represented by the minimum dimension of the region.
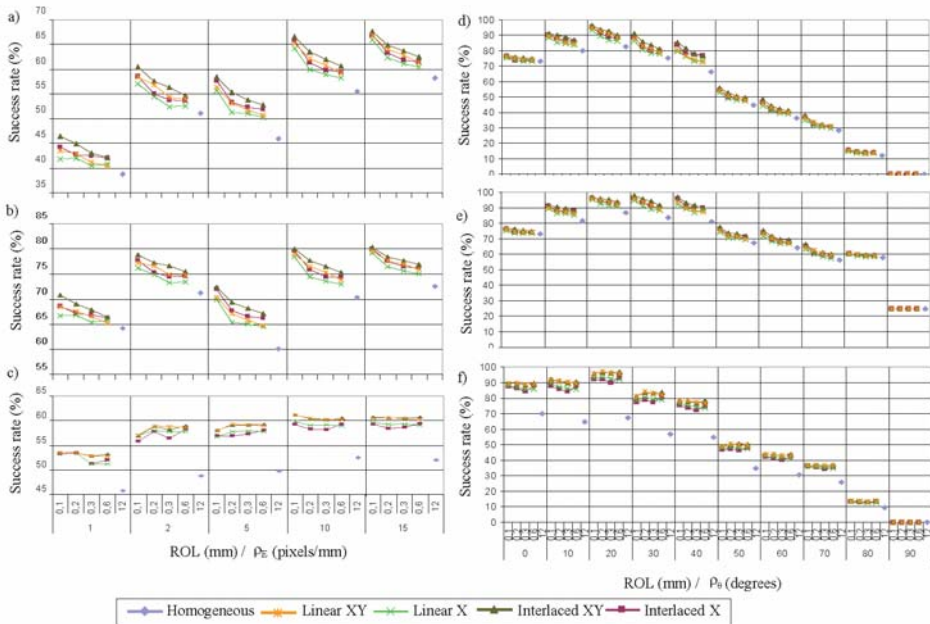
Figure 14. Success rates for inspecting a crater (**a**,**d**), crack (**b**,**e**) and chromatic defect (**c**,**f**) according to size of lighting regions and scale (**a**–**c**) and angle of perception (**d**–**f**).

Increasing the size of ROL decreases the rate of the regions inside the defect. This fact decreases the success rates of inspecting topographic defects using the transformations Y_Φ for all perception scales and angles. The average differences vary from 4% to 5% in the scale case. According to angle, the average differences are about 7% for the crack defect (Figure 14e) in the interval of the maximum sensitivity [10°, 40°]. The behaviour for the crater defect (Figure 14d) is analogous; it establishes an average difference of sensitivity in that interval of 5%. The differences decrease as the perception angle increases until they are 0 for 90°.

The capacity to modulate irradiance in the defect is inversely proportional to its size. In the extreme case where the size of the bands is greater than the size of the defects, the irradiance function will not be able to modulate the reflectance of the object, so it would behave in the same way as a the reference lighting.

The analysis of the chromatic defect shows that the size of lighting regions is independent of the success rate (see Figure 14c,f). In this case, it is practically constant. Therefore, a considerable increase in the perception capacity of the system is obtained.

6. Case Study: Automobile Logo

Simulation, as a previous step of experimentation in real manufacturing environments, allows a preliminary study that can be used to discern the best acquisition path planning or the equipment characteristics to perform the inspection of the specular surfaces. Finally, in this section an example of the application of the experimental results of the scale ρ_E and angle ρ_θ of perception on specular surface acquisition for inspecting a 100 mm diameter Mercedes Benz metallic logo is shown. The area to be inspected is 4676.15 mm^2. Given that the minimum threshold χ depends on the application, in this case a value of 80% is assumed.

The inspection conditions are restricted to only 5 values of the scene magnitudes ρ_i for a metallic object with an unstructured lighting (reference lighting). As can be seen in Figure 15 considering scale and angle (ρ_θ, ρ_E), the scene magnitudes are the following: (0°, 2 pixels/mm), (10°, 10 pixels/mm) and those with the considered highest scale ([0°, 10°, 20°], 15 pixels/mm). The best lighting configuration, Y_Φ '*Interlaced XY*', permits increase the tuning of the scene magnitudes system up to 20 points (see Figure 15): ([20°, 40°], 1 pixel/mm), ([0°, 10°, 20°, 30°], 2 pixels/mm), ([10°, 20°, 30°, 40°], 5 pixels/mm), ([0°, 10°, 20°, 30°, 40°], [10, 15] pixels/mm).
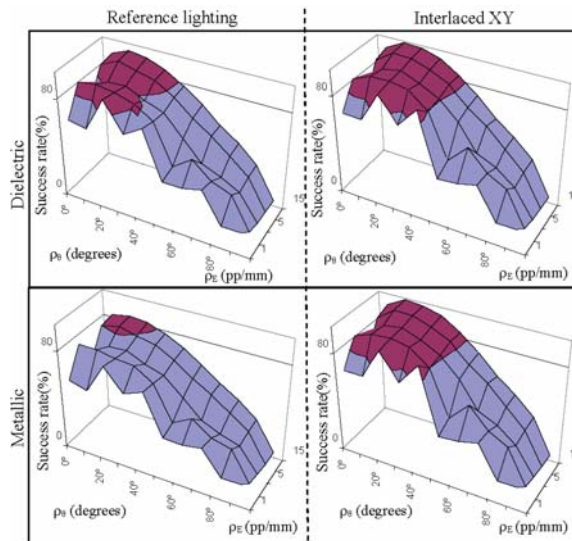
Figure 15. Success rates for the dielectric and metallic surfaces according to lighting of reference and the transformation Y_Φ with best results ('*Interlaced XY*').

The choice of the conditions (scale, angles, etc.) to capture the whole object for inspection is a complex problem. It is necessary to take into account the particularities of each solution. In this case as example, the scale ρ_E is established at 15 pixels/mm. Therefore, according to the previous scene magnitudes, the perception angle can deviate by up to 20° using the reference lighting and by up to 40° using the function '*Interlaced XY*'. In other words, the angle formed by the vector normal to the camera and the normal vector of the surface to be inspected must be from 0° up to 20° using the reference ligthing and from 0° up to 40° in case of '*Interlaced XY*' lighting is used.

In consequence, in order to apply the results, any point on the surface can be viewed as a point on a plane whose normal vector is the normal of the surface at that point. For example, Figure 16a shows an outline of this assumption. In this way, the surface can be analysed as a set of planes (a plane per point on the surface). Then, the experimental results (calculated for planes) can be extrapolated to calculate the perception scale and angle of surfaces of different curvature for any point on the surface. The use of planes for any point on the surface could be computationally expensive. Hence, as the function μ_S establishes the points on the surface in a coordinate system that is local to the object, in practice the function μ_S is defined as a triangle mesh: a collection of triangles that defines the surface shape of a polyhedral object in 3D computer graphics. The use of a triangle mesh allows the system to discretize the surface geometry as a reduced collection of planes. The number of planar faces will be determined by the geometry of the surface and the resolution used in the experiments (in this case more than 4000 polygons, although less triangles are enough, high details are not needed due to angle perception is discretized each 10°).
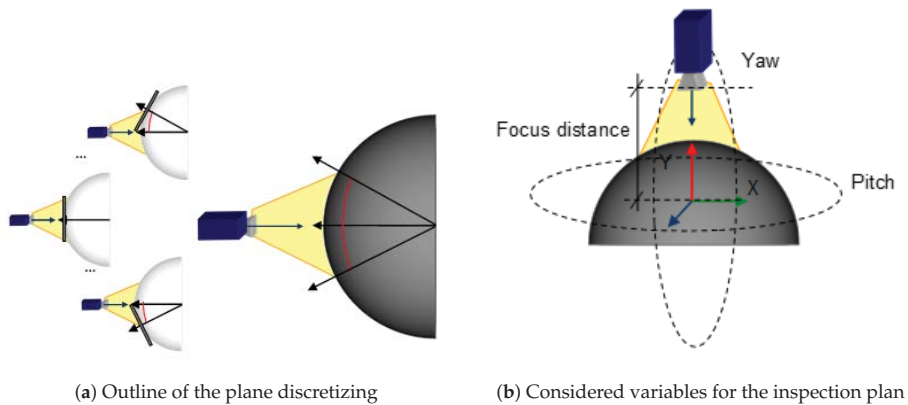

(**a**) Outline of the plane discretizing   (**b**) Considered variables for the inspection plan

Figure 16. Essential considerations for transferring the conclusions of experimental values to a specific inspection of an object.

Since the scene magnitudes are a function of the characteristics of the object, the environment and the camera, it is required to make decisions about the appropriate magnitudes to infer to provide them. In other words, it is necessary to decide which magnitudes of the environment or the camera have to be modified to establish the adequate values of scale and angles calculated in the experiments. The scale will be determined by the number of pixels available to the sensor by setting the camera position at a focus distance and at a constant focal length. The conditions of the angle of perception will be established by the movement on the X and Y axis of the origin of coordinates located in the center of the of the object: the Yaw and Pitch movements of the camera (see Figure 16b). Due to select the variables is a complex problem, an approximation to the optimum solution is proposed in order to determine the appropriate angles between camera and object surface. This increases the captured area of the logo and reduces the images to be captured. For this purpose, a search tree was designed using a branch and bound algorithm, in which the solutions space of each node is reduced to a maximum of five children and a maximum temporal processing level is established.

The logo inspection requires 26 captures (see Figure 17a and Table 4) using the lighting reference and a camera of 1452×1452 pixels to acquire a surface area of 4619.96 mm^2 (98.79% of the logo). Some captures can be made using a lower resolution of up to 510×510 pixels. According to the function '*Interlaced XY*', only 9 images are necessary (see Figure 17b and Table 5). The camera resolution varies between 826×826 and 1455×1455 in order to cover a total surface area of 4675.05 mm^2 (99.9%). The first capture obtains 45% of the total inspection of the logo whereas only 8.17% is obtained using

the reference lighting. This fact shows how the lighting system allow to perceive in more scene magnitudes. Specifically, more points on the surface accomplish the angle of perception ρ_θ.

Table 4. Environment and camera characteristics to acquire 26 images needed to inspect the metallic logo using the reference lighting.

Yaw/Pitch (degrees)	CCD (pixels)	Image Area (mm^2)	Inspected Area (mm^2)	Accumulated Area (mm^2)	Inspected (%)
(0, 0)	1452 × 1452	9377.45	381.97	381.97	8.17%
(30, −30)	1218 × 1218	6588.08	769.21	1151.19	24.62%
(0, 40)	1186 × 1186	5982.37	761.16	1912.35	40.90%
(−30, −30)	1215 × 1215	6503.9	758.21	2670.56	57.11%
(40, 20)	1167 × 1167	5691.68	222.63	2893.19	61.87%
(−40, 20)	1166 × 1166	5680.73	220.89	3114.07	66.60%
(0, −80)	550 × 550	707.26	143.46	3257.53	69.66%
(−60, 70)	570 × 570	1038.1	142.98	3400.51	72.72%
(30, 80)	510 × 510	630.42	142.09	3542.6	75.76%
(80, 40)	519 × 519	1127.5	141.35	3683.95	78.78%
(−60, −70)	570 × 570	1015.71	140.94	3824.89	81.80%
(−20, 80)	493 × 493	620.97	138	3962.89	84.75%
(0, −40)	1137 × 1137	4144.17	137.76	4100.66	87.69%
(50, −80)	543 × 543	615.36	131.11	4231.77	90.50%
(−30, −80)	560 × 560	705.12	88.05	4319.82	92.38%
(−80, −20)	576 × 576	1075.71	67.33	4387.15	93.82%
(50, 60)	635 × 635	1744.96	47.76	4434.91	94.84%
(0, 80)	550 × 550	695.4	45.85	4480.76	95.82%
(20, −70)	526 × 526	1213.01	32.72	4513.49	96.52%
(10, 30)	1280 × 1280	6672.88	31	4544.48	97.18%
(60, −60)	630 × 630	1538.01	29.37	4573.85	97.81%
(−30, 40)	1133 × 1133	5565.17	25.12	4598.97	98.35%
(−50, 70)	586 × 586	1177.92	11.57	4610.55	98.60%
(40, −30)	1164 × 1164	5933.79	4.83	4615.37	98.70%
(−40, −20)	1192 × 1192	6175.95	4.41	4619.79	98.80%
(10, 60)	749 × 749	1970.75	0.17	4619.96	98.80%

Table 5. Environment and camera characteristics to acquire nine images needed to inspect the metallic logo using '*Interlaced XY*' lighting.

| Yaw/Pitch (degrees)	CCD (pixels)	Image Area (mm^2)	Inspected Area (mm^2)	Accumulated Area (mm^2)	Inspected (%)
(0, 0)	1455 × 1455	9410.15	2079.07	2079.07	44.46%
(−10, −60)	1219 × 1219	4252.76	581.96	2661.04	56.91%
(−10, 60)	1221 × 1221	4234.16	580.29	3241.33	69.32%
(60, 10)	1263 × 1263	4574.35	560.57	3801.9	81.30%
(−70, 0)	1174 × 1174	2899.19	480.71	4282.61	91.58%
(40, −60)	1187 × 1187	4118.58	197.45	4480.06	95.81%
(30, 50)	1279 × 1279	5563.56	121.89	4601.95	98.41%
(−80, 50)	826 × 826	2711.22	36.55	4638.5	99.20%
(−80, −70)	923 × 923	1695.65	36.55	4675.05	99.98%
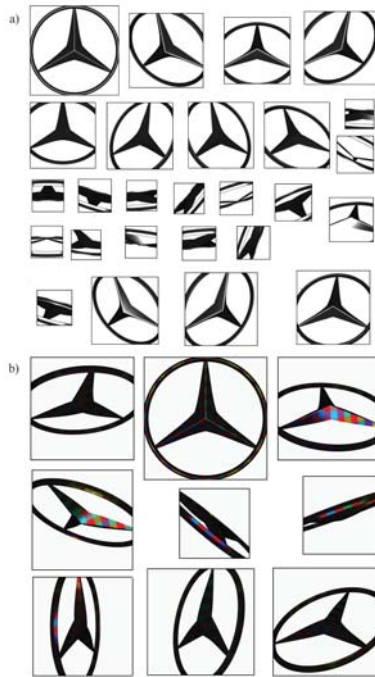
Figure 17. Images of a metallic logo to be inspected using the reference lighting (**a**) and the function *'Interlaced XY'* (**b**).

7. Conclusions

A novel active imaging model able to increase the perception capacity of the visual system is presented. The model provides solutions to vision problems in which it is difficult to perceive. It describes the parameters involved in a visual sensorization system in three main aspects: target object to perceive, environmental conditions, and sensor parameters. Each of them is individually parameterized as object size, color, etc.; environment parameters as light, objects-sensor orientations,...; and sensor parametrization such as focal length, monochrome/RGB/3D, etc. Moreover, the model describes the perception capabilities and the limits, and present solutions to these limits. In particular, the model is instantiated for the specular object inspection, as an example of a challenging situation for visual sensor perception. The specular surfaces problem means the device must operate in the intervals of low perception, related to reflections and shine. Traditionally, automated visual inspection requires a thorough analysis of the problem where solutions include everything from the acquisition equipment to the algorithm to recognize the possible defects. As a consequence, these vision systems are oriented to concrete applications and cannot be generalized. The model presented here deals with this lack by providing a general representation of vision systems and solutions for the perception limitations.

The solution proposal of the problem related to specular surfaces provides a normalization of the image in which different objects perceived as the same can be distinguished. First, *Measurement Enhancement* increasing the differences of the input magnitudes is performed. In this paper, the enhancement of measurements is carried out using environment variables, concretely controlling the lighting conditions. This enhancement spatially structures the lighting in order to set up regions on the surface of the object radiated using different spectral powers forming a grid. Finally, the system is tuned in order for the magnitudes of the scene (scale, angles, intensity lighting, etc.) to be properly perceived.

According to thorough analysis of the problem characteristics, use of virtual imaging simulations as a preliminary method step for validating hypotheses of the visual inspection systems is proposed. The validation of the conditions (point of view, scale, lighting, etc.) in which the inspection has to be performed, can be carried out in a flexible and low cost manner justifying the use of simulations in an early warning step of the system viability. Hence, the model can be pre-validated before the system is developed.

The use of simulations and knowledge bases and the generalist approach of the transformation provide a general solution that can be systematically applied. The method can be applied to the resolution of different inspection problems adapting the contents of the knowledge bases, avoiding a new design solution for each problem.

A realistic simulator has been designed to carry out the experimentation. This simulator recreates the conditions of the image formation and permits the validation of the inspection systems based on the model. The test shows the use of the transformations Y_Φ improves the capacity of perception of the system compared to a homogeneous environment (only one wavelength or colour). This is both for dielectric and metallic materials with regard to the tuning of the scale, angle or lighting. The use of transformation considering maximum amplitude gradient and maximum spatial differences ('*Interlaced XY*') obtains the best results to perceive the defects for all cases. In contrast, the function with minimum amplitude gradient and minimum spatial differences ('*Linear X*') offers the worst results of the transformations considered to perform the *Measurement Enhancement*. The functions ('*Linear XY*' and '*Interlaced X*') that use only one of the maximum proposed gradients, scale or amplitude, generally present similar behaviour. The function that performs a homogeneous lighting of the object surface, which is used as reference, obtains the minimum success rate in all cases. Hence, the results prove the improvement in the capacity of perception for different conditions of scale, angle and intensity lighting. The proposed method enables the detection of surface defects in a greater number of values of scale, angle of perception and lighting conditions than in normal conditions using uniform lighting. The immediate repercussion is that a smaller number of captures of the scene is needed by the system.

The research should continue by studying the *Measurement Enhancement* by means of the extension of the input magnitudes: using other transformations based on structured lighting (with different patterns and using the time domain) and using other parameters like variables of the capture system to provide great gradients in the image.

The simulation confirms the hypotheses. It would be wise to advance to physical experiments according to the concrete industry in which the system is to be technologically developed. Nowadays, the '*Interlaced XY*' function (in this case, the pattern is made up using the ordered sequence Δ^E of grey levels instead of using wavelengths) is being tested for increasing the perception capacity of an inspection system aimed to detect shape defects on the surfaces of ceramic tiles.

Author Contributions: J. Azorin-Lopez, A. Fuster-Guillo and J.M. Garcia-Chamizo conceived the model and designed the experiments; J. Azorin-Lopez and M. Saval-Calvo performed the experiments; J. Azorin-Lopez, H. Mora-Mora and J.M. Garcia-Chamizo analyzed the data; All authors wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Oren, M.; Nayar, S.K. A Theory of Specular Surface Geometry. *Int. J. Comput. Vis.* **1997**, *24*, 105–124.
2. Jin, H.; Soatto, S.; Yezzi, A.J. Multi-View Stereo Reconstruction of Dense Shape and Complex Appearance. *Int. J. Comput. Vis.* **2005**, *63*, 175–189.
3. Kim, S.; Kweon, I.S. Automatic model-based 3D object recognition by combining feature matching with tracking. *Mach. Vis. Appl.* **2005**, *16*, 267–272.
4. Klinker, G.; Shafer, S.A.; Kanade, T. The measurement of highlights in color images. *Int. J. Comput. Vis.* **1988**, *2*, 7–32.

5. Bajcsy, R.; Lee, S.W.; Leonardis, A. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *Int. J. Comput. Vis.* **1996**, *17*, 241–272.

6. Tan, R.T.; Nishino, K.; Ikeuchi, K. Separating Reflection Components Based on Chromaticity and Noise Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1373–1379.

7. Tan, R.T.; Ikeuchi, K. Separating Reflection Components of Textured Surfaces Using a Single Image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 178–193.

8. Wolff, L.B. Polarization-Based Material Classification from Specular Reflection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 1059–1071.

9. Umeyama, S.; Godin, G. Separation of Diffuse and Specular Components of Surface Reflection by Use of Polarization and Statistical Analysis of Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 639–647.

10. Bronstein, A.M.; Bronstein, M.M.; Zibulevsky, M.; Zeevi, Y.Y. Blind Separation of Reflections Using Sparse ICA. In Proceedings of the Symposium on Independent Component Analysis and Blind Signal Separation, Nara, Japan, 1–4 April 2003; pp. 227–232.

11. Xu, L.M.; Yang, Z.Q.; Jiang, Z.H.; Chen, Y. Light source optimization for automatic visual inspection of piston surface defects. *Int. J. Adv. Manuf. Technol.* **2016**, doi:10.1007/s00170-016-9937-x.

12. Irani, M.; Rousso, B.; Peleg, S. Computing occluding and transparent motions. *Int. J. Comput. Vis.* **1994**, *12*, 5–16.

13. Szeliski, R.; Avidan, S.; Anandan, P. Layer Extraction from Multiple Images Containing Reflections and Transparency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, 15 June 2000.

14. Swaminathan, R.; Kang, S.B.; Szeliski, R.; Criminisi, A.; Nayar, S.K. On the Motion and Appearance of Specularities in Image Sequences. In *Lecture Notes in Computer Science, Proceedings of the 7th European Conference on Computer Vision (ECCV), London, UK, 28–31 May 2002*; Heyden, A., Sparr, G., Nielsen, M., Johansen, P., Eds.; Springer: Berlin, Germany, 2002; Volume 2350, pp. 508–523.

15. Lin, S.; Li, Y.; Kang, S.B.; Tong, X.; Shum, H.Y. Diffuse-Specular Separation and Depth Recovery from Image Sequences. In *Lecture Notes in Computer Science, Proceedings of the 7th European Conference on Computer Vision (ECCV), London, UK, 28–31 May 2002*; Heyden, A., Sparr, G., Nielsen, M., Johansen, P., Eds.; Springer: Berlin, Germany, 2002; Volume 2350, pp. 210–224.

16. Shah, S.M.Z.A.; Marshall, S.; Murray, P. Removal of specular reflections from image sequences using feature correspondences. *Mach. Vis. Appl.* **2017**, *28*, 1–12.

17. Nayar, S.K.; Fang, X.S.; Boult, T.E. Separation of Reflection Components Using Color and Polarization. *Int. J. Comput. Vis.* **1997**, *21*, 163–186.

18. Lin, S.; Lee, S.W. Detection of Specularity Using Stereo in Color and Polarization Space. *Comput. Vis. Image Underst.* **1997**, *65*, 336–346.

19. Zhou, S.; Liang, D.; Wei, Y. Automatic detection of metal surface defects using multi-angle lighting multivariate image analysis. In Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA), Ningbo, China, 1–3 Augest 2016; pp. 1588–1593.

20. Ikeuchi, K. Determining surface orientation of specular surfaces by using the photometric stereo method. *IEEE Trans. Pattern Anal. Mach. Intell.* **1981**, *3*, 661–669.

21. Schultz, H.J. Retrieving Shape Information from Multiple Images of a Specular Surface. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 195–201.

22. Koenderink, J.J.; van Doorn, A.J. Photometric Invariants Related to Solid Shape. *J. Mod. Opt.* **1980**, *27*, 981–996.

23. Zheng, J.Y.; Murata, A. Acquiring a Complete 3D Model from Specular Motion under the Illumination of Circular-Shaped Light Sources. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 913–920.

24. Solem, J.E.; Aanæs, H.; Heyden, A. A Variational Analysis of Shape from Specularities using Sparse Data. In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece, 9 September 2004; pp. 26–33.

25. Wang, J.; Dana, K.J. A Novel Approach for Texture Shape Recovery. In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece, 9 September 2003; pp. 1374–1380.

26. Kutulakos, K.N.; Steger, E. A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation. *Int. J. Comput. Vis.* **2008**, *76*, 13–29.

27. Vasilyev, Y.; Adato, Y.; Zickler, T.; Ben-Shahar, O. Dense specular shape from multiple specular flows. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

28. Savarese, S.; Chen, M.; Perona, P. Local Shape from Mirror Reflections. *Int. J. Comput. Vis.* **2005**, *64*, 31–67.

29. Savarese, S.; Perona, P. Local Analysis for 3D Reconstruction of Specular Surfaces—Part II. In *Lecture Notes in Computer Science, Proceedings of the 7th European Conference on Computer Vision (ECCV), London, UK, 28–31 May 2002*; Heyden, A., Sparr, G., Nielsen, M., Johansen, P., Eds.; Springer: Berlin, Germany, 2002; Volume 2351, pp. 759–774.

30. Savarese, S.; Chen, M.; Perona, P. Recovering Local Shape of a Mirror Surface from Reflection of a Regular Grid. In *Lecture Notes in Computer Science, Proceedings of the European Conference on Computer Vision (ECCV), Prague, Czech Republic, 11–14 May 2004*; Pajdla, T., Matas, J., Eds.; Springer: Berlin, Germany, 2004; Volume 3023, pp. 468–481.

31. Lellmann, J.; Balzer, J.; Rieder, A.; Beyerer, J. Shape from Specular Reflection and Optical Flow. *Int. J. Comput. Vis.* **2008**, *80*, 226–241.

32. Fleming, R.W.; Torralba, A.; Adelson, E.H. Specular reflections and the perception of shape. *J. Vis.* **2004**, *4*, 798–820.

33. Tarini, M.; Lensch, H.P.A.; Goesele, M.; Seidel, H.P. 3D acquisition of mirroring objects using striped patterns. *Graph. Models* **2005**, *67*, 233–259.

34. Bothe, T.; Li, W.; von Kopylow, C.; Juptner, W.P.O. High-resolution 3D shape measurement on specular surfaces by fringe reflection. *Opt. Metrol. Prod. Eng.* **2004**, *5457*, 411–422.

35. Knauer, M.C.; Kaminski, J.; Hausler, G. Phase measuring deflectometry: a new approach to measure specular free-form surfaces. *Opt. Metrol. Prod. Eng.* **2004**, *5457*, 366–376.

36. Perard, D. Automated Visual Inspection of Specular Surfaces with Structured-Lighting Reflection Techniques. Ph.D. Thesis, Universitat Karlsruhe, Karlsruhe, Germany, 2001.

37. Huang, P.S.; Hu, Q.; Jin, F.; Chiang, F.P. Color-encoded digital fringe projection technique for high speed three-dimensional surface contouring. *Opt. Eng.* **1999**, *38*, 1065–1071.

38. Perard, D.; Beyerer, J. Three-dimensional measurement of specular free-form surfaces with a structured-lighting reflection technique. In Proceedings of the Three-Dimensional Imaging and Laser-Based Systems for Metrology and Inspection III, Pittsburgh, PA, USA, 14 October 1997; pp. 74–80.

39. Newman, T.S.; Jain, A.K. A Survey of Automated Visual Inspection. *Comput. Vis. Image Underst.* **1995**, *61*, 231–262.

40. Malamas, E.N.; Petrakis, E.G.; Zervakis, M.; Petit, L.; Legat, J.D. A survey on industrial vision systems, applications and tools. *Image Vis. Comput.* **2003**, *21*, 171–188.

41. Neubecker, R.; Hon, J.E. Automatic inspection for surface imperfections: Requirements, potentials and limits. In Proceedings of the Third European Seminar on Precision Optics Manufacturing, Teisnach, Germany, 12 April 2016; p. 1000907.

42. Cui, Z.; Lu, W.; Liu, J. Real-time Industrial Vision System for Automatic Product Surface Inspection. In Proceedings of the 2016 8th International Conference on Information Management and Engineering, New York, NY, USA, 2–5 November 2016; pp. 93–97.

43. Kulmann, L. On Automatic Visual Inspection of Reflective Surfaces. Ph.D. Thesis, Technical University of Denmark, Kongens Lyngby, Denmark, 1995.

44. Li, J.; Parker, J.M.; Hou, Z. An intelligent system for real time automatic defect inspection on specular coated surfaces. *Vis. Commun. Image Process.* **2005**, *5960*, 596043.

45. Parker, J.M.; Cheong, Y.L.; Gnanaprakasam, P.; Hou, Z.; Istre, J. Inspection Technology to Facilitate Automated Quality Control of Highly Specular, Smooth Coated Surfaces. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington, DC, USA, 11–15 May 2002; pp. 2567–2574.

46. Pernkopf, F.; OĹeary, P. Image acquisition techniques for automatic visual inspection of metallic surfaces. *NDT E Int.* **2003**, *36*, 609–617.

47. Chu, H.H.; Wang, Z.Y. A vision-based system for post-welding quality measurement and defect detection. *Int. J. Adv. Manuf. Technol.* **2016**, *86*, 3007–3014.

48. Werling, S.; Mai, M.; Heizmann, M.; Beyerer, J. Inspection of Specular and Partially Specular Surfaces. *Metrol. Meas. Syst.* **2009**, *16*, 415–431.

49. Kammel, S.; Leon, F.P. Head-mounted display for interactive inspection of painted free-form surfaces. *Helmet-Head-Mounted Disp. VIII Technol. Appl.* **2003**, *5079*, 254–264.

50. Lu, Y.; Lu, R. Using composite sinusoidal patterns in structured-illumination reflectance imaging (SIRI) for enhanced detection of apple bruise. *J. Food Eng.* **2017**, *199*, 54–64.

51. Sanderson, A.S.; Weiss, L.E.; Nayar, S.K. Structured Highlight Inspection of Specular Surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **1988**, *10*, 44–55.

52. Nayar, S.; Sanderson, A.C.; Weiss, L.; Simon, D. Specular surface inspection using structured highlight and Gaussian images. *IEEE Trans. Robot. Autom.* **1990**, *6*, 208–218.

53. Lai, W.W.; Zeng, X.X.; He, J.; Deng, Y.L. Aesthetic defect characterization of a polymeric polarizer via structured light illumination. *Polym. Test.* **2016**, *53*, 51–57.

54. Leon, F.P.; Kammel, S. Inspection of specular and painted surfaces with centralized fusion techniques. *Measurement* **2006**, *39*, 536–546.

55. Kammel, S.; Leon, F. Deflectometric Measurement of Specular Surfaces. In Proceedings of the 2005 IEEE Instrumentationand Measurement Technology Conference, Ottawa, ON, Canada, 16–19 May 2005; pp. 531–536.

56. Guo, H.; Feng, P.; Tao, T. Specular surface measurement by using least squares light tracking technique. *Opt. Lasers Eng.* **2009**, *48*, 166–171.

57. Zhang, X.; North, W.P. Analysis of 3-D surface waviness on standard artifacts by retroreflective metrology. *Opt. Eng.* **2000**, *39*, 183–186.

58. Zhang, X.; North, W.P. Retroreflective Grating Generation and Analysis for Surface Measurement. *Appl. Opt.* **1998**, *37*, 2624–2627.

59. Hoefling, R.; Aswendt, P.; Neugebauer, R. Phase reflection: A new solution for the detection of shape defects on car body sheets. *Opt. Eng.* **2000**, *39*, 175–182.

60. Hung, Y.Y.; Shang, H.M. Nondestructive testing of specularly reflective objects using reflection three-dimensional computer vision technique. *Opt. Eng.* **2003**, *42*, 1343–1347.

61. Hung, Y.Y.; Lin, L.; Shang, H.M.; Park, B.G. Practical three-dimensional computer vision techniques for full-field surface measurement. *Opt. Eng.* **2000**, *39*, 143–149.

62. Seulin, R.; Merienne, F.; Gorria, P. Simulation of specular surface imaging based on computer graphics: Application on a vision inspection system. *EURASIP J. Appl. Signal Process.* **2002**, *7*, 649–658.

63. Aluze, D.; Merienne, F.; Dumont, C.; Gorria, P. Vision system for defect imaging, detection, and characterization on a specular surface of a 3D object. *Image Vis. Comput.* **2002**, *20*, 569–580.

64. Virtual reality applications in manufacturing process simulation. *J. Mater. Proce. Technol.* **2004**, *155–156*, 1834–1838.

65. Seulin, R.; Merienne, F.; Gorria, P. Simulation of Specular Surface Imaging Based on Computer Graphics: Application on a Vision Inspection System. *EURASIP J. Adv. Signal Process.* **2002**, *2002*, 801489.

66. Nicodemus, F.E.; Richmond, J.C.; Hsia, J.J.; Ginsberg, I.W.; Limperis, T. *Geometrical Considerations and Nomenclature for Reflectance*; Technical Report; National Bureau of Standards: Washington, DC, USA, 1977.

67. Cook, R.L.; Torrance, K.E. A reflectance model for computer graphics. In Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques, New York, NY, USA, 3–7 August 1981; pp. 307–316.

68. Ngan, A.; Durand, F.; Matusik, W. Experimental Analysis of BRDF Models. In Proceedings of the Sixteenth Eurographics conference on Rendering Techniques, Konstanz, Germany, 29 June–1 July 2005.

69. Beckmann, P.; Spizzichino, A. *The Scattering of Electromagnetic Waves from Rough Surfaces*; Artech House, Inc.: Norwood, MA, USA, 1987; p. 1987.

Crack Detection in Concrete Tunnels Using a Gabor Filter Invariant to Rotation

Roberto Medina [1], José Llamas [1], Jaime Gómez-García-Bermejo [2,*], Eduardo Zalama [2] and Miguel José Segarra [3]

[1] CARTIF Foundation, Parque Tecnológico de Boecillo, 47151 Valladolid, Spain; robmed@cartif.es (R.M.); joslla@cartif.es (J.L.)
[2] ITAP-DISA, University of Valladolid, 47002 Valladolid, Spain; ezalama@eii.uva.es
[3] DRAGADOS S.A., Av. Camino de Santiago 50, 28050 Madrid, Spain; mjsegarra@dragados.com
* Correspondence: jaigom@eii.uva.es; Tel.: +34-983-546-504

Abstract: In this article, a system for the detection of cracks in concrete tunnel surfaces, based on image sensors, is presented. Both data acquisition and processing are covered. Linear cameras and proper lighting are used for data acquisition. The required resolution of the camera sensors and the number of cameras is discussed in terms of the crack size and the tunnel type. Data processing is done by applying a new method called Gabor filter invariant to rotation, allowing the detection of cracks in any direction. The parameter values of this filter are set by using a modified genetic algorithm based on the Differential Evolution optimization method. The detection of the pixels belonging to cracks is obtained to a balanced accuracy of 95.27%, thus improving the results of previous approaches.

Keywords: tunnel inspection; concrete surface; image processing; spatial frequency domain

1. Introduction

Traditional methods of tunnel inspection are based on the subjective criteria of a human operator that performs local visual inspection and sensing equipment that usually requires installation and contact with the tunnel surface (plaster indicators, crackmeters, etc.). The visual inspection of 100% of the tunnel surface would be preferred, but this requires the management of a massive amount of information. Therefore, the automatic acquisition and processing of images emerges as the only practical solution. To this end, the development of image processing algorithms capable of detecting the presence of defects in the surface of the tunnel is necessary.

In this paper, the automated acquisition and processing of images of 100% of the surface of the tunnel is proposed that improves upon the results obtained with traditional techniques. The management of a large amount of information provided by the image sensors on the entire surface of the tunnel does not allow the manual evaluation of data, so new methods for the automated processing of this data have been developed. The proposed methods allow for the inspecting and detecting of fissures and cracks in the surface of the tunnel. This type of defect is one of the most usual, but also one of the most difficult to detect.

The image sensors are mounted onto an inspection platform designed to inspect any rail, highway, or hydraulic tunnel whose surface is made from concrete by tunnel boring machines. The methodology for the inspection of different tunnel typologies and different crack width detection requirements is addressed. A new method based on Gabor filters for the detection of cracks, independently of their direction of growth, is also presented.

2. Related Work

A number of techniques have focused on the detection of defaults in continuous surfaces by using computer vision-based sensors. Some of these techniques are based on the segmentation of the original image, either by using a fixed threshold as in [1], a threshold computed from the previous images as in [2,3], or a different threshold for each image column, as in [4]. Other authors have worked on the previously processed images, as in [2,5,6]; some have worked on entropy [7], where the use of the Hough transform is proposed; or on morphological filters, as in [8]; or on different edge detection filters, as in [9–11].

More specifically, the detection of cracks has been addressed in many areas, such as the inspection of concrete and asphalt infrastructures. One of the most developed fields is the detection of cracks in roads, these being one of the infrastructures with the largest surface anywhere in the world. The simplest and most common detection method is direct image segmentation, given that cracks are usually darker than the surrounding areas. Morphological filters are then applied to the segmented images in order to join the cracks and remove the segmentation noise, as is done for example in [1]. Other works have addressed a more elaborate segmentation, such as [12], where the Neighboring Difference Histogram Method (NDHM), which compares each pixel with the histogram of the surrounding pixels, is used; or [6], where the images are divided into different zones, to which a modified Otsu thresholding algorithm is applied [13], and the resulting entropy is computed. Other methods are based on the search for seeds that are merged by taking into account the fact that the cracks appear as elongated structures [14]. Recently, Convolutional Neuronal Networks has been used to detect cracks, as can be found in [15], where cracks are detected at pixel level.

A significant number of works can also be found where the detection of cracks in concrete buildings and infrastructures, especially bridges, is addressed. Four methods for the detection of cracks in concrete bridges are compared in [16]: Sobel and Canny edge detectors, the Fourier transform, and the Haar wavelet transform. These authors conclude that the Haar wavelet transform is the best performer.

A method for detecting cracks in concrete structures is described in [17]. This method is based on the application of a large number of filters to a database that has previously been processed by hand. The optimal combination of filters to obtain the best possible result is selected using genetic algorithms. In [10], a method for detecting cracks in concrete surfaces by areascan color sensors, based on different edge detection algorithms, is described. The system can detect cracks with an opening greater than 0.25 mm.

The works where cracks are detected inside tunnels differ from the others mainly due to the use of artificial light. Some works are based on thresholding techniques, such as [4], where the value of each pixel is compared to the average value of the pixels in the column where that pixel is located. The work of Fujita et al. [11], cited above, can be included in this group. The main drawback of these techniques is how to choose the threshold value. The methods for setting the threshold value are usually based on prior knowledge, but these methods are not generalizable and do not usually work properly when applied to real tunnels, where shadows and humidity may occur.

Other techniques for detecting cracks in tunnels are based on models, such as the one described above in [3], using the concept of percolation, or the one presented in [18], where it is assumed that cracks are formed by line segments joined together. In the second case, the image resolution is 2.5 mm per pixel and can be considered a semi-automatic method, as it needs human intervention to choose the seeds of every crack.

There are also some methods based on pattern recognition, such as the one proposed in [19], where the SVM (Support Vector Machine) algorithm is used to determine whether the preselected areas are really cracks. An inspection system of concrete tunnels composed of nine linear color CMOS cameras, whose images are transformed to gray images for processing, is proposed in [8]. First, a smoothing filter to remove noise is applied to the gray images. The resulting image is filtered by the top-hat filter to enhance cracks, as they are supposed to be darker than the rest of the image. The threshold

for segmenting the filtered image is set by analyzing the distribution of the gray levels of the filtered image. Neighboring pixels are grouped, and shape and gray level features are calculated to determine which groups are really cracks and which are not. For this, four pattern recognition techniques are applied: two neural networks, the SVM algorithm, and the nearest neighbor algorithm. The accuracy of these algorithms is approximately 90%, which was calculated using 200 pixel groups, of which 64 were actually cracks. A total of nine linear cameras of 12,288 pixels acquire images in a subway tunnel covering 270° of the tunnel, but the tunnel diameter was not available, so the image resolution cannot be determined.

Table 1 summarizes the described related work. However, in general, the reviewed techniques are often not suitable for the detection of cracks in tunnels. The roughness of the concrete, the shadows derived from this roughness, and the fact that the crack brightness is often similar to that of the surrounding areas mean that these segmentation techniques can hardly provide the required results.

In the present paper, a segmentation process based on the localized space-frequency information of the images, obtained using Gabor filters, is presented. Gabor filters have also been used successfully by our research team for the detection of cracks in roads [20]. However, in this previous work, the sensed images are divided into pieces and the algorithm determines whether there is a crack inside each piece using a variant of the AdaBoost algorithm for training. The present work addresses the use of Gabor filters in a significantly different way. In particular, the classification process is made for every pixel of the image, and the Gabor filter parameters are set using a modified genetic algorithm based on the Differential Evolution optimization method. Moreover, the texture of asphalt roads and concrete surfaces are very different, and while road cracks exhibit longitudinal and transverse preferential directions, tunnel cracks can be oriented in any direction. Therefore, a new method based on Gabor filters, which we have called Gabor filter invariant to rotation, is proposed in the present work for the detection of cracks independent of their direction of growth.

Table 1. Summarized comparisons of related work.

References	Vision Techniques
[1–4]	Thresholding original images
[2,5–11]	Thresholding processed images
[6,12,13]	More elaborated threshold techniques
[14,18]	Seeds
[15]	CNN
[9–11,16]	Edge detectors
[17]	Genetic algorithms
[3,18]	Models
[19]	SVM
[20]	Gabor filters

3. System Description

There are multiple defects that may appear on the surface of the lining of a concrete tunnel. These defects can be classified into four main types: mechanical, chemical, water-related and joint-related defects.

Concerning the mechanical defect type, there are fissures that appear in the surface mainly due to the presence of tensions over its resistance capacity. When a fissure traverses the entire element thickness, from side to side, it becomes a crack. Other mechanical defects are detachments or ruptures on precasts, broken off corners, or seen steel frames. Some examples of these defects are shown in Figure 1. Finally, other defects are of a chemical nature, such as carbonation, efflorescence, and other defects related to the presence of water, such as moisture, dripping, and seepage.
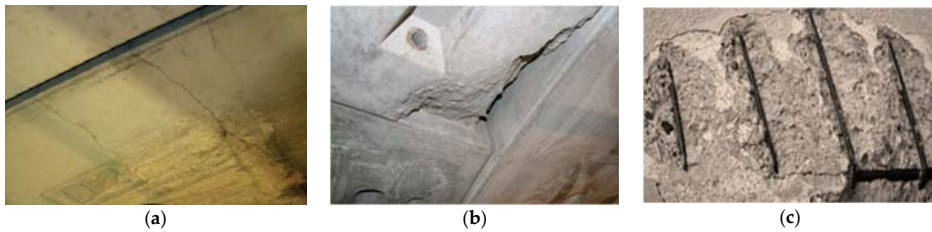
| (a) | (b) | (c) |

Figure 1. Mechanical defects in tunnels: (**a**) fissures; (**b**) detachments on precasts; (**c**) seen steel frames.

This paper aims to develop a system for the detection of fissures and cracks in tunnels. This type of defect can be identified through visual inspection, but often goes unnoticed during a routine human inspection. In addition, these defects appear at preliminary phases of tunnel deterioration, so their early detection allows preventive actions to be taken, before the problem becomes more severe. Moreover, both fissures and cracks exhibit a similar visual appearance. Therefore, from now on, we will use the term *crack* for referring to both fissures and cracks.

Cracks can be classified into different categories depending on their opening and the apparent damage of the precast, as shown in Table 2. Cracks of type 0 are those with an opening of less than 0.1 mm. Cracks of type A include those with an opening of between 0.1–0.4 mm, that do not reach the intrados steel, and so are of an aesthetic nature. Type B cracks have an opening smaller than 2 mm at their head, pass through the first reinforcement layer, and are stabilized. Finally, type C cracks correspond to the multicracked precasts. Subtype C1 corresponds to cracks without risk of fall or detachment, usually with the presence of moisture or water ooze, and subtype C2 correspond to multicracked precasts with risk of detachment.

Table 2. Crack classification on tunnels.

Crack Type	Subtype	Opening (mm)	Description
0		<0.1	Usually not taken into account
A		0.1–0.4	Do not reach the intrados steel They are of an aesthetic nature
B		0.4–2	Pass through the first reinforcement layer They are stabilized
C	C1	<0.4	Multicracked precasts Cracks without risk of fall or detachment
	C2	>0.4	Multicracked precasts Cracks with risk of detachment

A system for the automatic visual inspection of tunnels consists of three major blocks or stages: image acquisition, processing, and result exploitation. Figure 2 shows these major blocks and the main actions involved in each of them.

Image acquisition is addressed in depth in Section 4 of this paper, while the processing techniques for crack detection in concrete tunnels are described in Section 5. The obtained results can be visualized, and reports of the state of the tunnel can be made, allowing engineers to take the pertinent action to repair them.
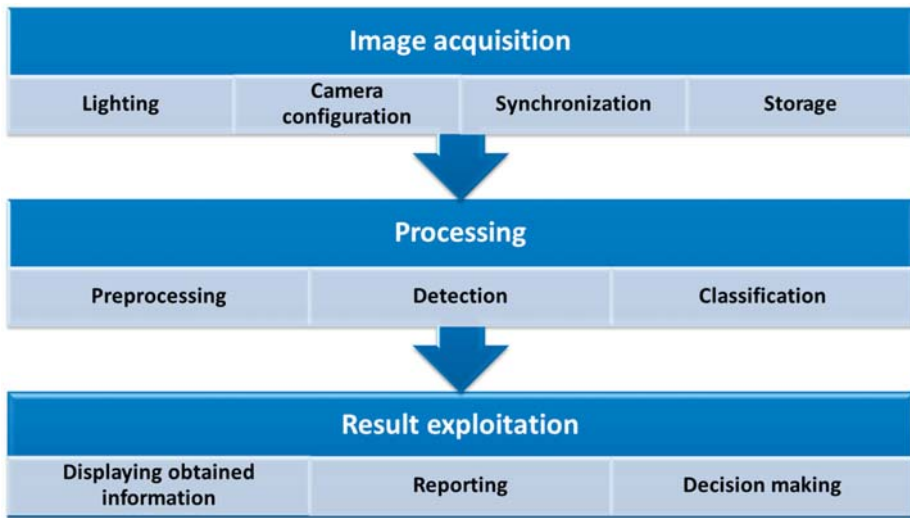
Figure 2. Stages of the automatic visual inspection process.

4. Image Sensors for Crack Detection

Concerning data acquisition, the use of linear cameras is recommended in order to simplify lighting and ulterior data processing. These cameras are equipped with a 1D sensor array (instead of the 2D array used in common cameras). The linear cameras must be attached to a mobile platform that travels along the tunnel. The number of cameras, their resolution, and their arrangement will depend on the size and characteristics of the tunnel and on the minimum size of the cracks to be detected.

In particular, in order to achieve a uniform resolution along the entire perimeter of the tunnel, the cameras must be arranged radially, i.e., placed on a circumference centered at the tunnel center (for the current tunnel section), and oriented so that their optical axes cross at the said center. Moreover, the inspection elements must be placed in the area of the tunnel free from obstacles, which we will call the Working Area (WA). The radius of the bigger circumference inside this area is the working area radius, R_{WA}.

The camera sensor type and the corresponding optics must be selected according to the minimum crack size, as will be explained later. These aspects will determine the working distance, W_D, namely the distance between the camera's optical center and the tunnel surface. In general, this working distance is recommended to be lower than the radius of the tunnel, R_T, plus the radius of the working area, R_{WA}, and greater than the radius of the tunnel, R_T, minus the radius of the working area, R_{WA}:

$$R_T - R_{WA} < W_D < R_T + R_{WA} \tag{1}$$

When the selected working distance is greater than the tunnel radius, the cameras should be placed along half the tunnel section so that their view is not obstructed by the opposite cameras, as described in Figure 3. Moreover, either the system should travel twice along the tunnel or the cameras should be arranged on two different planes.
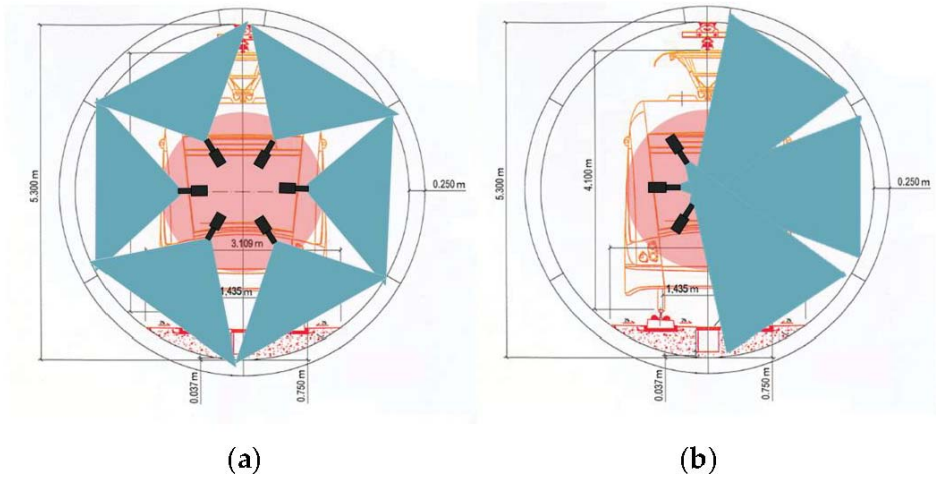
Figure 3. Two examples of camera position distribution, and the portion of the tunnel section covered by each camera. (The field-of-view of the cameras is shadowed in blue). (**a**) Working distance (W_D) is smaller than the radius of the tunnel (R_T) (**b**) Working distance (W_D) is greater than the radius of the tunnel (R_T), so only half of the cameras can be placed in the same plane.

Concerning camera sensor resolution, common linear cameras range from about 1024 pixels for an e2v AviiVA SM2 1010, for example, to 12,288 pixels for an e2v AviiVA UM8 camera. The number of cameras required for inspection is determined by the actual sensor resolution and the minimum crack size:

$$N_C > \frac{2P_T}{W_C^{min} N_P} \tag{2}$$

In this expression, N_C is the number of cameras, P_T is the tunnel perimeter, W_C^{min} is the minimum crack width, and N_P is the number of pixels of the camera sensor. Moreover, a (small) overlap between the field of view of two adjacent cameras is recommended to avoid gaps.

The necessary sensor resolution that allows the camera to measure these types of crack requires twice the resolution of the image, i.e., to detect cracks 1 mm wide, images with a resolution of at least 2 pixels per millimeter should be acquired. Typical minimum crack values are between 0.1 and 2 mm, which would include most types of cracking defects and cracks in tunnels.

Furthermore, appropriate optics must be selected for the chosen camera, so the image can be captured within the established working distance limits, and the tunnel surface can be properly focused.

Table 3 reports some examples of sensor camera resolution, number of cameras and focal length of the optics recommended for detecting cracks ranging from 0.1 to 2 mm, which cover most expected crack types, for three tunnel types: electricity tunnel, underground tunnel and high-speed train tunnel (sized 3.5 m, 5.3 m, and 10.40 m in diameter, respectively). Both one-pass and two-pass acquisition cases are shown.

Finally, the data captured by the cameras should be stored for later analysis. In the worst-case scenario, for the 10.4 m sized, high-speed train tunnel, with 0.1 mm cracks, about 8 TBytes are required to store images from a 1 km tunnel. These storage requirements drop to 0.5 Tbytes in the case of 0.4 mm cracks. The storage requirements drop significantly for the other resolution figures.

Table 3. Number of cameras and their recommended sensor resolution for the different tunnel types.

Crack Size	Sensor Resolution (Pixels)	Electricity Tunnel			Underground Tunnel			High-Speed Train Tunnel		
		No. of Cameras		Focal Length (mm)	Nb of Cameras		Focal Length (mm)	No. of Cameras		Focal Length (mm)
		1 Pass	2 Pass		1 Pass	2 Pass		1 Pass	2 Pass	
0.1 mm	12,288	14	7	80	22	11	105	–	–	–
0.4 mm	8194	6	3	50	8	4	50	–	7	135
	6144	7	4		11	6		–	9	
1 mm	6144	–	2	28	–	3	28	–	4	50
2 mm	4096	–	–		–	2	28	–	3	28
	1024	9	5	12	–	–	–	–	–	–

5. Methodology for Spatial-Frequency Image Analysis by Gabor Filters Invariant to Rotation

5.1. Gabor Filter

A two-dimensional Gabor filter is a complex sinusoidal wave modulated by a Gaussian envelope [21]. The filter performs a localized and oriented frequency analysis of a two-dimensional signal. The spatial domain formulation is,

$$G_{\sigma,F,\theta}(x,y) = g_\sigma(x,y)\exp[j2\pi Fx']$$
$$\text{where } g_\sigma(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\left(\frac{x'}{\sigma_x}\right)^2 + \left(\frac{y'}{\sigma_y}\right)^2\right)\right] \quad (3)$$
$$\text{and } x' = x\cos\theta + y\sin\theta \quad y' = -x\sin\theta + y\cos\theta$$

where x, y are the spatial coordinates, F is the central frequency, θ is the angle between the direction of the sinusoidal wave and the x-axis, and σ_x and σ_y are the smoothing parameters (standard deviations of the Gaussian envelope in the direction of the wave and orthogonal to it). The frequency domain formulation of the Gabor filter is,

$$G_{\sigma,F,\theta}(u,v) = \exp\left[\frac{-1}{2}\left(\frac{(u'-F)^2}{\sigma_u^2} + \frac{v'^2}{\sigma_v^2}\right)\right]$$
$$\text{where } \sigma_u = \frac{1}{2\pi\sigma_x}, \sigma_v = \frac{1}{2\pi\sigma_y} \quad (4)$$
$$u' = u\cos\theta + v\sin\theta \quad \text{and} \quad v' = -u\sin\theta + v\cos\theta$$

Four parameters must be set to appropriate values in order to apply a Gabor filter for processing an image: the center frequency, F, the orientation, θ, and two smoothing parameters, σ_x and σ_y. We actually aim to enhance the cracks in images, so the filtered image should offer high values in the presence of cracks and low values otherwise. This can be achieved by selecting the frequency value so the corresponding wavelength ($2\pi/F$) approaches the width of the crack, and an orientation value so it matches that of the crack. In turn, smoothing parameters serve to increase or decrease the range of frequencies and orientations to which the filter reacts.

5.2. Gabor Filters Invariant to Rotation

Tunnel cracks can exhibit any possible orientation in the image, and a single crack can even exhibit different orientations along its length. Therefore, the design of filters invariant to rotation (i.e., filters that have the same response regardless of the orientation of the cracks) has been addressed in the present work. To do this, a given number of different orientations is assumed, n_θ, and a set of values for F, σ_x, and σ_y parameters is selected. Then, a filtered image is obtained for each orientation in n_{θ_i}

and the remaining parameter values. The resulting filtered-images are combined to obtain a final, single image by taking the maximum value at each pixel,

$$I_{ij}^F = Max\left(I_{ij}^n\right) \tag{5}$$

where I^F is the final filtered image, I^n is the image filtered by the Gabor filter with the nth orientation, and i and j are the columns and rows of the image, respectively. Finally, this resulting image, I^F, is segmented by computing a suitable threshold value.

An example of the application of the rotation-invariant Gabor filter is shown in Figure 4. The original image and the result of applying the rotation-invariant Gabor filter to it are shown on the left. The Gabor filters applied along 16 orientations (from 0° to 180°) are shown on the right, along with their maximum values (images have been normalized for a better visualization). In this example, $F = 0.16$ pixels^{-1}, $\sigma_x = 2.08$ pixels and $\sigma_y = 17.16$ pixels (see Section 6 for further details on these values).
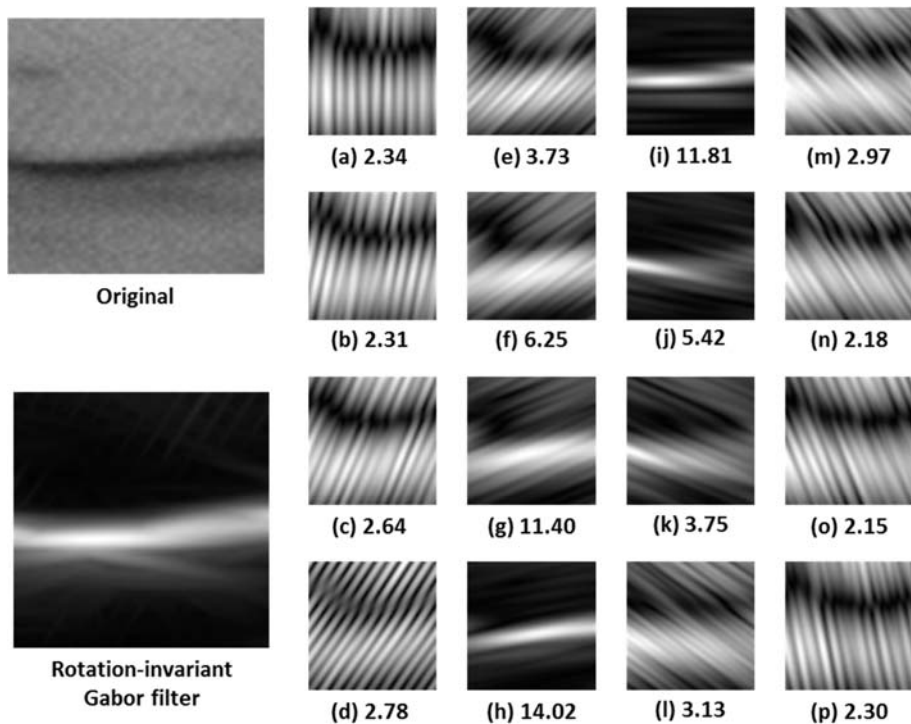

Figure 4. Example of the application of the rotation-invariant Gabor filter. The original image and the result of applying the rotation-invariant Gabor filter to it are shown on the left. The Gabor filters applied along 16 orientations (from 0° to 180°) are shown on the right, along with their maximum values (**a**–**p**).

5.3. Selection of Optimal Parameters by Differential Evolution Algorithm

A proper choice of the F, σ_x, and σ_y parameter values of the Gabor filters is essential to obtain accurate results. In the present work, the use of a genetic algorithm based on the so-called Differential Evolution optimization method is proposed. This algorithm was first introduced by Storn and

Price [22], and allows good convergence results to be achieved, to a large extent avoiding falls in local minimums. The algorithm operates in four main steps:

1. **Initialization**: Generate N_p vectors randomly within a given parameter space. Each vector is composed of the variables of the function to be minimized, i.e., the parameters of the image segmentation method whose optimization is looked for. The segmentation method is applied to a set of training images using each of these vectors, and the corresponding result is evaluated.

2. **Mutation**: N_p new vectors are created by applying,

$$v_{i,g} = x_{best,g} + F(x_{r1,g} - x_{r2,g}) \tag{6}$$

   where i is the vector number, g the generation number, x the current vector, and v the muted one. Moreover, the subscript *best* refers to the vector that leads to the best results, $r1$ and $r2$ are two different values randomly selected within the 0 to N_p-1 range, and F is a mutation scale vector. The original algorithm proposed by Storn and Price used another random vector, $r3$, instead of the one leading to the best result, *best*, and $r1$, $r2$, $r3$, and i should be different.

3. **Recombination**: In this step, the mutated vector is recombined with the unmutated one by applying,

$$u_{ji,g} = \begin{cases} v_{ji,g} & \text{if } \text{rand}_j[0,1] \leq CR \\ x_{ji,g} & \text{otherwise} \end{cases} \tag{7}$$

   where j is the variable number, u is the recombined vector and CR is the probability for a variable to mutate. It must be ensured that at least one variable is mutated, so one variable is initially mutated and the probability is applied to the others.

4. **Selection**: Finally, the recombined vectors are evaluated and the best performing one, from the original and recombined vectors, are assigned to the next generation as follows:

$$x_{i,g+1} = \begin{cases} u_{i,g} & \text{if } f(u_{i,g}) < f(x_{i,g}) \\ x_{i,g} & \text{otherwise} \end{cases} \tag{8}$$

In this expression, f is the function selected to evaluate the vector performance. The whole process is then repeated from point 2 until a stop criterion is matched.

The results of the algorithm are improved in the present work by modifying the initialization step of the above-described process (step 1) in such a way that a Gabor filter bank is used, computed as described in [20], instead of starting from a group of randomly selected vectors. This reduces the possibility of falls in local minimums and reduces the number of generations required to reach the stop criterion.

6. Experimental Results

6.1. Inspection Platform Prototype

A mobile platform that travels along the railroad track (see Figure 5) has been developed to test the feasibility of the proposed method. The platform is equipped with an Aviiva SM2 CL 1010 linear camera. Moreover, a Tamron AF 70–300 mm F/4–5.6 Di LD Macro 1:2 optics has been used, along with a 100 W halogen bulb with an AR111 reflector. The selected linear camera has a sensor resolution of 1024 pixels and can acquire up to 54,400 lines per second. The selected optics has a variable zoom that can be set from 70 to 300 mm, so it can be used in tunnels of different diameters. The reflector lighting system allows the light to be concentrated in an 8° cone. Both the camera and the light source can be oriented in multiple positions, which allows different areas of the tunnel to be acquired with different resolutions. The images of the tunnel surface obtained by using this system are then processed through the crack detection algorithms.

Figure 5. Prototype of the tunnel inspection platform.

The camera is configured in such a way that the average value of defect-free images is placed at roughly half the dynamic range of the camera, i.e., 128 in a 0 to 255 range. Thus, defects both darker and brighter than the background surface are correctly observed. This set up can be achieved by properly tuning three configurable parameters of the camera: exposure time, aperture, and sensitivity. These parameters, which determine the intensity level of pixels in the image, also affect other factors such as image sharpness or noise level.

Moreover, the image acquisition must be synchronized with the movement of the platform in order to correctly compose the image. An incremental encoder is then required on the platform wheels to measure the platform displacement. The encoder signal is activated once all the cameras on board the platform are ready to acquire images. Thus, the first encoder pulse received by every camera is the same pulse for all of them, so proper synchronization is achieved.

A software module has been developed in the present work to ease the system setup and image grabbing. This module provides the status of the different cameras, the current vehicle speed and the maximum admissible speed the vehicle can travel, which is calculated as follows:

$$S_{MAX} = \frac{R_A}{T_{Exp}} = \frac{W_C^{min}}{2T_{Exp}} \tag{9}$$

where S_{MAX} is the maximum admissible speed of the vehicle, R_A is the resolution in the direction of movement of the vehicle, T_{Exp} is the exposure time, and W_C^{min} is the minimum crack width. A correction factor of 0.9 is then applied to the result in order to ensure that the maximum speed is never exceeded. For example, if exposure time is set to 100 μs and W_C^{min} is 0.2 mm, it results in S_{MAX} being 1 m/s. Appling the correction factor means that the vehicle speed should never be greater than 0.9 m/s.

In addition, another software module has been developed that provides all the required information concerning the acquired images and the processed data. The human supervisor can navigate the stored information using this software and analyze the result of the tunnel crack detection.

6.2. Crack Detection Results

The proposed crack detection algorithm has been trained and tested using a number of images containing cracks, extracted from a 3k rail tunnel survey. The pixels of these images have been

manually classified as defect-free or cracks. A total of 100,000 pixels have been used, 20,000 of which are pixels with cracks. These pixels have been used to set the parameters of the proposed algorithms and verify the accurate operation of the system.

The parameters of the Gabor filter invariant to rotation have been set using the variant of the Differential Evolution algorithm described in the previous section. The following conditions have been assumed:

1. The first generation has been generated by defining a bank of filters covering the entire frequency space (instead of just taking a random selection). The parameters of the filter bank are: filter size $(N \times M) = 128 \times 128$ pixels, number of frequencies $(n_F) = 12$, number of orientations $(n_\theta) = 16$, frequency band $(B_F) = 0.5$ octaves, overlapping constant along x $(K_x) = 1$, overlapping constant along y $(K_y) = 1$, Max. Center Frequency $(F_M) = 0.427$ pixels^{-1}.

2. There are 12 individuals in each generation, $N_p = 12$. This value matches the number of frequencies of the filter bank described in the previous section.

3. Mutation probability has been established at 50%, $CR = 0.5$.

4. The mutation scale factor F is not a fixed value, but a function of the current generation and the number of generations without improvement, according to,

$$F = 1 - \frac{n}{n_{max}} + \frac{n^{wi}}{n_{max}^{wi}} \tag{10}$$

where n is the current generation, n_{max} is the maximum number of generations, n^{wi} is the number of generations without improvement, and n_{max}^{wi} is the maximum number of generations without improvement. Thus, the scale factor decreases as the number of current generation increases, and increases through non-improving generations. The value of F is always greater than 0 and smaller than 2.

5. The classifier threshold has been computed by minimizing the classifier weighted error (1 minus balanced accuracy). To do this, the response of all the training samples is ordered, and the weighted error is computed for each value between two samples. The threshold leading to the lowest weighted error is finally selected.

6. The stop criterion has been set to stop searching when the maximum number of generations, n_{max}, (set to 500) is reached, or the maximum number of generations without improvement, n_{max}^{wi}, (set to 50) is achieved. There is an improvement when the classifier weighted error of one individual in the present generation is smaller than the smallest previous classifier weighted error.

The confusion matrix with the results obtained after training is reported in Table 4; while the sensitivity, specificity, accuracy, precision [23], and balanced accuracy [24] values are reported in Table 5.

Table 4. Confusion matrix obtained in the detection of cracks in tunnels using the Gabor filters invariant to rotation.

Confusion Matrix		Expert Classification	
		Positive	Negative
Classifier classification	Positive	TP = 19,050	FP = 3772
	Negative	FN = 950	TN = 76,228

Table 5. Sensitivity, specificity, accuracy, precision, and balanced accuracy in the detection of cracks in tunnels using the Gabor filters invariant to rotation.

Sensitivity	Specificity	Accuracy	Precision	Balanced Accuracy
0.9525	0.9529	0.9373	0.9876	0.9527

Figure 6 shows the evolution of the weighted error through the increase in the number of generations when applying the proposed algorithm. The Gabor filter parameters obtained after applying the Differential Evolution algorithm, once the error has been stabilized (around 20th generation), are: frequency = 0.1599 (pixels^{-1}), dispersion along X = 2.0813 (pixels), dispersion along Y = 17.1607 (pixels), threshold = 3.7301.
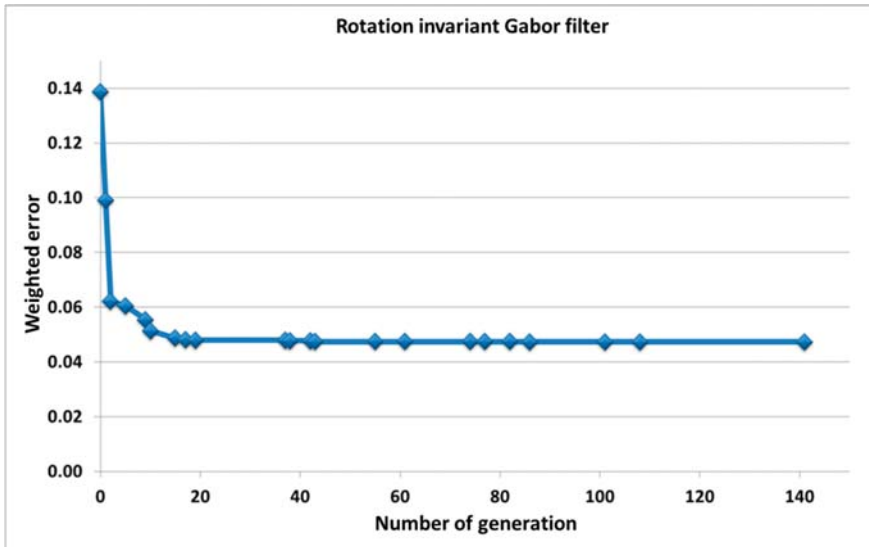

Figure 6. Evolution of the weighted error through the increase in the number of generations of the Gabor filter invariant to rotation.

The time required for the proposed algorithm to process a 1 Mbyte image is 65.6 s, which corresponds to a 15.6 kbytes/s processing speed using an Intel Pentium Core 2 Duo E8400 at 3 GHz. Figure 7 shows three examples where the described technique has been applied to 256 × 256 pixel images using the obtained parameters. The original normalized image, the filtered image, and the segmented image are shown. The values of the filtered image, in the upper row, range from 0 to 9.91, but have been rescaled to 0–255 for display purposes. Similarly, images in the middle row, ranging between 0 and 15.87, and images in the lower row, ranging between 0 and 7.89, have also been rescaled to 0–255.
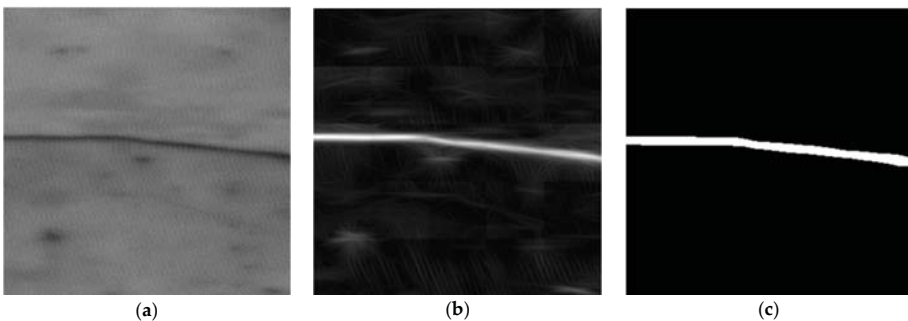

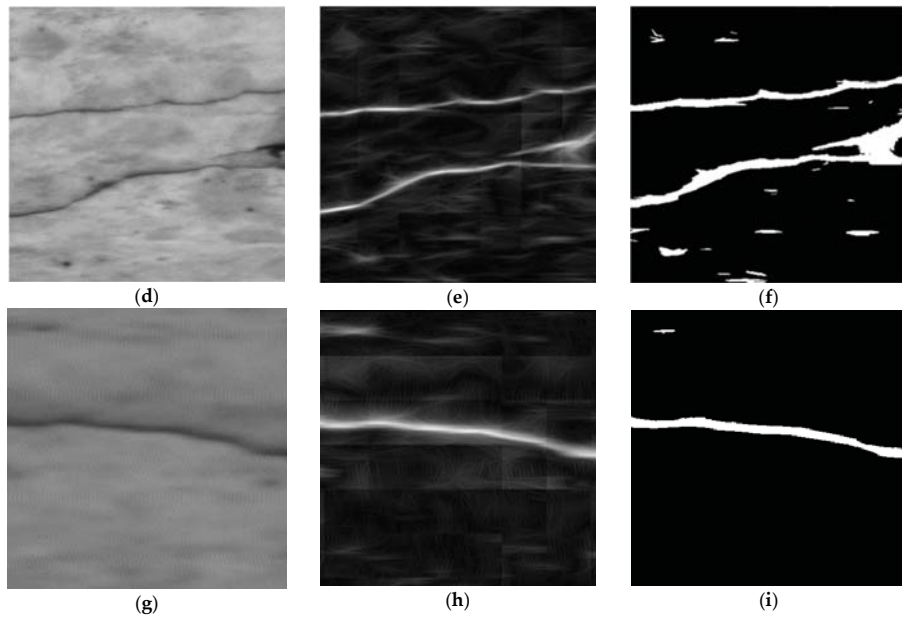(**a**)　　　　　　　　　　(**b**)　　　　　　　　　　(**c**)

Figure 7. *Cont.*

Figure 7. Example of application of the rotation-invariant Gabor filter. (**a**,**d**,**g**) are the normalized original images. (**b**,**e**,**h**) are the filtered images; (**c**,**f**,**i**) the segmented images.

A comparison of the obtained results with those obtained by other authors can be approached by taking into account the fact that, currently, there are no open datasheets available about tunnel cracks, and even the way to evaluate the results is not always the same. Some works use Gabor filters for the detection of cracks, as in [20,25]. In these works, instead of analyzing the result at pixel level, the images are divided into portions, and the detection of cracks is approached in each portion. A balanced accuracy of about 90% is reported in these works, to which the 95% obtained in this work compares favorably with [20,25]. Other works used Convolutional Neural Networks for the detection of cracks at pixel level, as can be found in [15]. The sensitivity obtained in [15] is 92.5%, while precision is 87.0%, which compares favorably to the 95.3% and 98.8%, respectively, obtained in the present paper. In the case of crack detection in the concrete surface of the tunnel, as in [8], the authors use different processing techniques and compare them to each other. The images are captured in a subway tunnel, and cracks wider than 0.3 mm and longer than 15 cm can be detected. The balanced accuracy value is not computed in this work, but the reported accuracy value, which is always equal to or greater than the balanced accuracy, ranges from 88.7% to 91.7%. Therefore, our method would compare favorably to these works.

The proposed algorithm has also been tested with larger regions of the tunnels not used in the training process. The results obtained are satisfactory when compared to the visual inspection carried out by an expert: all cracks have been detected with a limited number of false positives. Figure 8 shows these results for a 1024 × 1714 pixel image (the maximum value is 7.09).
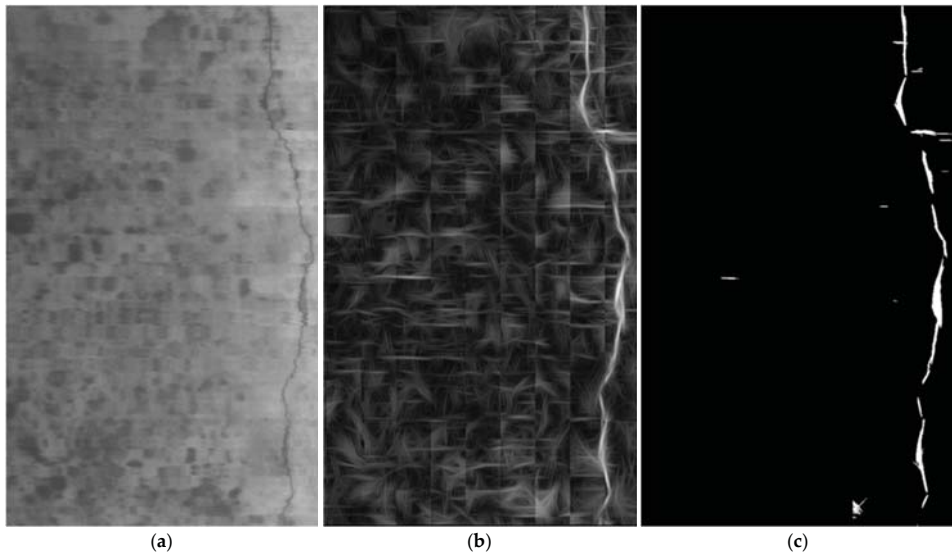
Figure 8. Result of applying the proposed algorithm to a large area of the tunnel. (**a**) Original image; (**b**) filtered image; (**c**) segmented image.

7. Conclusions

In this paper, a computer vision–based methodology for the inspection of fissures and cracks on tunnels whose surface is made of concrete is presented. A system architecture adapted to different tunnel types and a method for processing the acquired images toward detecting said cracks in the tunnel surface are proposed.

A system prototype based on the proposed principles has been developed that uses linear sensor cameras and proper lighting, along with the software required for the system configuration and control, the data acquisition, and the display of the processing results. This prototype has been used to acquire tunnel surface images that have been used for training and testing the proposed processing method.

This method composes a new image upon the maximum values of images filtered by a Gabor filter invariant to rotation proposed in this work, whose parameter values have been set using a modified genetic algorithm based on the Differential Evolution optimization method. The obtained results show that our method compares favorably to existing ones, including those using Gabor filters for detecting road cracks and other works focused on the detection of tunnel surface cracks.

Moreover, the eventual occurrence of shadows and humidity in real tunnels hamper those algorithms based on threshold values obtained from previous knowledge (due to the brightness variations). The algorithms based on the space-frequency analysis of the images are less sensitive to changes in image brightness, which allows better results to be obtained in these situations.

The current work is based on the analysis of the sensed visual spectrum images, but other technologies could complement the approach and widen the range of defects that could be detected. For example, the defects under analysis involve geometric anomalies that could be measured, in principle, using 3D sensors. Some advances are being made in this line, but the resolution and precision required for detecting fissures and cracks are difficult to obtain in practice, so these advances are mainly focused on obtaining the 3D profile of the tunnels [26]. On the other hand, thermographic sensors could also provide relevant information, especially in the detection of defects involving moisture, although the acquisition of such images to the required resolution and at a functional speed is still a challenge.

