

## Article

# Indoor-Guided Navigation for People Who Are Blind: Crowdsourcing for Route Mapping and Assistance

Darius Plikynas <sup>1,\*</sup>, Audrius Indriulionis <sup>1</sup>, Algirdas Laukaitis <sup>2</sup> and Leonidas Sakalauskas <sup>3</sup>

<sup>1</sup> Department of Business Technologies and Entrepreneurship, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania; audrius.indriulionis@gmail.com

<sup>2</sup> Department of Information Systems, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania; algirdas.laukaitis@vilniustech.lt

<sup>3</sup> Faculty of Natural Sciences, Klaipeda University, 92294 Klaipeda, Lithuania; leonidas.sakalauskas@mif.vu.lt

\* Correspondence: darius.plikynas@vilniustech.lt

**Abstract:** This paper presents an approach to enhance electronic traveling aids (ETAs) for people who are blind and severely visually impaired (BSVI) using indoor orientation and guided navigation by employing social outsourcing of indoor route mapping and assistance processes. This type of approach is necessary because GPS does not work well, and infrastructural investments are absent or too costly to install for indoor navigation. Our approach proposes the prior outsourcing of vision-based recordings of indoor routes from an online network of seeing volunteers, who gather and constantly update a web cloud database of indoor routes using specialized sensory equipment and web services. Computational intelligence-based algorithms process sensory data and prepare them for BSVI usage. In this way, people who are BSVI can obtain ready-to-use access to the indoor routes database. This type of service has not previously been offered in such a setting. Specialized wearable sensory ETA equipment, depth cameras, smartphones, computer vision algorithms, tactile and audio interfaces, and computational intelligence algorithms are employed for that matter. The integration of semantic data of points of interest (such as stairs, doors, WC, entrances/exits) and evacuation schemes could make the proposed approach even more attractive to BSVI users. Presented approach crowdsources volunteers' real-time online help for complex navigational situations using a mobile app, a live video stream from BSVI wearable cameras, and digitalized maps of buildings' evacuation schemes.

**Keywords:** electronic travelling aids; guided navigation; blind and severely visually impaired; social networking; crowdsourcing; computer vision



**Citation:** Plikynas, D.; Indriulionis, A.; Laukaitis, A.; Sakalauskas, L. Indoor-Guided Navigation for People Who Are Blind: Crowdsourcing for Route Mapping and Assistance. *Appl. Sci.* **2022**, *12*, 523. <https://doi.org/10.3390/app12010523>

Academic Editors: Martin J.-D. Otis and Alexandre Robichaud

Received: 2 November 2021

Accepted: 15 December 2021

Published: 5 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The growing number of people who are BSVI (blind and severely visually impaired) with smartphones, which are multifunctional, multisensory GSM networking devices, has provided the impetus for the development of considerably cheaper electronic traveling assistive (ETA) devices that can employ hardware and software gadgets (defined by the 3GPP standards collaboration) integrated into smartphones. Smartphones are equipped with a CPU, an operating system, various sensors (GPS, accelerometer, gyroscope, magnetometer, pedometer, and compass) and can run mobile apps for data processing. Mobile computing platforms offer standard APIs for general-purpose computing, providing both application developers and users a level of flexibility that is very conducive to the development and distribution of novel solutions [1,2]. In addition, smartphones possess real-time GSM connections with the mobile phone network and the internet and facilitate continuous wireless data transfer to external servers and cloud platforms for web-services-based data processing. This considerably enlarges smartphones' general usability and could also be employed for BSVI ETA solutions using social outsourcing, such as remote real-time visual assistance, route mapping, etc.

A few of the technologies involved are reviewed in this paper. First, it is important to note that BSVI individuals do not differ remarkably from the visually able population with regard to smartphone use. In fact, due to their condition, BSVI individuals are even more inclined to use handheld smartphones for social communication and mobility (making calls, chatting, using social media and many other apps, including GPS navigation, and so on). The screen reader interface integrated into modern mobile operating systems is accessible enough for people who are BSVI. The number of mobile apps tailored for blind users is also increasing, boosting the use of mobile devices and apps among people who are BSVI, and this usage is expected to continue to grow [3].

Relatively few studies have been conducted on mobile app use among people who are BSVI. In some preliminary studies, participants rated apps as useful (95.4%) and accessible (91.1%) tools for individuals with visual impairment. More than 90% of middle-aged adults strongly agreed that specifically tailored apps were practical. This shows that BSVI individuals frequently use apps that are specifically designed to help them accomplish daily tasks. Furthermore, it was found that the BSVI population is generally satisfied with mobile apps and is ready for improvements and new apps [3]. Thus, among the BSVI community, the multifunctional usage of smartphones for general and specialized tasks is widespread, and there is potential for them to be readily adapted for crowdsourced ETA solutions indoors.

Recent advances in computer vision, smartphone devices, and social networking opportunities have motivated the academic community and developers to find novel solutions that combine these evolving technologies to enhance the mobility and general quality of life of people who are BSVI. Unfortunately, this prospective research niche is not well covered in existing research papers. The only reviews we could identify were several that focused on existing mobile applications for the blind [1–3]. These findings suggest that electronic travel aids, navigation assistance modules, and text-to-speech applications, as well as virtual audio displays, which combine audio with haptic channels, are becoming integrated into standard mobile devices. Increasingly user-friendly interfaces and new modes of interaction have opened a variety of novel possibilities for the BSVI [4,5].

It is important to note that multifunctional mobile devices are increasingly being employed as interim embedded sensory processing units (exploiting IMU sensors, cameras, etc.) and BSVI user control and interface devices (sound and tactile feedback). In the embedded settings, they work in the context of larger ETA systems, where other often specialized and more powerful sensory controlling devices and local mini PC or remote processing units (such as web cloud servers) are used [1,6]. In this way, a mobile device becomes an embedded part of a larger networking system, where Web 2.0 services can be employed.

Admittedly, a wide range of general-purpose social networks, web 2.0 media apps, and other innovative ICT (information and communication technology) tools are developed to improve navigation and orientation. Although they are not destined to meet the specialized requirements of people who are BSVI, some mobile apps make them useful. For instance, text (and image) to voice, tactile feedback, and other additional enabling navigation software and hardware solutions are helpful for this matter. However, the complexity and abundance of general-purpose features pose a significant challenge for people who are BSVI. According to Raufi et al. [7], the increasing volume of visual information and other data from social networks confuses BSVI users. In this way, the expansion of general-purpose vision-based web 2.0 social networks leaves behind specialized digital (audio and tactile) content accessibility for BSVI users [8]. Some approaches that are more focused on the needs of BSVI are required.

In general, BSVI users are actively involved in social networks [9–11]. More than 90% of people who are BSVI actively use one or more general-purpose social networking means, such as Facebook, Twitter, LinkedIn, Instagram, and Snapchat [9,10,12–14]. However, only a few social networking platforms have additional features oriented for BSVI users. For instance, BSVI surveys have revealed that social networking apps are among the five most

popular mobile apps [3]. The majority of people who are BSVI who use social media use the Facebook social network [10,15,16]. The use of Twitter is also unusually high, and it is assumed that its simple, text-based interface is more accessible to screen readers [10].

Next to the general-purpose social networks, people who are BSVI frequently use apps that are specifically designed to allow them to accomplish daily activities. However, N. Griffin-Shirley et al. emphasize that persons with visual impairments would like to see improvements in existing apps as well as the development of new apps [3]. There are a number of examples of some of the most popular navigation apps used for path planning, navigation, and obstacle avoidance [1,2,6,9]. Unfortunately, navigation apps are mostly based on predeveloped navigational information and do not provide real-life support, user experience-centric approaches, or participatory Web 2.0 social networking. On the contrary, there are other real-life social apps, such as Be My Eyes, which give access to a network of sighted volunteers and company representatives who are ready to provide real-time visual assistance for orientation, navigation, and other tasks [17]. There are also many other R&D applications that can be applied for the outsourcing of navigational information [18–22].

The above overview of the related literature reveals technological and socially guided indoor navigation advancements, implications, and drawbacks. In this regard, our approach proposes a novel guided indoor navigation solution, which is user-centric, crowd-sourced, and does not require costly prior infrastructural indoor investments, such as the earlier-mentioned installation of Wi-Fi, RFID tags, beamers, etc. However, it does demand the involvement of some social networks, where volunteers walk in the chosen buildings, mark indoor routes, and carry out semantic tagging (using a mobile app with a voice recording or command line) of points of interest (POI) such as doors, exits, lifts, stairs, etc. For that matter, they use wearable ETA equipment with IMU sensors, stereo, and IR (depth) video cameras that record a visual stream and send it through Wi-Fi or GSM to the web cloud server where adapted SLAM algorithms produce clouds of characteristic points for each sequential video frame and use this information to form 3D routes. In our setting, we call this set of procedures the first operational modality.

In the proposed setting, the wearable ETA device undergoes some initial real-time data stream processing using Raspberry Pi4 and local mini PCUs such as NUC, but the main computational vision-based algorithms do the rest of the work offline in the web cloud server (video stream data from the wearable ETA device have to be transferred in advance). The server side analyzes video streams, recognizes objects, classifies them, measures distance and direction, and relates semantic POI tags with route coordinates.

In the second operational modality, the database of prearranged routes for people who are BSVI is maintained in the web cloud server and is available online for their use. People who are BSVI use the same wearable ETA equipment with a local mini PCU, such as the NUC. They can choose building and indoor routes from the online web cloud database using the mobile app. Wearable ETA equipment use computational vision-based algorithms such as SLAM to recognize clouds of points on the BSVI route and guides the BSVI user through bone-conductive headphones and tactile signals. The latter are displayed using a unique headband with a tactile interface display. In this way, people who are BSVI can arrive at their desired indoor destinations through visual-based guided navigation.

In the proposed approach, the third operational modality deals with extraordinary situations such as dead reckoning and unrecognized paths or objects. For that reason, BSVI can use the mobile app to call volunteers who are familiar with that building or who were involved with the production of its routing data in the first modality. The web cloud server provides information about the current BSVI position on the chosen route (or last known location) and on that building's digitalized evacuation scheme. A volunteer can help to recognize obstacles, read texts, and find a route in dead reckoning situations.

Compared with other earlier-reviewed approaches, this novel indoor navigation setting for people who are BSVI requires relatively more input from (i) social networking using the help of volunteers, (ii) AI-based computational intelligence algorithms, (iii) server and client-side processing power units, and (iv) a Wi-Fi or GSM Internet connection.

However, its main advantages are (1) the indoor route database is available 24/7 (a kind of Visiopedia), (2) it has an offline working mode, (3) there is no need for indoor infrastructural investments, (4) it has an autonomous and flexible, wearable ETA device with a tactile display and bone-conductive headphones, (5) real-time online volunteer help is available in complex situations using a mobile app, (6) it employs a user-centric approach, and (7) routes can be rated after each guided navigation to allow consequent improvement of the route database via BSVI feedback.

The remainder of the paper is organized as follows: Section 2 briefly describes the results of a survey conducted on people who are BSVI concerning their navigation and social networking needs and expectations; Section 3 layouts some insights on ETA enhancements of navigation and orientation using the advantages of participatory Web 2.0; Section 4 brings forth the wearable prototype R&D challenges; Section 5 provides web-based crowd-assisted social networking implications for navigation indoors; Section 6 presents the conclusions and discussion.

## 2. Survey of BSVI Social Networking in the Context of Navigation Needs

The presented literature review gives a better understanding of the object of research. However, it lacks practical R&D insights concerning the real-life needs of people who are BSVI for effective and innovative guided indoor navigational solutions. Thus, we had to look for first-hand experience-based and user-centric feedback from people who are BSVI. To define the social networking needs and expectations concerning indoor navigation help, we conducted a semi-structured survey of people who are BSVI (see Figure 1).

