SLAM for Visually Impaired People: A Survey

MARZIYEH BAMDAD, University of Zurich, Department of Informatics, Switzerland and Zurich University of Applied Sciences, Institute of Applied Information Technology, Switzerland

DAVIDE SCARAMUZZA, University of Zurich, Department of Informatics, Robotics and Perception Group, Switzerland

ALIREZA DARVISHY, Zurich University of Applied Sciences, Institute of Applied Information Technology, Switzerland

In recent decades, several assistive technologies for visually impaired and blind (VIB) people have been developed to improve their ability to navigate independently and safely. At the same time, simultaneous localization and mapping (SLAM) techniques have become sufficiently robust and efficient to be adopted in the development of assistive technologies. In this paper, we first report the results of an anonymous survey conducted with VIB people to understand their experience and needs; we focus on digital assistive technologies that help them with indoor and outdoor navigation. Then, we present a literature review of assistive technologies based on SLAM. We discuss proposed approaches and indicate their pros and cons. We conclude by presenting future opportunities and challenges in this domain.

Additional Key Words and Phrases: SLAM, visually impaired, navigation

ACM Reference Format:

Marziyeh Bamdad, Davide Scaramuzza, and Alireza Darvishy. 2022. SLAM for Visually Impaired People: A Survey. 1, 1 (December 2022), 26 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

One of the topics that many researchers have widely addressed in the past decades is developing assistive technology as a substitute for the sight of visually impaired people. The main goal of these studies is to guide and assist VIB people in navigating safely in unknown environments without the help of a sighted assistant. Navigation requires finding an optimal path to the desired destination, perceiving surroundings, and avoiding obstacles. All these functionalities need to localize the VIB user accurately in the environment. There are several approaches for localization such as global positioning system (GPS), radio frequency identification (RFID), and simultaneous localization and mapping (SLAM) [Alkendi et al. 2021; Panigrahi and Bisoy 2021]. Each has advantages and challenges and is used in different applications.

Authors' addresses: Marziyeh Bamdad, bamdad@ifi.uzh.ch, University of Zurich, Department of Informatics, Binzmuehlestrasse 14, Zurich, Zurich, Switzerland, 8050 and Zurich University of Applied Sciences, Institute of Applied Information Technology, Steinberggasse 13, Winterthur, Zurich, Switzerland, 8400; Davide Scaramuzza, sdavide@ifi.uzh.ch, University of Zurich, Department of Informatics, Robotics and Perception Group, Andreasstrasse 15, Zurich, Zurich, Switzerland, 8050; Alireza Darvishy, dvya@zhaw.ch, Zurich University of Applied Sciences, Institute of Applied Information Technology, Steinberggasse 13, Winterthur, Zurich, Switzerland, 8400.

Manuscript submitted to ACM

GPS is a localization technique employed in outdoor scenarios due to its affordability to the end user, wide coverage of the Earth, and easy integration with other technologies. However, this technique suffers from some restrictions that make it challenging in some scenarios. Satellite signal blockage, inaccuracy, and line-of-sight restrictions in places affected by walls and other obstacles and being affected by the weather condition are among these limitations [Alkendi et al. 2021].

Approaches based on RFID utilize small low-cost tags for localization. To localize an agent, a set of RFID tags should be installed in the environment [Panigrahi and Bisoy 2021]. Although localization can be accurately performed by an RFID scheme, taking advantage of this technique requires pre-installed infrastructure.

In scenarios where infrastructure-based solutions such as RFID or GPS do not provide adequate accuracy or are not available, SLAM techniques are used to provide reliable localization. SLAM is an innovative technique that consists of simultaneously constructing an environment model (map) and estimating the state of an agent moving within it [Cadena et al. 2016]. SLAM uses diverse types of sensors to determine the position, location, and velocity of an agent, as well as to detect and avoid obstacles even in a dynamically changing unknown environment. This technique widely uses infrared (IR), acoustic sensors, cameras, inertial measurement units (IMUs), ultrawide-band (UWB), LiDAR, RADAR, and RGB-D sensors [Khan et al. 2022b].

The SLAM community has made tremendous strides over the past 35 years, developing large-scale practical applications, and seeing a steady transition of this technology into the industry [Cadena et al. 2016]. Cadena et al. [Cadena et al. 2016] have divided the life cycle of SLAM into three periods: the classical age (1986-2004), the algorithmic-analysis age (2004–2015), and the robust-perception age (2015-present).

The evolution in portable computation and the availability of low-cost, highly accurate, and lightweight sensors such as cameras and IMUs made them appropriate for pedestrian navigation. By exploiting these advances, many researchers have recently adopted SLAM to develop assistive technology demonstrators to help VIB people navigate in unknown environments.

Since the first electronic travel aids (ETAs) emerged around 70 years ago, the development of navigation devices to guide VIB people through indoor and/or outdoor environments has remained a challenge and a key concern for researchers [Real and Araujo 2019].

From the traditional to deep-learning-based navigation approaches, researchers have always faced challenges ranging from technical issues to the limitations of users' capabilities. As VIB navigation approaches must improve real-time performance while reducing the size, weight, energy cost, and overall price of the assistive system, these works have put a lot of effort into coping with constraints in computational issues, sensory equipment, and portable devices. They also need to provide solutions to calculate the precise position and orientation of the user in a real-time manner. On the other hand, the challenges of different scenarios including complex and cluttered environments, noisy environments, and large spaces need be taken into consideration.

Furthermore, the problem of efficient and reliable obstacle detection both in indoor and outdoor environments has always been a concern. In this regard, other challenges include identifying static and dynamic obstacles, predicting the risk of collision, understanding moving objects' motion and estimating their speed, detecting small objects, and identifying obstacles in different levels of the user's body from drops in terrain to head-level. In addition, an intuitive, user-friendly, low-cognitive load method to provide accurate and sufficient environmental information to the user is also considered an important research target. These methods should be improved to provide adjustable and customized feedback on demand for different users.

Moreover, assistive technology ought to provide user safety and independence, hands-free operations, decreased effort, and backup in case of system failure. In addition to the aforementioned challenges, deep-learning-base solutions have also special issues such as designing lightweight neural network architecture to reduce computational expense and providing enough data for the training and validation of the models.

So far, many reviews have been carried out on assistive technologies developed for VIB navigation. Several papers have reviewed walking assistance systems [Fernandes et al. 2019; Islam et al. 2019; Khan et al. 2021; Manjari et al. 2020; Real and Araujo 2019; Romlay et al. 2021; Zhang et al. 2021b] and provided a detailed classification of developed approaches. Islam M. et al. [Islam et al. 2019] categorized walking assistants into three groups: sensor-based, computer vision-based, and smartphone-based. The authors explained the technologies used and inspected each approach and evaluated some important parameters of each approach such as type of capturing device, type of feedback, working area, cost, and weight.

Some studies focused on indoor navigation for VIB users [Façanha et al. 2020; Khan et al. 2022a; Plikynas et al. 2020a,b; Simões et al. 2020; Wang et al. 2021; Zvironas et al. 2019] and some focused on computer-vision–based navigation systems [Fei et al. 2017; Jafri et al. 2014; Khan et al. 2022a; Sivan and Darsan 2016; Valipoor and de Antonio 2022; Walle et al. 2022]. Among these papers, [Khan et al. 2022a] conducted a systematic literature review on state-of-the-art computer vision-based methods used for indoor navigation. The authors described the advantages and limitations of each solution under review and included a brief description of each method.

A number of review papers on wearable navigation systems have also been published [Chaudhary et al. 2019; Dakopoulos and Bourbakis 2009; Tapu et al. 2020]. The review given by [Tapu et al. 2020] provides a comprehensive review of computer vision and machine learning-based assistive methods. The paper divides the existing ETAs into two groups: active systems providing subject localization and object identification, and passive systems providing information about the users' surroundings using a stereo camera, monocular camera, or RGB-D camera.

Other similar studies include a survey of Inertial Measurement Units (IMUs) in assistive technologies for visually impaired people [Reyes Leiva et al. 2021], a review of urban navigation for VIB people [El-Taher et al. 2021], a survey paper that reviewed assistive tools based on white canes [Motta et al. 2018], and review papers exploring smartphone-based navigation devices [Budrionis et al. 2022; Khan and Khusro 2021].

To the best of our knowledge, there is no survey paper on SLAM-based navigation systems for VIB people. Our paper aims to bridge this gap. Section 2 presents the results of a survey that we conducted with VIB people to understand their experience and needs in indoor and outdoor navigation. Section 3 reviews SLAM-based solutions for VIB individuals. Section 4 presents their respective advantages and limitations. Finally, Section 5 proposes future opportunities and concludes the paper.

2 SURVEY ON VIB NAVIGATION EXPERIENCE AND NEEDS

Investigating and understanding the navigational needs and preferences of VIB individuals in real-life scenarios is essential to develop appropriate assistive technology to help them navigate safely and independently in unfamiliar environments. Occasional use of existing technologies by VIB people does not mean they do not need or are not willing to use these aids, it can indicate that there is a gap between their needs and preferences and the technologies that have been developed for this purpose. Therefore, to get an overview of the needs, experiences, and challenges for VIB navigation, we designed and implemented a fully accessible online questionnaire and distributed it to the visually impaired and blind individuals around the world through 76 persons and organizations.

The questionnaire was structured in three sections with 17 close-ended and 2 open-ended questions. In the first section, we asked about personal and sight-related information. In the second section, we asked about the participants' personal experience with the assistive technologies they use for navigation. In the third section, we asked about their needs and their preferences for assistive technologies for navigation. We also asked them to share their views about ideal assistive technologies.

Our audiences are people with vision impairment severity from mild to total blindness in the age range of 15 to 60+ years. Among them, 25 are male, 16 are female and 1 is non-binary. 18 of the respondents were born with a visual impairment, while 24 lost their visual acuity later. Participants' information is shown in Table 1

Table 1. Participants information

Feature	Number of participants
Region of residence	Africa: 11, Asia-Pacific: 4, Eastern Europe: 9, Latin America and Caribbean:1, Western Europe and North America : 17
Age range	15-30: 9, 30-45: 14, 45-60: 14, 60+: 5
Gender	male: 25, female: 16, non-binary: 1
Born with a visual impairment	yes: 18, no: 24
Vision impairment severity	mild: 4, moderate: 6, severe:17, total blindness: 15

Of the participants, 66.7% use navigation apps on a smartphone as a digital assistive technology. Two thirds of them are satisfied or very satisfied with this kind of technology. Almost all these users use smartphone applications for outdoor navigation. However, as one of participants reported, using a smartphone is not a suitable solution for people with physical disabilities such as polio. Two participants use smart cane, and two others only use white cane, while close to 24% of participants don't use any digital assistive technology. High cost, lack of availability, assistance from sighted people, and use of personal functional vision have been reported as the reasons. Close to 19% of participants who use assistive technologies for navigation, mainly smartphone apps, reported feedback provided by assistive technologies is enough to guide them in an unfamiliar place. But most of them considered the amount of feedback insufficient for places with many obstacles and outdoor places. Also, some of them have reported lack of adequate feedback in noisy places. Among the participants who use assistive technologies, 93.8% use them for outdoor or both outdoor and indoor navigation. This shows the importance of outdoor navigation for VIB people. However, more than 70% of the papers under our review have proposed solutions for indoor scenarios.