Author Contributions: Roberto Medina has contributed extensively to the entire work, in particular to the design of the Gabor-filter invariant to rotation and the differential evolution algorithm. José Llamas's contribution has focused on the system architecture, the data collection, and the system validation. Jaime Gómez-García-Bermejo and Eduardo Zalama have contributed to the work as scientific directors, monitoring the work progress, analyzing the results, and preparing the paper. Finally, Miguel José Segarra has brought his wide and valuable experience in tunnel construction to the problem statement and the analysis of the obtained results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, Z.; Wang, C.; Sha, A. Study of image-based pavement cracking measurement techniques. In Proceedings of the 9th International Conference on Measurement & Instruments (ICEMI09), Beijing, China, 16–19 August 2009; Volume 2, pp. 140–143.
2. Yazdchi, M.R.; Mahyari, A.G.; Nazeri, A. Detection and Classification of Surface Defects of Cold Rolling Mill Steel Using Morphology and Neural Network. In Proceedings of the 2008 International Conference on Computational Intelligence for Modelling Control & Automation, Viena, Austria, 10–12 December 2008; pp. 1071–1076.
3. Yamaguchi, T.; Hashimoto, S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Mach. Vis. Appl.* **2010**, *21*, 797–809. [CrossRef]
4. Miyamoto, A.; Konno, M.; Bruhwiler, E. Automatic crack recognition system for concrete structures using image processing approach. *Asian J. Inf. Technol.* **2007**, *6*, 553–561.
5. Zhao, J.; Yang, Y.; Li, G. The Cold Rolling Strip Surface Defect On-Line Inspection System Based on Machine Vision. In Proceedings of the Second Pacific-Asia Conference on IEEE Circuits, Communications and System (PACCS2010), Beijing, China, 1–2 August 2010; pp. 402–405.
6. Oliveira, H.; Correia, P.L. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the 17th European Signal Processing Conference (EUSIPCO'09), Glasgow, UK, 24–28 August 2009; pp. 622–626.
7. Martins, L.A.O.; Padua, F.L.C.; Almeida, P.E.M. Automatic detection of surface defects on rolled steel using Computer Vision and Artificial Neural Networks. In Proceedings of the IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; pp. 1081–1086.
8. Zhang, W.; Zhang, Z.; Qi, D.; Liu, Y. Automatic Crack Detection and Classification Method for Subway Tunnel Safety Monitoring. *Sensors* **2014**, *14*, 19307–19328. [CrossRef] [PubMed]
9. Mahler, D.S.; Kharoufa, Z.B.; Wong, E.K.; Shaw, L.G. Pavement distress analysis using image processing techniques. *Comput. Aided Civ. Infrastruct. Eng.* **1991**, *6*, 1–14. [CrossRef]
10. Rivera, J.P.; Josipovic, G.; Lejeune, E.; Luna, B.N.; Whittaker, A.S. Automated Detection and Measurement of Cracks in Reinforced Concrete Components. *ACI Struct. J.* **2015**, *112*, 397–406. [CrossRef]
11. Fujita, Y.; Hamamoto, Y. A robust automatic crack detection method from noisy concrete surfaces. *Mach. Vis. Appl.* **2011**, *22*, 245–254. [CrossRef]
12. Li, Q.; Liu, X. Novel Approach to Pavement Image Segmentation Based on Neighboring Difference Histogram Method. In Proceedings of the IEEE International Congress on Image and Signal Processing (CISP2008), Sanya, China, 27–30 May 2008; Volume 2, pp. 792–796.
13. Dong, L.; Yu, G.; Ogunbona, P.; Li, W. An efficient iterative algorithm for image thresholding. *Pattern Recognit. Lett.* **2008**, *29*, 1311–1316. [CrossRef]
14. Gavilán, M.; Balcones, D.; Marcos, O.; Llorca, D.F.; Sotelo, M.A.; Parra, I.; Ocaña, M.; Aliseda, P.; Yarza, P.; Amírola, A. Adaptive Road Crack Detection System by Pavement Classification. *Sensors* **2011**, *11*, 9628–9657. [CrossRef] [PubMed]
15. Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712.
16. Abdel-Qader, I.; Abudayyeh, O.; Kelly, M. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. *J. Comput. Civ. Eng.* **2003**, *17*, 255–263. [CrossRef]

17. Nishikawa, T.; Yoshida, J.; Sugiyama, T.; Fujino, Y. Concrete crack detection by multiple sequential image filtering. *Comput. Aided Civ. Infrastruct. Eng.* **2012**, *27*, 29–47. [CrossRef]

18. Paar, G.; Caballo-Perucha, M.P.; Kontrus, H.; Sidla, O. Optical crack following on tunnel surfaces. In Proceedings of the SPIE 6382, Two- and Three-Dimensional Methods for Inspection and Metrology IV, 638207, Boston, MA, USA, 12 October 2006.

19. Liu, Z.; Suandi, S.A.; Ohashi, T.; Ejima, T. Tunnel crack detection and classification system based on image processing. In Proceedings of the SPIE 4664, Machine Vision Applications in Industrial Inspection X, 145, San Jose, CA, USA, 14 March 2002.

20. Zalama, E.; Gómez-García-Bermejo, J.; Medina, R.; Llamas, J. Road Crack Detection Using Visual Features Extracted by Gabor Filters. *Comput. Aided Civ. Infrastruct. Eng.* **2014**, *29*, 342–358. [CrossRef]

21. Daugman, J.G. Uncertainty relation for resolution in space, spatial-frequency and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.* **1985**, *2*, 1160–1169. [CrossRef]

22. Storn, R.; Price, K. Differential Evolution—A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *J. Glob. Optim.* **1997**, *11*, 341–359. [CrossRef]

23. Powers, D.M.W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

24. Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. The balanced accuracy and its posterior distribution. In Proceedings of the 20th International Conference on Pattern Recognition (ICPR2010), Istanbul, Turkey, 23–26 August 2010; pp. 3121–3124.

25. Salman, M.; Mathavan, S.; Kamal, K.; Rahman, M. Pavement crack detection using the Gabor filter. In Proceedings of the 16th International IEEE Conference on Intelligence Transportation System (ITSC 2013), The Hague, The Netherlands, 6–9 October 2013; pp. 2039–2044.

26. Zhan, D.; Yu, L.; Xiao, J.; Chen, T. Multi-Camera and Structured-Light Vision System (MSVS) for Dynamic High-Accuracy 3D Measurements of Railway Tunnels. *Sensors* **2015**, *15*, 8664–8684. [CrossRef] [PubMed]

A Compact Energy Harvesting System for Outdoor Wireless Sensor Nodes Based on a Low-Cost In Situ Photovoltaic Panel Characterization-Modelling Unit

Diego Antolín, Nicolás Medrano *, Belén Calvo and Pedro A. Martínez

Group of Electronic Design (GDE), Aragón Institute for Engineering Research (I3A),
Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, C/ Pedro Cerbuna 12,
Zaragoza 50009, Spain; dantolin@unizar.es (D.A.); becalvo@unizar.es (B.C.); pemar2@unizar.es (P.A.M.)
* Correspondence: nmedrano@unizar.es; Tel.: +34-876-55-33-58

Abstract: This paper presents a low-cost high-efficiency solar energy harvesting system to power outdoor wireless sensor nodes. It is based on a Voltage Open Circuit (VOC) algorithm that estimates the open-circuit voltage by means of a multilayer perceptron neural network model trained using local experimental characterization data, which are acquired through a novel low cost characterization system incorporated into the deployed node. Both units—characterization and modelling—are controlled by the same low-cost microcontroller, providing a complete solution which can be understood as a virtual pilot cell, with identical characteristics to those of the specific small solar cell installed on the sensor node, that besides allows an easy adaptation to changes in the actual environmental conditions, panel aging, etc. Experimental comparison to a classical pilot panel based VOC algorithm show better efficiency under the same tested conditions.

Keywords: maximum power point tracking; solar panel; artificial neural network; PV modelling; energy harvesting; wireless sensor network; cyber-physical systems

---

1. Introduction

Energy harvesting (EH) is a main research field in cyber-physical systems reliability, especially in portable and unattended environmental sensing applications, such as wireless sensor networks (WSN) deployed on extensive outdoor areas monitoring ambient conditions as temperature, light, pressure, and humidity or air pollutants. Figure 1 shows a typical power consumption profile for a sensor node within a WSN intended for environmental monitoring. Most of the time the nodes are set to a low power state (sleep mode in Figure 1), being periodically wake up to acquire the desired environmental parameters, which are then conveyed to the network host. The slow variation of environmental conditions allows programming sleep times of tens of minutes, while the power-hungry measurement and radio frequency (RF) transmission processes take a few seconds or less. Depending on the network configuration, some of these wireless nodes can be configured to receive data from the neighbouring devices, thus extending the surveyed area using intermediate routers for data transmission in a multi-hop architecture, with an additional cost in energy due to the RF transmission. According to the current energy available in the node, a suitable power management requires the use of an optimized cross-layer framework [1] for a precise adjustment of several main parameters in the sensor node as duty cycle, RF power transmission, node data routing and energy harvesting from the environment. Actually, a suitable node sensor power management together with an efficient energy harvesting technique is critical to increase the sensor network autonomy while reducing the required maintenance operations.

In these WSN monitoring applications in natural areas, reviewing the different techniques to harvest energy from the environment, solar energy is the preferential choice because of its availability and predictability [2–6]. The solar power is converted to electric energy through low-cost and low-size photovoltaic (PV) panels. Thus, to successfully design an efficient EH system, one first key issue is having an accurate PV model to reliably predict the panel behaviour under the expected working conditions, with the objective of optimizing the energy transfer from the panel to the energy storage device, by selecting the suitable electric load connected to the panel as a function of the maximum generated power, which in turn depends on the physical solar cell characteristics and ambient working conditions.
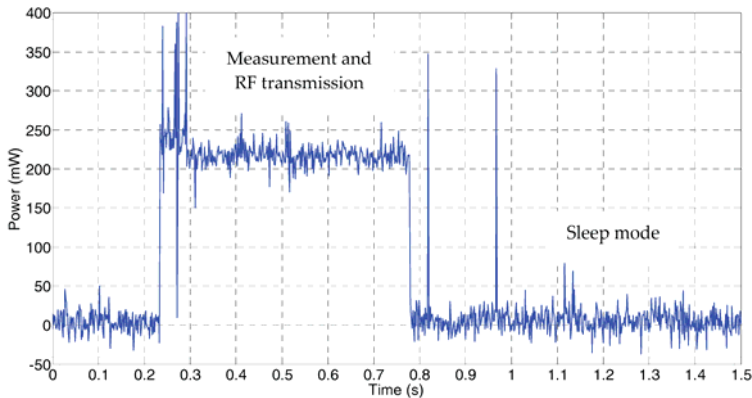

Figure 1. Typical power consumption profile for an environmental wireless sensor node.

The behavioural description of large-size high-power PV panels and PV arrays is typically based on simulation/modelling algorithms using Matlab and/or Simulink [7–10], specific high level mathematical simulation models [11,12], or electrical–based models [13] that extract the required parameters from the manufacturer datasheet, from experimental PV characterization data (this one being the optimum choice) or from a combination of both. A PV panel characterization system must measure the power-to-voltage characteristic under different irradiance and temperature values, at different working conditions obtained by changing the electric load connected to the panel. Thus, a complete PV panel characterization station requires the use of bulky and expensive instruments (solar simulators, pyranometers, programmable power loads), so that these infrastructures are only worthy for large power photovoltaic panels.

In the case of low-cost small PV cells, such as those usually used in wireless sensor nodes, models provided by manufacturers are limited to an approximate behaviour without an accurate description, thus jeopardizing the related efficiency in the power management and transmission system. To overcome this limitation, this work first introduces a low-cost and compact PV panel characterization system, designed to be included as a part of the sensor node. The acquired experimental data are then used to complete a PV panel modelling based on the use of artificial neural networks (ANN). With this approach, it is possible to characterize in situ the photovoltaic cell, and re-characterizing it periodically, upgrading the panel model according to the actual working parameters, and including effects of efficiency lost due to aging, solar radiation variations related to seasonality or dust deposition over the active area. The final goal is the implementation of a maximum power point tracking system based on a Voltage Open Circuit (VOC) algorithm that models the solar panel to reproduce its open-circuit behaviour under the real environmental working conditions in order to speed up the load of an energy storage system in low form factor wireless sensor nodes, maintaining as objectives maximum efficiency, low cost and minimization of visual impact.

The paper is organized as follows: Section 2 presents the compact and portable PV panel characterization system designed for locally accomplish parameter extraction in low-cost solar panels, utilizing a small form-factor PV device that powers a wireless sensor node. Section 3 addresses the modelling of the photovoltaic panel, with a multilayer perceptron neural network as fitting tool using the data collected from the previous characterization system. Section 4 presents the energy harvesting system for the sensor node mounting the panel, based on a DC-DC boost converter with a maximum power point tracking VOC algorithm that relies on the developed PV modelling approach to track the open circuit voltage. Experimental results confirm the achievement of better efficiency compared to a classical VOC technique based on the use of a pilot panel, thus allowing faster charge of the energy storage device.

2. In Situ Photovoltaic Panel Characterization

The proposed PV panel characterization system follows the classical approach, adapted for low-cost solar panels and it owns the characteristic of portability: it allows one to characterize the PV cell in the same site where the device is deployed, thus providing a local accurate behavior model that considers the real external parameters affecting the PV panel operation (light, temperature), as well as the real voltage and current provided by the panel at different programmable loads.

It consists of three main elements (Figure 2): a compact variable load connected to the device under test; a control and acquisition system that selects the load value connected to the PV cell at each moment and reads the data provided by the PV cell at actual temperature and light conditions; and the system software, which coordinates the full system operation. The PV panel selected is a MC-SP0.8-NF-GCS from Multicomp (Premier Farnell, Leeds, UK); its main parameters are: 800 mW maximum power, 4.8 V open circuit voltage and a temperature operation range from $-40\,°C$ to $+85\,°C$.
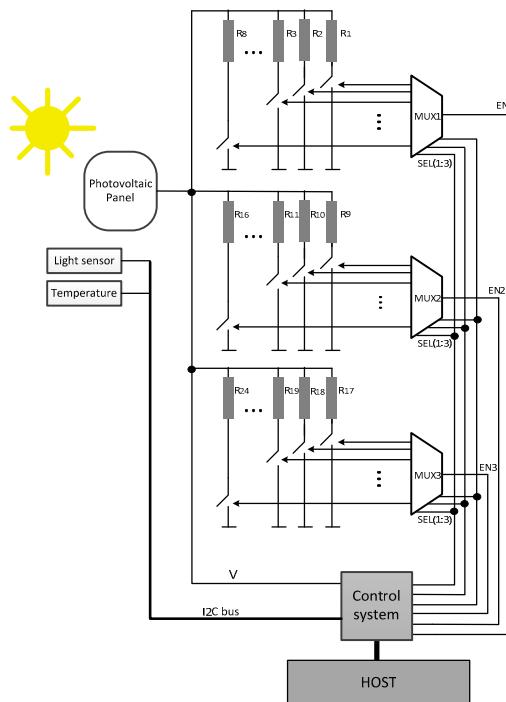

Figure 2. Proposed PV panel characterization system.

2.1. Variable Load Module

The custom variable load consists on a set of 24 fixed value resistors (Table 1), which can be selected by means of three 1:8 demultiplexers that enable or disable a set of ultra-low on-resistance MOSFET transistor switches in series with the loads, connecting the PV panel to each selected resistor at a time (Figure 2). The selection of resistor values attends to the PV panel behaviour, covering from near open circuit for high resistors, to short circuit for the lowest resistor value.

Table 1. Resistor values in the variable load module.

SEL(1:3)	R_{MUX1} (Ω)	R_{MUX2} (Ω)	R_{MUX3} (Ω)
000	10,000	1800	62
001	8200	1200	33
010	7500	1000	16
011	6200	820	8.2
100	5100	680	3.9
101	4300	430	2
110	3300	220	1
111	2200	130	0.1

Their overall real values have been measured using a Keysight (Santa Rosa, CA, USA) 34410A 6 $\frac{1}{2}$ digital multimeter (DMM) using four-wire technique. In this way, the accurate knowledge of the resistor values allows estimate the value of the power transferred to the load at a certain voltage without requiring the current measurement, simplifying the data acquisition. The operation of this module is controlled by a low-cost microcontroller that selects the panel load and reads the measurements, locally storing the values before being sent to a host system for its processing.

Note that a finer load variation resolution can be programmed. However, this possibility added high complexity at computing and managing level, while a very low increase of accuracy in the panel modeling was obtained and was therefore withdrawn.

2.2. Control and Acquisition System

The core of this block is a 16 bits MSP430FR5739 microcontroller from Texas Instruments (Dallas, TX, USA). It selects at each moment the resistor to be connected from the load array to the PV module and acquires the main variables involved in the photovoltaic panel characterization: voltages in the panel output, temperature and illuminance. The microcontroller clock rate is set to 24 MHz, letting assume that both temperature and illuminance keep constant while sweeping the 24 resistors in the load module, thus simplifying the measurement process. In addition, this microcontroller includes 10-bit analog-to-digital (ADC) converters.

Environmental temperature is measured by means of a TMP112 digital sensor (Texas Instruments, Dallas, TX, USA). Its measurement range spans from -40 °C to +125 °C, which covers the temperature modelling range (-10 °C to +70 °C) at a suitable accuracy (from 0 °C to 65°C the sensor accuracy is ±0.5 °C, so sensor readings keep this accuracy in almost the full working range, only giving some additional (acceptable) error at the upper and lower limits) and its bias voltage is compatible with the microcontroller requirements.

Considering the measurement of incident solar light, systems designed for industrial solar panel characterization monitor irradiance using pyranometers. Irradiance is a physical magnitude that measures incident solar power per unit area and considers all incident wavelengths. However, pyranometers are bulky and expensive instruments, what make them unsuitable for compact and low-cost systems. Hence, the proposed system incorporates a compact and low-cost MAX44009 illuminance sensor from Maxim Integrated (San Jose, CA, USA). Unlike irradiance, illuminance is a photometric parameter which depends on the human eye sensitivity to colors and has no proportional relation to irradiance. In local measurements, light spectral composition can be assumed as constant

during the most of the daylight duration and the relation between photometric and ratiometric magnitudes can be considered fixed [14]. Therefore, if the energy harvesting system is calibrated for a specific location, illuminance offers a suitable measurement of the relative incident solar power for that place. The proposed illuminance sensor presents an input range from 0.045 lux to 188,000 lux, widely covering the illuminance for an average sunny day. In addition, bias voltages are compatible with those of the microcontroller. Data are provided to the microcontroller in a 12-bits floating point representation via I2C protocol.

2.3. System Software

The control and acquisition system is designed to be controlled by an external host. Before its deployment, a first characterization can be performed using a computer as host system, though after placement in its definitive location the system is designed to be managed by the sensor node where it is connected. Figure 3 shows the dataflow implemented in the microcontroller. First, the microcontroller configures its peripherals and initializes variables, next going to a low-power state until the host device sends an interrupt request through the serial port. The first interrupt commands the microcontroller to run the measurement process, which starts acquiring temperature and illuminance. Next, the panel output voltage measurements are taken, consecutively selecting the 24 resistors in the variable load. Considering switching, voltage settling time and ADC conversion time using the microcontroller at a 24 MHz working frequency this process takes less than 20 ms, so that both temperature and illuminance can be considered constant along the measurement process.
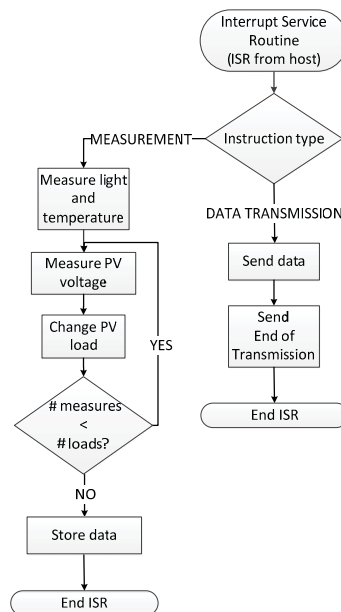

Figure 3. System software dataflow.

Figure 4 shows a photograph of the characterization board connected to the solar panel. Illuminance and temperature readings are acquired from the corresponding sensors located in the upper-right side of the PV panel. The power transmitted by the PV panel is estimated from the output voltage readings over the 24 selectable resistors in the variable load (which are distributed in the PCB, see Figure 4). Figure 5 shows the output voltage changes in the solar panel due to variations in the

connected load value (green channel). The PV output voltage changes with the resistor load as shown in this figure, where the non-linear region corresponds to resistor values in the range of few kilo-ohms to ohms, in the local irradiance and temperature conditions where the system was tested, being this region directly related to the maximum power point area of interest. The full measurement time is given as the time between cursors a and b in Figure 5, shown as the high level in the upper magenta signal. Though the first 12 ms the measured voltage (green channel) seems to be constant, this is an effect due to the oscilloscope scales; the values measured by the characterization system are different and provide valid data for the panel modelling. The non-monotonicity visible as two glitches in the non-linear region of the load signal appears random and occasionally. These errors can be avoided including a small additional delay between each measurement. However, because these measurements are collected to be used as training data for an artificial neural network to model the PV panel, and considering the robustness of these modelling techniques to local data errors, in order to ensure that environmental parameters (light and temperature) are constant we have preferred keep those times as low as possible. Additionally, this measurement times allows using 0.5 W power dissipation resistors, below the maximum power that the panel can provide (0.8 W) without integrity damage risk, reducing cost and size.
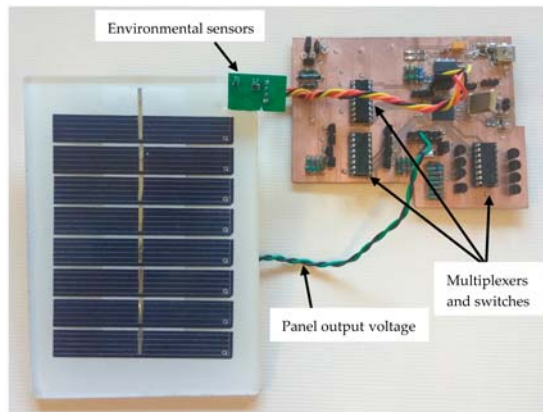

Figure 4. Characterization system connected to the solar panel (control host not shown).
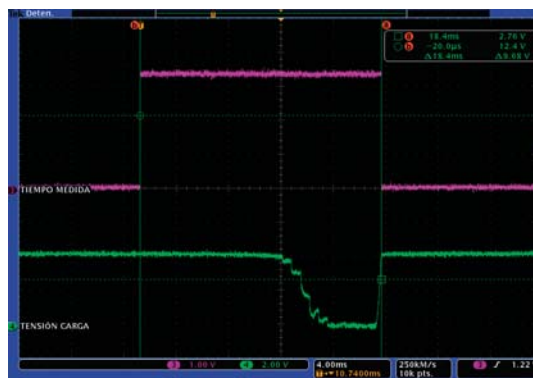

Figure 5. Panel output voltage values (green) in the characterization process sweeping resistor values in the PV panel load. Rising and descent slope in purple signal indicates the start and end time in the measurement. Time span is below 20 ms (see cursor measures).

After acquisition, data are locally stored waiting for a transmission command from the host. If a new measurement instruction is received from the host, previous data are overwritten.

2.4. Host Control

The panel characterization system operation is designed to be controlled by considering as host device the sensor node where the panel is connected after its deployment in the area to be monitored. In this manner, the periodic PV panel characterization can be included as an additional task in the sensor node operation, thus achieving an updated and accurate local model of the solar panel, with an in situ periodical model recalibration using real temperature, illuminance and transferred power. The re-characterization period can be set according to the user requirements, programming its periodicity from the WSN host by means of the corresponding command. In addition, defining warning conditions a non-periodical characterization can be triggered either by the sensor node itself or by the network host.

The host operation for the panel data acquisition process is the following: after configuration of the serial port that connects to the characterization system, the host sends a command to start the measurement process and waits for a time *t* to the end of this operation. In this time (stated 1 s), the characterization system performs the measurements of temperature, illuminance and power transferred to the resistors in the load array as indicated in Figure 3. Next, the host device sends a data transmission command to the control and acquisition system, receiving the data. This process is recurrently executed with a period of T s, which can be modified according to the requirements, for a sufficiently long time to ensure that measures are available with adequate variability in temperature and illuminance to construct a suitable panel model.

3. ANN-Based Photovoltaic Panel Modelling

Using the data collected by means of the proposed characterization system, a numerical model of the solar panel used in this work has been developed. The behavioral model obtained is locally valid under the ambient conditions in which measurements have been taken, and can be updated by collecting new measurements with the characterization system previously presented. This choice simplifies the model, allowing its implementation in compact low-power systems where a microcontroller is the computing core.

The PV panel model has been developed using an artificial neural network (ANN) as fitting technique. ANNs are suitable tools for system modelling in tasks where enough data are available, and no accurate relationship between input and output magnitudes in the modelled system is known. In fact, ANNs applied to power solar panel modelling have been recently reported in the literature [15–17].

Among the variety of ANN models, the multilayer perceptron (MLP) is the most common in system modelling due to its universal function estimator characteristic in non-linear fitting tasks, thanks to the use of sigmoid (nonlinear) output functions in its processing nodes [18], and the use of advanced parameter adjustment techniques, as the Levenberg-Marquardt algorithm [19].

Data used to perform the PV model were collected for a week. The ANN input and output data were normalized in the range [−1, +1], considering that environmental inputs (temperature and illumination) as well as output voltage are limited to those shown in Table 2, corresponding to the environmental limits in the region where the panel is working, and the useful panel voltage range. Measures have been acquired every minute, i.e., with a period T = 60 s.

Table 2. Range of parameters used in the solar panel modelling.

Parameter	Upper Limit	Lower Limit
Light (lux)	120,000	10,000
Temperature (°C)	70	−10
Voltage (V)	5	2.8
Panel load (Ω)	10,000	0.1

Figure 6 shows the experimental panel output voltage as a function of the load resistor for three different environmental conditions of illuminance and temperature and its corresponding simulated values provided by a Levenberg-Marquardt trained 3 input – 6 hidden – 1 output MLP, giving very similar results.
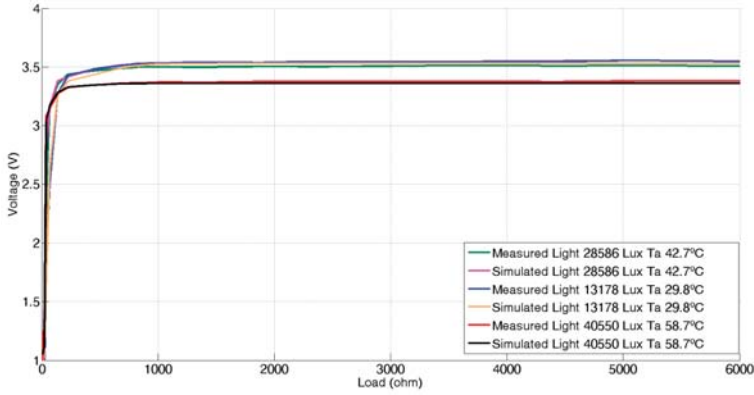

Figure 6. Real and simulated voltage-to-load curves at three different light and temperature conditions.

Once verified the correct fitting, the MLP has been tested to obtain the power-voltage curves that characterize the behaviour of a solar panel. For this, the environmental MLP model inputs (illuminance and temperature) are fixed at the desired values, sweeping the range of the ANN input that represents the panel load to obtain the voltage values from the ANN output. Power is calculated as the output voltage over the corresponding input load value. Figure 7 shows the obtained results. Note that in this modelling, the voltage shows a coupled dependence with illuminance and temperature. This is because, unlike the models developed in laboratory, where temperature and irradiance can be independently controlled, the experimental models based on in-situ environmental measurements account for coupling in temperature and illuminance (i.e., the more illuminance, the more temperature). Therefore, these simpler models are not general, but locally valid and suitable for its use in the proposed application: the development of a solar energy harvesting system based on a local behavioural model to control the maximum power point tracking.
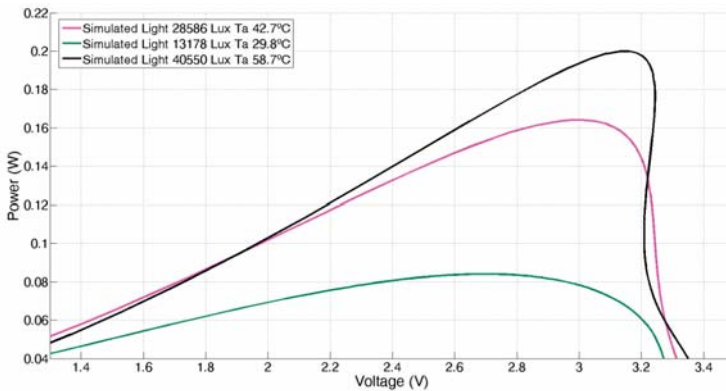

Figure 7. P-V curves obtained from the ANN-based solar panel model.

4. Solar Energy Harvesting System Architecture

Figure 8 shows the block diagram of a typical solar-based energy harvesting system. A photovoltaic device converts solar to electrical energy. The direct-current-to-direct-current (DC-DC) converter generates from the PV device a stabilized voltage, that feds the powered system, comprising an energy storage system (battery, supercapacitor), and the electronic system itself. A maximum power transfer circuit (called maximum power point tracking system, MPPT) tracks environmental and electrical magnitudes from the PV system, and modifies in real time the DC-DC architecture to keep an optimum energy transfer. The monitored variables are selected according to the implemented MPPT technique.
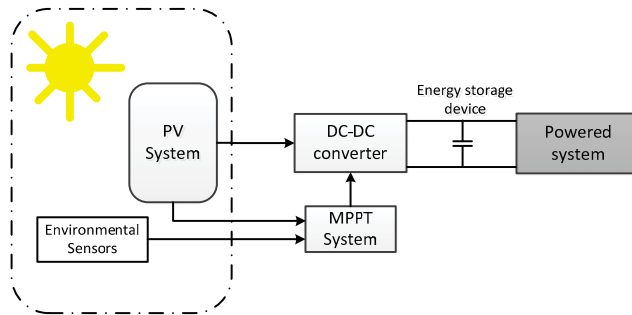

Figure 8. Generic solar energy harvesting block diagram.

The DC-DC topology depends on the input and output voltages, provided by the solar panel and required by the powered system, respectively. The solar panel provides a maximum voltage of 3.85 V. The energy storage device selected is a supercap, due to its simple control electronic requirements; conversely, batteries require complex control electronics to avoid overload and deep discharges that can damage the load device, as well as higher loading currents, that demand higher panels, unsuitable for the small form factor designs used in wireless CPS. The selected supercap, as will be detailed next, has a maximum input voltage of 5.5 V, suitable for the powered environmental nodes [20]. Under these conditions, the DC-DC architecture has to be a boost converter.

Figure 9 shows the boost DC-DC converter architecture. When switch S1 is ON, the diode is in reverse mode, thus the sensor node (R_{load}) is powered by the energy stored in the supercap SC, while the inductor L1 is energized from the PV output. When switch S1 is OFF, the output voltage V1 from the PV panel and the energy stored in the inductor load up the supercap voltage (V2), which provides energy directly to the load. Capacitor C_{in} filters abrupt changes in the output voltage V1 due to variations in the working conditions of PV panel. The component values (Table 3) were estimated using classical techniques [21,22], assuming the system works in continuous conduction mode (CCM) [23].

The value of the supercap must be selected to provide the required energy to the electronic system (the sensor node) for a minimum operating life, even in the absence of light. In our case, the selected device is a 1.5 F–5.5 V EECF5R5U155 supercap from Panasonic (Osaka, Japan), with electrical characteristics compatible with the low dropout (LDO) regulator used in the wireless nodes presented in [20]. The 1.5 F capacitance value guarantee a reduced node lifetime (about 12 h for the aforementioned node), but has been selected to allow a fast and complete characterization of the proposed energy harvesting system. The energy storage capacity can be increased just increasing the capacitance value.
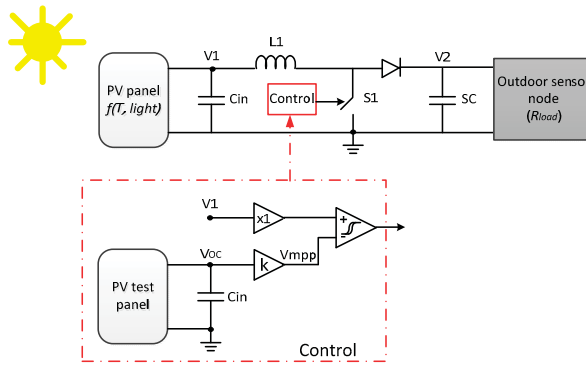
Figure 9. Boost DC-DC circuit. In red, a dual solar panel MPPT control block diagram.

Table 3. Component values for the DC-DC boost converter.

Component	Value
Cin	>25 μF
L1	>6.7 mH
Diode	HSMS-2800
Switch	IRFML8422 (MOSFET transistor)

4.1. Maximum Power Point Tracking

The maximum power transfer is given at impedance matching, i.e., when the DC-DC regulator presents an input impedance equal to the output impedance of the solar panel. Because the output impedance in the solar panel dynamically depends on its intrinsic characteristics and environmental conditions, impedance fitting requires a dynamic technique able to adapt the DC-DC topology. For this, energy harvesting systems make use of maximum power point transfer (MPPT) algorithms [24], which monitor several different electrical and environmental parameters to come to maximize the power transfer at all times.

The Voltage Open Circuit (VOC) is a widely used MPPT algorithm that tracks the voltage at which the power transfer is maximized. Experimentally, it is known that this voltage is proportional to the system at open circuit, according to:

$$V_{mpp} = k \cdot V_{oc} \tag{1}$$

where V_{mpp} is the voltage at maximum power point transfer, V_{oc} is the voltage at open circuit (which depends on the incident light and temperature), and k is a constant. The value of k depends on the panel model; its value is typically limited to 0.7–0.8 [25–27], and can be experimentally estimated by characterizing the solar panel (see Figure 7). In our case, $k = 0.8$.

Therefore, in the VOC MPPT algorithm the tracking of the open circuit voltage is required. For this, there are two different strategies: (1) to add an extra identical solar panel without load to act as a reference or (2) to periodically disconnect the solar panel from the converter and read its voltage at open circuit. Both options present drawbacks: in the first case, an additional cost and increase of system size; in the second, the DC-DC is periodically leaved without power for the time the voltage is measured, losing efficiency.

In the dual panel VOC technique (Figure 9), the control system monitors the open-circuit voltage in the reference PV panel (V_{oc}), estimating the maximum power transfer voltage V_{mpp} (Equation (1)), which is compared to the working PV panel voltage *V1*. When *V1* is lower than V_{mpp}, the switch S1 is turned off, increasing the load impedance, and thus increasing *V1*. Otherwise, if *V1* > V_{mpp}, the switch

is on, reducing the load impedance. To avoid fast and continuous changes in the switch state, the comparator module presents hysteresis.

4.2. Maximum Power Point Tracking Technique Based on ANN Modelling

As it was stated before, the use of an additional PV panel to control the switching in the MPPT algorithm presents several drawbacks, mainly cost, size and the need to accurately place both panels in the same orientation to allow right behaviour comparison. In this work, we propose to replace the extra solar panel by the model obtained from an in-situ characterization (Figure 10) achieved using the methodology shown in Section 3. In this way, the proposed tracking method allows continuous energy transfer from the solar panel to the energy storage without the need of additional PV device. This model will be implemented in a low-power microcontroller incorporated to the sensor node where the panel is connected, and it must be capable of simulating the full neural-based model fast enough.
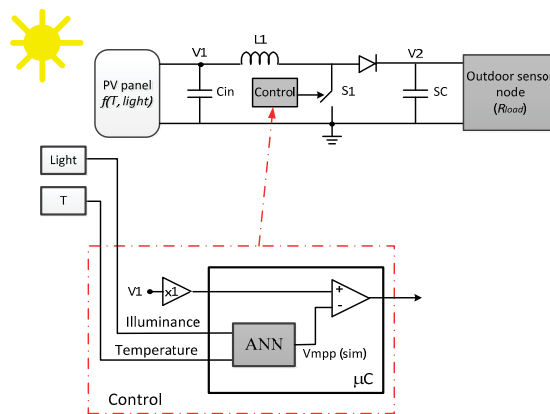

Figure 10. Proposed ANN-based VOC MPPT.

Following similar considerations to that presented in Section 3, the neural model selected is a MLP. In this case, in order to be implemented in a low-cost microcontroller, several constrains next detailed have been applied to the number of input parameters and non-linear function computational complexity, so as to simplify the model while preserving accuracy.

The simulator in the Control block (Figure 10) estimates the V_{mpp} of the panel from the readings of the sensors of illuminance and temperature; this limits the number of inputs to two. Thus, the PV panel behaviour is modelled by a 2-2-1 MLP with non-linear output function in the hidden layer and linear operation in the output layer. This neural network has been implemented in the same microcontroller used in the panel characterization system (Section 2), a MSP430FR5739 (Texas Instruments, Dallas, TX, USA), which includes a 32 bit hardware multiplier. In this way, the computational resources are shared with the solar panel characterization module. Weight values are represented in 16 bit floating point, and data are normalized in the $[-1, +1]$ range, limiting the training data to the values shown in Table 2. The training data set is generated from the complete ANN panel model previously shown in Section 3, just sweeping the load value for each illuminance-temperature couples (in the valid ranges of the panel location). Therefore, the maximum voltage V_{oc} is obtained for each pair of illuminance—temperature values and, hence, the corresponding V_{mpp} is straightforwdly derived.

The MLP has been trained in off chip mode, using hyperbolic tangent as non-linear operation in the hidden layer, leaving only the execution phase for its electronic implementation. Figure 11 shows the MLP architecture with the resulting weights after training.

Due to its inherent computational complexity, the most important limitation when a neural model is electronically implemented is the non-linear operation in the hidden processors. In this case, the hyperbolic tangent has been replaced by a five sections piecewise polynomial approach given by:

$$
y = f(x) = \begin{cases}
-1, \ x \leq -5.3 \\
0.0127 \cdot x^3 + 0.1435 \cdot x^2 + 0.5234 \cdot x - 0.3840, \ -0.9 > x > -5.3 \\
-0.2358 \cdot x^3 + 0.9834 \cdot x, \ -0.9 \leq x \leq 0.9 \\
0.0127 \cdot x^3 - 0.1435 \cdot x^2 + 0.5234 \cdot x + 0.3840, \ 5.3 > x > 0.9 \\
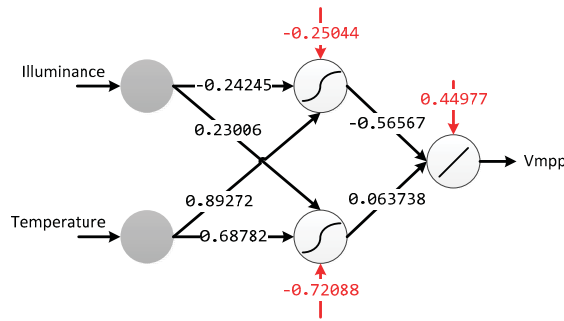+1, \ x \geq 5.3
\end{cases}
\tag{2}
$$

Figure 11. Multilayer perceptron weight values obtained to modelling the solar panel behavior. Red values correspond to bias weights.

And that has been implemented minimizing the operations, as shown in Figure 12, to reduce the computing time. The accuracy of this polynomial approach is shown in Figure 13. Left side shows the representation of a *tanh(x)* computed by Matlab running on a personal computer overlapped to the proposed polynomial approach in Figure 12 implemented on the used microcontroller. Right side in Figure 13 presents the relative error of this approach, which is lower than 2% (except in the zero crossing). Finally, the execution time has been experimentally estimated and compared to the time required for computing the *tanh(x)*, using the standard C mathematical library. Both operations have been implemented in the microcontroller, measuring the execution time switching an output pin from logical '1' to logical '0'. Figure 14 shows the operation time using the *tanh(x)* (a) and polynomial approach (b) for three different input values, corresponding to the input ranges shown in Figure 12. The *tanh(x)* function requires a constant time to perform each of the three operations ($T_A = T_B = T_C = 2.4$ ms), while the polynomial implementation requires a variable time of $T'_C = 740$ μs (worst case, second *if(.)* in Figure 12), $T'_B = 500$ μs (*else* condition) and $T'_A = 33$ μs (best case, first *if(.)*), always improving the execution time.

```
if(abs(x) ≥ 5.3)
        y = sign(x)
else if(abs(x) > 0.9)
        y = 0.0127*x³-sign(x)*0.1435*x²+0.5234*x+sign(x)*0.3840
else
        y = -0.2358*x³+0.9834*x
end
```

Figure 12. Pseudocode representing the piecewise polynomial implementation of equation.
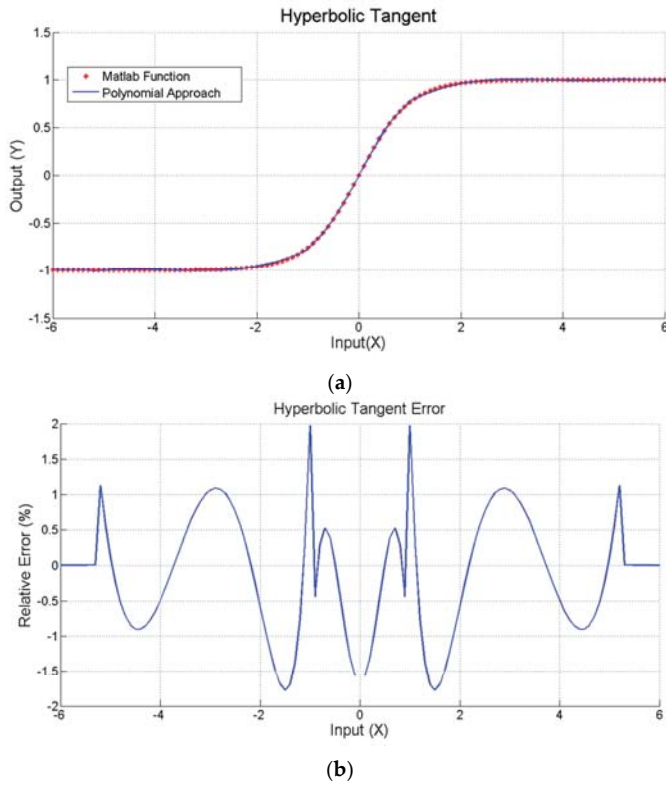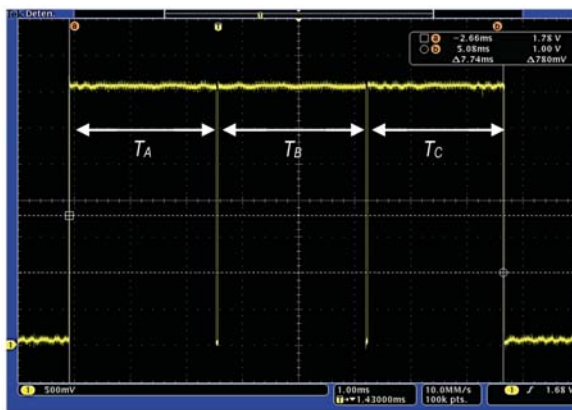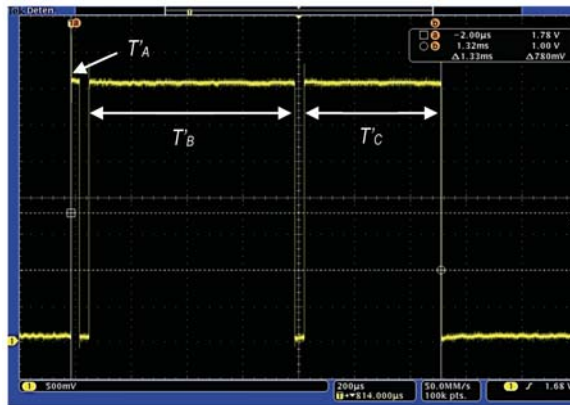
(**a**)

(**b**)

Figure 13. (**a**) Hyperbolic tangent computation using a standard C library *tanh(x)* function (red) and the piecewise approach (blue) on a microcontroller; (**b**) Relative error of the polynomial fitting compared to the library function.

(**a**)

Figure 14. *Cont.*

(**b**)

Figure 14. Computing time measurement for three different input values using (**a**) C library-based *tanh(x)* operation and (**b**) polynomial approach; Times correspond, respectively, to calculate the operations for inputs that meet: abs(x) $\geq$ 5.3 (T_A, T'_A); 5.3 > abs(x) > 0.9 (T_B, T'_B); and abs(x) $\leq$ 0.9 (T_C, T'_C).
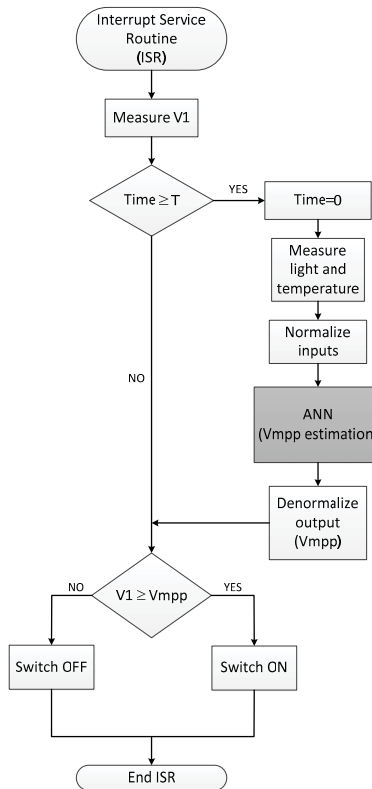

Figure 15. ANN-based VOC MPPT routine flow diagram.

Figure 15 shows the flow diagram of the VOC MPPT routine with ANN modelling of the PV panel. The Control block microcontroller enters every time *t* in the interrupt routine. If the time after the last V_{mpp} estimation is higher than a set time *T* (>*t*), the system updates its value by computing the new V_{mpp} from the ANN model using current illuminance and temperature readings. Next, if the voltage V1 at the panel output is higher than the V_{mpp}, the control switch is turn on reducing the input impedance and, hence, reducing the voltage V1. Otherwise, the switch turns off.

4.3. MPPT Comparison

Figure 16 shows the electronic implementation of the energy harvesting system based on the VOC technique with ANN-based panel modelling. Temperature and illuminance are monitored by the microcontroller (located at the characterization system on top of the figure), that compares the real panel output voltage (sent from the DC-DC system to the microcontroller through the voltage tracking wires), and the ANN-based maximum power point voltage estimation (from temperature and illuminance real time values) to properly switch the value of the Control line of the boost DC-DC converter (centre, bottom). An additional line (in parallel to the Control line in Figure 16) disables the DC-DC circuit to operate the characterization system when it is required. Load tests were performed using one supercap to speed up the comparison process, though a wireless sensor node can require several in parallel [20] for suitable energy storage.

The proposed technique has been compared to a classical analog VOC MPPT to evaluate its performance. In the proposed model, the MPPT management routine is executed every 10 μs (*t* in Figure 15), while the V_{mpp} estimation is upgraded every 100 ms (*T* in Figure 15). Both techniques are executed in parallel over two identical solar panels with the same orientation, seeking for the same working conditions.

Figure 17 shows the dynamic energy load process for a 1.5 F supercap using both classical (blue) and ANN-based (yellow) VOC MPPT techniques at four different environmental conditions. The load process stops when the supercap is fully loaded and its voltage reaches 5.5 volts. In Figure 17, the x axis represents time (1000 seconds full screen for Figure 17a–c, and 2000 seconds for Figure 17d), while y axis shows the voltage evolution with time. Both techniques have been implemented in parallel, placed in identical environmental conditions and orientation. Energy load processes are triggered at the same time, and sensors and DC-DC converters are identical for both systems, being the control architecture the only difference. The proposed system presents a more efficient behaviour under those conditions, considerably reducing the load time.
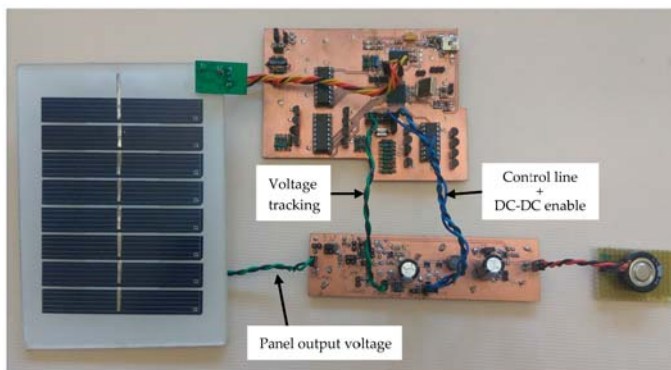

Figure 16. Full solar energy harvesting system based on ANN MPPT and energy storage system (supercapacitor): solar panel (left), control system (centre, on top, includes the characterization system), DC-DC system (centre, bottom) and supercap (right). In this case, only one of the 1.5 F supercaps is connected to the boost system for characterization purposes.
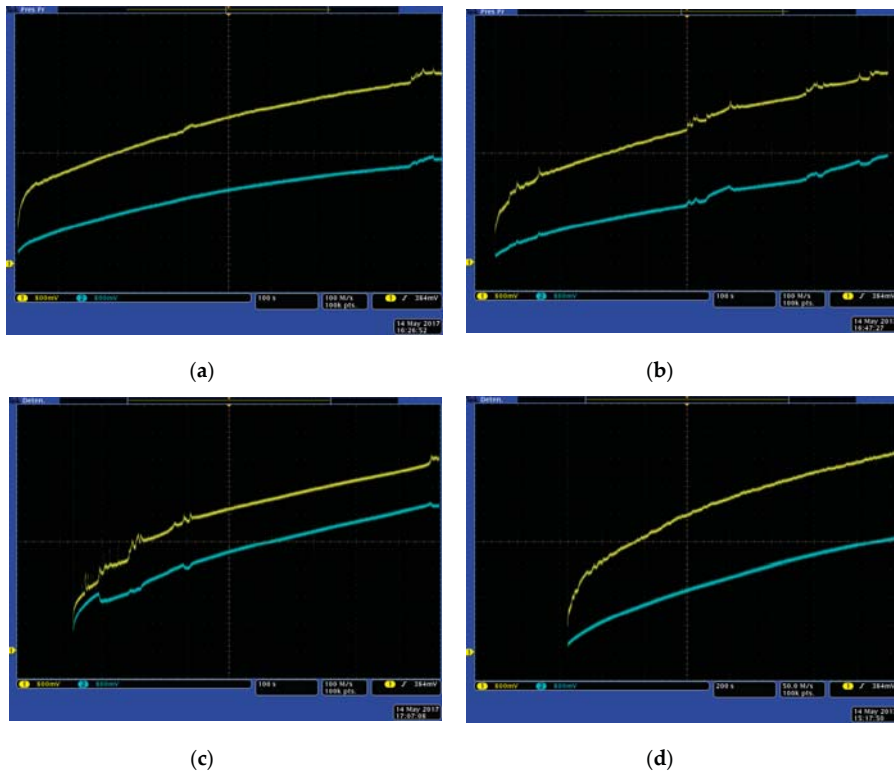
Figure 17. Charge curves for a 1.5 F supercap for a VOC algorithm using a reference panel (blue) and the ANN-based solar panel model (yellow). Environmental conditions (irradiance and temperature, average): (**a**) 850 W/m^2, 35 °C; (**b**) 1300 W/ m^2, 55 °C; (**c**) 1120 W/ m^2, 43 °C; (**d**) 500 W/ m^2, 26 °C.

5. Conclusions

The application of a characterization system for small solar panels powering wireless sensor networks after deployment provides an accurate local behavioural model, which allows monitoring the efficiency changes in the energy generation due to both environmental and aging effects. Taking this into account, this paper has proposed an in situ PV panel modelling system, complying with the severe constraints related to wireless sensor networks in size, cost and energy consumption. Based on the acquired data, a behaviour-based PV model is developed based on a multilayer perceptron; this ANN solution makes easier to incorporate new characterization measures into the model, adapting it to changes in the current operating conditions.

Finally, a boost DC-DC converter based on the maximum power point tracking using the VOC algorithm, where the pilot panel monitoring the open circuit voltage is replaced by a customized behavioural model based on the proposed multilayer perceptron approach, with reduced computing requirements to allow the software model to be implemented in a low-cost microcontroller embedded in the wireless node, that manages both the data acquisition and modelling. Compared to the classical VOC technique, experimental results show an improved performance in loading a supercap as energy storage device. In addition, this technique can be applied to different MPPT algorithms where a solar panel or its behaviour is required in its operation.

Author Contributions: Diego Antolín designed and implemented the system, developed the panel model and experimentally validated the system. Belén Calvo and Nicolás Medrano supervised all the work and analyzed the results. Pedro A. Martínez provided theoretical support to the work. Diego Antolín, Belén Calvo, Nicolás Medrano and Pedro A. Martínez wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, G.; Dong, Y.; Guo, H.; Shu, L.; Wu, D. Cross-layer optimized routing in wireless sensor networks with duty cycle and energy harvesting. *Wirel. Commun. Mob. Comput.* **2015**, *15*, 1957–1981. [CrossRef]

2. Cammarano, A.; Petrioli, C.; Spenza, D. Online Energy Harvesting Prediction in Environmentally Powered Wireless Sensor Networks. *IEEE Sens. J.* **2016**, *16*, 6793–6804. [CrossRef]

3. Zou, T.; Lin, S.; Feng, Q.; Chen, Y. Energy-Efficient Control with Harvesting Predictions for Solar-Powered Wireless Sensor Networks. *Sensors* **2016**, *16*, 53. [CrossRef] [PubMed]

4. Yoon, I.; Kim, H.; Noh, D.K. Adaptive Data Aggregation and Compression to Improve Energy Utilization in Solar-Powered Wireless Sensor Networks. *Sensors* **2017**, *17*, 1226. [CrossRef] [PubMed]

5. Mandourarakis, I.; Koutroulis, E. Optimal design of a boost-type DC-DC converter for PV power-supplied wireless sensor networks. In Proceedings of the 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 17–19 March 2015; pp. 1100–1105.

6. Penella-López, M.T.; Gasulla-Forner, M. *Powering Autonomus Sensors*; Springer: Dordrecht, The Netherlands, 2011.

7. Altas, I.H.; Sharaf, A.M. A Photovoltaic Array Simulation Model for Matlab-Simulink GUI Environment. In Proceedings of the International Conference on Clean Electrical Power, ICCEP '07, Capri, Italy, 21–23 May 2007; pp. 341–345.

8. Khatib, T.; Elemenreich, W. *Modeling of Photovoltaic Systems Using MATLAB: Simplified Green Codes*; John Wiley & Sons: Hoboken, NJ, USA, 2016.

9. Mahmoud, S.A.; Alsari, M.M.; Reda, E.I.; Alhammadi, R.M. MATLAB modeling and simulation of photovoltaic modules. In Proceedings of the 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), Boise, ID, USA, 5–8 August 2012; pp. 786–789.

10. Ropp, M.E.; Gonzalez, S. Development of a MATLAB/Simulink Model of a Single-Phase Grid-Connected Photovoltaic System. *IEEE Trans. Energy Convers.* **2009**, *24*, 195–202. [CrossRef]

11. Can, H.; Ickilli, D.; Parlak, K.S. A New Numerical Solution Approach for the Real-Time Modeling of Photovoltaic Panels. In Proceedings of the Power and Energy Engineering Conference (APPEEC), Shanghai, China, 27–29 March 2012; pp. 1–4.

12. Ulapane, N.N.; Dhanapala, C.H.; Wickramasinghe, S.M.; Abeyratne, S.G.; Rathnayake, N.; Binduhewa, P.J. Extraction of parameters for simulating photovoltaic panels. In Proceedings of the 2011 6th International Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 16–19 August 2011; pp. 539–544.

13. Castañer, L.; Silvestre, S. *Modelling Photovoltaic Systems Using PSpice*; John Wiley & Sons: Chichester, UK, 2002.

14. Bukshtab, M. *Applied Photometry, Radiometry, and Measurements of Optical Losses*; Springer: New York, NY, USA, 2012.

15. Saberian, A.; Hizam, H.; Radzi, M.A.; Ab Kadir, M.Z.; Mirzaei, M. Modelling and Prediction of Photovoltaic Power Output Using Artificial Neural Networks. *Int. J. Photoenergy* **2014**, *1–10*, 469701. [CrossRef]

16. Sánchez-García, J.L.; Espinosa-Juárez, E.; Tapia-Juárez, R. Photovoltaic panel characterization by using artificial neural networks and comparison with classical models. In Proceedings of the 2015 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 4–6 November 2015; pp. 1–6.

17. Karapete, E.; Boztepe, M.; Colak, M. Neural network based solar cell model. *Energy Conver. Manag.* **2006**, *47*, 1159–1178.

18. Haykin, S. *Neural Networks and Learning Machines*; Pearson Education: New Jersey, NJ, USA, 2009.

19. Principe, J; Euliano, N.; Lefebvre, W. *Neural and Adaptive Systems, Fundamentals through Simulations*; John Wiley & Sons: New York, NY, USA, 2000.

20. Antolín, D.; Medrano, N.; Calvo, B. Reliable Lifespan Evaluation of a Remote Environment Monitoring System Based on Wireless Sensor Networks and Global System for Mobile Communications. *J. Sens.* **2016**, *2016*, 1–12. [CrossRef]

21. Mohan, N.; Underland, T.M.; Robbins, W.P. *Power Electronics: Converters, Applications, and Design*; John Wiley & Sons: New York, NY, USA, 2003.

22. Hart, D.W. *Power Electronics*; McGraw-Hill: New York, NY, USA, 2011.

23. Texas Instruments. *Application Note-1994: Modeling and Design of Current Mode Control Boost Converters. Texas Instruments Application Report SNVA408B*; Texas Instruments: Dallas, TX, USA, 2013.

24. Balato, M.; Costanzo, L.; Vitelli, M. *Maximum Power Point Tracking Techniques*; Wiley Online Encyclopedia of Electrical and Electronics Engineering; John Wiley & Sons: New Jersey, NJ, USA, 2016; pp. 1–26.

25. Esram, T.; Chapman, P.L. Comparison of photovoltaic array maximum power point tracking techniques. *IEEE Trans. Energy Conver.* **2007**, *22*, 439–449. [CrossRef]

26. Gupta, A.K.; Saxena, R. Review on widely-used MPPT techniques for PV applications. In Proceedings of the 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), Noida, India, 3–5 February 2016; pp. 270–273.

27. Pressman, A.; Billings, K.; Morey, T. *Switching Power Supply Design*; McGraw-Hill: New York, NY, USA, 2009.

An Efficient Wireless Sensor Network for Industrial Monitoring and Control

**Juan Aponte-Luis [1], Juan Antonio Gómez-Galán [2,*], Fernando Gómez-Bravo [2],
Manuel Sánchez-Raya [2], Javier Alcina-Espigado [1] and Pedro Miguel Teixido-Rovira [1]**

[1] OnTech Security LLC, C/Hispano Aviación, 7-9, 41300 Sevilla, La Rinconada, Spain;
juan.aponte@ontech.es (J.A.-L.); javier.alcina@ontech.es (J.A.-E.); pedro.teixido@ontech.es (P.M.T.-R.)

[2] Department Electronic Engineering, Computers, and Automatic, University of Huelva,
Ctra Huelva-La Rábida, s/n, 21819 Huelva, Spain; fernando.gomez@diesia.uhu.es (F.G.-B.);
msraya@diesia.uhu.es (M.S.-R.)

* Correspondence: jgalan@diesia.uhu.es; Tel.: +34-959217650; Fax: +34-959217348

Abstract: This paper presents the design of a wireless sensor network particularly designed for remote monitoring and control of industrial parameters. The article describes the network components, protocol and sensor deployment, aimed to accomplish industrial constraint and to assure reliability and low power consumption. A particular case of study is presented. The system consists of a base station, gas sensing nodes, a tree-based routing scheme for the wireless sensor nodes and a real-time monitoring application that operates from a remote computer and a mobile phone. The system assures that the industrial safety quality and the measurement and monitoring system achieves an efficient industrial monitoring operations. The robustness of the developed system and the security in the communications have been guaranteed both in hardware and software level. The system is flexible and can be adapted to different environments. The testing of the system confirms the feasibility of the proposed implementation and validates the functional requirements of the developed devices, the networking solution and the power consumption management.

Keywords: remote industry monitoring; wireless sensor network; low power consumption

1. Introduction

Wireless sensor networks (WSN) have emerged as one of the most promising technologies for the future. Some of the potential application domains are healthcare, army, environment and agriculture, industry, transportation systems and security [1–10]. Unlike other networks, WSN are designed for specific applications, and thus, they must satisfy a set of requirements that differs from one application to another. Therefore, researchers must address a variety of challenges to facilitate the widespread deployment of WSN technology in real-world domains. As a result, great efforts have been devoted to overcome the energy-saving problem and reliability [11,12].

Industrial plants integrate sensors connected to the control station through wire and wireless methods for continuous sensing and monitoring the status of the system. In this context, wireless technology gives a suitable support to the industry offering advantages in terms of low installation cost, scalability, flexibility, lack of cabling, intelligent-processing capability, high mobility and ease of deployment compared to conventional wired solutions. These advantages are promising for industry, where a considerable growth in the near future is expected [13].

Nevertheless, the use of WSN sensors for industrial applications requires accomplishing a set of hard constraints. For instance, monitoring and control of a specific process, demands the development of particular network architectures, mechanisms and algorithms that guarantee a high communication quality and reliability of the system [14,15]. In addition, as environmental conditions may be hard,

industry implementations must ensure data reliability at all times. Finally, the design of an industrial communication network must ensure the availability of the data, authenticity and confidentiality [16].

It is not easy for the WSN designer to select the efficient solutions that should be considered in the design of application-specific WSN architectures. Although is difficult to propose a general approach for developing WSNs for industrial applications, this work considers some of the more common resource constraints of a WSN in this field related to: limited power consumption, robustness, processing capacity and storage.

The main contribution of the paper is related to the development of an efficient architecture in terms of power consumption and reliability, which involves both hardware and software design strategies. Moreover, the deployed WSN is easily scalable up. Authors have experimentally demonstrated the proposal feasibility and the energy efficiency of their approach.

The research covers several areas going from physical layer optimization to network layer solutions. IEEE 802.15.4 standard was chosen for communications because in the last few years it has become a reference for wireless applications with low data rates and high energy efficiency [17–23]. In fact, most of the WSN technologies are established by adapting IEEE 802.15.4 standard by using different frequencies and protocols. This standard, specifically designed for LR-WPAN networks, is a low-complexity protocol that offers low cost and energy features, as well as high versatility for the development of different topologies. We have devoted efforts to develop a specific stack since industrial applications require fast and timely responsiveness. The network architecture is particularly designed to address these industrial requirements.

The proposed system can be easily scaled up equipping various sensors on the designed sensor node or even including new sensor nodes to measure other physical magnitudes or parameters of interest. The wireless transmission was used to integrate all independent sensor signals that allowed for centralization and real-time control. Central management saves labor, increases efficiency and greatly reduces costs in industrial safety.

A particular case of study has been considered in order to validate the authors' theoretical hypothesis. An industrial WSN for detecting and monitoring different types of gases has been implemented. The article gives details about the design of the modules and highlights the implementation of the theoretical approaches.

The paper is organized as follows: in the next section the components of the proposed architecture are presented and the general way of working is described. Section 3 is devoted to explaining the main features of the net by giving details about the communication protocol and the role played by the different elements of the network. Finally, Section 4 illustrated the experimental validation of the elements implemented for the case of study. Finally, conclusions are given in Section 5.

2. System Description and Operation

A WSN consists of tiny sensor devices spatially distributed forming a low level self-organized network, which send the data gathered through wireless links to sink nodes. Then, the sink nodes process or analyze the data locally or are connected to other networks, realizing real-time monitoring and control. Usually, the intrinsic characteristics of the wireless sensor network present new challenges in hardware design, communication protocols and application design. This requires modifying legacy protocols for conventional wireless networks or designing new effective communication protocols and algorithms.

Figure 1 shows the three different subsystems of the proposed industrial WSN architecture: The sensor devices or motes, the base station and the information management system.
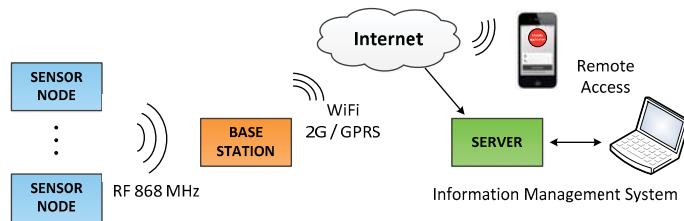
Figure 1. Architecture of the designed wireless sensor network.

Despite of being a common structure, each of the elements depicted in this figure requires a particular consideration for achieving the efficiency demanded by industrial applications. This section is devoted to describe the design of these elements in terms of: reliability, power consumption, data availability, authenticity and confidentiality, friendly and easy remote configuration. Theoretical approaches are first described and later turned into a real development by considering the proposed case of study: an industrial WSN for monitoring and detecting different types of gases.

2.1. Sensor Node

Several questions have been addressed in the design of the sensor nodes, concerning energy efficiency, reliability and latency reduction.

On the one hand, sensor nodes are designed to built a flexible solution and avoid complex cabling. Moreover, defining an energy efficient operation mode is essential for the target application. For this purpose, sensor nodes gather the data periodically but they alternate between working cycles (when the node is awake) and low power consumption cycles (when the node is slept). This issue allows the sensor nodes to keep the energy consumption to a minimum, extending their functionality for a reasonable period of time.

This operation mode together with an energy efficient protocol (described in detail in Section 3) has been proposed to save energy and guarantee that the data is reliably delivered to the sink node. To this aim, sensor nodes are deployed with a cluster-tree topology. Both the employed network topology and the protocol allow the designed system to increase its scalability, containing a large number of sensor nodes. In this way, low cost and small-sized sensor nodes have been implemented. Finally, latency has also been taken into account for practical operation. Latency is an important factor for the system reliability such as in the case of emergency response, and accuracy of data reporting in case of high frequency periodical data updates.

All these ideas have been reflected in the design of the sensor implemented in the case of study. Figure 2 shows the board of the developed sensing node. The designed hardware consists of an analog gas sensor, a data processing module (microcontroller), a wireless communication module (wireless transceiver) and an energy module. The gas sensor comprises of all circuitry necessary for the detection of butane, methane and propane. The detection system is completed including a carbon monoxide (CO) sensor. Both the gas sensor and the CO sensor are placed on the bottom of the board.

The analog gas sensor TGS6810 from Figaro (Arlington Heights, IL, USA) has been chosen. This is a compact sensor whose stability, quick response and linear output makes it suitable for detecting combustible gases. The signal-conditioning circuit is basically a Wheatstone bridge, whose differential output voltage is converted to a single-ended signal and amplified by a non-inverting amplifier stage. The high sensitivity of this sensor can detect very small gas leaks in very early stages, avoiding accidents due to explosions, fires or inhalations. Regarding the CO sensor, the TGS5342 also from Figaro (Arlington Heights, IL, USA) has been used. This sensor generates a small current, linear to CO concentration, which is converted into an output voltage by an op-amp/resistor combination. The output signal of the analog sensors is converted to a binary value using the analog to digital converter of the microcontroller.

Figure 2. Gas sensing node.

2.2. Base Station

The core of the proposed architecture is the base station which centralizes all system signals. This node is also known as gateway or sink node. It coordinates the network, receives the data from the sensor nodes, and processes and stores it into the information management system. For this purpose, the base station has to include several elements responsible for performing these main tasks. First of all, an efficient wireless circuit for implementing communication with the sensor nodes is needed. Secondly, a redundant communications system that supports outer communications has to be included. The base station should also provide a simple human interface to allow basic operation. Finally, although the base station is usually plugged to a power supply, it is also subject to energy constraint and the possibility of being powered by a battery has to be taken into account.

In this article a general purpose design for a base station is proposed. It can be applied either for the considered case of study or for another industrial application. In the following the proposed design is described in detail.

The hardware of the base station is divided into two printed circuit boards. The primary board is the core of the base station, and includes the following elements: a RF (radio frequency) communication circuit compatible with the one used by the sensor nodes, a power supply system (internal battery, battery charger), an outer communication system, control circuits and data processing (MCU). Figure 3 shows details of the hardware implementation of this main board.

The base station is structured in a multi-processor design with two independent control circuits as shown in Figure 4: One based on the 16-bit PIC24FJ128GA306 microcontroller and another based on the 32-bit PIC32MX795F512L microcontroller, both from Microchip (Chandler, AZ, USA). These microcontrollers meet all needed specifications, and the main tasks assigned to each are as follows: the 16-bit microcontroller handles the control of the capacitive touch keyboard and its Light Emitting Diode (LEDs), Serial Peripheral Interface (SPI) communications with the RF 868 MHz circuit for the wireless data transmission with the sensor nodes, and communications with the 32-bit microcontroller in order to synchronize with the external world all events occurring in wireless peripherals attached to the system. The 32-bit microcontroller performs two basic tasks: managing the main firmware, and controlling the outer communication system. These communications can be supported by three different technologies (Ethernet, Wi-Fi, and 2G/GPRS modules). By means of these mechanisms the base station is able to retransmit the information gathered by the sensor nodes to the external world. Thus, the versatility of the base station ensures a feasible communication with the server and Internet through wired and wireless connections. The communication is further assured

with the 2G/GPRS module for losses of Ethernet and Wi-Fi connections, making the system more robust and even avoiding sabotages.
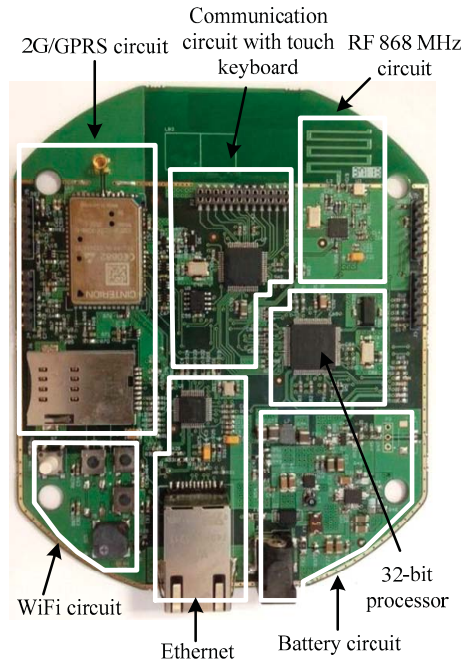

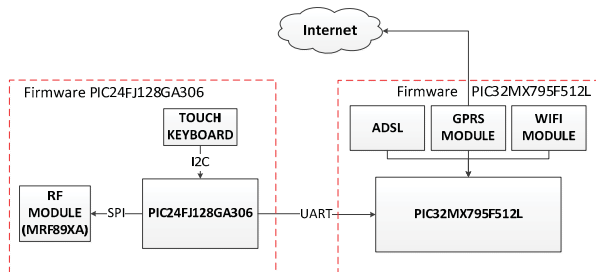**Figure 3.** Image of the main board of the base station.

Figure 4. Block diagram of the firmware structure of the base station.

The short-range radio module exchanges data via wireless at 868 MHz following a specific stack, which has been developed conciliating the most appropriate features of IEEE 802.15.4 and MiWi for the target application. Additionally, encryption algorithms have been implemented to improve the security of the RF communications, as will be discussed later. An electromagnetic shield to avoid interference problems between the RF circuit and adjacent circuits has also been included.

The radio module is based on the transceiver MRF89XA from Microchip (Chandler, AZ, USA), chosen due to the very low power consumption and for being particularly suitable for the sensor nodes. A low cost solution has been chosen designing a PCB trace antenna, avoiding the use of conventional antennas or commercial elements that would add cost to the device. Figure 5 shows the designed antenna and the scheme of the RF module.
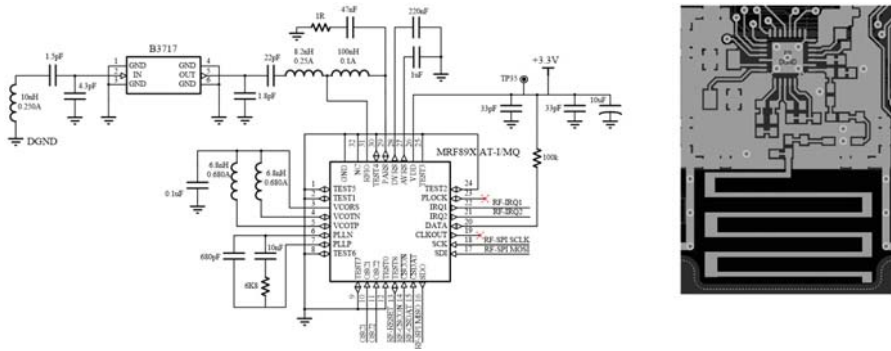
Figure 5. Designed antenna and scheme of the radio frequency (RF) module.

Simulation software for high performance antennas has been used. Results shown in Figure 6 guarantee the proper operation.
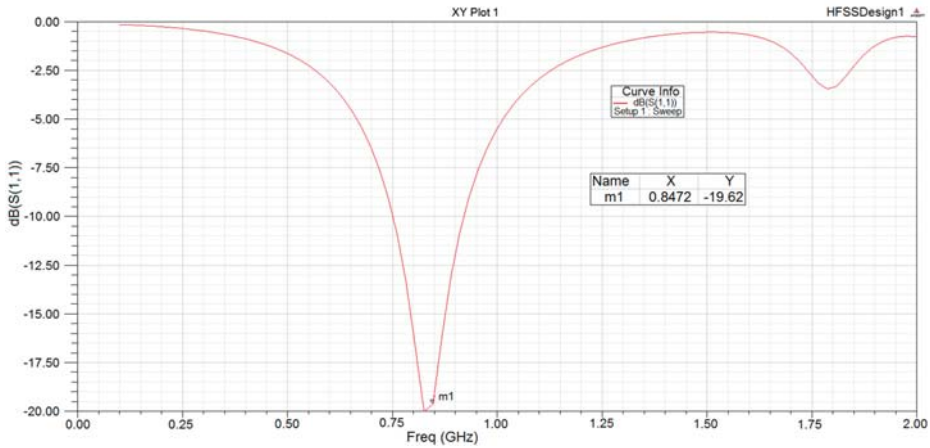

Figure 6. Simulated performance of the antenna.

The base station is plugged to a power supply but it can also be supplied by a battery. Additional circuitry has been included based on the integrated linear charger bq24070 from Texas Instruments (Dallas, TX, USA) as a system power-path management device. This combined approach allows the base station to be simultaneously powered while the battery is charging, and avoids charge-discharge cycles, greatly extending the battery life. The base station is able to continue working properly for an additional period of 10 hours, if an electrical power outage occurs.

The secondary board of the base station includes a capacitive touch keypad with associated electronics and LEDs as shown in Figure 7. The design of this board and the integration in the base station has been a challenge, as the capacitive sensing technology to detect one touch (only functional with a human finger), with the proper choice of the materials which permit the transmission of signals from the user's finger to the sensor of the board (copper pad), had to be merged. For the design of this board, the capacitive keyboard inputs that are available in the 16-bit microcontroller were initially used. After stringent preliminary tests, random pulses occurred due to external noise and the sensitivity was not entirely correct. Thus, the drivers CAP1166 with capacitive sensors especially suitable for

these applications were used, providing satisfactory results after repeating the tests. The inclusion of the touch keyboard on the base station increases the versatility of the system. Although the system is controlled remotely, the capacitive touch keypad also allows the user to manually enter codes to enable or disable different parts of the system.
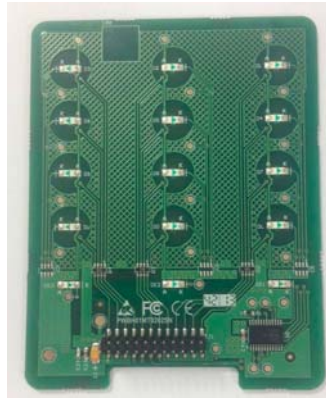

Figure 7. Touch keyboard of the base station.