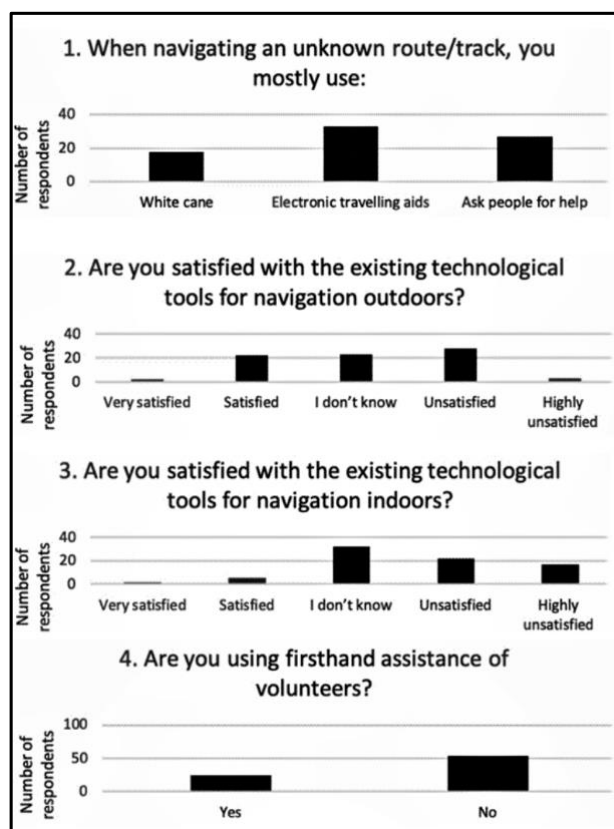
A semi-structured survey was conducted as a part of a research project titled “Complex research of augmented reality for the blind and weak-sighted people” (project No. 01.2.2-LMT-K-718-01-0060) funded by European Regional Development. The goal of the semi-structured survey of BSVI persons was to find out user-defined ETA development niche and specification of tasks to be achieved. This survey aimed to identify BSVI people’s requirements, problems, wishes, visions, and expectations for indoor and outdoor ETA technological solutions. The survey was anonymous.

We surveyed regular people who are BSVI using the online LymeSurvey web service platform at this link: <https://fts.vgtu.lt/survey/index.php/237934?lang=en> (accessed on 23 October 2021), where our questionnaire was available on the web. We used a psychometric Likert-type scale commonly involved in research that employs questionnaires [23]. We used the same anonymous web survey platform and the same questionnaire supplemented with open-ended and priority questions for the BSVI experts’ semi-structured survey. In the case of open-ended questions, we interviewed each expert. Interviews were audio-recorded for later analysis.

In the survey questionnaire, we used 19 questions: (a) 8 demographic questions (age, living location, education, etc.), (b) 11 sight-related questions (level of disability and behavioral patterns). Meanwhile, for the semi-structured interview, we used additional 20 questions: (a) 14 open-ended questions (needs, problems, expectations), (b) 6 priority questions (listing problems and needs in decreasing order of the personal preference). The whole questionnaire is available at this address.

We obtained 87 responses from regular people and 46 from experts. After filtering out the wrong questionnaires, we obtained 78 responses from regular people and 25 from experts. The main reason for wrongly filled responses was experts’ reluctance to fill up the longer questionnaire.

In total, the responses of 78 people who are BSVI located in the EU were analyzed, of which 25 were identified as blind experts (10+ years of experience or active interest in using ETAs for the blind). In the survey, some questions (out of 42 questions in total) concerned ETA navigation functionalities, and others dealt with social networking approaches.



**Figure 1.** Survey responses from 78 people who are BSVI.

As shown in Figure 1, our survey results of 78 people who are BSVI indicate that

1. Over 50% of people who are BSVI do not use firsthand assistance from volunteers;
2. They prefer to navigate an unknown route using ETA (over 35%);
3. They are unsatisfied or highly unsatisfied with the existing technological tools for navigation indoors (over 40%);
4. They are less unsatisfied with the existing technological tools for navigation outdoors (around 30%).

These results indicate that a considerable portion of people who are BSVI are quite autonomous, familiar with ETA, and are looking for better indoor ETA solutions. These user-centric findings shaped the next steps of our research.

To pinpoint future ETA needs, we filtered out the answers of 25 BSVI ETA experts (see Figure 2). Their answers provided an even more unambiguous indication that people who are BSVI need autonomous indoor navigation with better ETA solutions. For instance, over 60% of blind experts were unsatisfied or highly unsatisfied with existing technological tools for indoor navigation. They also placed more emphasis on the needs of people who are BSVI for self-reliant autonomy when using ETAs to navigate indoors (around 45% were inclined to use an ETA for navigation indoors instead of asking for help).

A semi-structured interview of 25 BSVI experts indicates that current ETAs are not sufficient for indoor navigation and orientation applications. They pointed out the lack of real-time, user-friendly, experience-centric, and participatory Web 2.0 technologies employed for specialized BSVI needs. In other words, the current situation concerning the high potential of modern social networks, web 2.0 media apps, smartphones, and other ICT (information and communication technology) tools does not meet the requirements of people who are BSVI.

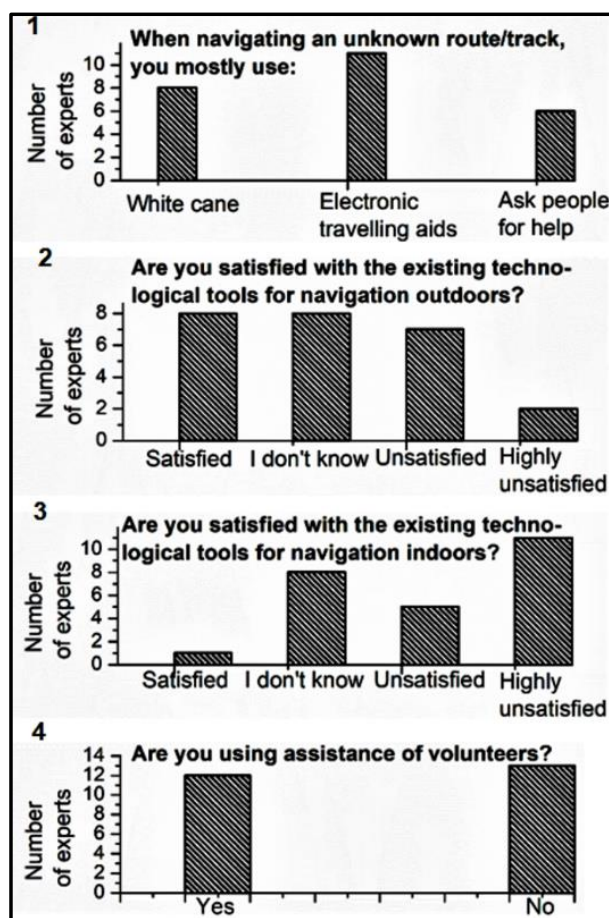


Figure 2. Survey responses from 25 BSVI experts.

BVSI experts were also interviewed concerning the usage of smartphone apps and Web portals. For instance, questions such as “What smartphone apps, web portals, and social networks do you know, and which of them do you use to communicate with sighted people?” and “What smartphone apps, web portals and social networks, specifically designed for the blind, do you know?” revealed that the most popular apps and social networks used by people who are BSVI are Facebook, Twitter, LinkedIn, and Snapchat. Additionally, people who are BSVI use Telegram, Youtube, Facetime, Google Hangouts, WhatsApp, Skype, Viber, Messenger, Zello, MySpace, Tinder, TeamTalk, and Eskimi apps. Only 5 out of 25 BSVI experts mentioned apps or websites specifically designed for the blind: Bee my eyes, Telelight—an accessible telegram client, Voreil, Talking Communities, FourSquare/BlindSquare, Playroom, Applevis.com, Elvis, blindhelp.net, Blindbargens, ACB network, and RNIB.

Blind experts also shared their opinions on more specific questions regarding social networking tools used for navigation. For instance, “Are you familiar with social networking tools that support sharing of navigation information (directions) between the blind and/or sighted volunteers?” Surprisingly, most BSVI respondents were not aware of such tools, and only a few mentioned WhatsApp, Be my eyes, and the Google Groups “Eyes-free group”. Thus, social networking tools used for navigation are not well known or popular. During additional interviews, we collected a few more discouraging details. For instance, social networking tools (i) are mostly in English and do not operate in other national languages; (ii) are run by casual volunteers who are not accustomed to dealing with the specific BSVI problems; and (iii) use applied technology that is not specialized enough to provide real-time help of sufficient quality.

Other questions included “Would you be willing to pay for the functionality of an electronic travel aid listed below? How much?” From these, we identified value-added and monetary estimations of each ETA functionality. About 80% of the top-ranked ETA needs included navigation and orientation functionality outdoors and indoors, including recognition of stairs, elevators, and doors; navigation directions; and assistance to return to a specific location. In terms of the total price for all twenty chosen ETA functionalities, people who are BSVI stated that they are willing to pay for the following services: 18.3% for outdoor navigation; 12.9% for indoor navigation; 12.5% for recognition of textual and numerical information; 8.4% for recognition of stairs, lifts/elevators, doors, passages and pavements/sidewalks; 6% for information about products with BAR and QR codes; 5.8% for assistance from remote volunteers to interpret sophisticated surroundings in the mother tongue; 3.1% for the ability (through social networking) to record, store and reuse outdoor navigation information; 2.7% for ability (through social networking) to record, store, and reuse indoor navigation information; 1.9% for the ability to share and exchange outdoor navigation directions through a specially designed social network; and 1.8% for the ability to share and exchange indoor navigation directions through a specially designed social network. The latter two estimates were low due to the lack of enabling technologies and BSVI user experiences. When we explained to the BSVI experts how the approach could work using specialized ETA means and participatory Web 2.0 outsourcing of indoor mapping and routing tasks, the BSVI experts’ opinion changed to be in strong favor of the approach.

In sum, around 32% of the total price that people who are BSVI were willing to pay was for ETA functionalities, which can be substantially enhanced using participatory Web 2.0 social networking. In the next section, the latter possibility is discussed in more detail.

We have presented just a few exemplary semi-structured interview questions. Based on the entire survey and interview analysis, we identified an R&D niche in the field of navigational ETA solutions. It mainly concerns navigation and orientation indoors and the exploitation of experience-centric and participatory Web 2.0 technologies for social outsourcing of indoor mapping and routine tasks. Based on these insights, we made some inferences regarding a combination of modern enabling technologies that could be employed successfully in this regard. In the next section, we give an exemplary case.

### **3. Indoor Navigation: Towards a Crowdsourced Approach**

In this section, we provide an overview of a few participatory Web 2.0 social networking solutions, which enhance the navigation capabilities of people who are BSVI for traveling, shopping, and other everyday BSVI mobility tasks. Afterwards, we provide a narrative about the proposed novel crowdsourced-based ETA solution that is tailored to conform to BSVI needs for indoor navigation following the BSVI survey and interview results provided in the section above.

#### *3.1. Indoor Navigation Approach*

Admittedly, regarding real-time assistance and guidance for routes, the primary objective of the specialized mobile apps with wearable services is to assist visually impaired or blind users to navigate indoors with online or offline help obtained from an online community. In the client–server model, a wearable client smartphone or another specialized sensory device (i) streams live video to a crowd server (working as a social navigation networking service) for sighted volunteers using the internet/WiFi and (ii) receives near-real-time feedback with assistance and guiding instructions from the crowd server.

Below, we provide a few examples of such social collaboration. For instance, SoNavNet allows connected users in the social network to share navigation information with the intention being to provide more personalized navigation methods and routes based on member experience rather than on the shortest distance. SoNavNet is based on an experience-based approach—through communication (using online social media) and collaboration (sharing and exchanging experiences), people who are BSVI can find suitable routes, both outdoors and indoors, that meet their specific needs and preferences. SoNavNet, as an online social navigation network system, facilitates the sharing and exchange of experiences with points of interest (POIs), routes of interest (ROIs), and areas of interest (AOIs) [24].

The authors [25] designed the Tales4Us platform to promote creativity, collaboration, and a learning process in which BSVI and other communities can share their shopping stories through a specialized social network. The application has the following major functionalities: (i) the user can play other users' shopping stories and (ii) users can record new stories and share them with the community.