To study the results of this survey and figure out the participants' needs and preferences, we analyzed their responses from four viewpoints: assistive technology characteristics, environment in which assistive technology is used, routing and collision avoidance, and feedback.

Assistive technology characteristics. The two most important characteristics of assistive technology reported by participants are ease of use and low cost. Next in order, the importance of low mental load, low weight of the device, identification and avoidance of obstacles, and low physical effort are mentioned almost equally. 6.5% of the respondents also mentioned receiving remote help as an important feature. These people are of different ages and genders. But apart from one of them, the rest have severe or total blindness vision impairment severity.

Interestingly, 61.9% of our sample prefer to use wearable assistive technology, 28.6% handheld assistive technology, and the remainder a white cane for navigation. These results show that, although wearable assistive technologies are most desired, some people still prefer to rely on their white cane as a part of an assistive device. Additionally, except for one respondent with mild visual impairment, the rest of the participants who have desire for handheld assistive technologies have severe or total blindness visual impairment. Table 3 shows that the above mentioned preferences have been addressed in several works. As this table demonstrates, half of the papers under our review proposed wearable assistive technology, and 9 out of 34 papers used canes as the base of their solutions.

Environment. As Figure 1 depicts, navigation in an unknown environment is challenging for VIB people in different aspects. Finding the objects and places in outdoor environments and finding the way to desired destination are the most challenging tasks for the participants in the survey when exploring an unfamiliar environment. Finding the objects and places in indoor environments and detecting the objects and obstacles ahead are among the other challenges reported in this regard. Other challenges that have been specifically described by respondents that can be considered as open challenges include:

- Finding the way around in large areas indoors such as big hallways and outdoors such as large avenues or squares. In these situations, the lack of auditory and tactile clues is mentioned as the most challenging.
- Differentiating obstacles in a dimly lit environment. Despite the importance of this problem, it has not been considered in the papers under our review. The participant also experiences difficulties tracing his way, especially for small paths at night.
- Managing to walk through vehicles when crossing streets and walking in crowded spaces. Although several papers focused on dealing with this problem and provided solutions such as [Son and Weiland 2022], it still remains a main challenge for VIB people navigation.
- Finding out the number of stairs to reach the destination while going up or down the staircase. In several papers such as [Katzschmann et al. 2018], the problem of detecting the first step of the staircase has been investigated, but according to the need stated by one of our survey participants, information about the number of stairs can be considered as a main challenge. Additionally, the size of the stairs can also be regarded as another challenge. Since all stairs are not the same size and some are much bigger than usual and one should walk each of them with two or three steps.
- Avoiding collision with pedestrians. Although this issue has been considered and investigated by some works such as [Kayukawa et al. 2020], it seems that the result of these studies has not yet been exploited for real-life scenarios.

Routing and collision avoidance. Obstacles in head level is reported as the most challenging. Meanwhile, the challenges of chest-; leg-; and ground-level obstacles, and drop-offs are reported almost equally by the participants. 7 out of 42 participants have declared the obstacle avoidance challenge at all levels. The vision impairment severity of all these people was severe or total blindness. It can be concluded that it is important to pay attention to obstacles at all levels. Since dealing with all these challenges at the same time complicates the assistive approach and causes new challenges, the importance of system customization is revealed again. Additionally, for choosing a path when traveling from one place to another, the participants emphasized routing through an optimal path as obstacle-free as possible. The options that are in the next priorities include:

(1) Providing route information
(2) Providing constant assistance for navigating from current user location to the desired destination
(3) Routing through the shortest distance
(4) Re-routing quickly from current user position to the desired destination once an obstacle has been detected in the suggested path

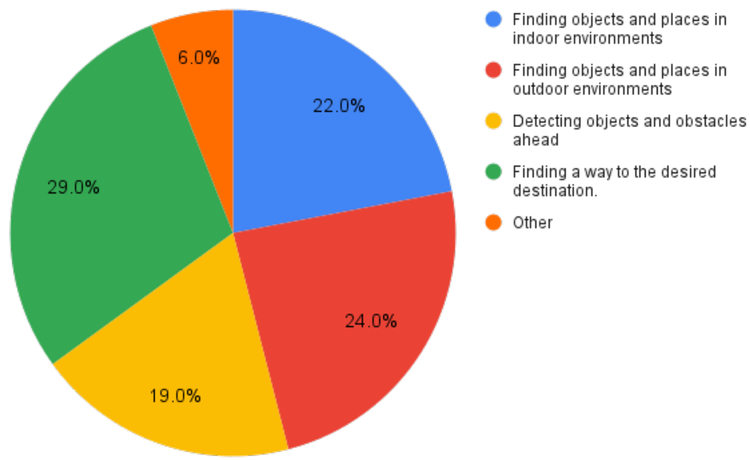

Fig. 1. Most challenging for VIB people when exploring an unfamiliar environment.

(5) Routing through a path with the fastest traveling time
(6) Routing through a path with low cost

Feedback. Over 71% of participants stated speech feedback as a preference to receive information about their surroundings. Among them, close to 31% wished to combine speech with other types of feedback such as tactile. 9.5% of the respondents with severe vision impairment chose beeping sound, and one of the participants mentioned beep sound must be enough loud to alert him. It can be seen that the amount and the intensity of the feedback should be adjusted by the target users according to their physical conditions and the environment in which they navigate. Of the respondents, 19% regarded tactile (vibration) as proper feedback. One of our audiences stated tactile feedback is somewhat limited unless users learn a code relevant to the feedback. He mentioned less complex information is fine through vibration. Another participant reported her preference for speech feedback for directions and haptic/vibrations for obstacles avoidance. In the case of tactile feedback, most respondents prefer to receive feedback on their arms and hands. Getting feedback on wrists, shoulders, forehead, feet, and waist are the next priorities. Only one of the participants reported the willingness to receive feedback on his legs. This participant reported does not matter on which part of the body he receives feedback.

Table 2 presents participants' responses to open-ended questions about the ideal assistive technology for navigation. Based on these comments we can conclude that, due to the variety of needs and preferences, the ideal system would be a system that can be adjustable and customized for each group of users according to the vision impairment severity, the environment in which they navigate, the user's physical health, etc. The participants stated that an ideal assistive technology is affordable, lightweight, easy to use, and in the form of a smartwatch or smartphone. However, some of them preferred the assistive technology to be in combination with a cane. Respondents in low- to middle-income countries emphasized that these technologies should be commercially available and applicable in the informal settlements. Participants preferred receiving surroundings information on demand during navigation.

Based on the classification of the World Health Organization, distance vision impairment is divided into four classes: 1) mild, 2) moderate, 3) severe, and 4) blindness. Results from our survey confirm each class of visual impairment should

Table 2. Participants' responses to the open-ended survey questions

Features of an ideal assistive technology for navigation

- Being able to navigate informal settlements and slums
- Providing surroundings information on demand during navigation
- Providing distance and nature of the obstacles ahead
- Affordable, easy to use, unobtrusive, light, small, intuitive, accurate, low-tech, solar powered, and off-line
- A watch or a phone
- Smart and if it has buttons, the buttons should have different colors
- In combination with a cane
- Based on mainstream products and services not a bespoke device
- Performing one function perfectly, rather than multiple functions that would be reached approximately
- Commercially available
- Smart glasses that integrate with the smart watch
- A robot that would go at guide dog speed, detect obstacles, and describe the environment on demand

Additional comments or suggestions

- The navigation aids should be designed keeping in mind the low to middle-income countries where access to government funding for the acquisition of assistive devices is either non-existent or limited.
- Efforts need to be made to ensure the promotion of open-source technology for further improvement in the future.
- The device should be low-cost for developing countries like Fiji as it is not available here.
- Technological advancement is a little slower in Africa compared to Europe and other parts of the world. Our physical environment is not well plotted/planned - which makes it difficult for some of the solutions to work efficiently.
- You did not cater to ordinary white canes, yet these are the ones blind people have used for years.
- Most smart canes end up becoming another distraction from what the cane already tells us. Focus on enhancing rather than duplicating that feedback.
- When creating assistive devices for navigation for the blind, not well-structured environments, as we have in most places in Africa should be put under serious consideration. Otherwise, they will be useless to most Africans with visual impairment.
- My guide dog is the best guide for my security outdoors.
- I am extremely interested in smart white cane.
- In combination with bone-conduction headphones, I should be able to use my smartphone as a navigation device.
- It would be great if any device could be operated easily with one hand or voice control.

have its own appropriate aid. In other words, a unique solution for all visually impaired users will not satisfy their needs. This is a point that was clearly stated by one of our survey participants. She emphasized not to develop an assistive technology that does everything, instead focus on the one function that the technology would run perfectly, rather than on multiple functions that would be reached approximately. Therefore, developers need to specify precisely which category of VIB people they want to provide a solution for. This affects the choice of sensors, the type and amount of feedback, and the type of assistive devices.

From this survey we also found that the prior visual experience of the people who were not born with visual impairment can help them to perceive the environment different than VIB people who were born with a visual impairment. Since the questionnaire was distributed from a limited number of channels and only the people who are familiar with English were the audiences of this questionnaire, our survey was based on a small sample of target audience, and not the entire VIB population. However, the result can be useful in various aspects of developing assistive technologies for the VIB user navigation and future research.

3 SLAM-BASED SOLUTION FOR VISUALLY IMPAIRED AND BLIND PEOPLE'S NAVIGATION

Figure 2 depicts the number of research papers (included in this review) that have focused on VIB navigation using the SLAM technique in recent years. The growth in the number of papers in this domain suggests an advancement of SLAM techniques and an increase in their usage for developing navigation technologies for VIB people. This section provides a review of proposed SLAM-based systems based on the three main components of these assistive systems: sensors used to acquire environmental information, computation units to process data and run navigation algorithms, and Human-Computer Interaction (HCI) for user feedback. Table 3 summarizes SLAM-based systems included in this review.

