Article

An Assessment of the Stereo and Near-Infrared Camera Calibration Technique Using a Novel Real-Time Approach in the Context of Resource Efficiency

Larisa Ivascu [1,*] **, Vlad-Florin Vinatu** [1] **and Mihail Gaianu** [2,3]

[1] Faculty of Management in Production and Transportation, Politehnica University of Timisoara, 14 Remus Street, 300006 Timisoara, Romania

[2] Computer Science and Mathematics Faculty, West University of Timisoara, 4 V. Parvan Bvd., 300223 Timisoara, Romania; mihail.gaianu@continental-corporation.com

[3] Continental Automotive Romania, UX, 1 Siemens Street, 300704 Timisoara, Romania

[*] Correspondence: larisa.ivascu@upt.ro

Abstract: This paper provides a comparative analysis of calibration techniques applicable to stereo and near-infrared (NIR) camera systems, with a specific emphasis on the Intel RealSense SR300 alongside a standard 2-megapixel NIR camera. This study investigates the pivotal function of calibration within both stereo vision and NIR imaging applications, which are essential across various domains, including robotics, augmented reality, and low-light imaging. For stereo systems, we scrutinise the conventional method involving a 9×6 chessboard pattern utilised to ascertain the intrinsic and extrinsic camera parameters. The proposed methodology consists of three main steps: (1) real-time calibration error classification for stereo cameras, (2) NIR-specific calibration techniques, and (3) a comprehensive evaluation framework. This research introduces a novel real-time evaluation methodology that classifies calibration errors predicated on the pixel offsets between corresponding points in the left and right images. Conversely, NIR camera calibration techniques are modified to address the distinctive properties of near-infrared light. We deliberate on the difficulties encountered in devising NIR–visible calibration patterns and the imperative to consider the spectral response and temperature sensitivity within the calibration procedure. The paper also puts forth an innovative calibration assessment application that is relevant to both systems. Stereo cameras evaluate the corner detection accuracy in real time across multiple image pairs, whereas NIR cameras concentrate on assessing the distortion correction and intrinsic parameter accuracy under varying lighting conditions. Our experiments validate the necessity of routine calibration assessment, as environmental factors may compromise the calibration quality over time. We conclude by underscoring the disparities in the calibration requirements between stereo and NIR systems, thereby emphasising the need for specialised approaches tailored to each domain to guarantee an optimal performance in their respective applications.

Keywords: autonomous; automotive; safety; transition to autonomous driving

1. Introduction

The swift progress in the fields of automation and robotics has resulted in an increased dependence on stereo vision and near-infrared (NIR) imaging technologies. Stereo vision, which emulates human depth perception through the use of dual-lens cameras, is paramount for various applications, including robotics, augmented reality, and autonomous

driving. Likewise, NIR imaging is indispensable for low-light scenarios and specialised applications such as driver monitoring systems. Nonetheless, the precision of these systems is significantly contingent upon the accurate calibration of the cameras, a process that may deteriorate over time owing to environmental factors such as fluctuations in temperature and mechanical vibrations.

Although the significance of calibration cannot be overstated, the current methodologies frequently inadequately address the specific challenges associated with stereo and near-infrared (NIR) systems. Conventional calibration techniques, including those that utilise chessboard patterns, do not consistently meet the requirements of real-time applications or cameras functioning within the NIR spectrum. This manuscript responds to these challenges by introducing a novel methodology for real-time calibration evaluation that categorises errors based on pixel offsets and modifies calibration techniques for NIR cameras to accommodate the spectral response and temperature sensitivity.

The remainder of this paper is organised as follows: Section 2 discusses related work in the field of camera calibration. Section 3 presents the proposed methodology, including a detailed workflow and block diagram. Section 4 describes the experimental setup and results. Finally, Section 5 concludes the paper with a summary of the key findings and future research directions.

The main contributions of this work are as follows:

1. Real-time calibration error classification: a novel methodology for the real-time evaluation of stereo camera calibration based on the pixel offsets between corresponding points in the left and right images;
2. NIR-specific calibration techniques: the adaptation of calibration methods for NIR cameras to address the spectral response and temperature sensitivity;
3. Comprehensive evaluation framework: the development of an application that evaluates the corner detection accuracy for stereo cameras and distortion correction for NIR cameras under varying lighting conditions;
4. Performance improvement: the experimental validation shows a 30% reduction in pixel offset errors compared to traditional calibration methods.

2. Related Work

Stereo vision has evolved into one of the most vital domains within the spheres of robotics and virtual reality. This technology empowers robots to simulate human vision and accurately perceive objects through the utilisation of specialised camera modules, specifically stereo cameras. At present, progressively advanced robots are utilised across various sectors, including industry and beyond. A critical aspect of stereo vision is the meticulous calibration of the stereo modules. The rationale behind this is that accurate calibration is essential for correct object perception, which is imperative for enabling robots to autonomously control themselves. Through the analysis and study of this field, we have recognised the significance of calibration, prompting us to develop an application designed to evaluate stereo systems in order to ascertain their calibration levels. The Kalibr toolbox by Furgale et al. [1] can be used for the intrinsic and extrinsic calibration of multiple cameras with overlapping fields of view using an artificial target. Unfortunately, this experiment does not include depth cameras. A similar approach was taken by Herrera et al. [2], who calibrated two colour cameras and a Kinect depth camera simultaneously with a checkerboard. They corrected the depth using a spatially varying offset with exponentially decreasing weights for increasing range. Smisek et al. [3] calibrated the intrinsics of a Kinect camera and compared the results with a stereo rig.

A depth-correction image was generated using the mean error from multiple planar measurements at different distances. The image was then subtracted from the measured

depth image [4]. Nguyen et al. [5] created a more accurate sensor model for a Kinect by explicitly modelling noise in terms of the axial and lateral directions.