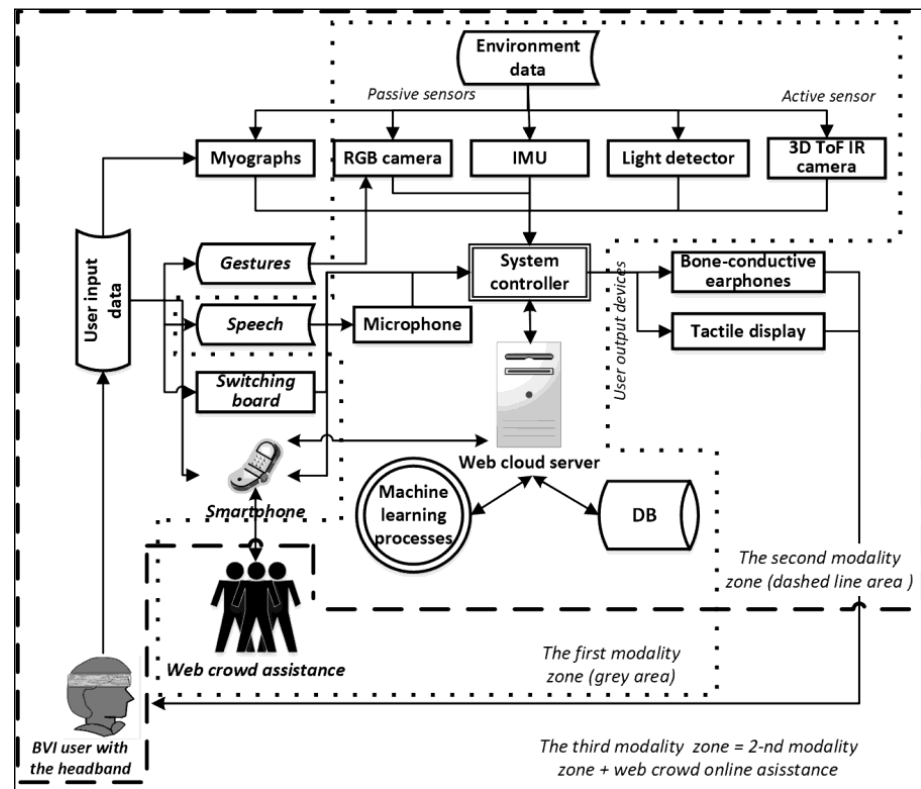
In the case of the "Seeing-eye person" proposed in [26], a crowdsourcing approach enables multimedia data sharing and services for BSVI navigation. The goal of this work is to provide user-accessible crowd services (uniquely tailored for people who are visually impaired) that are flexible (with a friendly HCI and APIs that make it easy to plug in new apps to motivate online volunteers to provide their services) and efficient (near real-time response and a balanced workload between the mobile phone, the back end system, and the different types of users).

The authors [27] designed their general-purpose social navigation approach to be available for any user, including people who are BSVI with impaired mobility. The system allows knowledge to be shared between users, and existing places can be reviewed freely and new ones can be uploaded to the global database, improving the application content. The ParticipAct infrastructure implements calls to different external API services, such as geocoding, localization, routing calculation, and the download of POI entities, enabling a new set of functionalities. However, the system does not include data quality support in the sense of automatic filtering-out of erroneous inputs as (possibly) fake entities.

In sum, after a semi-structured survey of people who are BSVI, an interview with BSVI experts (see section above), and a brief overview of the recent developments in the field (see above), we came up with some meaningful empirically based insights concerning the context of experience-driven indoor navigation and related R&D solutions. Consequently, based on this knowledge, we identified some R&D niches for the further enhancement of composite ETA solutions for people who are BSVI. Below, we describe our insights and the proof of concept, i.e., a proposed novel ETA prototype for indoor navigation using offline and online crowdsourced assistance from volunteers.

The novel ETA system presented below is a compound technology of innovatively adapted hardware devices, such as the 3D ToF IR camera, the RGB camera, a specially designed tactile display with EMG sensors, bone-conducting earphones, a controller, an IMU, GPS, a light detector, and compass sensors. GSM communication was implemented as a stand-alone device or smartphone that works as an intermediate processing device. Passive sensors passively collect environmental data, whereas active sensors, such as the 3D ToF IR camera, emit IR light to estimate distances to objects (see the principal scheme in Figure 3). Multi-sensory data are used to (i) find needed objects, (ii) locate obstacles, and (iii) infer users' locations in an indoor environment in order to help navigate. The devices and sensors observe the environment in real-time and send data via the controller to be machine learning processed, where feature extraction, object recognition, and data storage occur in the web cloud database server. Our approach integrates devices and interfaces using modern technology and methods from the machine learning and computational vision domain.





**Figure 3.** The principal scheme, where rectangular boxes indicate devices. The dotted line delineates the first modality (sighted volunteers mark indoor landscapes); the dashed line delineates the second modality (people who are BSVI obtain navigational instructions on the spot in real-time); the third modality includes the functionality of the second modality supplemented with the real-time web-crowd assistance.

We propose an algorithm for the automated pairing of camera images with motion classes extracted from raw wearable IMU data. This algorithm allows us to automatically label training sets for the training of imitation-learning controllers, whose outputs correspond to three movement classes (“forward”, “left”, “right”) and a prediction reliability estimate, which is important for our application. In this way, raw IMU time series of movement classes (“forward”, “left”, “right”) are collected by separately executing the corresponding movement. A convolutional LSTM classifier with four outputs (the first three outputs correspond to the classes “forward”, “left”, and “right”, and the fourth is the prediction reliability estimate) is used. Softmax activation is used for the first three outputs, and sigmoid activation is used for prediction reliability estimates. The model is trained using modified cross-entropy (*MCE*) loss:

$$MCE(x, y) = - \sum_{c=1}^3 y_c \log y_c(x) + \lambda \left( - \sum_{c=1}^3 y_c \log y_c(x) - y_4(x) \right)^2 \quad (1)$$

where  $y_c$  is the ground truth label, and  $y_c(x)$  is the corresponding class prediction (probability).

Another example concerns an algorithm that allows transformation between the camera image and tactile display coordinates. This transformation is required to represent rectangles in camera frames in tactile display vibro motor activations. We assume that the tactile display’s coordinate frame is located at the top-left element, and its orientation is the same as the orientation of the RGB camera image pixel coordinate frame. Since coordinate transformation between the tactile display coordinate frame and both the RGB camera coordinate frame and the depth camera coordinate frame is static, data can be stored in the configuration file or even omitted due to the insignificant differences between

coordinate origins. Therefore, the camera's rectangle  $(x_p, y_p, w_p, h_p)$  (top lowest point coordinates, width, and height) pixel space can be linearly mapped to a rectangle in the Vibro display matrix

$$(x_p, y_p, w_p, h_p) \rightarrow \left( \left[ \frac{x_p W_m}{W_p} \right], \left[ \frac{y_p H_m}{H_p} \right], \left[ \frac{w_p W_m}{W_p} \right], \left[ \frac{h_p H_m}{H_p} \right] \right), \quad (2)$$

where  $W_p$  and  $H_p$  are the image width and height, respectively, and  $W_m$  and  $H_m$  are the numbers of vibromotors in the columns and rows of the tactile display. Results are rounded to the nearest integer operator.

In general, from the point of view of the end-user, the presented approach is distinguished from other related wearable indoor navigational ETA novelties in the sense of (a) having intelligent user interface integrity based on its unique tactile display and audio instructions; (b) having a hands-free intuitive control interface that uses EMG (or alternatively mobile app and panel); (c) having a comfortable user-orientated headband design; (d) providing machine-learning-based real-time guidance and object recognition; and (e) using web-crowd assistance to map indoor navigational routes and solve problematic situations. For efficient indoor navigation performance, the presented ETA system is used in three consequently interconnected modalities (see Figure 3):

- (i) Web-crowd assistance when volunteers go through buildings and gather step-by-step indoor route information that is processed in the web cloud server and stored in the online DB;
- (ii) BSVI usage of indoor web cloud DB routes when guided navigational assistance is needed;
- (iii) in complex indoor situations (such as being lost, encountering unexpected obstacles and situations), the BSVI ETA system's multisensory data stream can be used in real-time to obtain voice-guided help from volunteers who are familiar with the particular route or building.

Each modality is composed of the same set of eight modes that are active or work in the background. The user (volunteer or person who is BSVI) can activate these modes via a control interface (see Figure 4):

1. Object detector mode is based on the Faster RCNN neural network. Any other object detection architecture (e.g., CenterNet) can also be used. It accepts color image input from an RGB camera and detects a set of trained object classes, which are essential for people who are BSVI (e.g., corridor, door, elevator, stairs, etc.). Each detection consists of rectangles in the camera image, paired with corresponding class labels and reliability scores. CNN (convolutional neural network) object detection is trained by a standard gradient descend method and a custom training data augmentation algorithm. Our application object class detector can be used to detect physical objects and regions in an image with specific properties (e.g., a traversable/non-traversable area).
2. Specific object detector mode also accepts color image input from the user's camera and detects a set of pre-trained objects. The main difference is that this module not only relies on a CNN object detector but also uses template matching, which allows it to learn new objects instantaneously. Because new object learning is performed as corresponding images are included in the object model's database, new objects can be included by the user or its assistants. This module is based on commercial software.
3. Scene description mode accepts color image input from the user's camera and provides a textual description of the depicted scene. This module utilizes CNN to extract features from the input image and LSTM RNN to map it to a textual representation. Afterwards, the textual description is transformed into the speech for people who are BSVI to use.

4. Face recognition mode accepts color images and detects and recognizes faces within them. The list of faces that can be recognized can be managed by the user. The module is based on commercial software.
5. Optical character recognition mode relies on the CNN object detector module and commercial OCR API composition. The CNN object detector is used to detect the user's hand gestures, which potentially contain useful text. Afterwards, this corresponding region is processed via OCR to extract text.
6. Obstacle recognition mode relies on information from the depth camera. It detects obstacles that would be hard to detect with a white cane (e.g., obstacles in the upper body region). Standard point cloud segmentation methods detect obstacles.
7. Navigation mode relies on imitation-learning deep neural networks and object detection components. The imitation learning component records and learns from a trajectory-conditioned controller, which accepts camera images as input and outputs motion commands (e.g., forward, left, or right turns). In this case, to prepare a navigation module for a particular trajectory, volunteers collect training data from the corresponding trajectory in modality#1 (see Figure 3), which consists of a set of images paired with the motion information (forward, left, or right turns), that is automatically extracted via the RNN LSTM classifier from the wearable component's IMU data. It automatically labels training data, which are further used to train the trajectory controller's neural network.
8. Social networking mode is activated in complex situations when the ETA guiding system cannot help. For instance, when a person who is BSVI is lost, encounters an unrecognizable obstacle, needs a real-time explanation, etc. Volunteers are called via the mobile app to assist. They obtain current route information, a 2D floor evacuation plan, and a camera view of the person who is BSVI to help them.

The inclusion of all modes in three functional modalities is a unique feature of the proposed guided ETA system. The abovementioned modes can be activated in each modality. Hence, in the first modality of the ETA system, indoor objects and routes in buildings are practically explored and recorded by sighted volunteers using our proposed ETA system. Volunteers go through the indoor routes, comment on objects, and mark key guidance points. In other words, sighted volunteers mark indoor landscapes, map navigational directions, and make comments using the ETA system's web-crowd-assisted interface (crowdsourced functionality). This option uses the ETA system when volunteers go through buildings and gather information on indoor routes stored in online DBs, and the machine learning processes take place in the Web cloud server (see Figure 3). In modality#1, the ETA system's software functionality is based on integrating data streams from the active modes (see Figure 4), where mode#1 plays a major role and the other modes work in the background or are inactive. In short, with volunteers' help, the ETA system can generate navigational routes for people who are BSVI. Machine learning algorithms (e.g., deep neural networks) are used to integrate data stemming from sensors, cameras, semantics (e.g., volunteers' comments), third-party geospatial floor plans, etc.