2.3. Information Management System

Several system status data need to be defined and gathered together with the sensor data. Moreover, it is important to remotely adjust system configurations and update and upgrade software programs. In addition, remote monitoring of system status is useful for system development, debugging, and maintenance purposes.

The information management system includes a web application and mobile applications for iOS and Android, which allow the control of all system parameters, and monitoring alarms and events. The web application allows configuring the system from a remote computer via Internet by authorized users, accessible through the outer communication system. The mobile application also provides alerts in real-time.

The complete monitoring and control for industry applications involves the measurement of several physical magnitudes to improve the quality of the processes. The designed wireless system is scalable, and thus the number of quantities to be measured is easily expandable.

3. Network Management

3.1. Preliminary Considerations

Initially we started from a known open source protocol proposed by the transceiver's manufacturer, called MiWi. This is a low cost solution for applications where low power consumption, few computational resources and low data rates are required. It is based on the IEEE 802.15.4 standard, specifically designed for LR-WPAN networks, which offers high versatility for the development of different topologies.

The preliminary tests performed to MiWi concluded that it was not sufficiently robust and stable for the target application. Specifically, it presents large latencies, which were also variable and never deterministic, it even sometimes blocked part of the hardware that needed immediate attention of the processor because MiWi makes intensive use of busy-wait cycles. Duty cycle is one of the most commonly used approaches to reduce the energy consumption during waiting time. Furthermore, it is one of the major sources of energy waste. Therefore, for the firmware of RF communications, we preferred to develop a specific stack. The main features of the developed stack are the following:

- SPI controller with an interrupt mechanism to minimize latency and the busy-wait cycle of the microcontroller
- Driver for the Microchip MRF89XA transceiver. The controller is ready to work in background, so it does not use busy-wait cycles, and therefore there is no latency
- A medium access (MAC) layer, which is responsible for the communication between nodes by routing through a unique fixed address, read from the Electrically Erasable Programmable Read Only Memory (EEPROM). This layer is also designed to minimize the busy-wait cycles and for most of the process is done in background, reducing latency as much as possible. For the design of this layer aspects of the IEEE 802.15.4, such as timeouts or frame formats were considered
- A logical link layer (LLC). This layer handles the logical addressing of the nodes, the allocation of roles within the network (PAN coordinator, coordinator or end device), assignment of the logical address (which is not fixed), route management, network discovery, traffic control and management of consumption modes. The design of this layer is also designed to minimize busy-wait times and latency
- AES128 encryption layer is the responsible for encrypting the messages that will be sent to the other devices. This layer improves the system safety
- Presentation layer. This layer (above the LLC layer) performs the functions of an interface between the stack and the node. It is responsible of configuring the stack, and of coordinating its operation
- Driver for EEPROM. It is responsible of EEPROM access without interfering with the radio module. This prevents a high coupling between them, since both may operate asynchronously, sharing the same SPI bus.

3.2. Protocol Description

The designed stack was performed in layers so that they can be used independently and disengaging the different levels of functionality, allowing an enhanced code, portability, adaptation to new hardware and better debugging and testing. The same addressing scheme proposed by MiWi and supported for cluster tree network topologies (provided by the LLC layer) have been used in the new approach.

The main aim of the design is to minimize the latency and the processor usage almost without busy-wait cycles. Thus, the entire design is based on the use of interruptions by the SPI controller. All functions are implemented as state machines in which each state is triggered by an interruption when the previous operation ends.

The communication of asynchronous events, such as the reception of new messages or network outage is performed through a set of callbacks added to the APIs of each layer. This structure (although the development has been somewhat more complex than in the case of stacks as in MiWi) has allowed to reduce the maximum processor usage, and the stack scarcely influences the other components of the nodes. The power consumption is also reduced because the CPU is free a longer period of time.

MiWi, similar to the specifications defined by IEEE 802.15.4, supports three network frameworks: Star, Cluster Tree and Mesh; and defines two types of devices: full-function devices (FFD) and reduced-function devices (RFD) as shown in Table 1. The roles of the nodes can be divided into PAN Coordinator, Coordinator and End Device as shown in Table 2. Sensor nodes could act as a full function device (FFD) or reduce function device (RFD). In FFD, the nodes never sleep, as they need to relay the information from the RFD to the sink node, while a RFD will only wake after a certain period of time to: gather the sensor data, forward it to a FFD node and go to the sleep mode.

Table 1. Class of devices by means of IEEE 802.15.4.

Device Class	Offered Services	Power Path	Low Power Configuration
Full Function Device (FFD)	All	Mains	Always ON
Reduced Function Device (RFD)	Limited	Batteries	Always in SLEEP

Table 2. Devices role (IEEE 802.15.4).

Device Role	Class of Device (IEEE)	Function in the Network
PAN Coordinator	FFD	Only one by network. Create network, assign addresses, and hold the connections table
Coordinator	FFD	Optional. Make bigger the range of the network and make possible to add more devices to the network. Control and monitoring functions
End Device	FFD o RFD	Control and monitoring functions

The sensor nodes must use low consumption operating modes in order to achieve good energy efficiency. For this purpose, a cluster-tree network has been deployed so that the sensor nodes (acting as End Devices) can be set as reduced-function devices (RFD) with the ability to enter in the sleeping mode in which the consumption can be considered negligible. Figure 8 shows the topology of a cluster-tree network. A PAN Coordinator node is responsible for managing the network. This PAN Coordinator assigns the addresses to the END Devices and to the other Coordinators. A coordinator node also assigns addresses to the END Device nodes located inside its transmission range. If a new node is switched on, it starts on the awake mode and takes contact with the FFD node inside its transmission range (either a Coordinator or the Pan Coordinator). Once the FFD node assigns the address, the END Device starts its cycle switching from awake to sleep mode. This mechanism facilitates the system scalability.
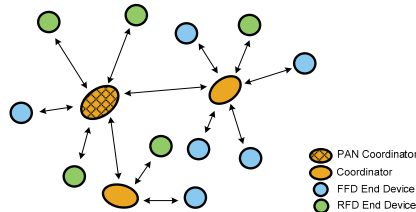
Figure 8. Cluster-tree network topology.

The medium access protocol that has been chosen corresponds to the IEEE 802.15.4 specification. The designed layer is located above the MAC layer and must specify the following mechanisms: (1) Assigning addresses to new nodes to be included in the network; (2) Determining the necessary elements to create a network topology based on a Cluster Tree topology; (3) Message routing inside the network, if several jumps are required to reach the destination; (4) Finding out when a node loses the connection with its coordinator; (5) Security configuration; (6) Configuring parameters for energy efficiency, such as transmission power, sensitivity and sleep states of the transceiver; (7) Preventing the loss of messages even if the target node is in sleep mode.

Security is critical when designing a wireless communication system. Intrusions may result in harmful or even disastrous situations in industrial plants. The designed network includes security mechanisms and message encryption to maintain data integrity and prevent interception of the transmitted data between the nodes of the network.

Regarding the flow control, an un-slotted CSMA/CA mechanism without beacons is used when a node wants access to the medium for message transmission, by means defined in IEEE 802.15.4.

3.3. Energy Conservation and Data Reliably Deliberation

The energy optimization can be made at different levels. As sensor nodes require efficient energy, optimized hardware for low power operation requires selectable power states (awake or sleep). As in sleep mode the transceiver is disabled, it is necessary to establish some sort of protocol to avoid the loss of messages.

As discussed above, only END Devices can enter sleep mode. The mechanism that prevents the loss of messages when END Devices are in sleep mode consists in making Coordinators store messages addressed to their END Devices. Once a node is awake it must explicitly request data to the Coordinators using Data Request commands defined by the IEEE 802.15.4 standard. Thus, we will also have a flow control for the case that an END Device takes too long to process a message before it can receive the following. The mechanism of energy management, which in principle only makes sense for the case of an END Device, allows controlling the way that they ask data to the Coordinators.

Obviously, the sink node requires high processing power to coordinate the whole network activity, thus, it requires the highest energy consumption amongst the devices.

A detailed power budget analysis of the transceiver was necessary to determine the capacity of the system. The power consumption is determined by measuring the current draw and the time spent in each of its operating modes. The tests were carried out for four functional states: Data transmission, acquisition, listening and sleep mode, for a transmission power of 13 dBm. The measured power consumption was equal to 12.8 mA for the transmission state, 6.52 mA for acquisition, 2.99 mA for listening and 17 µA for sleep state. The mean elapsed time used in receiving and answering the pending messages, measured between the awakening of the node and the return to sleep mode, was 140 µs (wakeup time). The mean current consumption for the transceiver was limited to 150 µA over a certain period of time. Thus, the following expression must be met:

$$F_{tw} \cdot I_w + F_{ts} \cdot I_s \leq I_{max} \qquad (1)$$

where F_{tw} is the fraction of time that the transceiver is turned on, and I_w is the current consumption of the transceiver in that state; whilst F_{ts} is the fraction of time that the transceiver is in sleep mode, and I_s is the current consumption in this same state. As $F_{ts} = (1\ F_{tw})$, the above expression can be rewritten, yielding,

$$F_{tw} \leq \frac{I_{max} - I_s}{I_w - I_s} \qquad (2)$$

Substituting the measured current values, $F_{tw} \leq$ (150 µA − 17 µA)/(12.8 mA − 17 µA) $\cong$ 0.0104. Considering the power consumption per hour, the time that the transceiver can remain in active mode is T_w = 0.0104 × 3600 s = 37.45 s. Thus, the transceiver allows accomplishing a very high number of requests per hour (37.45 s/140 µs = 2.67 × 10⁵). Therefore, at maximum power, the consumption of the transceiver allows a large number of requests to be made without breaking the time requirements.

Once the hardware was validated and correct operability was assured, the next step was to review the software incorporating low-consumption modes in all the motes.

Figure 9 shows the final implementation of the base station and the gas sensor, and the packing designed to protect them. The final dimensions of the sensor packing are 70 mm in diameter with a height of 30 mm. For the base station they are 140 mm in diameter and 35 mm in height.

Figure 9. Image of the base station and gas sensor, and packing design.

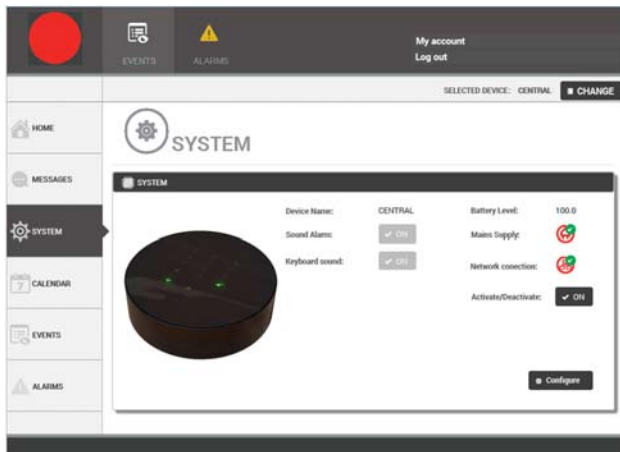**4. Experimental Validation: Monitoring and Control Application**

Considering the case of study, web and mobile applications that implement a graphical user interface have been developed. By means of this software, real time monitoring of physical magnitudes and battery level of all active nodes can be done. Furthermore, all information measured by the nodes can be preserved for further processing. This information allows the user to act on the system at any time.

The web and the mobile applications allow the user to configure all devices of the wireless system, as well as monitoring and receiving alarm warnings. Figure 10 shows the main views of the remote monitoring and control application, and the main features are as follows:

1. Registration and access to the control panel. Users may register through the web portal, which includes their user data and credentials. Once inside the application, the different system devices can be configured and monitored

2. The battery voltage level is closely monitored as a part of the remote system status monitoring service. Near-real-time monitoring of such a system status data is important in determining battery efficiency and early detection of severe battery degradation to prevent system failure and the loss of data

3. Methods of notification about alarms and events provided by the system (Short Message Service (SMS), Email and mobile notifications) can be changed as well as selecting which user should receive such notifications

4. The configuration of the base station and the peripherals that comprise the system allows the user to activate them, disable them, or make changes in the configuration of each device, such as changing detection thresholds, detection periods, etc.

5. Event monitoring. The various events generated in the system can be filtered and listed. These events refer to changes in the configuration of devices and accesses to the system

6. Alarm monitoring. The various alarms generated in the system can be filtered and listed.

The sensors proved to be robust and have very high performance. The gas sensor meets the UNE-EN 50194 standard, and enables the detection of toxic and explosive gases, such as, butane, propane, methane, town gas, natural gas and others. The gas detector is calibrated to detect the presence of gases at a level of concentration of 10% L.E.L. (Lower Explosion Limit), corresponding to: 5.000 p.p.m of methane, 2.000 p.p.m. of propane and 1.800 p.p.m. of butane following the mentioned standard.

Figure 11 shows three cases of measured values for butane, methane and propane. Tests were performed in a room of dimensions 3 m × 5 m and 2.5 m of height. The gases were introduced at three different pressures A, B and C (pressure A is the highest and pressure C is the lowest). Then, the volume of gas in the room was measured for 80 seconds. Note that for the case of higher pressure (A) the gas sensor responds faster and the measured gas concentration is also higher. In the opposite situation is the case of lower pressure (C). When any of the above threshold levels (dashed lines) is exceeded, the detector activates an alarm indicator light for 5 seconds to confirm the detection and prevent false alarms. When this time is up, if the level of detection is maintained, the alarm for the mobile application is activated. When the detector, in alarm state, detects a gas concentration level of less than 10% L.E.L for a certain time, it deactivates the alarm indicator light and the alarm output for the mobile application.

(**a**)

(**b**)

(**c**)

Figure 10. Main views of the monitoring application.

(**a**)

(**b**)

(**c**)
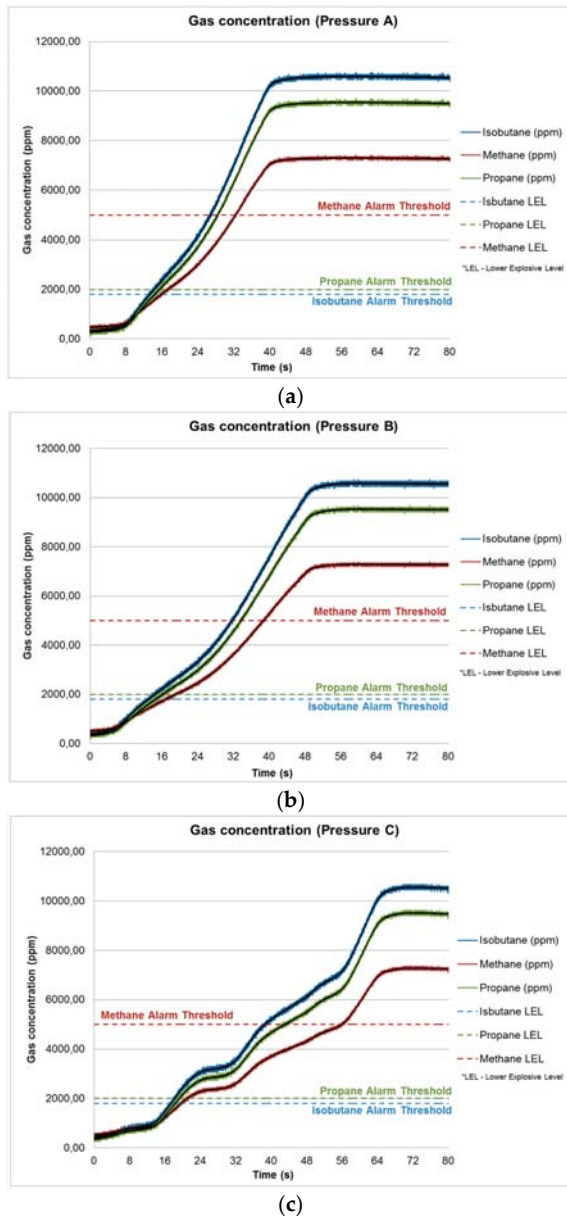
Figure 11. Time response of gas detector measurements in three cases. (**a**) Higher pressure, (**b**) Middle pressure and (**c**) Lower pressure.

5. Conclusions

This paper presents an original approach regarding the design of an efficient WSN for industrial monitoring and control in terms of power consumption, data availability, authenticity, confidentiality, and friendly and easy remote configuration. The proposal is aimed to accomplish industrial

requirements, and thus, involves hardware and software design strategies for achieving robust nodes and reliability and security in the communications. A real WSN designed to measure industrial parameters has been implemented, so that experimental validations can be performed.

The implemented wireless system includes the development of a low cost gas sensor node with optimized size, a base station and a graphical user interface for remote access to industrial conditions and real-time control and monitoring. The system has been designed to be easily scaled-up to measure other parameters of interest. All aspects of the monitoring system such as sensing activity, transmission from sensor nodes to base station, remote transmission from the base station to the control center or information management system, data storage, and visualization, have also been designed and implemented. The system requires low maintenance and includes energy-budget management. The flexibility of the system allows being used to realize remote monitoring in coal mine, buildings, etc.

Author Contributions: Juan Aponte-Luis, Javier Alcina-Espigado and Manuel Sánchez-Raya designed and developed the system. Juan Antonio Gómez-Galán, Fernando Gómez-Bravo and Pedro Miguel Teixido-Rovira proposed the ideas, performed the experiments and prepared the manuscript. Juan Antonio Gómez-Galán also made the supervision and direction of the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rawat, P.; Deep Singh, K.; Chaouchi, H.; Bonnin, J.M. Wireless sensor networks: A survey on recent developments and potential synergies. *J. Supercomput.* **2014**, *68*, 1–48. [CrossRef]
2. Jawad, H.M.; Nordin, R.; Gharghan, S.K.; Jawad, A.M.; Ismail, M. Energy-Efficient wireless sensor networks for precision agriculture: A review. *Sensors* **2017**, *17*, 1781. [CrossRef] [PubMed]
3. Yu, X.; Wu, P.; Han, W.; Zhang, Z. A survey on wireless sensor network infrastructure for agriculture. *Comput. Stand. Interfaces* **2013**, *35*, 59–64. [CrossRef]
4. Du, Y.-C.; Lee, Y.-Y.; Lu, Y.-Y.; Lin, C.-H.; Wu, M.-J.; Chen, C.-L.; Chen, T. Development of a telecare system based on ZigBee mesh network for monitoring blood pressure of patients with hemodialysis in health care centers. *J. Med. Syst.* **2011**, *35*, 877–883. [CrossRef] [PubMed]
5. Medina, J.; Sánchez-Rodríguez, T.; Gómez-Galán, J.A.; Delgado, A.; Gómez, F.; Jiménez, R. A wireless sensor system for real-time monitoring and fault detection of motor arrays. *Sensors* **2017**, *17*, 677–684.
6. Alippi, C.; Camplani, R.; Galperti, C.; Roveri, M. A robust, adaptive, solar-powered WSN framework for aquatic environmental monitoring. *IEEE Sens. J.* **2011**, *11*, 45–55. [CrossRef]
7. Gutiérrez, J.; Villa-Medina, J.F.; Nieto-Garibay, A.; Porta-Gándara, M.A. Automated irrigation system using a wireless sensor network and GPRS module. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 166–176. [CrossRef]
8. Capella, J.V.; Bonastre, A.; Ors, R.; Peris, M. In line river monitoring of nitrate concentration by means of a wireless sensor network with energy harvesting. *Sens. Actuators B* **2013**, *177*, 419–427. [CrossRef]
9. Anisi, M.H.; Abdullah, A.H. Efficient data reporting in intelligent transportation systems. *Networks Spat. Econ.* **2016**, *16*, 623–642. [CrossRef]
10. Suryadevara, N.K.; Mukhopadhyay, S.C.; Tebje Kelly, S.D.; Singh Gill, S.P. WSN-based smart sensors and actuator for power management in intelligent buildings. *IEEE/ASME Trans. Mecatronics* **2015**, *20*, 564–571. [CrossRef]
11. Rault, T.; Bouabdallah, A.; Challal, Y. Energy efficiency in wireless sensor networks: A top-down survey. *Comput. Networks* **2014**, *67*, 104–122. [CrossRef]
12. Aparicio, J.; Echevarria, J.J.; Legarda, J.A. Software defined networking approach to improve the energy efficiency of mobile wireless sensor networks. *IEEE Trans. Inter. Inf. Syst.* **2017**, *11*, 2848–2869.
13. Xu, L.D.; He, W.; Li, S. Internet of Things in Industries: A Survey. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2233–2243. [CrossRef]
14. Ege, I.; Gökhan-Sensoy, M.; Kalender, O.; Nazlıbilek, S.; Citak, H. A new wireless asynchronous data communications module for industrial applications. *Measurement* **2013**, *46*, 2672–2681. [CrossRef]
15. Yuan, Y.; Xu, Q.; Guan, X.; Liu, Z. Industrial high-speed wireless synchronous data acquisition system with real-time data compression. *Measurement* **2013**, *46*, 3482–3487. [CrossRef]

16. Arthi, K.; Vijayalakshmi, A.; Vanaja-Ranjan, P. Critical event based multichannel process control monitoring using WSN for industrial applications. *Int. Conf. Des. Manuf. IConDM* **2013**, *64*, 142–148. [CrossRef]

17. Kumar, A.; Hancke, G.P. Energy efficient environment monitoring system based on the IEEE 802.15.4 standard for low cost requirements. *IEEE Sens. J.* **2014**, *14*, 2557–2566. [CrossRef]

18. Toscano, E.; Lo Bello, L. Multichannel superframe scheduling for IEEE 802.15.4 industrial wireless sensor networks. *IEEE Trans. Ind. Inform.* **2012**, *8*, 337–350. [CrossRef]

19. Li, X.; Bleakley, C.J.; Bober, W. Enhanced beacon-enabled mode for improved IEEE 802.15.4 low data rate performance. *Wirel. Netw.* **2012**, *18*, 59–74. [CrossRef]

20. Kone, C.T.; Hafid, A.; Boushaba, M. Performance management of IEEE 802.15.4 wireless sensor network for precision agriculture. *IEEE Sens. J.* **2015**, *15*, 5734–5747. [CrossRef]

21. Du, W.; Navarro, D.; Mieyeville, F. Performance evaluation of IEEE 802.15.4 sensor networks in industrial applications. *Inter. J. Commun. Syst.* **2015**, *28*, 1657–1674. [CrossRef]

22. Torfs, T.; Sterken, T.; Brebels, S.; Santana, J.; Hoven, R.; Spiering, V.; Bertsch, N.; Zonta, D. Low power wireless sensor network for building monitoring. *IEEE Sens. J.* **2013**, *13*, 909–915. [CrossRef]

23. Lopez Riquelme, J.A.; Soto, F.; Suardíaz, J.; Sánchez, P.; Iborra, A.; Vera, J.A. Wireless Sensor Networks for precision horticulture in Southern Spain. *Comput. Electr. Agric.* **2009**, *68*, 25–35. [CrossRef]

Compact Embedded Wireless Sensor-Based Monitoring of Concrete Curing

Joaquín Cabezas [1], Trinidad Sánchez-Rodríguez [2,*], Juan Antonio Gómez-Galán [2], Héctor Cifuentes [3] and Ramón González Carvajal [1]

[1] Department of Electronic Engineering, University of Seville, Camino de los Descubrimientos, s/n, E-41092 Seville, Spain; joaquin.cabezas@adevice.es (J.C.); carvajal@us.es (R.G.C.)

[2] Department of Electronic Engineering, Computers and Automatic, University of Huelva, Ctra Huelva-La Rábida, s/n, 21819 Huelva, Spain; jgalan@diesia.uhu.es

[3] Department of Continuum Mechanics and Structural Analysis, University of Seville, Camino de los Descubrimientos, s/n, E-41092 Seville, Spain; bulte@us.es

* Correspondence: trinidad.sanchez@diesia.uhu.es; Tel.: +34-959-217-662; Fax: +34-959-217-348

Abstract: This work presents the design, construction and testing of a new embedded sensor system for monitoring concrete curing. A specific mote has been implemented to withstand the aggressive environment without affecting the measured variables. The system also includes a real-time monitoring application operating from a remote computer placed in a central location. The testing was done in two phases: the first in the laboratory, to validate the functional requirements of the developed devices; and the second on civil works to evaluate the functional features of the devices, such as range, robustness and flexibility. The devices were successfully implemented resulting in a low cost, highly reliable, compact and non-destructive solution.

Keywords: embedded sensor; concrete curing; monitoring; wireless networks; IEEE 802.15.4

1. Introduction

Monitoring of concrete properties introduces many challenges in the design of non-destructive sensing systems. The miniaturization of the sensors taking advantage of technological advances is crucial for their placement inside the concrete. In recent years the growing importance of the so-called structural health monitoring is demanding intelligent systems based on sensors embedded into concrete, capable of controlling and monitoring the curing process mainly by measuring the temperature, or both temperature and humidity [1,2]. Other traditional approaches for structural health monitoring are based on accelerometers, strain gauges and displacement transducers [3–5]. However, the amount of works related to sensors embedded into concrete is limited by the obvious unfavorable conditions under which they have to operate.

Digital temperature and/or humidity sensors are widely known and their evolution continues to progress. These small systems incorporate the sensor, conditioning electronics, analog to digital conversion and digital processing necessary for connecting to a digital bus in one integrated circuit of reduced dimensions. Digital sensors provide advantages in terms of size, cost, reduced wiring, high noise immunity and direct connection to digital buses, allowing for the connection of large number of sensors without auxiliary electronics. However, the available digital temperature and/or moisture sensors cannot be used directly into the concrete, since their packages are not able to withstand the adverse conditions, especially in terms of sealing and compression.

On the other hand, wired sensing increases the cost of the monitoring system [6]. Moreover, the added complexity would not make such a system feasible to be used for monitoring

purposes on a construction site. A wireless approach would help to enable its use for on-site applications [7–11].

This paper describes our experience with the introduction and development of an experimental sensor embedded into concrete. The sensor exhibits temperature measurement capability. According to most of the available thermo-chemical models for concrete [12–15], by monitoring the evolution of the internal temperature of the concrete bulk, the effect of other factors, such as humidity, water to cement ratio, composition of cement, etc., on concrete curing are also considered. A specific package was designed to protect the sensor from the aggressive environment within the concrete. This sensor also incorporates significant additional features in terms of connectivity, measuring, remote programming and operation. To avoid affecting the concrete structure the embedded sensor occupies as small a space as possible. Thus, the sensor only comprises very few and small components.

Concerning the integration of a wireless embedded sensor into an instrumentation system (generally into any measuring device), issues such as ease of connection or signal degradation due to the transmission process must be taken into account. The proposed sensor presents a number of features that are particularly useful for this application and that can be extended for structural health: (1) the energy consumption of the sensor node is optimized in such way that it can be powered by batteries. In this way, the energy supply is ensured throughout the lifetime of the node; (2) the simplicity of the proposed solution results in a low cost device, further guaranteeing robustness, accuracy and high resolution; (3) it can be connected directly to a standard instrumentation system. Thus, the developed embedded sensor includes an I2C full-duplex serial digital port (i.e., directly-transmitted and received signals are digital); (4) the sensor is equipped with a specifically designed virtual instrument, which allows acquiring and storing measurements in different formats, configuring the parameters of the sensor, reprogramming and controlling the sensor via Internet from a PC; (5) connection features in the developed sensor are significant, both in terms of quality (ease, robustness, immunity to noise, etc.) and the cost-savings involved in not having to transmit and condition analog signals outside the device; (6) a low cost and low power consumption solution with optimized size has been developed.

The paper is organized as follows: Section 2 describes the main properties of the wireless sensor network in terms of type of standard, operation frequency, network topology, and transmission power. Section 3 deals with the hardware and software of the devices developed for the implementation of the system. Section 4 is devoted to experimental calibration procedures, testing and measurement. Finally, some conclusions are drawn in Section 5.

2. Considerations on Used Wireless Communication

Advances in low power wireless technologies made possible the immersion of a wireless node completely surrounded by the building material. Several wireless technologies, such as Bluetooth Low Energy, IEEE 802.15.4, ZigBee, Ultra-Wide Band (UWB) and Radio Frequency IDentification (RFID) [16–18] could be used to implement a wireless sensor network (WSN) within concrete, overcoming the problems related to the propagation of the electromagnetic waves through the structure and the node autonomy.

Wireless sensor networks are composed of smart low rate devices or motes, which are generally battery-powered. The number of nodes can vary from a few to thousands. These networks allow for a dynamic network configuration, which is a feature particularly important for concrete curing, since a failure in one node is compensated by rerouting through other nodes.

IEEE 802.15.4 was chosen for communications because in the last few years it has become a reference for wireless applications with low data rates and high energy efficiency. This standard, specifically designed for LR-WPAN networks, is a low-complexity protocol that offers low cost and energy features, as well as high versatility for the development of different topologies.

Network Topology

The network consists of a number of wireless nodes embedded into concrete, and needs to merge particular and commercial solutions. Regarding the radio propagation, the high signal attenuation and the antenna detuning must be considered. These issues can result in very high error rates, so the information exchange mechanism must adapt to it. In addition, the optimization of the energy consumption is degraded since more errors in the data transmission will require more retransmissions, and thus higher energy consumption. Finally, the transmission power must be high to avoid problems of attenuation.

It is common practice to place steel reinforcement to provide strength against tensile stresses and avoid excessive cracking of concrete. Moreover, the wet concrete is shaped in a mould comprised of formwork, which is most commonly steel or wood, being more problematic the presence of steel for the signal transmission. Therefore, if we want to monitor the concrete curing since when it is first poured, a careful design of the wireless network topology must be accomplished since the presence of steel may degrade the transmission of data. This fact imposes a wireless network topology of linear type, which involves an asymmetry in the network. Thus, the nodes that are further away have to send data through a long list of intermediate nodes. In turn, these intermediate nodes will waste part of their energy in the tasks of rerouting data. Figure 1 illustrates the network topology and the problems related to the data traffic.
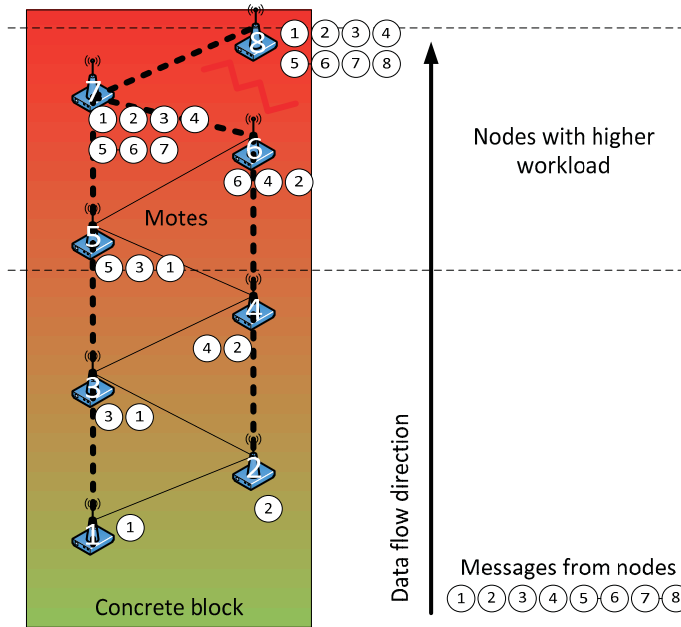
Figure 1. Network topology and associated problems.

Considering the former problems, routing and synchronization algorithms have been designed to minimize the following characteristics: energy consumption, error rate, complexity and the probability of loss of a node. The energy consumption is decreased by synchronizing messages, whereas the rest of the time the node is idle. The use of very short messages allows balancing the workload between nodes, improving the error rate. Moreover, when a node performs a data transmission to all nodes within its coverage range (broadcast transmission), reliability is quite high since several nodes will receive the data message, so the loss of information is rather unlikely. The transmitting node waits

for an acknowledgment message that validates its transmission. Then, it enters in low power state, reducing the time it is active. Concerning the complexity, the network is previously configured so that each node is assigned a task at the time when the network is created. The initial synchronization of the clocks is carried out in an outstation. Finally, to decrease the probability of losing a node an external watchdog is used, and different strategies will be applied depending on the level of the battery.

3. Embedded Sensor Description

The smart wireless devices or motes basically consist of a radio module including the antenna, a processor, a power supply and sensors mounted on the mote itself or connected to it. The radio module transmits and receives data through the IEEE 802.15.4 communication standard. The processor controls all the functions of the node. The power supply consists of a battery that uses a high performance power converter. The motes buried in concrete include sensors which measure temperature. In addition, an external environmental mote has also been implemented to measure the ambient temperature, since the study of concrete curing requires the analysis of both internal and external variables. Some of the nodes in the wireless sensor network may communicate with a computer through a gateway. Taking into account the environment in the concrete, the requirements that have been considered for the design of these devices are the following: (1) low cost and low power consumption; (2) use of long-lasting energy sources and reduced size; (3) a robust radio technology; (4) use of interfaces to connect different sensor nodes; (5) a package suitable to protect the wireless node from the concrete but that does not affect the measure of temperature.

3.1. Developed Sensor Node

The mote consists of a microcontroller, a radio transceiver, a lithium-thionyl dichloride primary unit cell, an embedded antenna and a waterproof enclosure. The requirements of the microcontroller were ultra-low power consumption and low voltage. Two candidates were evaluated: a CC430F6137 (Texas Instruments, Dallas, TX, USA) which includes a CC1101 transceiver that allows several frequency bands below 1 GHz (868 MHz and 433 MHz), and a CC1125 transceiver, also from Texas Instruments, alongside with an MSP430 microcontroller (Texas Instruments, Dallas, TX, USA). The power supply consists of a 3.6 V battery connected to a TPS62730 DC/DC converter (Texas Instruments, Dallas, TX, USA) of 2.1 V and efficiency of 95%, especially suitable for wireless applications. The mote equipped with a CC1125 transceiver, shown in Figure 2, benefits from the excellent receiver sensitivity at low rate modes, reaching −129 dBm at 300 bps, and therefore was selected for the deployment. The dimensions of the PCB are 56×54 mm.

The power consumption of each mote has been estimated considering the period of time at which measurements are taken as required by the concrete curing monitoring. The battery used was a LS14250 (SAFT, Levallois-Perret, France), with 1100 mAh capacity, a hermeticity guaranteed up to 110 degrees Celsius, and a small form factor (AA type, 50 mm length). This battery lasts around three months, which is more than enough time for the concrete curing monitoring.
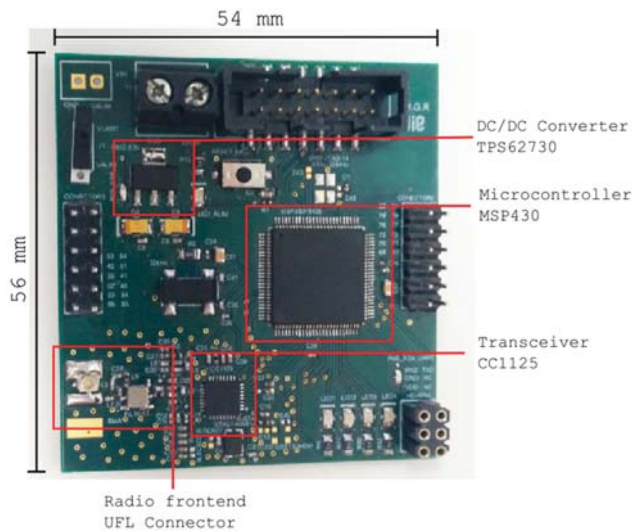
Figure 2. Wireless node (equipped with TI CC1125).

Regarding the sensor, a specific package assembly was developed initially, with details of the construction shown in Figure 3. Concrete is an extremely aggressive material during early stages, reaching a pH of up to 13. As a protective element for the sensor, and at the same time acting as a temperature and moisture diffuser, an acetal cover has been designed and manufactured. Acetal has been used because it is a thermoplastic with characteristics superior to all other options, with an extraordinary impact resistance (high resistance to deformation), and is able to regain its elasticity. It is also very resistant to abrasion and traction. It is a perfect dielectric, so it is recommended for use in electrical and telecommunications applications, and it does not absorb moisture. The package must be designed in such a way as to not affect the measurement and transmission of temperature and moisture, while also protecting the electronics. A polypropylene membrane has been included in the package, aimed to allow separating the sensor from direct contact with the concrete while allowing the sensor layer to be exposed to the concrete environment to measure the parameters of interest. This is the key for a successful functionality.

The sensor layer consists of a commercial SHT21P humidity and temperature sensor (Sensirion, Staefa, Switzerland). This is a digital sensor that can be easily integrated into any circuit due to its ability to be directly connected to any microprocessor system. The claimed performance of the sensor shows good accuracy, good repeatability and good response time. Besides the capacitive relative humidity sensor and a band gap temperature sensor, the sensor chip includes an amplifier, an A/D converter, OTP memory and a processing unit. The developed sensor generates an electrical signal directly related to the temperature and moisture, which is subsequently converted inside the same device into digital format. This sensor transmits data in two packets, the first for the humidity reading and the second for the temperature reading.
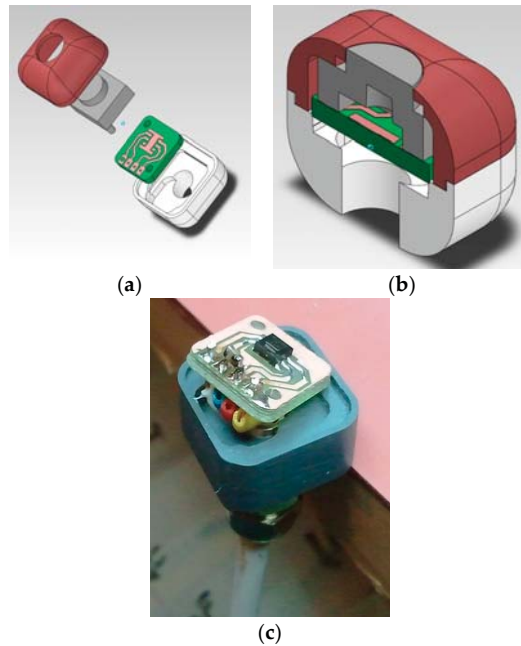
(**a**)   (**b**)

(**c**)

Figure 3. Embedded sensor (**a**,**b**) Detail of the construction (**c**) Real implementation.

A second alternative for the sensor was considered. The classical method of monitoring concrete curing is to only monitor the temperature profile. The DS18B20 digital temperature sensor from Maxim Integrated (San Jose, CA, USA) was chosen. This sensor is fairly precise, can provide up to 12 bits of resolution and works with any microcontroller using a single digital pin. The sensor covers the range of temperatures for the target application. Moreover, in order to avoid developing a specific package for the sensor, it was coated with a special lacquer and resin. The suitability of this protection was evaluated and tested to ensure accurate measurement of the temperature, but also protection from direct contact. This ensures a simpler solution which is crucial for developing a low cost and compact non-destructive sensing system suitable to be embedded into concrete.

Regarding the antenna, an internal antenna was always used, because of the risks related to maintaining the watertightness, and also impact risks with the aggregates of concrete (inert granular materials such as sand, gravel or crushed stone) being poured from the mixer truck to the moulds to be filled. Different antenna types were used, these being a dipole quarter wave (85 mm), a custom design on a PCB and a chip antenna. Our final design uses a dipole quarter wave antenna with an RP-SMA connector.

3.2. Software

The language used in programming is almost entirely a C variant for microcontrollers which intended to be as portable as possible, so excluding lower level parts (drivers and statement types, memory or basis functions) it can be read as ANSI C. In any case, this microcontroller comes with very limited memory, so it is often preferred to include assembly code directly, although this is usually limited to certain functions related to wireless transmission. The firmware of the microcontroller is programmed by Code Composer or IAR for a device from the MSP430 family.

The MSP FET430 programmer-debugger connects to the device via the JTAG or SWD protocol and allows for the step by step debugging and the use of breakpoints. Concerning the wireless controller,

it is programmed using the protocol stack from Texas Instruments (SimpliciTI). It has been used for the programming of the nodes in different phases, and for the implementation of network algorithms. Synchronization processes and waiting periods have had to be defined as the protocol stack does not implement them.

Figure 4 shows the simplified routing mechanism. First, a static routing has been chosen for power saving, but also considering the possibility of errors within the nodes. Thus, if the node 3 has low battery, it may decide not to acknowledge the message from node 1 to avoid having to route it, leaving this task to other nodes. These nodes could also reject it; then node 1 would detect that none of the nodes has accepted its message and would wait for the next turn, but this time it would include a field of alert, informing its neighbors that they must transmit the message in turns in order to spread the payload.
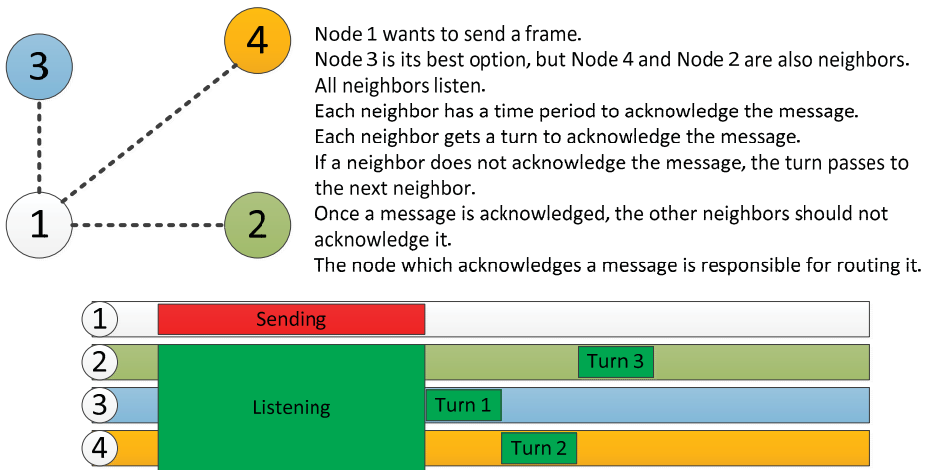
Figure 4. Simplified routing mechanism.

Sending messages will always be coordinated in stages as shown in Figure 5a, with an upstream direction according to their numbers as shown in Figure 1. As this network is only used for monitoring, it is only necessary to move upwards. This allows nodes with low identification number to enter the idle state first, providing a balance of energy consumption, since these nodes will be the first ones sending data (they are placed in the lower part of the concrete structure).

The time required to send data always corresponds to a very low percentage of the total, so that the duty cycle will be below 1%, which allows for the device to be inactive most of the time (deep sleep state) as shown in Figure 5b; only the clock circuitry is enabled in order to know when the devices must wake up again.

Thanks to these mechanisms a very low energy consumption is achieved and the use of small batteries becomes appropriate.
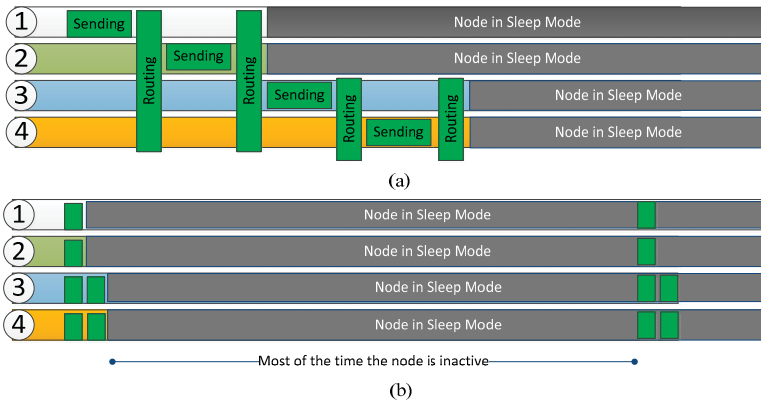
Figure 5. (**a**) Data sending spread out in intervals; (**b**) Power saving mechanism.

3.3. Monitoring and Control System

The monitoring application is a program that receives and stores data from the sensor nodes, and shows the data read from the sensors in a graphical user interface. The software has been developed with additional features in order to provide flexibility. Thus, the user can identify each node. The data from the sensors is periodically sent and stored within the database, and it is visualized in real time. In addition, the sampling period can be modified by the user. In fact, during the first days after the concrete is poured, the activity of the curing process is greater, requiring a large number of measurements. Some alarms received from the sensors, related to battery failure and voltage values lower than a preset threshold, have also been included.

Figure 6 shows the infrastructure required for the monitoring system. It consists of the motes inside the concrete, a gateway (NPort W2150A, Moxa, New Taipei City, Taiwan) between the measurement network and the data network, a WiFi router (AWK-4121, also from Moxa), a monitoring device and a storage device.

Figure 6. Scheme of the monitoring system based on the wireless sensor network.

4. Experimental Results

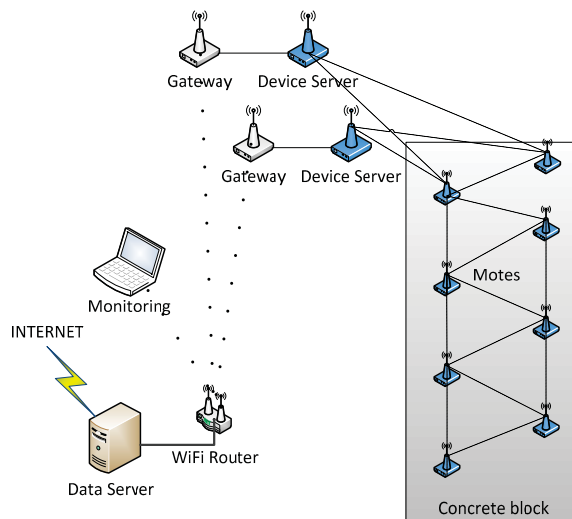Traditionally, for conventional Portland concrete a limit of 3 days for curing, 7 days for rapid hardening, and 28 days for the development of the characteristic strength of the concrete is considered. Considering a tradeoff between the accuracy of a thermo-chemical-mechanical model that could analyze the concrete curing and the use of a wireless sensor network with motes supplied from batteries, the sampling period of temperature for the sensors was set as follows: four measurements every hour for the first three days. From day four, one measurement every hour until the seventh day. Between day 7 and day 28, one temperature measurement is performed every three hours. To save power consumption, several sensor readings are stored and sent together in the same message. This approach provides lower power consumption in the radio module, but the longer frames increase the error rate. Thus, a maximum of 128 bytes per frame was set. The estimation of power consumption, including a safety margin, determines a minimum battery capacity of 707 mAh. The average power consumption of each mote has been estimated using a Simulink model and considering each component separately. After the model, real world estimations have been taken into account, like battery discharge caused by temperature, loss of efficiency on components and non-linear effects such as communication issues. Therefore, the selected battery has a capacity of 1100 mAh.

4.1. Calibration and Testing

Once the DS18B20 sensor was chosen all efforts were devoted to checking the suitability of the protection based on lacquer and resin. The accuracy and linearity of the sensor were monitored, as well as the effect that the protection had on the accuracy. Nevertheless, the suitability of the package in Figure 3 for the SHT21P sensor was also determined testing the ability of the acetal material and the poly propylene membrane to protect the sensor and also its effect on the measurement accuracy. This fact will allow evaluating the acetal material to be used to protect the wireless board in Figure 2. First, the effect of the lacquer and resin on the sensor response was tested within a water container, since it allows simultaneously checking the sealing of the sensors and the accuracy, as shown in Figure 7.

Figure 7. Sealing testing with sensors immersed in water.

Figure 8a,b show the thermocouples, named TerEle and TerEst respectively, used as reference for the calibration. The sensor was found to work accurately when compared to the temperature readings of the thermocouples with PT100 probes, as shown in Figure 9. Note that the sensor features a similar evolution to that of the references (0.99) with an offset of only 0.72 ± 0.5 °C. The results showed a proper operation without communication problems.

(**a**)   (**b**)
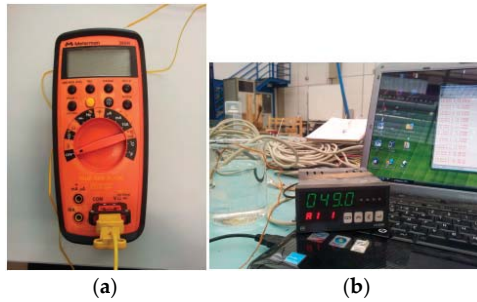
Figure 8. Thermocouples used for the sensor calibration. (**a**) TerEle, (**b**) TerEst.
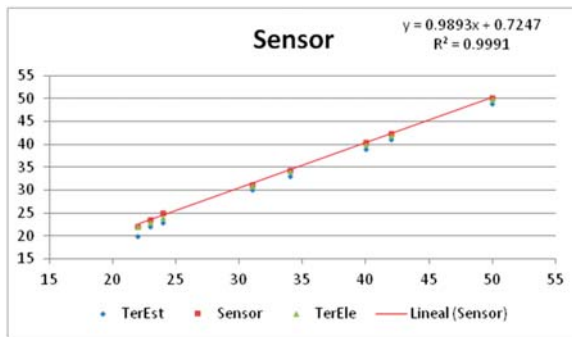
Figure 9. Calibration of a sensor with respect to the thermocouple references TerEst and TerEle.

On the other hand, the transmission of each node was analyzed under different situations, especially considering the power of the output signal, which is a critical parameter taking into account the attenuation due to the concrete and the steel. Figure 10 shows the mote and a FSIQ 3 spectrum analyzer (Rhode & Schwarz, Munich, Germany). The tests were carried out both open air and inside concrete.
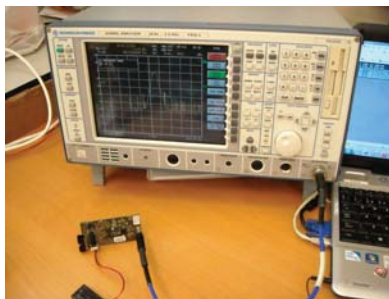
Figure 10. Test of output power level for the output signal of mote.

A preliminary test was performed to verify the validity of the specifications on several specimens of concrete as shown in Figure 11. Figure 11a also shows the data acquisition board used to gather physical signals from several sensors. The aim of this test is to determine if the sensors can perform inside a concrete specimen in order to replicate, as closely as possible, the construction site conditions. The specimens are created using a steel mould and filled with concrete. This test allows determining

the effect of the combination of formwork and concrete on the transmission of data. In addition, the test also examines the transmission distance of the sensors. The radio transmission was tested by an emission test at fixed intervals of a data frame. As expected, the effect of concrete, rebar and formwork reduced the transmission distance of the system. When the moulds were taken out of the specimens, the transmission distance doubled. The design has been optimized to achieve a percentage of packets successfully received above 95%. The data was transmitted since the first pouring of concrete for 28 days. These results are obtained for a transmission distance of approximately two meters. Results confirmed the feasibility of the proposed approach. There were no changes in the transmission of data in the early stages of concrete pouring when it is wetter.

A simple LabVIEW program was designed to interact with the sensors, logging the time, temperature readings and also from which sensor the data was received.

(**a**)

(**b**)
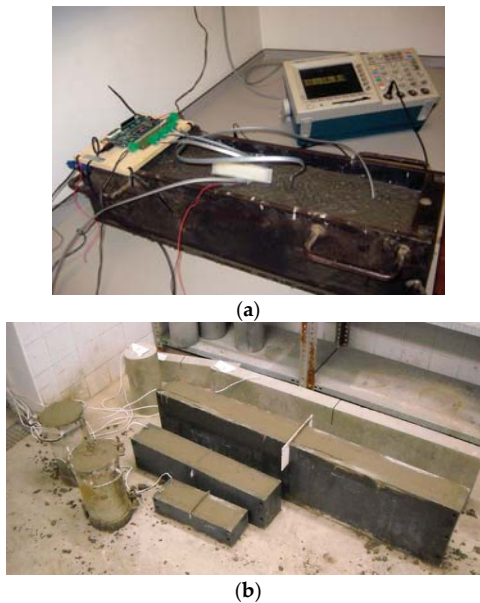
Figure 11. Preliminary test. (**a**) Data acquisition board to gather physical signals from several sensors (**b**) Mote buried in several specimens of concrete.

4.2. Construction Site

The designed motes and the wireless sensor network were introduced in a construction site. The place is located in a viaduct of the high-speed Antequera-Granada railway in the region of Andalucia, in the South of Spain.

The preliminary study of the location of the motes determined the use of the 869.4 MHz to 869.65 MHz band given by ETSI EN 300 220 and short-range device (SRD)of unspecified use, with a transmission power of 500 mW (+27 dBm). This need arises from the distances to be treated being around three meters, which are larger than the ones tested in the laboratory, and therefore the dynamic range should be increased. Thus, to reach the mentioned power level, it is necessary to include in the design an output power amplifier and a linear noise amplifier to improve the reception of signals.

Figure 12 depicts the construction, the dimensions in centimeters and the location of each sensor. It consists of a footing to support a shaft. The dimensions of the footing are $8 \times 11 \times 3$ m. The placement of the sensors is performed in the middle plane 1.5 m above the base. This location was carefully chosen from experimental results.

For a distance of 1.5 m the motes can transmit from within the concrete to the outside, with a signal loss between 48 dBm and 64 dBm. Given an initial signal of −36 dBm and an error margin, it's the maximum distance where a sniffer can receive packets in order to debug the deployment, considering a sensibility of −102 dBm. Eight points of measurement are taken, as indicated in Figure 12b. The devices must be installed prior to pouring the concrete, and placed in preset positions using clamps or flanges to ensure that the measuring point corresponds to the one required. Each sensor is installed and tested individually by placing the antenna in different positions. Due to the nature of the test and the inability to repeat the process, several repeaters where included in the structure, to assure transmission of the values and to provide redundancy.

Figure 13a shows an operator testing a mote on the structure, and Figure 13b shows a mote tied to a steel bar. For the wireless sensing node, a 4 mm thick acetal box has been used. The final dimensions of the prototype wireless node package are 85 mm width and length, and a height of 60 mm.

(**a**)

Figure 12. *Cont.*
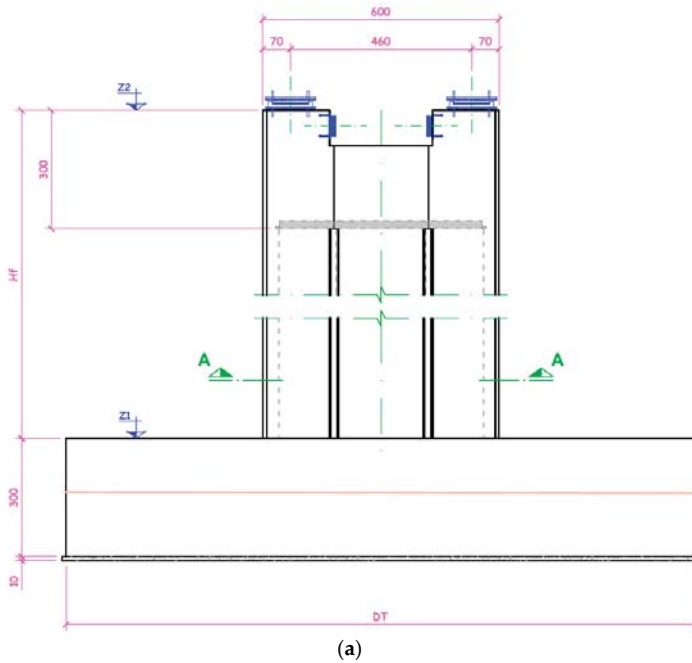
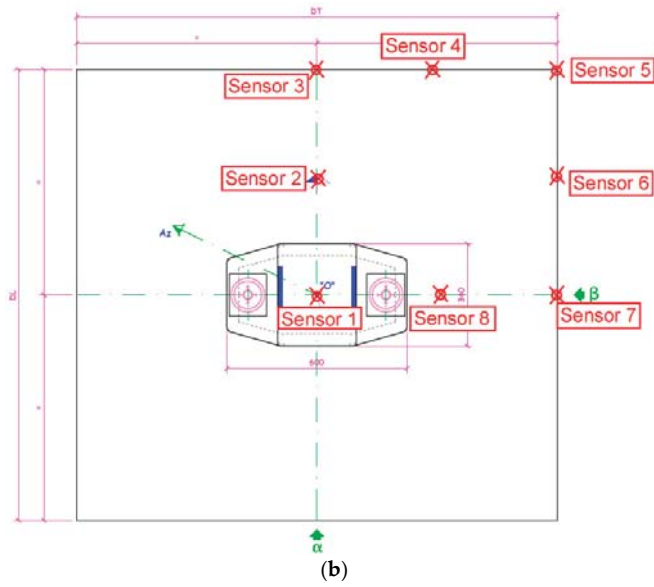**Figure 12.** (**a**) Construction element and (**b**) location of sensors.

Figure 13. (**a**) Operator testing a sensor on the steel structure; (**b**) Detail of the mote tied to a steel bar.

Figure 14 shows the ambient temperature profile since the first hours that the concrete is poured up to day 28, respectively. Note that the variations in temperature can reach almost 10 °C.
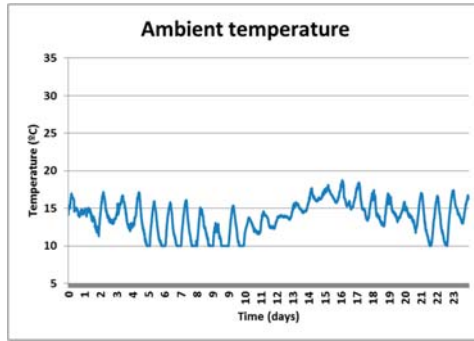
Figure 14. Ambient temperature profile.

Figure 15 shows the temperature profile of the sensors located into the concrete. Note how the temperature of the concrete increased during the first hours due to the reaction between cement and water (hydration process). Once the temperature reaches a peak, then it starts to drop following the fluctuations of the ambient temperature. It can also be seen that the sensors 3 and 4 provide a slightly lower temperature, probably because they are located at the edges of the structure and the heat was dissipated away.

The higher temperatures inside the concrete structure did not affect the operation of the wireless sensor network, and measurements were gathered with the expected sensor accuracy, which is protected from impacts produced during the pouring process.

The applicability of the developed monitoring system in civil engineering is evident. For example, by considering a thermo-mechanical model of the hydration kinetics of the cement (see for example [15]), the degree of hydration of the cement can be controlled by monitoring the evolution of the temperature with time. These experimental measurements provide useful information to validate the theoretical predictions and to check whether the hydration process inside the concrete bulk is developing correctly. Combining these measurements with those of the mechanical properties of concrete at early age and a convenient thermo-mechanical model of the structure (by using a finite elements model, for example) provides a powerful tool to determine the most convenient stripping time and, subsequently, the associated costs can be reduced. Table 1 shows the results of the main mechanical parameters, such as the compressive strength (f_c), the Young's modulus (E) and splitting tensile strength (f_{st}) of the employed concrete at early age and at 28 days. These results were obtained in laboratory by means of standard tests on cubic and cylindrical concrete specimens [19].

(**a**)

Figure 15. *Cont.*

(b)
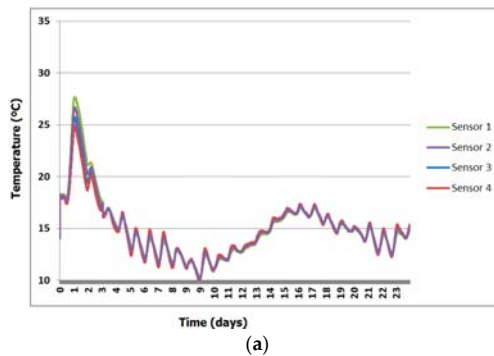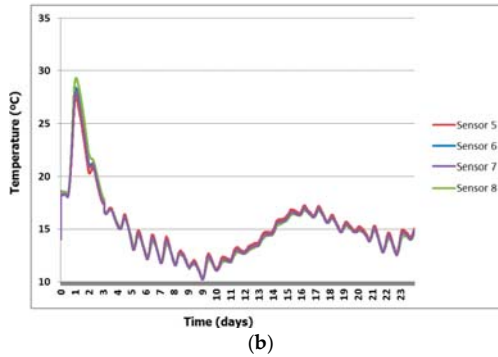
Figure 15. Temperature profile of concrete from sensors. (**a**) Sensors from 1 to 4. (**b**) Sensors from 5 to 8.

Table 1. Main mechanical parameters of the employed concrete at early ages and 28 days.

Age (Days)	f_c (MPa)	E (GPa)	f_{st} (MPa)
1	20.1	18.9	1.9
2	26.7	19.8	2.1
3	28.2	20.7	2.3
7	36.2	24.4	2.8
28	43.9	28.4	4.3

In order to confirm the reliability of the results, a thermo-chemical-mechanical finite element model (FEM) of the concrete footing has been implemented in ABAQUS. Details about the hydration kinetics modeling of the concrete employed in this analysis are given in [15]. The composition of the concrete in this work was 1/0.43/2.31/3.07/2.57/0.009 (cement/water/sand/gravel/superplasticizer) with a class II Portland cement content of 325 kg/m^3. The concrete parameters necessary to perform the thermal analysis were a density of 2240 kg/m^3, a coefficient of thermal expansion of 10^{-5} K^{-1}, a thermal conductivity of 2.5 W/mK, a specific heat of 922 J/kgK, a film coefficient between concrete-air of 9 W/m^2·K and a room temperature of 18 °C. The behavior of the material was considered linear and elastic with a Young modulus ranging with age according to the values indicated in Table 1. Figure 16a shows the geometry and discretization employed in the FEM model for the concrete and the steel rebars, as well as a view of the internal profile of temperatures obtained for concrete at 7 days. In Figure 16b the evolution of the temperature with age for the nodal points of the FEM model corresponding with the location of sensors 3 and 5 (Figure 12b) are shown.

(a)

Figure 16. *Cont.*

(**b**)
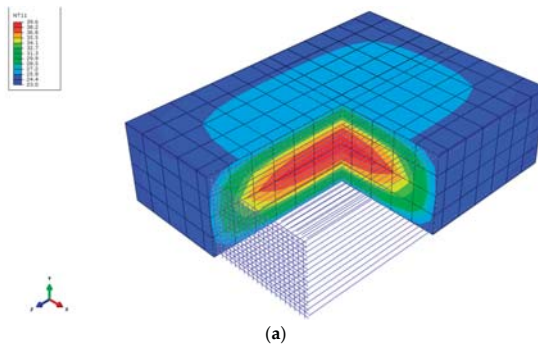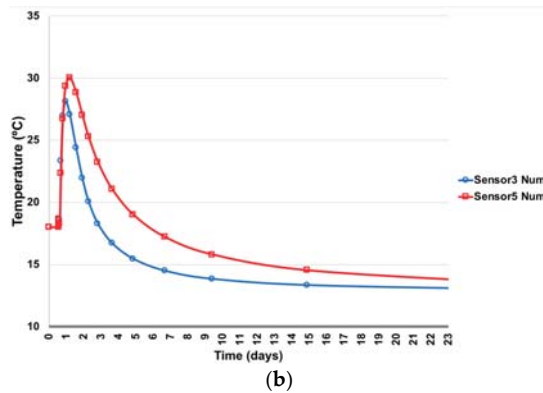
Figure 16. (**a**) Thermo-chemical-mechanical FEM model of the footing; (**b**) Numerical prediction of the evolution of temperature with age for sensors 3 and 5.

As observed from the results shown in Figure 16b, the evolution of the temperature with age predicted by the numerical model is in good agreement with that measured by the sensors placed in the structure. Obviously, the values of the numerical prediction are not exactly the same than those obtained experimentally because of several factors, which are difficult to consider in the model. For example, variation of the room temperature between day and night was not considered, the film coefficient was estimated from literature [19] and not experimentally obtained, and the temperature of the soil was also estimated, etc. However, the maximum values of the temperature given by the sensors, the age at which these values were obtained and the evolution of the concrete temperature after the peak value (cooling) are quite similar, which reflects the reliability of the experimental measurements.

5. Conclusions

A non-destructive wireless sensing system has been implemented, with a low cost and compact design suitable to be embedded into concrete for monitoring the curing process. This application demands small sized sensing nodes with low power requirements, complicating the design. Preliminary tests were accomplished in the laboratory to validate the accuracy and the protection of the developed sensors, and the data transmission to a data acquisition system located outside the concrete. Finally, the wireless system was used in a civil construction and the data from the sensors was successfully transmitted during the period required to monitor the concrete from the first pour.

Author Contributions: Joaquín Cabezas and Trinidad Sánchez-Rodríguez designed and developed the system. Juan Antonio Gómez-Galán and Héctor Cifuentes proposed the ideas, performed the experiments and prepared the manuscript, and Ramón González Carvajal made the supervision and direction of the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alam, M.N.; Bhuiyan, R.H.; Dougal, R.A.; Ali, M. Concrete Moisture Content Measurement Using Interdigitated Near-Field Sensors. *IEEE Sens. J.* **2010**, *10*, 1243–1248. [CrossRef]
2. Zhou, S.; Deng, F.; Yu, L.; Li, B.; Wu, X.; Yin, B. A Novel Passive Wireless Sensor for Concrete Humidity Monitoring. *Sensors* **2016**, *16*, 1535. [CrossRef] [PubMed]
3. Liu, Z.; Yu, Y.; Liu, G.; Wang, J.; Mao, X. Design of a wireless measurement system based on WSNs for large bridges. *Measurement* **2014**, *50*, 324–330. [CrossRef]

4. Chae, M.J.; Yoo, H.S.; Kim, J.Y.; Cho, M.Y. Development of a wireless sensor network system for suspension bridge health monitoring. *Autom. Constr.* **2012**, *21*, 237–252. [CrossRef]

5. Yu, Y.; Zhao, X.; Wang, Y.; Ou, J. A study on PVDF sensor using wireless experimental system for bridge structural local monitoring. *Telecommun. Syst.* **2013**, *52*, 2357–2366. [CrossRef]

6. Grasley, Z.C.; Lange, D.A.; D'Ambrosia, M.D. Internal relative humidity and drying stress gradients in concrete. *Mater. Struct.* **2006**, *39*, 901–909. [CrossRef]

7. Perveen, K.; Bridges, G.E.; Bhadra, S.; Thomson, D.J. Corrosion Potential Sensor for Remote Monitoring of Civil Structure Based on Printed Circuit Board Sensor. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 2422–2430. [CrossRef]

8. Qiao, G.; Sun, G.; Hong, Y.; Liu, T.; Guan, X. Corrosion in Reinforced Concrete Panels: Wireless Monitoring and Wavelet-Based Analysis. *Sensors* **2014**, *14*, 3395–3407. [CrossRef] [PubMed]

9. Mehta, W.; El Zarki, M. A Bluetooth Based Sensor Network for Civil Infrastructure Health Monitoring. *Wirel. Netw.* **2004**, *10*, 401–412. [CrossRef]

10. Shams, K.M.Z.; Ali, M. Wireless Power Transmission to a Buried Sensor in Concrete. *IEEE Sens. J.* **2007**, *7*, 1573–1577. [CrossRef]

11. Quinn, W.; Angove, P.; Buckley, J.; Barrett, J.; Kelly, G. Design and performance analysis of an embedded wireless sensor for monitoring concrete curing and structural health. *J. Civil Struct. Health Monit.* **2011**, *1*, 47–59. [CrossRef]

12. Ulm, F.-J.; Coussy, O. Modeling of thermochemomechanical couplings of concrete at early ages. *J. Eng. Mech.* **1995**, *121*, 785–794. [CrossRef]

13. Cervera, M.; Faria, R.; Oliver, J.; Prato, T. Numerical modeling of concrete curing, regarding hydration and temperature phenomena. *Comput. Struct.* **2002**, *80*, 1511–1521. [CrossRef]

14. Bentz, D.P. Influence of water-to-cement ratio on hydration kinetics: Simple models based on spatial considerations. *Cement Concr. Res.* **2006**, *36*, 238–244. [CrossRef]

15. Schindler, A.K.; Folliard, K.J. Heat of hydration models for cementitious materials. *ACI Mater. J.* **2005**, *102*, 24–33.

16. Sung, W.-T.; Hsu, Y.-C. Designing an industrial real-time measurement and monitoring system based on embedded system and ZigBee. *Exp. Syst. Appl.* **2011**, *38*, 4522–4529. [CrossRef]

17. Rault, T.; Bouabdallah, A.; Challal, Y. Energy efficiency in wireless sensor networks: A top-down survey. *Comput. Netw.* **2014**, *67*, 104–122. [CrossRef]

18. Every, E.V.; Deyhim, A. Embedded Sensors for Life-Time Monitoring of Concrete. In Proceedings of the 4th International Conference on Structural Health Monitoring on Intelligent Infrastructure (SHMII-4), Zurich, Switzerland, 22–24 July 2009.

19. Neville, A.M. *Properties of Concrete*, 4th and Final edition; Longman Group UK Limited: London, UK, 1995.

Vibroacoustic Impact on the Architectonic Heritage When Using Replicas of 16th Century Weapons

Angel Tomas Lloret [1], Sandra Sendra [2,3], Jaime Lloret [3,*], Romina del Rey [4] and Miguel Louis Cereceda [1]

[1] Construcciones Arquitectonicas, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, San Vicente del Raspeig 03690, Spain; atllm@hotmail.com (A.T.L.); miguel.louis@ua.es (M.L.C.)
[2] Signal Theory, Telematics and Communications Department (TSTC), Universidad de Granada, C/Periodista Daniel Saucedo Aranda s/n, Granada 18071, Spain; ssendra@ugr.es
[3] Integrated Management Coastal Research Institute, Universitat Politècnica de València, C/Paranimf, n° 1, Grao de Gandia 46730, Spain
[4] Centre for Physics Technologies, Universitat Politècnica de València, Camino Vera, s/n., Valencia 46022, Spain; roderey@fis.upv.es
* Correspondence: jlloret@dcom.upv.es; Tel.: +34-609549043

Abstract: The recreation of historical battles next to old buildings, walls, churches, fortifications or historical facades belonging to the historical heritage of a city, has always been a source of controversy and discussion. In the absence of a clear legislation about how these buildings can be affected by the use of blunderbusses and pyrotechnics, it is necessary to carry out practical experiments to test the effect of these celebrations on these buildings. For this reason, this paper presents a set of practical experiments where the vibroacoustic effect of using weapons such as blunderbusses and harquebuses is analyzed. To gather these measurements, we have used several sound level meters and 3-axis accelerometers placed on the facade of an old building. The tests have been carried out at the Moors and Christians festival of Villajoyosa (Spain) which is internationally famous for this festival. In order to carry out the tests, six harquebusiers shot their firearms and the sensors placed along the facade of the building at different height collected the data. The results of these devices allow us to study the vibroacoustic impact on the facade depending on the height.

Keywords: acoustic impact; structural movements; architectonic heritage; pyrotechnics; vibration; harquebus; blunderbuss; sensors

1. Introduction

The Corsican battles along the eastern coast of Spain, in the Mediterranean Sea, between the XIV and XVIII centuries, have left a trail of fortifications and watchtowers built for defense purposes. Nowadays, many of these buildings are considered part of the national historical heritage.

Many European countries (especially the ones that border the Mediterranean Sea) celebrate numerous festivals. Such celebrations often recreate battles and events that occurred centuries ago, and in these festivals, large amounts of gunpowder are used in the form of fireworks, firecrackers, and shots from weapons. These elements produce large rumblings, accompanied by a subsequent blast. In addition, these recreations are usually carried out in the old quarters of the towns taking advantage of the beauty of the scenery. The preservation of this heritage represents an interesting tourist attraction which implies an economic benefit for that country [1]. In addition, the use of fireworks and pyrotechnics are not unique to these celebrations. They are also widely used as a means of signaling and beaconing in the livestock industry, agriculture, fisheries, and mining, among others.

To monitor these activities we need to use sensor devices capable of perceiving a physical stimulus from the environment and translating it into a signal or a measurable parameter. In most cases, these sensors need additional electronics to show the measurements in an understandable way [2].

There are many applications where acoustic waves and sensor deployments provide great benefits [3]. Regarding to the use of acoustic waves, one of the most used applications is the detection of gas and water leaks [4,5]. Fluid leaks that flow through pipes often generate small vibrations [6] and even small acoustic signals. By capturing these parameters, it is possible to detect such leaks [7].

One the other hand, buildings are sometimes subjected to hazards such as earthquakes and strong winds, fire or vandalism, among others. In order to mitigate, or at least to monitor these facts, researchers generally use sensors and sensor networks. These architectures can measure acceleration, displacement, strain, temperature, smoke, acoustic pressure, etc. Buildings may face many risks that should be monitored, such as aging of structural performance, fatigue, damage, gas leaks, fires, etc. From the results of this monitoring, researchers or owners can take appropriate actions such as structural control, maintenance, evacuation guidance, warning, alarm, firefighting, and rescue or security measures, among others [8].

There are very few papers published regarding the use of sensors for monitoring and protecting our architectural heritage, although one field where sensors are often used is in the monitoring of artworks [9] and frescoes [10].

In this paper, we present the design and development of a series of experiments to study the vibroacoustic impact that the use of blunderbusses and harquebuses can cause on the historical heritage. The tests should help us to determine if these old buildings would be affected by the effect of pyrotechnics. The acoustic impact is understood as the magnitude of the mechanical wave that impacts the facade. Depending on the amplitude of this mechanic wave, which will be measured by using sound level meters, the facade may suffer some movements that can be perceived as a vibration by accelerometer sensors. With these two parameters, we will evaluate the vibroacoustic impact over the facade as a function of the height.

In order to carry out this work, we selected an old building located in Villajoyosa (Province of Alicante, Spain) which is famous for its Moors and Christians festival. Several sound level meters and vibration sensors were installed on the facade. The noise was generated using the synchronized shots of six harquebusiers. The firearms used are reproductions of actual weapons used in these ancient battles. From the results, we have drawn vibroacoustic impact maps of the facade. The results showed that the impact over the facade could have serious implications such as the detachment of elements of the facade that could both harm people and generate irreparable damage on these monuments that we want to protect.

The rest of this paper is structured as follows: Section 2 presents some papers where the authors carried out studies on the effects of acoustic impacts on building facades. Section 3 describes the scenario where the tests have been carried out and the tools and materials used during the tests. The experiments carried out and the measurement results are discussed in Section 4. Finally, Section 5 presents the conclusions and suggestions for future work.

2. Related Work

There are several studies related to the analysis of acoustic impacts on buildings. However, they are more focused on the study of the insulating characteristics of housing to ensure the welfare of its inhabitants. This is the case of a study presented by Park et al. [11] where the authors examined how residents in apartment buildings perceive and react to impact sounds coming from their upstairs neighbors' dwellings. The authors described the noise annoyance and non-acoustic factors. In addition, the authors stated that the noise sensitivity of a person had a direct impact on the perceived disturbance and an indirect impact on its annoyance. Finally, the results showed that the annoyance generated by the impact noise on the floor was associated with self-reported health complaints.

An interesting example related with the surveying of ancient buildings was presented by Fregonese et al. [12]. This work describes some techniques for surveying ancient buildings. In this case, the authors used a Terrestrial Laser Scanner (TLS) in combination with traditional topographic techniques such as geometric leveling and topographic networks for 3D control to survey and study the Palazzo del Capitano in Mantua (Province of Mantua, Italy). The main aim of this study was the detection of surface displacements in buildings. For the case of this specific building, the authors analyzed images taken over time to see if any part of the structure of this palace had undergone some structural movement. The authors conclude that nowadays, the known techniques still present serious limitations for monitoring ancient buildings. Costanzo et al. [13] used similar techniques for analyzing the St. Augustine Monumental Compound, located in the historical center of Cosenza (Calabria, Italy).