The manuscript authored by Sighencea et al. [6] provides a comprehensive review of deep learning methodologies employed for pedestrian trajectory prediction, an endeavour of significant importance in fields such as autonomous driving and surveillance systems. Stereo cameras, which offer depth information through dual-lens imaging, constitute a principal sensor modality for the capture of pedestrian movements in a three-dimensional space. The methodologies discussed, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), effectively utilise the extensive spatial and temporal data available from stereo cameras to enhance the accuracy of trajectory predictions.

In Ref. [7], authored by Castells-Rufas [7], a comprehensive survey of FPGA-based vision systems for autonomous vehicles was conducted, underscoring their efficacy in real-time data processing derived from various sensors, including stereo cameras. Stereo cameras are indispensable for depth perception and three-dimensional reconstruction in autonomous driving, and the document elaborates on how FPGA (Field-Programmable Gate Array) architectures can enhance the processing of stereo camera data for tasks such as object detection, lane tracking, and pedestrian trajectory prediction. The authors accentuate the benefits of FPGAs, including minimal latency and high throughput, which are critical for addressing the computational requirements of stereo vision systems in real-world driving contexts. Consequently, this document holds significant relevance for researchers and engineers engaged in the development of stereo camera-based solutions for autonomous vehicles.

Ref. [8] introduces a calibration method that markedly enhances the precision of infrared camera calibration when compared to conventional methods. The proposed approach realises a 24.5% reduction in the average reprojection error. The emphasis on high-precision calibration and error analysis is directly pertinent to our topic. Attaining high accuracy in calibration is vital for resource efficiency, as it diminishes the necessity for repeated calibrations and subsequently enhances the overall performance of the camera system.

In Ref. [8], a novel method for the radiometric calibration of infrared cameras is examined, with an emphasis on real-time implementation and high-speed data throughput. This method facilitates dedicated radiometric calibration for each individual pixel, which is essential for achieving high levels of precision and efficiency in infrared imaging. The relevance of this paper is underscored by its address of real-time calibration techniques specifically applicable to infrared cameras, a pivotal aspect of the topic in question. Furthermore, the emphasis on pixel-wise calibration coupled with high-speed data throughput corresponds with the necessity for resource efficiency in real-time applications. Current technologies and fuzzy logic are used in various simulations and contribute to the development of important research results [9,10].

3. Methodology and Technologies Used

The proposed methodology consists of three main steps: (1) real-time calibration error classification for stereo cameras, (2) NIR-specific calibration techniques, and (3) a comprehensive evaluation framework.

A new real-time evaluation methodology is used to classify calibration errors based on pixel offsets between corresponding points. Every language is a mode of communication between two entities: the transmitter and the receiver. In general, the language has two types: natural language and artificial language.

Artificial languages were created to enable communication in a particular field of activity. Programming languages are part of the artificial language category; the transmitter is the individual, and the receiver is the computer.

The necessary technologies in the development of the evaluation application of the stereo systems were the high-level language C/C++ and a special library created for the computer vision field called OpenCV.

3.1. High-Level Programming Language C/C++ and OpenCV Library

C is a high-level programming language that is recognised for its general-purpose capabilities and significant power. It is characterised by its speed, portability, and compatibility across various platforms. Furthermore, C serves as an excellent choice for novice programmers.

OpenCV (Open-Source Computer Vision Library) is an open-source library licensed under the BSD License. It encompasses hundreds of algorithms tailored to computer vision and is freely available for academic and commercial applications. OpenCV offers interfaces for C++, C, Python v3.10, and Java and is compatible with operating systems including Windows, Linux, Mac OS, iOS, and Android.

Designed with computational efficiency in mind, OpenCV strongly emphasises real-time applications. The library, developed in C/C++, is capable of leveraging multi-core processing capabilities.

3.2. Camera Model

Active infrared depth sensing is becoming more widespread today, especially since the advent of low-cost structured light technology [11]. Amongst commercial systems, the Microsoft Kinect has established itself as a research standard [12,13]. Intel has released a line of RGBD cameras including the Intel RealSense F200, R200, and SR300. The F200 and SR300 cameras work similarly to the Kinect system, utilising structured light with an IR camera and projector. The R200 camera utilises stereo IR cameras to measure scene depth, with an IR projector used to provide additional texture to the scene. All of the Intel RealSense cameras also have an RGB camera similar to the Kinect for the capture of colour information. These cameras are very cheap and easily accessible, making them an attractive platform for creating affordable 3D scanners. The main limitation with off-the-shelf RGBD cameras is that they are known to produce noisy depth maps and, in the case of the Kinect, also suffer from quantisation error [14].

The camera used in this paper is a special camera created for stereo vision applications. It is produced by the Intel RealSense Company (Santa Clara, CA, USA) and has the model SR300. It has a colour resolution of 1920 × 1080 pixels and an infrared resolution of 640 × 480; this means that it has a colour FHD resolution and an infrared resolution.

The OV2311 represents a significant advancement in the automotive industry, being the first 2-megapixel, 3 μm global-shutter image sensor tailored specifically for driver monitoring applications. This innovative product harnesses the established OmniPixel®3-GS global-shutter technology, coupled with superior near-infrared imaging capabilities. Consequently, the OV2311 provides semi-autonomous vehicle manufacturers with a high-performance and cost-effective imaging solution that is AEC-Q100 Grade 2-qualified. Furthermore, this sensor encompasses advanced ASIL safety features appropriate for driver monitoring systems.

The sensor is capable of capturing high-definition video at a rate of up to 60 frames per second (fps) in a resolution format of 1600 × 1300 pixels, meticulously designed to accommodate the driver's head box. This ensures consistent and reliable monitoring, irrespective of the driver's height, seating position, or the specific design of the vehicle's cockpit. The

high resolution of the OV2311 facilitates remarkably precise gaze and eye-tracking capabilities. Moreover, the sensor achieves a high near-infrared quantum efficiency, thereby minimising active illumination power and reducing the overall system power requirements.