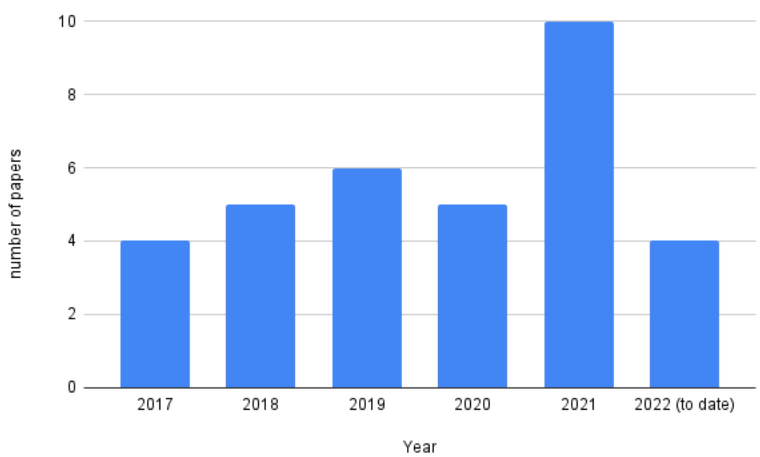

Fig. 2. Publications included in this review on SLAM-based VIB navigation, by year

3.1 Sensors

SLAM architecture (Figure 3) consists of a front-end and back-end. The front-end receives environmental information from sensors, abstracts them into amenable models for estimation, and sends them to the back-end for mapping, localization, and data fusion [Cadena et al. 2016]. Sensors are electronic devices used to detect changes in an environment, such as rotation, motion, light, and distance. There are various sensors used for human navigation systems, including vision sensors (e.g. depth camera and normal camera), distance sensors (e.g. LiDAR, ultrasonic), position sensors (GPS, encoder, Bluetooth beacons), and inertial measurement units (IMUs). Since pre-processing in the front-end depends on the information obtained from the sensors, they are very important for SLAM systems. The choice of sensor type for

Table 3. Summary of SLAM-based systems included in this review

Area	Ref.	Sensor(s)	HCI	Computation unit	Assistive tool	Carrying mode
Indoor	[Li et al. 2022]	RGB-D camera, Ultrasonic sensor	Audio, haptic	Embedded computer, remote server	Electronic glasses	Wearable
	[Plikynas et al. 2022]	ToF and RGB camera, light detector, IMU, compass sensors	Audio, tactile	Raspberry Pi4, Cloud server	Headband	Wearable
	[Ou et al. 2022]	RGB-D camera	Acoustic	Laptop	Smart glasses	Wearable
	[Lu et al. 2021b]	Camera	Audio	GPU cloud	Google glasses	Wearable
	[Xu et al. 2021]	RGB-D camera, Ultrasonic sensor	Audio, haptic	Embeded computer, Remote server	smart glasses	Wearable
	[Lu et al. 2021a]	LiDAR, RGB-D camera	Audio, haptic	Intel NUC, NVIDIA Jetson	Guiding robot	Handheld
	[Hakim and Fadhil 2021]	RGB-D camera, ultrasonic sensor	Audio	Raspberry Pi	Head mounted	Wearable
	[Liu et al. 2021]	OptiTrack cameras	Audio, tactile	PC server	Tactile compass	Handheld
	[Zhang et al. 2021a]	RGB-D camera, IMU	Audio, tactile	UP board computer	Cane	Handheld
	[Zhang and Ye 2020]	3D ToF camera, IMU	Audio	UP board computer	Smart cane	Handheld
	[Jin et al. 2020]	Monocular cameras	Audio	Not mentioned	Smart glasses	Wearable
	[Zhao et al. 2019]	ZED camera	Audio	Laptop	Wearable sensors	Wearable
	[O'Reilly et al. 2019]	Microphones	No device	Not mentioned	No device	No device
	[Zhang and Ye 2019]	RGB-D camera, IMU	Audio	UP board computer	Cane	Handheld
	[Zhang et al. 2019]	Smartphone sensors	Audio, haptic	Smartphone	Gloves	Wearable
	[Li et al. 2018]	Camera	Audio, haptic	Google Tango tablet	SmartCane	Handheld
	[Eden et al. 2018]	Smartphone and ZED camera	No device	Nvidia Jetson	No devie	No device
	[Yun et al. 2018]	LRF, camera, IMU	Steering by carrier robot	PC	Wheelchair	Person Carrier Robot
	[Hakim and Fadhil 2019]	LiDAR, ultrasonic sensor, Raspberry pi camera	Audio	Raspberry Pi	Cane	Handheld
	[Ramesh et al. 2018]	Monocular camera	Audio	Remote server	No device	No device
	[Bai et al. 2018]	Fisheye and depth camera, ultrasonic sensor	Audio or/and visual cues	CPU 1.44 GHz	Glasses	Wearable
	[Zhang and Ye 2017]	Camera	Audio	Tablet, laptop	Smart cane	Handheld
	[Chen et al. 2017]	Motion tracking camera, depth sensor	Audio, haptic	Lenovo Phab2	Smart cane	Handheld
	[Endo et al. 2017]	Wearable camera	No device	Not mentioned	No device	Wearable sensors
Outdoor or both in- & outdoor	[Son and Weiland 2022]	RGB-D camera, IMU	Audio	Jetson computer	Headband	Wearable
	[Rui et al. 2021]	RGB-D camera	Audio, tactile	Cloud server	Smart cane	Handheld
	[Slade et al. 2021]	2D LIDAR, monocular camera, GPS, IMU	Grounded kinesthetic, Audio, tactile	Raspberry Pi	Augmented cane	Handheld
	[Cheng et al. 2021]	RGB-D-IR camera, IMU	No feedback	Various platform	Auxiliary glasses	Wearable
	[Chen et al. 2021]	RGB-D camera	Audio	Cloud server	Head mounted	Wearable
	[Kayukawa et al. 2020]	RGB-D camera, IMU, LiDAR sensor	Audio, tactile	Laptop	Suitcase	Handheld
	[Duh et al. 2020]	Monocular camera	Audio	Laptop	Wearable camera, a smartphone	Wearable
	[Weiss et al. 2020]	Stereo camera	No device	Not mentioned	No device	No device
	[Bai et al. 2019]	RGB-D camera, IMU	Audio, tactile	Smartphone	Eyeglasses	Wearable
	[Bai et al. 2017]	Stereo cameras, microphone	Audio	Cloud server	Helmet	Wearable

human navigation systems depends on several factors, such as working area (indoor and/or outdoor), other sensors used, carrying mode (wearable or handheld), assistive services, time (day and/or night), reliability, range, latency, cost, size, weight, and many more. For indoor mapping and navigation, O'Reilly et al. [O'Reilly et al. 2019] proposed a novel

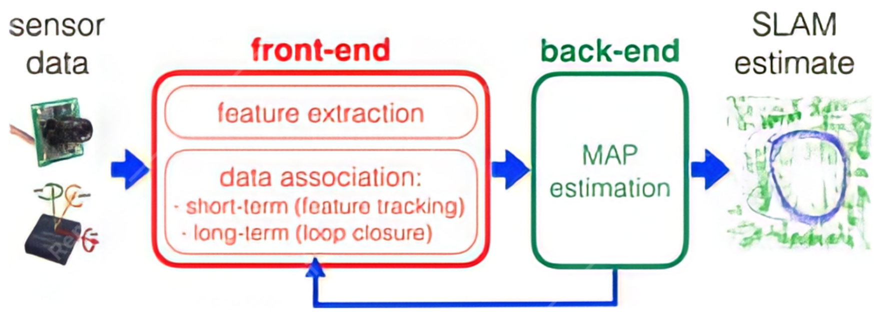

Fig. 3. Front- and back-end in a typical SLAM system [Cadena et al. 2016]

approach using an acoustic scene mapping technique. This method is used to represent the positional surrounding sound source information. This information is provided only by direction-of-arrival estimates of source directions [Evers and Naylor 2018]. The authors used adapted microphones as key sensors for accumulating information indoors, and mapped the environment before using it with the SLAM algorithm. Although this idea can be used for 3D scene mapping in low light conditions, and without the need for other types of sensors, this method can be still affected by environmental noise, which can cause errors in estimating source location.

With the exception of [O'Reilly et al. 2019] mentioned above, all papers under review adapted different types of cameras to receive surrounding information. Most recent works used RGB-D cameras as capturing devices. RGB-D is a sensing system with a color sensor and depth sensor that captures RGB images as well as pixel-wise depth information, used to identify objects in the environment and estimate their distance to the camera, respectively [Plikynas et al. 2020a; Tang et al. 2016]. Advantages such as high accuracy, high resolution, low cost, and light weight have made this type of sensor appropriate for human navigation. The weak points of these sensors are their limited measurement ranges and the increase in the depth measurement error with increasing distance from the camera. In addition, this type of sensor has a limited field of view [Tang et al. 2016].

In an effort to compensate for these limitations, [Rui et al. 2021] and [Kayukawa et al. 2020] employed dual RGB-D camera systems in their designs. The former used two cameras for local obstacle avoidance to identify and avoid obstacles to the front, left, right, above, and below the user, which provides safe real-time navigation. The latter utilized two RGB-D cameras mounted on a rolling suitcase to widen the field of view and detect and track surrounding pedestrians. In addition to the cameras, the rolling suitcase carries an IMU and LiDAR to be able to locate the user's position, determine the user's speed, and predict the risk of collision with pedestrians. [Li et al. 2022] and [Hakim and Fadhil 2021] used ultrasonic sensors in addition to RGB-D cameras. Ultrasonic sensors [Zhmud et al. 2018] are appropriate for perceiving surrounding objects (detecting obstacles and determining distance). An ultrasonic sensor is made up of a transmitter that emits ultrasonic waves and a receiver that perceives the echo. It measures the time it takes for a signal to travel the distance from the transmitter to the receiver. Since ultrasonic signals are not affected by sunlight or object color, they can be a good option in different illumination conditions. Although these sensors are capable of detecting even transparent and glassy objects, they have difficulty determining the distance to small objects.

The angle of wave radiation to objects is also important. If the ultrasonic sensor is perpendicular to an object, the distance measurement can be done accurately. But if this angle is so wide that the wave reflected from the objects does not reach the receiver, the distance is not measured correctly. Li et al. [Li et al. 2022] proposed a wearable cognitive assistance system for indoor navigation. The authors obtained 3D environmental information using a depth map and color images of the scene from a RealSense D435i depth camera at a rate of 30 frames/s. This information was used for navigation by applying the ORB-SLAM2 algorithm and multi-target recognition by applying artificial intelligence algorithms . They also adopted an ultrasonic module attached to the user's leg to help in detecting front obstacles in the range of one meter and more than 10 cm in height.

Some proposed solutions used an Inertial Measurement Unit (IMU) in addition to RGB-D cameras to develop Visual-Inertial Navigation Systems (VINS) for VIB people. VINS achieve low-cost high tracking accuracy, and offer one of the best solutions for localization and navigation on restricted platforms [Liu et al. 2020]. An IMU [Ahmad et al. 2013] is an electronic device consisting of motion and rotation sensors to measure inertial acceleration, angular rotation, and sometimes magnetic field using 3-axis accelerometers, 3-axis gyroscopes, and 3-axis magnetometers. An IMU with accelerometers and gyroscopes is considered a 6-axis IMU, while an IMU including magnetometers is considered a 9-axis IMU. Son & Weiland [Son and Weiland 2022] integrated an IMU into a wearable system to guide VIB users at a signalized crosswalk. This system works on the Robot Operating System (ROS) and is composed of five nodes (one sensing node, two perception nodes and two planning & feedback nodes). The sensing node sends color and depth images captured by a Realsense D435 camera, as well as 6-axis IMU values obtained by a BNOO55 compass sensor, to the perception node in order to run scene understanding and global pose estimation algorithms.

Bai et al. [Bai et al. 2019] proposed a similar wearable prototype for VIB navigation and obstacle recognition in both indoor and outdoor settings. For data acquisition, the authors used an RGB-D camera and IMU mounted on eyeglasses. The depth and color images captured by the Intel RealSense D435 camera are used for navigation and recognition tasks, while the IMU sensor, mounted on top of the camera, obtains the attitude angle of the camera.