Cuadra et al. [14] presented a study to estimate the dynamic characteristics of Inca stone structures due to microtremors that generate very small displacements. The authors assumed that stone structures present an elastic behavior capable to withstanding small strain conditions. From their results, the authors proposed a new method to evaluate the seismic behavior of these constructions and consequently the seismic vulnerability of these structures.

Ceriotti et al. [10] described the development of a wireless sensor network on Torre Aquila, a medieval tower located in Trento (Province of Trento, Italy). The system was based on a set of motes called TMotes, 3MATE! environmental nodes and fiber optic sensors stretching the length of the tower to measure its deformation. The tests were performed during four months. The authors concluded that WSN-based monitoring systems are effective tools to assess a tower's stability, and deliver the needed data to monitor this kind of structure.

As we can see in these previous works, there is significant interest in monitoring and analyzing the status of our cultural heritage, but this section tries to be more focused on previous works where the authors analyzed the effect of sound pressure on buildings.

Firstly, Karatzetzou [15] developed three different applications of ambient noise measurement methods to measure the foundation-soil stiffness in terms of shear-wave velocity, the fundamental frequency of the structure in its present state, the mechanical properties and the distribution and intensity of damage to masonry walls. The authors analyzed the ambient vibrations measurements on three monuments in the medieval city of Rhodes (Island of Rhodes, Greece) and the surrounding soil. From the results, the authors identified the dynamic characteristics of these structures and the foundation-soil where these structures were located.

We should also highlight that it is easy to find studies [16,17] where authors analyzed facade sound insulation and material response as a function of the working frequency and the source that generated the noise. Yu and Kang [18] also presented a study to analyze the differences in environmental impact between different architectural acoustic materials in residential buildings. The study was performed on five houses with different infrastructure, i.e., the authors used a bungalow, detached, semi-detached, and terraced houses and apartments to carry out their tests. For each building, the authors compared several common wall materials with similar sound transmission losses and three different glazing ratios. The results showed the importance of considering the environmental sustainability of acoustic materials when planning to build a house.

There are very few works where the practical study also includes measurements regarding to vibrations of structural movements. This is the case of Klos [19] who performed a study to analyze the vibroacoustic response of buildings due to sonic boom exposure. To do this, the authors used 112 transducers installed in three bedrooms of a house. Accelerometers were attached to the walls, windows, and ceilings of these three rooms and microphones were placed at random locations in each room and in the surroundings of the house to characterize the resulting noise indoors and the effect of diffraction of the boom around the house. The tests included shaker excitation of the walls and windows of the house and reverberation time measurements, among others, inside and outside the house at various locations.

Finally, regarding the use of weapons, one of the first works related to the acoustic characteristics of harquebuses and blunderbusses was performed by Marco Sanjuán in 1996 [20]. In this work, the author performed a set of measurements of the sound levels of different types of commercial gunpowders. This study does not cover the three types of explosive used for the different models and sizes of historical weapon reproductions, but the obtained values are an interesting indicator to start with our measurements of sound pressure over a facade.

As far as we know, there is no study that relates the use of weapons during re-enactments of these ancient battles and the effect of these on the cultural heritage. We think it is important to maintain these celebrations, their reproductions, and memory as faithful as possible to the actual facts, however, we also believe it is important to preserve the monumental remains and our historical heritage. It is, therefore, important to characterize how historical buildings may be affected by the use of this type of fireworks.

3. Scenario and Material Description

This section shows the scenario where the tests have been carried out and the tools and materials used during the different tests.

3.1. Scenario

To perform these tests, we have chosen an uninhabited house of ancient construction since we cannot expose old architectural elements to the stress of these experiments. This property is located in a pedestrian street in the historic city center. In the same area, we can find many historical sights of the city such as the defensive walls built between 1551 and 1565 to repel Berber attacks. Figure 1 shows a section of the urban plan of Villajoyosa (Alicante, Spain) and the area where the tests have been performed.

Figure 1. Section of the urban plan of Villajoyosa (Alicante, Spain).

The building is a block placed on the historical center of Villajoyosa, declared as Heritage of Cultural Interest (in Spanish *Bien de Interés Cultural-BIC*), a denomination given by the Spanish Government for not only material heritage, like monuments or movable works of art, but also intangible cultural heritage). The building presents a rectangular base and it is placed between other buildings. The foundation base around the perimeter is scarce or nonexistent. The structural support function

is performed by load-bearing walls that present higher thickness than normal. These are buried and mainly transmit the loads to the ground by friction. The building has a surface structural system based on load-bearing walls of masonry mortar, as has been observed in various historical documents. Load-bearing walls that form the facades typically have a coating of hydraulic cement and sand mortar, and a traditional plaster cover of lime and sand. The load-bearing walls decrease in thickness as the building height increases. Finally, as a vertical communication element, there is a staircase that connects the different floors. It is built as a brick vault where the stairs are placed.

Figure 2 shows a scale drawing of the scenario where the tests were carried out. This picture also shows the position of the harquebusiers, the positions where the sound level meters have been placed as well as the dimensions of the facade section exposed to our tests. The surface of our facade is 6 m wide and 6 m in height. To perform the tests, six harquebusiers and four sound level meters have been used. Each harquebusier is identified as "Font_X" where X indicates the harquebusier. Sound level meters are identified as "Sono_Y" where Y indicates the device. This nomenclature is used in the rest of paper.
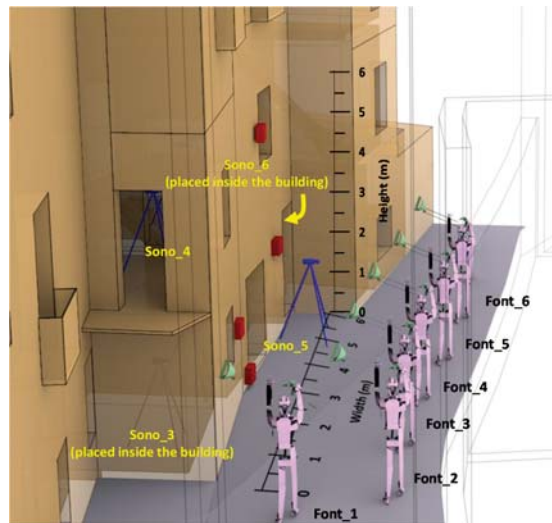
Figure 2. Scenario where measurements have been performed, and the position of harquebusiers and sound level meters.

3.2. Material Used to Perform the Tests

This subsection describes the features and sizes of the weapons used and technical data such as the maximum sound pressure levels that these weapons can reach and the physical properties of the black gunpowder used during the tests.

3.2.1. Weapons Used in the Replica of the Real 16th Century Battle

During the replica of these centenarian battles, harquebusiers use firearms without bullets which generate a large flare (See Figure 3). Harquebusiers usually use blunderbusses or harquebuses. Both are firearms with small differences between them:

Blunderbuss: These are short muzzle-loading firearms with a large caliber; it is destined to fire shrapnel or small pellets. This kind of firearm is generally used over short distances. The blunderbuss is usually activated by a spark, although there are some models that are activated by percussion or small detonators (pistons). They were often used by bandits and smugglers. The blunderbuss is the

successor of the earlier "petrol" used by bandits and Catalan supporters of Christians troops during the 17th century.

Harquebus: This is an old firearm with an iron gun barrel and a wooden box. The name "harquebus" is due to the arcuate shape of the mouth that was modified in the blunderbuss to facilitate the addition of gunpowder. The harquebus is a firearm that belongs to the muzzle-loading category, i.e., it is loaded through the muzzle. It is smaller in size than the blunderbuss, with a bore established between 14 and 18 mm. According to its construction and type of chamber, a harquebus can load from 15 g to 50 g of black powder. The fireplace primer and the firing pin driven by the crossbow are placed on the back of the gun barrel. The muzzle of the weapon has a flared look for easy loading.

Figure 3. Harquebusier in a replica of battles.

Figure 4 shows some weapons, explosive detonators and transport containers used in the battle replicas.

Figure 4. Weapons, explosives detonators and transport container used in the replicas of battles.

3.2.2. Industrial Explosive Mixtures Used in the Detonations

Mine gunpowder is not an actual explosive product since its reaction deflagrates instead of detonating. The feature of this product is mainly its large volume of combustion gases. It is a ternary mixture of sulfur, potassium nitrate and charcoal. Another important feature is its high combustion temperature. Its reaction rate is low compared with the rates given for the denotation of explosives. Depending on the caliber, the rate can reach speeds to 500 m/s.

Black gun powder is quite sensitive, so it is necessary to maintain strict safety measures while handling it. The mine powders are supplied in granulated form. At the end of the manufacturing process, graphite is added because its conductive nature prevents the formation of electrostatic charges. It also provides a lubricating action so that good fluidity is achieved.

The detonation of any explosive always produces a certain proportion of gases that are toxic to the human body. The greater or lesser effect depends on other factors such as the composition of the explosive itself and the amount of oxygen dissolved in it. Table 1 shows the physical properties of black gunpowder used in harquebuses and blunderbusses.

Table 1. Physical properties of black gunpowder.

Parameter	Pressure (P)	Kinetic Energy (F)	Volume (V)	Amount of Heat (Q)	Temperature (T)
Black Gunpowder	2970 kg/cm^2	210×10^3 kgm/kg	330 L	500 kcal/kg	2100 °C

Black gunpowder is a mixture of potassium nitrate:carbon:sulfur (KNO_3:C:S) which percentages are 75%:15%:10% (potassium nitrate can be replaced by sodium nitrate). It is usually prepared wet to ensure that the sodium or potassium nitrate particles impregnate the C and S particles. In this way, the mixing process offers better results. Only 50% of black gunpowder becomes hot gases when burned; the rest is burnt as very fine particles. Figure 5 shows the relationship between the black gunpowder used in shooting and the sound level pressure registered at 5 m.
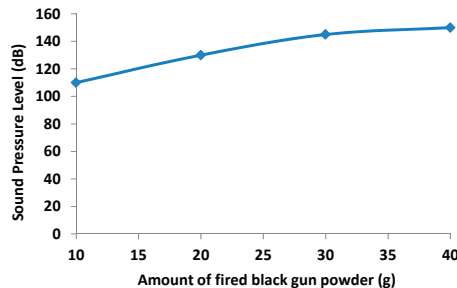
Figure 5. Sound pressure level as a function of black gun powder used.

To carry out the tests performed in these experiments, we have used black gunpowder marketed by the company Maxam Outdoors (Madrid, Spain) whose characteristics are [21]:

- Grain size: From 0.35 to 1.19 mm.
- Amount of KNO_3: 74%.
- Container Size: 1 kg.

3.3. Explanations about the Test Bench

In order to perform these tests, we have required the presence of six experimented harquebusiers placed as shown in Figure 2. In our case, the impact study is conducted in several rooms of a house that lacks furniture, which may modify the results. The measurements are made at different heights. The six noise sources are located outside the building under study. The distance between harquebusiers is 2 m and the distance between the building and harquebusiers is 3 m (see Figure 6).

In regard to the location of the sensors, the sound level meters are located within the building located at 2 m from the inner side of the facade, window or access door. With this distance, the reflections over the reflecting surfaces of the building itself do not affect the measurement. Furthermore, because the building does not contain furniture, we can be sure that no additional element will interfere with the measurements results. The vibration sensors take measurements at three axes. They are fixed to the outside part of the facade with bee wax that gives freedom of movement to the mobile parts of accelerometer sensors. Figure 6 also shows the location of the vibration sensors. They are identified as "Acc X" where X indicates the device. Table 2 lists the characteristics of the weapons used in this study.
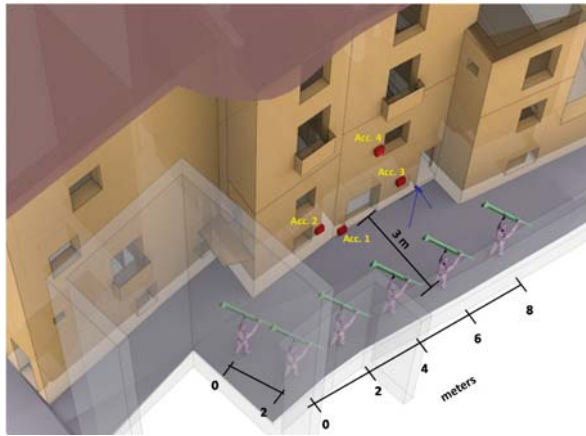
Figure 6. Distribution of harquebusiers and accelerometers.

Table 2. Features of weapons used during these tests.

Shooter	Kind of Weapon	Brand	Caliber (mm.)	Barrel Length (cm.)
1	Muzzle-loading Harquebus	BOPE	62	50
2	Blunderbuss	ARDESA	62	50
3	Muzzle-loading Harquebus	BOPE	28	41
4	Muzzle-loading Harquebus	GIL	28	40
5	Blunderbuss	BOPE	60	50
6	Blunderbuss	ARDESA	62	50

In the performed shots, we have not used metal projectiles. Therefore, to imitate the sound of a real shot, the amount of explosive used must be increased with amounts of up to 40 g of explosive deflagration inside the weapon. In our case, we have used several historical weapon reproductions the same type (muzzle-loading) with different sizes and dimensions. The amount of explosive has been the same in all cases.

The position adopted by the harquebusiers is conditioned by security aspects, i.e., the residue of shot and flame should not interfere on other harquebusiers or viewers. The second reason for adopting this position is to keep the weapon in a safe position after the recoil caused by the explosion inside it (see Figure 7). In this position, the weapon must point to the sky at an angle of 45° from the horizontal with the trigger. The gun must be kept over the head of the harquebusier for safety reasons.

Figure 7. Harquebusiers carrying out a shot where the explosive deflagration is observed.

Finally, the weapon must remain turned outwards on its longitudinal axis to avoid any kind of incident if the deflagration comes back down the barrel (see Figure 8).

Figure 8. Harquebusiers in the safety position ready to perform a shot.

The vibration sensors used in our study were Bosch BMA056-3 axis accelerometers [22] (see Figure 9). The encapsulation size is $3 \times 3 \times 0.95$ mm^3. It is one of most used models for virtual reality and navigation applications, motion tracking, shock and vibration detection in smartphones and mobile devices. This sensor is able to measure magnetic field levels in the range of $\pm$ 1000 µT (for X-, Y-axis) and $\pm$ 2500 µT (for Z-axis) with a resolution of 0.3 µT. Regarding the accelerometer measurements, the Bosch BMA056 has a resolution of 12 bits and a sensitivity tolerance of $\pm 4\%$. In addition, this sensor can be programmed in different ranges ± 2 g; ± 4 g; ± 8 g; ± 16 g and Zero-g offset of ± 80 mg. In order to collect and save the data from acceleration sensors, an ESP8266 node has been used. The acceleration sensors are connected to the node through I^{2}C bus. Finally, the sampling rate is 12.5 Hz.
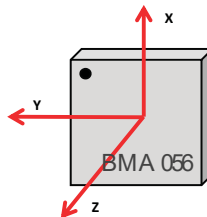
Figure 9. BMA056-3-Axis accelerometers.

In relation to the synchronization of the accelerometer sensors, the sensors are initialized using a real-time clock connected to the ESP8266 node. This device allows us to have a temporal reference to perform the measures. The real-time clock and the acceleration sensor are slave elements within the communication through the I^{2}C bus while the ESP8266 node acts as the master. Figure 10 shows a schematic of the connections.

The sound pressure measurements have been performed using professional sound level meters. In these tests, we have selected the Bruël & Kjaer Type 2250 sound level meter [23]. It performs Class 1 measurements according to the international standards. This sound level meter has a wide dynamic range that supports both the most intense noises and those barely perceived above the background noise. Moreover, the Bruël & Kjaer Type 2250 sound level meter is able to measure from the low-frequency option, ranges from 20 kHz down to infrasound for measurement of noise sources suspected of emitting very low frequency noise. The device can perform in 1/1 octave and 1/3 octave bands. Figure 11 shows one of the sound level meters used in the test bench.
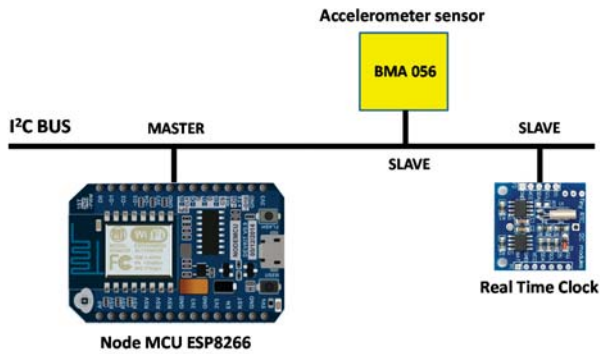
Figure 10. Schema of connections.

Figure 11. A Bruël & Kjaer Type 2250 sound level meter.

Finally, it is important to know the environmental conditions because acoustic waves are highly affected by the temperature. Moreover, very high wind speeds can mask the measurement results. As an important datum, when the speed wind is higher than 5 m/s, the resulting acoustic measurements are not valid [24]. In our case, the following environmental values were registered:

- Temperature: 29.1 °C.
- Relative Humidity: 75.3%.
- Speed wind: 1 m/s.

With these values, we can perform the tests and we can be sure that the registered values really represent the effect of the harquebuses and sound pressure on the facade.

4. Tests Performed in the Real Scenario and Measurement Results

This section shows the experiments carried out in the scenario presented in the previous section. We also discuss the measurement results. These results show the acoustic and seismic impact on the building.

4.1. Acoustic Impact and Building Movement Results

In order to perform the tests and take measurements of sound pressure levels and vibration recorded at different points of the facade, the harquebusiers made two sets of shots. This subsection shows the values registered by the sound pressure level and vibrations registered at different points of the facade during the two sets of shots.

4.1.1. Considerations over the Performed Measurements

When measuring the acoustic impact, we should take into account the effect we want to measure. This fact will determine the frequency weighting we should use. Frequency weightings correspond to acoustic sound issues and have a large psychoacoustic component.

Acoustic analyzers aim to approximate their responses to the human ear. While Sound Pressure Level (SPL) in dB describes the physical phenomenon, the weighted decibel level describes the existing loudness, i.e., the perception of a sound by a human. In our case, we want to measure the physical effect over the facade and therefore we will use linear weighting. Linear weighting is a flat frequency response between 10 Hz and 20 kHz $\pm$ 1.5 dB, excluding the microphone's response. The results directly represent the effect we want to measure. All information concerning this weighting is presented in the international standard IEC 61672: 2013. Our measurements are performed in Z weighting, according to Bruël & Kjaer nomenclature (measured in dB) in the 1/3 octave band [25–27].

4.1.2. Values of Sound Pressure Level Registered over the Facade

First we analyze the average sound pressure level received by the facade when there is no activity, i.e., without pyrotechnics or harquebusiers performing shots. Figure 12 shows the average value (calculated from the data from all sound level meters) of the background sound level in dB registered on the facade.
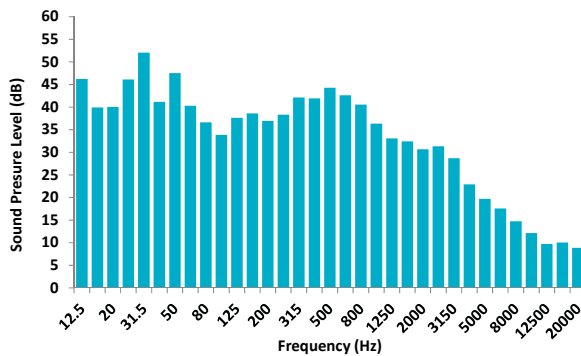
Figure 12. Average background sound pressure (dB).

We can see that maximum value of sound pressure is 52 dB at 31.5 Hz and around 45 dB at 12.5 Hz, 50 Hz and 500 Hz, i.e., the highest values are registered at low frequencies. Low frequencies are the ones which present the biggest impact over the buildings and are the ones which most disturb people.

Figure 13 shows the sound pressure level measured by the four sound level meters during the first shot. As we can see, the sound level meter that registers the lowest sound pressure level is the one placed at lowest altitude respect to the floor (103 dB). Moreover, the rest of devices record pressure sound levels around 110 dB. The biggest noise impact is generated in the range of 200–250 Hz.
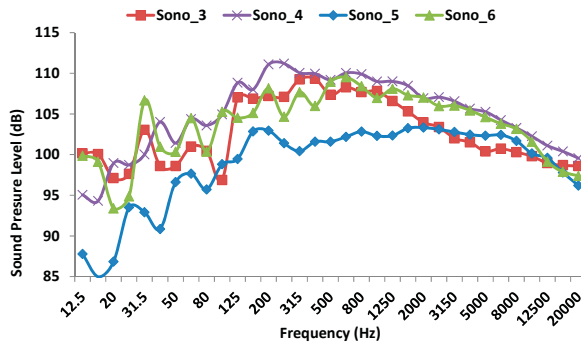
Figure 13. Sound pressure level measured by the four sound level meters during the first shot.

Figure 14 shows the sound pressure level measured by the four sound level meters during the second shot. Again, the sound level meter that registers the lowest sound pressure level is the one placed at lowest height (approximately 1 m) with respect to the floor (108 dB at 200 Hz). This is Sono_5. Moreover, the rest of sound level meters record maximum values around 111–113 dB in the frequency range of 125 to 300 Hz. From this value, the received signal is attenuated. The biggest sound impact is recorded at 125 Hz. by Sono_3 (with a value of 113.31 dB), which is approximately at 1.60 m height.
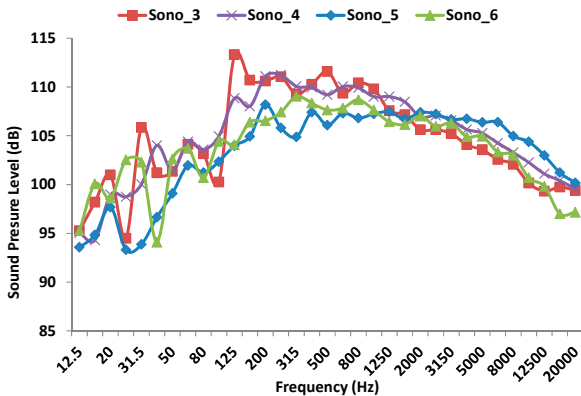
Figure 14. Sound pressure level measured by the four sound level meters during the second shot.

Finally, Figure 15 shows the map of maximum levels of sound pressure on the facade as a function of the height and width of facade. Figure 13 shows the maximum value registered simultaneously in all sound level meters which are placed on the positions specified in Figure 2. As we can see, the highest value is registered at 1.5–2 m (which corresponds to the height of the ground floor) with a value around 113 dB. We can also see that a sound pressure level peak is registered at 4 m height with a value of 109 dB.

As we can see, there are two areas of greatest sound pressure impact. This is because during a shot, the explosion generated by the weapon has approximately been generated at 1.5–2 m height, while the deflagration can reach several meters (up to 5–6 m) in height and also causes the propagation of sound waves. Figure 16 shows a drawing of a harquebusier shooting a weapon, where it is easy to see the explosion and deflagration.
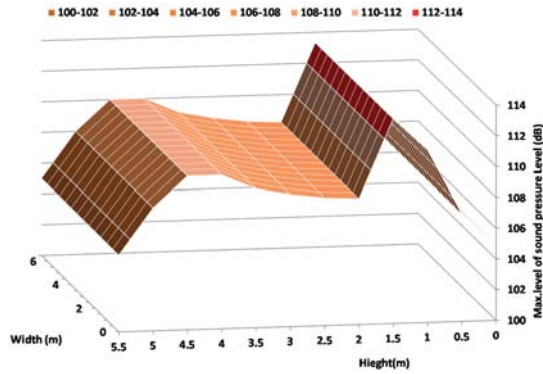
Figure 15. Map of maximum values of sound pressure level registered on the facade in (dB) as a function of the height.
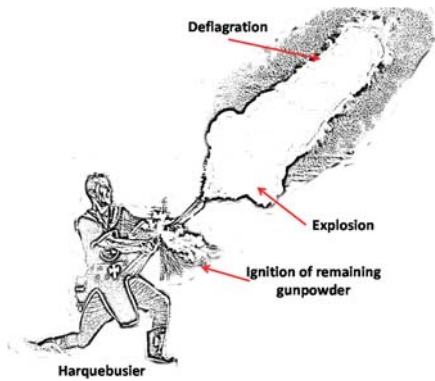
Figure 16. Harquebusier shooting a weapon.

4.1.3. Values of Vibration Index Registered over the Facade

To start with the measurements of vibration, we should take into account that a facade and a building, in general, only receive vibrations from seismic sources or from industrial machinery called seismic noise, although the human activity and natural phenomena can also generate structural movements.

Usually, when we talk about vibration measurements in buildings, this magnitude is usually represented as a vibration index (Law) in dB [28]. This value is obtained by Equation (1):

$$Vibration\ Index\ (\text{dB}) = 20 \cdot \log \frac{acc\left(\text{m/S}^2\right)}{acc_0\left(\text{m/S}^2\right)} \tag{1}$$

where, $acc\left(\text{m/s}^2\right)$ is the acceleration measured by the acceleration sensor and $acc_0\left(\text{m/s}^2\right)$ is the acceleration reference that depends on the device. In our case, the $acc_0 = 10^{-6}\ \text{m/s}^2$.

In our case, we are going to measure the effect of shooting on the facade cladding. The most affected elements with the biggest probability of coming off will be those that present a deficient adhesion to the structure. For this reason, the sensors are not embedded in the structure. To fix the sensors, we have used bees wax which is widely used when vibration dosimeters are employed. The sensors are in contact with the elements of structure that could be detached by the effect of the received

acoustic waves. This subsection shows the results of the vibrations detected over the facade cladding during the two shots.

To process the data from the 3-axis accelerometers, we have measured the components, i.e., the acceleration registered in X-axis, the acceleration registered in Y-axis and the acceleration registered in Z-axis. After that, these values are combined using Equation (2):

$$a_{Combined} = \sqrt{a_x{}^2 + a_y{}^2 + a_z{}^2}$$

(2)

where, a_x, a_y and a_z are the values of acceleration registered in each orthogonal axis, X, Y and Z.

The measurement time for each case is 1 min, but we have selected the time during which the movements has been detected. The measurement results are as follows: firstly, Figure 17 shows the vibration values registered by accelerometer 1 during the shots. We can observe that the maximum amplitude is around 128–130 dB for both cases while the rest of the time, the sensor registers an average value of 93–94 dB as ground vibration. The recorded impulse is narrow which implies very low structural movements. This is because this sensor is placed on the ground floor.
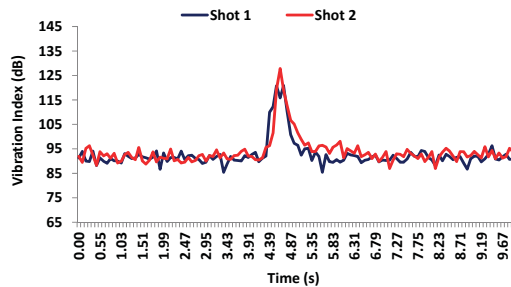
Figure 17. Vibration levels registered by accelerometer 1 on the facade during the shots.

Figure 18 shows the values registered by accelerometer 2 during the both shots. In this case, shot 2 registers a slightly bigger vibration (133 dB) than shot 1 which reaches 125 dB. The ground vibration is around 85 and 90 dB for both shots. The sensor results also show that shot 2 has needed around 3.2 s in order for the wall to totally absorb the vibration. For shot 1, this time is smaller since the maximum registered level has been lower. This sensor is placed at 2.20 m height.
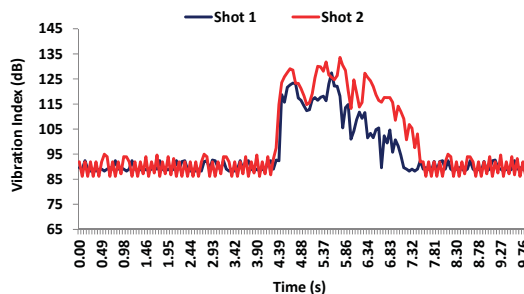
Figure 18. Vibration levels registered by accelerometer 2 on the facade during the shots.

Figure 19 shows the values registered by accelerometer 3 during both shots. In this case, shot 1 registers a bigger vibration (124 dB) than shot 2 which reaches 113 dB. The rest of time the vibration level is around 94 dB for both shots. We can also see that the vibration needs around 3 s to be totally absorbed. This is because this sensor is placed at 1 m height.
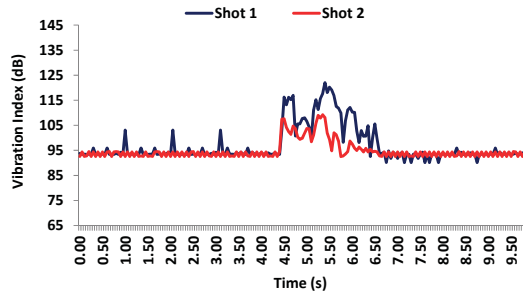
Figure 19. Vibration levels registered by accelerometer 3 on the facade during the shots.

Figure 20 shows the values registered by accelerometer 4 during both shots. Both shots register similar values, i.e., 125 dB). The ground vibration is around 92–93 dB for both shots. Again, the results have shown that the facade needed around 3.2 s in order to totally absorb the vibration. This sensor is placed at 4.5 m height.
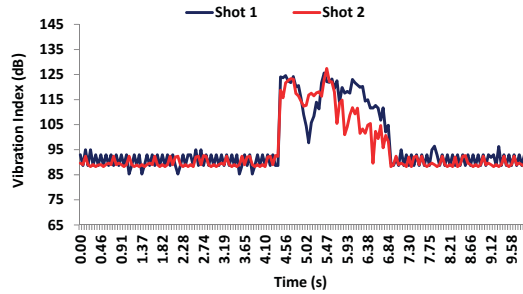
Figure 20. Vibration levels registered by accelerometer 4 on the facade during the shots.

Figures 21–23 show the vibration maps on the facade for the X-axis, Y-axis, and Z-axis of the sensors, respectively. To make them, we have taken the maximum value of the vibration index registered at the same time at different places on the facade. These places are the sensor positions and because they are placed at different heights, these figures show us the movements as a function of different heights.
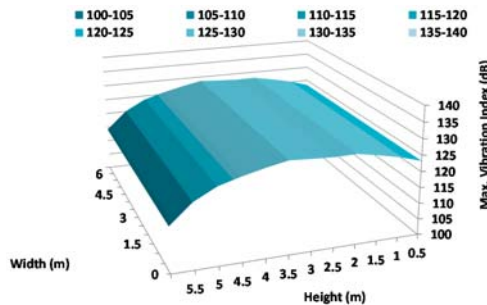
Figure 21. Maximum variation of vibration in the X-Axis.
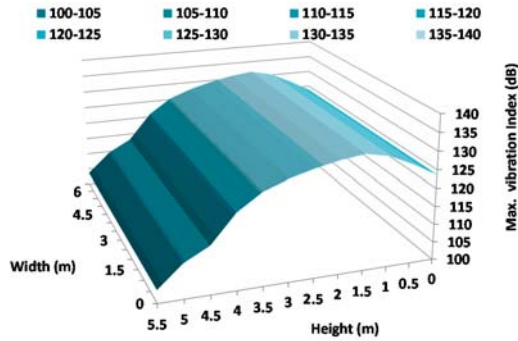
Figure 22. Maximum variation of vibration in the Y-Axis.
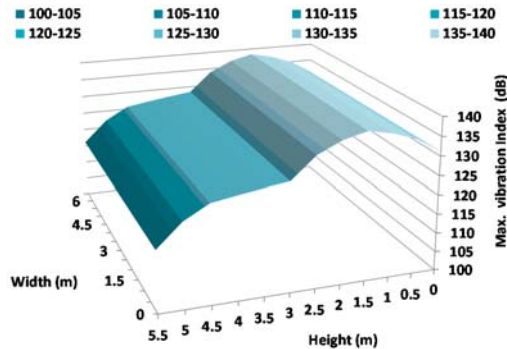
Figure 23. Maximum variation of vibration in the Z-Axis.

As we can see, the highest vibration index is registered in the Z-axis (Figure 23). In addition, the highest value is registered around 1.5 m with maximum values of 133 dB which corresponds with the height where the harquebuses generate the highest sound pressure level. The axis that registers the lowest vibration is the X-axis, with a maximum value of 128 dB at 3 m height. As we can conclude, the Z-axis registers the biggest vibration index since the sensors register the movements in the transversal sense of the wall which presents a larger freedom of movement.

4.2. Discussion

Initially, this study was carried out to find out if there is any risk to people and built heritage in order to ensure the perpetuity of a traditional celebration. In any case, the security of people should prevail over any type of patrimony. For this reason, the human risks derived from this activity in terms of the historical scenario are a result of the possible detachment of objects from the facades onto the people.

To carry out the test, the most unfavorable acoustic site has been established, i.e., a narrow street with tall buildings, and the most unfavorable situation of the shooters, i.e., a big concentration of harquebusiers with an excess of gunpowder in the weapon.

The purpose of this test was to detect the vibrations caused by the most unfavorable case of these festivals where elements such as the eaves, cornices, carpentry, balconies, and facade cladding can fall off and harm people and these detachments could be irreversible and consequently the historic heritage would remain damaged

From the results, we can make several observations. Firstly, it is clear that the use of these weapons causes significant movements on the building facade. In regard to the acoustic impact, we note that the highest level is recorded at 1.5 m. The sound pressure levels recorded show some values higher than 113 dB at 160 Hz. However, there is another area where another peak is detected. It is registered at 4.5 m height, whose value is lower than the peak located at 1.5 m. This is due to blast and deflagration generated by this kind of weapon.

We have also observed that the most pronounced structural movements are recorded at 1.5 m height. The time the structure needs to absorb and completely attenuate this movement is also maximum at this point. Another aspect also observed, and predictable, is that the structural movements are greater in the Z axis since it measures the transverse movements of the wall instead of the longitudinal movements along the length and width of the facade. Finally, we should denote that the highest vibration indexes are also registered around 1.5 m, which corresponds to the area of the greatest sound pressure level registered on the facade.

Although we should not compare the structural movements produced by pyrotechnics and captured at the level of the façade with those generated by a natural earthquake, we can perform a small comparison in relation to the damages observed in both cases. To do this, we will use the Mercalli scale [29]. The Mercalli scale quantifies the effects of an earthquake on the Earth's surface, humans, objects of nature, and man-made structures on a scale from I (not felt) to XII (total destruction). Figure 24 summarizes the ranges considered by this scale.
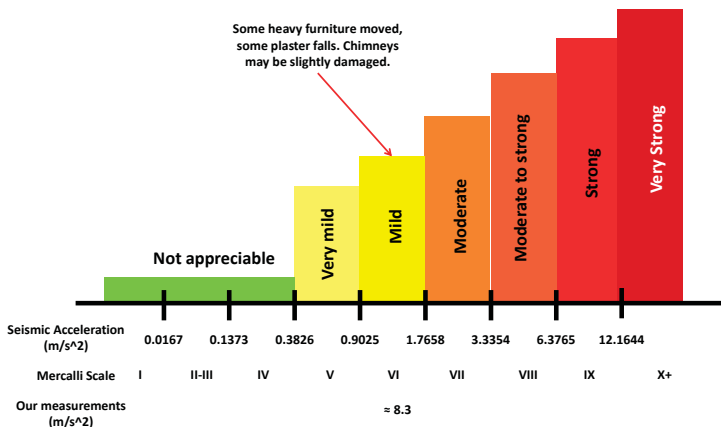
Figure 24. Mercalli scale and its damages.

As we have seen in our results, the biggest vibration has been detected in the Z axis in which case we have obtained amplitudes of 133 dB (which is equivalent to 8.341 m/s^2) for accelerometer 1. These values would imply a seismic intensity of grade IX in the Mercalli scale.

Obviously, these damages are mitigated because the effect and vibrations are only registered at the facade level. However, if we consider the damages generated, we would be a seismic intensity of grade V-VI in the Mercalli scale. As Figure 25 shows, after performing our tests some damages have been detected in elements such as windows and plastering of the facades. Finally, we can conclude this section by affirming that the use of these weapons in a scenario like the one used in this tests can generate some structural damage, especially in elements like windows and plastering of facades. Consequently, the results of this study can help the organizers of these festivals in terms of security and damage prevention.

Figure 25. Some damages detected after tests.

5. Conclusions

During the summer months, many European countries celebrate hundreds of festivals that commemorate historical events. Along the Spanish Mediterranean coast, the population often replicates ancient battles which involve plenty of shots of firearms and pyrotechnics along the historical center of the cities. These battles reproduce the events when Arab troops tried to conquer the Spanish land and how the Christian troops repelled these attacks. As a result of our history and all these facts, nowadays, we have many buildings like fortresses, watchtowers, churches and historic facades, which must be preserved to keep our history. However, so far we have not checked whether the reproduction of such celebrations near these historic buildings can cause some kind of irreparable damage to these buildings.

Because of this need, in this paper, we have presented a study about the vibroacoustic impact that buildings may suffer by the presence of pyrotechnics like the weapons typically used in these celebrations. To carry out this study, we used an old house where several sound level meters and vibration sensors have been placed over the facade. We have performed two sets of measurements. In both cases, six harquebusiers simultaneously shot their weapons.

From the results, we have seen that the greatest acoustic impact is recorded at 1.5 m height, although a higher sound pressure level is also recorded at 4.5 m height. As shown in Figure 16, when a weapon is fired, some deflagrations and explosions are generated. While the deflagration can reach up to 5–6 m height (this is reason of the high value of sound pressure level at 4.5 m), the explosion is generated next to the weapon mouth. In addition a small deflagration of the remaining gunpowder is registered near the handle of the weapon. Consequently, a high sound pressure level value has been registered at 1.5–2 m. These sound pressure level values has a consequence on the facade and in our case, this effect is the registration of some structural movements and vibrations which also increase with the height.

Finally, we can conclude that the height of the building must be taken into account when the vibroacoustic impact is measured. Obviously, another important factor is the height of the explosions, deflagrations and physical features of weapons. The results obtained demonstrate that the vibroacoustic effect on facade claddings in a bad state is high and it is capable of causing detachment, and human injuries (as Figure 22 shows). During the inventory of the damages caused, there were several broken windows, so these tests have served to determine the extent of the damage. The determination of human injuries, rather than the built heritage, will be decisive for limiting these traditional celebrations in historic centers of cities and towns.

These tests were carried out in old buildings but in the future, we would like to perform them on old architectural elements such as historical facades and watchtowers. In addition, we would like to implement a sensor network [30] able to take measurements for long periods using distributed database management techniques [31,32]. Moreover, we have in mind to create alarms [33] in cases where the building suffers some type of micro-vibration which can deteriorate the building structure somehow, when it has had a stressful event as the one reproduced in this work. Finally, we would like

to test how these buildings can be affected by the explosions of aerial fireworks and compare these results with those obtained in the current paper.

References

1. Ashworth, G.J.; Larkbam, P.J. *Building a New Heritage: Tourism, Culture and Identity in the New Europe*, 1st ed.; Routledge Library Editions: Tourism; Routledge: Abingdon, UK, 2013.

2. Garcia, M.; Bri, D.; Sendra, S.; Lloret, J. Practical deployments of wireless sensor networks: A survey. *Int. J. Adv. Netw. Serv.* **2010**, *3*, 170–185.

3. Saïd, M.N.A.; Brown, W.C.; Shirtliffe, C.J.; Maurenbrecher, A.H.P. Monitoring of the building envelope of a heritage house: A case study. *Energy Build.* **1999**, *30*, 211–219. [CrossRef]

4. Murvay, P.S.; Silea, I. A survey on gas leak detection and localization techniques. *J. Loss Prev. Process Ind.* **2012**, *25*, 966–973. [CrossRef]

5. Martini, A.; Troncossi, M.; Rivola, A.; Nascetti, D. Preliminary investigations on automatic detection of leaks in water distribution networks by means of vibration monitoring. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 535–544. [CrossRef]

6. Yazdekhasti, S.; Piratla, K.R.; Atamturktur, S.; Khan, A. Experimental evaluation of a vibration-based leak detection technique for water pipelines. *Struct. Infrastruct. Eng.* **2017**, 1–10. [CrossRef]

7. Martini, A.; Troncossi, M.; Rivola, A. Leak Detection in Water-Filled Small-Diameter Polyethylene Pipes by Means of Acoustic Emission Measurements. *Appl. Sci.* **2016**, *7*, 2. [CrossRef]

8. Kurata, N.; Spencer, B.F., Jr.; Ruiz-Sandoval, M. Risk monitoring of buildings using wireless sensor network. In Proceedings of the International Workshop on Advanced Sensors, Structural Health Monitoring, and Smart Structures, Tokyo, Japan, 10–11 November 2003; pp. 10–11.

9. D'Amato, F.; Gamba, P.; Goldoni, E. Monitoring heritage buildings and artworks with Wireless Sensor Networks. In Proceedings of the IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS 2012), Perugia, Italy, 28 September 2012; pp. 1–6.

10. Ceriotti, M.; Mottola, L.; Picco, G.P.; Murphy, A.L.; Guna, S.; Corra, M.; Pozzi, M.; Zonta, D.; Zanon, P. Monitoring heritage buildings with wireless sensor networks: The Torre Aquila deployment. In Proceedings of the 2009 International Conference on Information Processing in Sensor Networks, San Francisco, CA, USA, 13–16 April 2009; pp. 277–288.

11. Park, S.H.; Lee, P.J.; Yang, K.S.; Kim, K.W. Relationships between non-acoustic factors and subjective reactions to floor impact noise in apartment buildings. *J. Acoust. Soc. Am.* **2016**, *139*, 1158–1167. [CrossRef] [PubMed]

12. Fregonese, L.; Barbieri, G.; Biolzi, L.; Bocciarelli, M.; Frigeri, A.; Taffurelli, L. Surveying and monitoring for vulnerability assessment of an ancient building. *Sensors* **2013**, *13*, 9747–9773. [CrossRef] [PubMed]

13. Costanzo, A.; Minasi, M.; Casula, G.; Musacchio, M.; Buongiorno, M.F. Combined use of terrestrial laser scanning and IR thermography applied to a historical building. *Sensors* **2014**, *15*, 194–213. [CrossRef] [PubMed]

14. Cuadra, C.; Sato, Y.; Tokeshi, J.; Kanno, H.; Ogawa, J.; Karkee, M.B.; Rojas, J. Evaluation of the dynamic characteristics of typical Inca heritage structures in Machupicchu. *WIT Trans. Built Environ.* **2005**, *83*, 237–244. [CrossRef]

15. Karatzetzou, A.; Negulescu, C.; Manakou, M.; François, B.; Seyedi, D.M.; Pitilakis, D.; Pitilakis, K. Ambient vibration measurements on monuments in the Medieval City of Rhodes, Greece. *Bull. Earthq. Eng.* **2015**, *13*, 331–345. [CrossRef]

16. Scrosatia, C.; Scamonia, F.; Pratob, A.; Secchic, S.; Faustid, P.; Astolfi, A.; Barbaresi, L.; D'Alessandro, F.; Di Bella, A.; Zambon, G. Uncertainty of facade sound insulation by a Round Robin Test. Evaluations of low-frequency procedure and single numbers. *Build. Environ.* **2016**, *105*, 253–266. [CrossRef]

17. Secchi, S.; Cellai, G.; Fausti, P.; Santoni, A.; Martello, N.Z. Sound Transmission between Rooms with Curtain Wall Façades: A Case Study. *Build. Acoust.* **2015**, *22*, 193–207. [CrossRef]

18. Yu, C.-J.; Kang, J. Environmental impact of acoustic materials in residential buildings. *Build. Environ.* **2009**, *44*, 2166–2175. [CrossRef]

19. Klos, J. Vibro-Acoustic Response of Buildings Due to Sonic Boom Exposure: July 2007 Field Test. September 2008; Reference Document: NASA/TM-2008-215349. Available online: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20080037572.pdf (accessed on 30 July 2017).

20. Sanjuán, J.M. Estudios del Nivel Sonoro de Diferentes Tipos de Pólvoras Negras Para Arcabuces y Trabucos de Salvas. Revista Especial día 4 que Fuera de la Junta Central de Fiestas de Villena. 1996. Available online: http://server-die.alc.upv.es/Alcoy/Tuna/interest.htm (accessed on 30 July 2017).

21. Maxam Black Gunpowder Features. In Armeria Alberdi Website. Available online: http://www.armeriaalberdi.com/polvora-maxam-arcabuz-html.html (accessed on 1 July 2017).

22. Features of Bosch BMA056-3 Axis Accelerometer. Available online: http://www1.futureelectronics.com/doc/BOSCH/BMA250-0273141121.pdf (accessed on 30 July 2017).

23. Sound Level Meter Bruël & Kjaer Type 2250 Features. In Armeria Alberdi Website. Available online: https://www.bksv.com/en/products/sound-and-vibration-meters/sound-level-meters-and-vibration-meters/Type-2250-S (accessed on 1 July 2017).

24. Harris, C. *Manual de Medidas Acusticas y Control del Ruido*; Harris, C.M., Ed.; Mcgraw-Hill Interamericana De España S. A.: Milladoiro, Spain, 1995.

25. IEC 61672-1:2013-Electroacoustics—Sound Level Meters—Part 1: Specifications. Available online: https://webstore.iec.ch/publication/5708 (accessed on 30 July 2017).

26. IEC 61672-2:2013-Electroacoustics—Sound Level Meters—Part 2: Pattern Evaluation Tests. Available online: https://webstore.iec.ch/publication/5709 (accessed on 30 July 2017).

27. IEC 61672-3:2013-Electroacoustics—Sound Level Meters—Part 3: Periodic Tests. Available online: https://webstore.iec.ch/publication/5710 (accessed on 30 July 2017).

28. Crocker, M.J. (Ed.) *Handbook of Noise and Vibration Control*; John Wiley & Sons: Hoboken, NJ, USA, 2007.

29. Sendra, S.; Lloret, A.T.; Lloret, J.; Rodrigues, J.J.P.C. A wireless sensor network deployment to detect the degeneration of cement used in construction. *Int. J. Ad Hoc Ubiquitous Comput.* **2014**, *15*, 147–160. [CrossRef]

30. Wood, H.O.; Neumann, F. Modified Mercalli Intensity Scale of 1931. *Bull. Seismol. Soc. Am.* **1931**, *21*, 277–283.

31. Diallo, O.; Rodrigues, J.J.P.C.; Sene, M.; Lloret, J. Distributed database management techniques for wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 604–620. [CrossRef]

32. Meghanathan, N. A Generic Algorithm to Determine Maximum Bottleneck Node Weight-based *Data* Gathering Trees for Wireless Sensor Networks. *Netw. Protoc. Algorithms* **2015**, *7*, 18–51. [CrossRef]

33. Tornai, K.; Oláh, A.; Levendovszky, J. Monitoring Scheme for Event and Danger Detection in Wireless Sensor Networks. *Ad Hoc Sens. Wirel. Netw.* **2016**, *30*, 145–161.

Octopus: A Design Methodology for Motion Capture Wearables

Javier Marin [1] , Teresa Blanco [2,3] and Jose J. Marin [1,3,*]

[1] IDERGO (Research and Development in Ergonomics) Research Group, I3A (Aragon Institute of Engineering Research), University of Zaragoza, C/Mariano Esquillor s/n, 50018 Zaragoza, Spain; 647473@unizar.es

[2] HOWLab (Human Openware Research Lab) Research Group, I3A (Aragon Institute of Engineering Research), University of Zaragoza, C/Mariano Esquillor s/n, 50018 Zaragoza, Spain; tblanco@unizar.es

[3] Department of Design and Manufacturing Engineering, University of Zaragoza, C/María de Luna, 3, 50018 Zaragoza, Spain

* Correspondence: jjmarin@unizar.es; Tel.: +34-976-841-978

Abstract: Human motion capture (MoCap) is widely recognised for its usefulness and application in different fields, such as health, sports, and leisure; therefore, its inclusion in current wearables (MoCap-wearables) is increasing, and it may be very useful in a context of intelligent objects interconnected with each other and to the cloud in the Internet of Things (IoT). However, capturing human movement adequately requires addressing difficult-to-satisfy requirements, which means that the applications that are possible with this technology are held back by a series of accessibility barriers, some technological and some regarding usability. To overcome these barriers and generate products with greater wearability that are more efficient and accessible, factors are compiled through a review of publications and market research. The result of this analysis is a design methodology called Octopus, which ranks these factors and schematises them. Octopus provides a tool that can help define design requirements for multidisciplinary teams, generating a common framework and offering a new method of communication between them.

Keywords: design methodology; design requirements; wearables; MoCap; body positioning; body attachment; IMU; rigid bodies

1. Introduction

Movement is healthy, but whether from injury, illness, or ageing, we are all exposed to losing motor skills at some stage in our lives. In that case, we need to carry out a process of re-education, training, or rehabilitation that teaches us to move or exercise in a certain way. In this sense, motion capture (MoCap) is an opportunity that provides information, both directly (range of mobility of a joint) and indirectly (habits, physical inactivity, etc.) [1,2].

The exponential rate of technology development, and its extensive access through platforms, such as Arduino or Raspberry-Pi, has fostered a world of interconnected objects and connection to the cloud and the Internet of Things (IoT) [3]. In this context, existing objects take on new features, and new ones are inspired by the technology itself. In this way, the elements that we place in the body undergo this evolution, giving rise to a new generation of products, the wearables, whose appearance can be considered at the conceptual level as well as the user acceptance level, as one of the small revolutions within the IoT.

Based on the criteria proposed by Knight et al. [4], wearables can be defined as devices that allow user interaction and user data collection, while remaining attached to the body, regardless of the body's activity and without muscular effort required to hold them. Additionally, wearability describes the

interaction between the human body and the wearable device, and dynamic wearability includes the movement of the human body in the design [5].

For this text, the devices intended to capture human movement that meet the above-mentioned characteristics of wearability and dynamic wearability concepts will be referred to as motion capture wearables (MoCap-wearables). Specifically, we refer to MoCap systems based on the rigid bodies, concept defined by Skogstad et al. [6] as objects that will not deform and will simulate or monitor a body segment.

As shown in Figure 1, MoCap systems that are based on rigid bodies correspond to clusters of reflective spherical markers (Figure 1a) that can be univocally identified by infrared light emitting cameras [6,7], providing three rotations and three positions (one for each space axis). They can also be electronic devices—inertial measurement units (IMUs, Figure 1b)—that provide rotations (rotation matrix, euler angles, quaternions, etc.) through signal processing of the output data of different built-in sensors accelerometers, gyroscopes, and magnetometers) [8–10]. Furthermore by processing from the IMU acceleration data, more information can be extracted, such as speed and position [11,12] or even the moment of reaction on the ground [13]. For MoCap, the concept of a rigid body, in which we focus, matches for both cases, to associate a rigid body to a body segment.

| (a) | (b) |
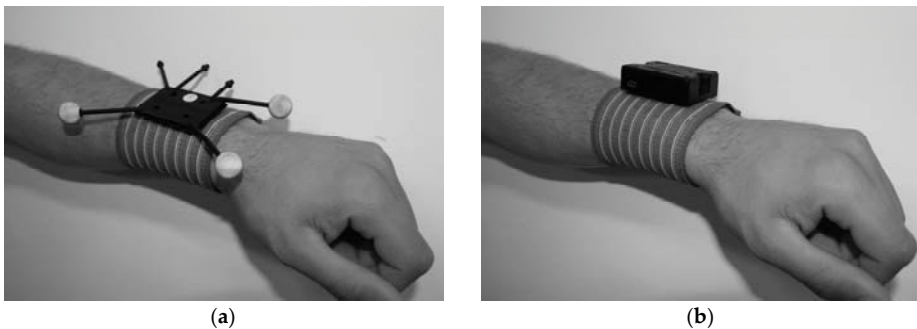
Figure 1. Motion capture (MoCap) rigid bodies: (**a**) optical rigid body [14]; (**b**) IMU rigid body [15].

Currently, MoCap applications are mainly restricted to research, medical rehabilitation, sports training, augmented reality systems, and 3D animation [7,16]. Unfortunately, the large numbers of applications that are possible with this technology are held back by many barriers related to various factors. Overcoming such barriers, we could improve existing applications through more usable, democratised, and higher quality technology. Alternatively, we can generate innovative methods of using MoCap through disruptive innovations based on its application in new fields.

There are barriers associated with errors of the technology itself (magnetic field disturbances, gyroscope drift, etc.) that have had a clear effect on the effectiveness of MoCap systems and therefore on the trust of the technology associated with them [17]. However, due to advances in technology, we observe that the errors directly linked to it are progressively diminishing, either through new or better signal processing or hardware upgrades [10,18–21].

Nevertheless, there are other barriers that are currently unavoidable, such as difficulty in measuring the skin, muscles, and soft tissue movements around the bones. This is one of the most problematic error sources in MoCap systems that use optical or IMU surface markers [20,22]. According to Capozzo et al. [23], isolated markers located directly on the skin at specific anatomical points (landmarks) may undergo relative displacements to the underlying bone in the range of 10–30 mm, which causes an error accumulation in the segment angle. Thus, it should be assumed that MoCap systems with surface markers do not represent bone movement [24].

On this basis, one of the main barrier in terms of error to overcome in MoCap applications is reproducibility [22,25]. Therefore, the accuracy of the body segment angle measurement does not matter, as the same results are obtained using a certain MoCap system with the same subject on different days with different operators in independent laboratories.

The reproducibility factor in a MoCap system is conditioned by the attachment of the markers or sensors to the body [20]. If the union is not constant, stable, and rigid, the measurement quality may worsen through the existence of relative movements between the devices and body [25–27]. Various authors have referred to errors derived for this reason in their research [6,8,11,28]. Although some authors consider the position of the IMUs in the body [29,30], positioning continues to be performed only by a visual check, by trying to align sensor axes with the segment [27,28,31]. Despite being decisive, union elements between the electronic device (IMU) and the body are usually one of the last issues considered in the design, when this should be a matter considered during product design.

We are therefore facing a design challenge, a problem that has been asserted by multiple authors and that transcends reproducibility and measurement accuracy. In this way, to offer products that expand the current sectorial limits (health, sports, 3D animation, etc.), factors such as body attachment, usability, device comfort [9], or accessible costs among others, should be considered, as these are the keys to an optimum MoCap-wearable design.

In view of these problems, this article collects design principles related to the physical aspects of MoCap-wearable systems. The challenge is to identify all those factors that affect the creation process to achieve a MoCap system design that includes both the most purely technical requirements and those that are more focused on the users and the environments. Collected factors are ranked and schematised in a design methodology for requirement definitions called Octopus. The aim of this tool is to facilitate and guide designers and other professionals involved in the development by generating knowledge that allows precisely considering, without omission, all factors from the initial stages of development. The tool is expected to improve existing applications or support the creation of new ones.

2. Materials and Methods

In the wake of the design needs detected, considering the high complexity of MoCap-wearable systems, this section presents a series of factors that are indispensable for study in the design process. These factors have been grouped by theme into five sub-sections: context (Section 2.1), technology (Section 2.2), body attachment (Section 2.3), physical properties (Section 2.4), and user interaction (Section 2.5).

A two-line research methodology has been carried out to identify the design factors. On one hand, a literature review has been conducted regarding wearables and MoCap, these articles are cited to describe the factors, and, on the other hand, an extensive study of the products currently available in the market has been done. Regarding this last point, Table 1 shows a selection of those studied products that will facilitate illustrating the factors proposed in this article since they have been considered representative of the different fields of application and/or provide innovative or unique solutions. Both full body MoCap systems and wearables products have interesting and prominent characteristics that have served as reference to find some of the following factors. Although not all selected wearable products are designed for MoCap purposes, we have used them as inspirational products, extrapolating the characteristics to the MoCap field.

For terminology in this document, *devices* will be understood as MoCap elements to be placed on the body and data processing points (DPPs) will be understood as elements to which they can be connected, such as computers, smartphones, tablets, consoles, etc. DPPs will carry out certain actions with the information collected from the devices. In this way, a set of devices, DPP, and the operation of use constitutes a system, which may refer to commercial systems (wearables or MoCap) or those that are intended to be designed (MoCap-wearables).

Table 1. Product and service market examples.

Full Body Systems MoCap			Wearable Products			
Wireless Inertial Products	Wired Inertial Products [1]	Optical Products	Services	Head	Chest	Extremities
Noraxon [32]	Perception Neuron [33]	Natural point [14]	MySwing [34]	Alex Posture [35]	Araig [36]	LEO Fitness Intelligence [37]
Notch [38]	Rokoko studios [39]	Vicon [40]	Run3D [41]	Google Glass [42]	MC10 [43]	Quell relief [44]
Perception Legacy [45]	Shadow [46]		Imaginarium Studios [47]	Jolt Sensor [48]	SenseOn [49]	Sensoria fitness [50]
Stt-Systems [51]	Technaid [52]			Melon Headband [53]	Tesla Suit [54]	Thalmic Labs [55]
Trivisio [15]	Xsens (suit) [56]			Reebok checklight [57]	UpRight [58]	
Xsens [56]				Thync [59]		

[1] Wired sensor to sensor products, but wireless communication with data processing point (DPP).

2.1. Contextual Factors

Contextual awareness addresses the study of scenarios in which the product is used [60] and allows developers to focus design decisions on the user's world, achieving greater acceptance and system usability. The contextual characteristics will translate into requirements and design opportunities, which allow cross-functional decisions that affect other factors. This section includes factors related to use, the user, and the environment of the product to be designed.

Environmental studies can provide information about available resources, protocols of use, level of hygiene required, etc., while user studies may allow considering both explicitly as latent needs related to psychological, physical, behavioural, or formative aspects.

Generally, and specifically in the case of MoCap-wearables systems, the environment defines the user. Table 2 lists the fields of existing applications—in line with Ahmad et al. [16] and Mayagoitia et al. [7]—and the users involved in each of them. Differentiation is made between two user profiles that can be identified in a MoCap-wearable system: the professional user, which is the person interested in obtaining the data and usually is the product purchaser or prescriber, and the actor user, which is the person captured by the system who may be interested in obtaining information. We assume that the concept of user profile may include one or more individuals as appropriate.

Table 2. Link between some MoCap environments and their users.

	Environment	Professional User	Actor User
Medicine	Diagnosis	Doctor	Patient
	Rehabilitation	Doctor, Physiotherapist	Patient
	Forensic	Forensic Doctor	Injured (may be uncooperative)
Sports	Performance	Coach, Physiotherapist	Athlete
	Rehabilitation	Coach, Physiotherapist	Athlete
Animation Simulation	Professional simulation	Coach, Technician, Others	Athlete, Military, Others
	Video game	Player	Player
	Cinema/theatre	Director, Technician, Others	Performer
Research	Laboratory	Developer, Researcher	Unknown

As seen in Table 2, in the medical and sports fields, the professional user may be a doctor, physiotherapist, or coach and the actor user a patient or athlete. In this area, patients are usually collaborators, except in the forensic environment, where expectations of a possible financial compensation for the damage or injury caused by an accident or other cause may call into question such collaboration. In 3D animation and simulation, user types depend on each application; for example, in video games, it is singular that the profiles of the professional and actor coincide in the same person:

the player. Finally, in the research sector, professional users are researchers or developers who are interested in obtaining information or implementing new applications.

It should be noted that, according to the mode of application or use of the system, the resulting niche can lead not only to a new commercial product development but also to a service or a product-service operation. Some manufacturers are already marketing MoCap systems as a service. This is the case of MySwing [34], where the customer has a personalised learning and analysis program of his or her golf practice. Likewise, a MoCap professional studio for the film industry or a gait analysis laboratory is intrinsically a service [41,47]. Figure 2 illustrates a real MoCap service with both described users involved.

Figure 2. MoCap service, biomechanics laboratory. System with simultaneous optical and IMU technology (optical full body MoCap and IMU upper body MoCap).

It is important to note that if a MoCap-wearable design is associated with a service, the system profitability and its purchase interest may depend on the developer or customer ability to perform an adequate servitisation of the product. In addition, the training, motivation, and resources available to the professional will be keys for the service to be adequate. In addition, the privacy and the required confidentiality level in each of the usage scenarios must be considered [60,61].

Therefore, assuming the business model as a service, defining the operational use from the initial development stages will allow applying specific design techniques to predict failures, extracting critical points and accompanying the physical design with the considerations related to the intangible part of the product.

2.2. User Interaction Factors

Achieving an appropriate interaction between users and the product is relevant to minimise the learning phase and avoid errors during use. In the literature review, it has been detected that, as a general rule, some adjectives that define a good user interaction with wearable devices are simplicity, subtlety, transparency, and intuitiveness [5,60].

The product-service system can interact with both the professional user and the actor user. Therefore, the information to be transmitted and its representation must be adequate to each profile mental model and to the context at the time of receiving the information [62]. Therefore, the user characterisation will greatly influence the interaction design, which will be a cross-functional theme for all product factors.

To classify the communication interfaces between users and technology, the two parts of the MoCap-wearable system can be used: devices and DPP. In Figure 3, the communication interfaces are illustrated: users with device and users with DPP. Note that there is also interaction between both types of users, but the definition of such interaction depends on the service design already discussed in Section 2.1.
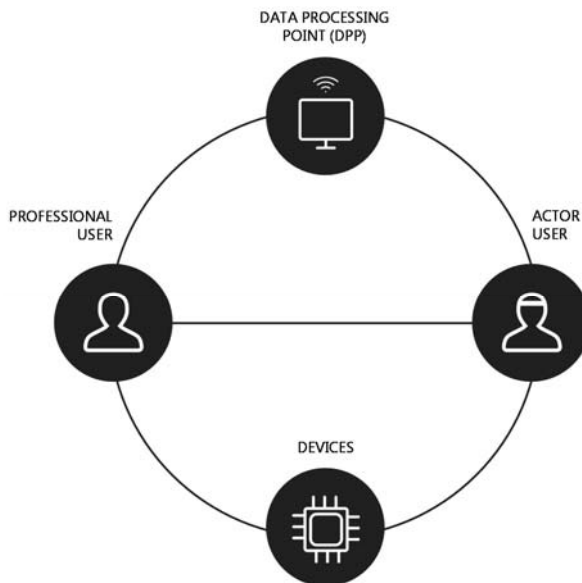

Figure 3. Communication interfaces scheme in a MoCap service. (Icons designed by Freepik and Alfredo Hernandez, from www.flaticon.com).

2.2.1. User Interface with MoCap-Wearable Device

MoCap-wearable devices require a user interface that should allow, facilitate, and optimise product usage, helping its cognitive interpretation. In the reviewed devices (Table 1), there are physical interaction components (power button, load connector, and light-emitting diodes (LEDs) to show the battery level and/or wireless connection status), and graphical components (X, Y, and Z axes on the surface to the device orientation in the body, and body segment labels). These interaction elements are generally aimed to the professional user; however, device designs should consider and take advantage of communication opportunities with the two types of users involved.

Through the interaction elements or other elements that can be incorporated in a bidirectional interaction process, the devices will be able to receive input stimuli (subject movements, change of configuration, etc.) and respond by feedback or output to the users. Such feedback can serve to communicate with the actor through biofeedback [63] (guide, correct movements, etc.) or to reach agreement with the professional (improper device positioning, wrong movements, low battery, etc.). The feedback types can be classified by human senses involved in communication:

- *Visual feedback*: Interaction with LEDs, images, or text. From the perspective of the user actor, an interesting point of interaction may be the upper face of the wrist, inspired by wearable wristwatch style designs or visible areas of the body, such as legs or arms.
- *Aural feedback*: Sounds, beeps, instructions, etc.
- *Haptic feedback*: Communication through non-visual and non-auditory sensations using vibration, temperature, or electrical impulses, which can be observed in some products, such as the Notch motion sensor [38], Araig jacket [36], or the Tesla suit [54].

2.2.2. User Interface with DPP

The DPP can also interact with both users at both hardware and software levels. From the hardware point of view, the most common design process is to select a smart device, such as a computer, smartphone, tablet, game console, etc. This can be recognised in most analysed products. Generally, in MoCap systems, a computer is employed, and in wearables, smartphone use is more frequent. The information input type (touch, voice, etc.) and the output characteristics (visual, sound, etc.) will depend on the selected DPP features to which the designer must adapt.

With this, the problem can be reduced to a software design issue and, from the standpoint of interaction, can be reduced to the screen interface or graphic user interface (GUI) design. Additionally, the software may be able to be supported in the cloud or even only operate on it; thus, the Internet communication will serve to store and download information or communicate with other systems.

2.3. Technological Factors

Technology, its possible configurations, and the main characteristics of the electronic components influence the device external parts that are in contact with the user [5]. Consequently, the technological requirement study includes the selection between optical or inertial systems, the electronic component structure that incorporates each capture device, and the possible configurations of interconnection between them.

2.3.1. Choice between Optical and Inertial MoCap Technology

The choice between inertial or optical MoCap technology depends on the application requirements, which should be contrasted with the characteristics of each technology features. In this paper, we focus on these products as they are the most common; however, it should be noted that these are not the only technologies used in MoCap systems [64].

Optical systems have a consolidated development and provide proven accuracy. However, they require a controlled space and environment, especially about lighting (preferably artificial), and a variable number of cameras that will depend on the area of the body to be monitored and the number of

simultaneous actors to capture. Normally, the technology is based on vision cameras that emit infrared light, which are capable of recording and processing the movement of reflective spherical markers placed at landmarks [6,7]. Its accuracy is high—on the order of 1% on the measurement taken [65] or 1 mm [22]. It is necessary to emphasise the problems derived from occlusions of the reflective markers by other body parts, other actors, or objects or devices that the actor handles or in the scene itself that may require a high number of cameras or important post-process work to recreate the actor's movement.

Moreover, inertial systems are generally more economical and require less infrastructure [7]. They can be used either in real time by a DPP signal-receiving device or autonomously with an internal memory storing the information. In contrast, integration errors and the presence of magnetic fields may reduce accuracy [8,10]. For MoCap commercial systems (Table 1), the IMU accuracy offered is in the following intervals: $0.2°$–$1°$ (roll/pitch), $0.4°$–$2°$ (yaw), and dynamic root mean square (RMS) $1°$–$2°$ RMS (roll/pitch/yaw).

2.3.2. Electronic Components (Building Blocks)

It is considered necessary to overcome the strict separation between design engineer and electronic engineer. Knowledge of electronics by the designer allows creation of more viable projects and, alternatively, generates more design opportunities due to the vision of possible technological options. It is true that, for an external device design, it is not necessary to know all the components in depth, but in line with [66], it is interesting to promote designers with minimal technological literacy that are aware and can collaborate in defining the main building blocks. This definition of blocks depends on the usage scenarios (context), especially regarding the times of use, the speed of movements, or the real-time capture needs.

The building blocks can be considered from the device, the DPP, or even the complete system point of view. In Figure 4, we see an example of the main blocks (building blocks) of a wireless IMU device. Under the scheme, the technological factors to consider for a wireless IMU are related to communication, storage, battery, intelligence, and interface.
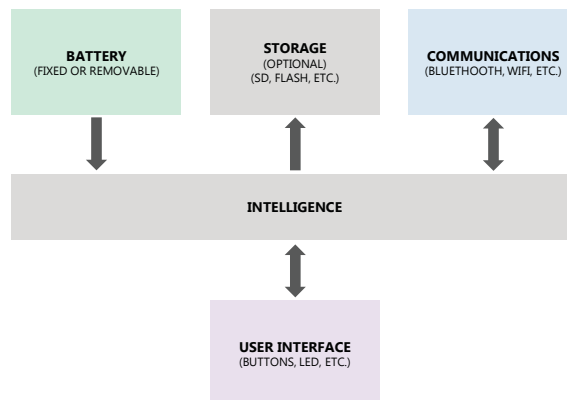
Figure 4. Building blocks example of a MoCap wireless IMU, created with the methodology of Blanco et al. [66].

In the case of revised IMUs for MoCap-based devices (Table 1), the following factors in relation to electronic components can be highlighted:

- *User interface*: Already mentioned in Section 2.2.
- *Battery*: MoCap wireless sensors typically have built-in non-removable batteries, recharging either in a charging socket or directly connecting to each sensor. The battery life (according to manufacturer's information) can vary between 3 h and 8 h. The battery selection has a special

interrelation with the other factors because it depends on the usage scenarios (life required), on the other components' consumption, on the DPP characteristics, and on the interconnection between the different components.

- *Storage*: With the option of including internal storage, it is not required to be in range of a wireless network, which increases versatility. In contrast, in this case, the ability to perform real-time processing is limited. Systems that work with internal storage usually have a secure digital (SD) card, such as Perception Neuron or Stt-Systems [33,51].

- *Communication*: The communication features between devices and the DPP depend on the selected wireless communication protocols and consequently on the selected DPP to process the data. However, sometimes, an external dongle communication is connected to the DPP, which relieves its selection requirements. Communication protocols are typically WiFi 2.4–5 GHz for local area networks (LANs), Bluetooth for personal area networks (PANs), or other proprietary protocols.

- *Intelligence*: All these actions are performed by the devices from raw measurement data and users' actions. In terms of intelligence, the quality of the measurements provided by the IMU, and therefore the restrictions for some applications, depends on the quality in the accelerometers, gyroscopes, and magnetometers and the quality of the signal processing. The IMU measurement ranges vary between ± 2 g and ± 50 g for accelerometers and between $\pm 150\,^\circ$/s and $\pm 1000\,^\circ$/s for gyroscopes [27]; however, the evolution of the technology is improving the quality of these aspects. In signal processing, Kalman filters are widely used [67]. Currently, this field is being improved by different authors. For example, Dejnabadi et al. (2006) and Favre et al. (2009) [19,20] have proposed other magnetic and compensation drift algorithms that have been shown to be effective.

2.3.3. Number of Devices and Interconnection of Them

A key factor in the technology definition is the number of devices be placed on the body areas, each application will require a varying number. The number of sensors influences and depends on the available computational resources and on how much the signal processing can be tailored to the application. In general, for full body MoCap, a total of 15 are required: three per each extremity, another to monitor the head, and two others for the chest. In addition, more sensors can be added in the phalanges or chest. We can see a MoCap application with fewer devices in Macard et al.'s paper [68]. In terms of interconnection and communication between devices, in the studied MoCap systems (Table 1), different possible configurations can be found:

- *Full body suits*: Typically made with Lycra, they contain devices in areas to be monitored. The sensors are wired to a hub placed on the waist or back. The hub communicates wirelessly with the DPP and includes a battery that powers the devices [39,56].

- *Wired independent elements*: These have the same operation and interconnection as the full body suit, but each sensor has an independent fixing support. The cables, sensors, and other devices are in sight [33,46,52].

- *Wireless devices*: Each sensor is placed with an individual fixing support, and each one has its own battery and communicates independently with the DPP [32,34,38,45,51,56].

2.4. Body Attachment Factors

Body attachment is one of the key factors in the MoCap-wearable device design, as it is a cross-functional concept. Thus, it directly affects system accuracy, reproducibility, user comfort, and, consequently, product market acceptance [9,25]. Although it could be analysed as a sub-point of the user study (ergonomics), it is considered that it goes further and is considered an independent issue, comprising the good characterisation of a considerable number of requirements.

Body attachment requires studying two sub-factors: positioning and the attachment method. The positioning refers to where the devices are located, that is, in which body segments and in which

zones within each segment. The attachment method defines how they are joined, that is, what elements are used to attach the devices to the body segments.

2.4.1. Device Positioning

Yang and Li [27] asserted that the effect of sensor location on measurement quality is not usually analysed; however, the collected data from acceleration and angular velocity are different in one location from another for the same body segment. Consequently, devices should be positioned so that they are oriented towards the body of a subject in the same way regardless of condition, context, or activity. Correct positioning should not depend mainly or exclusively on the professional user; in this sense, the product shape and the rules of use play a key role. Some authors have also detected this fact and have attempted to solve it through different software or hardware improvements [29,30,68,69], which is a complementary approach to the one presented here.

From the point of view of accuracy, to select the position of a body segment regarding where to affix the capture device, the potential effects of that place, and the quality of the movement measurement must be considered. In terms of comfort, devices should generally be non-intrusive and consider the body to be a dynamic structure in motion. Areas with relatively the same size should be selected in adults. Moreover, areas with the least movement, friction, and flexibility when the body is in motion should be selected. The anthropometry of the target users must be considered, and how the sections of each segment change according to age, morphology, and weight should be studied [5,26].

Following these rules and analysing the positioning areas of the commercial MoCap devices (Table 1), the most common placement areas for monitoring major segments of the body are shown in Figure 5.

| (a) | (b) | (c) | (d) |
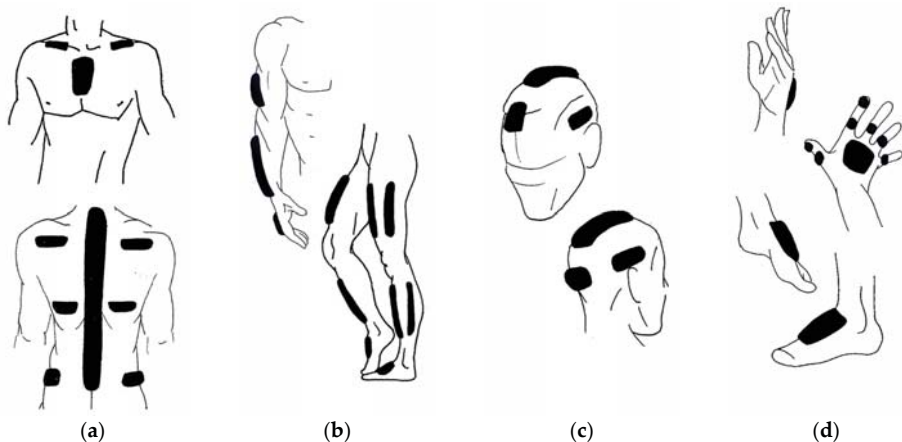
Figure 5. Main zones for positioning devices: (**a**) chest; (**b**) extremities; (**c**) head; (**d**) hands and feet.

In Figure 5a, areas of chest placement are seen. In this multi-segment zone, we can opt for different positioning alternatives: sternum, spine, hips, ribs, shoulder blades, or clavicles. Most systems place one sensor on the sacrum and another on a higher area of the torso, either the anterior or posterior part. To homogenise the stem analysis, Yu et al. [31] studied the movements of the vertebral column to determine the optimal placement of the sensor placed in the upper part. The results showed that the most representative points to monitor displacement of the medial or lateral trunk were the T7-T8 vertebrae. In these parts, the smallest error occurs (0.5°). However, using more sensors in this area may be beneficial, so some systems place a sensor in the mid-spine zone to improve the approximation of its curvature [32], others also add sensors on the sides (clavicles, shoulder blades,

or ribs), which allow monitoring the movements of the shoulders and/or the asymmetrical twists of the body [33,39,45,46,52,56].

Figure 5b shows the placement above the extremities. It should be noted that no consensus has been reached regarding the height at which to place the elements in each segment; nevertheless, there is more agreement about the segment faces used; the outermost zones that do not interfere with the movements are suggested. In the forearm, the upper surface of the wrist is used, just above the joint. In the legs, the frontal or outer zones are used. Sometimes, the same system combines both options, placing elements externally on the thigh and frontally on the shin. A resource that can help positioning the limbs is to take advantage of areas where muscles are inserted into the joint, such as the section between the calf muscle and the knee, which provides a curvature that prevents the device from slipping.

Figure 5c shows areas of head placement, with the most common placement being the front and the back of the head, and sometimes the sides or over the top. It is interesting how the head wearables show other positioning resources, using the natural supports that provide the upper part of the ears and the nasal septum [35,42,53,59].

Additionally, Figure 5d depicts areas where devices are placed on the hands and feet, generally on the top of these body parts. These zones provide a relatively extensive and flat surface without movement impairments. When the fingers are monitored, sensors are placed on each of the proximal phalanges (first segment) and on the distal phalanx (last segment) of the thumb and forefinger.