With the help of the machine learning processes, the route is generated as a sequence of interconnected location ID places with associated images and audio information that can be tracked on an interactive map (see Figure 5). Routes with guiding navigational information can be accessed offline or online. This type of information is stored on a web cloud server that can be accessed and used by third parties through a convenient XML or other data format. The depersonalized geospatial data on efficient indoor navigation routes can be accessed by other open-source intelligence platforms.

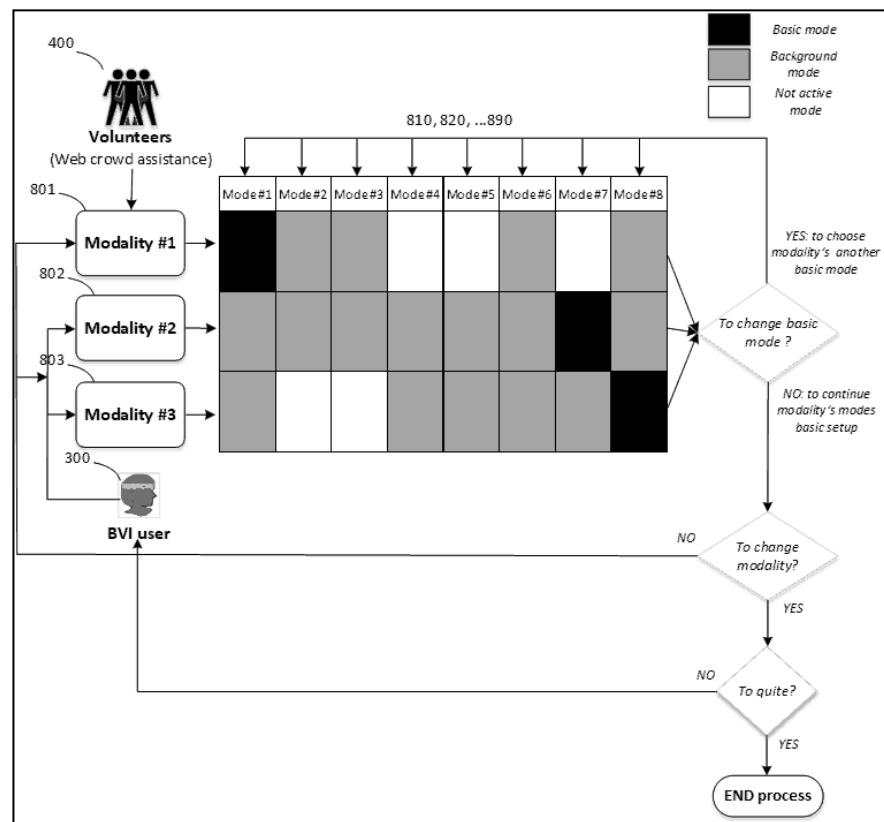


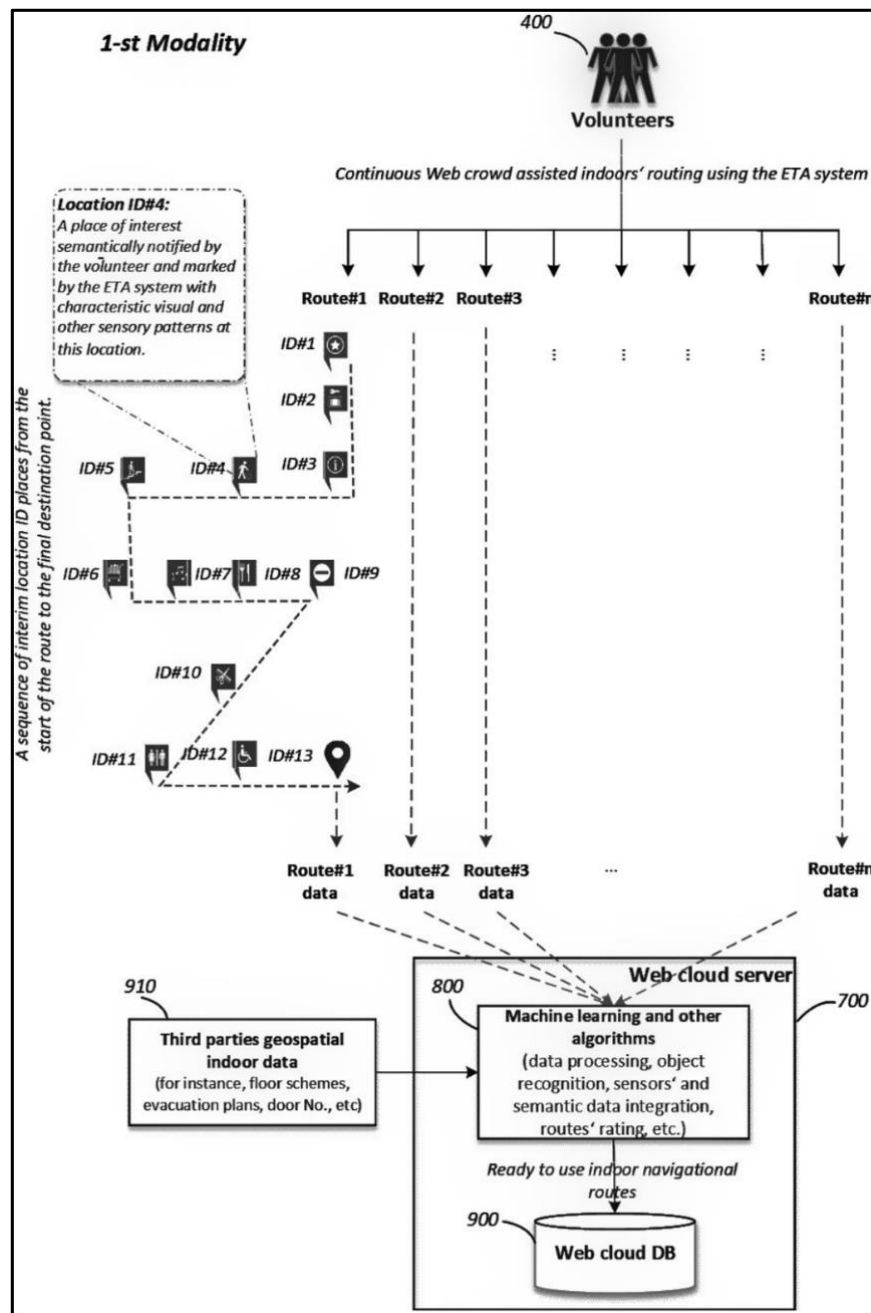
Figure 4. Invocation of the ETA system’s modalities and modes.

### 3.2. Crowdsourcing for Route Mapping

Sequentially, in the second modality, the Web cloud server’s navigational route information (stored in online DB) is used by people who are BSVI for indoor navigational purposes in chosen buildings (see Figure 5). Based on the user’s preferences (faster, shorter, stair-free, most-used, best-rated, most recent, or other options), the machine learning software suggests the best route. The ETA system provides analyzed, semantically enriched, interpreted, and statistically validated indoor routes using information gathered in the first modality. In this way, people who are BSVI can use navigational instructions to (i) become acquainted with the chosen route and (ii) while orientating and navigating indoors. Machine learning and robot navigation approaches are innovatively adapted for this task.

After the trip, BSVI users’ feedback is used to evaluate, improve, and rate navigational route information in the web cloud DB. When a person who is BSVI receives navigational help from the ETA system, they can make additional comments and provide location IDs of landmarks. People who are BSVI can also rate the route. This user-centric feedback helps the ETA system to estimate the route and improve its validity. For this reason, ant colony or other SWARM optimization algorithms can be adapted. SWARM intelligence solves computational problems, which can be reduced to finding the right routing paths, where volunteers and people who are BSVI serve as SWARM agents who help the ETA system find the best routes.

In the third modality, while using the ETA system for navigation, BSVI users can obtain online help in complex, out-of-the-ordinary indoor orientation situations, such as when they are lost or encounter unexpected obstacles. The ETA system can be used in real-time to obtain voice-guided help (through bone-conductive headphones) from a volunteer who is familiar with the particular route or building. It is important to note that, in this modality, mode#8 is active, while modes#1–7 (see Figure 4) can work in the background, optionally informing the BSVI user about the environment.



**Figure 5.** Modality#1: Web-crowd-based mapping process of indoor navigational routes for people who are BSVI using the presented ETA system’s functionality.

By using a smartphone’s app and web-crowd-assisted interface, BSVI users can call registered volunteers who are familiar with that building and indoor route (see Figure 3). Volunteers can help the user to interpret the route, current position, obstacles, or other complex surrounding circumstances. They can do so using information sent from web cloud DB and the BSVI ETA system’s cameras. In the latter case, the current camera view is provided to volunteers. However, as practice shows, that is often not enough for volunteers to make meaningful supporting decisions. They need to understand the contextual grand view of the building’s interior passages. Therefore, in the proposed solution, volunteers can obtain additional information from the ETA system and online DB about the current or last confirmed position of the person who is BSVI on their route map. The ETA system can also provide the time, speed, and movement direction since

the last confirmed position to suggest where the person who is BSVI went astray and is currently located. Additionally, volunteers can obtain building floor schemes and other relevant information from third parties in the online DB. Equipped with this information, volunteers can be much more helpful to BSVI users in complex indoor navigational situations, especially if the ETA guiding system can select volunteers who are familiar with a particular building or route.

Data input on indoor routes are processed in the web cloud server using a machine learning approach. Instructional information on guiding routes is collected in the web cloud database (DB). The best statistical options for successful navigation are estimated each day in the web cloud DB using deep neural networks or other computational intelligence-based methods. In this way, BSVI users can later choose faster, shorter, stair-free, most-used, best-rated, most recent, or other route options. Route updates are constantly sent from the volunteers and people who are BSVI. Such assistance works through social networking when relatives, neighbors, friends, and other people voluntarily and periodically use the ETA system to record the indoor routes that are most important for people who are BSVI. Therefore, even ever-changing indoor situations, such as renovations, furniture movements, closed doors, etc., can be recorded and updated continuously through social networking. The integration of the social networking approach drives a new perspective R&D frontier that will change ETA applications for people who are BSVI. In the next section, an R&D framework that enables the crowdsourced BSVI navigation approach is provided.

#### 4. Bringing Forth Current R&D Challenges

The presented prototype is still in the development stage, and we can only provide initial evaluations of stand-alone modes and associated technologies. There is still some R&D left to go for field testing outside the lab with people who are BSVI.

##### 4.1. Vision-Based Modes

Here is a brief summary of technologies and their evaluations used for the vision-based modes (see Section 3):

[Mode#1] Our object detection subsystem is based on a Faster RCNN algorithm at this link: <https://arxiv.org/pdf/1506.01497.pdf> (accessed on 23 October 2021) Object detectors usually are evaluated in terms of mean average precision (mAP). Faster RCNN with Resnet-101 backbone achieves 48.4% mAP on MSCOCO validation set [28] and is able to process VGA-resolution images in near-real-time (~10 Hz on Nvidia 2080 Ti GPU). Although these are not SOTA characteristics at the moment, we selected Faster RCNN due to satisfactory empirical performance when tested on our prototype and efficient, publicly available implementations.

[Mode#2] Specific (or a few-shot) object detection subsystem is based on Neurotechnology's SentiBotics Navigation SDK 2.0 software (Neurotechnology, Vilnius, Lithuania). In our set up we use the BRIEF descriptor, which according to the tested data set, provides an accuracy of 84.93% [29]. After tuning thresholds, the SentiBotics algorithm allowed practically eliminate false positive recognitions, while providing fast matching and rapid learning of user-specified objects.

[Mode#3] Scene captioning subsystem relies on the [28] algorithm. In terms of the BLEU-4 score, its accuracy is 27.7–32.1 when tested on the MSCOCO data set.