For real-time localization used in wayfinding algorithms, Zhang et al. [Zhang et al. 2021a] designed a Robotic Navigation Aid (RNA) prototype composed of an RGB-D camera and an IMU installed on a white cane. The authors presented a new Visual-Inertial Odometry (VIO) method for 6-DoF RNA pose estimation with high accuracy in an architectural floor plan (an architectural CAD drawing). To do so, they first extracted the geometric features such as walls, doors, junctions, and corners from the camera's depth data, and then tightly coupled this with visual features and IMU values in a graph optimization framework. In another publication [Zhang and Ye 2020], the same authors presented a similar work with a time-of-flight (ToF) camera (for 3D perception) and IMU (for motion measurement). The proposed navigation system provides both wayfinding and obstacle avoidance functionalities. The authors introduced a method called plane-aided visual-inertial odometry (PAVIO), which extracts plane features from the camera's 3D point cloud, tracks them over the camera's data frames, and associates these features between frames.

A ToF camera delivers 3D image information and has the potential to capture both intensity (to detect objects) and depth (to estimate distance) information simultaneously [Jeon et al. 2016; Katzschmann et al. 2018]. This camera is small in size and requires relatively low computational complexity to process data received from the environment. The travel time of the infrared light emitted by the IR ToF sensor is used to measure the distance between the object and the ToF camera. The data provided by this camera is appropriate for creating a 3D point cloud map used in pose estimation problems and object recognition algorithms. Attributes such as simplicity, affordability, ease of use, high accuracy, and high frame rate have made this sensor a viable option for use in human navigation systems.

Plikynas et al. [Plikynas et al. 2022] benefited from the ToF camera in combination with an RGB camera and IMU to provide an indoor routing process that is crowdsourced to volunteers. Multi-sensory environmental data is sent to a web cloud server to be processed by machine learning algorithms for the required navigation services. In the navigation procedure, the ToF camera is used as an "active sensor" (i.e. using an external energy source) , emitting IR light and providing depth information for obstacle detection. RGB images captured by the RGB camera as a "passive sensor" (i.e. receiving energy from the environment) are employed for object detection tasks. An object detection algorithm detects a set of trained object classes in these images. Additionally, these images can be used for scene description and face recognition assistive services. In this solution, sighted volunteers use a mobile application to mark indoor routes and manually perform semantic tagging of indoor landmarks such as doors, elevators, and corridors. This information is sent to the web cloud server and stored in an online database. The VIB user then uses web cloud database routes to navigate in that specific building.

One of the lesser-used sensor types among the reviewed papers was the ZED camera. A ZED camera is a binocular vision system that captures two images at the same time and can be used to provide a 3D perception (i.e. the depth of objects) within a range of 1 to 20 meters and at 100 FPS [Ortiz et al. 2018]. Compared to other stereo cameras, this camera has a smaller size. ZED stereo cameras can be used in assistive devices for VIB navigation, as presented by Zhao et al. [Zhao et al. 2019]: With the aim of enhancing VIB corridor navigation, the authors obtained information on indoor surroundings via a ZED camera for obstacle avoidance and path planning tasks. Using the proposed prototype, users were able to identify specific points of interest in the corridor such as offices, staircases, toilets, and exits.

As the abovementioned review shows, in the development of navigation systems for VIB people, camera-based solutions have received by far the most attention from the academic community and developers. Different types of cameras used in such approaches have advantages such as high accuracy, high resolution, low cost, light weight, low size, and low power consumption. Even so, vision sensors have drawbacks that originate from the sensitivity of these sensors to light changes, environments with low texture, or the need for high computational power for image processing [Debeunne and Vivet 2020]. Such drawbacks can be overcome by combining them with other sensors or by optimizing the algorithms that use sensor data.

3.2 Computational unit

A SLAM-based navigation assistive system dedicated to VIB people provides a variety of functionalities such as mapping, localization, path planning, and obstacle detection. To process data and run navigation algorithms, the reviewed papers adopted two classes of computational resources: local and remote. Local calculations are performed in situ on devices such as smartphones, tablets, laptops, portable microcontrollers, UP board computers, and in some cases in the initial development stage, navigation algorithms are carried out on PCs [Liu et al. 2021; Yun et al. 2018]. Smartphones are widely used as communication gadgets and their technology is continuing to grow, to the point that it is possible for smartphones to implement functional navigation systems. Since smartphones integrate diverse sensors such as IMU, GPS, and cameras, they can be used as a convenient tool for collecting environmental information. Additionally, their computational power can be exploited to perform various navigation operations. For example, the system proposed by [Bai et al. 2019] implemented all algorithms relevant to data acquisition, ground segmentation, moving direction search, global path planning, indoor and outdoor localization, and object detection on a smartphone and achieved real-time performance. Without an additional depth sensor, [Zhang et al. 2019] took advantage of an ARCore-supported smartphone to track pose and to build a map of the surroundings in real time. However, despite the significant advantages of smartphones such as small size, low weight, easy portability, and low cost, their computing power is not enough for some approaches.

Some reviewed approaches [Zhao et al. 2019] and [Zhang and Ye 2017] perform all or part of the required calculations locally on a portable computer such as a laptop. Despite higher computing power compared to a smartphone, and greater security compared to remote computational resources, the laptop's heavier weight and large size are considered major disadvantages, especially during long trips.

An alternative solution is to transfer all or part of the calculations to remote computing resources. [Li et al. 2022] utilized remote servers and [Bai et al. 2017; Chen et al. 2021; Lu et al. 2021b; Plikynas et al. 2022] benefited from cloud servers. In order to reduce local computing costs, [Li et al. 2022] adopted an embedded computer as well as a remote server. In the proposed vision-based assistance system, before transferring input images to the server, the images were time-stamped and encrypted on the embedded computer. The remote server was equipped with CPU as well as GPU in order to run parallel ORB-SLAM2 and artificial intelligence algorithms for indoor navigation and object detection, face recognition, and scene text recognition . The experiments conducted by the above-listed authors confirmed that the use of remote servers under a smooth network connection such as 4G or WiFi can meet the computational needs of the presented system. However, although the high computing power of remote servers is considered a significant advantage, constant Internet access over a secure connection is required. Moreover, the performance of the entire system would be affected by the network condition.

3.3 Human-Computer interaction for feedback

VIB people need alternatives to eyesight in order to understand the surrounding environment, detect desired objects and places, find the path to their destination, and move safely and independently in the environment. Many use their sense of hearing to sense the surroundings, and/or rely on a white cane for navigation in an unknown environment. For this reason, some papers [Chen et al. 2017; Li et al. 2018; Slade et al. 2021; Zhang et al. 2021a; Zhang and Ye 2017] mounted a Human-Computer Interaction (HCI) module onto a white cane to enhance their guidance. With the aim of increasing walking speed and reducing mental burden, [Slade et al. 2021] proposed an augmented cane as a navigation aid for VIB people. The authors investigated the accuracy of different feedback methods including grounded kinesthetic, vibrotactile, and audio feedback. They provided audio feedback through earbuds to convey audio instructions, grounded kinesthetic feedback with a motorized omni wheel on the cane which turns the tip of the cane to steer the user, and vibrotactile feedback using two vibrating motors. The authors noticed that grounded kinesthetic feedback is faster and more accurate than audio and vibrotactile feedback. A push button was installed on the cane, with which users could adjust the speed of the grounded kinesthetic feedback or turn it off. In this way, users could customize the settings to suit their walking speed.

In some studies, the Human-Computer Interface is placed on user body parts to convey feedback to the user. [Li et al. 2022] mounted a haptic module onto the legs to inform the user about the distance to the obstacles ahead of them. Although only 1 out of 42 participants in our survey was willing to receive feedback on his legs. This participant has reported it does not matter on which part of the body he receives feedback. Feedback should convey environmental information to the user in a timely manner. Additionally, it should allow the user to adapt to it quickly and the received signals should be effortlessly interpretable. Most research has used audio to transport environmental information to the VIB user (see Table 3). Despite the advantages of this method such as clear transmission of instructions, its disadvantages should not be ignored. Devices such as earbuds, earphones, and headsets that transmit audio to the user cover the ears and prevent the user from hearing ambient sounds. Additionally, audio feedback loses its effectiveness in noisy places.

Audio feedback can also take longer to guide the user than some other techniques. To tackle this problem, most of the publications used a combination of audio and tactile feedback. In addition to the aforementioned feedback techniques, some other types of feedback such as three-dimensional auditory [Fathi et al. 2022] and visual cues [Bai et al. 2018] have been investigated, which can be adopted to guide VIB individuals depending on their vision impairment severity and the environment in which they navigate.

4 REVIEW OF PROPOSED SOLUTIONS

Although a real-time uniform assistive system for indoor and outdoor navigation is necessary for human navigation, due to the variety and difference of challenges in different environments, most of the papers have focused on either indoor or outdoor settings. There are few solutions for both indoor and outdoor navigation, and a limited number have provided general solutions for VIB navigation regardless of the environment. Therefore, we categorized the papers under review based on their working environment, including indoor; outdoor; and combined indoor and outdoor, and discussed the solutions provided and relevant open issues. Overall, indoor navigation was the most commonly explored among the reviewed papers. Table 4 shows the research issues, advantages, and disadvantages of each proposed solution for indoor navigation. The pros and cons of other solutions are given in table 5.

4.1 Indoor navigation

The main goal of autonomous navigation is to safely navigate through an unfamiliar environment from the start to the target position while maintaining a safe path and generating an optimal path length [Gul et al. 2019]. Currently, numerous solutions have been proposed for VIB navigation in an unknown scenario, some of which have used SLAM techniques. In scenarios where positioning infrastructure-based solutions such as beacons or GPS do not provide adequate accuracy or are not available, SLAM techniques are adopted to provide reliable localization. Importantly, these systems must have high efficiency and high reliability to guide users safely and independently in unseen environments. To navigate collision-free in an indoor unknown environment, five major challenges should be considered: scene understanding, object detection/obstacle avoidance, localization, wayfinding, and key object localization. In what follows, we summarize the reported solutions in the papers under review for these five issues.

4.1.1 Scene understanding. In order to navigate safely and efficiently, it is necessary for VIB users to have a high-level understanding of their surroundings. To enable this, assistive devices use sensors to collect environmental information which is then conveyed to the used via feedback; this is referred to as scene understanding. In a novel approach, [Plikynas et al. 2022] collects environmental and indoor route information by collaborating with sighted volunteers. Using an android mobile application, volunteers load a 2D floor evacuation plan, walk through indoor paths, and perform semantic tagging of places of interest such as elevators and stairs. Additional data collected by IMU, stereo, and IR (depth) cameras is transferred to a cloud server, where an ORB-SLAM3 algorithm uses it to generate paths which are then saved in cloud storage. In navigation mode, VIB users specify the desired destination using an android mobile application and are guided by verbal commands and tactile feedback. If users encounter an unknown obstacle or lose their way, they can contact the registered volunteers and ask for help. By receiving the user's current location and 2D floor map, volunteers are able to guide users. The proposed solution is still in the study stage and does not yet have the necessary accuracy or efficiency. Additionally, this approach has not been evaluated in a real-world situation, and the use of the cloud platform and its relevant challenges have not been investigated. Moreover, the authors have not mentioned whether this system can be used in navigation between floors, considering that stairs are often one of the