In relation to the above and from a contextual point of view, a factor related to the positioning that is frequently not considered with the necessary detail is the device interaction with the garments. It is necessary to study what kind of clothes users will wear, the possibility of changing their clothes, and how it will affect their comfort and emotions as well as the intimacy and privacy feeling that may influence the individual acceptance and behaviour. For some applications, it may be necessary to design clothes that are compatible with the devices, either integrating the sensors in them or leaving space for the sensors to be positioned in the appropriate way, all respecting the requirements of hygiene and healthiness.

2.4.2. Device Attachment Methods

Multiple authors have stated the importance of using an appropriate attachment method; this is because union is a key factor in all wearables, but even so more in those that monitor movements [6,8,9,11,25–28]. In a research environment, the professional can use ad-hoc solutions (adhesive tape, dressings, medical plaster, etc.) that can ensure precise positioning and individual adjustment for each subject. However, in commercial applications, a proposal is required that ensures a constant, stable, rigid, comfortable attachment with minimal preparation requirements [9].

In the MoCap-wearable device body attachment design, it is advisable to try to keep the device as close as possible to the body and bones. In general, it is preferable to secure the device by completely or partially surrounding the body region involved, rather than using single-point fastening systems. Furthermore, the long-term effects of carrying the device should be considered, and its effects on the user from the psychological (comfort) and physical (sweating, tiredness, etc.) points should be analysed. In addition, the diversity of body size should be considered, allowing a certain size and shape customisation [5,60]. Thus, from the reviewed products (Table 1), the following types of body attachment or fixing supports can be distinguished:

- *Fabric fixing supports*: The union with fabric is made from different widths of bands or tape or tight garments. The elastic tape can be closed or open, and the latter will be closed with Velcro, clips, or buckles. The fabric characteristics will influence the union accuracy, perspiration, comfort, and wear resistance. The connection between the fabric and the device can be made with Velcro, pockets, dedicated housing (plastic base), or metal pressure clips that are used by some wearables [38,50]. In addition, it may be beneficial to use wide Velcro areas above the fabric, which allow certain variability at the point of attachment to suit each subject.

- *Disposable adhesive fixing supports*: This support type groups different types of unions: hypoallergenic double-sided adhesive, which is economical although relatively weak, bandage or kinesio-tape, which has high adhesion but with some preparation time required, and disposable electrodes, which are used in some wearables [43,49,58]. The latter can be standard or manufactured specifically for the product, being able to use one or several connection points and allowing measurement of biometric signals. Note that, if we choose disposable adhesive fixing supports, although hygiene is maximum, the cost is higher due to the material expenditure. In addition, body hair and sweating will significantly worsen the adhesion.

- *Semi-rigid fixing supports*: This union type is mainly found in wearable devices that, due to the flexibility of some of the parts, the product stands by itself, wrapping around the body or the garments that the subject wears. This solution is observed in the Thalmic Labs product [55] that has elastic zones to fit the arm, surrounding it as a bracelet, the Jolt Sensor wearable [48] that can be attached to clothing with flexible flaps, or others such as Alex Posture, Google Glass, Melon Headband, or Thync [35,42,53,59] that take advantage of their elasticity to hold onto the head as if they were hair headbands. Although it is a method that does not provide a strong union as others and it may be difficult to apply in all body segments, it must be considered to solve some problems, such as fungible expenditures or hygiene.

It should be noted that the different attachment methods can be combined. In fact, some wearables that use tape to surround body segments combine it with non-slip surfaces [37] or disposable electrodes that adhere the tape to the skin [44].

From these types of bindings, in Table 3, a weighting method for choosing the ideal type in a given application is proposed. It collects the body unions identified in the products and articles reviewed as well as a selection of seven factors that characterise each of them. Assuming that there is no better method of fixing than another, the selection will depend on the weight or value given to each of the seven factors, considering the particularities of each application and context.

The score of each factor varies from 1 to 3, with 1 being the worst and 3 being the best. The 'Result' column can be completed by applying Equation (1), whose highest value would correspond to the method of union most recommended for a given application and context:

$$R_j = \sum P_{ij} \times W_i \qquad (1)$$

In Equation (1), *R* is the result obtained by each binding method, *P* is the score assigned to each element of the matrix (1, 2, 3), *W* is the weight assigned to each factor (from 1 to 10), while *j* is the rows and *i* is the columns.

In addition to Table 3 and the attachment methods described, we note that as device fixing support, elements that are already carried by the user in the application environment are a resource to consider. This is the case for some wearables that use everyday items, such as glasses, swimming caps, or jackets [36,42,57]. The use of these elements contains known interaction patterns, which is beneficial for all types of interaction, but it makes special sense when referring to the body attachment factor since it can favour technology learning and confidence. Thus, new union concepts require greater learning time and create a new interaction language; however, this does not mean a worse alternative, as long as the designer considers and plans how to transmit it and teach it to the users.

The most commonly used attachment methods in the reviewed products vary according to the anatomical area involved: on the torso (pelvis and thorax), disposable adhesives, harnesses, or adjustable and elastic tape is used; on extremities, elastic tape or disposable adhesive is used; on the head, hair headbands, elastic tape, helmets, hats, or semi-rigid elements that rest in the anatomical references, such as the ears, nasal septum, front, or nape, are used; on the hands, tape or gloves that surround the thumb and/or fingers to prevent the device slipping are used; and on the feet, disposable adhesive or elements attached to parts of the shoe (cords, tongue, or sole) are used.

Table 3. MoCap-wearable device body attachment methods valued by factors (1—worst, 3—best).

Attachment Method	Area Selection	Preparation Speed	Washing	Adapt-Ability	Fungible Restrictions	Union Distribution	Union Strength	Result (R)
Weight (W)	(5) [1]	(4) [1]	(9) [1]	(10) [1]	(2) [1]	(8) [1]	(8) [1]	-
Closed elastic tape [2]	3	1	1	2	3	3	3	(102) [1]
Open elastic tape [2]	3	2	1	3	3	3	3	(116) [1]
Garment (fixed pockets, clips) [2]	1	2	1	1	3	3	3	(86) [1]
Garment (Velcro areas) [2]	3	2	1	1	3	3	3	(96) [1]
Commercial electrode [3]	3	3	3	3	1	1	2	(110) [1]
Custom electrode [3]	3	3	3	3	1	2	2	(118) [1]
Double-sided tape [3]	3	2	3	3	1	1	2	(106) [1]
Bandage kinesiotape [3]	3	1	3	3	1	3	3	(126) [1]
Semi-rigid (bracelet, flaps) [4]	2	3	3	2	3	3	1	(107) [1]

[1] Example of scoring for a sanitary elderly rehabilitation application; [2] Fabric fixing supports; [3] Disposable adhesive fixing supports; [4] Semi-rigid fixing supports.

2.5. Physical Property Factors

The physical properties of the device, such as the shape, dimensions, or weight distribution, affect the user both physically and mentally. Therefore, the needs and restrictions of both user profiles in contact with the product are appreciated even more in this section. These factors have effects on the actor user in the aspects of energy expenditure, biomechanics, posture, movements, and perceived comfort [4]. From the professional user's point of view, they have effects on comfort, acceptance, use experience, and perceived accuracy.

2.5.1. Shape

The shape is one of the key product elements since it must be intimately linked to its function. The MoCap devices must have forms that facilitate positioning, which helps the professional find the proper fixation area and get and maintain its correct orientation, which are requirements that will vary according to each body part. Furthermore, a smooth and subtle transition must be ensured from the body surface to the device; this can be achieved with a concavity in the inner surface in contact with the body and with convexity in the outer surface to avoid blows and hooks [5]. The studied wearable products (Table 1) mostly follow these guidelines; however, the MoCap IMU devices do not. The latter are generally a rectangular box with slightly rounded corners or other polygonal shapes.

2.5.2. Dimensions

The device dimensions should be adequate to the areas to be monitored and should suit different morphologies, considering the diversity of body sizes [60]. Although the surface they occupy in each body segment is not a critical factor, the thickness is. According to Gemperle et al. [5] there is an intimate space or aura of 0 to 127 mm around the body for the devices, and as a rule, one should try to minimise the thickness as much as possible so that it feels like a part of the body.

2.5.3. Weight

The weight should be distributed so that the maximum load is placed near the body centre of mass [5,60]. In this way, if we have several elements to place, heavier ones should be in the torso, and the lighter ones in the extremities.

2.5.4. Flexibility

It is necessary to consider the flexibility of the MoCap-wearable device components. In this sense, the possibility of integrating flexible electronic-printed circuit boards would be the optimal solution for body adaptation, but this would possibly increase technical complexity and cost. Another option is to use rigid areas coupled with flexible areas. Thus, the flexible areas would be between the solid forms, extending it like wings [5]. This is shown in wearable market products, which integrate rigid areas combined with flexible ones in the same structure [35,43,48,55,59]. The MoCap devices are composed by the IMU zone, are completely rigid, and are near the fixing support area, which is usually more flexible and adaptable. The Shadow MoCap system [46] differs from the others because its sensors are embedded in fabric pads (neoprene type), which makes the entire system flexible.

2.5.5. Material

The material selection depends largely on the context of use, which determines different aspects to consider, such as weather resistance agents, abrasion, impacts, temperature, humidity, required maintenance, washing, hygiene, comfort, and breathability [5,26,60]. The most commonly used material in MoCap device housings is plastic. In wearables, however, we find other materials such as silicone or intelligent fabrics. It is important to note that the use of metallic (ferromagnetic) materials, that alter the magnetic field close to the IMU sensors can disorient its magnetometers, so some manufacturers use a shielded packaging [33].

2.5.6. Comfort

Comfort is not a tangible property by itself, but it depends on the physical aspects, and in MoCap-wearable device design, it is necessary to prioritise this factor [5,60,61] because it increases system usability and acceptance. According to Knight et al. [70], wearable comfort depends on several dimensions or factors: first, from the body attachment, the movement restriction, and the user concern about how the device moves in his or her body. Subsequently, the pain (itching, burning, pricking, heat, etc.), the perceived change (feeling physically different, strange, or uncoordinated), the anxiety that can be caused by the system, and the emotions that the device causes when it is worn; which are issues that link to the next point.

2.5.7. Psychological Aspects

The study of the emotional and psychological aspects that people have towards the products helps deepen understanding of the user and thus should also be present in the design process. According to Spagnolli et al. [61], the users should be characterised according to a series of psychological factors, among which we highlight the attitude towards technology, the perception about the product usefulness, the expected learning effort, the social influence for purchase, and the expected pleasure perception during use.

2.5.8. Aesthetics

In addition, in terms of user comfort, aesthetics can be considered one of the most relevant aspects in the MoCap-wearable device design [5,61]. Aesthetics is a more important issue than it may seem, even more so than in other technological products, because people consciously or unconsciously express themselves with the garments and objects worn, as they define us from the social and relational point of view [71]. In this sense, it may be beneficial to consider the user preferences, interests, and desires, and allow a certain personalisation, as Motti and Caine proposed [55]. Nevertheless, the options at this point are endless depending on the new applications.

3. Discussion

MoCap systems have been used for more than two decades, and their usefulness and application in various fields such as health, sports, or leisure are widely recognised; therefore, this is increasing its inclusion in current wearables. However, to adequately capture human motion requires addressing requirements that are difficult to satisfy. Among other issues to consider, devices must maintain their body position independently of the subject movements and not be invasive in order to facilitate natural body movements. With all this and because of its prohibitive costs in the past, MoCap expansion has slowed and unfortunately has been restricted to laboratories or specialised studies.

Today's society already has broad access to technology (cost reduction, free software development platforms, etc.). Therefore, in the context of interconnected objects (IoT) and constant technological development, it is also necessary to have greater access to the complete process of technology creation (design guidelines, methodologies, etc.). Thus, having tools that help the MoCap-wearables systems development process can introduce an entire range of possibilities and allow extending them to new areas where they are not present or not widespread, solving real problems of users and society.

Faced with this problem, a study of the factors that are critical to MoCap-wearable system designs has been conducted. Neglecting any of them may involve creating unprofitable products either because they are not viable or have low effectiveness or low acceptance. In this discussion, a specific design methodology called Octopus is proposed, which aims to manage all these factors and facilitate the requirement definitions of developers, designers, and multidisciplinary design teams. The Octopus is a metaphor for how the devices must be attached to the body, using the eight steps established below. Its scheme is shown in Figure 6.
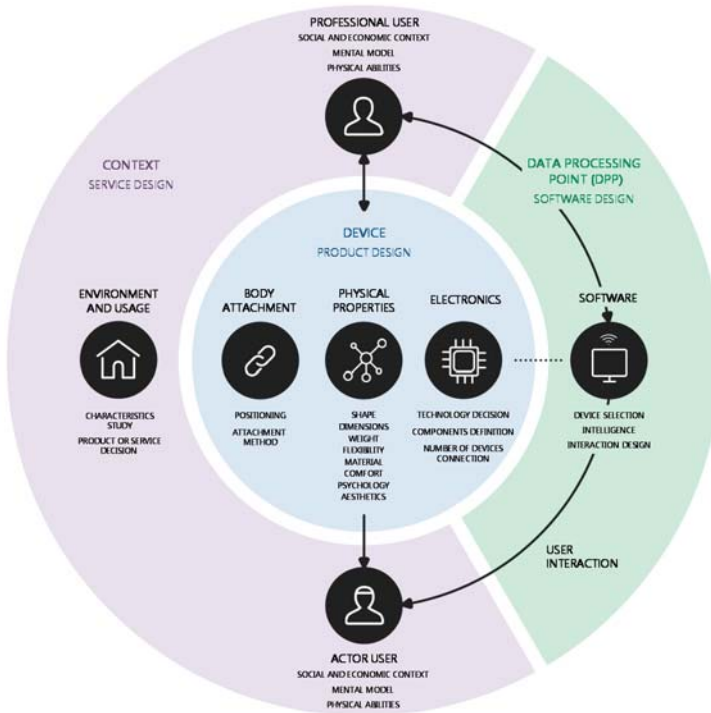

Figure 6. Octopus methodology for MoCap-wearable system designing. (Icons designed by Freepik, recep-kutuk, madebyoliver, gregor-cresnar, EleanorWang, cursor-creative, from www.flaticon.com).

The Octopus methodology is the result of ad-hoc research and analysis, of the different disciplinary author's views (mechanical engineering, biomedical, and design), and of the experience in MoCap systems use in the research group, which has conferred a lead user vision that, as Lilien et al. [72] indicated, benefits the design process.

In a general human-centred design methodology, the following iterative studies or stages can be identified: planning, context of use, requirements, design, and evaluation. For each, there are distinctive design methods [73]. The Octopus methodology approaches the first three steps: plan, contextualise, and extract requirements—in this case, the requirements of MoCap-wearable systems.

The scheme in Figure 6 is a representation of the MoCap-wearable product/service and its ecosystem. It is composed of three zones (context, device, and DPP), each of which houses a series of elements directly related to the factors described in the second section (materials and methods). The Octopus methodology proposes to approach the design sequentially, according to the following steps: (1) design goal; (2) context study; (3) service design; (4) user interaction; (5) technology; (6) body attachment; (7) physical properties; and (8) DPP. Although it is linearly represented in principle, it will usually be necessary to perform successive iterations and made a sequence accommodation to each situation.

3.1. Step 1: Design Goal

The initial methodology hypothesis starts from the design goal definition. Obviously, the design process evolution—even more in iterative orientations—can slightly or substantially modify this initial objective, especially when one of the main aspirations of the present methodology is linked to the latent need to search for new applications, a necessity for MoCap and wearable evolution. As mentioned, the methodology is flexible enough to adapt to the project imponderables, in line with the iterative design methods, returning to the needed point and taking advantage of the experience gained. In this sense, establishing an abstract character of the design goal will depend on the project philosophy. That is, the challenge can be placed in a concrete improvement (e.g., creating a MoCap system for older people for rehabilitation in their homes) or placing it in more intangible horizons (e.g., look for application fields using new materials). The goal level of abstraction will also influence the starting point, the number of iterations required, and the linear character of the process.

3.2. Step 2: Context Study

Once the objective has been established, the next stage is context characterisation, definition and modelling of user types, and use of the product (use, user, and environment analysis). Figure 6 includes the two most common user profiles in MoCap, the professional user and actor user; nevertheless, as many user profiles or sub-types as necessary may be added. Observational methods, interviews, focus groups, and role play among others can be used for this study [73]. Due to these, context characteristics, opportunities, and problems as well as user qualities, capacities, behaviour models, and desires can be extracted. To synthesise the information collected, some tools such as the 'person-method', user targets, or archetypes can be applied [74]. Moreover, in terms of use, developers can establish analogies with related products to approach the user's base knowledge.

3.3. Step 3: Service Design

This step will be included if the product is considered a service, a growing area for MoCap; in this case, service design tools [75,76] would be an indispensable basis for the project. One of the most interesting methods is 'Blueprint' [77], a design tool and dialogue and training media for professional users, because it marks the necessary key moments in the service for each of the users involved through a temporary schematisation.

In a MoCap-wearable service, the key moments or actions that are usually carried out are the following: (1) explanation to the actor user; (2) device body placement; (3) anatomical calibration; (4) capture; (5) data and result generation; (6) data processing (by software and/or by user); and (7) final

actions (e.g., medical reports, real-time feedback, etc.). This generic use protocol must be considered in the physical device design and if applicable, in the service definition. It will influence both users experience and, consequently, the final solution success.

3.4. Step 4: User Interaction

User interaction flows are defined in the Figure 6 diagram by arrows for all items that relate to users. These interaction lines can be unidirectional (feedback or biofeedback) or bidirectional (information exchange), which may vary depending on the device, the user, and the context.

In the early design iterations, these lines can be used to define some needs and specifications, based on the work already done; thus, a relational needs methodology can be followed. This tool is applied in cases like ours, where it is intended to design solutions that involve relationships between several users to detect those needs that each specific user has with respect to the others [78]. Here, it can be proposed to apply in the rest of the items, contemplating the human-machine and machine-machine relations.

With respect to human-machine interactions, as detailed in Section 2.2, the system interaction with the professional user can be bidirectional, from devices and from DPP, including interaction with devices regarding proper placement, maintenance, or connection and with DPP regarding capture configuration of possible incidents, annotation, or result study, analysis, and interpretation.

The human-machine interaction with the actor user are typically unidirectional, from the device and/or the DPP; it can be either as feedback or as biofeedback [63] and in a visual, auditory, or haptic way, showing different events like movement changes on a screen with a virtual scene, or the start and end sounds, among others. In design, it will be important to consider the elements that are necessary to allow and favour the mentioned user interaction.

3.5. Step 5: Technology

It is necessary to study the technological content, as it affects the physical product conception. To do this, it is proposed to approach it through some points, which, despite being presented sequentially, will require iterations due to their strong interrelation between them and with the rest of the sections:

- Make decisions about the technology type to be used (described in Section 2.3.1): optical, inertial, or even other MoCap technologies depending on the application. Thus, we can find solutions such as the one proposed by Shiratori et al. [64], which uses cameras fixed on the body.
- Define the main electronics that are needed, for which the electronic building blocks can be made and apply electronic design methodologies focused on designers or multidisciplinary teams, as Blanco et al. proposed [66]. Defining the building blocks (Figure 4) is necessary to ensure product viability, and to anticipate factors like the space that can occupy the electronics or the associated requirements and restrictions; for example, for IMU design, zones free of ferromagnetic materials and a minimum distance from the human body are required to not alter the magnetic fields and facilitate radiofrequency communications. The main blocks in the case of MoCap-wearables devices would be (Section 2.3.2) communication (dashed line of the schematic), storage, battery, intelligence, and interface (defined in Step 4). Note that, to be able to define some points of the electronics, it will be necessary to at least have selected the DPP that will process and register the data (Step 8) because the DPP will also influence the communications, functionalities, size of components, etc.
- Select the required number of MoCap-wearable devices to be placed on the body and the most appropriate interconnection between them (Section 2.3.3): full body suits, wired independent elements, or wireless devices.

3.6. Step 6: Body Attachment

It is necessary to consider the body attachment, a key factor in the design of wearables. At this point, the positioning and the attachment method must be determined. In relation to the positioning, the exact area must be determined, explaining the reasons for its selection, and investigating and testing the best rules to help the professional user determine its correct location. Regarding the attachment method, the proposed decision table can be used (Table 3), in which more attachment methods can be included as this scope progresses. In any case, prototyping at this stage will be crucial.

3.7. Step 7: Device Physical Properties

It is also required to define the physical MoCap-wearable device properties. For this purpose, criteria are presented in Section 2.5 related to shape, dimensions, weight, flexibility, material, comfort, psychological aspects, and aesthetics, which are crucial factors to achieve the necessary precision measurement for good user adaptation.

3.8. Step 8: DPP

Having selected the DPP product to perform processing, at this stage, it is not necessary to define the interaction elements because they are given by the product. Thus, this stage requires the software design, which, according to the purposes of each MoCap application, will utilise data processing and will communicate the appropriate and accurate information to carry out the user interaction. For this design, software development and interaction methodologies [62] can be followed. The development of this point depends on the system purpose, which, due to Octopus, is expected to extend to other areas not yet explored.

3.9. Case Study, Methodology Assessment

To illustrate the methodology and assess its applicability, a case study is included in Table 4, where the basic requirements of a shoulder rehabilitation service for elderly individuals are defined.

Table 4. Case study, shoulder rehabilitation service for elderly individuals.

Step 1: Design Goal
Shoulder rehabilitation service for elderly in private clinics.
Step 2: Context Study
Professional User Target: Occupation: Physiotherapist. Age: 35. Technology level: Medium—High. Verbatim: 'I am a person who wants to improve and learn every day in my work. I have little time since I attend about six patients a day in 45-min sessions'. **Actor User Target (patient):** Occupation: Retired. Age: 78. Technology level: Low. Verbatim: 'At my age, I appreciate tranquillity and patience; it makes me feel safer'. **Environmental Characteristics:** Indoor, bright, hygienic, and clean. Furniture: work tables, chairs, and stretchers.
Step 3: Service Design
1. Explain the test to the patient. 2. Place the devices on the patient and warm up. 3. Record the target movement cyclically (e.g., flexor-extension or internal-external rotation), while the physiotherapist mobilises the patient's shoulder according to the appropriate rehabilitation schedule. 4. The patient repeats the movement alone while receiving biofeedback from the recorded motion. 5. Report of the results (success or failure regarding the recorded movement cycle).

Table 4. *Cont.*

Step 4: User Interaction	

DPP—Professional Interaction:
Allow to start and pause motion recording (possibility to do it remotely).
Set the number of repetitions of the exercise and grade the threshold to consider whether the movement is correct for rehabilitation purposes in the session.
Observe the results.
DPP—Patient Interaction: Display an avatar that moves in real time according to the patient's movement. Display a superimposed avatar that plays the pre-recorded motions during the professional's mobilisation. Use gamification techniques to facilitate tracking of target movement.
Device—Professional Interaction: During the recording period, the user interacts with the devices through the visual sense: green LED—transmitting and no movement recorded, red LED—recording, and blue LED—there is movement recorded.
Device—Patient Interaction: During the task of repetitive movements, the patient interacts with the haptic sense (vibration) to indicate breaks, repetitive failures, phase changes, etc.

Step 5: Technology

Technology Selection: IMU devices are used. No special lighting or space requirements are necessary. During the physiotherapist manipulation, there are no problems with marker occultations.
Building Blocks: Shown in Figure 7.
Number of Sensors: Five are placed if a single upper extremity is examined and seven for both.
Connection: The sensors are wireless and connect to the DPP via Wi-Fi.

Step 6: Body Attachment

Placement: Sacrum, on vertebra D2, the upper area of the head, outside of the arm just above the elbow, and on the forearm on the upper face of the wrist.
Attachment Method: Device on the skin, attached to the body by pre-cut kinesio-tape bands with hole to protrude the device and leave the LED visible. (In case of excessive hair, this would be removed. The area is cleaned with disinfectant). The choice of this attachment method is justified in the Table 3 example.

Step 7: Device Physical Properties

Shape: Oval, rounded edges with wings in the skin contact area, where the tape is fixed.
Dimensions: 4×3 cm.
Weight: 40 grams per sensor.
Housing material: Rigid plastic and resistant to humidity. It is painted with silicone texture for easy sanitising.
Colour: It is white, adapts to the environment, and transmits appropriate values.

Step 8: DPP

Elements: Computer, Wi-Fi dongle for communications, and an external monitor.
Basic software operation: It has two modes, the configuration for the professional user and the display for the patient user. It calculates the hits and misses regarding the target movement and displays them in real time. It connects to the Internet to store data in the cloud at the end of the session.

The example summarised in Table 4 and shown in the building-block scheme (Figure 7) considers the main requirements for the case of a rehabilitation service. It has been observed that the structuring carried out systematises the development process from the design goal. In the absence of being applied in more cases, it is considered to be a useful tool for new or improved product development.
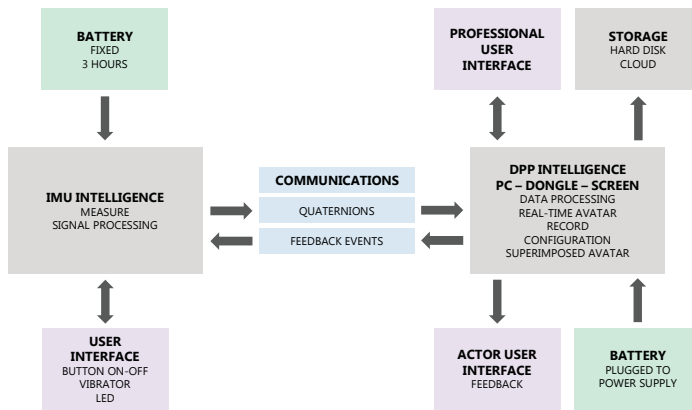
Figure 7. Case study building blocks.

4. Conclusions

This article presents a problem about the MoCap-wearable systems that affects product and service designers and electronics developers. This problem is related to the difficulty of considering all human, ergonomic, technological, or material factors, among others. Thus, MoCap-wearable designs require studying the following critical aspects that have been identified: context, user interaction, technology, body attachment, and physical property factors.

In response to this problem, a bibliographical and representative commercial product review has been carried out, which allowed extracting the factors directly related to design requirements, which have been collected and ranked. As a result, the Octopus methodology is proposed, which aims to help the MoCap-wearable system requirement extraction process, due to the factor schematisation and a visual representation that allow this to be studied and evaluated sequentially.

Octopus begins by studying the context to later develop the product or service. The method allows the creation of systems aimed towards MoCap from one, several, or all parts of the body. It has several characteristics that imply an improvement in this type of product design process, as follows:

- It is flexible and adaptable. Without a closed and immovable scheme that limits creativity, the design team, which is necessarily multidisciplinary, can eliminate or add factors and elements according to each case.
- The tool makes the job easier. Due to the visual representation of Figure 6 and the different steps proposed, it is expected to improve organisation, structuring, synthesis, and facilitate decision making, providing a global view of all the necessary factors.
- Due to the case study, it has been observed that the tool allows us to generate innovative ideas in an effortless way and to consider the main specifications and design problems.
- It is a communication tool between professionals or researchers from different disciplines involved in the design team, illustrating the common objectives to be fulfilled, mapping the status of the project, and allowing the assessment and recognition of the contributions of each of the team members.

Due to the simplification of the MoCap-wearable system creation process, this article introduces different fields and possibilities:

- In the current context of the technological progress, electronics miniaturisation, cost reduction, and the IoT, Octopus can contribute to facilitating the success of new products aimed at the MoCap area. It is expected that developed MoCap-wearable systems will be more suited to users

and environments, so that MoCap can be used more extensively solving current problems in existing applications and allowing its implementation in more and new ones.

- In relation to this search for new applications, it is observed that the assumption of the MoCap as a service can be an improvement and an opportunity to cover real needs.

- It is expected that the study increases knowledge of optical rigid body MoCap systems that, according to Baker et al. [22], have been largely ignored in the literature as of 2006, and it has been verified in further review that it has not increased in the last years.

- MoCap-wearables devices can be considered products of high or very high complexity in relation to the design requirements, so the study can also be extrapolated to other wearables and to other areas with less complexity. In fact, the described factors can be used not only for advanced technological devices but also for other, more basic products that need to be precisely and comfortably placed on the body.

As future work, the methodology should be implemented in more cases to improve it. Additionally, it is proposed to create objective methods to adequately evaluate the prototypes developed by operating in real situations, mainly in relation to the union attachment. In this way, it is proposed to investigate new types of body attachment. This may be interesting to study regarding the behaviour of the skin with the body movement and the design of the contact surfaces, considering anthropometric and morphological factors of the subject under study. In addition, there is no doubt that there is a latent need to look for new MoCap applications and areas, improving the movement and consequently the quality of life of more sectors of the population.

Author Contributions: J.J.M.: literature review, synthesis, ideation, and paper writing; J.M.: planning, ideation, and critical review of the article with special attention to motion capture technology and ergonomics; T.B.: planning, ideation, and critical review of the article with special attention to product design and design methodology aspects.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

MoCap	Motion Capture
IoT	Internet of Things
IMU	Inertial Measurement Unit
DPP	Data Processing Point
LED	Light-Emitting Diode
GUI	Graphic User Interface
RMS	Root Mean Square
SD	Secure Digital
LAN	Local Area Network
PAN	Personal Area Network

References

1. Zijlstra, W.; Aminian, K. Mobility Assessment in Older People: New Possibilities and Challenges. *Eur. J. Ageing* **2007**, *4*, 3–12. [CrossRef] [PubMed]

2. Majumder, S.; Mondal, T.; Deen, M.J. Wearable Sensors for Remote Health Monitoring. *Sensors* **2017**, *17*, 130. [CrossRef] [PubMed]

3. Perera, C.; Zaslavsky, A.; Christen, P.; Georgakopoulos, D. Context Aware Computing for the Internet of Things: A Survey. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 414–454. [CrossRef]

4. Knight, J.F.; Deen-Williams, D.; Arvanitis, T.N.; Baber, C.; Sotiriou, S.; Anastopoulou, S.; Gargalakos, M. Assessing the Wearability of Wearable Computers. In Proceedings of the 10th IEEE International Symposium on Wearable Computers, Montreux, Switzerland, 11–14 October 2006; pp. 75–82.

5. Gemperle, F.; Kasabach, C.; Stivoric, J.; Bauer, M.; Martin, R. Design for Wearability. In Proceedings of the 2nd International Symposium on Wearable Computers, Digest of Papers, Pittsburgh, PA, USA, 19–20 October 1998; pp. 116–122.

6. Skogstad, S.A.; Nymoen, K.; Høvin, M. Comparing Inertial and Optical Mocap Technologies for Synthesis Control. In Proceedings of the Intertional Sound and Music Computing Conference, Padova, Italy, 6–9 July 2011; pp. 421–426.

7. Mayagoitia, R.E.; Nene, A.V.; Veltink, P.H. Accelerometer and Rate Gyroscope Measurement of Kinematics: An Inexpensive Alternative to Optical Motion Analysis Systems. *J. Biomech.* **2002**, *35*, 537–542. [CrossRef]

8. Cloete, T.; Scheffer, C. Benchmarking of a Full-Body Inertial Motion Capture System for Clinical Gait Analysis. In Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 4579–4582.

9. Mooney, R.; Corley, G.; Godfrey, A.; Quinlan, L.R.; ÓLaighin, G. Inertial Sensor Technology for Elite Swimming Performance Analysis: A Systematic Review. *Sensors* **2016**, *16*, 18. [CrossRef] [PubMed]

10. Cooper, G.; Sheret, I.; McMillian, L.; Siliverdis, K.; Sha, N.; Hodgins, D.; Kenney, L.; Howard, D. Inertial Sensor-Based Knee Flexion/Extension Angle Estimation. *J. Biomech.* **2009**, *42*, 2678–2685. [CrossRef] [PubMed]

11. Roetenberg, D.; Luinge, H.; Slycke, P. Xsens MVN: Full 6DOF Human Motion Tracking using Miniature Inertial Sensors. Available online: https://www.researchgate.net/profile/Per_Slycke/publication/239920367_Xsens_MVN_Full_6DOF_human_motion_tracking_using_miniature_inertial_sensors/links/0f31752f1f60c20b18000000/Xsens-MVN-Full-6DOF-human-motion-tracking-using-miniature-inertial-sensors.pdf (accessed on 14 August 2017).

12. Kok, M.; Hol, J.D.; Schön, T.B. An Optimization-Based Approach to Human Body Motion Capture Using Inertial Sensors. *IFAC Proc. Vol.* **2014**, *47*, 79–85. [CrossRef]

13. Riaz, Q.; Tao, G.; Krüger, B.; Weber, A. Motion Reconstruction using very Few Accelerometers and Ground Contacts. *Graph. Model.* **2015**, *79*, 23–38. [CrossRef]

14. NaturalPoint—Optical Tracking Solutions. Available online: https://www.naturalpoint.com/ (accessed on 20 June 2017).

15. Trivisio—Inertial Motion Traking. Available online: https://www.trivisio.com/inertial-motion-tracking (accessed on 20 June 2017).

16. Ahmad, N.; Ghazilla, R.A.R.; Khairi, N.M.; Kasi, V. Reviews on various Inertial Measurement Unit (IMU) Sensor Applications. *Int. J. Signal Proc. Syst.* **2013**, *1*, 256–262. [CrossRef]

17. Cloete, T.; Scheffer, C. Repeatability of an Off-the-Shelf, Full Body Inertial Motion Capture System during Clinical Gait Analysis. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 5125–5128.

18. Bellusci, G.; Roetenberg, D.; Dijkstra, F.; Luinge, H.; Slycke, P. Xsens MVN Motiongrid: Drift-Free Human Motion Tracking using Tightly Coupled Ultra-Wideband and Miniature Inertial Sensors. Available online: http://www.dwintech.com/MVN_MotionGrid_White_Paper.pdf (accessed on 14 August 2017).

19. Favre, J.; Aissaoui, R.; Jolles, B.; De Guise, J.; Aminian, K. Functional Calibration Procedure for 3D Knee Joint Angle Description using Inertial Sensors. *J. Biomech.* **2009**, *42*, 2330–2335. [CrossRef] [PubMed]

20. Dejnabadi, H.; Jolles, B.M.; Casanova, E.; Fua, P.; Aminian, K. Estimation and Visualization of Sagittal Kinematics of Lower Limbs Orientation using Body-Fixed Sensors. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1385–1393. [CrossRef] [PubMed]

21. Vlasic, D.; Adelsberger, R.; Vannucci, G.; Barnwell, J.; Gross, M.; Matusik, W.; Popović, J. Practical Motion Capture in Everyday Surroundings. *ACM Trans. Graph. (TOG)* **2007**, *26*. [CrossRef]

22. Baker, R. Gait Analysis Methods in Rehabilitation. *J. NeuroEng. Rehabil.* **2006**, *3*, 4. [CrossRef] [PubMed]

23. Capozzo, A.; Catani, F.; Leardini, A.; Benedetti, M.; Della Croce, U. Position and Orientation in Space of Bones during Movement: Experimental Artefacts. *Clin. Biomech.* **1996**, *11*, 90–100. [CrossRef]

24. Benoit, D.L.; Ramsey, D.K.; Lamontagne, M.; Xu, L.; Wretenberg, P.; Renström, P. Effect of Skin Movement Artifact on Knee Kinematics during Gait and Cutting Motions Measured in Vivo. *Gait Posture* **2006**, *24*, 152–164. [CrossRef] [PubMed]

25. Haratian, R.; Twycross-Lewis, R.; Timotijevic, T.; Phillips, C. Toward Flexibility in Sensor Placement for Motion Capture Systems: A Signal Processing Approach. *IEEE Sens. J.* **2014**, *14*, 701–709. [CrossRef]

26. Andreoni, G.; Standoli, C.E.; Perego, P. Defining Requirements and Related Methods for Designing Sensorized Garments. *Sensors* **2016**, *16*, 769. [CrossRef] [PubMed]

27. Yang, S.; Li, Q. Inertial Sensor-Based Methods in Walking Speed Estimation: A Systematic Review. *Sensors* **2012**, *12*, 6102–6116. [CrossRef] [PubMed]

28. Sabatini, A.M.; Martelloni, C.; Scapellato, S.; Cavallo, F. Assessment of Walking Features from Foot Inertial Sensing. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 486–494. [CrossRef] [PubMed]

29. Taetz, B.; Bleser, G.; Miezal, M. Towards Self-Calibrating Inertial Body Motion Capture. In Proceedings of the 19th International Conference on Information Fusion (FUSION), Heidelberg, Germany, 5–8 July 2016; pp. 1751–1759.

30. Müller, P.; Bégin, M.; Schauer, T.; Seel, T. Alignment-Free, Self-Calibrating Elbow Angles Measurement using Inertial Sensors. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 312–319. [CrossRef] [PubMed]

31. Yu, B.; Bao, T.; Zhang, D.; Carender, W.; Sienko, K.H.; Shull, P.B. Determining Inertial Measurement Unit Placement for Estimating Human Trunk Sway while Standing, Walking and Running. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 4651–4654.

32. Noraxon-3D Motion Capture Sensors. Available online: http://www.webcitation.org/6rMGpRqTm (accessed on 20 June 2017).

33. Perception Neuron-MoCap System. Available online: https://neuronmocap.com/ (accessed on 20 June 2017).

34. My Swing-Golf Motion Captrure Service. Available online: http://www.webcitation.org/6rMHJKUtu (accessed on 20 June 2017).

35. Alex Posture. Available online: http://www.webcitation.org/6rMHl1we4 (accessed on 20 June 2017).

36. Araig-Sensory Feedback Suit for Video Game. Available online: http://www.webcitation.org/6rMHwg3Hu (accessed on 20 June 2017).

37. LEO Fitness Intelligence. Available online: http://www.webcitation.org/6rMI2Yhj5 (accessed on 20 June 2017).

38. Notch-Smart Motion Capture for Mobile Devices. Available online: http://www.webcitation.org/6rMI4aW6g (accessed on 20 June 2017).

39. Rokoko-Motion Capture System-Smartsuit Pro. Available online: http://www.webcitation.org/6rMIBBlnf (accessed on 20 June 2017).

40. Vicon-Motion Captrure Systems. Available online: http://www.webcitation.org/6rMINb14O (accessed on 20 June 2017).

41. Run3D-3D Gait Analysis. Available online: http://www.webcitation.org/6rMIX3thp (accessed on 20 June 2017).

42. Google Glass. Available online: http://www.webcitation.org/6rMIaF0iB (accessed on 20 June 2017).

43. MC10-Wearable Healthcare Technology & Devices. Available online: http://www.webcitation.org/6rMIvMYAG (accessed on 20 June 2017).

44. Quell-Wearable Pain Relief That's 100% Drug Free. Available online: http://www.webcitation.org/6rMJ27WzR (accessed on 20 June 2017).

45. Perception Legacy-Motion Capture System. Available online: http://www.webcitation.org/6rMJRSnZb (accessed on 20 June 2017).

46. Shadow-Motion Capture System. Available online: http://www.webcitation.org/6rMJWrKgp (accessed on 20 June 2017).

47. The Imaginarium Studios-Performance Capture Studio and Production Company. Available online: http://www.webcitation.org/6rMJbEqrb (accessed on 20 June 2017).

48. Jolt Sensor-A Wearable head Impact Sensor for Youth Athletes. Available online: http://www.webcitation.org/6rMJmWx9e (accessed on 20 June 2017).

49. SenseON-A Clinically Accurate Heart Rate Monitor. Available online: http://www.webcitation.org/6rMJsFsGa (accessed on 20 June 2017).

50. Sensoria Fitness-Garments Monitor Heart Rate with Embedded Technology for the Most Effective Workout. Available online: http://www.webcitation.org/6rMJxDGEo (accessed on 20 June 2017).

51. Stt Systems. Available online: http://www.webcitation.org/6rMKMGBto (accessed on 20 June 2017).

52. Technaid-Motion Capture System. Available online: http://www.webcitation.org/6rMKQqNHf (accessed on 20 June 2017).

53. Melon-A Headband and Mobile App to Measure Your Focus. Available online: http://www.webcitation.org/6rMKxkcY0 (accessed on 20 June 2017).

54. Teslasuit-Full Body Haptic VR Suit. Available online: http://www.webcitation.org/6rML4RSOd (accessed on 20 June 2017).

55. Myo Gesture Control Armband-Wearable Technology by Thalmic Labs. Available online: http://www.webcitation.org/6rMLGM7sy (accessed on 20 June 2017).

56. Xsens-3D Motion Traking. Available online: http://www.webcitation.org/6rEoDr7rq (accessed on 15 June 2017).

57. Reebok CheckLight. Available online: http://www.webcitation.org/6rMLTqeWx (accessed on 20 June 2017).

58. Upright-the World Leader in Connected Posture Trainers. Available online: http://www.webcitation.org/6rMLZ4Sic (accessed on 20 June 2017).

59. Thync Relax. Available online: http://www.webcitation.org/6rMLoUfF1 (accessed on 20 June 2017).

60. Motti, V.G.; Caine, K. Human Factors Considerations in the Design of Wearable Devices. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2014**, *58*, 1820–1824. [CrossRef]

61. Spagnolli, A.; Guardigli, E.; Orso, V.; Varotto, A.; Gamberini, L. Measuring User Acceptance of Wearable Symbiotic Devices: Validation Study Across Application Scenarios. In Proceedings of the International Workshop on Symbiotic Interaction, Helsinki, Finland, 30–31 October 2014; pp. 87–98.

62. Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*; Pearson Education: London, UK, 2010.

63. Mckee, M.G. Biofeedback: An Overview in the Context of Heart-Brain Medicine. *Clevel. Clin. J. Med.* **2008**, *75*, 31–34. [CrossRef]

64. Shiratori, T.; Park, H.S.; Sigal, L.; Sheikh, Y.; Hodgins, J.K. Motion Capture from Body-Mounted Cameras. *ACM Trans Graph. (TOG)* **2011**, *30*. [CrossRef]

65. Thewlis, D.; Bishop, C.; Daniell, N.; Paul, G. Next Generation Low-Cost Motion Capture Systems can Provide Comparable Spatial Accuracy to High-End Systems. *J. Appl. Biomech.* **2013**, *29*, 112–117. [CrossRef] [PubMed]

66. Blanco, T.; Casas, R.; Manchado-Pérez, E.; Asensio, Á.; López-Pérez, J.M. From the Islands of Knowledge to a Shared Understanding: Interdisciplinarity and Technology Literacy for Innovation in Smart Electronic Product Design. *Int. J. Technol. Des. Educ.* **2017**, *27*, 329–362. [CrossRef]

67. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]

68. Von Marcard, T.; Rosenhahn, B.; Black, M.J.; Pons-Moll, G. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Comput. Graph. Forum* **2017**, *369*, 349–360. [CrossRef]

69. Kalkum, E.; Van Drongelen, S.; Mussler, J.; Wolf, S.I.; Kuni, B. A Marker Placement Laser Device for Improving Repeatability in 3D-Foot Motion Analysis. *Gait Posture* **2016**, *44*, 227–230. [CrossRef] [PubMed]

70. Knight, J.F.; Baber, C.; Schwirtz, A.; Bristow, H.W. The Comfort Assessment of Wearable Computers. *ISWC* **2002**, *2*, 65–74.

71. Sonderegger, A. Smart Garments—The Issue of Usability and Aesthetics. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, Zurich, Switzerland, 8–12 September 2013; pp. 385–392.

72. Lilien, G.L.; Morrison, P.D.; Searls, K.; Sonnack, M.; Hippel, E.V. Performance Assessment of the Lead User Idea-Generation Process for New Product Development. *Manag. Sci.* **2002**, *48*, 1042–1059. [CrossRef]

73. Maguire, M. Methods to Support Human-Centred Design. *Int. J. Hum.-Comput. Stud.* **2001**, *55*, 587–634. [CrossRef]

74. Abras, C.; Maloney-Krichmar, D.; Preece, J. User-Centered Design. In *Encyclopedia of Human-Computer Interaction*; Bainbridge, W., Ed.; Sage Publications: Thousand Oaks, CA, USA, 2004; Volume 37, pp. 445–456.

75. Tassi, R. Dervice Design Tools Communication Methods Supporting Design Processes. Ph.D. Thesis, Politecnico di Milano, Milano, Italy, October 2008.

76. Stickdorn, M.; Schneider, J.; Andrews, K.; Lawrence, A. *This is Service Design Thinking: Basics, Tools, Cases*; Wiley Hoboken: Hoboken, NJ, USA, 2011.

77. Shostack, G.L. Designing Services That Deliver. *Harv. Bus. Rev.* **1984**, *62*, 133–139.

78. Blanco, T.; Berbegal, A.; Blasco, R.; Casas, R. Xassess: Crossdisciplinary Framework in User-Centred Design of Assistive Products. *J. Eng. Des.* **2016**, *27*, 636–664. [CrossRef]

Performance Study of a Torsional Wave Sensor and Cervical Tissue Characterization

Antonio Callejas [1], Antonio Gomez [2], Juan Melchor [1,3,*], Miguel Riveiro [1], Paloma Massó [1,3], Jorge Torres [1], Modesto T. López-López [4] and Guillermo Rus [1,3]

[1] Department of Structural Mechanics, University of Granada, 18071 Granada , Spain; acallejas@ugr.es (A.C.); mriveiro@ugr.es (M.R.); pmasso@ugr.es (P.M.); jtp31@correo.ugr.es (J.T.); grus@ugr.es (G.R.)

[2] Department of Mechanical Engineering, University College London, WC1E 7JE London, UK; aj.gomez@ucl.ac.uk

[3] Biosanitary Research Institute, 18012 Granada, Spain

[4] Department of Applied Physics, University of Granada, 18071 Granada, Spain; modesto@ugr.es

* Correspondence: jmelchor@ugr.es; Tel.: +34-958-240-037

Abstract: A novel torsional wave sensor designed to characterize mechanical properties of soft tissues is presented in this work. Elastography is a widely used technique since the 1990s to map tissue stiffness. Moreover, quantitative elastography uses the velocity of shear waves to achieve the shear stiffness. This technique exhibits significant limitations caused by the difficulty of the separation between longitudinal and shear waves and the pressure applied while measuring. To overcome these drawbacks, the proposed torsional wave sensor can isolate a pure shear wave, avoiding the possibility of multiple wave interference. It comprises a rotational actuator disk and a piezoceramic receiver ring circumferentially aligned. Both allow the transmission of shear waves that interact with the tissue before being received. Experimental tests are performed using tissue mimicking phantoms and cervical tissues. One contribution is a sensor sensitivity study that has been conducted to evaluate the robustness of the new proposed torsional wave elastography (TWE) technique. The variables object of the study are both the applied pressure and the angle of incidence sensor–phantom. The other contribution consists of a cervical tissue characterization. To this end, three rheological models have fit the experimental data and a static independent testing method has been performed. The proposed methodology permits the reconstruction of the mechanical constants from the propagated shear wave, providing a proof of principle and warranting further studies to confirm the validity of the results.

Keywords: torsional wave sensor; tissue mimicking phantom; cervical tissue; rheological model; rheometry experiment; sensitivity study; complex shear modulus

1. Introduction

Several authors have suggested that the mechanical functionality of soft tissue is a highly relevant clinical parameter to a broad range of pathologies. For example, Berghella [1] hypothesized that the elasticity of the cervical tissue below typical values conditioned delayed term birth and failure of the required delivery induction, which is the leading cause of fetal suffering. To this end, in the last years, measuring shear viscoelastic constants in soft tissues has become a challenge. There are existing commercial shear sensors available, although some limitations, like the upper threshold of detectable stiffness, have been reported [2].

In this field, two main elastography techniques have been developed for the measurement of tissue stiffness: Static elastography (SE) and dynamic elastography (DE). On one hand, SE provides strain maps. The tissue is manually compressed with the probe until no further tissue deformation is observed in the B-mode image (maximum deformation). Then, the maximum deformability of

the cervix is quantified as the ratio of the anterior-posterior distance (thickness), before (reference configuration) and after compression application [3]. A qualitative description of the relative strain distribution is obtained, which is not a quantitative description of the real stiffness of the tissue. Despite the simplicity and compatibility of the technique with standard equipment, measurements are highly dependent on the pressure applied by the practitioner. The lack of a standard is currently one of the challenges for the clinical use of SE [4–6].

On the other hand, DE techniques are based on shear waves propagation and provide a quantitative description of tissue stiffness compared to static methods [7–10]. A great advantage of some of these techniques is the ability to depict the spatial distribution of mechanical properties of the biological tissues [2]. However, some limitations have been identified such as limitation of the acquisition frame rate to recover an entire image of the medium, and the deposited energy in the medium [11,12].

Recently, a new family of sensors based on torsional waves has been patented, such as the US 5321333 A [13], which belongs to the field of low frequency sensors for the generation or reception of torsional shear waves in a substantially solid medium. This new technique solves the problem of the strong attenuation of induced shear waves, providing a new way for the development of new noninvasive tissue characterization techniques [14]. In this work, a new torsional wave sensor, designed and manufactured by our group, has been used [15]. The probe used in this work has been developed to provide quick measurements of the shear stiffness of the cervix. It has been optimized reducing the spurious waves contamination (P-waves), using a similar procedure as depicted in [15], the excitation energy deposited in the tissue and the dependence with the applied pressure sensor-tissue. The simplicity of the method ensures an easy clinical procedure and a short learning curve for the operator. Although further investigations are required, the work hypothesis considers that the averaged and local measurement of shear stiffness might be sufficiently significant for characterizing the elastic condition of the cervix.

Compared with other DE techniques such as SSI (SuperSonic Imagine, Aix-en-Provence, France) [2,16], the torsional wave elastography (TWE) offers a higher range of frequencies, up to 10 kHz, due to a high acquisition rate up to 100,000 fps compared with the 6000 fps of the SSI technique. This provides a broader spectrum of the mechanical behavior of the cervical tissue, which yields a finer viscoelastic characterization. At the same time, this high acquisition rate allows the analysis of perturbations of higher order by obtaining more data points per wavelength. Another advantage of the TWE technique is based on the lower energy deposited in the tissue during the procedure. The low energy delivery favors a high repeatability of the measurements, which, combined with the capacity of testing different pressure level scenarios, might minimize the inter and intra-operator deviations.

In the field of elasticity imaging, quantifying viscoelastic properties in soft tissues is a challenging problem. To date, several rheological models have been used to characterize different soft tissues and tissue mimicking phantoms. The Kelvin–Voigt (KV) model has been used specifically to quantify viscoelastic properties in soft tissues. Apart from this method, for characterization of the viscoelastic media, some authors have used the Maxwell (M), generalized Maxwell (GM) model, the Zener (Z) model and the Kelvin–Voigt fractional derivative (KVFD) model [17–24]. Nowadays, little is known about the mechanical behavior of cervical tissue. Genisson et al. [25] have shown that the pathological cervix seems to be a softer than normal cervix. A study in pregnant ewes carried out by Peralta et al. [26] indicates that stiffness of the uterine cervix changes throughout the maturation process. Miklos et al. in a study performed on 24 patients under small compressions, and a range of frequencies from 0.1 to 100 Hz [27] concluded that normal cervical tissue exhibited values of complex modulus over the same range. Finally, Jiang et al. applied 3D multifrequency MR elastography (3DMMRE) to study the viscoelasticity of the uterus and cervix, showing sensitivity to structural and functional changes of the endometrium and myometrium during the menstrual cycle [28].

As it was stated in this paper, we propose a novel torsional wave sensor capable of emitting and receiving shear waves. This sensor can measure the shear wave speed in tissue mimicking phantoms

and soft tissues. The structure of the sensor probe is described in detail, with an optimal design necessary to avoid the resonant frequency in the emitter and the receiver. Then, the experimental setup and protocol are presented. Experimental tests of the mechanical properties of the tissue mimicking phantoms were performed to confirm that the measurements are independent of the applied pressure and the angle between probe and phantom surface. Finally, a human cervical tissue characterization study is carried out. Ex vivo samples have been tested and three rheological models have been studied to understand the behavior of cervical tissues under shear wave propagation. Finally, an alternative testing method based on measuring viscoelastic parameters is used for comparison with the TWE technique.

2. Materials and Methods

2.1. Structure of the Sensor

The torsional device is composed of an emitter, a receiver and a casing that holds both. The emitter, responsible for the transmission of the wave which travels through the phantom, consists of a PLA (polylactic acid) disk, printed in 3D, whose rotational movement is due to an electromechanical actuator (see Figure 1b). This actuator is covered by a Faraday's cage of aluminum paper to nullify any effect of external electromagnetic fields. To improve the attachment between the aluminum and the actuator, a silver conductive epoxy, used as adhesive that increases the conductivity, is used. The actuator is excited electrically by a wave generator, where some parameters like frequency, voltage or working cycle are controlled.

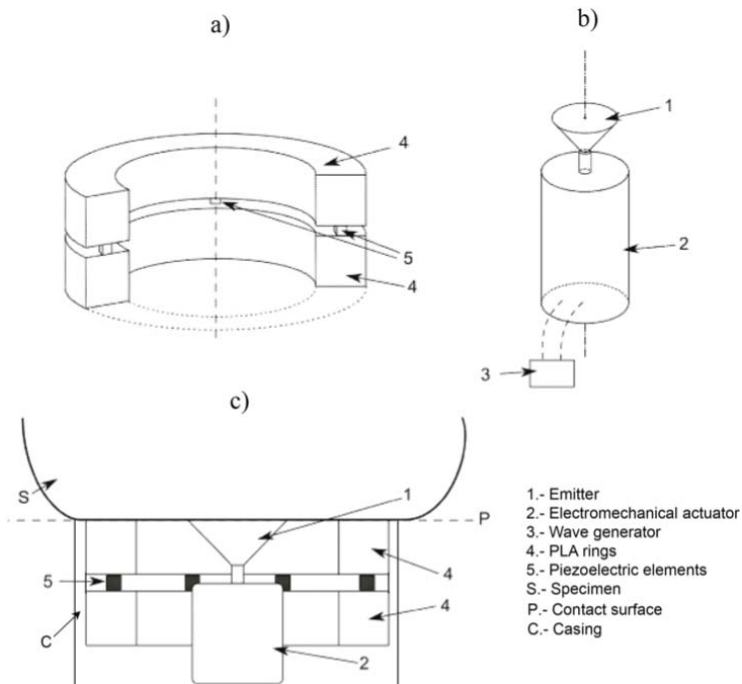**Figure 1.** Schematic view of the sensor elements. (**a**) the receiver; (**b**) the emitter; (**c**) contact sensor–phantom.

The receiver, showed in Figure 1a, is formed by two PLA rings with four slots in the inner face of the ring, where the four ceramic piezoelectric elements, whose material is NCE51, are fitted. All the piezoelectric elements are connected to each other with a copper rod, which works as a bracket too.

The rings mentioned above provide the inertia to reduce the resonant frequency and the stiffness to reduce dilatational waves. The array of transversely-polarized piezoceramic elements is connected to the ring by silver conductive epoxy. The piezoceramic elements are responsible for transforming the mechanical movement into an electric signal. Thus, each piezoelectric element is in contact with two electrodes of different charges.

The third part is a casing with geometrical and material selection to control the mechanical cross-talk and hold the emitter and the receiver (see Figure 1c). This configuration eliminates the masking p-waves that systematically arises at the boundaries of the regular contact or comb transducers [29].

2.2. Simplified Analytical Model

To respond to the principles of an optimal design of a torsional wave sensor considering the extraction of biomechanical properties, it is necessary to compute Finite Element Models (FEM) as a first step [15,29,30]. Thus, the resonant frequency can be determined for torsional waves for the emitter and the receiver using Finite Element Analysis Program (FEAP) software [31]. Therefore, a simplified model with a disk as a transmitter and a ring as a receiver (to facilitate the accessibility of experimental measurements) was developed to validate the output of resonant frequency of FEM.

The fundamentals of a simplified analytical model for an oscillatory movement of a torsional sensor are shown as follows. Consequently, the piezoelectricity coupling is introduced as a strain law uniformly distributed and directly proportional to the electrical field [32].

A single degree of freedom system is established, where the eigenproblem is reduced to a single frequency and a single mode. The movement is assumed to be subjected to the torsion rotation Θ in radians. In addition, the dynamic equilibrium of torsional moment is defined as:

$$K\Theta + I\frac{d^2\Theta}{dt^2} = 0, \tag{1}$$

where K is the stiffness coefficient measured in [Nm/rad] and I is the inertia moment. The steady-state solution has the following expression:

$$\Theta = \Theta^0 \sin(\omega t), \tag{2}$$

where ω is the natural frequency valid for the equilibrium Equation (1).

Considering the subsystem eigenfrequency in the case of cylinder mass and analyzing the relationship between the equations of equilibrium and the moment of inertia for a cylinder, the frequency obtained is derived as

$$f = \frac{1}{2\pi}\sqrt{\frac{nabd^2\mu}{\frac{\pi}{2}l^{\text{eff}}hr^4\rho}}. \tag{3}$$

Subsequently, in the case of ring mass, analyzing the subsystem eigenfrequency with the same equilibrium equation and the moment of inertia of a ring shape, we extract the next frequency

$$f = \frac{1}{2\pi}\sqrt{\frac{nabd^2\mu}{2\pi l^{\text{eff}}hmr^3\rho}}, \tag{4}$$

where n is the number of piezoelectric elements, a and b are the plane dimensions of the piezoelectric ceramic, d the distance from the center of rotation, μ the shear modulus, l^{eff} the effective length between piezoelectric ceramics, h the height of the cylinder or the ring, r the radius of the cylinder or the ring, m the thickness of the ring in the radial direction and ρ the density of the mass blocks.

The sensor is designed with a cylinder mass and a ring mass for each inertia subsystems of emission and reception. Then, their eigenfrequencies are fitted to optimize the resonant amplification.

Finally, the frequency of resonance obtained from FEM simulation is analyzed measuring the time between peaks of cycles. The result coincides with the resonance frequencies of the transducer described in the Table 1 with a percentage of error below 1% in each case.

Table 1. Validation analytic design vs. FEM.

Disc Frequency [Hz]	2.80×10^4
Ring Frequency [Hz]	2.82×10^4
FEM Frequency [Hz]	2.82×10^4
Error 1 (%)	0.68
Error 2 (%)	0.04

2.3. Experimental Setup

The gelatin gels and cervical tissues were excited by a low-frequency sine-burst at different frequencies (from 300 Hz to 1000 Hz). Excitation signals were generated by an arbitrary wave generator (Agilent 33220A, Santa Clara, CA, United States) and amplified (Radio Frequency Power Amplifier 150 A, 150 W, 100 MHz) before reaching the disk emitter. Measurements were acquired for 16 V mechanical actuator voltage amplitude that provides particle displacements at the emitter surface. The transmitted shear wave motion was propagated across the homogeneous phantoms and recorded by the receiver. Pre-amplification of received signal (Olympus, 576, 172 × 42.5 dB, Waltham, MA, United States) and low-pass filtering is necessary for recording a signal of greater amplitude than the signal noise level. The wave generator is synchronized with the oscilloscope (HDO 4034, 350 MHz, 2.5 GS/s, Santa Clara, CA, United States) to record the start of the signal. As a final step, a Matlab (Release 2014b, MathWorks, Natick, United States) optimization algorithm yields the shear wave speed. The experimental setup of the specimens is shown in Figure 2.
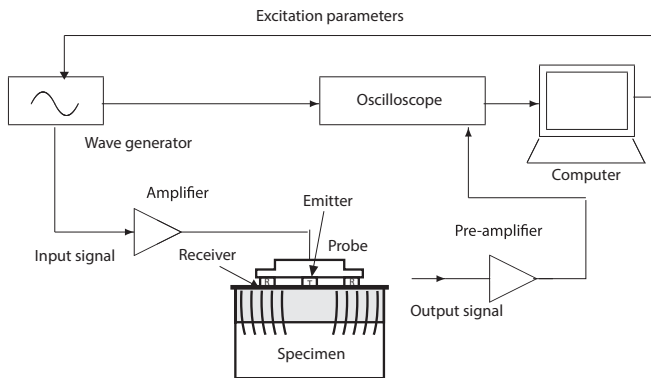
Figure 2. Experimental configuration of the excitation–propagation measurement system.

2.4. Specimens and Experimental Protocols

2.4.1. Sensitivity Study

Gelatin samples from porcine skin (Type A, 300 Bloom, G2500, Sigma-Aldrich, St. Louis, MO, USA) were constructed to test the sensor sensitivity under different applied pressures, sensor–specimens and different angles of incidence. Six homogeneous phantoms with 8% and 10% (*w/w*) gelatine concentrations were tested. Gelatine powder and purified water at 90 °C were stirred for 10 min to ensure gelatine dissolution. Molten gelatine was cast in cylindrical molds (6.8 cm diameter, 1.35 cm height) and kept at room temperature for 2 h before being stored in the refrigerator at

5 °C. The phantoms were taken out one day after being stored in the fridge and were tested when they reached the laboratory temperature (22 ± 1 °C).

All measurements were carried out with the same experimental protocol. A counterweight device was specifically designed to control the applied pressure and the angle of incidence phantom surface–sensor (see Figure 3). The phantoms were excited with a range of frequencies from 300 to 1000 Hz to avoid resonance phenomena in the rings that mimic shear waves. The phantoms were positioned at the balance to quantify the applied force during the measurement. Measurements from four applied pressures (from 4.81 to 24.06 kPa) and two angles of incidence phantom surface-sensor (0 and 7.5°) were performed. The material density used was 1000 kg/m^3.
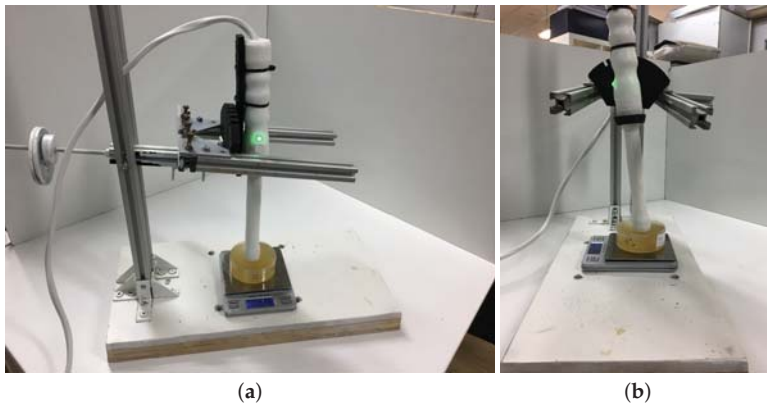

(**a**)   (**b**)

Figure 3. A counterweight device to control the applied pressure (phantom positioned on a balance) and the angle of incidence phantom-sensor. (**a**) angle of incidence 0°; (**b**) angle of incidence 7.5°.

2.4.2. Cervical Tissue Characterization

To obtain the viscoelastic properties, a test ex vivo on human cervical tissue samples from 3 healthy women aged from 55 to 74 years old were carried out using TWE and rheometry techniques (see Figure 4). Patient consent (the study was conducted according to the Declaration of Helsinki Principles and the agreement of the ethical committees of the University of Granada and the Universitary Hospital Complex of Granada) was obtained before carrying out the measurements. All patients were diagnosed with uterine prolapse (II and III grade) that required a partial or complete hysterectomy. The surgical procedures were performed at the San Cecilio University Hospital, Granada (Spain). The organs were placed in Phosphate Buffered Saline (PBS) and transferred to the Pathological Anatomy Laboratory in a refrigerator. The specimens were tested with the torsional wave sensor before being excised for rheometry experiments. Two pins were placed along the contours to prevent distortion of the tissue, and the cervix was placed on top of a rectangular piece of sound-absorbing rubber. Both measurements were performed at room temperature and typically within 2.5 h after being transferred to the laboratory, measuring three times each sample to obtain the mean and standard deviation. Given the independence of the applied pressure and the angle of incidence sensor-specimen in the sensitivity study, the cervical tissue samples were tested with the torsional wave sensor using the same range of frequencies (from 300 to 1000 Hz). After that experiment, circumferential samples were cut by a pathologist for rheometry experiments, 20 mm in diameter and 3–5 mm thick.
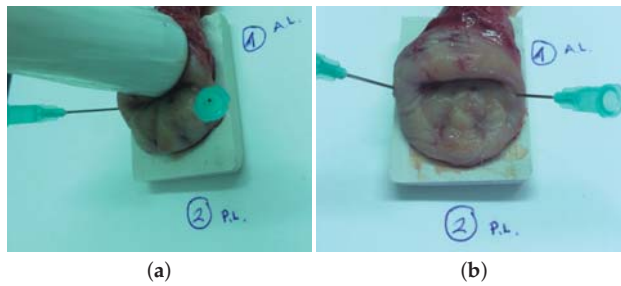
Figure 4. (a) measurement with the torsional wave sensor of a cervical tissue sample; (b) cervical tissue sample.

2.5. Time of Flight—Signal Processing

Time of flight calculation is conceptually simple, although it is necessary to introduce simplifying hypotheses. Physically, S-type waves are originated by the actuator and are transmitted through the specimen to the piezoelectric sensor, where they produce the deformation thereof and, consequently, an electric potential capturable by an oscilloscope.

In a purely theoretical approach, the signal in the sensor channel has null amplitude until the shear wave reaches the sensor producing a non-zero amplitude response. In practice, the presence of undesired interference signals from multiple sources requires a measurement and processing protocol to improve the quality of the resultant signal.

The signal processing includes the average of multiple measurements to decrease noise. Thereby, multiple identical measurements spaced by a relaxation time to completely dissipate previous waves were used. In addition, a low-pass filter in frequency domain with a cutoff frequency of 8000 Hz was applied to eliminate distortions generated from the electronics that controls the equipment.

Averaging and filters did not completely eliminate the electromagnetic coupling between emission and reception channels. This situation means that the received signal had an instant response time. Several tests suggest that the electromagnetic coupling component depends, on first approximation, only on the emitted signal power and in certain cases on external environmental components, like a remarkable increase of the humidity. In this work, the "calibration" measurement (without contact with the specimen) and the measurement of the specimen had the same coupling component. Thus, subtracting the calibration measurement from the measured signal, this effect can be eliminated.

The analysis of the resultant signal allows obtaining a first time of flight time measurement as the difference between the starting time of the received signal and the starting time of the excitation signal. Certain simplifications must be assumed for this analysis since there are phenomena such as imperfect centering of the transmitter with respect to the four piezoelectric receivers and transient effects associated with the engine start, high frequencies transmission and other effects whose quantification are beyond the scope of this work. Taking this into account, several approaches have been analyzed.

1. To normalize the signal with the first maximum (in absolute value) and to calculate the start from this maximum, using a reference value (typically 5% or 0.05).
2. To normalize the signal with the first maximum (in absolute value) and to calculate the start from the crossing of the horizontal axis with the tangent on the linear zone of the first sinusoidal section of the curve.
3. To calculate the frequency using the first two maximum and minimum and to extend a quarter of a cycle from the first function extremum.
4. To generate a parameter-dependent reference signal that simulates the expected signal. By means of an inverse problem (a combination of genetic algorithms and quasi-Newton type optimization algorithms), to perform and to adjust the experimental and the reference signal minimizing a residue from weight functions.

In this work, the last approach was used.

2.6. Shear Speed from Time of Flight

Ideally, the shear velocity can be calculated directly from the time of flight

$$Velocity = \frac{Distance}{Time\ of\ flight}. \tag{5}$$

However, the time of flight measured directly from the sensor signal is not the correct wave time of flight time because there is a delay associated with multiple effects (transmission in the PLA, delays related to electronics and piezoelectric devices...). These types of effects can be combined into a single delay time factor, so the Equation (5) results in

$$Velocity = \frac{Distance}{Time\ of\ flight - delay}. \tag{6}$$

This internal delay is characteristic of the measurement configuration and must be estimated previously. Assuming that the internal delay depends only on the material and the frequency excitation, time of flight of measurements at different distance emitter-receiver can be extended to zero distance obtaining a non-zero value of time that is the delay searched.

2.7. Rheological Models

Similarly to many soft tissues, the behavior of cervical tissue can be modeled as linear viscoelastic media and several models have been already studied [17,33,34]. Here, we propose the Kelvin–Voigt, Kelvin–Voigt Fractional Derivative, Maxwell and Zener models, which are the most widely used (see Figure 5).
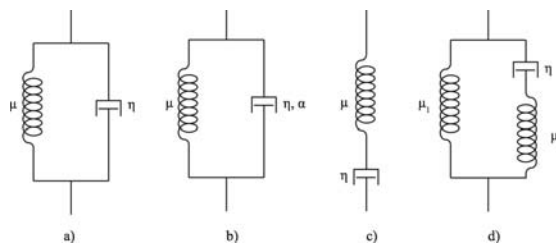
Figure 5. Rheological models. (**a**) the Kelvin–Voigt model; (**b**) the Kelvin–Voigt Fractional Derivative model; (**c**) the Maxwell model; (**d**) the Zener model.

The viscoelastic modulus (G^*) (Equation (7)), which is composed of a real part (G'), elastic or storage modulus and an imaginary part (G''), viscous or loss modulus, characterizes the viscoelasticity of the material [35]:

$$G^* = G'(\omega) + iG''(\omega). \tag{7}$$

The following rheological models relate the elasticity and viscosity by the complex shear modulus.

The Maxwell model (M) can be represented by a spring and a dashpot in series, as shown in the Figure 5. The complex shear moduli for this model is

$$G^{*M}(\omega) = \frac{\mu\eta\omega^2}{\mu^2 + \omega^2\eta^2} + i\frac{\mu^2\eta\omega}{\mu^2 + \omega^2\eta^2}. \tag{8}$$

The Kelvin–Voigt (KV) rheological model is characterized by elastic μ and viscous η coefficients. This rheological model consists of a spring and dashpot placed in parallel, which describes tissues

as solids that can creep but show little stress relaxation. The corresponding complex shear moduli for the KV model is

$$G^{*KV}(\omega) = \mu_1 + i\omega\eta, \tag{9}$$

where μ_1 is the shear elasticity (in *Pa*) and η is the shear viscosity (in Pa · s).

The Kelvin–Voigt Fractional Derivative model (KVFD), obtained by a generalization of the Kelvin–Voigt model, is a three-parameter fractional-order model. The first derivative in time in the KV model is replaced by a fractional-order derivative of order α [36]. The complex shear moduli for the KVFD model is [37,38]

$$G^{*KVFD}(\omega) = \mu_1 + \eta\omega^\alpha cos(\frac{\alpha\pi}{2}) + i\eta, \omega^\alpha sin(\frac{\alpha\pi}{2}), \tag{10}$$

where α is the fractional derivative power ($0 \leq \alpha \leq 1$).

The Zener model (Z) expresses both the stress relaxation and the creep in linear viscoelastic polymeric solids [39]. The complex shear modulus for the Zener model in the frequency response is given by [24]

$$G^{*Z}(\omega) = \frac{\mu_1\mu_2^2 + \omega^2\eta^2(\mu_1 + \mu_2)}{\mu_2^2 + \omega^2\eta^2} + i\frac{\mu_2^2\eta\omega}{\mu_2^2 + \omega^2\eta^2}, \tag{11}$$

where μ_1 and μ_2 are the elasticities (in Pa) and η is the shear viscosity (in Pa · s).

Dispersion curves are plots of shear wave speed (c_s) as a function of angular frequency ω. For mass density ρ, the relationship between c_s and the complex shear modulus is well known to be [40]

$$c_s(\omega) = \sqrt{\frac{2(G'^2 + G''^2)}{\rho(G' + \sqrt{G'^2 + G''^2})}}. \tag{12}$$

2.8. Rheometry Experiments

Rheometry is a technique that studies the relationship between stress–strain in materials that are capable of flowing, thus allowing to get mechanical properties related to elasticity and viscosity, which usually leads to the development of a constitutive relation. In this work, these properties were evaluated under shear stress with a controlled-rate magnetorheometer (MCR 300 Physica-Anton Paar, Graz, Austria) (see Figure 6). As inertia of the sample was negligible in oscillatory shear deformation, it required that a small phantom thickness compared with the wavelength of shear waves propagated through the medium at the frequency of oscillation [41]. The measuring system geometry was a 20 mm diameter parallel plate set for a gap width dependent on the sample thickness, as a rule 3–5 mm.
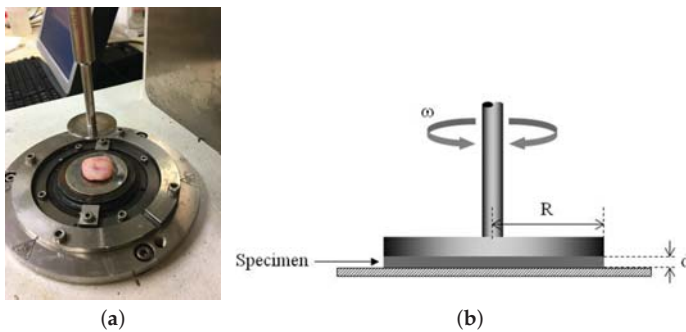

(**a**) (**b**)

Figure 6. (**a**) A controlled-rate magnetorheometer MCR 300 Physica-Anton Paar, Graz, Austria. Pink matter corresponds to the cervical tissue sample. (**b**) schematic view of the rheometer.

All measurements were conducted at room temperature to avoid loss of consistency. The experiment started with a squeezing compression of the rotor plate, no less than 0.2 N, for preventing slippery conditions. To ensure the reproducibility of the measurements, a pre-shear (10 cycles at 1 Hz) was applied with a waiting time of 10 s before measuring. The samples were subjected to a sinusoidal shear stress and the corresponding shear strain was measured, according to the constitutive equation for linear viscoelastic materials where

$$\sigma = \sigma_0 e i\omega t, \tag{13}$$

$$\gamma = \gamma_0 e^{i(\omega t - \delta)}, \tag{14}$$

are, respectively, the shear stress and strain, σ_0 is the stress amplitude, γ_0 the strain amplitude, ω the angular frequency, and δ the stress–strain phase lag. Two oscillatory tests were planned to obtain the viscoelastic modulus (G^*).

The experiments were [42,43]: (i) amplitude sweep: frequency was kept constant at 1, 2, 5, 10 and 20 Hz, and shear amplitude was varied up to values in which nonlinear behavior was reached. Nonlinearity was developed and was clearly distinguishable when G' and G'' were not only functions of ω but also strongly depended on γ_0. The frequency range was not higher owing to the linearity of oscillations that can not be ensured when the rotor (upper plate) and the sample inertia effects become significant at high frequencies [41]; and (ii) frequency sweep: shear amplitude was kept constant into the viscoelastic limit region, determined by the previous test and the frequency was swept from 0.5 Hz to 100 Hz logarithmically to obtain the oscillograms (G' and G'' as a function of ω).

2.9. Viscoelastic Parameters Reconstruction

The characteristic parameters of a viscoelastic material were obtained fitting the experimental results with the rheological models indicated in the previous section. Two different techniques (rheometry and TWE) were used in each sample of cervical tissue. For each sample and methodology, velocities at several frequencies were calculated and the mean of the velocities over all the measurements were used for further analysis.

In the rheometry study, the real and the imaginary parts of the viscoelastic modulus were obtained from the rheometer and the shear wave speed was calculated directly using Equations (7) and (12). For the torsional wave sensor, the shear wave speeds were obtained following the approach depicted in Section 2.5.

This allowed the construction of a speed-frequency chart for each tissue sample. In the next section will be represented the mean and standard deviation of the shear wave speeds using data from all the tissue samples. The high differences on the frequency ranges of the two techniques should be noted.

Every set of measurements was fitted to the models to obtain the parameter values. This adjustment was made by means of an inverse problem using a combination of genetic algorithms and quasi-Newton type optimization algorithms.

Three approaches were used and represented varying the fitted data using (a) shear wave speeds exclusively from rheometry, (b) speeds exclusively from the torsional sensor, and (c) all of the available data. The fitting procedure was applied to each tissue sample to obtain the average adjustment curves (including standard deviations). Two options were available: using the average values of the parameters to generate average curves or directly averaging the values of the curves. It should be highlighted that these adjustments are not equal since the models were nonlinear. The first approach was used in this work.