[Mode#4] Face recognition subsystem is based on Neurotechnology's Verilook 12.2 SDK (Neurotechnology, Vilnius, Lithuania). It was evaluated using three publicly available data sets (MEDS-II, LFW, NIR-VIS 2.0). According to obtained estimates, corresponding equal error rate (EER) values:

MEDS: 0.0455% 0.4149%

LFW: 0.0080% 0.0147%

NIR\_VIS 2.0: 0.0367% 0.0662%

It shows that algorithm is capable of high accuracy face recognition.

[Mode#5] OCR subsystem relies on Google Cloud Platform. According to [30] experiments, the average accuracy on the tested data set of 1227 images from 15 categories (after preprocessing) was 81%.

[Mode#6] We employed the PointCloud Library (<https://pointclouds.org/>, accessed on 23 October 2021), which is based upon [https://pcl.readthedocs.io/en/latest/cluster\\_extraction.html](https://pcl.readthedocs.io/en/latest/cluster_extraction.html) (accessed on 23 October 2021) and 3D depth camera images. We used Faster RCNN resnet101/resnet 50 or ssd mobilenet neural network architectures. It is worth noting that the Ligh backbone is much faster.

4.2. Data Structure

At the current R&D stage, the buildings’ 2D floor plans and SLAM trajectories are stored in a local server for online working mode and in the mini PC (such as Intel Nuc) for offline working mode. The points of interest and information about them are stored there as well. Web connection is needed for synchronization. However, at the later stages of the prototype R&D, we plan to store information in the web cloud server, using an enterprise platform for internal and external users with an application, core, and service layers. The structure of information storage will employ relational, file, and object databases.

The main projected data flows and storage are depicted in Figure 6. However, in the current R&D stage, we are testing stand-alone modes and modalities using a simple Wi-Fi connection with a portable mini PC (Intel Nuc) and a server.

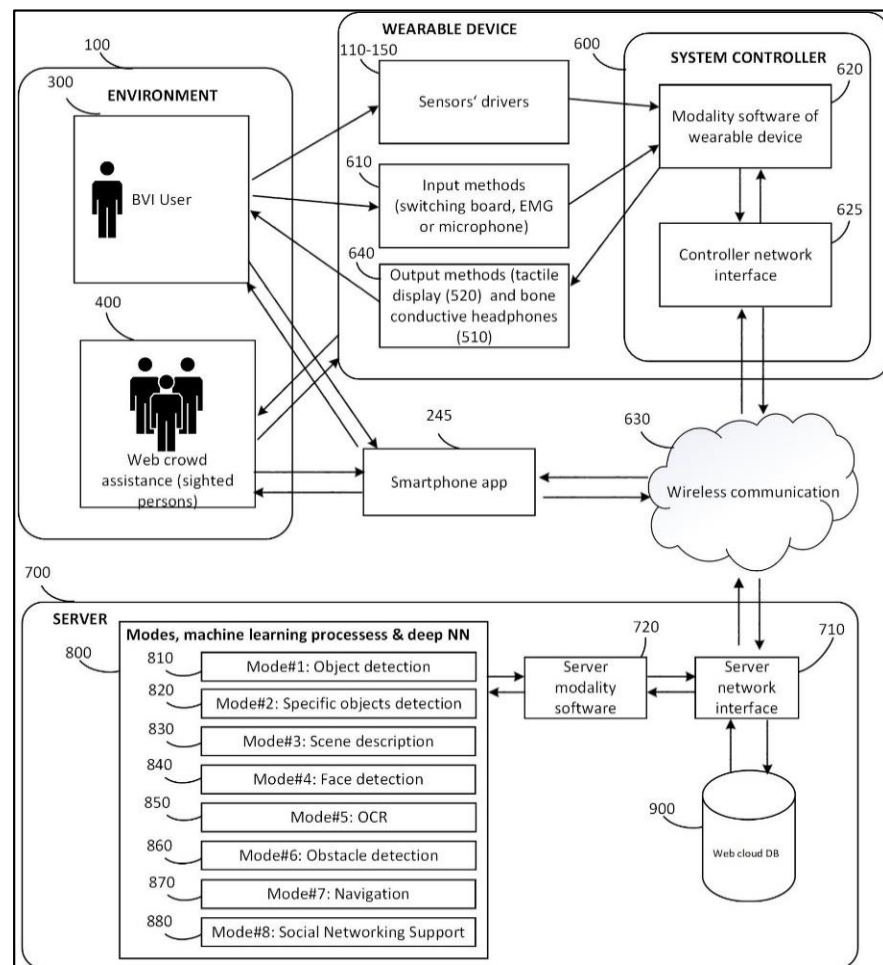


Figure 6. Main communication, data flows and storage.

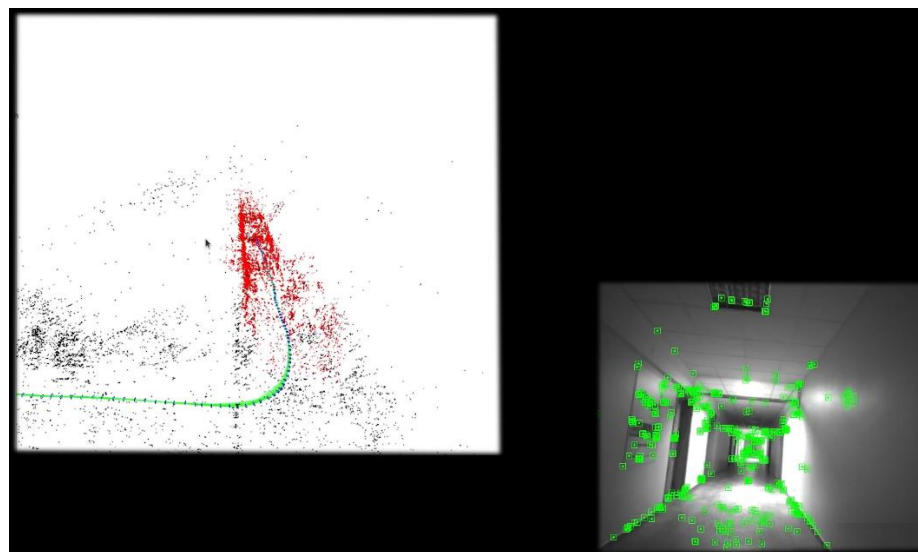
For the functional prototype development, we used software packages with their databases or libraries such as:

- Facial recognition mode's (Verilook, <https://www.neurotechnology.com/verilook.html>, accessed on 20 October 2021) database is stored in the server;
- SLAM database of points of interest, location, routes, and maps is stored in the server;
- accurate open-source Visual SLAM Library
- <https://arxiv.org/abs/2007.11898>; [https://github.com/UZ-SLAMLab/ORB\\_SLAM3](https://github.com/UZ-SLAMLab/ORB_SLAM3), accessed on 20 October 2021);
- scene description—image caption model, based on Google Tensorflow “im2txt” models;
- specific object detection database is stored in the server;
- for object classes (such as doors, lifts, stairs, etc.) classification and recognition we used Faster RCNN resnet101, resnet 50, ssd mobilenet training and validation sets based on our own and open source databases;
- for obstacle recognition, we used the Point Cloud library
- <https://pointclouds.org/>, [https://pcl.readthedocs.io/en/latest/cluster\\_extraction.html](https://pcl.readthedocs.io/en/latest/cluster_extraction.html) (accessed on 20 October 2021);
- for OCR we employed Google OCR API (<https://cloud.google.com/vision/docs/ocr>; (accessed on 20 October 2021));
- for Social Networking we are employing audio-video streaming to a web browser (Android Apps integrated with WEB Cloud DB).

The communication protocol is handled by ROS topics (ROS nodes, topics, and messages | ROS Robotics By Example—Second Edition (packtpub.com; accessed on 10 October 2021); Practical Example—ROS Tutorials 0.5.2 documentation ([clearpathrobotics.com](http://clearpathrobotics.com); accessed on 10 October 2021)), which organizes the data flow and ROS service interaction.

#### 4.3. Crowdsourcing for Route Mapping

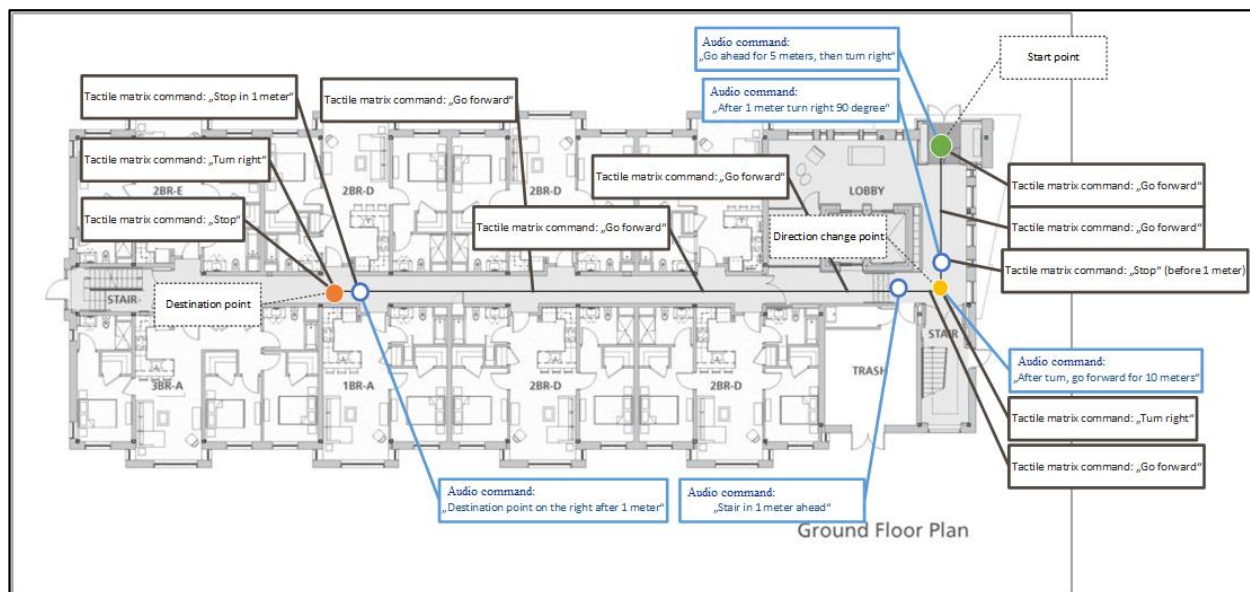
We use the state-of-the-art Visual SLAM algorithm Orb-slam-3 [31] for creating the trajectories that are later used to guide visually impaired users. We use camera poses returned by the SLAM algorithm as the points along the trajectory, see Figure 7. Loop closing implemented within the SLAM algorithm ensures that trajectory drifts are corrected once the volunteers visit the places where they have been before. The recorded trajectories are overlaid on the 2D building floor plans and navigational instructions are built, see Figure 8.



**Figure 7.** Point of clouds (see on the right) and route of such clouds (see on the left) generated with the SLAM algorithm and depicted from the screen.



Seeing the recorded trajectories on the 2D plan the volunteers can verify that all of the places in the building have been visited. The redundant trajectories are generated when the volunteer returns to the place that has been visited before. This process is very important for loop closing and is part of the reliable workflow. On the other hand, this will generate redundant trajectories close to each other. The generated trajectories are converted to a pose graph and we use post processing to merge graph nodes and edges that are spatially close to each other. The processed graph is then used for path planning. At the moment we do not have any mechanism to remove faulty parts of the trajectory graph. If the volunteer sees that created graph has some faulty trajectories, they have to rerecord the floor map again to obtain the correct pose graph. In the future, we plan to add a GUI tool that would allow the volunteer to modify the trajectories after they have been recorded.