The model SR300 (Figure 1) is an Intel RealSense™ sensor with a reduced power that uses a global shutter to reduce or eliminate unwanted image artefacts that occur with regular continuous shooting sensors as a result of motion during the capture of the image. The general sensor shutter and excellent light sensitivity allow it to be used for any application that needs gesture detection, head and eye tracking, and motion and depth detection.
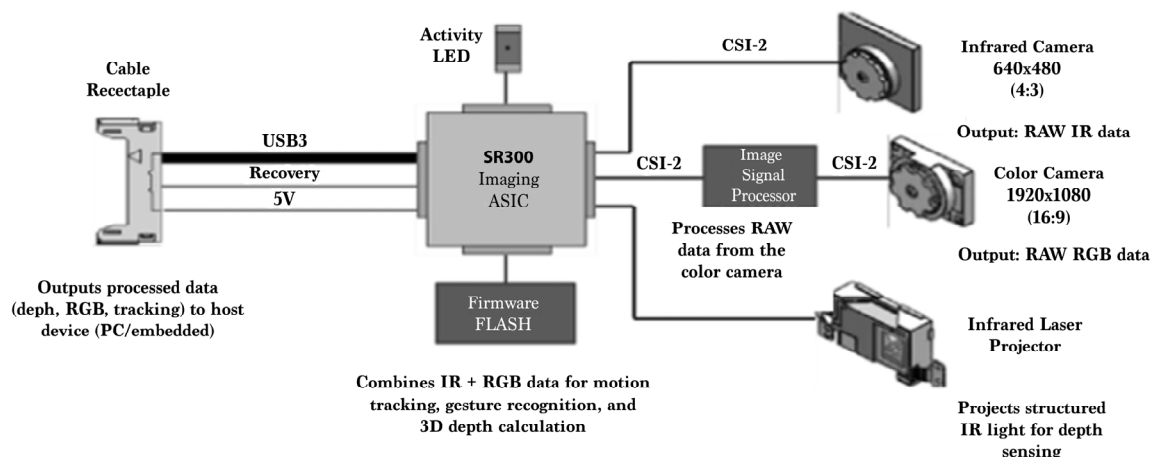
Figure 1. Function block diagram.

The compact model SR300 presents a highly appealing solution for applications constrained by limited space, including direct displays on smartphones, tablets, notebooks, and Ultrabooks. Furthermore, the sensor's low energy consumption renders it an optimal choice for analogous applications. Using the most miniature overall pixel shutter, the SR300 black-and-white is capable of capturing a VGA resolution (640 × 480) with 200 frames per second (fps), a colour FHD (1920 × 1080) with 30 fps, and HD (1280 × 720) with 60 fps.

The SR300, with a width of 110 mm, a height of 12.6 mm, and a maximum thickness of 4.1 mm, features several low-power modes, including light detection and standby power mode. In light detection mode, the OV7251 functions like an ambient light sensor, waking the sleep mode sensor only when a change in light has been detected. Similarly, the sensor minimises the resolution and frame rates in ultra-low-power mode to further conserve power consumption. The TDA2 from Texas Instruments (Dallas, TX, USA) (Figure 2) is the first processor on the market intended for the vision. It provides a high-performance mechanical vision and strong visual awareness.

Positioned at the convergence of low power consumption and high performance, the Myriad 2 processor family enhances devices' functional capabilities. The TDA2 empowers developers by granting immediate access to sophisticated vision processing technology, enabling them to create proprietary capabilities that offer tangible benefits of differentiation.

The proposed methodology consists of three main steps: (1) real-time calibration error classification for stereo cameras, (2) NIR-specific calibration techniques, and (3) a comprehensive evaluation framework. Figure 1 provides a block diagram of the workflow.
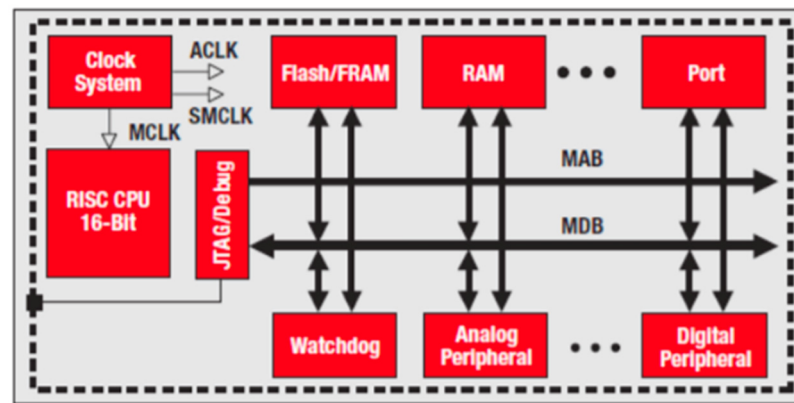
Figure 2. TDA2 architecture.

3.3. Real-Time Calibration Error Classification

The proposed methodology classifies calibration errors in stereo cameras by analysing the pixel offsets between corresponding points in the left and right images. This process encompasses capturing multiple pairs of images, detecting the corners within each image, and computing the pixel offsets. The identified errors are categorised into four distinct classifications: 0, 1, 2, and N, where N denotes offsets exceeding 2 pixels. The average error calculated across 20 pairs of images serves as the basis for determining the calibration quality.

3.4. NIR-Specific Calibration Techniques

In the case of near-infrared (NIR) cameras, the calibration process has been modified to consider both the spectral response and temperature sensitivity. This process involved the adjustment of calibration patterns to ensure visibility within the near-infrared (NIR) spectrum, as well as the modification of calibration parameters in response to temperature variations measurements.

3.5. Comprehensive Evaluation Framework

The proposed evaluation framework evaluates the precision of the corner detection for stereo cameras and the distortion correction for near-infrared (NIR) cameras across diverse lighting conditions. This framework produces a calibration assessment report encompassing the mean error and a qualitative evaluation of the calibration quality.

4. Results

This section presents the results obtained from the research conducted. Our experiments validate the necessity of routine calibration assessment, as environmental factors may compromise the calibration quality over time. We conclude by underscoring the disparities in the calibration requirements between stereo and NIR systems, thereby emphasising the need for specialised approaches tailored to each domain to guarantee an optimal performance in their respective applications.