3. Results

Firstly, we explored the influence of the applied pressure in shear wave speed. Shear wave signals measured with TWE method at different applied pressures phantom-sensor and with the same frequency excitation (300 Hz) and gelatin concentration (10%) are shown in Figure 7.
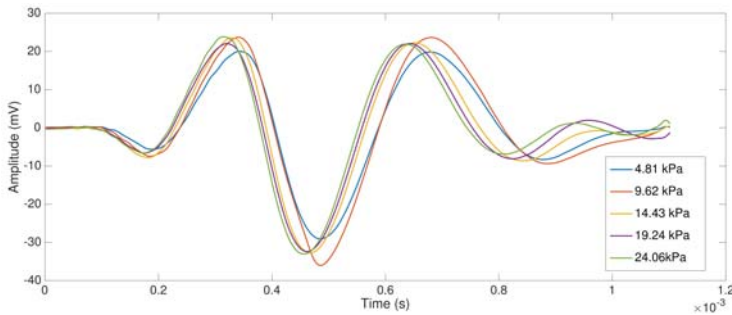

Figure 7. Shear wave signals at different pressures' phantom–sensors (from 4.81 to 24.06 kPa), frequency: 300 Hz; gelatin concentration: 10%

The shear wave speeds calculated from measurements using the proposed time of flight method are shown in Figure 8. Box and whisker plots are represented for the designed phantoms at 8 and 10% gelatin concentration (three samples per concentration), 300 Hz frequency and for the five different pressures. It can be extracted from the plots that there is no correlation between the applied pressure and the shear wave speed.

Figure 8. Box and whisker plots of shear wave speed measurements at different applied pressures phantom-sensor. Mean (lines within boxes), interquartile range (IQR, boxes) and extreme values (whiskers) are shown. (**a**) 8% gelatin; (**b**) 10% gelatin.

Secondly, to assess the other outcome of the sensor sensitivity study, two shear wave signals have been plotted at two different angles of incidence: 0 and 7.5° (see Figure 9).

Shear wave speed measurements are represented in Figure 10. Box and whisker plots are shown for the three designed phantoms at 8% and 10% gelatin concentration, 300 Hz frequency and for the two different angles of incidence. The same conclusion is extracted when varying the angle of incidence. No significant variations were obtained.
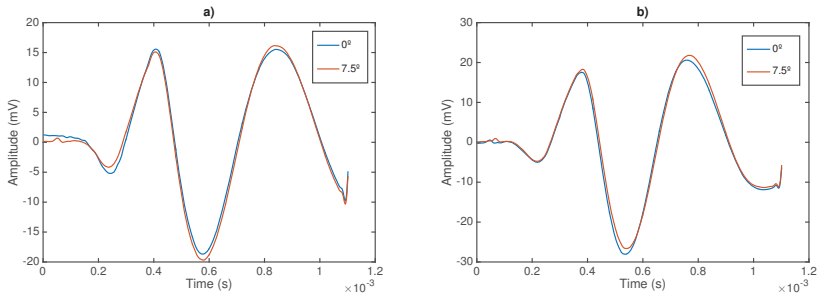
Figure 9. Measurement of shear wave speed at different angles of incidence sensor–phantom; frequency: 300 Hz. (**a**) 8% gelatin; (**b**) 10% gelatin.
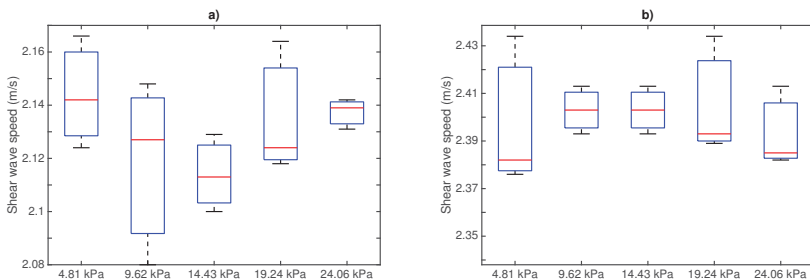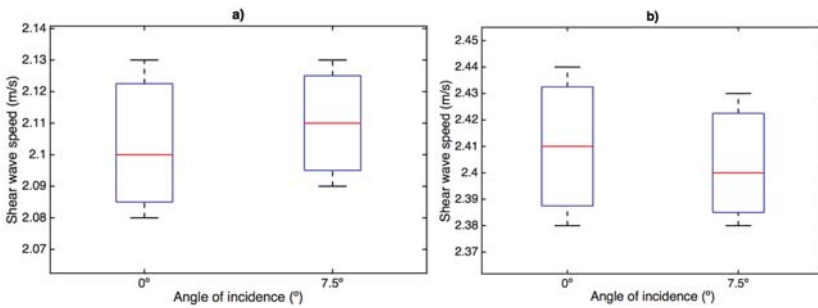

Figure 10. Box and whisker plots of shear wave speed measurements at different angles of incidence phantom–sensor. Mean (lines within boxes), interquartile range (IQR, boxes) and extreme values (whiskers) are shown; frequency: 300 Hz. (**a**) 8% gelatin; (**b**) 10% gelatin.

The real and the imaginary part of the complex shear modulus measured in cervical tissue by rheometry are shown in Figure 11. The measurements were taken three times for each cervix to obtain the mean and the standard deviation.
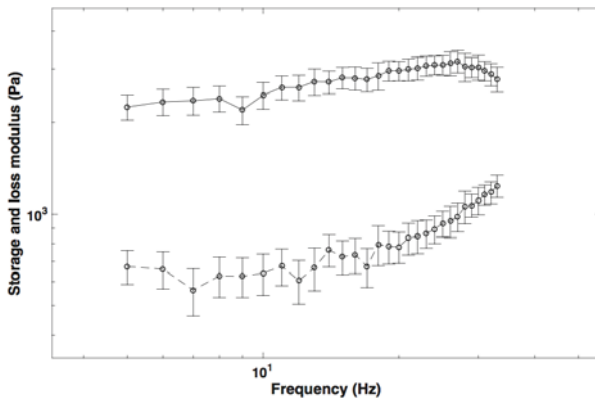
Figure 11. The storage and loss shear modulus measured by rheometry in cervical tissue.

Figure 12 shows the fitted curves using data from rheometry, elastography and both sources of data simultaneously, for each rheological model. Frequencies for rheometry measurements ranged from 4 to 30 Hz, while, for the elastography technique, there was a wider range from 300 to 1000 Hz.
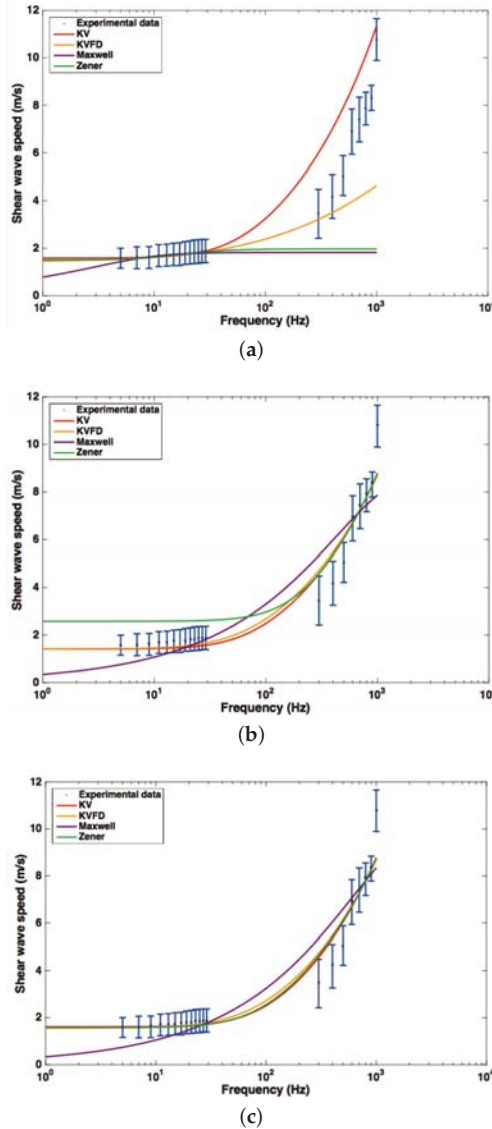
Figure 12. Fitted curves using data from rheometry, from elastography and using the combined data from the rheometry and elastography for each model. The circles are the mean values over the three cervix and the horizontal bars are the standard deviations. The curves for the Kelvin–Voigt (solid red line), Kelvin–Voigt Fractional Derivative (solid yellow line), Maxwell (solid purple line) and Zener model (solid green line) are shown. (**a**) data from rheometry; (**b**) data from elastography; (**c**) data from rheometry and elastography.

Tables 2 and 3 show the viscoelastic parameters using the data from rheometry, elastography and a combination of the two methods for Kelvin–Voigt and Kelvin–Voigt Fractional Derivative models. The high-frequency components of the dispersive curve play a much larger role than the low-frequency components. The curve shape or dispersive pattern is primarily determined by the high frequency components. KV and KVFD models matched satisfactorily all the data from the high to the low-frequency regime. The findings can also be confirmed by the results of Tables 2 and 3, showing that the elasticity and viscosity values obtained from the high-frequency TWE data were very similar those from the overall rheometry + TWE data, although they differ from those obtained from the low-frequency rheometry data.

Table 2. Viscoelastic parameters using the data from rheometry, TWE, and a combination of the two methods for the Kelvin–Voigt model.

Cervix Number	Elasticity μ (kPa)			Viscosity η (Pa·s)		
	Rheometry	TWE	R + TWE	Rheometry	TWE	R + TWE
1	1.69	2.13	1.82	5.31	4.32	4.21
2	1.83	2.52	2.10	6.52	4.55	4.64
3	1.85	2.64	1.84	7.19	4.9	4.65
Mean	1.79	2.43	1.92	6.34	4.59	4.5
Standard Deviation	0.08	0.26	0.15	0.95	0.29	0.25

Table 3. Viscoelastic parameters using the data from rheometry, TWE, and a combination of the two methods for the Kelvin–Voigt Fractional Derivative model.

Cervix Number	Elasticity μ (kPa)			Viscosity η (Pa·s)			Fract. Deriv. Power α		
	Rheometry	TWE	R + TWE	Rheometry	TWE	R + TWE	Rheometry	TWE	R + TWE
1	1.10	2.13	2.07	12	4.02	4.54	0.42	0.98	0.99
2	0.80	1.93	2.22	31	4.21	4.73	0.13	0.99	0.96
3	0.86	2.12	1.74	26	4.46	4.65	0.20	0.94	0.99
Mean	0.92	2.06	2.01	23	4.23	4.64	0.25	0.97	0.98
Standard Deviation	0.15	0.11	0.24	9.84	0.22	0.09	0.15	0.02	0.01

4. Discussion

Characterization of the viscoelastic properties of soft tissues is a key step for the development in many medical applications based on elastography imaging. In the case of cervical tissue, a special interest lays on the study of the evolution of the viscoelastic properties during pregnancy, in addition to their correlation with the gestational age. Little data for the case of viscoelasticity of human cervical tissue can be found in the literature [26].

The increasing interest in elastography techniques for measuring viscoelastic parameters is demanding appropriate validation studies. Currently, the traditional standard for evaluating the viscoelastic properties of soft tissues is based on rheological methods. However, these techniques are limited to in vitro and ex vivo samples. In addition, whereas the trustful range of frequency in the rheological test is from the quasi-static regime to no more than 50 Hz, dynamic elastography techniques usually use a range from 100 Hz to 500 Hz. This discrepancy makes the comparison process a challenge. However, due to the lack of alternatives, the validation between both techniques has been accepted in previous studies [44,45].

The presented TWE method is based on the transmission and detection of shear waves. Shear waves travel axisymmetrically from the center of the probe, where the mechanical actuator is placed, towards the outer side where the piezoceramic receivers are located. In this study, the TWE technique was proved capable of successful transmitting and receiving shear waves from 300 Hz to 1 kHz. This range of frequencies represents an obvious advantage when compared with commercial elastography devices [7], since a broader range of frequencies provides more significant information

about the viscoelastic behavior of the tissue, in particular for the viscous component, which is more sensitive to high frequencies [44].

Here, a study with respect of two variables involved during the testing procedure was presented: the normal pressure applied and the angle of incidence between the probe and the normal direction of the tissue surface. This simplified study may fall into a preliminary analysis of the intra-operator dependency of the method. The range of pressures (ranging from 4.81 kPa to 24.06 kPa) and angles (0° and 7°) were chosen in close relation with those experienced during a normal testing procedure. Gelatine phantoms with 8% and 10% concentration (w/w) were selected as an adequate representation of the viscoelastic behavior of soft tissues. As expected, higher concentration of gelatine produced higher shear wave speed due to higher stiffness. Variable levels of pressure produced a minimal variation in the shape of the collected signal, as can be seen in Figure 7. The standard deviation values of shear wave speed varied between 0.018 to 0.031 m/s, including all measurements for the range of frequencies studied (300–1000 Hz). The impact of this variation on the reconstructed values of shear wave speed was non-significant for all the frequencies. Therefore, no correlation between the level of pressure and shear wave speed was found. The same effects were observed when varying the angle of incidence. Figure 9 shows hardly noticeable variations in the measured signals. These observations were found consistent all over the different gelatine concentration phantoms tested, which seems to indicate that the TWE technique is, at least in gelatine-based phantoms, a robust and low user-dependent elastography method. However, as observed in the experiments, simple gelatine-based phantoms shows a low viscous behaviour, which is not proper of soft tissues. This low viscous behavior might be responsible for such stable results against the two variables of the user-dependency test. Different recipes for obtaining viscous elastography phantoms have been published, for instance by adding castor oil to the gelatine-based original recipe [46]. In further studies, elastography phantoms with higher viscosity will be tested. These analysis must be also extended to tissues exhibiting characteristics more similar to biological tissues, which may not be homogeneous [4] to define the validity limits of these studies and to develop clinical practice protocols to correct deviations when using the designed device.

After the positive outcomes of the sensitivity analysis, a series of experiments were performed on ex vivo human cervical tissue. The experiments aimed to combine information from rheometry and TWE providing two sources of data to find the most suitable rheological model to fit the cervical tissue behavior.

Figure 12 shows the dispersion curves of shear wave speed regarding the frequency. Since both rheometry and TWE techniques obtain shear wave speed values at different ranges of frequency, rheological models are required to combine the dispersive data. In this study, the classical models of Maxwell, Kelvin–Voigt, Zener and Kelvin–Voigt Fractional Derivative were employed to fit the dispersive data. As proposed by Lin et al. [44], three types of fitted dispersion curves were obtained depending the data: (a) by using only the data from low-frequency regime generated by rheometry, (b) by using only the data from the high-frequency data provided by TWE, and (c) by fitting the data from both low and high-frequency ranges.

Whereas the fitted curves obtained by using the TWE data were overall in good agreement with the data from rheometry (Figure 12b), the curves using the rheometry data diverged from those using TWE (Figure 12a). Additionally, the dispersion curve after fitting the whole set of data, this is rheometry + TWE (Figure 12c), practically coincided with the curve generated by using only the TWE data. These findings agree with the observations by Lin et al. [44], and point out the relevancy of the viscous effects at higher frequencies. Furthermore, it uncovers the difficulty of characterizing the viscous effects of soft tissue by limiting the experiments to classical rheometry techniques.

When comparing the four rheological models, it can be observed that the Maxwell model showed difficulties to represent the dispersion of shear wave speed along the whole range of frequency. The Zener model adjusted the data successfully when considering both the rheometry and the

TWE outcomes. However, when fitting only data from TWE, it showed some divergence at the low-frequency range.

On the other hand, Kelvin–Voigt and its fractional derivative version matched all the data satisfactorily from the high to the low-frequency regime. The fact that both models produced practically the same dispersive curve can be understood by observing Tables 2 and 3. The values obtained for the α parameters of the Kelvin–Voigt Fractional Derivative model were close to 1, which transforms the model directly into a classical Kelvin–Voigt. A two-parameter model as the Kelvin–Voigt usually will be preferred against a three-parameter model as the fractional version, due to practicality and faster computation. Further studies must be performed to analyze the consistency of all the findings above shown.

5. Conclusions

A TWE technique for characterizing the viscoelastic properties of soft tissue, and, in particular cervical tissue, is proposed. The shear wave curve dispersions for cervical tissue produced by the TWE method were in good agreement with the data obtained from rheometry. However, these results from rheometry were not able to reproduce the viscous effect on the speed dispersion at high frequencies. Further studies must be carried out to analyze the consistency of these observations. The Kelvin–Voigt model and its fractional derivative version fit the cervix experimental data along the whole range of frequency successfully, and this is considering rheometry and TWE.

The results obtained in this study pave the way to carry out a cross sectional study, which aims the correlation between the gestational age and the viscoelastic properties of the cervix during pregnancy.

Author Contributions: Antonio Callejas, Guillermo Rus and Juan Melchor conceived and designed the experiments. Antonio Callejas, Miguel Riveiro, Paloma Massó and Jorge Torres performed the experiments. Antonio Callejas, Miguel Riveiro, Antonio Gomez and Modesto López analyzed the data. Antonio Callejas, Guillermo Rus, Antonio Gomez, Juan Melchor, Miguel Riveiro wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SE	Static elastography
DE	Dynamic elastography
TWE	Torsional wave elastography
3DMMRE	3D multifrequency MR elastography
FEM	Finite Element Models
FEAP	Finite Element Analysis Program
M	Maxwell rheological model
GM	Generalized Maxwell rheological model
KV	Kelvin–Voigt rheological model
KVFD	Kelvin–Voigt Fractional Derivative rheological model
Z	Zener rheological model
ERDF	The European Regional Development Fund

References

1. Berghella, V. Post-term pregnancy. In *Obstetric Evidence Based Guidelines*; Informa: Philadelphia, PA, USA, 2007; p. 183.
2. Bercoff, J.; Tanter, M.; Fink, M. Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2004**, *51*, 396–409.
3. Fruscalzo, A.; Londero, A.; Schmitz, R. Quantitative cervical elastography during pregnancy: Influence of setting features on strain calculation. *J. Med. Ultrason.* **2015**, *42*, 1–8.
4. Feltovich, H.; Hall, T. Quantitative imaging of the cervix: Setting the bar. *Ultrasound Obstet. Gynecol.* **2013**, *41*, 121–128.
5. Maurer, M.; Badir, S.; Pensalfini, M.; Bajka, M.; Abitabile, P.; Zimmermann, R.; Mazza, E. Challenging the in-vivo assessment of biomechanical properties of the uterine cervix: A critical analysis of ultrasound based quasi-static procedures. *J. Biomech.* **2015**, *48*, 1541–1548.
6. Fruscalzo, A.; Schmitz, R. Quantitative cervical elastography in pregnancy. *Ultrasound Obstet. Gynecol.* **2012**, *40*, 612–613.
7. Bamber, J.; Cosgrove, D.; Dietrich, C.; Fromageau, J.; Bojunga, J.; Calliada, F.; Cantisani, V.; Correas, J.; D'Onofrio, M.; Drakonaki, E.; et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography. Part 1: Basic Principles and Technology. *Ultraschall Med.* **2013**, *34*, 169–184.
8. Booi, R.; Carson, P.; O'Donnell, M.; Roubidoux, M.; Hall, A.; Rubin, J. Characterization of cysts using differential correlation coefficient values from two dimensional breast elastography: Preliminary study. *Ultrasound Med. Biol.* **2008**, *34*, 12–21.
9. Muller, M.; Gennisson, J.; Deffieux, T.; Tanter, M.; Fink, M. Quantitative viscoelasticity mapping of human liver using supersonic shear imaging: Preliminary in vivo feasibility study. *Ultrasound Med. Biol.* **2009**, *35*, 219–229.
10. Peralta, L.M. Feasibility of Using Ultrasonic Shear Waves to Assess Cervical Remodelling during the Gestation Period. Ph.D. Thesis, Civil Engineering, Universidad de Granada, Granada, Spain, May 2015.
11. Fahey, B.; Palmeri, M.; Trahey, G. Frame rate considerations for real-time abdominal acoustic radiation force transient imaging. *Ultrason. Imaging* **2006**, *28*, 193–210.
12. Palmeri, M.; Nightingale, K. On the thermal effects associated with radiation force imaging of soft tissue. *Ultraschall Med.* **2007**, *51*, 551–565.
13. Walden, A.; Howarth, T. Torsional Shear Wave Transducer. U.S. Patent 5,321,333 A, 14 June 1994.
14. Ouared, A.; Montagnon, E.; Cloutier, G. Generation of remote adaptive torsional shear waves with an octagonal phased array to enhance displacements and reduce variability of shear wave speeds: Comparison with quasi-plane shear wavefronts. *Phys. Med. Biol.* **2015**, *60*, 8161–8185.
15. Melchor, J.; Rus, G. Torsional ultrasonic transducer computational design optimization. *Ultrasonics* **2014**, *54*, 1950–1962.
16. Peralta, L.; Molina, F.; Melchor, J.; Gómez, L.F.; Massó, P.; Florido, J.; Rus, G. Transient elastography to assess the cervical ripening during pregnancy: A preliminary study. *Ultraschall Med. Eur. J. Ultrasound* **2017**, *38*, 395–402.
17. Catheline, S.; Gennisson, J.; Delon, G.; Fink, M.; Sinkus, R.; Abouelkaram, S.; Culioli, J. Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach. *J. Acoust. Soc. Am.* **2004**, *116*, 3734–3741.
18. Chen, S.; Urban, W.; Pislaru, C.; Kinnick, R.; Zheng, Y.; Yao, A.; Greenleaf, J. Shear wave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2009**, *56*, 55–62.
19. Vappou, J.; Maleke, C.; Konofagou, E. Quantitative viscoelastic parameters measured by harmonic motion imaging. *Phys. Med. Biol.* **2009**, *54*, 3579–3594.
20. Orescanin, M.; Qayyum, M.; Toohey, K.; Insana, M. Complex shear modulus of thermally-damaged liver. In Proceedings of the 2009 IEEE International Ultrasonics Symposium (IUS), Rome, Italy, 20–23 September 2009; pp. 127–130.
21. Zhang, M.; Castaneda, B.; Wu, Z.; Nigwekar, P.; Joseph, J.; Rubens, D.; Parker, K. Congruence of imaging estimators and mechanical measurements of viscoelatic properties of soft tissues. *Ultrasound Med. Biol.* **2007**, *33*, 1617–1631.

22. Sack, I.; Beierbach, B.; Wuerfel, J.; Klatt, D.; Hamhaber, U.; Papazoglou, S.; Martus, P.; Braun, J. The impact of aging and gender on brain viscoelasticity. *NeuroImage* **2009**, *46*, 652–657.

23. Balocco, S.; Basset, O.; Courbebaisse, G.; Boni, E.; Frangi, A.; Tortoli, P.; Cachard, C. Estimation of the viscoelastic properties of vessel walls using a computational model Doppler ultrasound. *Phys. Med. Biol.* **2010**, *55*, 3557–3575.

24. Orescanin, M.; Insana, F. Shear modulus estimation with vibrating needle stimulation. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2010**, *57*, 1358–1367.

25. Gennisson, J.; Muller, M.; Ami, O.; Kohl, V.; Gabor, P.; Musset, D.; Tanter, M. Shear wave elastography in obstetrics: Quantification of cervix elasticity and uterine contraction. In Proceedings of the 2011 IEEE International Ultrasonics Symposium (IUS), Orlando, FL, USA, 18–21 October 2011; pp. 2094–2097.

26. Peralta, L.; Mourier, E.; Richard, C.; Charpigny, G.; Larcher, T.; Ait-Belkacem, D.; Balla, N.; Brasselet, S.; Tanter, M.; Muller, M.; et al. In Vivo Evaluation of Cervical Stiffness Evolution during Induced Ripening Using Shear Wave Elastography, Histology and 2 Photon Excitation Microscopy: Insight from an Animal Model. *PLoS ONE* **2015**, *10*, e0133377.

27. Kiss, M.; Hobson, M.; Varghese, T.; Harter, J.; Kliewer, M.; Hartenbach, E.; Zagzebski, J. Frequency-dependent complex modulus of the uterus: Preliminary results. *Phys. Med. Biol.* **2006**, *51*, 3683–3695.

28. Jiang, X.; Asbach, P.; Streitberger, K.; Thomas, A.; Hamm, B.; Braun, J.; Sack, I.; Guo, J. In vivo high-resolution magnetic resonance elastography of the uterine corpus and cervix. *Eur. Soc. Radiol.* **2014**, *24*, 3025–3033.

29. Rus, G.; Munoz, R.; Melchor, J.; Molina, R.; Callejas, A.; Riveiro, M.; Massó, P.; Torres, J.; Moreu, G.; Molina, F.; et al. Torsion ultrasonic sensor for tissue mechanical characterization. In Proceedings of the IEEE International 2016 Ultrasonic Symposium (IUS), Tours, France, 18–21 September 2016.

30. Melchor, J.; Muñoz, R.; Rus, C. Torsional ultrasound sensor optimization for soft tissue characterization. *Sensors* **2017**, *17*, 1402.

31. Taylor, R. *FEAP-Ein Finite Element Analysis Program;* Ing.-Gemeinschaft Klee & Wrigges: Laatzen, Germany 1987.

32. Kim, J.; Kwon, O. Vibration characteristics of piezoelectric torsional transducers. *J. Sound Vibr.* **2003**, *264*, 453–473.

33. Klatt, D.; Hamhaber, U.; Asbach, P.; Braun, J.; Sack, I. Noninvasive assessment of the rheological behaviour of human organs using multifrequency MR elastography: A study of brain and liver viscoelasticity. *Phys. Med. Biol.* **2007**, *52*, 7281–7294.

34. Chen, S.; Fatemi, M.; Greenleaf, J. Quantifying elasticity and viscosity from measurement of shear wave speed dispersion. *J. Acoust. Soc. Am.* **2004**, *115*, 2781–2785.

35. Macosko, C.; Ronald, L. *Rheology: Principles, Measurements and Applications;* Wiley-VCH: New York, NY, USA, 1994; 550p.

36. Eldred, L.; Baker, W.; Palazotto, A. Kelvin-voigt vs fractional derivative model as constitutive relations for viscoelastic materials. *AIAA J.* **1995**, *33*, 547–550.

37. Urban, M.; Nenadic, I.; Mitchell, S.; Chen, S.; Greenleaf, J. Generalized response of a sphere embedded in a viscoelastic medium excited by an ultrasonic radiation force. *J. Acoust. Soc. Am.* **2011**, *130*, 1133–1141.

38. Amador, C.; Kinnick, R.; Urban, M.; Fatemi, M.; Greenleaf, J. Viscoelastic tissue mimicking phantom validation study with shear wave elasticity imaging and viscoelastic spectroscopy. In Proceedings of the 2015 IEEE International Ultrasonics Symposium, Taipei, Taiwan, 21–24 October 2015; pp. 1–4.

39. Carcione, J. *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic and Porous Media;* Elsevier Science: Oxford, UK, 2001.

40. Malkin, A.; Isayev, A.I. *Rheology: Concepts, Methods, and Applications;* Elsevier: Amsterdam, The Netherlands; ChemTec Publishing: Toronto, ON, Canada, 2011; 528p.

41. Chan, R.; Titze, I. Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results. *J. Acoust. Soc. Am.* **1999**, *106*, 2008–2021.

42. Campos, F.; Bonhome, A.; García, L.; Durán, J.; López, M.; Alaminos, M.; Sánchez, M.; Carriel, V. Ex vivo characterization of a novel tissue-like cross-linked fibrin-agarose hydrogel for tissue engineering applications. *Biomed. Mat.* **2016**, *11*, 055004.

43. Rodríguez, I.; López, M.; Oliveira, A.; Sánchez, M.; Campos, A.; Alaminos, M.; Durán, J. Rheological characterization of human fibrin and fibrin-agarose oral mucosa substitutes generated by tissue engineering. *J. Tissue Eng. Regen. Med.* **2012**, *6*, 636–644.

44. Lin, H.; Shen, Y.; Chen, X.; Zhu, Y.; Zheng, Y.; Zhang, X.; Guo, Y.; Wang, T.; Chen, S. Viscoelastic properties of normal rat liver measured by ultrasound elastography: Comparison with oscillatory rheometry. *Biorheology* **2016**, *53*, 193–207.

45. Bernal, M.; Gennisson, J.L.; Flaud, P.; Tanter, M. Correlation between classical rheometry and supersonic shear wave imaging in blood clots. *Ultrasound Med. Biol.* **2013**, *39*, 2123–2136.

46. Man, N.; Shiwei, Z.; Jean-Luc, R.; Vijay, S.; Hua, X. Development of oil-in-gelatin phantoms for viscoelasticity measurement in ultrasound shear wave elastography. *Ultrasound Med. Biol.* **2013**, *40*, 168–176.

Model-Based Real-Time Non-Rigid Tracking

Sebastián Bronte, Luis M. Bergasa * , Daniel Pizarro and Rafael Barea

Electronics Department, University of Alcalá, Campus Universitario, 28805 Alcalá de Henares, Spain; sebastian.bronte@depeca.uah.es (S.B.); daniel.pizarro@uah.es (D.P.); rafael.barea@uah.es (R.B.)

* Correspondence: luism.bergasa@uah.es

Abstract: This paper presents a sequential non-rigid reconstruction method that recovers the 3D shape and the camera pose of a deforming object from a video sequence and a previous shape model of the object. We take PTAM (Parallel Mapping and Tracking), a state-of-the-art sequential real-time SfM (Structure-from-Motion) engine, and we upgrade it to solve non-rigid reconstruction. Our method provides a good trade-off between processing time and reconstruction error without the need for specific processing hardware, such as GPUs. We improve the original PTAM matching by using descriptor-based features, as well as smoothness priors to better constrain the 3D error. This paper works with perspective projection and deals with outliers and missing data. We evaluate the tracking algorithm performance through different tests over several datasets of non-rigid deforming objects. Our method achieves state-of-the-art accuracy and can be used as a real-time method suitable for being embedded in portable devices.

Keywords: SfT; tracking; PTAM; non-rigid reconstruction; SfM; NRSfM

1. Introduction

The problem of 3D reconstruction and camera localization from images is known as Structure-from-Motion (SfM). 3D awareness from visual cues is a natural task for a human being, but it is still a very challenging problem in computer vision. During the last few decades, SfM has been widely studied [1–7]. Thanks to modern computing capabilities (multicore CPUs, GPUs, etc.), current SfM algorithms are considered mature and capable of dealing with big volumes of data in real time [8,9].

The general assumption in SfM is based on the rigidity of the scene, where changes in the images are caused by the relative motion between the camera and the scene. Rigidity strongly links camera motion with image motion, making SfM a well-posed problem. Rigid SfM fails in scenarios where the rigidity assumption is violated. For instance, it fails to reconstruct scenes with multiple objects that move independently or with deformable objects, such as the human body, articulated objects, wires, flags, sheets, flesh, fabric, etc.

Reconstruction of deformable objects from images is known as Non-Rigid Structure from Motion (NRSfM). It has been actively studied in recent years [10–15]. When a prior model of the object is available, this problem is commonly known as Shape from Template (SfT) [16–19]. Current SfT/NRSfM methods lack the level of maturity of SfM.

In both SfM and NRSfM/SfT, we distinguish two categories of methods: batch approaches [10–13,20] and sequential approaches [21–24]. In the former, all data (images) are available beforehand and are jointly processed to obtain the 3D reconstruction. Those approaches are highly demanding in terms of processing time and memory, but achieve accurate reconstructions. In the second group, data are collected and processed online. This is usually a harder problem where less data are available for reconstruction than in batch approaches, resulting in less accurate reconstructions. However, sequential

methods are the base for real-time implementations required in some interesting applications, such as Augmented Reality (AR) [25–27].

3D reconstruction from images is a key technology in many applications such as Human-Machine Interfaces (HMI), Augmented Reality (AR) and robotics [4,8,27]. These applications require accurate and stable reconstructions. Robust, sequential and real-time reconstruction methods are thus a priority in most of these applications. In addition, new methods must be able to deal with real conditions such as the data association problem between images, which includes outliers and missing data.

2. Related Works

This section includes a summary and discussion of the most recent methods for 3D reconstruction from images. Section 2.1 discusses the SfM methods. The SfT and NRSfM methods are presented in Sections 2.2 and 2.3 respectively.

2.1. Structure-From-Motion

Structure-from-motion can be defined as the problem of jointly inferring the 3D geometry of the scene and camera motion using images as inputs. The rigidity of the scene is a prior condition for SfM. The geometry of multiple views of a rigid scene has been known for centuries, and the basic results for solving SfM are well known in photogrammetry and computer vision [1]. Modern SfM methods cope with large sequences from both uncalibrated and calibrated cameras.

Sequential approaches in SfM are very important in robotics and were mainly developed as solutions to the visual Simultaneous Localization and Mapping (vSLAM) problem, where the robot pose and a map of the 3D environment are sequentially obtained from a camera mounted on a mobile robot. Given the improvements of modern SfM algorithms, vision-based SLAM quickly became predominant in robotics and is usually fused with other sensors installed in the robot.

One of the first monocular vSLAM systems was proposed in [4]. The SLAM problem was posed as a sequential Bayesian inference problem, using the Extended Kalman Filter (EKF) as the inference core and a sparse set of salient feature tracks as image observations. This method was implemented in real time using a low cost hardware and was suitable to reconstruct small- to medium-sized environments.

Years later, another sparse SLAM method was proposed, based on sequential Bundle Adjustment (BA) and known as Parallel Tracking and Mapping (PTAM) [8]. This method showed very accurate and stable pose estimation and reconstructions, suitable for AR applications. It remarkably improved vSLAM methods based on statistical filtering. PTAM heavily influenced modern SLAM methods and changed the SLAM processing paradigm from pure sequential to a parallel mapping and tracking algorithm.

Recently, dense SfM methods have been proposed, such as Dense Tracking and Mapping (DTAM) [9], further extended to depth sensors, such as Kinect fusion [28]. SfM methods fail to recover the shape of non-rigid objects. This motivates the methods discussed next.

2.2. Shape-From-Template

Shape-from-Template (SfT) methods estimate the 3D shape of a non-rigid object from a single image and a template of the object. A template consists of the reference shape of the object, a texture map and a deformation model. The template can be computed from samples or even generated by other reconstruction algorithms from a training sequence. The objective of SfT is to find the deformed 3D shape given a single image or a sequence of images and the object template. There are two main categories of SfT methods based on the deformation model: statistics-based and physics-based.

Regarding the first group, the most commonly-known algorithm is represented by the Active Appearance Models (AAM), which are widely used for the tracking of the human face. The first AAM method was introduced by [29]. In the context of non-rigid reconstructions, the AAM evolved into the so-called 3D Morphable Models (3DMM) [30]. A revisiting example of these models including an extensive learning and their application over a database of 10,000 faces is studied in [31].

AAMs and 3DMMs reconstruct objects that belong to a specific category whose deformations and appearance are low dimensional. This makes them very suitable for 3D reconstruction of face gestures. They need large labeled databases for a proper training of the shape and appearance models. Hand-made labeled databases are prone to errors, not only due to the "human factor", but also as different people do this task differently. In [32], the authors take advantage of some of the approaches and dense correspondences to infer the point correspondences and predict the labeling.

Most of remaining SfT methods are based on physics-based deformation priors and in particular the isometric model [17,33,34]. The work in [17] describes the problem as a Partial Differential Equations (PDE) system and proves that imposing the isometry prior makes SfT a well-posed problem. Methods in SfT can be divided into local solutions, mainly based on solutions of a PDE system, and global solutions, based on convex relaxations of isometry [18,19].

The registration between the template and the model is needed in SfT. In most of the SfT papers, the registration is assumed to be known or obtained with some automatic methods, like in [35], based on graph matching, or dense optical flow methods, such as [36]. These methods are however not suitable for real-time applications.

2.3. Non-Rigid Structure from Motion

In NRSfM, the objective is to recover the 3D shape of an object under deformations from a sequence of images. Each image captures the combination of rigid motion and shape change in the object. Thus, rigid SfM is not applicable in this case. NRSfM is ill-posed, unless priors on the possible deformations are imposed. According to the deformation prior, existing NRSfM methods can be divided into two main groups: (i) physics-based and (ii) statistics-based methods.

In physics-based models, the deformation model is taken from the field of continuum mechanics, and it models how materials behave under the action of forces. The most popular models used in this category are the isometric model [14–16,34,37] and the elastic (linear and non-linear) model [23,38–40]. The isometric model has been thoroughly studied, and [15] proved that isometric NRSfM is a well-posed problem. However, isometric priors are not accurate to describe deformations suffered by soft materials. Elastic models have been proposed in [23,38–40], using the Finite Element Modeling (FEM) approximations of linear and non-linear elastic materials. They estimate both camera pose and the 3D shape of deformable objects from monocular scenes in real time. These methods require in general object-dependent physical parameters, except for [39] in which the deformation modes are estimated sequentially. Besides, existing NRSfM is not well posed with elastic models as it has been empirically confirmed that these methods require additional constraints to limit ambiguities.

Some methods assume that the shape of the object is composed of pieces, following simple local deformation models such as rigid or quadratic [41,42]. The main challenge in these approaches is how to split the data into a set of models automatically and how to preserve global coherence of the model.

In statistics-based approaches, the object shape space is assumed to be low-dimensional and is represented as the weighted sum of a set of basis shapes. This idea was first proposed by [10], and it is based on the assumption that the shape matrix is low-rank. This prior has been studied by many researchers in the NRSfM literature [11,13,43,44]. It has proven to be successful in many real-world scenarios such as in the gestures of the human face. NRSfM based on the low-rank model is ill-posed, as the solution space is very ambiguous. Some methods have added priors, such as temporal smoothness or spacial smoothness [11], physical priors such as limiting stretching or extension in the surface [33,34] or the use of trajectory bases [45]/shape-trajectory bases [20], which constrain the trajectory of point tracks along the sequence.

Most of aforementioned methods are based on tracking a sparse set of image correspondences, recovering a sparse 3D model of the object. Dense NRSfM approaches have been recently investigated. The first one is [43], which performs dense optical flow combined with low-rank modeling and local smoothness priors. In [46], the segmentation and reconstruction of local rigid models is proposed. On the other hand, Kinect fusion was further extended to handle deformable objects in dynamic

fusion [47] for depth cameras. However, these approaches need the use of specialized hardware (GPU-based) to be run in real time.

The majority of the NRSfM methods use orthographic projection. This is an advantage for statistics-based algorithms [10–12,20,21,44], as it allows one to pose NRSfM as a low-rank matrix factorization problem, which counts with global and efficient solvers. However, the orthographic camera can be very inaccurate for images taken under certain imaging conditions. It also includes ambiguities in the reconstructions due to the convex/concave shape flips and the depth ambiguity problem. Some NRSfM methods include the perspective camera such as [22,48–50], which is a more accurate model. However, it makes the problem harder to solve.

2.4. Proposal

This paper proposes a sequential solution to SfT able to be run in real time (e.g., 15–30 frames per second) using a low-cost hardware based on a CPU. Our algorithm is efficient and very suitable for embedded systems. These features are available in late SfM methods [51], but not in the context of non-rigid reconstruction.

We take PTAM, a state-of-the-art sequential real-time SfM engine, and we upgrade the tracking thread to solve SfT using a low-rank shape basis model. As in PTAM, we use a calibrated perspective camera model. As previously mentioned, the perspective camera is an accurate model and does not suffer from some ambiguities present in the orthographic model, widely used for statistics-based non-rigid reconstruction methods.

Most of the current state-of-the-art non-rigid reconstruction algorithms do not compute the tracking and data association, whereas in a real applications, this is one of the most important challenges to face. Our solution deals with data association in real time. This paper extends the work published by the authors in [52] where a preliminary version of our tracking algorithm was presented. In this work, an improved version is implemented, and a deeper analysis is carried out. We hereinafter refer to this method as the IROS-based approach.

This paper is organized as follows: The description of our method is given in Section 3. In Section 3.1, we describe the data-association model, and in Sections 3.2 and 3.3, we describe the feature matching method and the camera motion models, respectively. We present our shape inference method in Sections 3.4 and 3.5. In Section 4, a summary of the results is described. Finally, Section 5 presents the main conclusions of the paper.

3. Algorithm Description

Our algorithm estimates the camera pose and the shape of the object in each frame of a video sequence. Image cues are based on tracking a sparse set of salient features over the images. To that end, it needs an initial estimation of the object's rigid shape, the camera pose and the object shape basis. We refer to this method as the tracking thread, as it is based on the tracking thread of PTAM. A flowchart of the tracking thread and the initialization process is presented in Figure 1.

To model the deformations, we use a linear shape basis model [10]. The shape of the object can be expressed as the weighted sum of a finite number of shape vectors:

$$S_{NR}(f) = S_R + \sum_{k=1}^{K} L_k(f) B_k \tag{1}$$

where f represents the frame number, K is the number of shapes given in the model, $S_{NR}(f) \in \mathbb{R}^{3 \times P}$ is the computed shape for frame f and P represents the number of points of the shape. $S_R \in \mathbb{R}^{3 \times P}$ is the rigid shape average. $L(f) \in \mathbb{R}^K$ are the shape coefficients, and $B \in \mathbb{R}^{3K \times P}$ are the basis shapes, which were previously known.
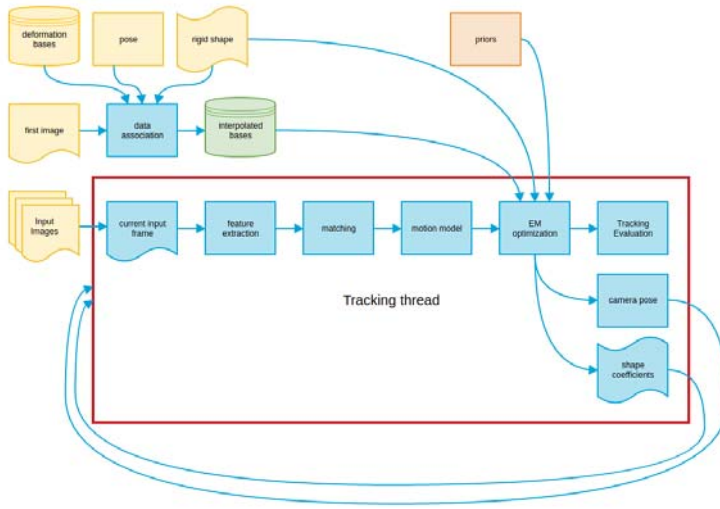
Figure 1. Tracking thread flowchart. The main modules of our tracking thread are included inside the red box. The yellow modules represent inputs; the orange modules are configurations (priors); the green ones are estimations from the input data; and finally, the blue ones are implemented tasks.

The transformation from world to camera coordinates for each point is given by:

$$X_i = \begin{pmatrix} x_i & y_i & z_i \end{pmatrix}^T = [R|T] \times S_{NR_i} \tag{2}$$

where the camera rotation is given by $R \in \mathbb{R}^{3\times3}$ and the camera translation is represented by $T \in \mathbb{R}^3$. Each 3D point in camera coordinates is projected using the perspective camera model:

$$U_i^S = \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \alpha \begin{pmatrix} f_u & 0 \\ 0 & f_v \end{pmatrix} \begin{pmatrix} x_i/z_i \\ y_i/z_i \end{pmatrix} \tag{3}$$

where $U^S \in \mathbb{R}^{2\times P}$ represents the set of shape points, $(u_0, v_0)^T$ are the camera center coordinates, (f_u, f_v) is the focal length and α is the radial distortion function. These parameters are estimated in an offline calibration process described in [8].

3.1. Measurement Model/Data Association

The vertices of the shape model S_{NR} do not match in general with the points detected in the image using feature detectors described in Section 3.2.2. To establish a link between the shape model and the measurements, we use a triangular mesh model. The mesh is estimated by using the Delaunay triangulation on the projected coordinates of our shape model (S_{NR}) in the first frame of the sequence U_0^S. This gives a mesh of connected vertices v_i. This mesh is fixed for the rest of the sequence. Given a point detected in the image $U_i' = (u_i' v_i')^T$, we find the mesh triangle it belongs to (the three vertices of the triangle $v_{i,1..3}$) and compute the barycentric coordinates (a_i, b_i, c_i), following [53]:

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = pinv \begin{pmatrix} v_{i,1x} & v_{i,2x} & v_{i,3x} \\ v_{i,1y} & v_{i,2y} & v_{i,3y} \end{pmatrix} \times \begin{pmatrix} u_i' \\ v_i' \end{pmatrix} \tag{4}$$

where $pinv(\cdot)$ is the matrix pseudo-inverse operator. An example of the data association is represented in Figure 2, where the projected mesh vertices of the triangle are represented as black points, and the detected point is represented as a red point inside the triangle. The barycentric coordinates are displayed as edges connecting the detected point and the vertices of the triangle.
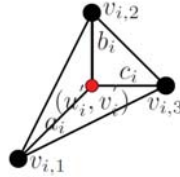

Figure 2. Data association from detected points (red) with model points in mesh (black).

We can define a new deformation basis B', adapted to the detected points and defined in world coordinates, by using the barycentric coordinates. The 3D position of the interpolated model point X'_i that corresponds to the 2D feature U'_i relates to the model basis shapes as:

$$X'_i = a_i X_{v1} + b_i X_{v2} + c_i X_{v3} \tag{5}$$

$$X'_i = \sum_k L_k \left(a_i B_{v1k} + b_i B_{v2k} + c_i B_{v3k} \right) = \sum_k L_k B'_{ik} \tag{6}$$

where X'_i is the interpolated point corresponding to the detected point U'_i and the triangulation defined by the barycentric coordinates (a_i, b_i, c_i) and X_{v1}, X_{v2}, X_{v3}, which are the 3D coordinates of the 2D associated coordinates vertices $(v_{i,1}, v_{i,2}, v_{i,3})$ and $(B_{v1k}, B_{v2k}, B_{v3k})$, are k-th bases of each of the vertices of the triangle. Equation (6) shows the duality of working with model or detected points and interpolated 3D points, once the data association is set.

3.2. Feature Matching

We detect and match point features between the reference image and the rest of the images in the video sequence. The quality of feature detection and matching is crucial for obtaining accurate reconstructions. To achieve real-time performance, we face a trade-off between the quality of matching and the complexity of the existing feature detection and matching methods.

Two different feature detection and matching methods are studied. The first one is based on PTAM, which uses the FAST detector and was presented by the authors in [52]. This is considered as the baseline for the comparison. A second approach is based on descriptor matching and uses different state-of-the-art visual detectors and descriptors.

3.2.1. PTAM-Based Matching Approach

Hereinafter, we assume that the points detected in the image (U'), e.g., using the FAST [54] detector, correspond to the projection of the points described by the interpolated bases B'. This saves computation time and makes equations simpler.

We use the same process described in [8] to track points in the video sequence. However, due to shape deformations, the appearance in a local region around each point can suffer variations. We thus propose the following adaptations:

- Affine warping described in [8] is kept to handle points whose local appearance is affected by a rigid motion.
- If the deformation causes significant changes in the appearance, affine warping is expected to fail. Then, we search features in a coarse-to-fine hierarchical correlation-based approach. To discard false positives, a married matching is applied in the lowest level of the pyramid in which the

feature has been found. In case the correlation is too low or the distance is too high, the matching is considered not found. For a deeper explanation, we refer the readers to our publication in [52].

This algorithm could fail for points with a highly deformed texture or deformations that imply large movements between images. Rejected points are considered as outliers and are discarded. This prevents tracking from degenerating in the following frames.

3.2.2. Descriptor-Based Matching Approach

The detection and matching layer of the IROS-based approach is substituted by descriptor-based features implemented in OpenCV [55]. Some of them are directly available, like KAZE [56], AKAZE [57], ORB [58] and BRISK [59], and others are not directly included, but can be accessed as third party software for research purposes, like SIFT [60] and SURF [61]. Binary descriptors, such as ORB, AKAZE and BRISK, are preferred for a real-time implementation as they are faster.

Once the features are detected and described, we match them across frames. Several matching algorithms are used such as: brute force, L1, L2 and Hamming distance. These techniques are prone to establish incorrect matches.

Matches are searched inside a circular area defined in the image domain. The criterion we follow is to select matches of the minimum descriptor distance that do not violate the maximum radius condition. The radius parameter is configurable; thus, it can be tuned if the number of tracking failures is high.

Feature descriptors, such as SIFT or KAZE, are not completely affine invariant and, thus, can fail with high deformations. Specialized feature descriptors exist in the literature for deformable registration, such as in [62,63]. We assume in this paper that strong deformations are not present in local patches, so these options are not considered.

3.3. Motion Modeling

A linear decaying motion model is used to improve camera tracking convergence. The pose is updated using the following motion model:

$$
\begin{aligned}
vel_t &= \beta/2 \left(vel_{t-1} + \mu\right) \\
\left[\widehat{R|T}\right]^t &= exp\left(vel_t \triangle t\right) \left[R|T\right]^{t-1}
\end{aligned}
$$

(7)

In order to apply the equation, the camera speed vel_t, between frames $t-1$ and t, is computed by applying the ESMhomography method proposed in [64]. β is the factor that balances the influence of the estimated camera speed on the camera motion update vector μ. This motion model was proposed in the original PTAM algorithm [8], where the shape was assumed to be rigid. We adapt it by using the updated shape obtained in the last frame, as will be explained in Section 3.4.

3.4. EM Optimization

We present in this section the shape-inference algorithm that estimates the shape deformation coefficients and the camera pose of the current frame. We use an optimization approach where we minimize the reprojection error between the detected points and the deformed shape projected using the perspective camera. This is a non-convex cost that cannot be solved in closed form. Previous approaches, like [11,24,65,66], implement Expectation-Maximization (EM) algorithms to split the optimization of a similar cost function into two parts: deformation weights, keeping the camera parameters fixed and vice versa. This strategy is easy to implement and has faster convergence time, contributing to the real-time constraint. We have adapted this method to our problem.

The Maximum Likelihood Estimation (MLE) function is defined as,

$$E_{data} = f_{MLE}(R, T, L) \propto \sum_i \left| U'_i - proj\left(X'_i\right) \right|^2 \tag{8}$$

where E_{data} represents the 2D reprojection error. It is minimized w.r.t. the following two sets of parameters:

- Pose parameters (6 DoF): rotation R and translation T defined by the vector of parameters $\mu = [\phi_x, \phi_y, \phi_z, t_x, t_y, t_z]$.
- Deformation parameters (K DoF): the set of K shape deformation coefficients L.

Both sets of parameters are grouped in the state vector $\theta = \begin{bmatrix} \mu & L \end{bmatrix}$, which has $6 + K$ DoF. The matrices B' and S_R are assumed to be fixed during the sequence. A maximum of ten iterations is used for each frame in order to estimate camera pose and deformation coefficients by using the alternation given in the EM algorithm. The actual iterations are obtained checking the RMS reprojection error according to Equation (9). Two consecutive increases on the error or low error reductions makes the minimization stop. Among all the estimations, the best one is kept as the solution for the current frame.

$$e_{RMS} = \sqrt{\frac{\sum_i \|e_i\|^2}{\sum_i \|U'_i\|^2}} = \sqrt{\frac{\sum_i \|U'_i - proj(X'_i)\|^2}{\sum_i \|U'_i\|^2}} \tag{9}$$

3.4.1. E-Step: Deformation Estimation

The main objective in this step is to minimize f_{MLE} as a function of the deformation parameters L, fixing the pose parameters μ. We denote by $\widehat{[R|T]}^{t-1}$ the current camera pose, derived from the previous M-step, and we recall that S_R and B' remain fixed along the process. Given:

$$e_i^L = \begin{pmatrix} u'_i \\ v'_i \end{pmatrix} - proj\left(\widehat{[R|T]}^t \sum_k L_k^{t-1} B'_{k,i} \right) \tag{10}$$

where e_i^L represents the re-projection error of the i-th point in the current frame and L^{t-1} is the current value for the deformation weights. An M-estimator using the Tukey bi-weight function [67] is used to help with outlier rejection [8]. The computed weights depend on each point reprojection error. Only matched points are processed. In order to estimate the shape coefficients, we start undistorting the projections, from Equation (**??**). The orthographic projection has as the main advantage that the factorization can be directly applied to the tracking matrix to obtain the coefficients, as the projection of the sum is the sum of the projections. However, for the perspective projection case, this does not hold, as it depends on the depth. We use Equations (2) and (**??**) to obtain the following result:

$$\begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = \begin{pmatrix} f_u\left(\sum_k L_k \vec{r}_x B'_k + t_x\right) + u_0 \lambda \\ f_v\left(\sum_k L_k \vec{r}_y B'_k + t_y\right) + v_0 \lambda \\ \sum_k L_k \vec{r}_z B'_k + t_z \end{pmatrix} \tag{11}$$

where λ is the perspective projection scale, $R = \begin{pmatrix} \vec{r}_x & \vec{r}_y & \vec{r}_z \end{pmatrix}^T$ is the rotation matrix decomposed in vectors and $T = \begin{pmatrix} t_x & t_y & t_z \end{pmatrix}^T$ is the translation vector expressed in its three components.

Grouping the terms in L_k on one side and writing the equivalent system in a Linear Least Squares (LLS) form $Ax = C$, where $x = L$, $\Delta u_p = u_p - u_0$, and $\Delta v_p = v_p - v_0$, we obtain:

$$A = \begin{pmatrix} (\Delta u_1 \vec{r}_z - f_u \vec{r}_x) B'_{11} & \cdots & (\Delta u_1 \vec{r}_z - f_u \vec{r}_x) B'_{1K} \\ (\Delta v_1 \vec{r}_z - f_v \vec{r}_y) B'_{11} & \cdots & (\Delta v_1 \vec{r}_z - f_v \vec{r}_y) B'_{1K} \\ \cdots & \cdots & \cdots \\ (\Delta u_P \vec{r}_z - f_u \vec{r}_x) B'_{P1} & \cdots & (\Delta u_P \vec{r}_z - f_u \vec{r}_x) B'_{PK} \\ (\Delta v_P \vec{r}_z - f_v \vec{r}_y) B'_{P1} & \cdots & (\Delta v_P \vec{r}_z - f_v \vec{r}_y) B'_{PK} \end{pmatrix} \tag{12}$$

$$C = \begin{pmatrix} f_u t_x - t_z \Delta u_1 \\ f_v t_y - t_z \Delta v_1 \\ \vdots \\ f_u t_x - t_z \Delta u_P \\ f_v t_y - t_z \Delta v_P \end{pmatrix} \tag{13}$$

If an average shape is given in the initialization (S_R), it must be taken out from the estimation. If this is the case, it can be described as a function of the basis shapes:

$$S_R = \sum_k L_{R_k} B_k \tag{14}$$

Then, in Equation (14), the average rigid shape S_R is described as a function of an additional set of coefficients, the rigid coefficients $L_R \in \mathbb{R}^K$. Doing $L_{NR} = L - L_R$ removes the rigid shape influence from the estimations.

If the previous condition does not hold, the rigid shape must be introduced in the deduction of the matrices A and C, as the 3D model and then the projection equations need to be adapted. The A matrix remains the same as in Equation (12), but C gets more complex, as indicated in Equation (15).

$$C = \begin{pmatrix} f_u t_x - t_z \Delta u_1 + S_{R,1} \left(f_u \vec{r_x} - \vec{r_z} \Delta u_1 \right) \\ f_v t_y - t_z \Delta v_1 + S_{R,1} \left(f_v \vec{r_y} - \vec{r_z} \Delta v_1 \right) \\ \vdots \\ f_u t_x - t_z \Delta u_P + S_{R,P} \left(f_u \vec{r_x} - \vec{r_z} \Delta u_P \right) \\ f_v t_y - t_z \Delta v_P + S_{R,P} \left(f_v \vec{r_y} - \vec{r_z} \Delta v_P \right) \end{pmatrix} \tag{15}$$

3.4.2. M-Step, Pose Estimation

The goal in this minimization step is to compute the camera pose. This is done by maximizing the likelihood of the observed data, as shown in Equation (8). In this step, the 3D shape is kept fixed, minimizing the following reprojection error expression:

$$e_i^\mu = \begin{pmatrix} u'_i \\ v_i \end{pmatrix} - proj \left(\exp \left(\mu^t \right) [\widehat{R|T}]^t \sum_k L_k^t B'_k \right) \tag{16}$$

Equations (10) and (16) are similar, but in this case, L is known from the last E-step, and μ (the camera update) is unknown. We compute μ as in [8]. Similarly to the E-step, the missing correspondences are taken out from the pose estimation on the current frame.

Since the pose parameters are not linear, the estimation cannot be done in closed form. Instead, it is based on several steps of gradient descent, for which the pose update w.r.t. the error must be obtained. To that end, Equation (??) is decomposed into several matrices to facilitate the derivation of the expression using the chain rule.

After several steps, the following expression is reached:

$$\left.\frac{\partial e_i^\mu}{\partial \mu_p}\right|_{2\times 6} = J_{cam}|_{2\times 2} \left(J_{D_i}|_{2\times 3} \begin{array}{c} 0 \\ 0 \end{array} \right) G_p|_{4\times 4, p=1..6} \, E_{CW}|_{4\times 4} \, X_W|_{4\times 1} \tag{17}$$

where $p \in 1\ldots 6$ denotes the index in the pose parameters, G_p the generation matrix of the $\mathbb{SE}(3)$ Lie group and E_{CW} is the pose matrix composed by the rotation matrix R and the translation vector T. $J_{cam}|_{2\times 2}$ is defined as follows:

$$J_{cam}|_{2\times 2} = J_A|_{2\times 2} \, J_C|_{2\times 2} \tag{18}$$

where J_A is the derivative w.r.t the 3D point in camera coordinates of Equation (**??**) (removing distortion) and J_C is related to the derivative of the distortion part of the calibration matrix w.r.t the 3D point in camera coordinates:

$$J_A = \begin{pmatrix} f_u & 0 \\ 0 & f_v \end{pmatrix} \tag{19}$$

$$J_C = \begin{pmatrix} \hat{B} & 0 \\ 0 & \hat{B} \end{pmatrix} + \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} J_{\hat{B}} \tag{20}$$

$\hat{B} = \frac{\tilde{r}}{r}$ with $r = \sqrt{C_1^2 + C_2^2}$, $\tilde{r} = r - \beta_1 r^3 - \beta_2 r^5$, $C_1 = \frac{x_c}{z_c}$, $C_2 = \frac{y_c}{z_c}$, (x_c, y_c, z_c) are the 3D coordinates of the points in camera reference system, and:

$$J_{\hat{B}} = \begin{pmatrix} -2\beta_1 C_1 - \beta_2 \left(4C_1^3 + 4C_1 C_2^2 \right) \\ -2\beta_1 C_2 - \beta_2 \left(4C_2^3 + 4C_2 C_1^2 \right) \end{pmatrix}^T \tag{21}$$

where β_1 and β_2 denote the distortion coefficients computed from camera calibration.

Lastly, J_D is the derivative of the point w.r.t. its coordinates in the camera reference:

$$J_D = \begin{pmatrix} \frac{1}{z_c} & 0 & \frac{-x_c}{z_c^2} \\ 0 & \frac{1}{z_c} & \frac{-y_c}{z_c^2} \end{pmatrix} \tag{22}$$

The Jacobian of Equation (17) is computed for each point X_i' in world coordinates (X_W) and then inserted into a WLS minimizer together with an M-estimator, for outlier rejection.

Once the update μ for the pose is obtained, as the pose belongs to the $\mathbb{SE}(3)$ Lie group, the pose is then updated as:

$$E_{CW}' = \exp\left(\mu\right) E_{CW} \tag{23}$$

3.5. Priors

In order to improve the method, we include regularization and smoothing priors in the optimization process. These tracking priors allow us to obtain better 3D error by smoothing the pose and shape solutions over the sequence.

After a revision of the related works, we found out that the most suitable priors for this tracking problem are temporal and shape smoothness, as they are easily adapted to a least squares minimization scheme in the E-step.

Some more complex priors were considered, such as the isometric prior. We discarded them as they generate non-convex costs, which do not fit well in our EM approach. Isometry is also a strong prior that cannot be used for some objects that undergo elastic deformations.

The consequence of including the priors is the modification of the error function to be minimized. Instead of relying only on the 2D reprojection error, the cost function takes into account the following terms:

$$E = E_{data} + \rho_{temp} E_{temp} + \rho_{shape} E_{shape} \tag{24}$$

E_{data} is the 2D reprojection error described in Equation (8):

$$E_{data} = \sum_{i=1}^{P} \left\| U_i' - proj\left(X_i'\right) \right\| \tag{25}$$

E_{temp} imposes temporal smoothness:

$$E_{temp} = \sum_{i=1}^{P} \left\| X_i'(f) - X_i'(f-1) \right\| \tag{26}$$

and E_{shape} is an extra restriction of the shape where deviations from the reference shape are penalized:

$$E_{shape} = \sum_{i=1}^{P} \left\| X_i(f) - \hat{X}_i(f) \right\| \tag{27}$$

where ρ_{temp} is the weight for temporal smoothness and ρ_{shape} is the weight for shape smoothness.

E_{temp} imposes a temporal dependence between the current 3D shape $X_i'(f)$ and the previous $X_i(f-1)$. Decomposing $X_i'(f) = B_i'L(f)$, the expression could be directly solved as it can be put as a least squares expression function of L.

E_{shape} forces the current shape $X_i(f)$ to be closer to the shape $\hat{X}_i(f)$, obtained by predicting each node of the mesh using its neighbors:

$$\hat{X}_i(f) = \sum_{k,m,n \in \aleph\left(\hat{X}_i\right)} \alpha_i X_k + \beta_i X_l + \gamma_i X_m \tag{28}$$

where α_i, β_i and γ_i are the barycentric coordinates of vertex i in the mesh with respect to three neighboring vertices X_k, X_l and X_m. Note that the barycentric coordinates are computed using the reference shape. This term imposes spatial smoothness by penalizing shapes with points that do not match with their neighbors. This equation can be drawn as a least square expression with Las a free parameter, so Equation (24) can be optimized in the E-step.

4. Results

Firstly, a general overview of the performance metrics is described followed by a revision of the different datasets used for the assessment of the algorithm and a detailed report of the results for each of the sequences.

4.1. Performance Metrics

This section presents the metrics used to evaluate the performance of our algorithm for different test sequences (motion captured or synthetically generated). After a sequence is processed by the algorithm, the output results are post-processed to compare the performance of the algorithm using the ground truth. The way that the sequences are adapted to our system is also explained.

We evaluate reconstruction error using the following indicators:

- 2D error:

$$2D\,err\,(px) = \frac{1}{F} \sum_{f=1}^{F} \frac{\sqrt{\sum_p \left\| x_{est}(f,p) - x_{gt}(f,p) \right\|_2^2}}{\sqrt{\sum_p \left\| x_{gt}(f,p) \right\|_2^2}} max\left(x_{gt}\right) \tag{29}$$

where F is the number of frames of the sequences, P is the number of points, $x_{est} = (u_{est}, v_{est})$ are the re-projected features on the image and $x_{gt} = (u_{gt}, v_{gt})$ represent the image coordinates of the ground-truth points.

- 3D error:

$$3D\ err\ (\%) = \frac{1}{F} \sum_{f=1}^{F} \frac{\sqrt{\sum_p \left\| S_{est}\,(f,p) - S_{gt}\,(f,p) \right\|_2^2}}{\sqrt{\sum_p \left\| S_{gt}\,(f,p) \right\|_2^2}} \times 100 \tag{30}$$

$S_{est} = (x_{est}, y_{est}, z_{est})$ is the estimated shape and $S_{gt} = (x_{gt}, y_{gt}, z_{gt})$ the ground truth shape.

The 2D reprojection error and the 3D reconstruction error were defined in [21]. When applied to an entire sequence, the error indicators are either averaged or presented individually for each frame. The former is useful to analyze the evolution of the error over time.

Regarding the 3D error, a previous Procrustes analysis [68] is carried out before applying Equation (30). It rotates and scales the shapes to align them w.r.t. the ground-truth. This is a common procedure in the state-of-the-art works, such as in [12,21]. Unless the opposite is explicitly indicated, this alignment is always applied before computing 3D errors.

4.2. Setup

In order to run the algorithm, we need a set of basis shapes, a rigid shape and an initial pose. These are estimated from the ground truth data as follows:

The rigid shape S_R is computed by extracting the average of the set of shapes used to obtain the bases. The initial pose is retrieved from the projection on the initial frame. We use PCA to retrieve the set of shape bases. The rigid shape is previously subtracted from the shapes, and no Procrustes alignment is applied before factorization. The basis shapes are then computed using PCA from this expression:

$$S - S_R = \begin{pmatrix} x_{1,1} & y_{1,1} & z_{1,1} & \cdots & x_{1,P} & y_{1,P} & z_{1,P} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{F,1} & y_{F,1} & z_{F,1} & \cdots & x_{F,P} & y_{F,P} & z_{F,P} \end{pmatrix} = UDV \tag{31}$$

After PCA is computed from this matrix, the bases ($B = \sqrt{D}V$) are obtained, but only the K most relevant components are taken. The number of bases used for all the experiments are at least $K = 15$, unless the opposite is indicated. We always ensure that the value of K represents at least 85% of the total deformation energy obtained from the PCA decomposition following the expressions in Equations (32) and (33). A higher number of bases can lead to over-fitting when dealing with real data, not to mention the memory and computational costs of handling a large number of bases. These are the main reasons why a fixed value (unless specified) of bases is set.

$$D_{normalized} = \frac{diag(D)}{\sum diag(D)} \tag{32}$$

$$D_{cumulated}[i] = \sum_{j=0}^{i} D_{normalized}[j] \ \ \forall i \tag{33}$$

It must be noted that no Procrustes alignment is applied before the SVD factorization.

To handle the data from motion captured datasets, without any kind of visual information, a special version of the tracking thread is developed, to accept as input text files containing the point projections for each frame, the 3D bases, the pose initialization, the visibility masks, etc. The ground truth 3D points are projected using the perspective camera (Equation (**??**)), choosing a convenient pose to see all the sequence points over time.

To perform a thorough comparison and to simulate real tracking conditions for these datasets, the following sets of outlier percentage values $outl = [0, 5, 10, 20, 30, 40]\%$ and noise strength values $\sigma = [0, 1, 2, 3, 4]$ are added to the set of projected points. This is similar to the evaluation proposed in [53]. With this setup, an initial evaluation is performed to check the tracking robustness under

these conditions. The added noise follows a normal distribution $N\left(0,\sigma^2\right)$. A percentage of the image points is marked as outliers. For those points both spacial directions are randomly deviated 20 pixels. Point visibility is also introduced as a variable to analyze algorithm robustness. To that end, randomly distributed visibility masks in an increasing percentage are generated.

All the experiments are computed using an i7 laptop with 8 GB RAM and eight virtual core processor at 2.4 GHz. The algorithm is implemented in C++ and runs on Ubuntu Linux, using PTAM-derived libraries, OpenMP, CGAL and OpenCV. If it is not specified, the tests for the methods of [11,12,20,21] are executed with the default parameters given by the authors. We set the motion movement constant β to 0.9; the radius limit of the IROS-based feature detection is 30 px; and the one for the rest of the matching algorithms is 70 px.

A general overview of the algorithm performance is presented with two motion capture datasets. We start evaluating the tracking with the assumption that the matching is perfect. After that, the results are re-evaluated for tracking under visibility degradation, noise, outliers and varying the number of basis shapes. Smoothing priors are not included in these experiments.

For real images, we analyze the performance by testing some matching methods, varying the number of basis shapes, and finally, we analyze the performance using the smoothing priors. Using real images implies that missing data, outliers and noise are already present in the data.

4.3. Flag Sequence

A motion captured flag bending with the wind is represented in the flag sequence. The deformations are so significant that they make this dataset challenging. Moreover, the amount of points in the sequence (540 in 450 frames projected on 640×480 images) is significantly higher than in other synthetic datasets used in the state-of-the-art, which implies a challenge to factorization-based algorithms. The deformation energy kept by the first 15 bases ($K = 15$) is 87.33%, which follows the imposed criteria of keeping 85% of the deformation energy.

4.3.1. Performance Evaluation Based on Perfect Matching

Some examples of both sequences are shown in Figure 3a. 3D reconstruction results without any degradation on the matching are shown in Figure 3b. They are compared to the ground truth, which is represented in blue, whilst the reconstruction result is represented in red. Two views are shown to better display the 3D reconstruction. For most of the frames, the reconstruction is very accurate. The average performance is shown in the row for $K = 15$ bases shown in Table 1.

In Frame #200, the reconstruction result shows more error on one of the corners, which corresponds to high frequency deformations that might not have been properly modeled. The results can be improved, in this case of perfect matching, by increasing the number of bases.

The reprojection and reconstruction errors for a perfect matching over time are shown in Figure 4. Increasing the number of bases minimizes these errors, as will be seen in Section 4.3.4. However, as will be discussed in Section 4.6.2, this trend will not be valid when correspondences are obtained with the matching of real images.

Table 1. Flag sequence error comparison with the number of bases, for K = 5, 7, 15, 25 and 30.

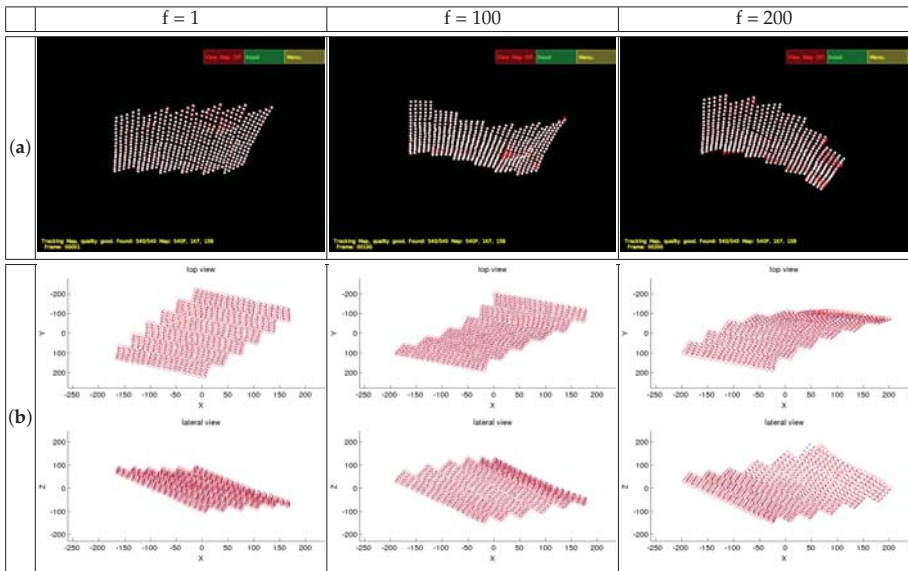| #Bases (K)	2D Error (pix)	3D Error (%)
5	4.56	3.84
7	3.62	3.21
15	2	2.63
25	1.34	2.1
30	1.18	1.93

Figure 3. Flag sequence snapshots (**a**) and reconstruction results (**b**) for certain frames.
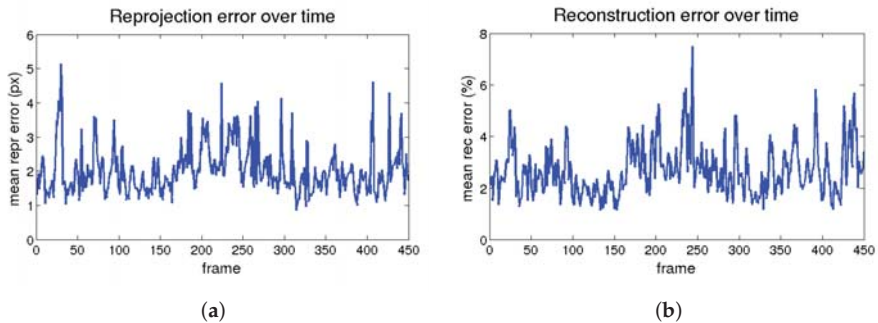

Figure 4. Flag sequence error figures through time. (**a**) The reprojection error and (**b**) the 3D reconstruction error.

4.3.2. Performance Evaluation Based on Visibility Degradation

Missing detection of features is simulated by using visibility masks. These masks are generated according to a percentage of missing data. The visibility distribution of the masks for each point and among frames follows a uniform random distribution.

Some results within a range of visibility from 5%–100% are carried out. The 2D reprojection and the 3D reconstruction errors are computed for both sequences. For the flag sequence, the errors are stable from 20–100%, which represents a minimum of 108 points. They have the following values on average: two pixels for 2D reprojection error and 2.6% for 3D reconstruction error. We highlight that the visibility distribution follows here a random uniform distribution. In real images, missing data usually affect entire areas of the image, due to occlusions or lack of texture. In some detectors, the visibility percentage varies from 50%–30%.

4.3.3. Performance Evaluation Based on Noise and Outliers

Similarly to the previous test, robustness analysis due to outliers and noise is performed. Different values for noise and outliers have been evaluated, as mentioned in Section 4.1. The experiments are repeated 10 times, and the average is taken, to achieve representative results, as a standard Monte Carlo experiment.

The results are shown in Figure 5. It can be seen that the 2D reprojection error is not significantly increased for the lowest σ values as the number of the outliers grows, meaning that the M-estimator is capable of dealing with them.For higher values of σ, the error gap between the methods is approximately the same as the number of outliers increases, the limit being $\sigma = 2$. With respect to the 3D reconstruction error, the trend is the same as in 2D reprojection error, starting at 2.6%.
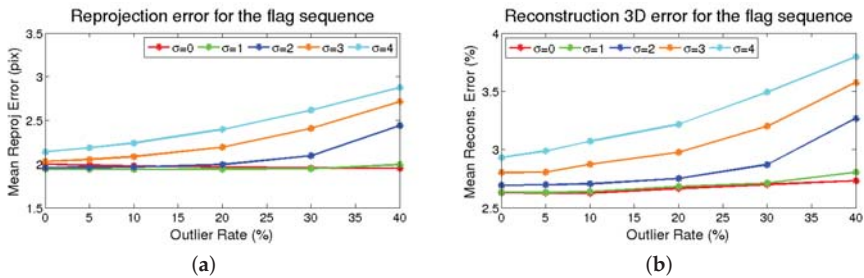

Figure 5. Flag sequence results for noise and outliers' degradation on tracking. (**a**) The column depicts the error variation with the *o* outlier points' percentage of the reprojection error. Each trace corresponds to a different noise σ level. (**b**) The column represents the same for the 3D reconstruction error.

4.3.4. Performance Evaluation Based on the Number of Bases

The dependence of the accuracy results with the number of bases is checked as well. For that purpose, a test using tracks without degradation and with 5, 7, 15, 25 and 30 bases is presented in Table 1.

Using perfect matching, as the number of bases increases, the error decreases for the presented sequence and so do all the error indicators. In addition, the processing time improves as the number of bases is reduced, as well as the memory requirements.

4.3.5. Comparison with Other Methods of the State-Of-The-Art

Table 2 shows a summary of results with the most representative methods of the state-of-the-art for the flag sequence. Processing time for the whole sequence, 3D accuracy, 2D error, rank, modeling type (auto/a priori) and procedure type (batch/sequential) are depicted. In this case, our algorithm gets close to the best ones in terms of 3D accuracy and 2D reprojection error for this sequence. In our approach, no priors are applied, in contrast to the best ones, in which physical priors such as isometry [34] or partial modeling [42] are used. Our approach can be better compared with sequential algorithms [21,24], as the estimation scheme follows a similar (sequential) approach. Regarding [21], our algorithm outperforms their 2D and 3D error. With respect to [24], our algorithm obtains comparable results using a lower amount of bases and without physical priors.

Regarding computation times, our algorithm is the best among all compared methods, thus reaching the best trade-off between performance and processing time. However, it must be highlighted that some of the competing methods are not optimized to run in real time, and some of them are implemented in MATLAB, while our method is implemented and optimized in C++.