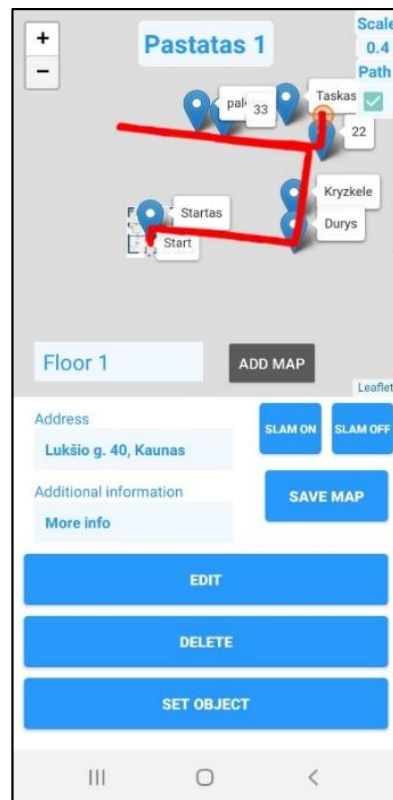


**Figure 8.** Navigation instructions (voice and tactile output) and current location projected on the 2D evacuation scheme, using stored route data.

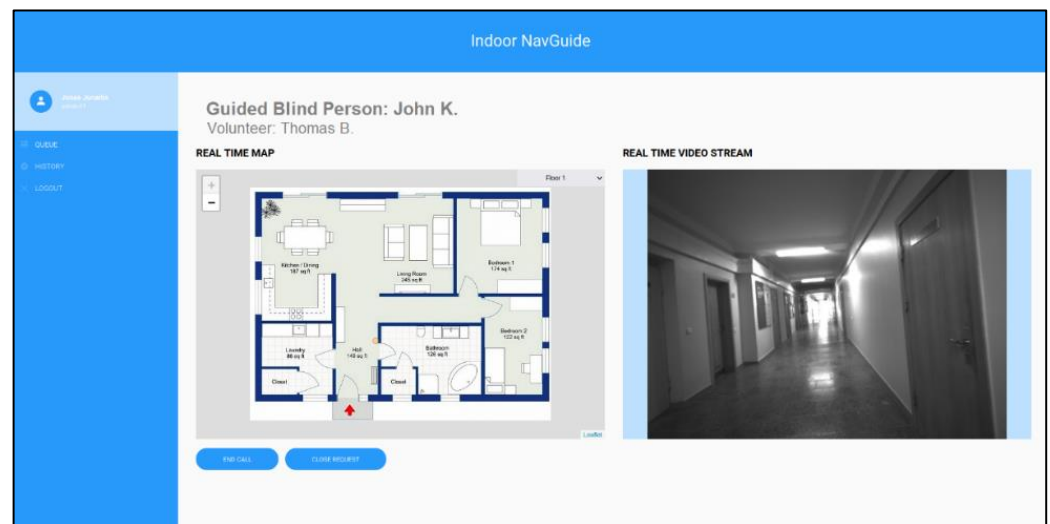
#### 4.4. Verification and Mobile App

At the moment the collected data is only verified by the volunteers. They have to visually inspect the collected data to make sure that it is correct, see Figure 9. We created a specialized android app for the volunteers. This app is used during the data collection stage. The app displays a 2D plan of each building floor and trajectories collected by the Visual SLAM algorithm are overlaid on this map. It is the volunteer's responsibility to verify that collected trajectories cover the whole building floor. Each volunteer would also be trained before performing the building mapping to make sure that they are able to verify the data. At the moment we do not use any automatic trajectory validation algorithms. In the future, this could be conducted using covered area validation algorithms.

In the third operational modality, a BSVI person can use the mobile app to call volunteers who are familiar with that building or who were involved with the production of its routing data in the first modality. The web cloud server provides information about the current BSVI position on the chosen route (or last known location) and on that building's digitalized evacuation scheme, see Figure 10. A volunteer can help to recognize obstacles, read texts, and find a route in dead reckoning situations.



**Figure 9.** Mobile app view used by volunteers to record/delete routes and mark points of interest (Modality#1).



**Figure 10.** Volunteers’ mobile app view (Modality#3: while assisting BSVI person in real-time).

At the current stage, we are working to achieve integrity of the overall system (modalities and modes), improve stability (failures occur in 36% cases), and accuracy of the SLAM algorithm (sometimes occur scaling issues) while building routes. We also have to resolve other technical issues such as more reliable detection of potential obstacles, stable web connection for data transfer, intuitive and user-friendly tactile-voice interface, minimizing time lags for vision-based real-time processing (currently about 2 sec time lag for vision-based recognition of objects), etc.

Thus, our prototype is still in the development stage. Besides, as we stated in the title of the paper, this manuscript is mainly dedicated to shedding light on the innovative web-crowd outsourcing method for indoor routes' mapping and assistance. At the current stage, we are working to achieve integrity of the overall system (modalities and modes), improve stability (failures occur in 36% cases), and accuracy of the SLAM algorithm (scaling issues) while building routes. We also have to resolve a number of other technical issues such as effective detection of potential obstacles, stable web connection, intuitive and user-friendly tactile-voice interface, minimizing time lags for vision-based real-time processing (currently about 2 sec time lag for vision-based recognition of objects), etc.

The literature review and patent DB analysis (see Section 1) indicate that the proposed web crowd assisted indoor navigation setup is unique and hardly comparable to other indoor navigation approaches that use infrastructural installations such as Wi-Fi routers, RFID tags, beamers, etc. Some live photos of the prototype development are presented in Figures 11 and 12.



**Figure 11.** A live prototype (Indoor NavGuide) development photo (headband with tactile display, controller, RealSense stereo camera, power bank, and mini PC).



**Figure 12.** Modality#2: lab testing of guided navigation indoors, using ETA prototype.

#### 4.5. Tactile Display

The tactile display Figure 13 consists of at least 27 vibrating motors (matrix of 3 rows and 9 columns). The base of the tactile display is 3D printed out of silicone (or elastomer). Each vibrating motor is immersed in a silicone cell with different stiffness compared to the base. In this way, the amplitude of the vibration is maximized, and the vibration energy is not transmitted to the other cells. The system controller can output a pulse-width modulated signal Figure 14 to drive vibration motors. The vibrating motor in the cell moves orthogonally to the forehead skin surface. The matrix is covered with a human-friendly elastic material.

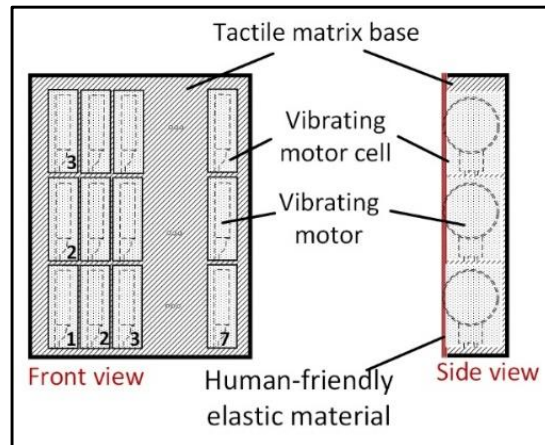


Figure 13. Tactile display.

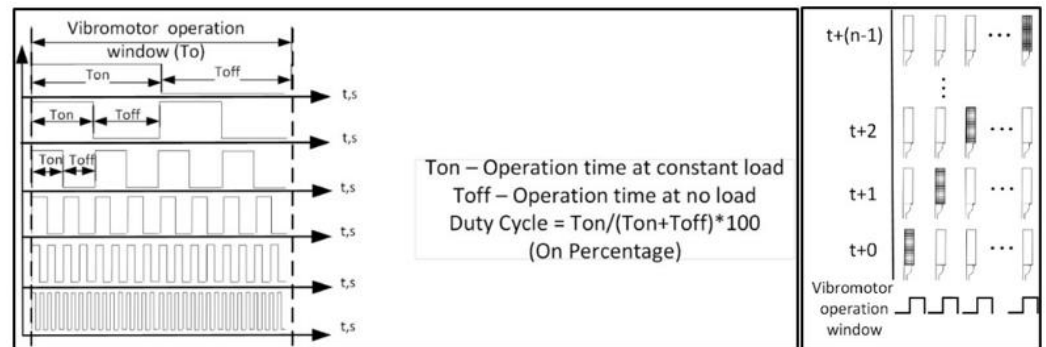


Figure 14. The system controller output to drive vibration motors.

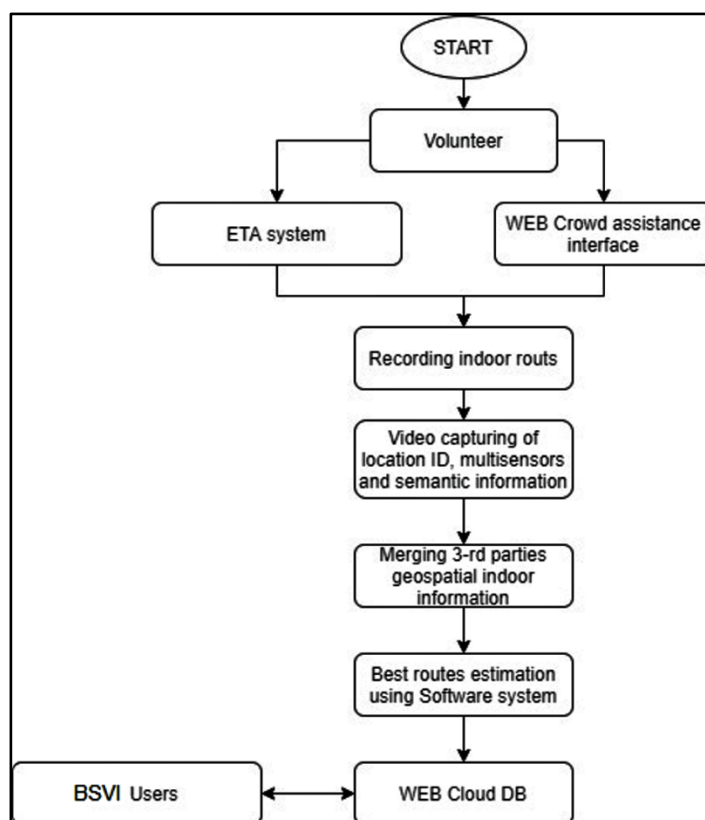
### 5. Web-Crowd-Assisted Social Networking Implications for Navigation Indoors

In the presented approach, following the information provided in the previous section, indoor objects and routes in the building can first be explored and recorded by sighted volunteers using the proposed ETA (electronic traveling aid) system’s interface (see Figures 3–5). That is, volunteers go through the indoor routes, comment on objects, and mark key guidance points (i.e., provide visual and semantic comments of location ID points). In this way, sighted volunteers mark indoor landscapes, map navigational directions, and make comments using the system’s web-crowd-assisted interface (mode#8). Data are collected, for instance, in the web cloud database (DB). Route updates are constantly sent from the volunteers and people who are BSVI using the proposed ETA system. This novel system works through social networking as relatives, neighbors, friends, and other people voluntarily and periodically use the ETA system to record the indoor routes that are most important for people who are BSVI (see Figure 5).

It is important to note that even various daily changing indoor situations, such as renovations, furniture movements, closed doors, etc., can be recorded and updated continuously by volunteers through social networking in the web cloud DB. In unrecognized situations, the trained ETA system can either guide the user around an obstacle or suggest another route. Thus, the presented innovative web-crowd-assisted method enables BSVI users to obtain the latest information about indoor route suitability.

In the web cloud DB, routes are analyzed, summarized, and enhanced using volunteers' records of multisensory data (location points' visual and semantic ID) and third-party information (e.g., building floor plans, indoor maps such as OpenStreetMaps, etc.; see Figure 15). The best statistical options for successful navigation are estimated each day in the web cloud DB using deep neural networks or other computational-intelligence-based methods (see Equation (1)). People who are BSVI can utilize processed navigational routes stored in the web cloud DB using the ETA system. They obtain interactive indoor maps enhanced by the third parties' geospatial data, such as digitalized floor escape plans. In this way, BSVI users, based on their preferences, can choose faster, shorter, stair-free, most-used, best-rated, most recent, or other route options.

After practical experience with a route, BSVI users (and correspondingly the ETA system) can rate the route's validity, making personal averaged ratings ascribed to the route (ascribed to the volunteer who recorded it). This allows other BSVI to choose from the best-rated routes and to obtain offline guidance from the best-rated volunteers (sighted users; see Figure 15).