Table 4. Advantages and limitations of proposed solutions for indoor navigation

Ref.	Research issue	Advantages	Limitations
[Li et al. 2022]	Multi-target recognition	Obstacle detection with more than 10 cm height, low-cost, reduce mental burden	Bulky, requiring stable secure internet connection
[Plikynas et al. 2022]	Crowdsourced routing process	Lightweight, easy to carry, no need for indoor infrastructural, volunteer help, obstacle detection in upper body	Need to update periodically route information, requiring relatively more input, reducing user privacy
[Ou et al. 2022]	Dynamic obstacles	Identifying dynamic and high speed objects	Highly dependency of computational complexity on the number of dynamic objects
[Lu et al. 2021b]	Surrounding understanding	Easy to carry, integrable in a comprehensive assistive system	Designed to work in a room not for navigation in a building
[Xu et al. 2021]	Navigation and object finding	Selecting appropriate mode to perceive the environment, low computational complexity	No backup in case of internet connection lost, bulky, not mentioned the efficiency in the face of moving obstacles
[Lu et al. 2021a]	Navigation with dynamic pedestrian presence	Navigating on flat surfaces and a few stairs	Bulky, confusing feedback, not able to navigate on rough terrain, low maximum speed of the robot
[Hakim and Fadhil 2021]	Finding final goal	Detecting the transparent objects, low cost	Heavy, not able to detect the object in low height
[Liu et al. 2021]	Directional feedback	Reducing mental burden, lightweight, easy to carry	Confusing by Tactus+Audio feedback, requiring the establishment of infrastructure
[Zhang et al. 2021a]	Pose estimation	Reducing the pose estimation error	Heavy, not explored the accuracy in complex conditions
[Zhang and Ye 2020]	Pose estimation	Satisfied pose estimation performance	Ambient sound blockage due to the user guidance only through audio
[Jin et al. 2020]	Moving-person tracking	Low-cost, satisfied effectively in a cluttered environment	Decreasing performance in different lighting conditions and speed, limited guiding instructions
[Zhao et al. 2019]	Indoor mobility	Low-energy cost	Failing to recognize classes of objects which are not trained
[O'Reilly et al. 2019]	Navigating using sounds	Efficient in low light conditions	Expensive
[Zhang and Ye 2019]	Robotic navigation aid	Selecting guiding mode by user	No evaluation on VIB people
[Zhang et al. 2019]	Localization	Reducing mental burden, lightweight, easy to carry, exploiting existing framework	Haptic gloves would affect holding objects in daily life
[Li et al. 2018]	Vision-based navigation	Reducing cognitive burden, lightweight, supporting multi-floor transitions	Not mentioned the efficiency in crowded environments and different lighting conditions
[Eden et al. 2018]	Text for localization	Useful for navigating in text-rich environment, simple user setup, no need to specific marker, small corpus of data	Need to repeat the data collection process, decreasing localization accuracy by changing the location of the products
[Yun et al. 2018]	Robot for indoor navigation	Being able to implement a terrestrial robots approach operating on uneven terrain	Problem in mapping due to reflection and transmission, collision due to slow processing speed of Salient algorithm
[Hakim and Fadhil 2019]	Navigation and object recognition	Low-cost, temporary object detection	No multifloor transitions
[Ramesh et al. 2018]	Localization	Computationally less expensive, low-cost	Dealing only with static object
[Bai et al. 2018]	Route following	Low-cost, displaying the guiding and surrounding information on OST-glasses	Ambient sound blockage due to the user guidance only through audio
[Zhang and Ye 2017]	Independent mobility	Reducing the 6-DOF pose error	System failure due to fast walking speed
[Chen et al. 2017]	Pathfinding	Failure backup	Inadequate position and rotation accuracy
[Endo et al. 2017]	Navigation	Potential to be small and lightweight	Dealing only with static or temporary objects, no evaluation on VIB people

biggest indoor challenges users face. However, knowing that the number of routes in an indoor environment is limited and it is possible to build trajectories related to all possible paths, this method shows potential for indoor navigation – especially where additional infrastructure-related costs (such as installing Bluetooth beacons and RFID tags) would be untenable.

[Li et al. 2018] also takes help from sighted people to collect indoor semantic environmental information. This paper presents an application called ISANA, which is implemented on a Google Tango tablet containing an RGB-D camera, a wide-angle camera and a 9-axis IMU. To provide visual positioning, a Google Tango tablet receives IMU and fisheye camera data and saves a visual features model in an Area Description File (ADF). ISANA leverages this positioning service to localize the user on a semantic map. To do so, the authors developed a map editor that receives Computer-Aided Design models (CAD) of a building as input and produces a semantic map (the indoor map with spatial context-aware information); the map consists of three geometric layers including a global 2D traversable grid map layer, an area and point of interest layer, and a topological layer. An algorithm designed by the authors then aligns the semantic map layers with the ADF. This paper additionally developed an obstacle avoidance solution, which is able to identify dynamic objects in the scene and guide the user on a safe path to the desired destination . The assistive system guides the user along the route via speech-audio and haptic interaction through a CCNY SmartCane [Chen et al. 2017].

4.1.2 Obstacle avoidance. Obstacle avoidance refers to the task of walking through the desired path without colliding with static and dynamic obstacles. Several publications only consider static obstacles, while some papers also explore dynamic obstacle avoidance. By adopting the panoptic segmentation technique, [Ou et al. 2022] proposes an assistive aid to help VIB users become aware of dynamic objects and perceive their motion. In order to estimate the ego motion of dynamic objects on the scene, static keypoints are obtained via sparse-feature visual SLAM and are combined with dynamic keypoints. Dynamic keypoints are obtained by optical flow tracking and are used to identify and track dynamic objects. This method uses the PanopticFCNl network to perform panoptic segmentation. Although the computational complexity of this system depends on the number of dynamic objects, this method is able to provide a variety of data to the user including the number of dynamic objects, average distance, position, speed, and possible direction of movement. Identifying objects that move quickly in the environment is one of the significant advantages of this solution.

In another work, [Lu et al. 2021a] proposed a deep reinforcement learning-based guiding robot that can effectively avoid dynamic obstacles. In addition, the authors compared two localization methods: SLAM localization and UWB localization. They found that navigation using UWB positioning was more stable and efficient than when SLAM localization was used, and resulted in significantly shorter trial times. SLAM results were also more vulnerable to errors when the user encountered a passing pedestrian. According to the results reported in this paper, although SLAM trajectories are smoother than UWB trajectories as long as pedestrians have not passed, SLAM is vulnerable in certain dynamic scenarios. This work integrated on-handle and on-beacon speakers to provide verbal feedback to the user. However, such speakers may not always be audible in crowded environments, and in real-life scenarios it may disturb other pedestrians. Considering that the location of the guiding robot is estimated by the positioning algorithm, and the robot moves a certain distance ahead of the user, the authors have not determined whether there is a possibility of user collision when a dynamic object or pedestrian enters a scene. Furthermore, due to the large size of the robot, navigation through narrow passages is challenging, especially when more than one user is in the scene.

4.1.3 Localization. Real-time localization of a user in an environment is crucial to enabling different functionalities such as obstacle avoidance and pathfinding. The higher the accuracy of user pose estimation, the higher the performance of locating the user in the environment, his/her position relative to obstacles, and guiding him/her safely to the destination on a collision-free path. The accumulation of pose estimation errors in the navigation process can lead to the failure of the navigation system. Therefore, reducing pose errors is very important.

In order to estimate accurate and real-time user pose in an indoor environment, [Zhang et al. 2021a] presented a new visual positioning system consisting of a novel visual-inertial odometry method and a particle filter localization method. Using a particle filter technique and 2D floor map of the environment, the proposed method reduced pose estimation error and increased localization accuracy. To evaluate the proposed positioning method, this study designed a prototype system for VIB navigation. In order to collect environmental and inertial data, an RGB-D camera and an IMU were mounted on a cane whose movement is controlled by a motor to guide the user to the correct direction. From the point of view of the assistive device, installing all the hardware needed for the navigation system, including sensors, guiding motor, and UP Board computer on the cane makes it heavy. Additionally, it is necessary to explore the accuracy of this method in complex conditions such as changing illumination conditions. However, considering that the proposed method reduces the accumulated pose error, it could be used in large-scale indoor navigation.

Exploiting existing frameworks can be a good solution for developing navigation aids. To solve the problem of localization drifts caused by accumulative pose errors, [Zhang et al. 2019] used Google ARCore vision SLAM to track user's poses and build a map of the environment in real time. ARCore is Google's platform with three key capabilities including motion tracking, environmental understanding, and light estimation which can be used for localization, grasping the real-world environment, and evaluating ambient lighting conditions, respectively. The proposed system performs localization based on the Area Learning technique. Area Learning is a mechanism to understand and remember surroundings [Marques et al. 2018]. This work labeled the area of interest in CAD maps of the building and then, with the help of sighted people, created a sparse map of the environment using Google ARCore. Finally, mapping between the points in these two representations was acquired. As the authors reported, benefiting from ARCore, the proposed solution showed acceptable mapping and tracking results. In addition, using ARCore area learning, the authors proposed a path planning algorithm that provides safe and efficient indoor navigation for users. Along the path, users are guided forward and warned of obstacles by a dual-channel feedback system consisting of haptic gloves and an audio interface. In terms of feedback, test users declared that haptic gloves are cumbersome. Additionally, in different weather conditions, such as hot weather, the gloves can be uncomfortably warm and the system can lose efficiency due to sweating of the user's hands.

4.1.4 Wayfinding. Wayfinding is the process of determining and following a path or route between an origin and a destination. The assistive system should be able to find the optimal path from the current user's location to his/her desired destination and to determine the target direction in real time. In addition, it is of great importance to design a guidance technique that helps users to go through the path smoothly and accurately. To increase path-following performance, [Liu et al. 2021] focused on guiding the user to follow a path and explored two types of directional feedback methods including Tactus-Only and Tactus+Audio. They have designed a Tactile Compass by which users can recognize their deviation from the path and perceive the direction. In order to develop an assistive navigation system, this paper considers the use of SLAM technique for positioning and path planning as the best solution to integrate with the proposed feedback method. Although this approach was in the study stage at the time of publication, the idea of this work can be integrated with other tasks relevant to VIB navigation to improve path-following tasks and reduce the user's cognitive load.

On the other hand, [Zhang and Ye 2017] focused on the wayfinding algorithm and guided the user via speech feedback. This work presented a new pose estimation method with client-server architecture for real-time wayfinding.

This solution uses a 3D camera mounted on a computer-vision-enhanced white cane for environment perception and obstacle detection. The authors compared the proposed method with planar SLAM and RGBD-SLAM and reported that the proposed method shows better results for pose estimation with less computation time. Considering that the focus of each paper is on one of the wayfinding issues, it is good to take advantage of each method and try to resolve the challenges of the navigation system by integrating the proposed methods.

4.1.5 Key object localization. Finding a desired object or place such as a chair or stairs is a core task in VIB navigation and can improve the quality of user navigation in unknown environments. In this regard, [Lu et al. 2021b] uses Google glasses to perform interactive indoor navigation. After the user gives a spoken description of the desired target, the objects in the scene are identified from images captured by the Google glasses using the YOLOv4 algorithm. Then the system associates the user's description with existing objects in the scene. Using a classic model of image captioning, the user receives a description of the detected objects and is asked whether the object matches the target of interest. Based on the users' answers, the recognition model is re-trained and will be used in subsequent similar navigations. Using a pre-built map, this work uses OpenVSLAM for user localization. It seems that the presented system only works in a room and is not designed for navigation through buildings. Extending the capabilities of this solution and adding services such as detecting dynamic objects or face recognition could contribute to a comprehensive navigation system for VIB users.