The trials of [44,69] did not finish processing the sequence after the referenced time. For some of the methods, where the source code is available [11,12,20,21,44,69], the same hardware is used to run

the experiments. In [69], they provide results with this dataset. We present them in a second column inside the 3D error column as we found some discrepancy betweenour tests and the ones provided in the original paper.

Algorithms where the code and processing times are not available, such as the results of [34,42], are not reported in Table 2. Batch algorithms implemented in MATLAB such as [11,12,20,21,24,44,69] take more time than those developed in the C++ language, like [40]. A C++ implementation does not automatically imply real-time performance, as the code needs to be optimized, but also the data volume needs to be restricted to be properly handled in real time.

Comparing the MATLAB processing time in Tables 2 and 3, it gives an idea of the complexity of the flag dataset. Table 2 shows that current solutions based on model-free methods (NRSfM) cannot recover the correct 3D shape for the flag sequence with a non-normalized error, unless the approach includes some extra prior (we recommend to the readers the review of [69]). As our method is model-based, it provides acceptable error values and is faster than the rest of model-free approaches, reaching 70–90 fps, which is compatible with real-time constraints. It must be noted that no visual information processing is applied in any of the tested approaches.

Table 2. Flag sequence results summary. Some results comparing several state-of-the-art algorithms and the presented one are depicted.

Method	2D Error (px)	3D Error (%)	Rank	Tproc(fps)	Model	ProcType	Impl.
[21]	6.14	91.65	7 (max)	20 min (0.37)	auto	seq.	MATLAB
[20]	29.79	66.65/16.09[1]	5	36 min (0.21)	auto	batch	MATLAB
[11]	11.65	15.59/13.25[1]	4	10 h (0.01)	auto	batch	MATLAB
[12]	**1.9**	29.724/17.61[1]	9	19 seg(23.68)	auto	batch	MATLAB
[34]	-	**1.79**	-	-	auto	batch	MATLAB
[42]	-	**1.29**/3.25[1]	-	-	auto	batch	MATLAB
[69]	~2	n.a./3.8	-	»28 h	auto	batch	MATLAB
[44]	-	n.a./17.41[1]	-	»16 h	auto	batch	MATLAB
[24]	-	3.28 (K = 10) 2.81 (K = 40)	10–40	708 seg (K = 10) (0.64) 1045.5 seg (K = 40) (0.43)	auto	seq.	MATLAB
Our approach	**2**	**2.63**	15	**5 seg. (90)**	priory	seq.	C++

[1] In [69], the authors report other values of 3D reconstruction methods [11,12,20,42,44]. They claim they tuned the setup to be optimal, but they neither specify how long the experiments took, nor the setup of their trials, nor the configuration parameters.

4.4. CMUfaceSequence

This sequence presents a face of a person while changing the pose and talking. This sequence is used in other methods such as in [21]. The sequence is mostly subject to rigid motions, although it presents deformations on the mouth area, when the person is talking. The amount of points is 40 over 315 frames projected on an image size of 640 × 480. The deformation energy kept by the first 15 bases ($K = 15$) is 95.8%, which follows the minimum energy criteria explained in Section 4.2.

The same tests have been carried out for this sequence, and the same trends observed in previous experiments have been seen here. As the number of points is significantly fewer than in the flag sequence, the minimum amount of points to consider for visibility is 50%. The same trend and conclusions can be extracted for the analysis with noise, outliers and bases, only varying the average values. The average value for K = 15 bases is 0.26 for 2D reprojection error and 1% of 3D reconstruction error.

Table 3 shows a summary of the results obtained with the CMUfacecompared to the sparse approaches [11,12,20,21]. The same format as in Table 2 is used. In this case, the presented algorithm gets the best results among the tested algorithms for the error parameters and the time performance.

In a sequence in which most of the behavior is rigid except for the mouth movements, the overall performance of the algorithms is good as they all exploit this circumstance. The presented model-based reconstruction algorithm outperforms the results of model-free approaches.

Table 3. CMUfacesequence results summary. Some results comparing different state-of-the-art methods against the presented one are depicted.

Method	2D Error (px)	3D Error (%)	Rank	Tproc (fps)	Model	Proc Type	Impl.
[21]	1.06	3.18	8 (max)	14 min (0.37)	auto	seq.	MATLAB
[20]	0.6	3.19	5	43 seg. (7.35)	auto	batch	MATLAB
[11]	6.42/32.39	9.9/56.06	5/15	78/606 seg. (4.05/0.52)	auto	batch	MATLAB
[12]	1.06	2.43	12	13 seg (24.3)	auto	batch	MATLAB
Our approach	**0.26**	**1.01**	15	**4.5 seg. (69.78)**	priory	seq.	C++

4.5. Point-Wise CVLab's Kinect Paper

This dataset is derived from the Kinect paper dataset presented in [70]. It consists of a video sequence of 99 points on 191 frames showing a deforming piece of paper, which is captured with a depth sensor. The projections of the mesh for each frame have been obtained by interpolation from the set of original matches. The projections are further adjusted by estimating the relative camera-object pose and making sure that they match with the original SIFT images.

We see the same trend observed in the previous datasets when the number of bases is incremented. This dataset needs a bigger amount of bases to achieve good reconstruction results. For K = 15, the 2D error is about 9 px, and the 3D error almost reaches 11.92%, whereas, for K = 80, the 2D error becomes lower (8 px approximately), and the 3D error goes to 11.8%.

4.6. Rendered Flag Sequence

The flag sequence depicted in the previous tests is interpolated to get a dense surface and then is rendered using orthographic projection, as presented in [71]. As our model-based tracking works with the perspective projection, this sequence was re-rendered with perspective projection. The sequence contains 450 frames.

This sequence was chosen in other works [24,43,52,71] as it consists of a set of images and a dense 3D ground truth to evaluate the performance of tracking methods. It is useful for our experiments because, even though our tracking is sparse, the ground truth is appropriate to perform a thorough comparison among different methods.

4.6.1. Evaluation of Visual Descriptors

First of all, some frames of the sequence processed with the IROS-based tracking published by the authors in [52] are shown in Figure 6a together with reconstructions (Figure 6c). The same frames processed with the SIFT tracking and their reconstructions are shown in Figure 6b,d. The detected points (Figure 6a,b is shown in magenta, the matched points in red, and the mesh model is overlaid in cyan, to give an idea of the tracking performance. Regarding the 3D reconstructions (Figure 6c,d), the ground truth points are represented in blue and the reconstruction in red.

In our method, the feature detection and matching methods are configured to prioritize speed over accuracy. The frames are ordered in the columns and the tracking methods in the rows. It can be seen that in Frame #200, there are some accuracy problems for the presented approaches. This is due to the sequence presenting very abrupt movements combined with an interval of possible tracking losses. The detection of features in these areas is a difficult task, even for the most advanced feature descriptors, so there are very few points detected in this area, yielding a poor local estimation.

The projection of the model on the screenshots (cyan) and the reconstruction results differ because the shape in the latter is obtained after applying Procrustes alignment between the ground truth and the reconstruction.

In order to have a general performance overview for all the analyzed descriptors, a benchmark is shown in Table 4. This comparative includes number of points, descriptor, matcher, processing time (for the whole sequence and fps), 2D error and 3D error.

A similar benchmark, but including the matching technique used for each feature descriptor, is shown in Figure 7. This studies the influence of the matching method on the error metrics. For binary descriptors such as AKAZE, ORB and BRISK, the use of Hamming distance slightly improves the results of 2D reprojection and 3D reconstruction.
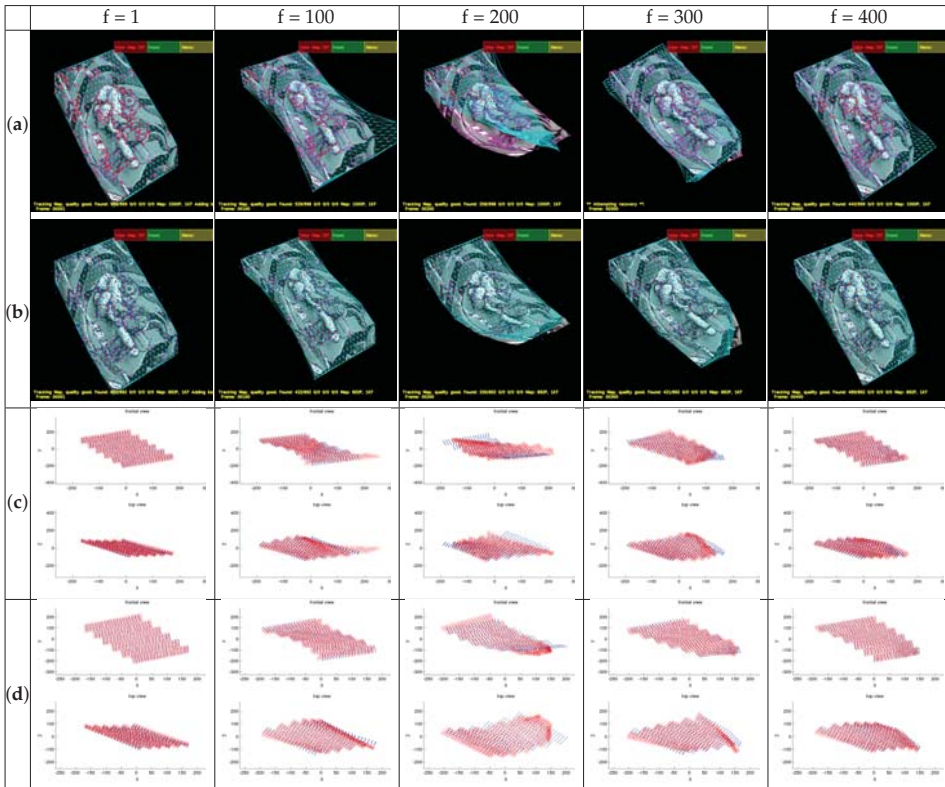

Figure 6. Rendered flag sequence screenshots and reconstructions for IROS and SIFT matching. (**a**) IROS screenshots, (**b**) SIFT screenshots, (**c**) IROS reconstruction results, (**d**) SIFT reconstruction result.

Table 4. Rendered flag sequence descriptor results comparison. As references, the original PTAM algorithm, the PTAM-based tracking and the descriptor-based results are presented.

Descriptor	Matcher	2D Error (px)	3D Error (%)	tp (s/fps)	Map Pts
PTAM	PTAM	40.12	104.54	102/4.4	350
IROS	PTAM-like	9.51	16.65	**26/17.3**	972
AKAZE	Brute force	5.68	12.42	**27/16.6**	607
BRISK	Brute force	5.98	14.27	56/8	1000
ORB	Brute force	5.67	12.71	29/15.5	1000
KAZE	Brute force	5.55	13.12	49/9.2	347
SIFT	Brute force	**5.27**	**12.04**	87/5.2	892
SURF	Brute force	5.4	14.11	29/15.5	650

Figure 7 shows that the best method according to the reprojection error is not necessarily the best in terms of the reconstruction error, as we pointed out before. SIFT gets the best performance for both

error estimations and among all the matching algorithms (brute force L1, L2). However, the second best is not the same for both figures (SURF in 2D and AKAZE in 3D).

AKAZE performs almost as well as SURF in terms of 2D reprojection error and is almost as good as SIFT in terms of 3D reconstruction error. Looking at Table 4, it can be seen that AKAZE is also one of the fastest approaches, which makes it a good candidate for a final implementation.

Regarding this table, it can be seen that the fastest approaches are IROS, AKAZE, ORB and SURF, and the ones that handle the most points are IROS, BRISK, ORB and SIFT. We remind that state-of-the-art approaches are only capable of working with a very small amount of points obtained from interpolated models. In this case, all the detected points are used to perform the estimation. It is important to remark that most of the reported time is consumed by the feature detection algorithm. The obtained times show the influence of the visual processing, compared with the reported results on the flag sequence in Table 2, in which the visual detection stage is not applied. PTAM results are included as a baseline for rigid SfM.
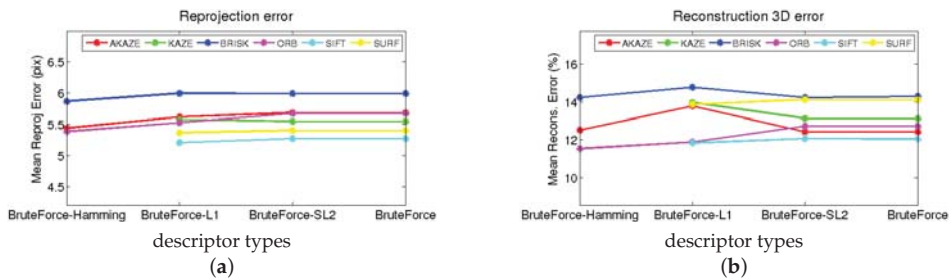

Figure 7. Rendered flag sequence rank based on the error results of the different descriptor methods. (**a**) Reprojection error; (**b**) 3D reconstruction error.

4.6.2. Performance Evaluation Based on the Number of Bases

As was done previously with the point-wise flag sequence dataset, the influence of the number of bases is tested for real conditions.

We show averaged results in Table 5. It can be seen that the trend observed with ideal matching is reversed here. Increasing the number of bases does not necessarily increases the 3D error. For IROS-based and SIFT-based tracking, the best results are obtained for seven bases.

Table 5. Rendered flag sequence: performance vs. the number of bases for IROS- and SIFT-based tracking.

Matching	#Bases	2D Error (pix)	3D Error (%)
IROS	**7**	**8.83**	**9.19**
IROS	15	9.51	16.65
IROS	30	13.2	34.23
SIFT	7	5.76	**6.53**
SIFT	15	5.27	12.04
SIFT	30	**5.23**	17.91

This clearly means that with real matches, the DoF of the shape model plays an important role in the reconstruction accuracy. With more bases, the model becomes more capable of adapting to large matching errors and more prone to erroneous reconstructions in areas of the object with a small amount of detected points. This effect is fixed by introducing priors as we explain in the next section and as displayed in Table 6.

4.6.3. Performance Evaluation with Time and Shape Smoothing Priors

Working with real images increases uncertainty in matching, which causes ambiguities and errors in the 3D shape. In order to reduce these effects, priors must be added. Due to the nature of our optimization algorithm (linear and split), time and shape smoothness priors can be implemented. The priors are studied at three levels, depending on the strength at which they are imposed, which corresponds to the value of their hyperparameters in the optimization (low, medium and high). The values for each type of prior are obtained heuristically.

The SIFT descriptor is chosen for the matching, as it obtained the best performance results in our previous tests (Table 4). Then, the results are compared to the IROS approach because it serves as a baseline of our first implementation [52].

After 3D reconstructions and 2D projections have been studied, a temporal analysis of the tracking for both priors regarding the baseline without priors is carried out.

The first evaluated prior is time smoothness, depicting the errors shown in Figures **??** and 8, which correspond to reprojection and 3D reconstruction errors. The figures are divided into two parts: the upper part corresponds to the IROS tracking, whereas the lower part corresponds to the SIFT tracking.
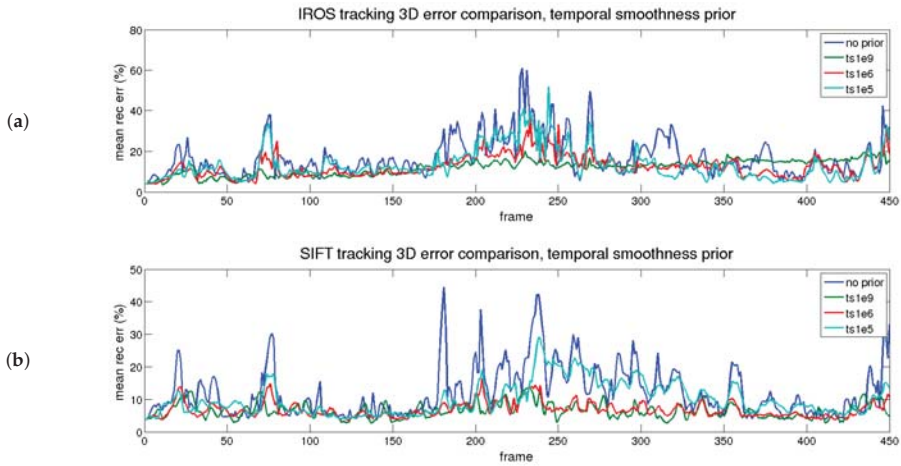
(a)

(b)
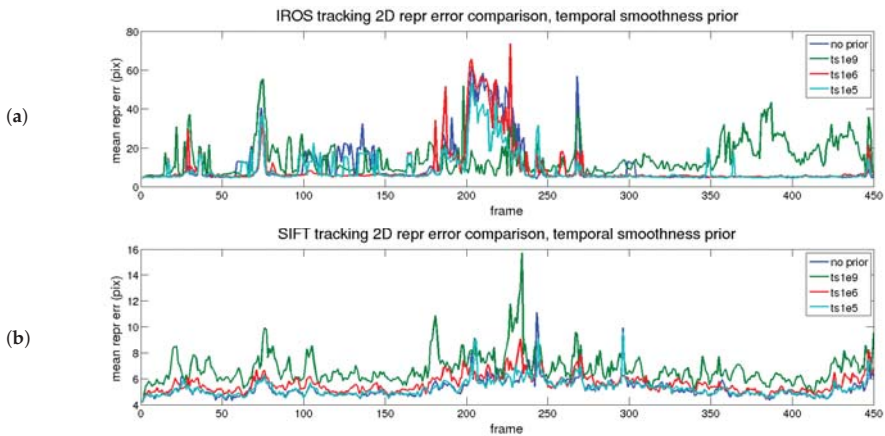
Figure 8. Rendered flag sequence: 3D reconstruction error over time with time smoothness. Results are shown for two types of descriptors. (**a**) IROS; (**b**) SIFT.

In Figure **??**, the reprojection errors are presented for the different priors with different colors. The blue trace represents the result without priors. In terms of error, cyan (low) and red (medium) traces are close to the blue trace. The green trace (high) presents high error for both approaches, except when the tracking is lost, as an excessive value of the prior makes the shape too rigid. The cyan trace, as expected, has little effect on the reprojection error, as the regularization value is low. The red trace has a slightly higher reprojection error, although in general terms, it is maintained along the sequence. Therefore, for the reprojection error, the prior does not imply an improvement.

The beneficial effect of the prior is seen in terms of reconstruction error, shown in Figure 8. For both approaches, when the prior is applied (even for small values), the 3D reconstruction error is reduced.

The best option corresponds to the red trace (medium) for both approaches. In that case, the increment in the reprojection error is small and the improvement in the 3D reconstruction error is large with respect to the blue trace (no priors applied). The effect is even more noticeable when the tracking is better, comparing the IROS and SIFT approaches.

To sum up, the errors reached after applying different priors are depicted in Table 6. The best results for each prior are marked in bold.The best option is obtained for a combination of both priors with medium values.

Table 6. Summary of the results for the rendered flag sequence with different numbers of bases, priors and descriptors.

Desc.	Prior Type	Value	2D Error (px)	3D Error (%)
PTAM	none	0	40.12	104.54
Point-wise ideal Matching	none	0	2	2.63
IROS	none	0	9.51	16.65
IROS	time	1e5	8.92	13.38
IROS	**time**	**1e6**	**9.76**	**12.42**
IROS	time	1e9	14.37	12.04
IROS	shape	1e4	24.05	18.31
IROS	**shape**	**1e5**	**9.38**	**9.54**
IROS	shape	1e9	11.62	6.2
IROS	**both**	**1e5/1e5**	**8.72**	**8.95**
SIFT	none	0	5.27	12.04
SIFT	time	1e5	5.29	9.93
SIFT	**time**	**1e6**	**5.64**	**6.79**
SIFT	time	1e9	6.79	6.51
SIFT	shape	1e4	5.18	8.48
SIFT	**shape**	**1.5e5**	**5.45**	**6.56**
SIFT	shape	1e9	6.80	6.2
SIFT	**both**	**1e5/1.5e5**	**5.48**	**6.46**

The results of the temporal analysis for the shape smoothness prior are depicted in Figure 9 (reprojection) and Figure 10 (3D reconstruction). A similar conclusion to the temporal smoothness prior can be extracted in this case. The trace in red (medium) reflects an equilibrium between reprojection and reconstruction errors, for both matching methods (IROS- and SIFT-based).

PTAM results are included to have a rigid SfM baseline in this sequence. On the other hand, just on the row below, the results from the (point-wise) flag sequence with perfect matching (Table 2) are displayed to show the impact of dealing with real images.

According to Table 6, the difference between the IROS and SIFT approaches is easy to explain: tracking quality is decisive to get accurate results, as the 2D reprojection error is about 4.3 pixels lower for SIFT before priors' application. Regarding the 3D reconstruction error, it is about 4% lower (16 down to 12) before priors, and after priors, it is reduced an additional 3% with respect to the best mark (9 down to 6).

Processing time due to the addition of priors is not included in the table, because it does not add a substantial change.
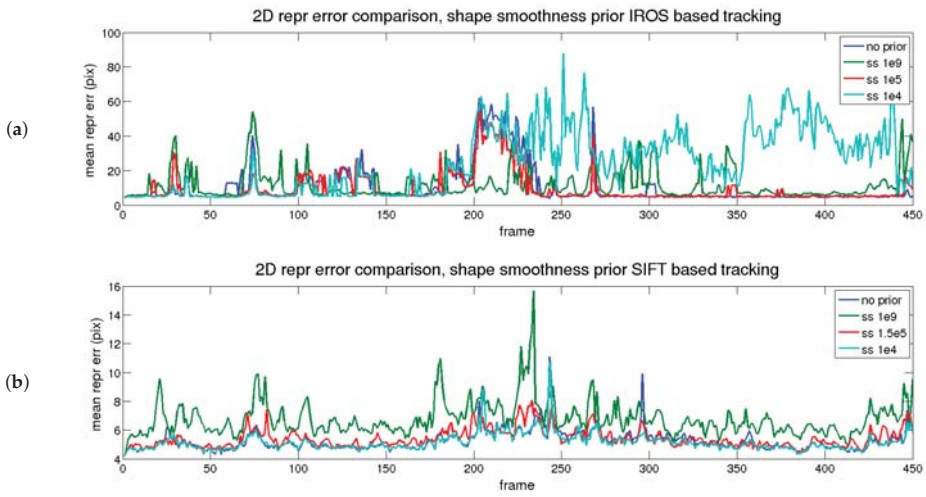
Figure 9. Rendered flag sequence: reprojection error over time with shape smoothness. Results are shown for two types of descriptors. (**a**) IROS; (**b**) SIFT.
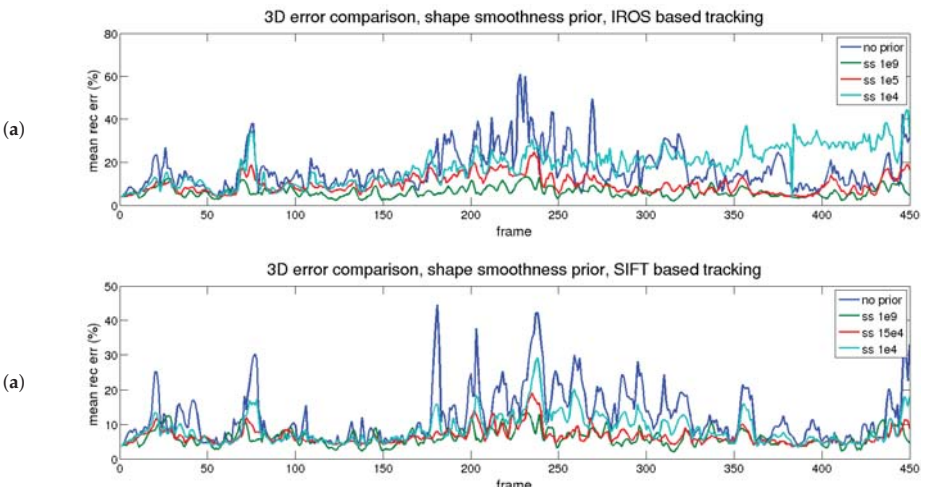

Figure 10. Rendered flag sequence: 3D reconstruction error over time with shape smoothness. Results are shown for two types of descriptors. (**a**) IROS; (**b**) SIFT.

4.7. CVLab's Kinect Paper

This sequence was first proposed in [70]. It consists of 191 frames of a well-textured piece of paper undergoing deformations and pose changes. The dataset includes a dense ground truth for each frame in the sequence. These data are acquired using the Kinect sensor, commonly used to test SfT algorithms.

One of the main differences of this sequence compared to the rendered flag is that, in this case, the background contains texture. Therefore, the area of the object is segmented at the beginning of the sequence to reject background features.

Regarding the results, they are similar to the ones in the rendered flag sequence. As the number of bases is increased, the 2D error is reduced, and the 3D error is increased. This time, there are few differences between the results using IROS and SIFT matching, as we show in Table 7.

The results are this time slightly better for the IROS approach compared to SIFT as the search radius is greater in the latter. The possibility of having incorrect matches due to a bigger radius is increased even though SIFT matching is better.

Finally, a comparative table with other state-of-the-art algorithms is shown in Table 8. We show our results without smoothing priors, which is a conservative result (check Table 7). It can be seen that our results are comparable with those obtained in [72].

Table 7. Summary of the results for the Kinect paper sequence with different numbers of bases, priors and descriptors.

Desc.	Prior Type	Value	#Bases	2D Error (px)	Depth Error (mm)
Point-wise ideal Matching	none	0	15	9.06	10.39
Point-wise ideal Matching	none	0	50	7.98	10.35
IROS	none	0	15	26.28	10.34
IROS	time	1e3	15	26.45	14.34
IROS	shape	1e4	15	27.48	10.38
IROS	none	0	50	24.77	11.44
IROS	time	1e3	50	24.82	11.37
IROS	shape	1e4	50	24.35	10.68
SIFT	none	0	15	23.34	10.63
SIFT	time	1e3	15	23.36	10.60
SIFT	shape	1e4	15	23.43	10.48
SIFT	none	0	50	24.47	13.64
SIFT	time	1e3	50	24.45	13.67
SIFT	shape	1e4	50	23.62	11.17

Table 8. Comparative results with other state-of-the-art algorithms in the Kinect paper sequence.

Method	Depth Error (mm)	Rank	Tproc (fps)	Model	Proc Type
[72]	7.23	-	-	priory	batch
[72] (PCA)	11.68	-	-	priory	batch
[73]	6.9	-	-	priory	sequential
[74]	5.57	-	959 (0.2)	priory	sequential
Our approach (SIFT)	13.65	50	90 (2)	priory	sequential
Our approach (IROS)	11.44	50	17 (11.24)	priory	sequential

5. Conclusions

This paper has presented a real-time and model-based reconstruction algorithm for non-rigid objects that offers a good trade-off between speed and quality. Our algorithm is based on PTAM, a well-known state-of-the-art SfM method that works for rigid objects. We upgrade PTAM's tracking thread to solve SfT using a low-rank shape basis model. We keep PTAM's real-time performance.

We demonstrate that our tracking algorithm is comparable in terms of accuracy to some of the most representative non-rigid reconstruction methods. It deals with challenging imaging conditions such as visibility reduction, outliers and noise. These conditions are not tackled by the majority of the state-of-the-art methods as they work with pre-registered image data. Our outlier rejection algorithm, based on [8], is effective for datasets showing mostly rigid or isometric deformations. Further tests on elastic sequences are thus required.

Our method includes different feature matching methods that improve tracking quality in real images. SIFT has been demonstrated to be the best matching method in terms of 2D reprojection

and 3D reconstruction error. However, it is very slow to comply with real-time constraints. The best trade-off between accuracy and complexity is given by the AKAZE matching method.

Our method uses priors for spatial and temporal smoothness. We show that a combination of the two priors helps to reduce the flickering, which is noticeable on the 3D reconstruction error and stabilizes the estimated trajectory.

Our current implementation requires further code optimizations to reduce CPU load (redundancy checks and code profiling). The objective is to reach stable 30 or even 60 fps for VGA images with shapes consisting of thousands of points.

Future research might include dense matching and shape representation. Our tracking method is also a key component for proposing a sequential NRSfM algorithm.

Author Contributions: Sebastián Bronte developed the algorithm, implemented it, conducted the experiments and wrote the paper. Luis M. Bergasa and Daniel Pizarro collaborated in the algorithm development, paper writing and reviews. Rafael Barea helped with conducting the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hartley, R.I.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004; ISBN: 0521540518.
2. Sturm, P.F.; Triggs, B. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In Proceedings of the 4th European Conference on Computer Vision-Volume II—Volume II, ECCV '96, Cambridge, UK, 15–18 April 1996; Springer-Verlag: London, UK, 1996; pp. 709–720.
3. Dellaert, F.; Kaess, M. Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing. *Int. J. Robot. Res.* **2006**, *25*, 1181–1203.
4. Davison, A.J.; Reid, I.D.; Molton, N.D.; Stasse, O. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1052–1067.
5. Agarwal, S.; Snavely, N.; Simon, I.; Seitz, S.M.; Szeliski, R. Building Rome in a day. In Proceedings of the IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, 27 September–4 October 2009; pp. 72–79.
6. Kyriakaki, G.; Doulamis, A.; Doulamis, N.; Ioannides, M.; Makantasis, K.; Protopapadakis, E.; Hadjiprocopis, A.; Wenzel, K.; Fritsch, D.; Klein, M.; et al. 4D Reconstruction of Tangible Cultural Heritage Objects from Web-Retrieved Images. *Int. J. Herit. Digit. Era* **2014**, *3*, 431–451.
7. Makantasis, K.; Doulamis, A.; Doulamis, N.; Ioannides, M. In the wild image retrieval and clustering for 3D cultural heritage landmarks reconstruction. *Multimed. Tools Appl.* **2016**, *75*, 3593–3629.
8. Klein, G.; Murray, D. Parallel Tracking and Mapping for Small AR Workspaces. In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 13–16 November 2007; pp. 1–10.
9. Newcombe, R.A.; Lovegrove, S.J.; Davison, A.J. DTAM: Dense Tracking and Mapping in Real-Time. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2320–2327.
10. Bregler, C.; Hertzmann, A.; Biermann, H. Recovering Non-Rigid 3D Shape from Image Streams. In Proceedings of the IEEE Converence on Computer Vision and Pattern Recognition, Havana, Cuba, 9–12 September 2008; pp. 690–696.
11. Torresani, L.; Hertzmann, A.; Bregler, C. Nonrigid Structure-from-motion: Estimating Shape and Motion with Hierarchical Priors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 878–892.
12. Paladini, M.; Del Bue, A.; Stosic, M.; Dodig, M.; Xavier, J.; Agapito, L. Factorization for Non-Rigid and Articulated Structure Using Metric Projections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2898–2905.

13. Del Bue, A.; Lladó, X.; Agapito, L. Non-Rigid Metric Shape and Motion Recovery from Uncalibrated Images Using Priors. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, NY, USA, 17–22 June 2006; pp. 1191–1198.

14. Chhatkuli, A.; Pizarro, D.; Bartoli, A. Non-Rigid Shape-from-Motion for Isometric Surfaces using Infinitesimal Planarity. In Proceedings of the British Machine Vision Conference, BMVC 2014, Nottingham, UK, 1–5 September 2014.

15. Parashar, S.; Pizarro, D.; Bartoli, A. Isometric Non-Rigid Shape-From-Motion in Linear Time. In Proceedings of the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

16. Bartoli, A.; Collins, T. Template-Based Isometric Deformable 3D Reconstruction with Sampling-Based Focal Length Self-Calibration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013.

17. Bartoli, A.; Gérard, Y.; Chadebecq, F.; Collins, T.; Pizarro, D. Shape-from-template. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 2099–2118.

18. Fua, P.; Salzmann, M. Linear Local Models for Monocular Reconstruction of Deformable Surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 931–944.

19. Ngo, D.T.; Östlund, J.; Fua, P. Template-Based Monocular 3D Shape Recovery Using Laplacian Meshes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 172–187.

20. Gotardo, P.F.; Martinez, A.M. Non-Rigid Structure from Motion with Complementary Rank-3 Spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 20–25 June 2011; pp. 3065–3072.

21. Paladini, M.; Bartoli, A.; Agapito, L. Sequential non-rigid structure-from-motion with the 3D-implicit low-rank shape model. In Proceedings of the 11th European Conference on Computer vision, Crete, Greece, 5–11 September 2010; pp. 15–28.

22. Agudo, A.; Moreno-Noguer, F.; Calvo, B.; Montiel, J. Sequential Non-Rigid Structure from Motion using Physical Priors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 979–994.

23. Agudo, A.; Calvo, B.; Montiel, J.M.M. Finite Element based Sequential Bayesian Non-Rigid Structure from Motion. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1418–1425.

24. Agudo, A.; Montiel, J.M.M.; Agapito, L.; Calvo, B. Online Dense Non-Rigid 3D Shape and Camera Motion Recovery. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014.

25. Collins, T.; Pizarro, D.; Bartoli, A.; Canis, M.; Bourdel, N. Computer-Assisted Laparoscopic Myomectomy by Augmenting the Uterus with Pre-operative MRI Data. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 10–12 September 2014.

26. Maier-Hein, L.; Groch, A.; Bartoli, A.; Bodenstedt, S.; Boissonnat, G.; Chang, P.L.; Clancy, N.; Elson, D.S.; Haase, S.; Heim, E.; et al. Comparative validation of single-shot optical techniques for laparoscopic 3-D surface reconstruction. *IEEE Trans. Med. Imag.* **2014**, *33*, 1913–1930.

27. Collins, T.; Bartoli, A. Realtime Shape-from-Template: System and Applications. In Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2015, Fukuoka, Japan, 29 September–3 October 2015; pp. 116–119.

28. Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Pushmeet, K.; Shotton, J.; Hodges, S.; Fitzgibbon, A. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Basel, Switzerland, 26–29 October 2011.

29. Cootes, T.; Edwards, G.; Taylor, C. Active Appearance Models. In *IEEE European Conference on Computer Vision*; Springer: New York, NY, USA, 1998; pp. 484–498.

30. Muñoz, E.; Buenaposada, J.M.; Baumela, L. A Direct Approach for Efficiently Tracking with 3D morphable models. In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 1615–1622.

31. Booth, J.; Roussos, A.; Zafeiriou, S.; Ponniah, A.; Dunaway, D. A 3D Morphable Model Learnt From 10,000 Faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

32. Zhou, Y.; Antonakos, E.; Alabort-i Medina, J.; Roussos, A.; Zafeiriou, S. Estimating Correspondences of Deformable Objects "In-The-Wild". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

33. Brunet, F.; Hartley, R.; Bartoli, A.; Navab, N.; Malgouyres, R. Monocular Template-based Reconstruction of Smooth and Inextensible Surfaces. In Proceedings of the Tenth Asian Conference on Computer Vision (ACCV 2010), Queenstown, New Zealand, 8–12 November 2010; Volume 6494, pp. 52–66.

34. Vicente, S.; Agapito, L. Soft Inextensibility Constraints for Template-Free Non-rigid Reconstruction. In Proceedings of the Computer Vision—ECCV 2012-12th European Conference on Computer Vision Part III, Florence, Italy, 7–13 October 2012; pp. 426–440.

35. Pizarro, D.; Bartoli, A. Feature-Based Deformable Surface Detection with Self-Occlusion Reasoning. *Int. J. Comput. Vis.* **2012**, *97*, 54–70.

36. Sundaram, N.; Brox, T.; Keutzer, K. Dense point trajectories by GPU-accelerated large displacement optical flow. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2010; pp. 438–451.

37. Chhatkuli, A.; Pizarro, D.; Collins, T.; Bartoli, A. Inextensible Non-Rigid Shape-from-Motion by Second-Order Cone Programming. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 1719–1727.

38. Agudo, A.; Calvo, B.; Montiel, J.M.M. 3D Reconstruction of Non-Rigid Surfaces in Real-Time Using Wedge Elements. In Proceedings of the Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (ECCVW), Firenze, Italy, 7 October 2012; pp. 113–122.

39. Agudo, A.; Agapito, L.; Calvo, B.; Montiel, J.M.M. Good Vibrations: A Modal Analysis Approach for Sequential Non-Rigid Structure from Motion. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1558–1565.

40. Agudo, A.; Moreno-Noguer, F.; Calvo, B.; Montiel, J. Real-Time 3D Reconstruction of Non-Rigid Shapes with a Single Moving Camera. *Comput. Vis. Image Underst.* **2016**, *153*, 37–54.

41. Fayad, J.; Agapito, L.; Del Bue, A. Piecewise Quadratic Reconstruction of Non-Rigid Surfaces from Monocular Sequences. In Proceedings of the 11th European Conference on Computer Vision, Computer Vision—ECCV 2010, Part IV, Heraklion, Crete, Greece, 5–11 September 2010; pp. 297–310.

42. Russell, C.; Fayad, J.; Agapito, L. Energy based multiple model fitting for non-rigid structure from motion. In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 3009–3016.

43. Garg, R.; Roussos, A.; Agapito, L. Dense Variational Reconstruction of Non-rigid Surfaces from Monocular Video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1272–1279.

44. Dai, Y.; Li, H.; He, M. A simple prior-free method for non-rigid structure-from-motion factorization. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2018–2025.

45. Akhter, I.; Sheikh, Y.; Khan, S.; Kanade, T. Nonrigid structure from motion in trajectory space. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA; London, UK, 2009; pp. 41–48.

46. Russell, C.; Yu, R.; Agapito, L. Video Pop-up: Monocular 3D Reconstruction of Dynamic Scenes. In *Computer Vision—ECCV 2014*; Lecture Notes in Computer Science; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Berlin, Germany, 2014; Volume 8695, pp. 583–598.

47. Newcombe, R.A.; Fox, D.; Seitz, S.M. DynamicFusion: Reconstruction and Tracking of Non-Rigid Scenes in Real-Time. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

48. Hartley, R.I.; Vidal, R. Perspective Nonrigid Shape and Motion Recovery. In Proceedings of the 10th European Conference on Computer Vision, Computer Vision—ECCV 2008, Marseille, France, 12–18 October 2008; pp. 276–289.

49. Bartoli, A.; Gay-Bellille, V.; Castellani, U.; Peyras, J.; Olsen, S.I.; Sayd, P. Coarse-to-fine low-rank structure-from-motion. In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK, USA, 24–26 June 2008.

50. Yu, R.; Russell, C.; Campbell, N.D.F.; Agapito, L. Direct, Dense, and Deformable: Template-Based Non-Rigid 3D Reconstruction From RGB Video. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015 .

51. Klein, G.; Murray, D. Parallel Tracking and Mapping on a Camera Phone. In Proceedings of the Eigth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09), Orlando, FL, USA, 19–22 October 2009.

52. Bronte, S.; Paladini, M.; Bergasa, L.M.; Agapito, L.; Arroyo, R. Real-time sequential model-based non-rigid SFM. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 1026–1031.

53. Moreno-Noguer, F.; Porta, J.M. Probabilistic Simultaneous Pose and Non-Rigid Shape Recovery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1289–1296.

54. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2006; Volume 1, pp. 430–443.

55. Itseez. Open Source Computer Vision Library, 2015; Available online: https://github.com/itseez/opencv (accessed on 30 September 2017).

56. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE features. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2012; pp. 214–227.

57. Alcantarilla, P.F.; Nuevo, J.; Bartoli, A. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In Proceedings of the British Machine Vision Conference (BMVC), Bristol, UK, 9–13 September 2013.

58. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.

59. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.

60. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

61. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2006; pp. 404–417.

62. Joseph Tan, D.; Holzer, S.; Navab, N.; Ilic, S. Deformable Template Tracking in 1 ms. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014.

63. Simo-Sierra, E.; Torras, C.; Moreno-Noger, F. DaLI: Deformation and Light Invariant Descriptor. *Int. J. Comput. Vis.* **2015**, *115*, 136–154.

64. Benhimane, S.; Malis, E. Homography-based 2D Visual Tracking and Servoing. *Int. J. Robot. Res.* **2007**, *26*, 661–676.

65. Torresani, L.; Hertzmann, A.; Bregler, C. Learning Non-Rigid 3D Shape from 2D Motion. In *Advances in Neural Information Processing Systems 16*; MIT Press: Cambridge, MA, USA, 2003; pp. 1555–1562.

66. Dellaert, F. *The Expectation Maximization Algorithm*; Technical Report; Georgia Institute of Technology, College of Computing: Atlanta, Georgia, 2002.

67. Tukey, J.W. A survey of sampling from contaminated distributions. *Contrib. Probab. Stat.* **1960**, *2*, 448–485.

68. Gower, J.C.; Dijksterhuis, G.B. *Procrustes Problems*; Oxford Statistical Science Series; Oxford University Press: New York, NY, USA, 2004.

69. Lee, M.; Cho, J.; Oh, S. Consensus of Non-Rigid Reconstructions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

70. Varol, A.; Shaji, A.; Salzmann, M.; Fua, P. Monocular 3D Reconstruction of Locally Textured Surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1118–1130.

71. Garg, R.; Roussos, A.; de Agapito, L. Robust Trajectory-Space TV-L1 Optical Flow for Non-rigid Sequences. In Proceedings of the Energy Minimization Methods in Computer Vision and Pattern Recognition, Saint Petersburg, Russia, 25–27 July 2011; Volume 6819, pp. 300–314.

72. Varol, A.; Salzmann, M.; Fua, P.; Urtasun, R. A constrained latent variable model. In Proceedings of the IEEE Computer Society, Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2248–2255.

73. Chhatkuli, A.; Pizarro, D.; Bartoli, A. Stable Template-Based Isometric 3D Reconstruction in All Imaging Conditions by Linear Least-Squares. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, 23–28 June 2014; pp. 708–715.

74. Salzmann, M. Continuous Inference in Graphical Models with Polynomial Energies. In Proceedings of the IEEE Computer Society, Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 1744–1751.

sensors

Article

Enhancing the Trajectory Generation of a Stair-Climbing Mobility System

Jose Abel Chocoteco [1], Rafael Morales [2,*] and Vicente Feliu-Batlle [3]

[1]   Instituto Tecnológico de Ciudad Guzmán (ITCG), Tecnológico Nacional de México (TecNM),
     Ciudad Guzmán 49100, Mexico; jachocoteco@itcg.edu.mx
[2]   School of Industrial Engineering, University of Castilla-La Mancha (UCLM), 02071 Albacete, Spain
[3]   School of Industrial Engineering, University of Castilla-La Mancha (UCLM), 13071 Ciudad Real, Spain;
     Vicente.Feliu@uclm.es
*   Correspondence: Rafael.Morales@uclm.es; Tel.: +34-967-599-200 (ext. 2542)

Abstract: Recent advances in mobile robotic technologies have enabled significant progress to be made in the development of Stair-Climbing Mobility Systems (SCMSs) for people with mobility impairments and limitations. These devices are mainly characterized by their ability to negotiate those architectural barriers associated with climbing stairs (curbs, ramps, etc.). The development of advanced trajectory generators with which to surpass such architectural barriers is one of the most important aspects of SCMSs that has not yet been appropriately exploited. These advanced trajectory generators have a considerable influence on the time invested in the stair climbing process and on passenger comfort and, consequently, provide people with physical disabilities with greater independence and a higher quality of life. In this paper, we propose a new nonlinear trajectory generator for an SCMS. This generator balances the stair-climbing time and the user's comfort and includes the most important constraints inherent to the system behavior: the geometry of the architectural barrier, the reconfigurable nature of the SCMS (discontinuous states), SCMS state-transition diagrams, comfort restrictions and physical limitations as regards the actuators, speed and acceleration. The SCMS was tested on a real two-step staircase using different time-comfort combinations and different climbing strategies to verify the effectiveness and the robustness of the proposed approach.

Keywords: architectural barriers; stair-climbing mobility system; trajectory generation; assistive technology

1. Introduction

Stair-Climbing Mobility Systems (SCMSs) are assistive devices for people whose mobility is impaired and limited. These devices are mainly characterized by their ability to negotiate those architectural barriers associated with climbing stairs. In this respect, SCMSs whose locomotion systems combine more than one mechanism of a different nature have a greater ability to surpass stairs than other models. These SCMSs, which are known as hybrid SCMSs, generally combine wheeled and legged systems (see, for example, [1–5]), wheel clusters attached to powered linkages (see, for example, [6–8]) and wheeled and tracked systems (see, for example, [9–12]). However, despite the fact that hybrid SCMSs are the most suitable systems with which to surpass stairs, they still have problems as regards guaranteeing the user a high factor of safety and comfort, signifying that most are not completely autonomous and external assistance is therefore required (see recent reviews by [13,14]).

Hybrid SCMSs can be divided into two major categories: semi-autonomous and fully-autonomous. Semi-autonomous SCMSs require a great deal of user interaction to surpass stairs (in some cases, external assistance, such as another person or a fixed structure, is required to surpass the obstacle).

Semi-autonomous SCMSs are characterized by the fact that their stair-climbing modes are not based on trajectory planning. Abrupt movements or collisions usually occur because the users do not control the SCMSs correctly. In this case, the climbing process may be uncomfortable for the passenger, and security cannot be guaranteed without appropriate assistance. Two examples of this type of SCMS are the iBOT mobility system [8] and the TopChair [9]. Unlike semi-autonomous SCMSs, fully-autonomous SCMSs require minimal user interaction and do not require external assistance when negotiating stairs. With fully-autonomous SCMSs, a trajectory planning is considered for the stair-climbing operation. These models generally utilize algorithms to detect objects and avoid collisions and, in some cases, algorithms with which to estimate the dimensions of the stairs in order to derive a path using trajectory planning and trajectory optimization strategies. Little research in the literature deals with trajectory planning and trajectory optimization for fully-autonomous SCMSs. Three examples of this type of SCMS are presented in [4,15,16]. These fully-autonomous SCMSs employ strategies that allow them to know the environment (they estimate the size of the stairs using laser sensors).

Electric Powered Wheelchair (EPW) comfort has been identified as a priority in many studies, some major examples of which include [17–26]. Yet, there seems to be little agreement among ergonomists or disability researchers as how best to quantify discomfort or produce mechanisms that can reliably link feelings of discomfort with qualitative indicators. Some indicators and criteria regarding the assessment of human comfort in electric powered wheelchairs are proposed in [27–32]). Something similar occurs with current SCMSs, which do not adequately meet the need for comfort, mainly during the stair-climbing task. Relatively few research efforts have, therefore, been focused on the discomfort associated with SCMSs, although three recent examples of this include [5,33,34].

In this paper, we propose a trajectory planning approach based on the optimization of both stair-climbing execution time and user comfort for our fully-autonomous SCMS, which is shown in Figure 1. We propose to deal with the problem of generating optimal and smooth trajectories under various kinematic and dynamic constraints by using a simple numerical method. The trajectory generation problem is formulated as a nonlinear optimization problem by parameterizing both the SCMS path and the associated motion profile using a set of control points which are fitted by means of B-spline functions [35,36]. These control points are directly extracted from a reference trajectory initially based on the current knowledge of the size of the stair. Trajectory planning with which to control the position of the center of mass of the SCMS, denoted as $\mathbf{P}_g$, can then be performed using smooth trajectories, which have to be continuous and at least twice differentiable. The objective of the proposed method is to minimize a cost function, which is a weighted balance of stair-climbing time and the user's comfort. This optimization problem is solved by using a Sequential Quadratic Programming (SQP) technique. Moreover, the behavior of the SCMS and the maintenance of the vertical position of its chassis are notably improved when using the proposed trajectory planning. Experiments have been carried out in order to evaluate the efficiency of the proposed approach.

The remainder of the paper is organized as follows: Section 2 presents a brief description of the mechanical design (mechanism and operating modes) and the environment recognition of the reconfigurable SCMS. Section 3 is devoted to the description of the kinematics and dynamics of the SCMS. The statement of the trajectory generator is depicted in Section 4. Section 5 briefly describes the experimental platform and the experiments performed to demonstrate the robustness of the proposed approach by carrying out several trials with different time-comfort combinations and different climbing strategies. Finally, Section 6 is devoted to the conclusions of the work.
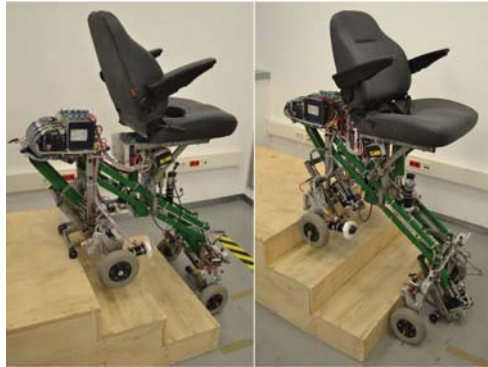
Figure 1. The proposed Stair-Climbing Mobility System (SCMS) is characterized by the fact that it can climb up and down stairs of different sizes and maintain a stable equilibrium on the stair at all times without assistance. It is also characterized by the fact that it can estimate the size of the stair using laser distance sensors and generate safe and comfortable trajectories for the central platform when carrying a user.

2. Experimental System

The SCMS presented in this paper was built by a team of researchers at the School of Industrial Engineering at the UCLMuniversity. The SCMS developed is considered to be a hybrid SCMS since it combines two different types of locomotion to solve the stair-climbing problem. The SCMS uses wheeled and sliding support systems.

2.1. Mechanical Description

The general design of the SCMS is based on splitting the obstacle-climbing problem into two different problems of a lower order of difficulty. These problems are: (1) posture control of the vehicle and accommodating the wheels to the staircase; and (2) traction control of the wheels and surpassing the step. Each problem is solved by using two independent mechanisms; the first problem is solved by using a positioning mechanism while the second is solved by using two similar climbing mechanisms (see Figure 2). These latter mechanisms are those which combine the aforementioned wheeled and sliding support systems. A key aspect of the mechanical design is that both mechanisms can act independently of each other during the locomotion, which makes it possible to use different stair-climbing strategies.
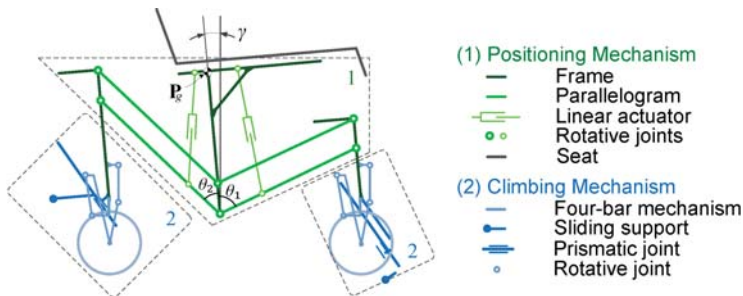
Figure 2. Kinematic scheme of the SCMS: (1) positioning mechanism, which is responsible for ensuring the posture of the entire vehicle; (2) climbing mechanism, which is responsible for surpassing the obstacle.

The design of the positioning mechanism is based on a closed-loop mechanism. This mechanism is composed of three frames: a central one on which the seat where the user sits is placed and a rear and a front one to hold the proposed climbing mechanisms. The frames are connected to each other by two parallelograms that prevent relative rotation occurring between them. The central frame is also called the SCMS frame. Its inclination with regard to the direction of gravity is the controlled variable and is represented using γ. The positioning mechanism needs four stable support points at each instant to maintain the SCMS posture and to position the vehicle wheels on the staircase. These points are supplied by the two climbing mechanisms by means of their four wheels and two sliding supports, two wheels and one support for each mechanism. The positioning mechanism has two Degrees of Freedom (DOFs), each of which is driven by a linear actuator. The positioning mechanism can adapt to different stair geometries, which includes positioning both axles in order to maintain the stability and accommodating the front wheels on the staircase.

The climbing mechanisms are designed to surpass curbs or stairs. Our hybrid SCMS has two similar and independent climbing mechanisms; one front and the other rear. Each mechanism combines two wheels, one on each side, and one sliding support. The driving wheels are coupled to the rear climbing mechanism and they are controlled by direct current motors. The sliding supports are the supports in charge of surpassing the obstacles and they are controlled by linear actuators, similar to those used in the positioning mechanism. These mechanisms make it possible for the SCMS to surpass different sized stairs. The climbing mechanism combines four-bar mechanisms to control the wheels when the SCMS confronts the steps. These mechanisms are designed to be used with either the front or the rear wheels. This design also enables the same control technique to be used for all the wheels.

The operation of the proposed climbing mechanism when climbing a step is outlined in the sequence of pictures in Figure 3. The weight is initially supported by the wheel, which rolls up closer to the step. It stops when it is close to the stair riser (Figure 3a). The sliding support is deployed until the instant at which the wheels are off the ground, as shown in Figure 3b. sliding support makes contact with the step tread. The wheel then moves upwards to climb the step in order to avoid interference with the corner of the step (Figure 3c). When the step has been surpassed (Figure 3d) the wheel moves backwards to its initial point, which is referred to as the frame (Figure 3e). The sliding support continues to maintain contact with the step tread while the wheel climbs. Finally, the support is retracted (Figure 3f). This last operation is performed in order to once again transfer the weight from the sliding support to the wheel. The descending operation runs similarly and conversely to the ascending operation. During the climbing/descent process, the wheels do not need traction since the movement and stability are guaranteed by the sliding support, which assumes the greatest responsibility for the obstacle being surpassed. The positioning mechanism operation is independent of the wheels while the climbing mechanism climbs/descends an obstacle. This latter mechanism can negotiate stairs of different sizes, even with geometric disturbances, and maintain a stable equilibrium on the staircase at all times.
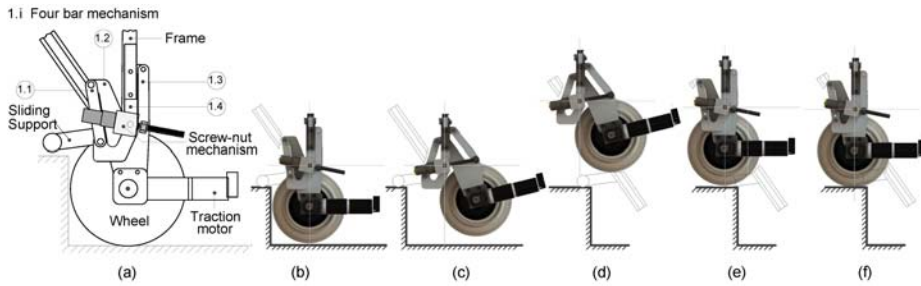
Figure 3. Actuating sequence of the rear climbing mechanism. . In (**a**) the wheel makes contact with the ground. In (**b**) the sliding support slides and makes contact with the step tread and the wheel is off the ground. In (**c**) the screw-nut mechanism is again actuated and the sliding support continues to slide. In (**d**) the sliding support is completely deployed. In (**e**) the screw-nut mechanism is actuated and the wheel moves backwards to its initial point. In (**f**) the wheel again makes contact with the ground and the sliding support is retracted.

2.2. Operating Modes

The SCMS has four possible operating modes, which we call configurations (see Figure 4). Configuration 1 (the most frequently used) is performed when the device is supported on four wheels. This configuration can be used for several purposes: (1) moving not only on level terrain, but also on tilted terrains, such as operating an usual EPW; (2) adjustment of the seat height, thus allowing the user to move around at eye level or/and reach items on high shelves and; (3) surpassing typical architectural barriers such as curbs, ramps or staircases. Configurations 2 and 3 are performed when the device is supported on the front wheels and the rear sliding support and the rear wheels and the front sliding support, respectively. These configurations are used to surpass curbs or staircases. Finally, Configuration 4, which is performed when the device is supported on the two sliding supports, is used only to surpass stairs with more than two steps.
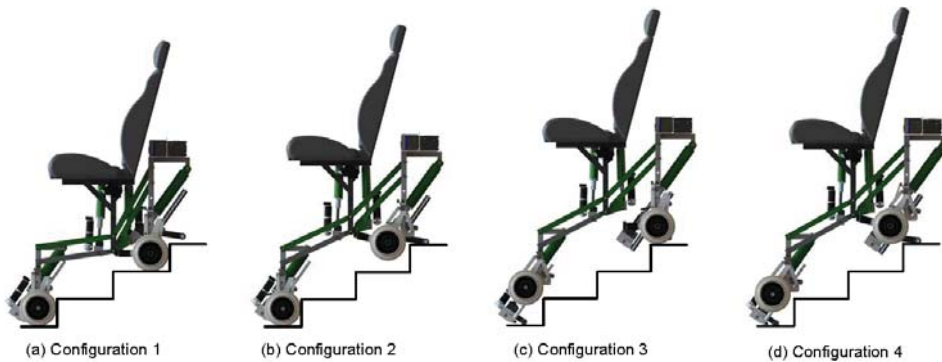
Figure 4. Possible configurations of the SCMS.

2.3. Laser Distance Sensors for Environment Recognition

Two laser distance sensors are used to perform the strategy employed to estimate the size of the stair. These sensors (BOD 63MLA04-S115) are manufactured by Balluff. They have a working range of between 200 mm and 6000 mm; a nominal response time of less than 2 ms; and a nominal resolution of less than 1 mm. This type of sensors works on the principle of Time-Of-Flight (TOF)

measurement, which is an excellent cost-effective tool for many modern mapping problems. All the technical specifications and characteristics of these sensors are described in [37].

The sensors are strategically positioned on the lateral parts of the central frame of the SCMS (see Figure 5). They are mechanically coupled to a pivot shaft, which is powered by a DC motor with a belt-pulley transmission. The motor (Maxon Motor 226774, maxon motor ag, Sachseln, Switzerland) combines a gearbox (Maxon Motor 166174, reduction 246:1, maxon motor ag, Sachseln, Switzerland) and an optical incremental encoder with which to sense position. The motor is controlled by means of an EPOS2 positioning controller (maxon motor ag, Sachseln, Switzerland).

Figure 5. Scanning system positioned on the SCMS frame.

From the sensors' strategic position, the architectural barrier can be scanned in both backward and forward directions without the chassis interfering with the laser pointers. They point backward when climbing up and forward when climbing down (see Figure 6). In addition, thanks to this strategic position, the system does not need any additional sensors to detect obstacles or to avoid obstruction during either the climbing or the descent processes. The estimation strategy greatly simplifies the control and sensory systems.
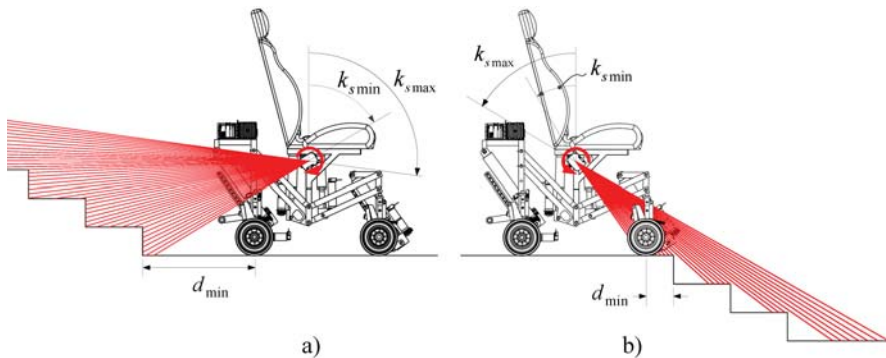
Figure 6. Side view of scanning on the staircase surface; (**a**) before climbing up and (**b**) before climbing down.

The estimation is obtained using a geometric analysis based on the scanning results and the odometry data. Figure 6a shows the scanning of the staircase surface that takes place before climbing up. The scanning is performed by rotating the two laser distance sensors in a clockwise direction. The angular position of the sensors is defined by the angle k_s. The field of view of the sensors allows the laser beams to project on the steps' surfaces without interference from the SCMS chassis. The scanning is executed only once, and this is sufficient to perform the geometric analysis and obtain the stairs' dimensions. In practice, the SCMS must carry out a new scan every four steps.

Figure 7 shows the results obtained from the scanning of the staircase surface before climbing up. The overlapping of the resulting trajectories demonstrates that the SCMS is aligned with the staircase.
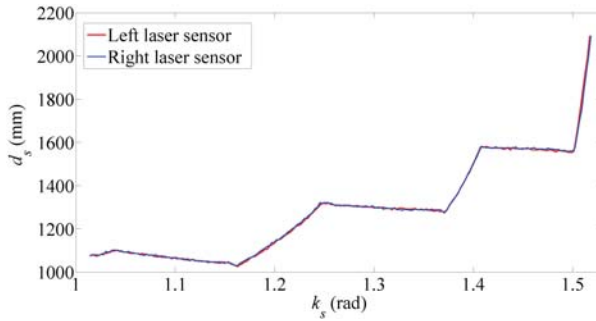
Figure 7. Distance measured during the scanning on the staircase surface before climbing up.

The components of the distance measured can be calculated using expressions:

$$distance(x,y) = \begin{cases} d_{sx} & = d_s \sin(k_s) \\ d_{sy} & = d_s \cos(k_s) - d_{so} \cos(k_{so}) \end{cases}$$

where d_s is the distance measured by the laser sensors, d_{sx} is the horizontal and d_{sy} is the vertical component, d_{so} is the initial measured distance and k_{so} is the initial angle of the laser beams. The results calculated are illustrated in Figure 8.

The estimation of the stairs' dimensions is achieved by partitioning the d_{sx} and d_{sy} measurements into groups, where a group is started whenever the distance between neighboring measurements exceeds a threshold of 50 mm. Figure 8 thus shows that there are three groups in the horizontal component (at approximately 944, 1244 and 1544 mm) and four in the vertical component (at approximately 0, 150, 300 and 450 mm). Finally, the median is calculated for each group, and the stairs' dimensions and the number of stairs are therefore obtained.
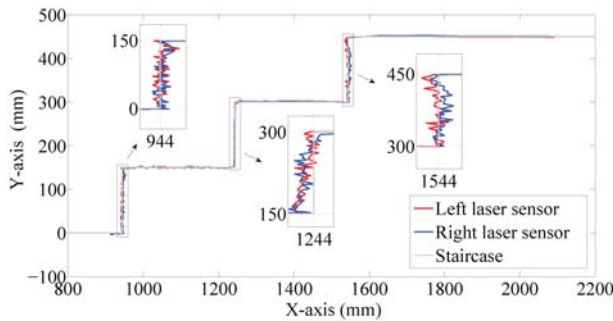
Figure 8. Estimated shape and size of the staircase.

Another important advantage of our SCMS is its capacity to estimate the stairs's dimensions when it goes down staircases. Figure 6b shows the scanning on the staircase surface before descent. Figure 9 shows the results obtained from the scanning of the staircase surface before climbing down. Note that in this case, the vertical part of the stairs cannot be scanned. Again, the overlapping among the resulting trajectories demonstrates that the SCMS are correctly aligned with the staircase.
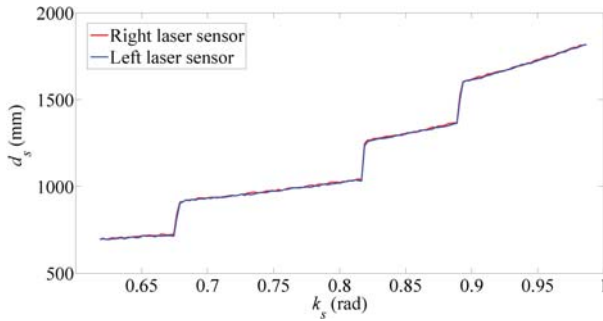
Figure 9. Distance measured during the scanning on the staircase surface before climbing down.

The dimension estimation is achieved using the same approach as shown above. Figure 10 indicates that there are only four groups in the vertical component; at approximately 0, -150, -300 and -450 mm. The median is calculated for each group to obtain the height of the stairs and the number of steps. The estimation of the stairs' width is based on the abrupt jumps of the laser beams, which occur three times. The right-hand ends of each group are calculated to determine the stairs width. The ends occur at proximity 445, 745 and 1045 mm.
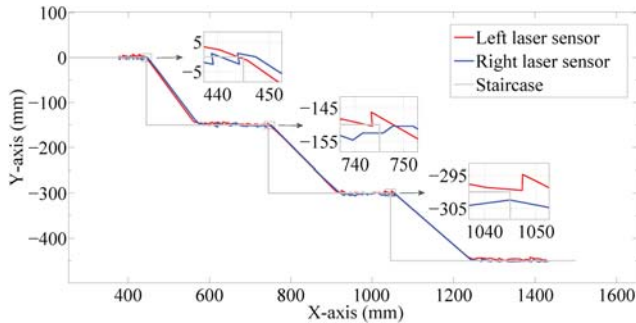
Figure 10. Estimated shape and size of the staircase.

The proposed strategy is easy to apply during both the ascent and descent of stairs. The estimation errors achieved are about 2–3 mm in both cases. These errors should not have a significant impact on the user's safety and comfort. More details about the strategy used to estimate the size of the stairs are provided in [38].

3. Kinematics and Dynamics Modeling

The literature related to modeling SCMSs is currently limited, and one of the reasons for this is that these models are not required for a simple control law, which is in fact that most commonly adopted by electric powered wheelchairs. In the case of SCMSs, the definition of kinematic and dynamic models plays an important role not only in the design of control algorithms but also in the generation of optimal and smooth trajectories. This section is focused on the development of the kinematic and dynamic models for our hybrid SCMS, which takes into consideration its reconfigurable nature.

3.1. Kinematics Modeling

The kinematic model is used to define the relation among the position variables when the vehicle moves on either a continuous smooth profile or a discontinuous profile, such as a flat floor or a staircase.

Figure 11 is very useful as regards understanding this relationship and formulating the kinematic equations for the four configurations that occur during stair-climbing. We shall begin by defining the notations used in the model and then determining the kinematic equations used in each configuration. The kinematic equations are expressed in complex notation, which has been chosen for mathematical and computational convenience.
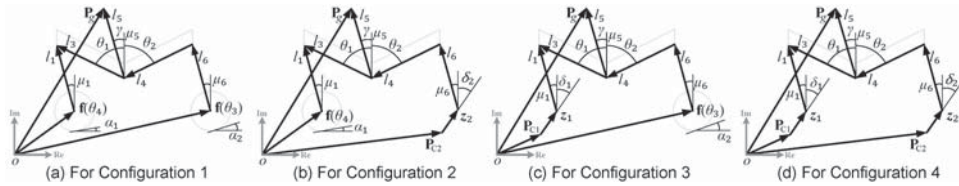

Figure 11. General kinematic scheme for the different configurations of the SCMS.

The position of the center of mass of the SCMS is defined as $\mathbf{P}_g = \left[P_{gx}, P_{gy} \right]^T$, where P_{gx} and P_{gy} are the horizontal and vertical components. The inclination angle of the SCMS frame is defined as γ. These values will be grouped into the vector $\mathbf{p} = \left[\mathbf{P}_g^T, \gamma \right]^T$. The front and rear joint variables of the positioning mechanism are denoted by θ_1 and θ_2, respectively. The positions of the front and rear axles of the wheels are represented using $\mathbf{f}(\theta_4)$ and $\mathbf{f}(\theta_3)$, where θ_4 is the movement of the front wheels and θ_3 is the movement of the rear wheels (the driving wheels). The positions of the front and rear sliding supports are denoted as z_1 and z_2. They form the angles δ_1 and δ_2 with the imaginary axis, respectively. θ_1, θ_2, θ_3 and θ_4 are rotational DOFs and z_1 and z_2 are translational DOFs. The actuated DOFs can be joined to the vector $\mathbf{q} = [\theta_1, \theta_2, \theta_3, z_1, z_2]^T$. The reference trajectories for the vectors $\mathbf{p}$ and $\mathbf{q}$ are defined as $\mathbf{p}^* = \left[\mathbf{P}_g^{*T}, \gamma^* \right]^T$ and $\mathbf{q}^* = \left[\theta_1^*, \theta_2^*, \theta_3^*, z_1^*, z_2^* \right]^T$, respectively.

Moreover, the components $\mathbf{P}_{C1}$ and $\mathbf{P}_{C2}$ correspond to the instant at which the wheels are off the ground during locomotion. The links that form a joint between the front climbing mechanism and the positioning mechanism are represented by the constants l_1 and l_3. The links that form a joint between the rear climbing mechanism and the positioning mechanism are denoted by the constants l_4 and l_6. The distance between the central joint and the position of the center of mass of the chassis ($\mathbf{P}_g$) is, meanwhile, represented by l_5. The angles μ_i are defined to connect the vectors that comprise the general kinematic scheme. Each of the angles have a different constant value. In the kinematic model it is assumed that the rear and front wheels roll on both a flat surface ($\alpha_i = 0$) and an uneven surface ($\alpha_i \neq 0$), and in both cases $i = 1, 2$. The direct kinematic model for each configuration is shown in Table 1. For a full description of the model, see [39–41].

Table 1. Initial expressions that define the current position of the SCMS.

Configurations	Expressions	
1 (see Figure 11a)	$\mathbf{P}_g = \mathbf{f}(\theta_3) + l_6 e^{j(\gamma + \frac{\pi}{2} + \mu_6)} + l_4 e^{j(\gamma + \frac{3\pi}{2} - \theta_2)} + l_5 e^{j(\gamma + \frac{\pi}{2} + \mu_5)}$	(1)
	$\mathbf{P}_g = \mathbf{f}(\theta_4) + l_1 e^{j(\gamma + \frac{\pi}{2} + \mu_1)} - l_3 e^{j(\gamma + \frac{\pi}{2} + \theta_1)} + l_5 e^{j(\gamma + \frac{\pi}{2} + \mu_5)}$	(2)
2 (see Figure 11b)	$\mathbf{P}_g = \mathbf{P}_{C2} + z_2 e^{j(\gamma + \frac{\pi}{2} - \delta_2)} + l_6 e^{j(\gamma + \frac{\pi}{2} + \mu_6)} + l_4 e^{j(\gamma + \frac{3\pi}{2} - \theta_2)} + l_5 e^{j(\gamma + \frac{\pi}{2} + \mu_5)}$	(3)
	Equation (2)	
3 (see Figure 11c)	Equation (1)	
	$\mathbf{P}_g = \mathbf{P}_{C1} + z_1 e^{j(\gamma + \frac{\pi}{2} - \delta_1)} + l_1 e^{j(\gamma + \frac{\pi}{2} + \mu_1)} - l_3 e^{j(\gamma + \frac{\pi}{2} + \theta_1)} + l_5 e^{j(\gamma + \frac{\pi}{2} + \mu_5)}$	(4)
4 (see Figure 11d)	Equation (3)	
	Equation (4)	

Upon computing the difference between the two equations that define the current position of the SCMS in each of the four configurations and taking the imaginary part, an implicit expression (this

implicit expression has a Jacobian form, which we call an implicit posture Jacobian model) that defines the posture of the SCMS can be obtained in a compact form as:

$$\mathbf{F}_k(\mathbf{q}) = \mathbf{0} \tag{5}$$

where, in this equation and those that follow, index k indicates the SCMS configuration ($k = 1, 2, 3$ or 4).

In the kinematic model, it is assumed that the velocities of the active DOFs of the SCMS, $\dot{\mathbf{q}} = \left[\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3, \dot{z}_1, \dot{z}_2\right]^T$, are the input control variables, and the derivative of the inclination angle of the SCMS frame, $\dot{\gamma}$, is the output. The differential relationship between the output γ and the variable $\mathbf{q}$ is therefore given by:

$$\frac{\partial \mathbf{F}_k}{\partial \gamma}\dot{\gamma} + \frac{\partial \mathbf{F}_k}{\partial \mathbf{q}}\dot{\mathbf{q}} = 0 \tag{6}$$

The locomotion control is achieved by means of actuation on joints θ_3, z_1 and z_2, while the posture control is achieved by means of actuation on joints θ_1 and θ_2. The relationship between the derivative of the controlled variable γ and the control signals $\dot{\theta}_1$ and $\dot{\theta}_2$ is therefore:

$$\frac{\partial \mathbf{F}_k}{\partial \gamma}\dot{\gamma} + \frac{\partial \mathbf{F}_k}{\partial \theta_1}\dot{\theta}_1 + \frac{\partial \mathbf{F}_k}{\partial \theta_1}\dot{\theta}_2 + \mathbf{g}_k(\mathbf{q}, \dot{\mathbf{q}}) = 0 \tag{7}$$

in which:

$$\mathbf{g}_k(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\partial \mathbf{F}_k}{\partial \theta_3}\dot{\theta}_3 + \frac{\partial \mathbf{F}_k}{\partial z_1}\dot{z}_1 + \frac{\partial \mathbf{F}_k}{\partial z_2}\dot{z}_2$$

The terms $\frac{\partial \mathbf{F}_k}{\partial \gamma}$, $\frac{\partial \mathbf{F}_k}{\partial \theta_1}$, $\frac{\partial \mathbf{F}_k}{\partial \theta_2}$ and $\mathbf{g}_k(\mathbf{q}, \dot{\mathbf{q}})$ for the four configurations are shown in Table 2.

Table 2. Terms of differential equations of the implicit Jacobian for the four configurations (where $\sin(\cdot) \equiv S_{(\cdot)}$ and $\cos(\cdot) \equiv C_{(\cdot)}$).

k	$\frac{\partial F_k}{\partial \gamma}$	$\frac{\partial F_k}{\partial \theta_1}$	$\frac{\partial F_k}{\partial \theta_2}$	$g_k(\mathbf{q}, \dot{\mathbf{q}})$
1	$l_1 S_{(\gamma+\mu_1)} - l_6 S_{(\gamma+\mu_6)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)}$	$-l_3 S_{(\gamma+\theta_1)}$	$-l_4 S_{(\gamma-\theta_2)}$	
2	$l_1 S_{(\gamma+\mu_1)} - l_6 S_{(\gamma+\mu_6)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)} - z_2 S_{(\gamma-\delta_2)}$	$-l_3 S_{(\gamma+\theta_1)}$	$-l_4 S_{(\gamma-\theta_2)}$	$+C_{(\gamma-\delta_2)}\dot{z}_2$
3	$l_1 S_{(\gamma+\mu_1)} - l_6 S_{(\gamma+\mu_6)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)} + z_1 S_{(\gamma-\delta_1)}$	$-l_3 S_{(\gamma+\theta_1)}$	$-l_4 S_{(\gamma-\theta_2)}$	$-C_{(\gamma-\delta_1)}\dot{z}_1$
4	$l_1 S_{(\gamma+\mu_1)} - l_6 S_{(\gamma+\mu_6)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)} + z_1 S_{(\gamma-\delta_1)}$ $-z_2 S_{(\gamma-\delta_2)}$	$-l_3 S_{(\gamma+\theta_1)}$	$-l_4 S_{(\gamma-\theta_2)}$	$-C_{(\gamma-\delta_1)}\dot{z}_1$ $+C_{(\gamma-\delta_2)}\dot{z}_2$

3.2. Dynamics Modeling

We develop a dynamic model under quasi-static conditions since the SCMS is assumed to be slow moving. The dynamic model is also different for each of the SCMS's configurations and is expressed in terms of generalized coordinates. The relation between these coordinates and the system coordinates in each particular configuration is also different, and this difference is considered in order to adapt different equations to different configurations. We first obtain explicit expressions of the forces that appear in the mechanism in terms of the generalized coordinate variables, $\mathbf{r}_1 = [r_1, r_3]^T$ and $\mathbf{r}_2 = [r_1, r_4]^T$.