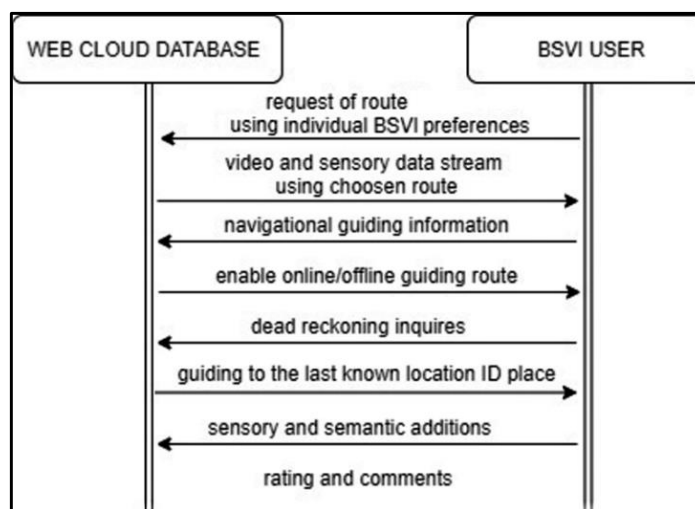
**Figure 15.** Web-crowd-assisted method for indoor routing enhancement and optimization in the first modality using ETA system functionality.

From an extended search of related patents and a literature review, we found that this ETA-guided system enhancement method for indoor navigation a priori using a web-crowd-assisted interface for indoor route mapping with users' feedback is unique [12,13].

Next, it is important to note that indoor route mapping by volunteers with the ETA system allows route optimization for navigation in the first modality (see Figures 3 and 4). Consequently, in the next (second) modality, people who are BSVI can use the enhanced and optimized routes for orientation and navigation indoors. In the proposed approach, based on individual BSVI needs and preferences, the navigational ETA system helps to choose suitable route options (e.g., shortest, fastest, stair-free, guided by top-rated volunteers, etc.) using deep neural networks that provide the route’s object classes, location IDs, destinations, scenes, and semantic information step-by-step. The route, which is adjusted to the individual’s needs, can work online or offline. The latter is needed when there is no internet connection.

In navigational mode, the BSVI user’s wearable ETA system generates a video and sensory data stream, which are provided to the web cloud database and machine learning algorithms for analysis. In this way, objects, location IDs, scenes, and sensory data recognition occur almost in real-time to give navigational support to the BSVI user.

When a BSVI user becomes lost or disoriented, the system can work using a dead reckoning method, i.e., guide the user to the last known location ID place (see Figures 16 and 17). For that matter, the system continually tracks movement using accelerometer, magnetometer, gyroscope, and compass information. This allows for the route of a person who is BSVI to be traced back to the last known location ID. Disorientation cases can be recorded, depersonalized, and processed to warn prospective BSVI users and improve the route guiding quality.

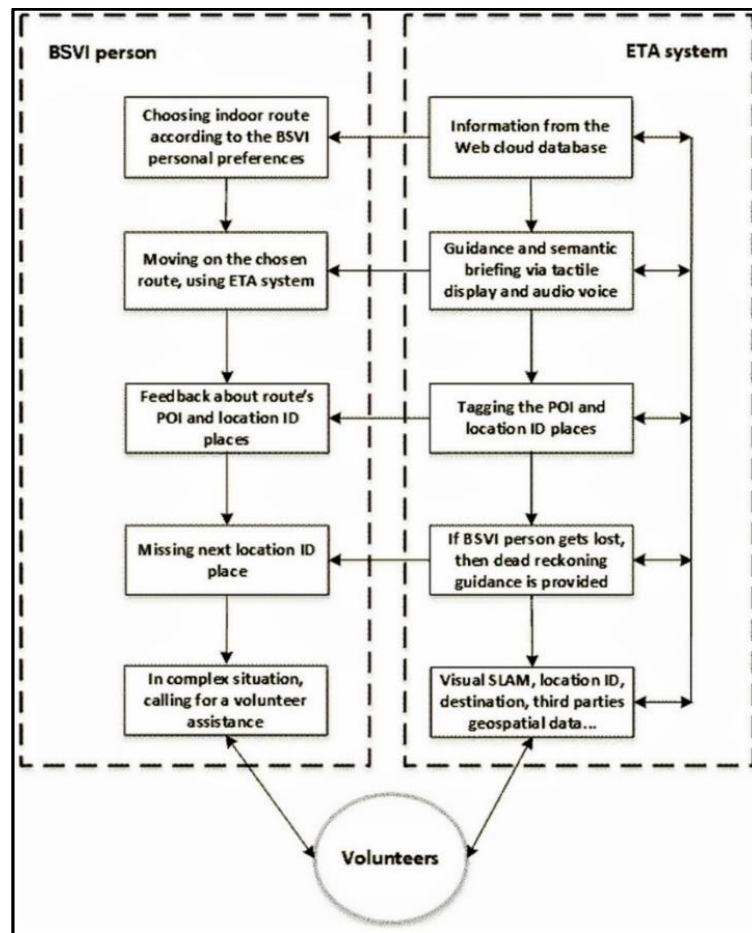


**Figure 16.** BSVI interface with the ETA system’s web cloud database while choosing route and navigating indoors.

While navigating indoors with the ETA guiding system, BSVI users can approve and make estimates and additions to the route’s DB navigation and orientation information (for instance, they can mark new objects, provide voice comments, make new location IDs, etc.). This information is used for improvements to and validation, credibility, and rating of routes. Similarly, BSVI users can add comments about the route regarding observed difficulties, inaccuracies, and errors.

Wayfinding and indoor navigation services for the BSVI population generally have to perform one or more of the following functions: familiarization, localization, route planning, and communication with the user in a meaningful manner through an accessible interface.

The specific abovementioned web-crowd interface advantages only work well in the context of the whole ETA system (see Figure 17).



**Figure 17.** The functionality of the second and third modalities using volunteers' assistance.

Consequently, all three modalities are used where:

1. Proposed visual odometry and SLAM (simultaneous localization and mapping) methods used for sequential recognition of route views have a competitive edge (comparing with other technologies) as they use modern advancements of deep neural networks (such as the convolutional NN);
2. The proposed navigation and orientation method works as an augmented reality decision support system that enables a better perception of indoor environments and does not interfere with the natural senses of the user;
3. It eliminates the need for infrastructural installations such as special marks on the floors, WiFi signal triangulation, 5G signals, beacons, installed Bluetooth devices, RFID, etc.;
4. People who are BSVI can take an active part not only in the rating of passed routes but also in creating and improving them with the help of the ETA system and dedicated software subsystem.

The integration of third-party geospatial indoor information gives additional information needed for matchmaking with the visual SLAM information. However, the proposed innovative experience-centric methods and interfaces present some challenges:

1. Developing routes that can reflect the exact needs and preferences of each individual and their disabilities, given the range of disability conditions and individual preferences;
2. Capturing and adequately quantifying all sensory parameters that affect wayfinding choices and navigation preferences;
3. Building accurate web cloud databases in a scalable and affordable manner;

4. Updating the route database with frequent changes (such as construction) in a scalable and affordable manner;
5. Mapping indoor spaces in affordable and scalable ways while preserving the privacy of relevant information as needed. However, these challenges can be mitigated by adopting an experience-centric approach, where communication and collaboration among members of social navigation networks and other trusted sources form the basis of providing wayfinding assistance.

It must be pointed out that, in real-life situations, even regularly updated navigational web cloud databases of indoor routes cannot account for unpredictable and complicated indoor daily changes caused by renovations, other humans, machines, accidents, and people who are BSVI themselves. However, unlike other similar in-kind devices and systems, the proposed integral ETA system can provide real-time help in such complex situations. For that matter, in the third modality, when a person who is BSVI encounters a complicated indoor navigation situation such as (a) a deviation from the chosen route, (b) unpredicted obstacles that do not allow the person to pass farther, (c) a missing location ID, etc., the person who is BSVI can make a real-time video call to volunteers for online help to resolve the indoor problem. In this way, using a mobile app, volunteers can obtain almost real-time access to the BSVI user's camera view (see Figures 1 and 17).

However, before calling a volunteer, the ETA system can propose a way back to the last identified location ID when the person who is BSVI is lost. For this reason, the ETA system records the recent multisensory data stream (walking directions and speed, distances of each straight walking segment) to allow it to make dead reckoning instructions back to the last known location ID point. For that matter, machine learning algorithms process the situation, reexamine the route validity, and make propositions to include new location IDs or recognizable objects.

If this method does not help, the person who is BSVI can call a volunteer for real-time help using the third modality's ETA system functionality (see Figure 17). In this way, with the consent of the person who is BSVI, a volunteer can see:

- (a) The current interactive indoor navigational route map stored in the online database;
- (b) the BSVI user's progress on that route;
- (c) passed and next expected views of location ID places;
- (d) third-party information stored a priori regarding the building's floor plan (e.g., escape plan), indoor maps such as OpenStreetMaps, or use other geospatial orientation systems.

This information enables a volunteer to be more informative and better understand the context of the BSVI user's problem. That is, it helps them to see the problem from a grand view perspective. This saves mobile connection time and makes assisting efforts more effective. The ETA system (while a volunteer guides the person who is BSVI) scans visual and sensory data to give feedback and help to navigate the next location ID place or the destiny place. It is important to note that the proposed ETA system can provide a ranked list of volunteers who are most familiar with the place or problem the person who is BSVI is facing.

The above-described web-crowd-assisted social networking support methods employed for indoor guided navigation have been described in detail with particular reference to certain aspects, but it should be understood that variations, combinations, and modifications can be made within the spirit and scope of the presented novel approach.

## 6. Conclusions and Discussion

This paper is dedicated to the R&D implications of the ETA wearable system (prototype) for navigation indoors for people who are BSVI. It emphasizes the novel outsourcing method of mapping and assistance for indoor routes.

The logic and structure of the presented research follows certain steps, which can be narrated in the following way:



1. After the overview of related research papers and patents, we found an evident lack of operationalization of the Web 2.0 social networking advantages for guided navigation indoors (see Introduction). That is, our investigation revealed that the indoor routing process could be crowdsourced to volunteers, avoiding costly infrastructural investment in RFID tags, Wi-Fi, beamers, etc.
2. To define the real social networking abilities and expectations concerning indoor navigation ETA assistance of people who are BSVI, we conducted a semi-structured survey of people who are BSVI and interviews of experts in the field (see Section 2). It clearly shaped and targeted our R&D efforts towards the development of a wearable ETA prototype with some key crowdsourced functions such as an indoor routing process and assistance from volunteers in complex situations.
3. Following the insights mentioned above and the needs of real BSVI users, we are in the process of constructing a unique computational, vision-based, wearable ETA prototype with crowdsourced functions. The presented wearable ETA prototype can be used in three consistent operational modalities: (i) sighted volunteers mark indoor routes using our wearable computational vision-based ETA prototype and help to maintain the web cloud DB of indoor routes; (ii) people who are BSVI can employ our wearable ETA prototype for guided navigation indoors using a chosen route from the route DB; and (iii) people who are BSVI can receive real-time web-crowd assistance (online volunteers' help via the mobile app) in complex situations (such as being lost and encountering unexpected obstacles and situations) using the wearable ETA prototype (see Sections 3–5).

The applied research structure is user-centric and oriented toward actual BSVI needs. In this way, we avoided academic biases towards specific technology-centric approaches. Thus, the presented ETA indoor guided navigational system uses crowdsourcing when volunteers (sighted users) go through buildings and gather step-by-step visual and other sensory information on indoor routes that is processed using machine learning algorithms in the web cloud server and stored for BSVI usage in the web cloud DB.

The proposed novel adaptation of crowdsourcing helps indoor routing services to be obtained from a large group of sighted participants (neighbors, friends, parents, etc.), who voluntarily map indoor routes and place them in the online web server database. Here, indoor routes are processed using computational intelligence methods, enriched with semantic data, rated, and later exploited by people who are BSVI for indoor navigational guidance.