In addition to an indoor navigation approach, [Xu et al. 2021] proposed a lightweight object detection solution and provided the identification of the object of interest functionality. By leveraging YOLOV4 network, the proposed system estimates the 3D position and semantic information of objects of interest and transmits this information to the user through audio and haptic feedback. In order to optimize the calculation speed of the YoloV4 network and also to improve the detection performance, the authors modified the structure of this network. For this purpose, they used MobilenetV3 instead of CSP-Darknet53 as the backbone of YOLOV4. Furthermore, 3x3 traditional convolution was replaced with depthwise separable convolution. Additionally, they added attention modules to the feature map acquired by MobilenetV3. For positioning functionality, the proposed method constructed the indoor 3D sparse point cloud map by leveraging the ORB-SLAM2 algorithm. One of the notable advantages of this system is the possibility of selecting the appropriate system mode including navigation or finding a specific object. In this case, unnecessary feedback will not be delivered to the user and the amount of cognitive load is reduced. However, this paper does not mention the efficiency of the system in the face of moving obstacles and complex situations such as changing light, which are two of the most important challenges in navigation.

4.2 Outdoor navigation

Most of the publications under review used the SLAM technique for VIB navigation in indoor settings. However, some of them used SLAM techniques to deal with outdoor navigation challenges. Since the challenges in complex outdoor environments are many and various, each paper focuses on a specific challenge such as crossing the street, walking seamlessly with nearby pedestrians, and decreasing cognitive burden. Among them, [Son and Weiland 2022] introduced a system to guide users to cross the street independently. This system adopted RGB-D camera and IMU to collect environmental information. With its appropriate verbal feedback, the user can choose their walking speed when crossing the pedestrian lane. Even though the hardware used is commercially available, it is not applicable in real-life scenarios since this system does not have the ability of object detection and safety still remains an issue. Another limitation of this system is that it can only be used for crosswalks with Tenji bricks (red sidewalk tiles with raised bumps), but no solution has been provided for other types such as crosswalks without zebra stripes, without Tenji bricks, and without traffic lights. Although this paper only examines a very specific problem with many constraints which cannot be used in real-world situations, the ideas and algorithms proposed can be used in the development of a system for navigation in both indoor and outdoor environments.

With the aim of navigating the last few meters from the sidewalk to a building entrance, [Weiss et al. 2020] introduces an open-source and novel simulated environment. This is a serious challenge that was also reported by one of our survey participants. This study examined the use of Reinforcement Learning methods for learning agents aiming to help with VIB navigation tasks. The authors profited from ORB- SLAM2 method to obtain camera poses from which a 2-D graph is built. This graph can be used to determine the shortest path between any two poses which can be exploited for pathfinding tasks. Despite the navigation task was successfully done with a rate of 74.9% using the proposed trained model, the solution only works in the training region and is not generalized to other regions. The authors also did not mention the other challenges such as occlusion, for example when the entrance door of the target building is blocked by a big truck.

Faced with appearance variation issues, including changes in lighting and viewing angle, dynamic objects, and motion blur, [Cheng et al. 2021] proposed an open-source solution for visual localization. In this approach, a unified descriptor network called Dual Desc is designed to extract both global and 2D-3D local descriptors from multimodal images. Images are captured by an RGB-D-IR camera integrated into a commercialized wearable assistive device called Intoer. As the authors reported an acceptable performance for their localization approach on real-world datasets under the outdoor scenario of assistive navigation, this idea can be integrated with other navigation aspects such as scene understanding and feedback methods to develop a navigation solution.

In terms of localization challenges, both methods proposed in [Cheng et al. 2021] and [Duh et al. 2020] are reliable under lighting changes. [Duh et al. 2020] proposed a new vision-based localization method for relative and global pose estimation. This method uses an ORB-SLAM algorithm to estimate the relative pose of every frame from a monocular camera. Then it runs a model-based localization algorithm to estimate global pose. These two estimations are performed on different computational resources that can extend the computational latency.

[Cheng et al. 2021] profited from a unified deep network to extract both local and global descriptors from RGB, infrared, and depth images. The global descriptors are used to retrieve the coarse candidates of query images from which the optimal results are selected in the geometric verification process. Inspired by SeqSLAM, this approach has also added a sequence matching step to robustify the localization results by synthesizing the verified results of successive frames. [Duh et al. 2020]'s solution will be described further in section 4.3.

4.3 Combined indoor and outdoor navigation

Although the assistive aids that simultaneously tackle all the challenges related to indoor and outdoor navigation as well as navigating between indoor and outdoor environments are crucial for VIB people, few studies have addressed this issue. The most impressive recent research study, [Slade et al. 2021], proposed an augmented cane that has significantly improved the user's mobility and overcomes more navigational challenges than similar works before it. In addition to indoor and outdoor navigation, this paper has also worked on combined indoor and outdoor navigation. Another significant advantage is the use of a white cane as the basis of the system for obstacle avoidance in the event of a system failure. The abovementioned strengths, as well as being an open-source framework and the low cost of augmented cane, have made this work a foundation for further research. However, it still has disadvantages such as being heavy and not being hands-free.

Table 5. Advantages and limitations of proposed solutions for outdoor and both indoor and outdoor navigation

Ref.	Research issue	Advantages	Limitations	Working area
[Son and Weiland 2022]	Street crossing	Using a prior map with semantic info., reducing mental fatigue, selecting preferred walking speed	Not able to detect objects, used only for crosswalks with Tenji bricks, heavy	Outdoor.
[Rui et al. 2021]	Multi-sensory feedback	Multi-sensory feedback, detecting obstacles in left, right, front, upper and lower sides of the user	Not considered dynamic object	General
[Slade et al. 2021]	Decreasing cognitive burden, increasing walking speed	Failure backup, low cognitive load, low-cost, open source, easy to carry, selecting preferred walking speed by user	Heavy, requiring mechanical assembly	Both
[Cheng et al. 2021]	Confronting variations between query and database images	Open source, confronting appearance variations between query and database images	No feedback to the test user	Outdoor.
[Chen et al. 2021]	Surrounding understanding	Surrounding understanding in real-time, low-cost	Heavy, obtrusive	General
[Kayukawa et al. 2020]	Walking seamlessly with nearby pedestrians	Reducing mental burden, low-cost, capturing images without significant motion-induced blur	Bulky, risk of hitting against pedestrians, only for flat spaces, user and suitcase should face in the same direction	Both
[Duh et al. 2020]	Spatial awareness	Reducing mental burden, low-cost, lightweight, easy to carry, recognizing accurately key elements	Implementing different algorithms on two remote servers	Both
[Weiss et al. 2020]	Visuospatial perception	Open source, sidewalk navigation	Not generalized train policy, not mentioned occlusion challenges	Outdoor.
[Bai et al. 2019]	Navigation and obstacle avoidance	Reduced mental burden, low-cost, activating recognition module according to the user requirement, detecting hanging obstacles, implementing all algorithm on a smartphone	Not satisfactory when moving across different floors, heavy glasses	Both
[Bai et al. 2017]	Environment perception	Using parallel computing ability and storage of the cloud resources, low cognitive load	Requiring of building map of the whole world, weak perception module	Both

While most assistive systems propose an alternative path around the detected obstacles, [Kayukawa et al. 2020] offers an additional walking mode, called on-path, that guides VIB users to avoid moving obstacles without changing their path by adjusting their walking speed. Attaching a camera, IMU and LiDAR as well as feedback interfaces on the handle of a suitcase, the system guides the user for navigation in public places such as train stations or shopping centers equipped with tactile paving. In order to localize the user in real time, this design uses a SLAM-based localization method that receives IMU and LiDAR sensor information and compares it with a 3D point cloud map previously generated. It also profited from two RGB-D cameras to detect and track pedestrians. Like [Slade et al. 2021], this study also uses various sensors to obtain environmental information, with the difference that [Slade et al. 2021] used a white cane as a sensor carrier and a backup in case of system failure. In this work, the users carry their white canes in addition to the suitcase.

Compared to [Slade et al. 2021], this system can only be used in flat spaces with tactile paving, while such assumptions do not limit the functionality of the system presented in [Slade et al. 2021]. In addition, this system considers moving pedestrians as dynamic objects and standing pedestrians as static objects, and does not consider other types of obstacles at different levels of the user's body. Both prototypes, proposed in [Slade et al. 2021] and [Kayukawa et al. 2020], install the tactile interface on the assistive device. While [Slade et al. 2021] adopts a grounded kinesthetic attached at the tip of the white cane to steer the user, to show the correct direction, [Kayukawa et al. 2020] attaches a directional lever on the suitcase handle. Another constraint of this work is that the direction of the suitcase affects the accuracy of collision prediction, and the suitcase should be moved in the direction of the user's movement, while using [Slade et al. 2021]'s augmented cane, the user can move the cane freely in any direction. However, combining the ideas of these two works can lead to providing a comprehensive navigation system for VIB people.

To deal with the challenge of moving obstacles in the environment and estimating the exact position and orientation of VIB users to select an optimal path during navigation, [Duh et al. 2020] presented a new localization method called VB-GPS, which uses Visual SLAM and image-based localization methods for relative and global pose estimation. This method outperformed three state-of-the-art positioning methods including ORB-SLAM, MBL, and PoseNet. The proposed solution distributed calculations on a local computer and two remote servers. Although using servers has its advantages, it also brings challenges such as network security, constant network connectivity, and user privacy, especially in this work with two separate servers these challenges become more prominent. To evaluate the system, the authors designed a real-life scenario based on which the users perform indoor, outdoor, and between indoor and outdoor navigation. Although the participants reported safe and independent navigation, considering the feedback is transmitted to the user through voice, the paper did not investigate the efficiency of the system in crowded places and did not evaluate the amount of feedback load in such scenarios. In addition, like most papers, this work did not consider the user's navigation status in case of assistive system or remote server failure. Moreover, the authors have not considered the problem of obstacles at different levels.

Focusing on navigation and scene perception, [Bai et al. 2019] proposed a wearable assistive aid for navigation in both indoor and outdoor environments. The proposed prototype, which senses the environment with an RGB-D camera and IMU mounted on a pair of glasses, can estimate the user's location by using an indoor map provided by visual SLAM and smartphone GPS module, in indoor and outdoor scenarios, respectively. This information will be used for the path planning stage. The experiments show that, compared to a white cane, this approach is more efficient while also is able to detect hanging objects that are difficult for a white cane to detect. However, it is not able to detect small objects. In this system, the user receives feedback only through audio, which has some limitations as mentioned in section 3.3. This work takes advantage of a smartphone to store the trained object-recognition models and indoor maps, to run navigation and perception algorithms, as well as to transfer feedback to the user while achieving real-time performance.

In case of encountering an obstacle during the path following process, this publication suggests that instead of changing the path, the user can receive more information about the obstacle, so that if it is moveable, he/she moves it, clears the path, and follows the current path. Taking advantage of this idea to develop a comprehensive assistive device, reduces navigation time and the risk of becoming disoriented.