Figure 12 shows the notation used to obtain the dynamics model for each of the SCMS's configurations. With this notation, the position, velocity and acceleration of the center of mass can be written in the following general form:

$$\overline{\mathbf{OP}}_g = \begin{bmatrix} \overline{\mathbf{OP}}_{gx} \\ \overline{\mathbf{OP}}_{gy} \end{bmatrix}_k = \begin{bmatrix} x_{1,2} \\ y_{1,2} \end{bmatrix}_k, \quad 1 \le k \le 4 \tag{8}$$

$$\dot{\overline{\mathbf{OP}}}_g = \begin{bmatrix} \dot{\overline{\mathbf{OP}}}_{gx} \\ \dot{\overline{\mathbf{OP}}}_{gy} \end{bmatrix}_k = \begin{bmatrix} \dot{x}_{1,2} \\ \dot{y}_{1,2} \end{bmatrix}_k, \quad 1 \le k \le 4 \tag{9}$$

$$\overline{\dddot{OP}}_g = \begin{bmatrix} \overline{\dddot{OP}}_{gx} \\ \overline{\dddot{OP}}_{gy} \end{bmatrix}_k = \begin{bmatrix} \ddot{x}_{1,2} \\ \ddot{y}_{1,2} \end{bmatrix}_k, \quad 1 \le k \le 4 \tag{10}$$

where subscripts 1 and 2 represent two ways in which to connect with the center of mass; subscript 1 is associated with r_1, which is generated on the front axle, and subscript 2 is associated with r_2, which is generated on the rear axle. The expressions $[x_{1,2}, y_{1,2}]^T$, $[\dot{x}_{1,2}, \dot{y}_{1,2}]^T$, $[\ddot{x}_{1,2}, \ddot{y}_{1,2}]^T$ for Configuration 1 are shown as follows:

$$
\begin{aligned}
x_1 &= (l_1 + l_5)C_{r_1} + l_3 C_{(r_1+r_3)} \\
y_1 &= (l_1 + l_5)S_{r_1} + l_3 S_{(r_1+r_3)} \\
\dot{x}_1 &= -(l_1 + l_5)S_{r_1}\dot{r}_1 - l_3 S_{(r_1+r_3)}(\dot{r}_1 + \dot{r}_3) \\
\dot{y}_1 &= (l_1 + l_5)C_{r_1}\dot{r}_1 + l_3 C_{(r_1+r_3)}(\dot{r}_1 + \dot{r}_3) \\
\ddot{x}_1 &= -(l_1 + l_5)(C_{r_1}\dot{r}_1^2 + S_{r_1}\ddot{r}_1) - l_3[C_{(r_1+r_3)}(\dot{r}_1 + \dot{r}_3)^2 + S_{(r_1+r_3)}(\ddot{r}_1 + \ddot{r}_3)] \\
\ddot{y}_1 &= (l_1 + l_5)(-S_{r_1}\dot{r}_1^2 + C_{r_1}\ddot{r}_1) + l_3[-S_{(r_1+r_3)}(\dot{r}_1 + \dot{r}_3)^2 + C_{(r_1+r_3)}(\ddot{r}_1 + \ddot{r}_3)]
\end{aligned}
\tag{11}
$$

or:

$$
\begin{aligned}
x_2 &= (l_6 + l_5)C_{r_1} + l_4 C_{(r_1+r_4)} \\
y_2 &= (l_6 + l_5)S_{r_1} + l_4 S_{(r_1+r_4)} \\
\dot{x}_2 &= -(l_6 + l_5)S_{r_1}\dot{r}_1 - l_4 S_{(r_1+r_4)}(\dot{r}_1 + \dot{r}_4) \\
\dot{y}_2 &= (l_6 + l_5)C_{r_1}\dot{r}_1 + l_4 C_{(r_1+r_4)}(\dot{r}_1 + \dot{r}_4) \\
\ddot{x}_2 &= -(l_6 + l_5)[C_{r_1}\dot{r}_1^2 + S_{r_1}\ddot{r}_1] - l_4[C_{(r_1+r_4)}(\dot{r}_1 + \dot{r}_4)^2 + S_{(r_1+r_4)}(\ddot{r}_1 + \ddot{r}_4)] \\
\ddot{y}_2 &= (l_6 + l_5)[-S_{r_1}\dot{r}_1^2 + C_{r_1}\ddot{r}_1] + l_4[-S_{(r_1+r_4)}(\dot{r}_1 + \dot{r}_4)^2 + C_{(r_1+r_4)}(\ddot{r}_1 + \ddot{r}_4)]
\end{aligned}
\tag{12}
$$

In these equations and those that follow, $\sin(\cdot) \equiv S_{(\cdot)}$ and $\cos(\cdot) \equiv C_{(\cdot)}$.

(a) Configuration 1   (b) Configuration 2   (c) Configuration 3   (d) Configuration 4
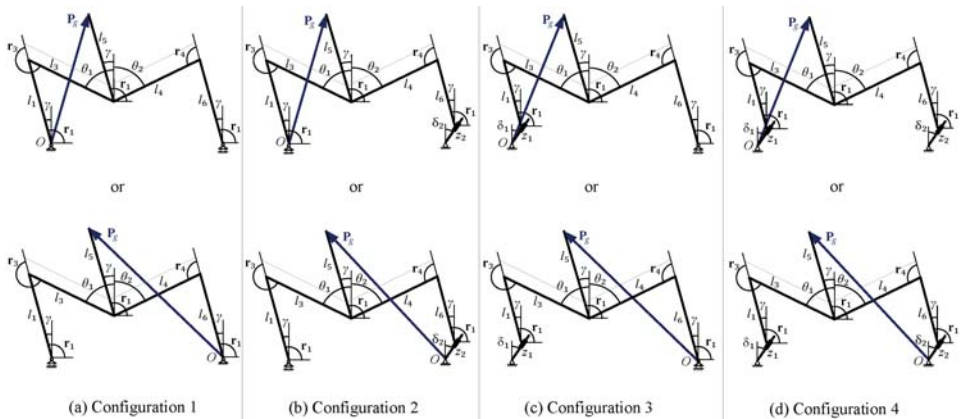
Figure 12. Definition of the generalized coordinate variables for each of the configurations of the SCMS.

Let $\mathbf{F}_m$ and $\mathbf{F}_g$ be the vectors of forces acting on the center of mass of the SCMS ($\mathbf{P}_g$), generated as a result of the acceleration of the system and gravity, respectively. $\mathbf{F}_m$ and $\mathbf{F}_g$ can be obtained as follows:

$$\mathbf{F}_m = m \cdot \overline{\dddot{OP}}_g = m \cdot [\ddot{x}_{1,2} \; \ddot{y}_{1,2}]^T, \tag{13}$$

$$\mathbf{F}_g = m[0 \quad -g]^T, \tag{14}$$

where m is the mass of the whole system (SCMS + passenger). It is important to mention two considerations regarding the dynamic model: first, that the mass of the system is invariant; and second, that the position of the center of mass barely changes during the stair-climbing process (it is assumed that the user will not make abrupt movements [42,43]).

If the results of the forces provided by Expressions (13) and (14) are grouped, and Newton's Second Law is applied, the generalized forces exerted on the center of mass of the mechanism, $\mathbf{F}_{Pg}$, are obtained from the following expressions:

$$\mathbf{F}_{P_g} = m \cdot \left[\mathbf{A}_1 \cdot \begin{bmatrix} \ddot{r}_1 \\ \ddot{r}_3 \end{bmatrix} + \mathbf{B}_1 \cdot \begin{bmatrix} \dot{r}_1 \\ \dot{r}_3 \end{bmatrix} + \begin{bmatrix} 0 \\ -g \end{bmatrix} \right] \tag{15}$$

where:

$$\mathbf{A}_1 = \begin{bmatrix} -(l_1+l_5)S_{r_1} - l_3 S_{(r_1+r_3)} & -l_3 S_{(r_1+r_3)} \\ (l_1+l_5)C_{r_1} + l_3 C_{(r_1+r_3)} & l_3 C_{(r_1+r_3)} \end{bmatrix}$$

$$\mathbf{B}_1 = \begin{bmatrix} -[(l_1+l_5)C_{r_1} + l_3 C_{(r_1+r_3)}]\dot{r}_1 & -l_3 C_{(r_1+r_3)}(2\dot{r}_1 + \dot{r}_3) \\ -[(l_1+l_5)S_{r_1} + l_3 S_{(r_1+r_3)}]\dot{r}_1 & -l_3 S_{(r_1+r_3)}(2\dot{r}_1 + \dot{r}_3) \end{bmatrix}$$

or:

$$\mathbf{F}_{P_g} = m \cdot \left[\mathbf{A}_2 \cdot \begin{bmatrix} \ddot{r}_1 \\ \ddot{r}_4 \end{bmatrix} + \mathbf{B}_2 \cdot \begin{bmatrix} \dot{r}_1 \\ \dot{r}_4 \end{bmatrix} + \begin{bmatrix} 0 \\ -g \end{bmatrix} \right] \tag{16}$$

where:

$$\mathbf{A}_2 = \begin{bmatrix} -(l_6+l_5)S_{r_1} - l_4 S_{(r_1+r_4)} & -l_4 S_{(r_1+r_4)} \\ (l_6+l_5)C_{r_1} + l_4 C_{(r_1+r_4)} & l_4 C_{(r_1+r_4)} \end{bmatrix}$$

$$\mathbf{B}_2 = \begin{bmatrix} -[(l_6+l_5)C_{r_1} + l_4 C_{(r_1+r_4)}]\dot{r}_1 & -l_4 C_{(r_1+r_4)}(2\dot{r}_1 + \dot{r}_4) \\ -[(l_6+l_5)S_{r_1} + l_4 S_{(r_1+r_4)}]\dot{r}_1 & -l_4 S_{(r_1+r_4)}(2\dot{r}_1 + \dot{r}_4) \end{bmatrix}$$

Moreover, it is well known that under quasi-static conditions, the relationship between the torques exerted on the joints and the forces and torques exerted on the center of gravity of the mechanism are related by mean of the following expression:

$$\tau = \mathbf{J}^T(\mathbf{r}_{1,2})\mathbf{F}_{P_g}, \tag{17}$$

in which the Jacobians, $\mathbf{J}(\mathbf{r}_1)$ and $\mathbf{J}(\mathbf{r}_2)$, are obtained directly from Equation (8) as follows:

$$\mathbf{J}(\mathbf{r}_1) = \begin{bmatrix} \frac{\partial x}{\partial r_1} & \frac{\partial x}{\partial r_3} \\ \frac{\partial y}{\partial r_1} & \frac{\partial y}{\partial r_3} \end{bmatrix}_k \quad 1 \le k \le 4 \tag{18}$$

or:

$$\mathbf{J}(\mathbf{r}_2) = \begin{bmatrix} \frac{\partial x}{\partial r_1} & \frac{\partial x}{\partial r_4} \\ \frac{\partial y}{\partial r_1} & \frac{\partial y}{\partial r_4} \end{bmatrix}_k \quad 1 \le k \le 4 \tag{19}$$

and the values of the partial derivatives are provided by the following expressions:

$$\mathbf{J}(\mathbf{r}_1) = \begin{bmatrix} -(l_1+l_5)S_{r_1} - l_3 S_{(r_1+r_3)} & -l_3 S_{(r_1+r_3)} \\ (l_1+l_5)C_{r_1} + l_3 C_{(r_1+r_3)} & l_3 C_{(r_1+r_3)} \end{bmatrix} \tag{20}$$

or

$$\mathbf{J}(\mathbf{r}_2) = \begin{bmatrix} -(l_6+l_5)S_{r_1} - l_4 S_{(r_1+r_4)} & -l_4 S_{(r_1+r_4)} \\ (l_6+l_5)C_{r_1} + l_4 C_{(r_1+r_4)} & l_4 C_{(r_1+r_4)} \end{bmatrix} \tag{21}$$

Upon substituting Equations (20) and (21) in Equation (17), and after certain algebraic manipulations, the following dynamics model is obtained for the SCMS:

$$\boldsymbol{\tau} = \mathbf{B}(\mathbf{r})\ddot{\mathbf{r}} + \mathbf{C}(\mathbf{r},\ \dot{\mathbf{r}})\dot{\mathbf{r}} + \mathbf{G}(\mathbf{r}), \tag{22}$$

where the values of the matrices $\mathbf{B}(\mathbf{r})$, $\mathbf{C}(\mathbf{r},\ \dot{\mathbf{r}})$ and $\mathbf{G}(\mathbf{r})$ are provided by the following expressions:

$$\mathbf{B}(\mathbf{r}_1) = m \begin{bmatrix} (l_1+l_5)^2 + l_3^2 + 2l_3(l_1+l_5)C_{r_3} & l_3^2 + l_3(l_1+l_5)C_{r_3} \\ l_3^2 + l_3(l_1+l_5)C_{r_3} & l_3^2 \end{bmatrix}$$

$$\mathbf{C}(\mathbf{r}_1,\ \dot{\mathbf{r}}_1) = m \begin{bmatrix} 0 & -l_3(l_1+l_5)S_{r_3}(2\dot{r}_1 + \dot{r}_3) \\ l_3(l_1+l_5)S_{r_3}\dot{r}_1 & 0 \end{bmatrix}$$

$$\mathbf{G}(\mathbf{r}_1) = mg \begin{bmatrix} (l_1+l_5)C_{r_1} + l_3 C_{(r_1+r_3)} \\ l_3 C_{(r_1+r_3)} \end{bmatrix}$$

or:

$$\mathbf{B}(\mathbf{r}_2) = m \begin{bmatrix} (l_6+l_5)^2 + l_4^2 + 2l_4(l_6+l_5)C_{r_4} & l_4^2 + l_4(l_6+l_5)C_{r_4} \\ l_4^2 + l_4(l_6+l_5)C_{r_4} & l_4^2 \end{bmatrix}$$

$$\mathbf{C}(\mathbf{r}_2,\ \dot{\mathbf{r}}_2) = m \begin{bmatrix} 0 & -l_4(l_6+l_5)S_{r_4}(2\dot{r}_1 + \dot{r}_4) \\ l_4(l_6+l_5)S_{r_4}\dot{r}_1 & 0 \end{bmatrix}$$

$$\mathbf{G}(\mathbf{r}_2) = mg \begin{bmatrix} (l_6+l_5)C_{r_1} + l_4 C_{(r_1+r_4)} \\ l_4 C_{(r_1+r_4)} \end{bmatrix}$$

The computation of articular variables and their corresponding derivatives ($\mathbf{r}_1 = [r_1, r_3]^T$, $\mathbf{r}_2 = [r_1, r_4]^T$, $\dot{\mathbf{r}}_1 = [\dot{r}_1, \dot{r}_3]^T$, $\dot{\mathbf{r}}_2 = [\dot{r}_1, \dot{r}_4]^T$, $\ddot{\mathbf{r}}_1 = [\ddot{r}_1, \ddot{r}_3]^T$ and $\ddot{\mathbf{r}}_2 = [\ddot{r}_1, \ddot{r}_4]^T$), which are expressed in terms of the system variables ($\mathbf{q} = [\theta_1, \theta_2, \theta_3, z_1, z_2]^T$, $\dot{\mathbf{q}} = [\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3, \dot{z}_1, \dot{z}_2]^T$ and $\ddot{\mathbf{q}} = [\ddot{\theta}_1, \ddot{\theta}_2, \ddot{\theta}_3, \ddot{z}_1, \ddot{z}_2]^T$), will depend on the particular configuration of the mechanism. It is, therefore, necessary to find a relationship between the articular variables and the system variables for all the possible configurations and to use the model provided by Equation (22) in order to obtain a new model as a function of $\mathbf{q}$ rather than $\mathbf{r}$, i.e.,

$$\boldsymbol{\tau} = \mathbf{B}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q},\ \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{G}(\mathbf{q}), \tag{23}$$

If the proposed dynamics model and the system geometry are used and we take advantage of the fact that the front and rear frames of the positioning mechanism do not rotate with regard to the central frame, the following relationships are obtained (see Figure 12):

$$r_1 = \frac{\pi}{2} + \gamma; \quad r_3 = \pi + \theta_1; \quad r_4 = \pi - \theta_2 \tag{24}$$

The relationships for velocities and accelerations are obtained by differentiating the expressions shown above:

$$\dot{r}_1 = \dot{\gamma}; \quad \dot{r}_3 = \dot{\theta}_1; \quad \dot{r}_4 = -\dot{\theta}_2 \tag{25}$$

$$\ddot{r}_1 = \ddot{\gamma}; \quad \ddot{r}_3 = \ddot{\theta}_1; \quad \ddot{r}_4 = -\ddot{\theta}_2 \tag{26}$$

This system property makes it necessary to obtain mathematical relationships between the articular variables and control variables of the actuated DOF of the SCMS for all possible configurations. As is shown in Figure 12, the vertical component of the center of mass is the same in both cases in each of the configurations. We can therefore assume that:

$$\overline{\mathrm{OP}}_{g_{y_1}} = \overline{\mathrm{OP}}_{g_{y_2}} \tag{27}$$

$$\dot{\overline{OP}}_{g_{y_1}} = \dot{\overline{OP}}_{g_{y_2}} \tag{28}$$

$$\ddot{\overline{OP}}_{g_{y_1}} = \ddot{\overline{OP}}_{g_{y_2}} \tag{29}$$

Upon taking the aforementioned considerations into account, substituting relations (24)–(26) in Equations (11) and (12), and after certain algebraic manipulations, the first and second derivatives of the inclination angle of the SCMS frame with regard to time ($\dot{\gamma}$ and $\ddot{\gamma}$) are obtained:

$$\dot{\gamma} = \frac{-l_3 S_{(\gamma+\theta_1)}\dot{\theta}_1 - l_4 S_{(\gamma-\theta_2)}\dot{\theta}_2}{(l_1-l_6)S_{(\gamma)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)}} \tag{30}$$

$$\ddot{\gamma} = \frac{(l_1-l_6)C_{(\gamma)}\dot{\gamma}^2 + l_3[C_{(\gamma+\theta_1)}(\dot{\gamma}+\dot{\theta}_1)^2 - S_{(\gamma+\theta_1)}\ddot{\theta}_1] - l_4[C_{(\gamma-\theta_2)}(\dot{\gamma}-\dot{\theta}_2)^2 + S_{(\gamma-\theta_2)}\ddot{\theta}_2]}{(l_1-l_6)S_{(\gamma)} + l_4 S_{(\gamma-\theta_2)} - l_3 S_{(\gamma+\theta_1)}} \tag{31}$$

4. Definition of the Trajectory Generator

During the stair-climbing process, the SCMS must realize a trajectory from an initial state to a final one, both of which are characterized by null velocities and null accelerations (this occurs every time that the SCMS changes from one configuration to another). Solving the optimal trajectory planning problem involves the determination of the transfer time T, the trajectory $\Im(t) = (P_{gx}(t), P_{gy}(t), \gamma(t))$ and the corresponding input controls $\Gamma(t) = (\theta_1(t), \theta_2(t), \theta_3(t), z_1(t), z_2(t))$ such that the initial and final states are matched, constraints are respected and a cost function is minimized. Hereafter, we exclusively adopt the following cost function:

$$F^{obj} = \alpha \int_0^T dt + (1-\alpha) \frac{1}{T} \int_0^T \gamma^2(t)dt \tag{32}$$

where $0 \le \alpha \le 1$. This function provides a compromise between the stair-climbing time process and the passenger's comfort, thus allowing the threshold of accuracy to be selected according to the specific needs of application.

The boundary conditions of position, velocity and acceleration inherent to the achievement of the optimal trajectory planning are those imposed on the SCMS:

$$\Im(0) = \Im_{ini} \text{ and } \Im(T) = \Im_{fin} \tag{33}$$

$$\dot{\Im}(0) = \vec{0} \text{ and } \dot{\Im}(T) = \vec{0} \tag{34}$$

$$\ddot{\Im}(0) = \vec{0} \text{ and } \ddot{\Im}(T) = \vec{0} \tag{35}$$

Other boundary conditions that need to be fulfilled during the stair-climbing movement are provided below:

- A function $Tran(\Im(t))$ is used to indicate whether it is possible to change the configuration of the SCMS:

$$Tran(\Im(t)) = true \tag{36}$$

- Bounds on the SCMS configurations:

$$\Im_{min} \le \Im(t) \le \Im_{max} \tag{37}$$

These bounds make it possible for the SCMS to surpass stairs of different sizes and to maintain an inclination with regard to the direction of gravity of its chassis ($\gamma = 0$).
- Bounds on the actuator velocities and accelerations:

$$\dot{\theta}_{imin} \le \dot{\theta}_i(t) \le \dot{\theta}_{imax}, \quad i = 1, 2 \text{ and } 3 \tag{38}$$

$$\ddot{\theta}_{imin} \leq \ddot{\theta}_i(t) \leq \ddot{\theta}_{imax}, \quad i = 1, 2 \text{ and } 3 \tag{39}$$

$$\dot{z}_{imin} \leq \dot{z}_i(t) \leq \dot{z}_{imax}, \quad i = 1 \text{ and } 2 \tag{40}$$

$$\ddot{z}_{imin} \leq \ddot{z}_i(t) \leq \ddot{z}_{imax}, \quad i = 1 \text{ and } 2 \tag{41}$$

which arise from the fact that each motor or actuator has a limited speed of operation. These bounds ensure that the inertial forces are smaller than the gravitational force. The mass center accelerations and velocities of the whole system (SCMS + passenger) must therefore be less than the maximum comfort acceleration and velocity, respectively. The inclination angle must have the acceptable comfort ranges, $\pm 10°$.

- Bounds on the comfort velocities and accelerations:

$$\dot{\gamma}_{min} \leq \dot{\gamma}(t) \leq \dot{\gamma}_{max} \tag{42}$$

$$\ddot{\gamma}_{min} \leq \ddot{\gamma}(t) \leq \ddot{\gamma}_{max} \tag{43}$$

- A function $Col(\Im(t))$ is also used to indicate whether or not a given wheel or sliding support of the SMCS is accidentally colliding with an obstacle (Information regarding the environment is obtained by using the strategy designed to estimate the size and shape of the stairs, which was presented in Section 2.3. In this case, knowledge about the environment allows the proposed prototype to determine whether or not a collision will occur):

$$Col(\Im(t)) = false \tag{44}$$

It is possible to use Equations (30) and (31) to express all the elements of the optimization problem (cost function and constraints) as a function of the time evolution of only five configuration parameters, namely $\{\theta_1(t), \theta_2(t), \theta_3(t), z_1(t), z_2(t)\}$ and their time derivatives. In fact, it is possible to define a trajectory candidate by defining the evolution of $\{\theta_1(t), \theta_2(t), \theta_3(t), z_1(t), z_2(t)\}$, for $t \in [0, T]$, while accounting for boundary conditions (33)–(35) and bounds (36)–(44) in the SCMS configurations.

The remaining question is how to generate the trajectory candidates. We therefore propose considering that any trajectory candidate is a composition of: (i) a parametric form $\mathbf{P}(v) = (\theta_1(v), \theta_2(v), \theta_3(v), z_1(v), z_2(v))$, $v \in [0, 1]$, which defines the SCMS path; and (ii) a monotonically increasing function $v(t)$, $t \in [0, T]$, which specifies the motion on this path, i.e.,

$$\Im(t) = \mathbf{P}(v) \circ v(t) \tag{45}$$

The optimal evolutions of $\mathbf{P}(v)$ and $v(t)$, which are approximated using B-spline functions by fitting a set of control points, are found using a nonlinear optimization technique. It is possible to construct $\mathbf{P}(v)$ and $v(t)$ as follows:

- Generate $\mathbf{P}(v)$, $v \in [0, 1]$ (Figure 13a) by means of a quintic B-spline model. The fifth degree ensures the continuity of the second order derivatives of angles $\theta_1, \theta_2, \theta_3, z_1$, and z_2. This B-spline is built by using a set of N_p control points (five points at least), and is generated within the admissible workspace defined by (37) and by accounting for constraints (33) and (44).
- Build the motion profile $v(t)$ on the interval $[0, T]$ using a quintic B-spline model (Figure 13b), generated by N_m control points that are uniformly distributed throughout the time scale, and by accounting for the following boundary conditions:

$$v(0) = 0 \quad \dot{v}(0) = 0 \quad \ddot{v}(0) = 0 \quad \dddot{v}(0) = 0$$

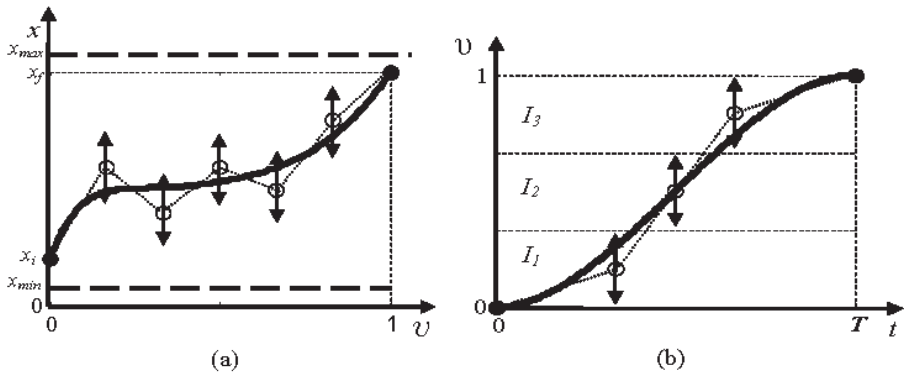$$v(T) = 1 \quad \dot{v}(T) = 0 \quad \ddot{v}(T) = 0 \quad \dddot{v}(T) = 0$$

Figure 13. Path and motion profiles. (**a**) A path component profile and (**b**) the motion profile.

These conditions ensure the compatibility of the resulting trajectory $\Im(t)$ with constraints (34) and (35). The increasing monotony of the motion profile is ensured by generating the N_m control points in I_{N_m} intervals, as shown in (Figure 13b):

$$I_i = \left[\frac{i-1}{N_m}, \frac{i}{N_m} \right] \quad i = 1, ..., N_m$$

The design parameters of $\mathbf{P}(v)$ and $v(t)$, which are the coordinates of the B-spline function control points and the value of T, become the sole unknown values of the trajectory generation problem.

Optimization and the SCMS's Control Scheme

The control scheme shown in Figure 14 was designed and implemented in order to achieve an accurate tracking of the desired inclination angle ($\gamma^* = 0$) and the desired center of mass trajectory ($\mathbf{P}_g^*$) during the stair-climbing process. In this scheme, the trajectory generator of the behavior diagram module was designed without taking trajectory optimization into account.
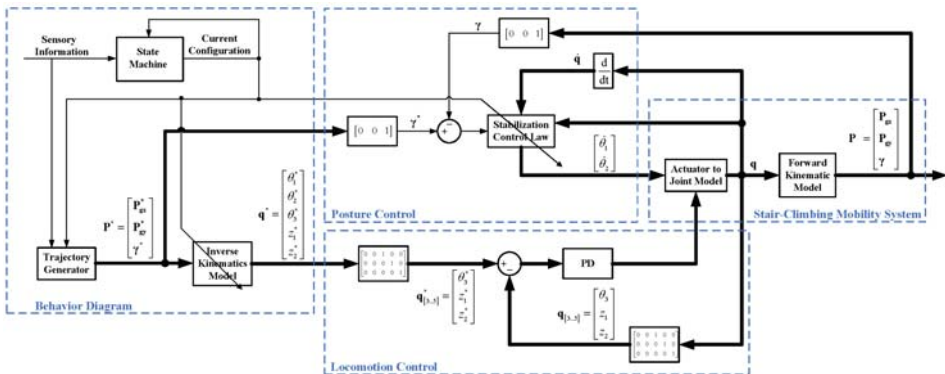
Figure 14. Control scheme of the proposed SCMS.

In order to solve the trajectory optimization problem proposed in this paper, a sub-module, called trajectory optimization, has been incorporated into the behavior diagram module, as shown in Figure 15.
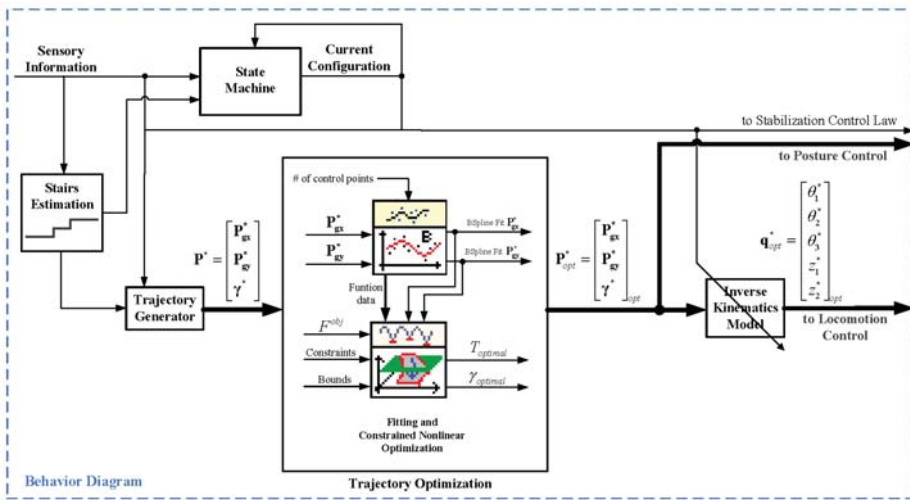
Figure 15. Scheme of the optimal trajectory generator based on constrained nonlinear optimization.

In the modified module, the State Machine controls the sequence of states that the SCMS should pass through. The trajectory generator provides the desired trajectory profiles $\mathbf{P}^*$ that the SCMS will track. Both the machine and the generator depend directly on the information provided by the sensory system and on the previous and current configurations. The generator also depends on the knowledge of the dimensions of the stairs. In the trajectory optimization sub-module, the reference trajectories $\mathbf{P}^*$ are then optimized by using the approach shown in Figure 15. In this sub-module, the reference trajectories $\mathbf{P}^*_{opt}$ are then generated and these are used to control the system.

5. Experimental Results

This section provides details of the experimental setup used to evaluate the development of the hybrid SCMS using the trajectory optimization proposed in this paper.

5.1. Experimental Setup

Two stair-climbing experiments were configured in order to make the SCMS climb up a two-step staircase. In both experiments, the SCMS was controlled using the control scheme shown in Figure 14 in conjunction with a PI controller proposed in [44]. The position profile of the center of mass ($\mathbf{P}_g$) used in the first experiment consisted of straight lines with slopes of 0 and 5° (see Figure 16a). The horizontal lines are attributed to Configurations 1 and 3 while the others are attributed to Configuration 2. This approach has two advantages: the control is simplified and the power consumption is minimized. The position profile used in the second experiment consisted of two straight lines, one vertical and the other horizontal (see Figure 16b). The vertical line is generated at the beginning of the stair-climbing process (with Configuration 1), while the horizontal line is generated in the rest of the process (with all the configurations). The main advantage of this approach is observed during locomotion, during which the effect of climbing stairs is minimized. Both profiles have been generated by considering the information regarding stair geometry.

In the experiments, a volunteer was seated on the hybrid SCMS in order to obtain a real interaction between the user and the system. The experiments were performed using the parameters shown in Tables 3–5.
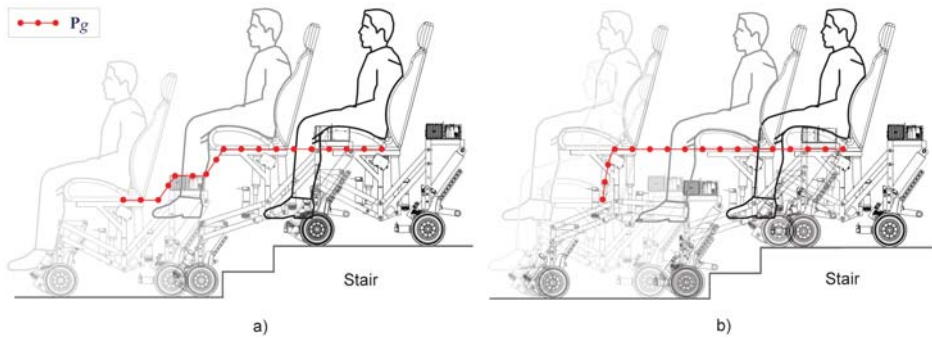
Figure 16. Position profiles of the center of mass of the SCMS used in the (**a**) first and (**b**) second experiments.

Table 3. Specifications of the SCMS.

Variable	Value
User's weight	70 kg
Vehicle weight	60 kg
Height steps	150 mm
Wide steps	300 mm
Sampling time	15 mm

Table 4. Dimensions of the SCMS's links.

Variable	Value
l_1	270 mm
l_3	410 mm
l_4	420 mm
l_5	323 mm
l_6	390 mm
δ_i	35 deg
r_{wheel}	100 mm

Table 5. Specifications of the motors and actuators.

Variable and Value
$\gamma_{mim} = 10°$
$\gamma_{max} = -10°$
$\acute{\gamma}_{max} = \pm 0.22\,°/\text{s}$
$\ddot{\gamma}_{max} = \pm 0.1\,°/\text{s}^2$
$\theta_{1mim} = 1.02$ rad
$\theta_{2mim} = 0.7$ rad
$\theta_{1max} = 2.32$ rad
$\theta_{2max} = 1.97$ rad
$\dot{\theta}_{1max} = \dot{\theta}_{2max} = \pm 0.12$ rad/s
$\ddot{\theta}_{1max} = \ddot{\theta}_{2max} = \pm 0.03$ rad/s^2
$\dot{\theta}_{3max} = \pm 4$ rad/s
$\ddot{\theta}_{3max} = \pm 0.24$ rad/s^2
$z_{1max} = z_{2max} = 260$ mm
$\dot{z}_{1max} = \dot{z}_{2max} = \pm 33$ mm/s
$\ddot{z}_{1max} = \ddot{z}_{2max} = \pm 0.6$ mm/s^2

The control system of the proposed SCMS incorporates a reliable and robust hardware that has been chosen by taking into account the technical aspects of control and communication. The physical hardware of the control system and its general architecture are illustrated in Figure 17.
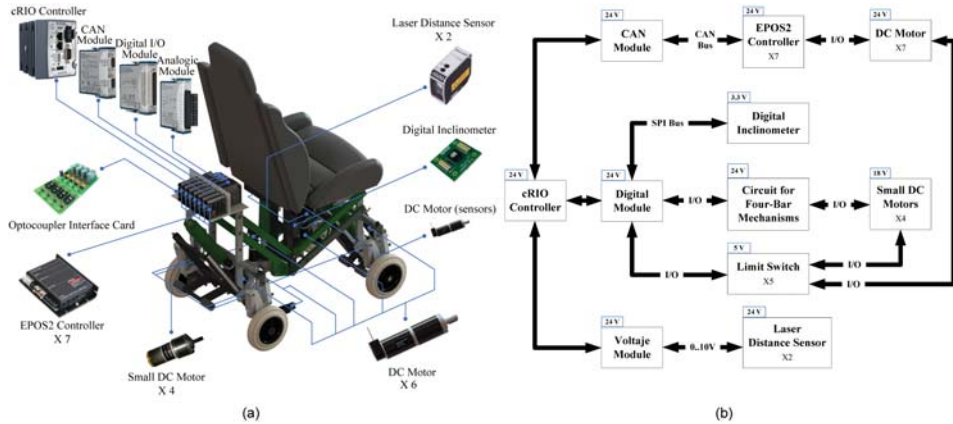
Figure 17. (**a**) Physical and (**b**) functional perspective of the hardware control system.

The control system architecture and the interface for the communication system have been developed for the purpose of solving the stair-climbing problem in an efficient and autonomous manner that ensures the passenger's safety and comfort. The proposed SCMS is controlled by an advanced reconfigurable control and acquisition system, CompactRIO, which is manufactured by National Instruments (NI). It consists of a real-time processor that is connected to a reconfigurable chassis using a local Peripheral Component Interconnect (PCI) bus interface. The overall control has been developed using the CompactRIO system with LabVIEW based control programs. The Qualisys Optotrack motion capture system was also used to capture the position of the center of mass ($\mathbf{P}_g$).

Functions such as "Quadratic Programming.vi", "B-Spline Fit.vi" and "Constrained Nonlinear Optimization.vi" are software units coded in LabVIEW and copyrighted from National Instruments used for testing when solving the optimization problems, along with other programming tools and solvers of differential equations. In this work, the optimization task is performed by using both the "B-Spline Fit" and the "Constrained Nonlinear Optimization" Virtual Instruments (VIs). The documentation of the functions is referred to in [45,46], respectively.

The Virtual Instrument (VI) (LabVIEW is a graphical programming language that uses icons rather than lines of text to create applications. LabVIEW programs are called Virtual Instruments (VIs) because their appearance and operation imitate physical instruments, such as generators, oscilloscopes, analyzers and multimeters) "B-Spline Fit" (see Figure 18a), one of the LabVIEW fitting routines, is used to approximate $\mathbf{P}(v)$ and $v(t)$. Their optimal values are approximated by fitting a set of control points that are easily found using an SQP technique. In these experiments, $\mathbf{P}(v)$ and $v(t)$ are generated by means of a quintic B-spline model. The fifth degree ensures the continuity of the second derivative of the paths. In $\mathbf{P}(v)$, the B-spline is built by using a set of five control points, while in $v(t)$, it is generated by N_m control points that are uniformly distributed throughout the time scale. Both B-splines are generated within the admissible workspace and by accounting for constraints and boundary conditions. Solutions based on SQP techniques have been proposed to generate collision-free smooth paths for unmanned aerial vehicles (see, for example, [47–49]), autonomous ground vehicles (see, for example, [50–53]) and electric powered wheelchairs (see, for example, [54–58]). This is, however, to the best of our knowledge, the first time that this sort of trajectory generator has been developed

for systems of a reconfigurable nature (discontinuous states) and applied to surpass non-continuous architectural barriers (stairs).

The VI "Constrained Nonlinear Optimization" (see Figure 18b), one of the LabVIEW optimization routines is, meanwhile, used to solve the optimization problem (cost function and constraints). This VI solves a general nonlinear optimization problem with nonlinear equality and nonlinear inequality constraint bounds, also using an SQP method. This function implements the SQP algorithm in a generic form. The function expects a continuous-time formulation of the objective function and constraints, limits on the decision variables and initial estimation values, as regards both the decision variables and the state of the algorithm. Besides the SQP algorithm, confronting the formulation of the nonlinear optimization problem involves solving the set of equations in the model, and evaluating the solution. This numerical approach is widely used when direct analytical estimates cannot be determined [59–61].
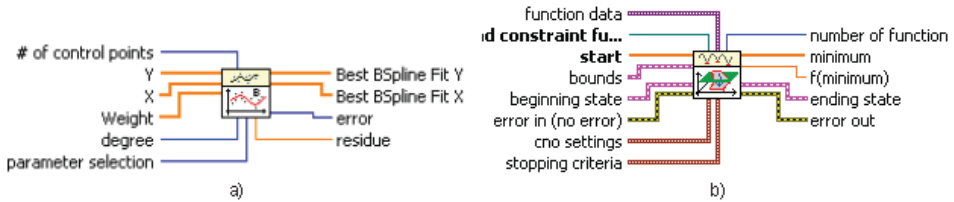
Figure 18. The two main Virtual Instruments (VIs) used to solve the optimization problem; (**a**) the B-Spline Fit VI and (**b**) the Constrained Nonlinear Optimization VI.

Finally, visual sequences (and Videos S1 and S2 referenced in the Supplementary Materials) that exhibit the climbing and descending process of a three-step staircase are illustrated in Figures 19 and 20. We would like to mention that videos associated with descending stairs are not usually shown because in these types of tests the non-commercial experimental prototypes are susceptible to having stability problems and therefore a greater risk of falls. In the case of our experimental prototype, this problem is completely solved because it always transits between mechanically stable configurations. Our prototype is characterized by the fact that it maintain a stable equilibrium on the stair when it changes from one configuration to another.
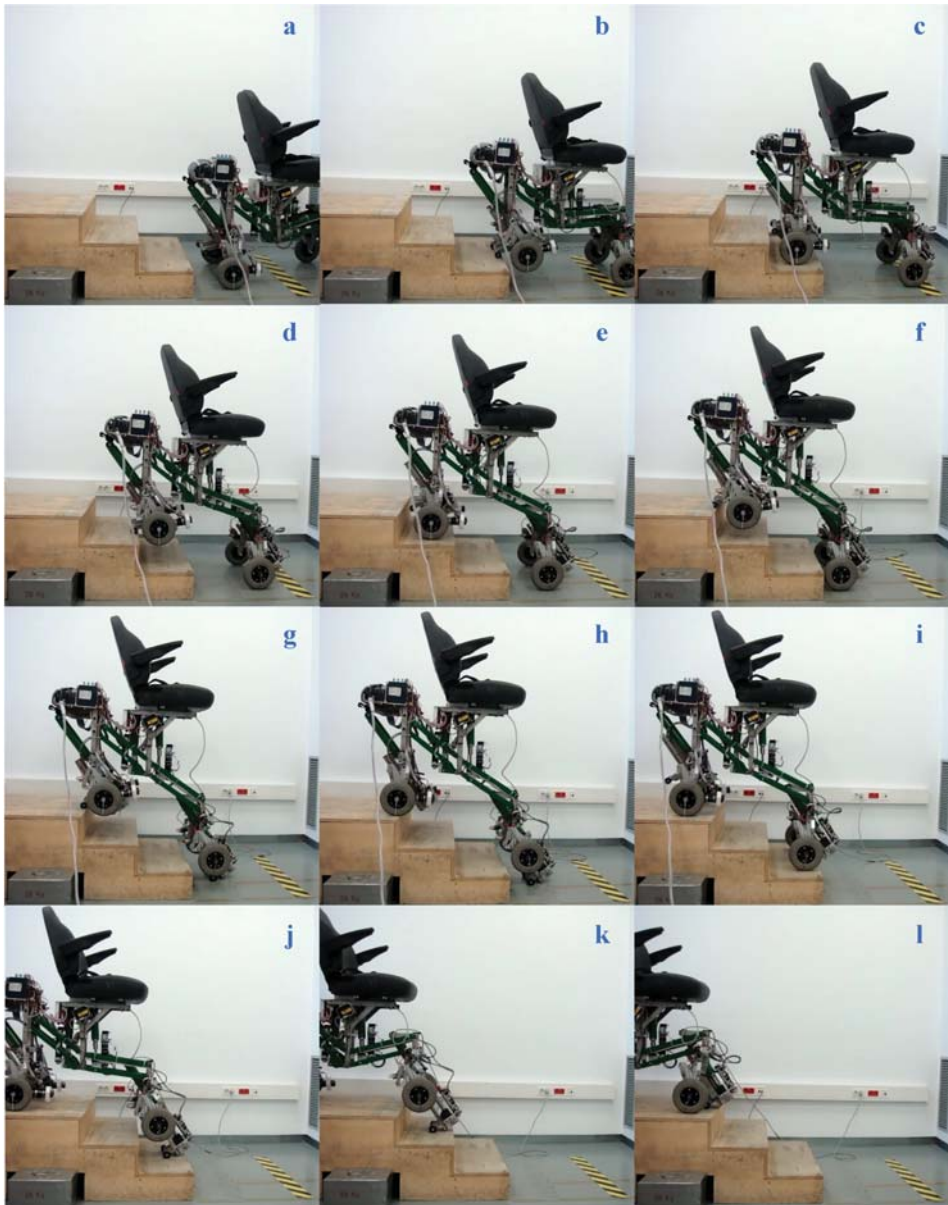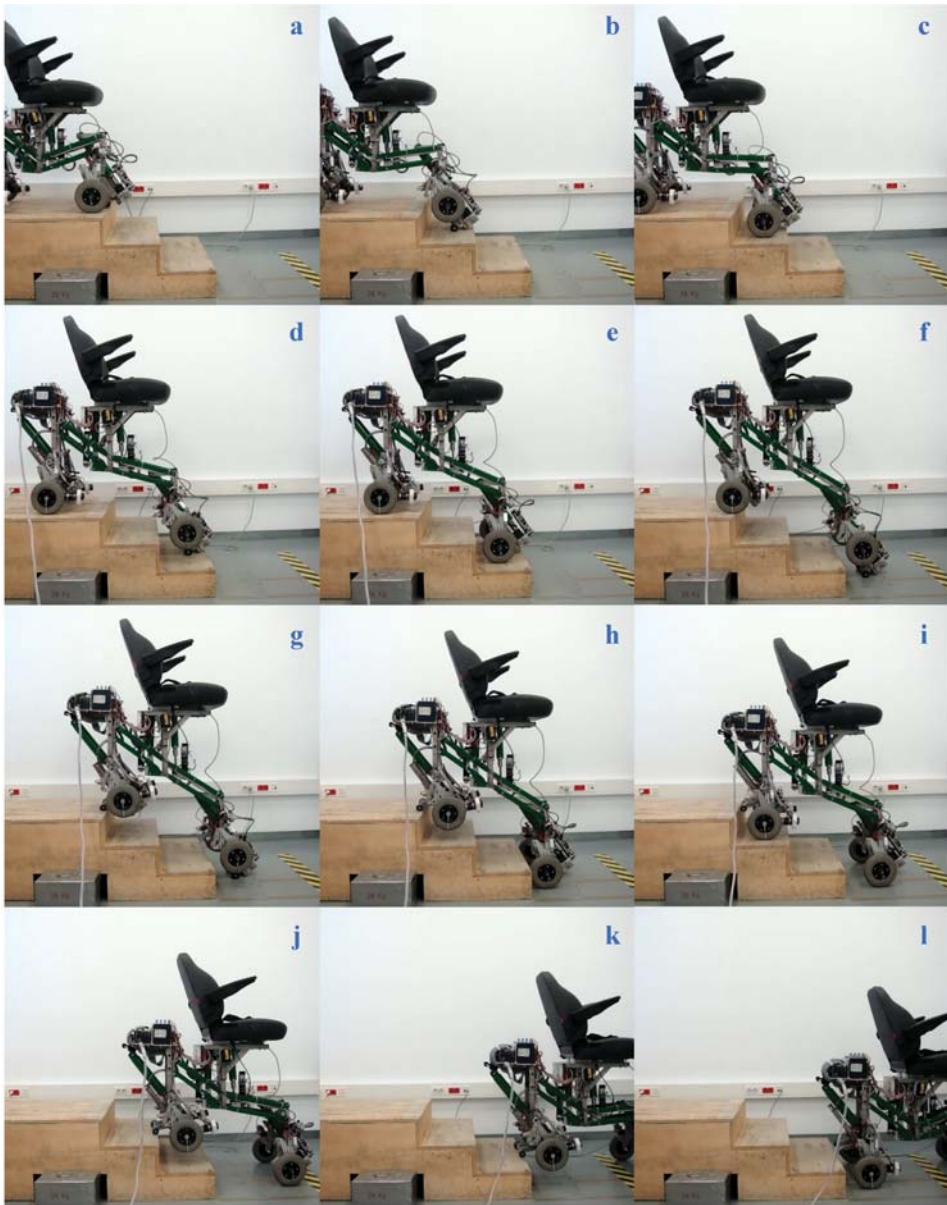
Figure 19. Sequence of climbing process when it surpasses a three-step staircase. All possible configurations of the SCMS are included; Configuration 1 in (**a**,**c**,**e**,**i**,**l**); Configuration 2 in (**b**,**d**,**f**); Configuration 3 in (**h**,**j**,**k**); and Configuration 4 in (**g**). See the video referenced in the Supplementary Materials.

Figure 20. Sequence of descent process when it surpasses a three-step staircase. All possible configurations of the SCMS are included; Configuration 1 in (**a**,**c**,**e**,**i**,**l**); Configuration 2 in (**h**,**j**,**k**); Configuration 3 in (**b**,**d**,**f**); and Configuration 4 in (**g**). See the video referenced in the Supplementary Materials.

5.2. Results

The results of the experiments are presented in two parts; the first shows the results of the first experiment (Stair-Climbing Mode 1), while the second depicts the results of the second experiment (Stair-Climbing Mode 2).

5.2.1. First Experiment

Figures 21–23 illustrate the experimental measurements of the inclination angle (γ) during the first stair-climbing process (Stair-Climbing Mode 1). These figures show a comparison between the inclination obtained when the desired trajectory profiles ($\mathbf{P}^*$) that the SCMS will track are and are not optimized (in the first case, the methodology proposed in this paper is applied to optimize $\mathbf{P}^*$, while in the second case, it is not applied and the profiles are not, therefore, optimized). When the optimization is carried out and the α of the cost function (32) is equal to zero, the variation in inclination is less than $0.7°$ (see Figure 21), signifying that an excellent control posture is achieved, which is associated with greater comfort and safety for the user. When the optimization is carried out with α equal to 0.5, the variation in inclination is less than $1.1°$ (see Figure 22), signifying that a good control posture is also achieved. When the optimization is carried out with α equal to 1, the variation in inclination is increased (less than $1.4°$). However, note that the time spent on the climbing process is notably minimized (see Figure 23).
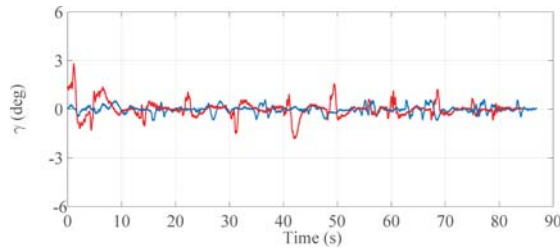
Figure 21. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to zero (—) and when the optimization is not considered (—).
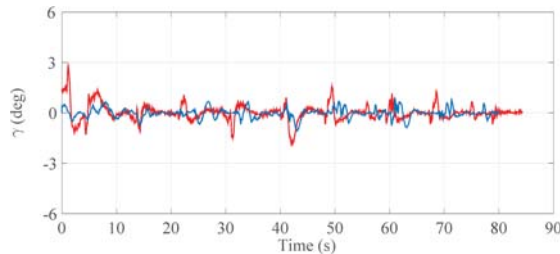
Figure 22. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to 0.5 (—) and when the optimization is not considered (—).
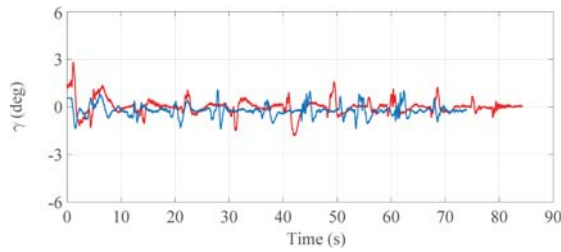
Figure 23. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to one (—) and when the optimization is not considered (—).

Figures 24 and 25 show the comparisons of the angular position of the driving wheels (θ_3) and the angular position of the joint connecting the SCMS frame (θ_1), respectively. The comparisons are also made when the desired trajectory profiles ($\mathbf{P}^*$) are and are not optimized. In this climbing mode, the responsibility for the maintenance of the vertical position of the SCMS frame is entirely supported by the linear actuator related to θ_1. This climbing strategy allows the angular position (θ_2) to be kept constant during the stair climbing task ($\theta_2 = 1.1$ rad).
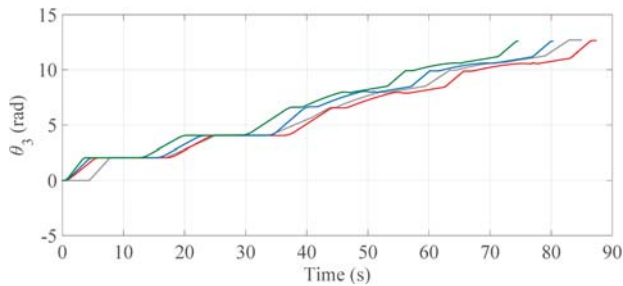
Figure 24. Behavior of the angular position of the driving wheels (θ_3) when the optimization is carried out with $\alpha = 0$ (—), $\alpha = 0.5$ (—) and $\alpha = 1$ (—) and when the optimization is not considered (—).
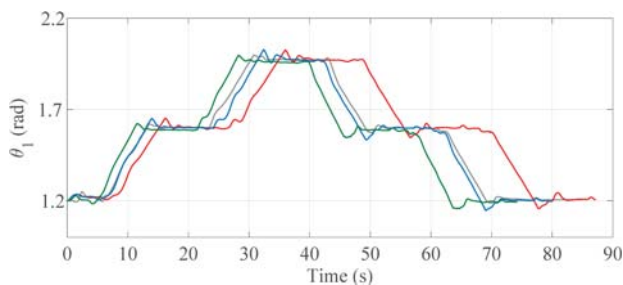
Figure 25. Behavior of θ_1 when the optimization is carried out with $\alpha = 0$ (—), $\alpha = 0.5$ (—) and $\alpha = 1$ (—) and when the optimization is not considered (—).

Figure 26 shows the position of $\mathbf{P}_g$ when the optimization is carried out with α equal to zero and when it is not considered. Note that in the first case there is only a small deviation in the trajectory, which is barely perceived by the passenger. The results obtained from these stair-climbing experiments are shown in Table 6.
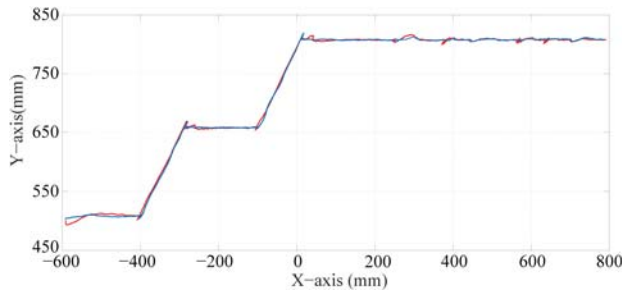
Figure 26. Path of the center of mass (**P**$_g$) when the optimization is carried out with α equal to zero (—) and when the optimization is not considered (—).

Table 6. Results obtained.

α	Climbing Time (s)	Comfort (Std)	Algorithm Execution Time (s)
0	86	0.0019	2.6
0.5	80	0.0032	3.1
1	74	0.0047	4.0
No optimization	85	0.0053	

5.2.2. Second Experiment

Figures 27–29 illustrate the experimental measurements of the inclination angle (γ) during the second stair-climbing process (Stair-Climbing Mode 2). These figures also show a comparison between the inclination obtained when the desired trajectory profiles (**P***) that the SCMS will track are and are not optimized. In this stair-climbing mode, the optimization is carried out with α equal to 0, 0.5 and 1. In the first case, the variation in inclination is relatively small (less than 0.82°), signifying that an excellent control posture is also achieved (see Figure 27). In the second case, the variation in inclination is less than 1.2° (see Figure 28). When α is equal to 1one, the variation in inclination is increased (less than 1.4°). However, note that the time spent on the climbing process is also notably minimized (see Figure 29).
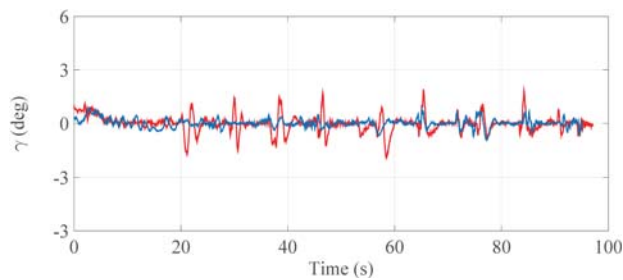

Figure 27. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to zero (—) and when the optimization is not considered (—).
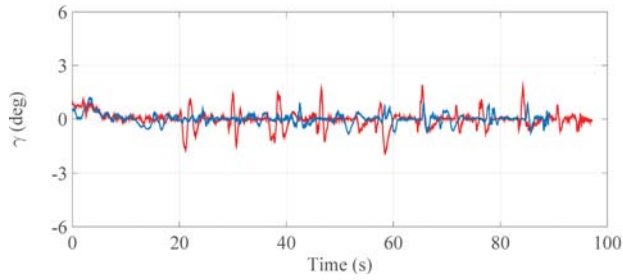
Figure 28. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to 0.5 (—) and when the optimization is not considered (—).
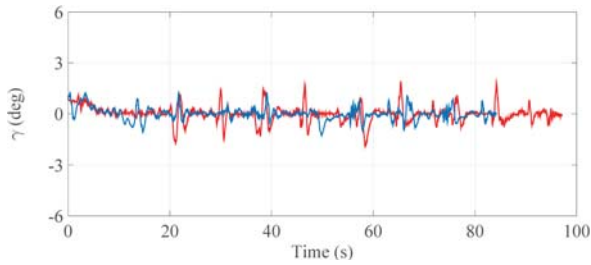

Figure 29. Inclination of the SCMS frame (γ) when the optimization is carried out with a α equal to one (—) and when the optimization is not considered (—).

The comparisons of the angular position of θ_3, θ_1 and θ_2 are shown in Figures 30–32, respectively. Unlike the first climbing strategy, in this strategy the responsibility for the maintenance of the vertical position of the SCMS frame is supported by both the linear actuators related to these angles (θ_1 and θ_2).
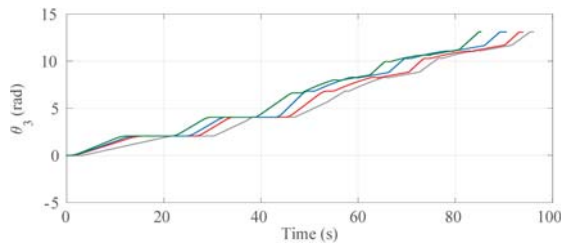
Figure 30. Behavior of the angular position of the driving wheels (θ_3) when the path is optimized with $\alpha = 0$ (—), $\alpha = 0.5$ (—) and $\alpha = 1$ (—) and when the optimization is not considered (—).
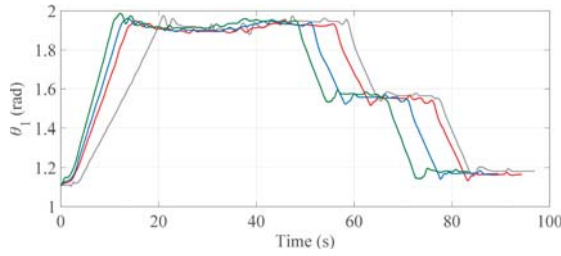
Figure 31. Behavior of the angular position (θ_1) when the path is optimized with $\alpha = 0$ (—), $\alpha = 0.5$ (—) and $\alpha = 1$ (—) and when the optimization is not considered (—).
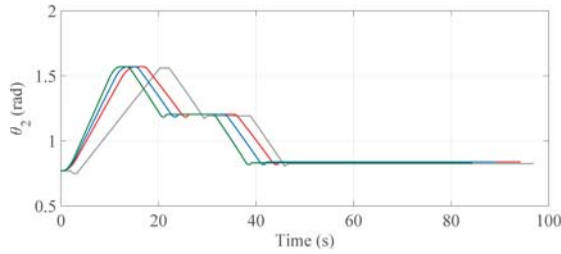

Figure 32. Behavior of the angular position (θ_2) when the path is optimized with $\alpha = 0$ (—), $\alpha = 0.5$ (—) and $\alpha = 1$ (—) and when the optimization is not considered (—).

Figure 33 shows the position of $\mathbf{P}_g$ when the optimization is carried out with α equal to zero and when it is not considered. As before, observe that when the optimization is considered there is a small deviation in the trajectory (note the regularity of the position of the center of mass during the maneuver). The results obtained after carrying out these stair-climbing experiments are shown in Table 7.
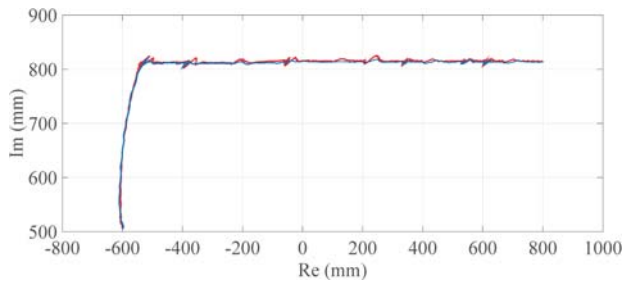

Figure 33. Path of the center of mass ($\mathbf{P}_g$) when the optimization is carried out with α equal to zero (—) and when the optimization is not considered (—).

Table 7. Results obtained.

α	Climbing Time (s)	Comfort (Std)	Algorithm Execution Time (s)
0	95.2	0.0021	1.1
0.5	90.0	0.0033	1.9
1	84.3	0.0042	3.6
No optimization	97.3	0.0062	

6. Conclusions

In this paper, we have developed an advanced trajectory generator with which to surpass those architectural barriers associated with climbing stairs. We have proposed a trajectory optimization technique that is used to improve the operation performance of the SCMS during the stair-climbing process. Two optimal reference trajectories (two different climbing strategies) for the platform carrying a user have been generated before confronting the architectural barriers. They have been generated by considering the information about stair geometry. This has allowed the control problem to be simplified, the time needed to climb/descend staircases to be reduced and the user's safety and comfort to be improved.

The proposed technique makes it possible to minimize a cost function when subjected to the most important constraints inherent to the system behavior, such as the geometry of the architectural barrier, its reconfigurable nature, comfort and limits as regards actuator torques and speeds. Two sets of experiments were carried out using different time-comfort combinations and different climbing strategies in order to evaluate the proposed methodology when the autonomous SCMS surpasses staircases. The results show that this technique can be successfully used to generate optimal trajectories under various constraints inherent in the whole system. When the α of the cost function is equal to zero, the trajectory tracking is successfully achieved and a comfortable motion is generated. When the α is equal to one, the trajectory tracking is fairly good and the time spent on the climbing process decreases. The work presented in this paper once again demonstrates the feasibility and advantages of the proposed concept. When compared with other SCMSs, our prototype is characterized by the fact that it can climb up and down stairs of different sizes, including those with geometric disturbances, and maintain a stable equilibrium on the stair at all times without assistance.

Author Contributions: José Abel Chocoteco, Rafael Morales and Vicente Feliu-Batlle conducted the trajectory generator design, and they wrote the paper and performed all the experiments. All authors were involved in modifying the paper, the literature review and the discussion of the results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wellman, P.; Krovi, V.; Kumar, V.; Harwin, W. Design of a wheelchair with legs for people with motor disabilities. *IEEE Trans. Rehabil. Eng.* **1995**, *3*, 343–353, doi:10.1109/86.481974.
2. Hashimoto, K.; Hosobata, T.; Sugahara, Y.; Mikuriya, Y.; Sunazuka, H.; Kawase, M.; Lim, H.O.; Takanishi, A. Realization by biped leg-wheeled robot of biped walking and wheel-driven locomotion. *ICRA* **2005**, 2970–2975, doi:10.1109/ROBOT.2005.1570565.
3. Chen, C.-H.; Pham, C.H. Design and fabrication of a statically stable stair-climbing robotic wheelchair. *Ind. Robot Int. J.* **2009**, *36*, 562–569, SSN 0143-991X.
4. Yuan, J.; Paisley, R.; Song, Y.; Zhang, W. Virtual realization of automatic stair-climbing motion by leg-wheeled hybrid mobile robot. *IEEE ROBIO* **2010**, 1352–1357, doi:10.1109/ROBIO.2010.5723526.
5. Lee, C.H.; Lee, K.M.; Yoo, J.; Kim, I.S.; Bang, Y.B. A compact stair-climbing wheelchair with two 3-DOF legs and a 1-DOF base. *Ind. Robot Int. J.* **2016**, *43*, 181–192.
6. Tomo Co. Ltd.; Tamagawa University. Freedom. Available online: www.tomo-co.co.jp/free.htm (accessed on 19 January 2017).
7. Sugahara, Y.; Yonezawa, N.; Kosuge, K. A novel stair-climbing wheelchair with transformable wheeled four-bar linkages. *IEEE/RSJ IROS* **2010**, 3333–3339, doi:10.1109/IROS.2010.5648906.
8. Independence Technology, L.L.C. The ibot Website. Available online: urlm.co/www.ibotnow.com (accessed on 17 September 2017).

9.   The Company TopChair. Available online: www.topchair.fr/ (accessed on 17 September 2017).
10. Galileo Mobility Instruments Ltd. The Galileo Mobility Website. Available online: www.galileomobility.com (accessed on 17 September 2017).
11. Quaglia, G.; Nisi, M.; Franco, W.; Bruzzone, L. Dynamic simulation of an electric stair-climbing wheelchair. *IJAT* **2017**, *11*, 472–480, doi:10.20965/ijat.2017.p0472.
12. Creative DNA Austria. Available online: www.creativednaaustria.com/ (accessed on 17 September 2017).
13. Leaman, J.; La, H.M. A comprehensive review of smart wheelchairs: Past, present and future. *arXiv* **2017**, arXiv:1704.04697.
14. Sundaram, S.A.; Wang, H.; Ding, D.; Cooper, R.A. Step-climbing power wheelchairs: A literature review. *Top. Spinal Cord Inj. Rehabil.* **2017**, *23*, 98–109, doi:10.1310/sci2302-98.
15. Bang, Y.; Lee, C.; Yoo, J.; Lee, K.; Kim, I. Two-legged stair-climbing wheelchair and its stair dimension measurement using distance sensors. In Proceedings of the 2011 11th International Conference on Control, Automation and Systems (ICCAS), Gyeonggi-do, Korea, 26–29 October, 2011; pp. 1788–1791.
16. Yamamoto, T.; Takemori, F.; Itakura, R. Development of stair locomotive wheelchair with adjustable wheelbase. In Proceedings of the 2012 SICE Annual Conference (SICE), Akita, Japan, 20–23 August 2012; pp. 629–633.
17. Solea, R.; Nunes, U. Robotic wheelchair trajectory control considering user comfort. *ICINCO* **2009**, 113–125, doi:10.1007/978-3-642-00271-7_8.
18. Huang, J.J.; Chiang, H.H.; Lee, T.T.; Kou, K.Y. Riding comfort improvement by considering passenger's behavior suppression on powered wheelchairs. *IEEE ICSSE* **2012**, 131–136, doi:10.1109/ICSSE.2012.6257163.
19. Chocoteco, J.A.; Morales, R.; Feliu, V. Improving the climbing/descent performance of stair-climbing systems confronting architectural barriers with geometric dusturbances. *Mechatronics* **2015**, *30*, 11–26.
20. Morales, R.; Feliu, V.; González, A. Optimized obstacle avoidance trajectory generation for a reconfigurable staircase climbing wheelchair. *Robot. Auton. Syst.* **2010**, 58, 97–114.
21. Chocoteco, J.A.; Morales, R.; Feliu, V.; Sira-Ramírez, H. Robust control for the trajectory tracking of robotic wheelchairs. *Robotica* **2015**, *33*, 41–59.
22. Morales, R.; Sira-Ramírez, H.; Somolinos, J.A. Robust control of underactuated wheeled mobile manipulators using GPI disturbance observers. *Multibody Syst. Dyn.* **2015**,*32*, 511–533.
23. Morales, R.; Feliu, V.; González, A.; Pintado, P. Coordinated motion of a new staircase climbing wheelchair with increased passenger comfort. In Proceedings of the 2006 IEEE Internacional Conference on Robotics and Automation (ICRA 2006), Orlando, FL, USA, 15–19 May 2006; pp. 3995–4001.
24. Leishman, F.; Monfort, V.; Horn, O.; Bourhis, G. Driving assistance by deictic control for a smart wheelchair: The assessment issue. *IEEE Trans. Hum. Mach. Syst.* **2014**, *44*, 66–77, doi:10.1109/TSMC.2013.2287792.
25. Morales, Y.; Abdur-Rahim, J.A.; Even, J.; Watanabe, A.; Kondo, T.; Hagita, N.; Ishii, S. Modeling of human velocity habituation for a robotic wheelchair. *IEEE/RSJ IROS* **2014**, 3284–3290, doi:10.1109/IROS.2014.6943019.
26. Wu, B.F.; Chen, P.Y.; Lin, C.H. A new criterion of human comfort assessment for wheelchair robots by Q-Learning based accompanist tracking fuzzy controller. *Int. J. Fuzzy Syst.* **2016**, *18*, 1039–1053, doi:10.1007/s40815-015-0120-6.
27. Wolf, E.; Cooper, R.A.; Pearlman, J.; Fitzgerald, S.G.; Kelleher, A. Longitudinal assessment of vibrations during manual and power wheelchair driving over select sidewalk surfaces. *J. Rehabil. Res. Dev.* **2007**, *44*, 573–580, doi:10.1682/JRRD.2006.05.0049.
28. Garcia-Mendez, Y.; Pearlman, J.L.; Boninger, M.L.; Cooper, R.A. Health risks of vibration exposure to wheelchair users in the community. *J. Spinal Cord Med.* **2013**, *36*, 365–375, doi:10.1179/2045772313Y.0000000124.
29. Chénier, F.; Aissaoui, R. Effect of wheelchair frame material on users' mechanical work and transmitted vibration. *Biomed. Res. Int.* **2014**, *2014*, 1–13, doi:10.1155/2014/609369.
30. Hirata, K.; Murakami, T. Assisted control of step passage motion and its comfort evaluation in two-wheel wheelchair systems. *EEJ* **2015**, *192*, 50–58, doi:10.1002/eej.22680.
31. Hobson, D.A.; Crane, B. State of the science white paper on: Wheelchair seating comfort. In Proceedings of the Conference on Seating Issues for Persons with Disabilities, Orlando, FL, USA, 19–20 February 2001; pp. 29–33.

32. Geyer, M.J.; Brienza, D.M.; Bertocci, G.E.; Crane, B.; Hobson, D.; Karg, P.; Schmeler, M.; Trefler, E. Wheelchair seating: A state of the science report. *Assist. Technol.* **2003**, *15*, 120–128, doi:10.1080/10400435.2003.10131896.

33. Shino, M.; Tomokuni, N.; Murata, G.; Segawa, M. Wheeled inverted pendulum type robotic wheelchair with integrated control of seat slider and rotary link between wheels for climbing stairs. *IEEE ARSO* **2014**, 121–126, doi:10.1109/ARSO.2014.7020991.

34. Dobrzyński, G.; Choromański, W.; Grabarek, I. Analysis of ride comfort on the stairs climbing wheelchair. In *Advances in Human Aspects of Transportation: Part II*; Springer: Berlin, Germany, 2014; Volume 8, pp. 114–125.

35. Cox, M.G. The numerical evaluation of B-splines. *IMA J. Appl. Math.* **1972**, *10*, 134–149, doi:10.1093/imamat/10.2.134.

36. De Boor, C. *A Practical Guide to Splines*; Springer: New York, NY, USA, 1978; p. 325.

37. Balluff Company. Photoelectric Distance Sensor, BOD 63M-LA04-S115. Available online: http://www.balluff.com/balluff/MUS/en/products/product_detail.jsp#/152994 (accessed on 22 October 2017).

38. Chocoteco, J.; Morales, R.; Feliu, V.; Sánchez, L. Trajectory planning for a stair-climbing mobility system using laser distance sensors. *IEEE Syst. J.* **2016**, *10*, 944–956, doi:10.1109/JSYST.2014.2309477.

39. Morales, R.; Gonzalez, A.; Feliu, V.; Pintado, P. Environment adaptation of a new staircase-climbing wheelchair. *Auton. Robots* **2007**, *23*, 275–292, doi:10.1007/s10514-007-9047-6.

40. Morales, R.; Gonzalez, A.; Feliu, V.; Pintado, P. Kinematic model of a new staircase climbing wheelchair and its experimental validation. *Int. J. Robot. Res.* **2006**, *25*, 825–841.

41. Gonzalez, A.; Morales, R.; Feliu-Batlle, V.; Pintado, P. Improving the mechanical design of new staircase wheelchair. *Ind. Robot Int. J.* **2007**, *34*, 110–115.

42. Morales, R.; Somolinos, J.A.; Cerrada, J.A. Dynamic model of a reconfigurable stair-climbing mobility system and its experimental validation. *Multibody Syst. Dyn.* **2012**, *28*, 349–367.

43. Morales, R.; Somolinos, J.A.; Cerrada, J.A. Dynamic control of a reconfigurable stair-climbing mobility system. *Robotica* **2013**, *31*, 295–310.

44. Morales, R.; Chocoteco, J.; Feliu, V.; Sira-Ramírez, H. Obstacle surpassing and posture control of a stair-climbing robotic mechanism. *Control Eng. Pract.* **2013**, *21*, 604–621, doi:10.1016/j.conengprac.2013.01.006.

45. National Instruments Corp.: LabVIEW 2016 Help. B-Spline Fit VI. Available online: zone.ni.com/reference/en-XX/help/371361N-01/gmath/bspline_fit/ (accessed on 17 September 2017).

46. National Instruments Corp.: LabVIEW 2016 Help. Constrained Nonlinear Optimization VI. Available online: zone.ni.com/reference/en-XX/help/371361L-01/gmath/constrained_nonlinear_optimization/ (accessed on 17 September 2017).

47. Foo, J.L.; Knutzon, J.; Kalivarapu, V.; Oliver, J.; Winer, E. Path planning of unmanned aerial vehicles using B-splines and particle swarm optimization. *J. Aerosp. Comput. Inf. Commun.* **2009**, *6*, 271–290, doi:10.2514/1.36917.

48. Besada-Portas, E.; De la Torre, L.; Jesus, M.; De Andrés-Toro, B. Evolutionary trajectory planner for multiple UAVs in realistic scenarios. *IEEE Trans. Robot.* **2010**, *10*, 619–634, doi:10.1109/TRO.2010.2048610.

49. Zhao, S.; Wang, X.; Zhang, D.; Shen, L. Model predictive control based integral line-of-sight curved path following for unmanned aerial vehicle. In Proceedings of the 3rd AIAA Guidance, Navigation, and Control Conference, Grapevine, TX, USA, 9–13 January 2017; p. 1511, doi:10.2514/6.2017-1511.

50. Connors, J.; Elkaim, G. Manipulating B-Spline based paths for obstacle avoidance in autonomous ground vehicles. *ION NTM* **2007**, *5*, 1081–1088.

51. Berglund, T.; Brodnik, A.; Jonsson, H.; Staffanson, M.; Soderkvist, I. Planning smooth and obstacle-avoiding B-spline paths for autonomous mining vehicles. *IEEE Trans. Autom. Sci. Eng.* **2010**, *7*, 167–172, doi:10.1109/TASE.2009.2015886.

52. Gu, T.; Snider, J.; Dolan, J.M.; Lee, J. Focused trajectory planning for autonomous on-road driving. *IEEE Intell. Veh. Symp.* **2013**, 547–552, doi:10.1109/IVS.2013.6629524.

53. Elbanhawi, M.; Simic, M.; Jazar, R.N. Continuous path smoothing for car-like robots using B-spline curves. *J. Intell. Robot. Syst.* **2015**, *80*, 23–56, doi:10.1007/s10846-014-0172-0.

54. Zeng, Q.; Burdet, E.; Rebsamen, B.; Teo, C.L. Evaluation of the collaborative wheelchair assistant system. *IEEE ICORR* **2007**, 601–608, doi:10.1109/ICORR.2007.4428486.

55. Rebsamen, B.; Burdet, E.; Guan, C.; Zhang, H.; Teo, C.L.; Zeng, Q.; Laugier, C.; Ang, M.H., Jr. Controlling a wheelchair indoors using thought. *IEEE Intell. Syst.* **2007**, *22*, 18–24, doi:10.1109/MIS.2007.26.

56. Zhou, L.; Teo, C.L.; Burdet, E. A nonlinear elastic path controller for a robotic wheelchair. *IEEE ICIEA* **2008**, 142–147, doi:10.1109/ICIEA.2008.4582495.

57. Carlson, T.; Demiris, Y. Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload. *IEEE Trans. Syst. Man Cybern. B* **2012**, *42*, 876–888, doi:10.1109/TSMCB.2011.2181833.

58. Ammouri, A.H.; Hamade, R.F. On the selection of constitutive equation for modeling the friction stir processes of twin roll cast wrought AZ31B. *Mater. Des.* **2014**, *57*, 673–688, doi:10.1016/j.matdes.2014.01.057.

59. Yoon, S.H.; Park, C.S.; Augenbroe, G.; Kim, D.W. Self-calibration and optimal control of a double-skin system. In Proceedings of the 11th IBPSA Conference (International Building Performance Simulation Association), Glasgow, UK, 27–30 July 2009; pp. 80–87.

60. Kim, J.H.; Woo, L.D.; Cho, K.R.; Jo, S.Y.; Kim, J.H.; Min, C.; Han, D.; Cho, S. Development of an electro-optical system for small UAV. *Aerosp. Sci. Technol.* **2010**, *14*, 505–511, doi:10.1016/j.ast.2010.03.006.

61. Chica, J.A.V.; Torres, A.G.D.; Acosta, D.A. NMPC controller applied to the operation of an internal combustion engine: Formulation and solution of the optimization problem in real time. *IJIDeM* **2016**, 1–22, doi:10.1007/s12008-016-0343-2.