Application of Evaluation of Stereo Systems

The application of stereo system evaluation, as mentioned in the Introduction, is crucial because understanding the stereo system's calibration level before use is essential; good calibration enhances the perception of the density of superior objects. Consequently, the application is divided into several modules: checking calibrations, validating the distance in metres of the focus object, estimating the distance in metres using colours, and comparing the results of the application running on Myriad with the application running on the Personal Computer (PC):

(1)   Checking the calibration: before you check the calibration, you must effectuate the calibration process.

The calibration of the camera represents the calculation of the camera's intrinsic and extrinsic parameters.

The first step in the camera calibration process is obtaining the set images using a calibrated camera to determine the intrinsic and extrinsic parameters of the camera. The calibration process is made using a tool from Continental Automotive (Hannover, Germany). For this, we need a pattern, which is a 9 × 6 chessboard. For a good calibration, it is essential to take the images from different distances and different angles, as in Figure 3.
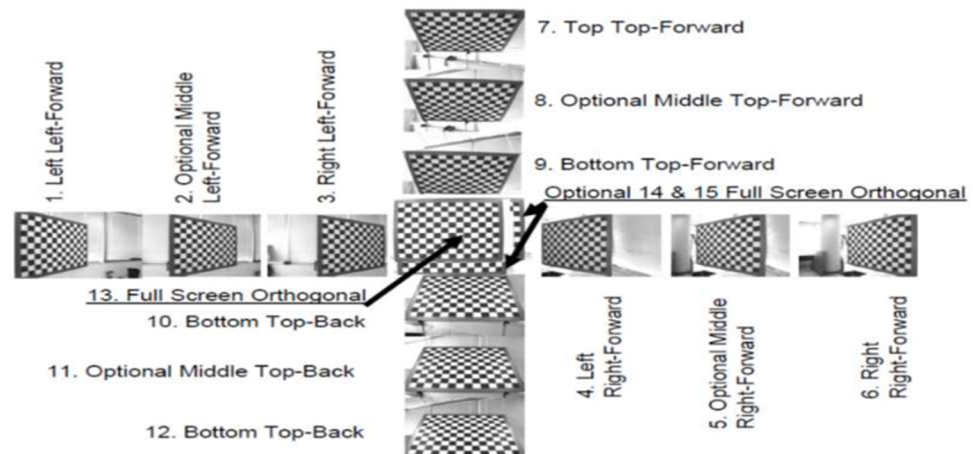
Figure 3. Models positions of pattern for camera calibration.

The next step is the determination of the camera's intrinsic parameters by detecting the corners of the pattern. In Figure 4, only two of the set images needed to determine the intrinsic parameters can be seen.
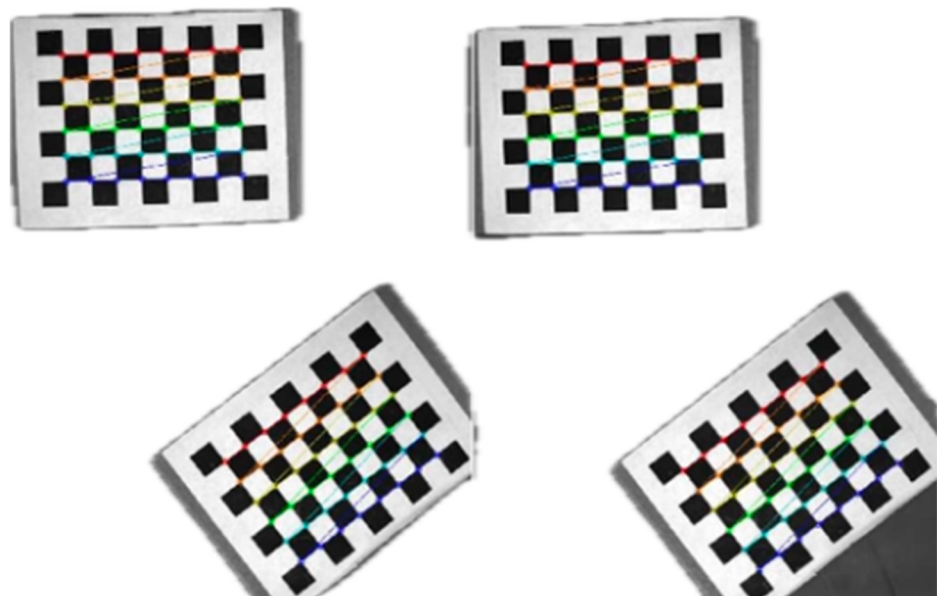
Figure 4. Determining the intrinsic parameters of the camera.

The extrinsic parameters are determined by detecting the pair points from the images taken from the two cameras, left and right (Figure 5).
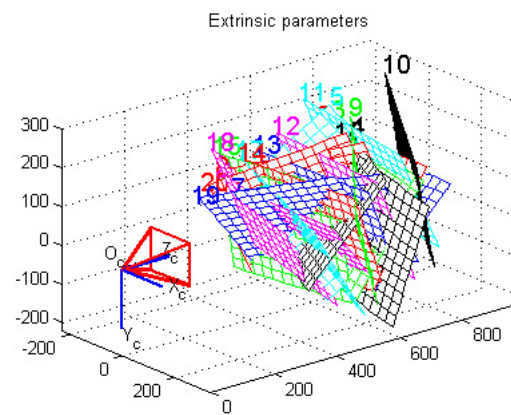
Figure 5. Determining the extrinsic parameters of the camera.