5 FUTURE OPPORTUNITIES AND CONCLUSION

Despite significant advances in autonomous navigation, pedestrian navigation remains challenging, particularly for VIB people. An effective navigation system for VIB people needs to meet mobility metrics such as decreasing navigation time, decreasing navigation distance, decreasing contact with the environment, and increasing walking speed. These systems must have high accuracy and efficiency in complex situations such as crowded places and changing light and weather conditions. At the same time, assistive aids should be comfortable, easy to use, unobtrusive, cost-effective, lightweight, and reduce cognitive load. In this review paper, we studied the publications that have used SLAM technique in their approaches. One of the advantages of SLAM is that this technique can be used in any location, and there is no need for pre-built maps or additional infrastructures like Bluetooth beacons or RFID tags. However, there is still much room for improvement, not only in user localization in an unknown environment but also in other aspects of navigation systems.

Due to the rapid development of hardware technologies, computer vision, and deep learning techniques, the existing challenges in this area will be solved over time. Future studies should pay more attention to diverse aspects of VIB navigation, including technologies for sensing the environment and obtaining surrounding data, techniques to process ambient data and provide semantic information for the users, solutions to perform real-time positioning and pathfinding, techniques to detect and recognize surrounding objects, and techniques to deliver environmental information to the user with low concentration and low cognitive load. Object detection and recognition are especially essential, as they are needed for obstacle avoidance. However, obstacle avoidance remains a significant challenge since it must factor in obstacles at different body levels, drop-offs, uneven steps, pillars, moving obstacles, and small-size obstacles. Recognition accuracy and the number of object classes that can be realized by the system are other issues that need to be addressed.

One of the most important aspects of assistive devices is computing resources. This must be powerful enough to perform calculations in real time, at the same time it must be light, small and portable. To deal with this challenge, some studies have transferred computing to the cloud and remote servers. Although this solution significantly reduces the weight of assistive devices and provides access to powerful computing resources, it also raises new challenges, such as the need for constant access to the Internet and security as well as privacy issues. Given that localization is a requisite for navigation that must be performed with high accuracy in real time, it should be considered carefully to improve the stability, robustness, and accuracy of relevant algorithms and to minimize delay for real-time processing.

Feedback modules also require special attention, since they are the only part of an assistive device that directly communicates with the users and must convey information in an efficient and intuitive manner. Although many works (e.g. [Liu et al. 2021; Rui et al. 2021; Slade et al. 2021]) have studied enhancements in human-computer interaction, it remains an unsolved problem, especially in crowded and noisy environments where there is a lot of environmental data. In such a situation, an adequate method is required to provide intuitive and user-friendly feedback to deliver accurate and sufficient information to the user in real time.

Wherever possible, assistive devices should not create added challenges in terms of weight, size, portability, and physical demands. For this reason, reducing size/weight and enhancing appearance still need to be better addressed.

Another issue that has been less focused on is system performance in complicated scenarios such as different lighting, seasonal, and weather conditions.

We noticed that most of the papers under review employed traditional SLAM approaches such as ORB-SLAM2. These techniques have difficulty dealing with low-texture and no-texture areas, low-light and excessively bright conditions [Li et al. 2021], and SLAMIDE issues (i.e. SLAM in dynamic environments) [Bibby and Reid 2007]. Deep learning-based SLAM solutions replace traditional hand-made features with deep learning-based features to improve the accuracy and stability of the SLAM system [Li et al. 2020, 2021]. Deep learning can be also used to replace loop closure and pose estimation modules in traditional SLAM to improve on traditional methods [Chen et al. 2022]. Deep learning technology can be adopted to detect dynamic objects at a semantic level. The advantages of this integration can be used to increase pedestrian navigation systems' efficiency in challenging environments.

We also observed that all the reviewed approaches are prototypes in the earlier stages of research and not yet at the stage of a practical product. One possible reason for this is a lack of a unified community or group to invest in and focus on solving VIB navigation challenges. Most work in this area is carried out by academic groups or small companies, and as such it often fails to result in a feasible final product.

ACKNOWLEDGMENTS

We would like to thank all survey participants.

REFERENCES

Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M Khairi, and Vijayabaskar Kasi. 2013. Reviews on various inertial measurement unit (IMU) sensor applications. *International Journal of Signal Processing Systems* 1, 2 (2013), 256–262.

Yusra Alkendi, Lakmal Seneviratne, and Yahya Zweiri. 2021. State of the art in vision-based localization techniques for autonomous navigation systems. *IEEE Access* 9 (2021), 76847–76874.

Jinqiang Bai, Shiguo Lian, Zhaoxiang Liu, Kai Wang, and Dijun Liu. 2018. Virtual-blind-road following-based wearable navigation device for blind people. *IEEE Transactions on Consumer Electronics* 64, 1 (2018), 136–143.

Jinqiang Bai, Dijun Liu, Guobin Su, and Zhongliang Fu. 2017. A cloud and vision-based navigation system used for blind people. In *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies*. 1–6.

Jinqiang Bai, Zhaoxiang Liu, Yimin Lin, Ye Li, Shiguo Lian, and Dijun Liu. 2019. Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics* 8, 6 (2019), 697.

Charles Bibby and Ian Reid. 2007. Simultaneous localisation and mapping in dynamic environments (SLAMIDE) with reversible data association. In *Proceedings of Robotics: Science and Systems*, Vol. 66. 81.

Andrius Budrionis, Darius Plikynas, Povilas Daniušis, and Audrius Indrulionis. 2022. Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. *Assistive Technology* 34, 2 (2022), 178–194.

Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics* 32, 6 (2016), 1309–1332.

Amit Chaudhary, Dr Verma, et al. 2019. State of Art on Wearable Device to Assist Visually Impaired Person Navigation in Outdoor Environment. In *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*.

Qingtian Chen, Muhammad Khan, Christina Tsangouri, Christopher Yang, Bing Li, Jizhong Xiao, and Zhigang Zhu. 2017. CCNY smart cane. In *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, 1246–1251.

Weifeng Chen, Guangtao Shang, Aihong Ji, Chengjun Zhou, Xiyang Wang, Chonghui Xu, Zhenxiong Li, and Kai Hu. 2022. An overview on visual slam: From tradition to semantic. *Remote Sensing* 14, 13 (2022), 3010.

Zhuo Chen, Xiaoming Liu, Masaru Kojima, Qiang Huang, and Tatsuo Arai. 2021. A wearable navigation device for visually impaired people based on the real-time semantic visual slam system. *Sensors* 21, 4 (2021), 1536.

Ruiqi Cheng, Weijian Hu, Hao Chen, Yicheng Fang, Kaiwei Wang, Zhijie Xu, and Jian Bai. 2021. Hierarchical visual localization for visually impaired people using multimodal images. *Expert Systems with Applications* 165 (2021), 113743.

Dimitrios Dakopoulos and Nikolaos G Bourbakis. 2009. Wearable obstacle avoidance electronic travel aids for blind: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1 (2009), 25–35.

César Debeunne and Damien Vivet. 2020. A review of visual-LiDAR fusion based simultaneous localization and mapping. *Sensors* 20, 7 (2020), 2068.

Ping-Jung Duh, Yu-Cheng Sung, Liang-Yu Fan Chiang, Yung-Ju Chang, and Kuan-Wen Chen. 2020. V-eye: A vision-based navigation system for the visually impaired. *IEEE Transactions on Multimedia* 23 (2020), 1567–1580.

Jake Eden, Thomas Kawchak, and Vijaykrishnan Narayanan. 2018. Indoor navigation using text extraction. In *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE, 112–117.

Fatma El-Zahraa El-Taher, Ayman Taha, Jane Courtney, and Susan Mckeever. 2021. A systematic review of urban navigation systems for visually impaired people. *Sensors* 21, 9 (2021), 3103.

Yuki Endo, Kei Sato, Akihiro Yamashita, and Katsushi Matsubayashi. 2017. Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM. In *2017 International Conference on Biometrics and Kansei Engineering (ICBAKE)*. IEEE, 158–162.

Christine Evers and Patrick A Naylor. 2018. Acoustic slam. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26, 9 (2018), 1484–1498.

Agebson Rocha Façanha, Ticianne Darin, Windson Viana, and Jaime Sánchez. 2020. O&M indoor virtual environments for people who are blind: A systematic literature review. *ACM Transactions on Accessible Computing (TACCESS)* 13, 2 (2020), 1–42.

Kiavash Fathi, Alireza Darvishy, and Hans Wernher Van de Venn. 2022. Augmented Reality for the Visually Impaired: Navigation Aid and Scene Semantics for Indoor Use Cases. In *2022 IEEE 12th International Conference on Consumer Electronics (ICCE-Berlin)*. IEEE, 1–6.

Zixiang Fei, Erfu Yang, Huoheng Hu, and Huiyu Zhou. 2017. Review of machine vision-based electronic travel aids. In *2017 23rd International Conference on Automation and Computing (ICAC)*. IEEE, 1–7.

Hugo Fernandes, Paulo Costa, Vitor Filipe, Hugo Paredes, and João Barroso. 2019. A review of assistive spatial orientation and navigation technologies for the visually impaired. *Universal Access in the Information Society* 18, 1 (2019), 155–168.

Faiza Gul, Wan Rahiman, and Syed Sahal Nazli Alhady. 2019. A comprehensive study for robot navigation techniques. *Cogent Engineering* 6, 1 (2019), 1632046.

Heba Hakim and Ali Fadhil. 2019. Indoor Low Cost Assistive Device using 2D SLAM Based on LiDAR for Visually Impaired People. *Iraqi Journal for Electrical & Electronic Engineering* 15, 2 (2019).

Heba Hakim and Ali Fadhil. 2021. Indoor Wearable Navigation System Using 2D SLAM Based on RGB-D Camera for Visually Impaired People. In *Proceedings of First International Conference on Mathematical Modeling and Computational Science*. Springer, 661–672.

Md Milon Islam, Muhammad Sheikh Sadi, Kamal Z Zamli, and Md Manjur Ahmed. 2019. Developing walking assistants for visually impaired people: A review. *IEEE Sensors Journal* 19, 8 (2019), 2814–2828.

Rabia Jafri, Syed Abid Ali, Hamid R Arabnia, and Shameem Fatima. 2014. Computer vision-based object recognition for the visually impaired in an indoors environment: a survey. *The Visual Computer* 30, 11 (2014), 1197–1222.

Dongsuk Jeon, Nathan Ickes, Priyanka Raina, Hsueh-Cheng Wang, Daniela Rus, and Anantha Chandrakasan. 2016. 24.1 A 0.6 V 8mW 3D vision processor for a navigation device for the visually impaired. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 416–417.

SongGuo Jin, Minhaz Uddin Ahmed, Jin Woo Kim, Yeong Hyeon Kim, and Phill Kyu Rhee. 2020. Combining obstacle avoidance and visual simultaneous localization and mapping for indoor navigation. *Symmetry* 12, 1 (2020), 119.

Robert K Katzschmann, Brandon Araki, and Daniela Rus. 2018. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 3 (2018), 583–593.

Seita Kayukawa, Tatsuya Ishihara, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2020. Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–22.

Akif Khan and Shah Khusro. 2021. An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. *Universal Access in the Information Society* 20, 2 (2021), 265–298.

Dawar Khan, Zhanglin Cheng, Hideaki Uchiyama, Sikandar Ali, Muhammad Asshad, and Kiyoshi Kiyokawa. 2022a. Recent advances in vision-based indoor navigation: A systematic literature review. *Computers & Graphics* (2022).