We believe that this integration of crowdsourcing methods using social networking will drive a new R&D frontier that can make more efficient ETA indoor navigational applications for people who are BSVI. Social networking and outsourcing can facilitate the sharing and exchange of experiences with points of interest (POI) (such as stairs, doors, WC, entrances/exits), routes of interest (ROI), and areas of interest (AOI) indoors. This form of social networking could initiate the formation of a self-organized community of people who are BSVI and volunteers.

In this way, participatory Web 2.0 social networking systems can integrate intelligent algorithms with the best experiences of people who are BSVI and sighted people while traveling, navigating, and orientating in indoor environments. This will help to build and continuously update real-time metrics of reachable POI, ROI, and AOI. This will allow routes to be averaged, erroneous routes eliminated, and optimal user-experience-based solutions found using various optimization approaches.

It is important to note that even indoor situations that change daily, such as renovations, furniture movements, closed doors, etc., can be recorded and updated continuously by sighted volunteers using social networking and the web cloud DB. In unrecognized environments, a trained ETA guiding system can either guide the user around the obstacle or suggest another route to continue on next.

Thus, the presented innovative web-crowd-assisted method enables BSVI users to obtain the latest information about the suitability of indoor routes. After experiencing a

route, BSVI users (and correspondingly ETA system) can rate the route's validity and make personal averaged ratings ascribed to the route (that are ascribed to the volunteer who recorded it). This allows other BSVI users to choose the best-rated routes and obtain offline guidance from the best-rated volunteers (sighted users).

In short, from the point of view of the end-user, the presented wearable prototype is distinguished from other related wearable indoor navigational ETA approaches in the sense that (a) it has an intuitive hands-free control interface that uses EMG (or mobile app and panel) and forehead tactile display; (b) it has a comfortable, user-orientated headband design; (c) it provides machine-learning-based real-time guided navigation and recognizes objects, scenes, and faces as well as providing OCR (optical character recognition); and (d) it provides web-crowd assistance while mapping indoor navigational routes and solving complex situations on the way using volunteers' help.

It is important to note that each modality is composed of the same set of eight modes (object detector, specific object detector, scene description, face recognition, OCR, obstacle recognition, navigation, social networking) that are active or work in the background. The inclusion of all of these modes in three functional modalities is a unique feature of the proposed guided ETA system.

Thus, the specific advantages of the proposed experience-centric indoor guided navigation in the sense of user interface and social networking are as follows:

1. People who are BSVI acquire more confidence through human-based wayfinding experiences (through interactions with trusted volunteers or other people who are BSVI with similar needs and preferences) than by using computer-generated models and algorithms.
2. Floor plans or evacuation schemes provided by third parties or volunteers are digitalized, scaled, and matched with the computational vision (SLAM)-generated routes. This helps people who are BSVI when they become involved in complex situations (such as being lost, encountering unexpected obstacles, etc.) and need real-time assistance from volunteers.
3. Indoor situations that change daily can be recorded and uploaded continuously by volunteers through social networking in the web cloud DB.
4. Wayfinding experiences can be effectively rated and shared using social networking.

In time, a participatory web 2.0 social networking platform could emerge for people who are BSVI—something similar to a worldwide “Visiopedia” with a rated, crowdsourced, and publicly available indoor guided navigational web cloud database that is updated in almost real-time. This would expand the set of available indoor routes considerably and enable a much more efficient and reliable rating.

In summary, this paper presented a unique approach to the provision of technological and operational know-how on real-time guided indoor navigation improvements for people who are BSVI that does not require prior expensive investment in Wi-Fi, RFID, beamers, or other indoor infrastructure. The provided insights could help researchers and developers to exploit social Web2.0 and crowdsourcing opportunities for computer vision-based ETA navigation developments for people who are BSVI.

## 7. Patents

Patent application ‘Hands-Free Crowd Sourced Indoor Navigation System and Method for Guiding Blind and Visually Impaired Persons’. Application Number: 17/401,348; Date: 8 August 2021; Number of priority application: US 17/401,348.

**Author Contributions:** Conceptualization, D.P, A.I. and A.L.; methodology, D.P. and A.I.; software, D.P. and A.L.; validation, L.S.; formal analysis, D.P., A.I. and L.S.; investigation, D.P. and A.I.; resources, A.L.; data curation, D.P. and A.I.; writing—original draft preparation, D.P. and A.I.; writing—review and editing, D.P., A.I. and A.L.; visualization, D.P.; supervision, D.P.; project administration, D.P.; funding acquisition, D.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by European Regional Development Fund (project No 01.2.2-LMT-K-718-01-0060) under grant agreement with the Research Council of Lithuania (LMTLT).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

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

## References

- Plikynas, D.; Žvironas, A.; Budrionis, A.; Gudauskis, M. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors* **2020**, *20*, 636. [[CrossRef](#)] [[PubMed](#)]
- Csapó, Á.; Wersényi, G.; Nagy, H.; Stockman, T. A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research. *J. Multimodal User Interfaces* **2015**, *9*, 275–286. [[CrossRef](#)]
- Griffin-Shirley, N.; Banda, D.R.; Ajuwon, P.M.; Cheon, J.; Lee, J.; Park, H.R.; Lyngdoh, S.N. A survey on the use of mobile applications for people who are visually impaired. *J. Vis. Impair. Blind.* **2017**, *111*, 307–323. [[CrossRef](#)]
- Chessa, M.; Noceti, N.; Odone, F.; Solari, F.; Sosa-García, J.; Zini, L. An integrated artificial vision framework for assisting visually impaired users. *Comput. Vis. Image Underst.* **2016**, *149*, 209–228. [[CrossRef](#)]
- Tapu, R.; Mocanu, B.; Bursuc, A.; Zaharia, T. A smartphone-based obstacle detection and classification system for assisting visually impaired people. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2–8 December 2013; pp. 444–451. [[CrossRef](#)]
- Plikynas, D.; Žvironas, A.; Gudauskis, M.; Budrionis, A.; Daniušis, P.; Sliesoraitytė, I. Research advances of indoor navigation for blind people: A brief review of technological instrumentation. *IEEE Instrum. Meas. Mag.* **2020**, *23*, 22–32. [[CrossRef](#)]
- Raufi, B.; Ferati, M.; Zenuni, X.; Ajdari, J.; Ismaili, F. Methods and techniques of adaptive web accessibility for the blind and visually impaired. *Procedia Soc. Behav. Sci.* **2015**, *195*, 1999–2007. [[CrossRef](#)]
- Ferati, M.; Raufi, B.; Kurti, A.; Vogel, B. Accessibility requirements for blind and visually impaired in a regional context: An exploratory study. In Proceedings of the 2014 IEEE 2nd International Workshop on Usability and Accessibility Focused Requirements Engineering (UsARE), Karlskrona, Sweden, 25 August 2014; pp. 13–16.
- Elbes, M.; Al-Fuqaha, A. Design of a social collaboration and precise localization services for the blind and visually impaired. *Procedia Comput. Sci.* **2013**, *21*, 282–291. [[CrossRef](#)]
- Brady, E.L.; Zhong, Y.; Morris, M.R.; Bigham, J.P. Investigating the appropriateness of social network question asking as a resource for blind users. In Proceedings of the Conference on Computer Supported Cooperative Work, San Antonio, TX, USA, 23 February 2013; pp. 1225–1236.
- Brinkley, J.; Tabrizi, N. A desktop usability evaluation of the facebook mobile interface using the jaws screen reader with blind users. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 23 September 2017; SAGE Publications: Los Angeles, CA, USA, 2017; Volume 61, pp. 828–832.
- Wu, S.; Adamic, L.A. Visually impaired users on an online social network. In Proceedings of the Sigchi Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 3133–3142.
- Voykinska, V.; Azenkot, S.; Wu, S.; Leshed, G. How blind people interact with visual content on social networking services. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, CA, USA, 27 February–2 March 2016; pp. 1584–1595.
- Qiu, S.; Hu, J.; Rauterberg, M. Mobile social media for the blind: Preliminary observations. In Proceedings of the International Conference on Enabling Access for Persons with Visual Impairment, Athens, Greece, 12–14 February 2015; pp. 152–156.
- Della Libera, B.; Jurberg, C. Teenagers with visual impairment and new media: A world without barriers. *Br. J. Vis. Impair.* **2017**, *35*, 247–256. [[CrossRef](#)]
- Tjostheim, I.; Solheim, I.; Fuglerud, K.S. The importance of peers for visually impaired users of social media. In (746) *Internet and Multimedia Systems and Applications/747: Human-Computer Interaction*; Norwegian Computing: Oslo, Norway, 2011; p. 21.
- Avila, M.; Wolf, K.; Brock, A.; Henze, N. Remote assistance for blind users in daily life: A survey about be my eyes. In Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments, Corfu Island, Greece, 29 June–1 July 2016; pp. 1–2.
- Watanabe, T.; Matsutani, K.; Adachi, M.; Oki, T.; Miyamoto, R. Feasibility Study of Intersection Detection and Recognition Using a Single Shot Image for Robot Navigation. *J. Image Graph.* **2021**, *9*, 39–44. [[CrossRef](#)]
- Hasegawa, R.; Iwamoto, Y.; Chen, Y.W. Robust Japanese Road Sign Detection and Recognition in Complex Scenes Using Convolutional Neural Networks. *J. Image Graph.* **2020**, *8*, 59–66. [[CrossRef](#)]
- Alkowitz, M.T.; Becerra, V.M.; Holderbaum, W. Estimation of visual motion parameters used for bio-inspired navigation. *J. Image Graph.* **2013**, *1*, 120–124. [[CrossRef](#)]
- Lo, E. Target Detection Algorithms in Hyperspectral Imaging Based on Discriminant Analysis. *J. Image Graph.* **2019**, *7*, 140–144. [[CrossRef](#)]

22. Utamingrum, F.; Prasetya, R.P. Rizdania Combining Multiple Feature for Robust Traffic Sign Detection. *J. Image Graph.* **2020**, *8*, 53–58. [[CrossRef](#)]
23. Likert, R. A Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *22*, 1–55.
24. Sheepy, E.; Salenikovich, S. Technological support for mobility and orientation training: Development of a smartphone navigation aid. In Proceedings of the E-Learn.: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Las Vegas, NV, USA, 21 September 2013; Association for the Advancement of Computing in Education (AACE): San Diego, CA, USA; pp. 975–980.
25. Karlsson, N.; Di Bernardo, E.; Ostrowski, J.; Goncalves, L.; Pirjanian, P.; Munich, M.E. The vSLAM algorithm for robust localization and mapping. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 24–29. [[CrossRef](#)]
26. Karimi, H.A.; Zimmerman, B.; Ozcelik, A.; Roongpiboonsopit, D. SoNavNet: A framework for social navigation networks. In Proceedings of the 2009 International Workshop on Location Based Social Networks, Seattle, WA, USA, 3 November 2009; pp. 81–87.
27. Vitek, S.; Klíma, M.; Husník, L.; Spirk, D. New Possibilities for Blind People Navigation. In Proceedings of the 2011 International Conference on Applied Electronics, Pilsen, Czech Republic, 7–8 September 2011; pp. 1–4.
28. Vinyals, O.; Toshev, A.; Bengio, S.; Erhan, D. Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 652–663. [[CrossRef](#)] [[PubMed](#)]
29. Laganieri, R.; Akhoury, S. Training Binary Descriptors for Improved Robustness and Efficiency in Real-Time Matching. In Proceedings of the International Conference on Image Analysis and Processing, Naples, Italy, 9–13 September 2013; Springer: Berlin/Heidelberg, Germany; pp. 288–298. [[CrossRef](#)]
30. Tafti, A.P.; Baghaie, A.; Assefi, M.; Arabnia, H.R.; Yu, Z.; Peissig, P. OCR as a service: An experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym. In *International Symposium on Visual Computing*; Springer International Publishing: Cham, Switzerland, 2016; pp. 735–746.
31. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.; Tardós, J.D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. [[CrossRef](#)]