Step 3 is the most important. Now that we have the intrinsic and extrinsic parameters of the camera, this is the moment to rectify the captured images. This is achieved by bringing the epipolar lines to the same level, which is visible in Figure 6. As can be seen in Figure 6, the epipolar lines are at the same level, indicating that we currently have a calibrated camera. Currently, if we possess a calibrated camera, as previously indicated at the beginning of this section, it is imperative to assess the calibration level of said camera. Consequently, the following description will elucidate the application module through which we will conduct the calibration verification process. At this stage, it is necessary to provide a pattern to the camera once more, ensuring that all the corners of the chessboard are visible to the camera. This process is predicated on the precise determination of the coordinates corresponding to each corner of every square on the chessboard, relative to the x and y axes.
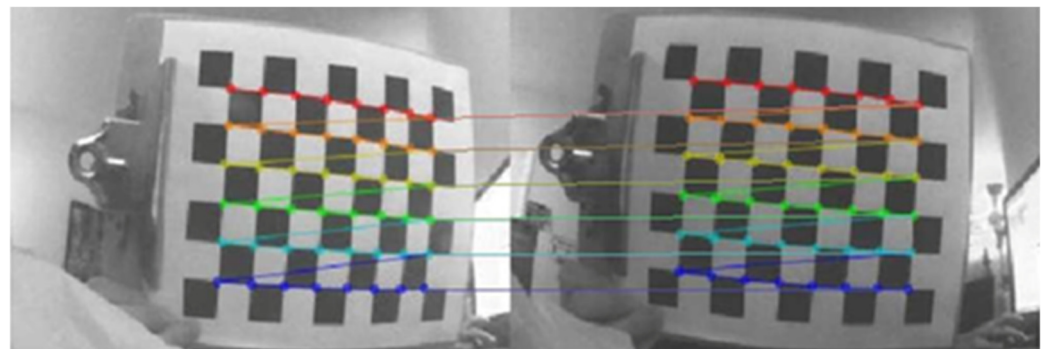

Figure 6. The process of calibrating images by matching the corners.

First, it is necessary to obtain the first pair of pictures so that they can be coordinated x1 and y1 related to the left picture and then coordinated xr and y1 related to the right picture. The first step within the calibration evaluation module (calibration check) is to detect the corners of every square from the left picture and right picture as points PL (xL, yL) and Pr (xr, yr). The coordinates of each square corner are saved in a vector-type point for every image, left and right.

So, we find the coordinates of each corner of the pattern and save them in the vector mentioned above. The next steps represent the conversion of the coordinated values of each corner into whole values. We check the difference in coordinates x and y for each point of the corners detected in the pictures. The error represents the absolute difference value between two points; the error variable values are as follows: 0, 1, 2, or a value greater than 2, which we note with N. We declare what variable to use to stock each error value so that we have err_0, err_1, err_2, and err_N. In the variable err_0, we stock the number of

points with the error value 0; i.e., the pixel in the right picture is at the exact coordinates as those of the pixel from the reference picture, the left picture. This is not offset in any direction. Err_1 stocks the number of points with the error value 1, which means that the point from the right picture to the reference picture is offset with one position to the left, right, up, or down. Err_2 stocks the number of points with the offset difference with two points of the point PR (xR, yR) to the point PL (xL, yL). Any other offset value greater than 2 we consider an N value, which represents that the offset point is at least 3. Figure 7 shows the schematic representation of a limited number of pixels from a picture, with an offset of zero positions in the corner from the right picture and the reference picture.
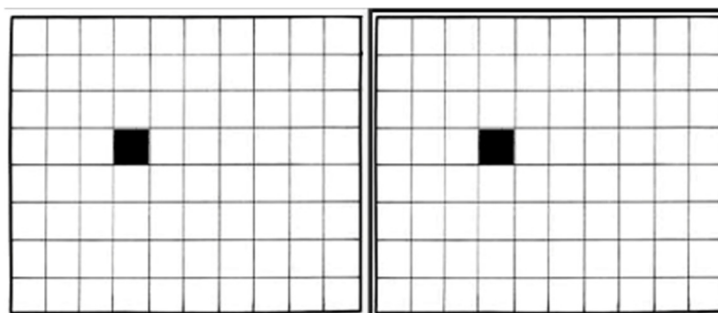
Figure 7. The offset with zero positions.

Figure 8 represents the offset of one position with a corner position, which may be at one end of the red lines.
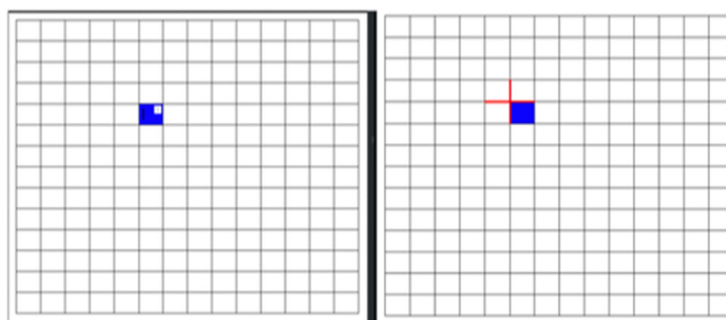

Figure 8. The offset with position: up, down, left, or right.

Figure 9 represents the offset of two positions; the corner of this one can be positioned at one of the red lines (i.e., at an offset that equals 2).
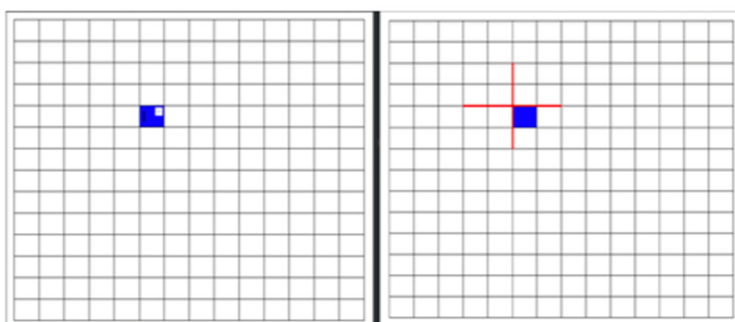

Figure 9. The offset with two positions: up, down, left, or right.