Muhammad Shahzad Alam Khan, Danish Hussain, Kanwal Naveed, Umar S Khan, Imran Qayyum Mundial, and Anas Bin Aqeel. 2022b. Investigation of Widely Used SLAM Sensors Using Analytical Hierarchy Process. *Journal of Sensors* 2022 (2022).

Sulaiman Khan, Shah Nazir, and Habib Ullah Khan. 2021. Analysis of navigation assistants for blind and visually impaired people: A systematic review. *IEEE Access* 9 (2021), 26712–26734.

Bing Li, Juan Pablo Munoz, Xuejian Rong, Qingtian Chen, Jizhong Xiao, Yingli Tian, Aries Arditi, and Mohammed Yousuf. 2018. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing* 18, 3 (2018), 702–714.

Dongjiang Li, Xuesong Shi, Qiwei Long, Shenghui Liu, Wei Yang, Fangshi Wang, Qi Wei, and Fei Qiao. 2020. DXSLAM: A robust and efficient visual SLAM system with deep features. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4958–4965.

Guoxin Li, Jiaqi Xu, Zhijun Li, Chao Chen, and Zhen Kan. 2022. Sensing and Navigation of Wearable Assistance Cognitive Systems for the Visually Impaired. *IEEE Transactions on Cognitive and Developmental Systems* (2022).

Guangqiang Li, Lei Yu, and Shumin Fei. 2021. A deep-learning real-time visual SLAM system based on multi-task feature extraction network and self-supervised feature points. *Measurement* 168 (2021), 108403.

Guanhong Liu, Tianyu Yu, Chun Yu, Haiqing Xu, Shuchang Xu, Ciyuan Yang, Feng Wang, Haipeng Mi, and Yuanchun Shi. 2021. Tactile compass: Enabling visually impaired people to follow a path with continuous directional feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

Wenxin Liu, David Caruso, Eddy Ilg, Jing Dong, Anastasios I Mourikis, Kostas Daniilidis, Vijay Kumar, and Jakob Engel. 2020. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters* 5, 4 (2020), 5653–5660.

Chen-Lung Lu, Zi-Yan Liu, Jui-Te Huang, Ching-I Huang, Bo-Hui Wang, Yi Chen, Nien-Hsin Wu, Hsueh-Cheng Wang, Laura Giarré, and Pei-Yi Kuo. 2021a. Assistive navigation using deep reinforcement learning guiding robot with UWB/voice beacons and semantic feedbacks for blind and visually impaired people. *Frontiers in robotics and AI* (2021), 176.

Jun-Li Lu, Hiroyuki Osone, Akihisa Shitara, Ryo Iijima, Bektur Ryskeldiev, Sayan Sarcar, and Yoichi Ochiai. 2021b. Personalized Navigation that Links Speaker's Ambiguous Descriptions to Indoor Objects for Low Vision People. In *International Conference on Human-Computer Interaction*. Springer, 412–423.

Kanak Manjari, Madhushi Verma, and Gaurav Singal. 2020. A survey on assistive technology for visually impaired. *Internet of Things* 11 (2020), 100188.

Bernardo Marques, Raphaël Carvalho, Paulo Dias, Miguel Oliveira, Carlos Ferreira, and Beatriz Sousa Santos. 2018. Evaluating and enhancing Google Tango localization in indoor environments using fiducial markers. In *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 142–147.

Gianmario Motta, Tianyi Ma, Kaixu Liu, Edwige Pissaloux, Muhammad Yusro, Kalamullah Ramli, Jean Connier, Philippe Vaslin, Jian-jin Li, Christophe de Vaulx, et al. 2018. Overview of smart white canes: connected smart cane from front end to back end. In *Mobility of Visually Impaired People*. Springer, 469–535.

Joseph O'Reilly, Silvia Cirstea, Marcian Cirstea, and Jin Zhang. 2019. A novel development of acoustic SLAM. In *2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*. IEEE, 525–531.

Luis Enrique Ortiz, Elizabeth V Cabrera, and Luiz M Gonçalves. 2018. Depth data error modeling of the ZED 3D vision sensor from stereolabs. *ELCVIA: electronic letters on computer vision and image analysis* 17, 1 (2018), 0001–15.

Wenyan Ou, Jiaming Zhang, Kunyu Peng, Kailun Yang, Gerhard Jaworek, Karin Müller, and Rainer Stiefelhagen. 2022. Indoor Navigation Assistance for Visually Impaired People via Dynamic SLAM and Panoptic Segmentation with an RGB-D Sensor. *arXiv preprint arXiv:2204.01154* (2022).

Prabin Kumar Panigrahi and Sukant Kishoro Bisoy. 2021. Localization strategies for autonomous mobile robots: A review. *Journal of King Saud University-Computer and Information Sciences* (2021).

Darius Plikynas, Audrius Indriulionis, Algirdas Laukaitis, and Leonidas Sakalauskas. 2022. Indoor-guided navigation for people who are blind: Crowdsourcing for route mapping and assistance. *Applied Sciences* 12, 1 (2022), 523.

Darius Plikynas, Arūnas Žvironas, Andrius Budrionis, and Marius Gudauskis. 2020a. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors* 20, 3 (2020), 636.

Darius Plikynas, Arūnas Žvironas, Marius Gudauskis, Andrius Budrionis, Povilas Daniušis, and Ieva Sliesoraitytė. 2020b. Research advances of indoor navigation for blind people: A brief review of technological instrumentation. *IEEE Instrumentation & Measurement Magazine* 23, 4 (2020), 22–32.

Kruthika Ramesh, SN Nagananda, Hariharan Ramasangu, and Rohini Deshpande. 2018. Real-time localization and navigation in an indoor environment using monocular camera for visually impaired. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. IEEE, 122–128.

Santiago Real and Alvaro Araujo. 2019. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors* 19, 15 (2019), 3404.

Karla Miriam Reyes Leiva, Milagros Jaén-Vargas, Benito Codina, and José Javier Serrano Olmedo. 2021. Inertial measurement unit sensors in assistive technologies for visually impaired people, a review. *Sensors* 21, 14 (2021), 4767.

Muhammad Rabani Mohd Romlay, Siti Fauziah Toha, Azhar Mohd Ibrahim, and Ibrahim Venkat. 2021. Methodologies and evaluation of electronic travel aids for the visually impaired people: a review. *Bulletin of Electrical Engineering and Informatics* 10, 3 (2021), 1747–1758.

Chufan Rui, Yichen Liu, Junru Shen, Zhaobin Li, and Zaipeng Xie. 2021. A multi-sensory blind guidance system based on YOLO and ORB-SLAM. In *2021 IEEE International Conference on Progress in Informatics and Computing (PIC)*. IEEE, 409–414.

Walter CSS Simões, Guido S Machado, Andre Sales, Mateus M de Lucena, Nasser Jazdi, and Vicente F de Lucena. 2020. A review of technologies and techniques for indoor navigation systems for the visually impaired. *Sensors* 20, 14 (2020), 3935.

Shankar Sivan and Gopu Darsan. 2016. Computer vision based assistive technology for blind and visually impaired people. In *Proceedings of the 7th International Conference on Computing Communication and Networking Technologies*. 1–8.

Patrick Slade, Arjun Tambe, and Mykel J Kochenderfer. 2021. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. *Science Robotics* 6, 59 (2021), eabg6594.

Hojun Son and James Weiland. 2022. Wearable System to Guide Crosswalk Navigation for People With Visual Impairment. *Front. Electron. 2: 790081. doi: 10.3389/felec* (2022).

Shengjun Tang, Qing Zhu, Wu Chen, Walid Darwish, Bo Wu, Han Hu, and Min Chen. 2016. Enhanced RGB-D mapping method for detailed 3D indoor and outdoor modeling. *Sensors* 16, 10 (2016), 1589.

Ruxandra Tapu, Bogdan Mocanu, and Titus Zaharia. 2020. Wearable assistive devices for visually impaired: A state of the art survey. *Pattern Recognition Letters* 137 (2020), 37–52.

Mohammad Moeen Valipoor and Angélica de Antonio. 2022. Recent trends in computer vision-driven scene understanding for VI/blind users: a systematic mapping. *Universal Access in the Information Society* (2022), 1–23.

Hélène Walle, Cyril De Runz, Barthélemy Serres, and Gilles Venturini. 2022. A Survey on Recent Advances in AI and Vision-Based Methods for Helping and Guiding Visually Impaired People. *Applied Sciences* 12, 5 (2022), 2308.

Jie Wang, Erwu Liu, Yuanzhe Geng, Xinyu Qu, and Rui Wang. 2021. A Survey of 17 Indoor Travel Assistance Systems for Blind and Visually Impaired People. *IEEE Transactions on Human-Machine Systems* 52, 1 (2021), 134–148.

Martin Weiss, Simon Chamorro, Roger Girgis, Margaux Luck, Samira E Kahou, Joseph P Cohen, Derek Nowrouzezahrai, Doina Precup, Florian Golemo, and Chris Pal. 2020. Navigation agents for the visually impaired: A sidewalk simulator and experiments. In *Conference on Robot Learning*. PMLR, 1314–1327.

Jiaqi Xu, Haisheng Xia, Yueyue Liu, and Zhijun Li. 2021. Multi-functional Smart E-Glasses for Vision-Based Indoor Navigation. In *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 267–272.

Youngjae Yun, Taeyang Gwon, and Donghan Kim. 2018. The Design of Person Carrier Robot using SLAM and Robust Salient Detection. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 196–201.

He Zhang, Lingqiu Jin, and Cang Ye. 2021a. An RGB-D Camera Based Visual Positioning System for Assistive Navigation by a Robotic Navigation Aid. *IEEE/CAA Journal of Automatica Sinica* 8, 8 (2021), 1389–1400.

He Zhang and Cang Ye. 2017. An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 9 (2017), 1592–1604.

He Zhang and Cang Ye. 2019. Human-robot interaction for assisted wayfinding of a robotic navigation aid for the blind. In *2019 12th International Conference on Human System Interaction (HSI)*. IEEE, 137–142.

He Zhang and Cang Ye. 2020. Plane-aided visual-inertial odometry for 6-DOF pose estimation of a robotic navigation aid. *IEEE Access* 8 (2020), 90042–90051.

Xiaochen Zhang, Xiaoyu Yao, Lanxin Hui, Fuchuan Song, and Fei Hu. 2021b. A Bibliometric Narrative Review on Modern Navigation Aids for People with Visual Impairment. *Sustainability* 13, 16 (2021), 8795.

Xiaochen Zhang, Xiaoyu Yao, Yi Zhu, and Fei Hu. 2019. An ARCore based user centric assistive navigation system for visually impaired people. *Applied Sciences* 9, 5 (2019), 989.

Yu Zhao, Ran Huang, and Biao Hu. 2019. A multi-sensor fusion system for improving indoor mobility of the visually impaired. In *2019 Chinese Automation Congress (CAC)*. IEEE, 2950–2955.

VA Zhmud, NO Kondratiev, KA Kuznetsov, VG Trubin, and LV Dimitrov. 2018. Application of ultrasonic sensor for measuring distances in robotics. In *Journal of Physics: Conference Series*, Vol. 1015. IOP Publishing, 032189.

Arunas Zvironas, Marius Gudauskis, and Darius Plikynas. 2019. Indoor electronic traveling aids for visually impaired: systemic review. In *2019 International conference on computational science and computational intelligence (CSCI)*. IEEE, 936–942.