Figure 10 represents an offset of a minimum of three positions in the four directions: up, down, left, or right.
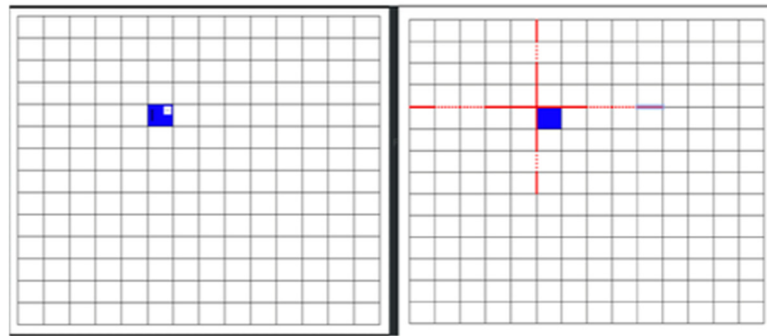
Figure 10. The offset with at least three positions: up, down, left, or right.

The calibration evaluation is achieved in real time, and all we need is to steer the pattern toward the camera and start the evaluation module. Because we want a very high accuracy, we calculate the average of the errors for a set of 20 shots of two pictures each, left and right. The final result is the average of the errors calculated for each pair of images.

Then, we sum up the value of the errors for each pair of pictures, and finally, after we have the final results, we can decide the calibration of the cameras. Upon the completion of this module and the acquisition of the collection of images, a report is automatically generated (refer to Figure 11) detailing the number of errors associated with each pair of images. The report concludes with a summary of the average errors across the 20 pairs of images, accompanied by an interpretation of the findings. Depending on the results, we can have messages like "Very good calibration!", "Good calibration!", and "Weak calibration!".

```
Calibration check mode active
 Please point camera to a checkboard pattern in order to evaluate the calibration
 20 images needed for evaluation
|
1 image(s) captured! Move the pattern to capture a new pose and press SPACE to continue...
1. 46 8 0 0
2 image(s) captured! Move the pattern to capture a new pose and press SPACE to continue...
2. 41 13 0 0
3 image(s) captured! Move the pattern to capture a new pose and press SPACE to continue...
3. 38 16 0 0
    .....................
20 image(s) captured! Move the pattern to capture a new pose and press SPACE to continue...
20. 42 12 0 0
Avg. errors:
        avg. error_0 : 41
        avg. error_1 : 13
        avg. error_2 : 0
        avg. error_N : 0
Exit Calib Check mode.


    Good calibration!
```

Figure 11. Calibration assessment report.

These messages are automatically generated from the code. We consider a calibration "Very good calibration!" if the average error of 0, i.e., err_1, is greater than 50 and the average error of 1, i.e., err_1, is greater than 4. "Good calibration" is when the average of zero errors and one error is greater than 52. If the sums of err_0 + err_1 is less than 52, it means that we have the calibration type "Weak calibration!". A very good calibration or good calibration leads to correcting a clear perception of the densities of the objects in the picture, and a weak calibration leads to an ambiguous perception of the objects from the picture. To easily understand the importance of the calibration, it presents the results of the experiments with the correlation between a calibrated camera and a less calibrated one (Figure 12).
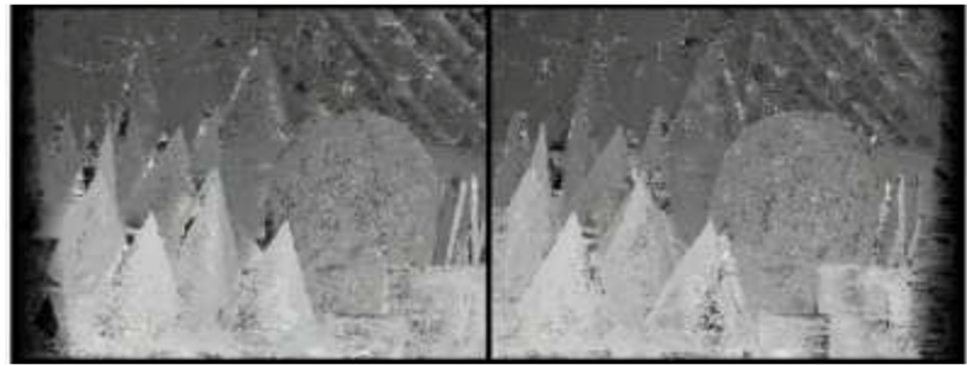
Figure 12. Disparity map before applying the Semi-Global (Block) Matching algorithm.

(2)   Validating the distance from the camera to the object

The first step in developing the mode in validating the distance from the camera to the focused object is the creation of the map of disparity.

The map of disparity represents the correlation between the pixel positions from the left picture and the right picture. The left picture is the reference picture, and finding the correlation between the left picture, reference picture, and right picture goes as follows: we select a window from the left picture, the centre of the window being the same as the centre of the pixel, the correspondence of which we want to find in the right picture. This pixel is compared to more windows from the right picture, starting with the window at the same location (disparity 0), and it moves away to the left, one pixel at a time, until it finds the least matched spot with the left picture, and the moving value represents the disparity value for that pixel. This process is performed for all the pixels in the left picture (see Figure 12).

A frequent approach for real-time systems, given the limited hardware resources today, is the Semi-Global Matching (SGM) algorithm proposed by Hirschmuller [15]. The original approach uses intensity differences, mutual information (Figure 12), a disparity map before applying the Semi-Global (Block) Matching algorithm, and an optimisation method integrating different paths through the image.

The Semi-Global Matching algorithm is applied to refine and perceive a much smoother density. To form the map of disparity and smooth the density of the objects, which can be seen in Figure 13, it is necessary to transform the map of disparity into the depth map. This is possible using Formula (1):

$$depth = \frac{baseline}{tang(disipq * Q)}, \ Q = \frac{H \ FOV}{H \ Pixels} \tag{1}$$

where baseline is the displacement between the left and right cameras; *disp* is the current pixel value in the disparity map; *H FVO* is the horizontal field of view; *H Pixels* 640 is the width resolution.

Due to the generation of the disparity map and our comprehensive understanding of the procedures involved in the formation of the density map, the steps taken to complete this module can be outlined. A defined area of interest was established, for which the distance from the camera to the object contained within this specified zone was calculated. It is important to note that it did not compute the distance for the entirety of the map, as this would have allowed for the calculation of the depth across the entire image. However, this did not align with our objective, as it would have yielded an average depth for the entire picture rather than the distance from the camera to a particular subject object.

Figure 13. The final disparity map after applying the entire Semi-Global Matching algorithm.

We have defined the interest zone as a square added to the disparity map to determine whether the object is correctly framed and whether the distance is perceived accurately (Figure 13). An analysis of this module will be discussed in the next section presenting the results. Calculating the distance from the camera to the object is closely connected with the entire paper, as the calibration process was incorrectly executed. This implies that the object density in the image was not smooth enough, indicating an error when calculating the distance to the object. The reported distance from the application is shown near the square representing the zone of interest in the disparity image. Figure 14 provides an example of this module within the stereo system evaluation application.
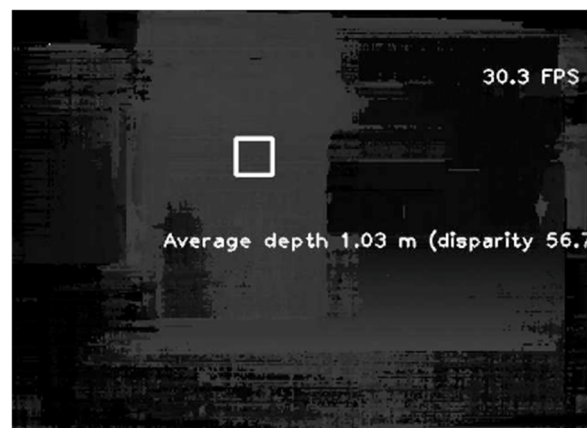

Figure 14. Reporting distance from the camera to the object.

(3)   Representing the distances in the range of colours according to the proximity of the camera objects

Because the disparity map is based on intensities, which means that the objects close to the camera have lower intensities, that is, a lighter colour, and the objects that are far from the camera have higher intensities, that is, they are darker in colour, we apply colour to the disparity map because it is in greyscale format, thereby obtaining a colourless disparity map in which the nearby objects are coloured in red, and depending on the intensity increase, the palette differs in the applied colours, going to dark blue for distant objects. Therefore, depending on the intensity, the colours will also vary (Figure 15).

Upon completing the disparity map representation, we linked it to the image on the left. This connection allowed us to examine both the actual image captured by the camera and the coloured disparity map, facilitating our analysis to determine the accuracy of the application's performance reporting. We have developed an information bar to facilitate the analysis of distances based on colour. This bar serves as a tool for evaluating the distances of objects according to their colour attributes. We established an intensity bar correlated with the maximum intensity value found in the image, which is 255. In greyscale

imaging, a pixel is represented using 8 bits, allowing for values ranging from 0 to 255. Consequently, our intensity bar encompasses values from 0 to 255, with each column in the information bar representing a singular value per row. As one transitions to another column, the intensity of the colour increases. This method enables a correlation between the disparity map comprising the intensities and the information bar formed from the equivalent intensity values. Thus, objects in close proximity in the image will manifest as intensities in the beginning segment of the information bar; conversely, as the distance of the objects increases and the intensities rise, they will appear toward the opposite end of the information bar.
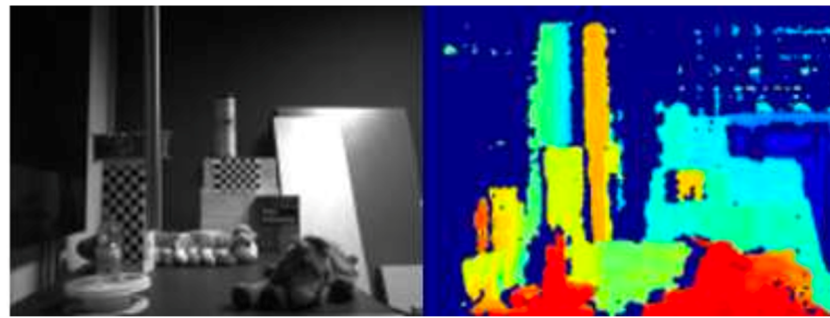
Figure 15. The left image and the colourless disparity map.

In conclusion, we have applied a consistent colour palette to the information bar, ensuring a proper alignment between the colourless disparity map and the information bar, which allows for a more accurate approximation of object distances based on colour (Figure 16). In order to easily distinguish the colours and densities of the objects from the pictures, we applied a stretch to the disparity map, which means the intensification of the values from the disparity map according to a predetermined threshold. If the pixel value is smaller than the threshold value, this will multiply with a smaller scaling factor, and if the pixel value is bigger, this will multiply with a bigger scaling factor. The stretch function was also applied to the information bar. In the two pictures, the left picture and the coloured disparity map, we have added it to the bottom side through concatenation and the information bar. In the lower part of the information bar, the distance in metres is added according to the corresponding colour from far away in blue and for very near in red. The final result, which is displayed after the run of the distance approximation module based on colour, can be seen in Figure 17.

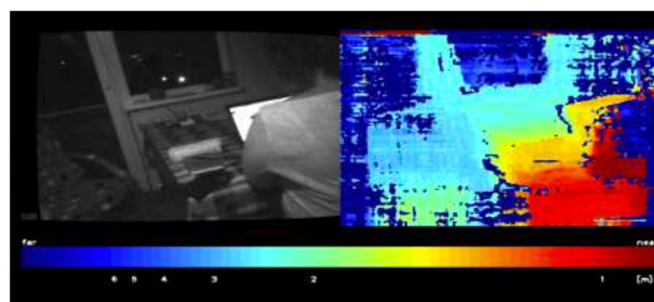**Figure 16.** The colour information bar.

Figure 17. Representing distances in the colour range according to the proximity of camera objects.

(4)   Comparison of Implementations Embedded with PC Implementation

This application module establishes a parallel between disparity maps that are computed at two distinct levels. One of the maps is computed at the chip level on the Myriad 2 platform, with the algorithm being implemented in the assembly language. In contrast, the second implementation utilises the C programming language to compute the disparity map on the computer, rather than on the chip. The primary distinction between these two implementations lies in the representation of the algorithm: the algorithm executed on the Myriad 2 chip utilises floating-point representation, whereas the algorithm executed on the computer does not employ floating-point representation. We created this module, in which we placed the two disparity maps in parallel, and in the middle of them, we produced a new image that displays the difference between the disparity map from the chip and the disparity map from the computer. We compared each pixel from the chip disparity map with each pixel from the disparity map from the computer, representing it in red if the difference between the current pixel on the chip map and the pixel from the computer map exceeded the threshold we set using the track bar, which allows real-time adjustments to the tolerance threshold. Figure 18 illustrates the comparison with a tolerance of 25, meaning that the difference between the pixels in the right map may have a value greater than or equal to 25 units compared to the current pixels in the left image.
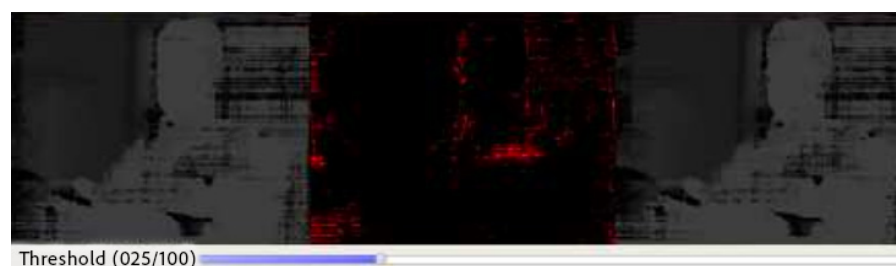

Figure 18. Tolerance accepted: 25 units.

5. Conclusions

The Intel RealSense SR300 merges RGB and infrared features, equipped with a 2-megapixel near-infrared (NIR) camera for improved near-infrared imaging. This camera offers a higher resolution of 1600×1300 versus 640×480, with OmniPixel®3-GS Technology, enhancing low-light performance and reducing noise. Calibration prioritises optimising the performance in the near-infrared range over balancing RGB and infrared.

In summary, while foundational calibration principles persist, NIR camera calibration requires adjustments for higher resolution and specialised technology. The evaluation framework can be adapted for NIR needs, potentially advancing calibration for high-resolution sensors with advanced pixel technologies.

In conclusion, although core calibration principles remain, NIR camera calibration needs unique modifications for resolution and focus. The discussed evaluation method offers a thorough, adaptable framework for NIR requirements, potentially enhancing calibration in high-resolution sensor applications.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Mihail Gaianu was employed by Continental Automotive Romania. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Furgale, P.; Rehder, J.; Siegwart, R. Unified temporal and spatial calibration for multi-sensor systems. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 1280–1286.
2. Daniel Herrera, C.; Kannala, J.; Heikkilä, J. Joint depth and color camera calibration with distortion correction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2058–2064. [CrossRef] [PubMed]
3. Smisek, J.; Jancosek, M.; Pajdla, T. 3D with Kinect. In *Consumer Depth Cameras for Computer Vision*; Springer: London, UK, 2013; pp. 3–25.
4. Hirschmuller, H. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Los Alamitos, CA, USA, 20–25 June 2005; Volume 2, pp. 807–814.
5. Nguyen, C.V.; Izadi, S.; Lovell, D. Modeling Kinect Sensor Noise for Improved 3D Reconstruction and Tracking. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Zurich, Switzerland, 13–15 October 2012. [CrossRef]
6. Castells-Rufas, D.; Ngo, V.; Borrego-Carazo, J.; Codina, M.; Sanchez, C.; Gil, D.; Carrabina, J. A Survey of FPGA-Based Vision Systems for Autonomous Cars. *IEEE Access* **2022**, *10*, 132525–132563. [CrossRef]
7. Zeng, C.; Wei, R.; Gu, M.; Zhang, N.; Dai, Z. High-Precision Calibration Method and Error Analysis of Infrared Binocular Target Ranging Systems. *Electronics* **2024**, *13*, 3188. [CrossRef]
8. Tremblay, P.; Belhumeur, L.; Chamberland, M.; Villemaire, A.; Dubois, P.; Marcotte, F.; Belzile, C.; Farley, V.; Lagueux, P. Pixel-Wise Real-Time Advanced Calibration Method for Thermal Infrared Cameras. In Proceedings of the 3rd International Symposium on Structural Health Monitoring and Nondestructive Testing, Quebeq City, QC, Canada, 25–26 November 2020; Volume 25.
9. Pislaru, M.; Vlad, C.S.; Ivascu, L.; Mircea, I.I. Citizen-Centric Governance: Enhancing Citizen Engagement through Artificial Intelligence Tools. *Sustainability* **2024**, *16*, 2686. [CrossRef]
10. Dușa, C.P.; Bejan, V.; Pislaru, M.; Starcea, I.M.; Serban, I.L. A Multimodal Fuzzy Approach in Evaluating Pediatric Chronic Kidney Disease Using Kidney Biomarkers. *Diagnostics* **2024**, *14*, 1648. [CrossRef] [PubMed]
11. Freedman, B.; Shpunt, A.; Machline, M.; Arieli, Y. Depth Mapping Using Projected Patterns. Patent US20 100 118 123A1, 23 August 2013. Available online: https://patents.google.com/patent/US20100118123A1/ (accessed on 15 February 2025).
12. Han, J.; Shao, L.; Xu, D.; Shotton, J. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Trans. Cybern.* **2013**, *43*, 1318–1334. [PubMed]
13. Webster, D.; Celik, O. Systematic review of kinect applications in elderly care and stroke rehabilitation. *J. Neuroeng. Rehabil.* **2014**, *11*, 108. [CrossRef] [PubMed]
14. Fanello, S.R.; Rhemann, C.; Tankovich, V.; Kowdle, A.; Escolano, S.O.; Kim, D.; Izadi, S. Hyperdepth: Learning depth from structured light without matching. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5441–5450.
15. Sighencea, B.I.; Stanciu, R.I.; Căleanu, C.D. A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction. *Sensors* **2021**, *21*, 7543. [CrossRef] [PubMed]