Recent trends in computer vision-driven scene understanding for VI/blind users: a systematic mapping

Mohammad Moeen Valipoor¹ · Angélica de Antonio¹

Accepted: 10 January 2022
© The Author(s) 2022

Abstract

During the past years, the development of assistive technologies for visually impaired (VI)/blind people has helped address various challenges in their lives by providing services such as obstacle detection, indoor/outdoor navigation, scene description, text reading, facial recognition and so on. This systematic mapping review is mainly focused on the scene understanding aspect (e.g., object recognition and obstacle detection) of assistive solutions. It provides guidance for researchers in this field to understand the advances during the last four and a half years. This is because deep learning techniques together with computer vision have become more powerful and accurate than ever in tasks like object detection. These advancements can bring a radical change in the development of high-quality assistive technologies for VI/blind users. Additionally, an overview of the current challenges and a comparison between different solutions is provided to indicate the pros and cons of existing approaches.

Keywords Computer vision · Visually impaired · Assistive technology · Systematic mapping · Scene understanding

1 Introduction

According to the study published by the World Health Organization (WHO), there are approximately 285 million visually impaired (VI) people around the world of which 39 million are completely blind [1]. Many of these people have difficulty handling some of their daily tasks. These tasks include navigating in an environment, knowing about obstacles in their path and identifying objects around them. During the past years, many researchers have been working on assistive solutions which use technologies like RGB-D cameras, ultrasonic sensors, optical beacons, LiDAR, RFID tags or WiFi access points, to ease the tasks of everyday life for the VI/blind.

Additionally, during the past few years, the use of computer vision and deep learning in assistive technologies for VI/blind people has noticeably increased in popularity among researchers. These technologies have provided

game-changing possibilities for making more effective and useful assistive tools. In this systematic mapping, a study of 105 recent papers is conducted to give an overview of the state-of-the-art computer vision-based assistive solutions. Moreover, a comparison of these papers is performed to identify advantages and disadvantages and for proposing possible improvements.

2 Related work

This section provides a brief discussion of relevant literature reviews, surveys and research on assistive technologies for the VI/blind. To the best of our knowledge, there are not many thorough literature reviews or systematic mappings about this topic. Nonetheless, there are some articles that to some extent review existing state-of-the-art assistive tools. For example, in one of the most recently published papers, Aileni et al. [2] presented several assistive technologies for people with low visual acuity. They discussed how these technologies could be integrated into wearable systems and transportation systems using the Internet of Mobility (IoM) and Internet of Mobile Things (IoMT). The review is about assistive solutions that are based on vision, audio and tactile senses. However, only a limited number (9 solutions) of

✉ Mohammad Moeen Valipoor
moeen.valipoor@alumnos.upm.es

Angélica de Antonio
angelica@fi.upm.es

¹ Madrid HCI Lab, Universidad Politécnica de Madrid, ETS Ingenieros Informáticos, Madrid, Spain

computer vision-based solutions are mentioned in the paper. The solutions they reviewed mostly utilize computer vision and speech recognition to bring an audio-based augmented reality experience for the users to assist them. Orcam or Horus [3], for instance, are assistive tools that have a camera and voice control to help users detect faces, recognize objects and read texts. An assistive app called Seeing AI [4] was also mentioned. It has functionalities similar to the aforementioned solutions, but using the power of smartphones to help the VI/blind users. Aira [5] is a different kind of assistive solution discussed in the review that provides live access to real agents that can see through the camera of the blind user's smartphone and provide different kinds of assistance, like scene description or text reading.

Chanana et al. [6] conducted a systematic review of assistive technologies for pedestrians with visual impairment that were in the market in 2017. Their work is mainly focused on the systems that use laser, infrared or ultrasonic sensors for obstacle detection. Solutions mentioned in the paper are obstacle detection devices that mount on a white cane. These devices can detect obstacles that are not ground level and warn the user using low and high pitch sounds or haptic feedback on their body. These solutions mostly have the limitation of obstacle detection of up to 1 or 2 meters of distance. Sadly, only 2 assistive devices out of the 9 mentioned in the paper were commercially available at the time of our review. One obstacle detection system mentioned was EYECane [7] that uses computer vision for assistive purposes, but it was at the prototype stage and could not be applied to a real-world environment at the time the paper was published. Two more recent assistive canes are discussed in [8]. WeWalk [9], which can be connected to smartphones via Bluetooth and uses ultrasonic sensors for obstacle detection, and assistive canes that use radio frequency identification (RFID) tags like [10]. However, due to the necessity for RFID tags, their use is confined to small or indoor environments.

Kau and Garg [11] published a systematic review in which they briefly discuss researches about camera-based and sensor-based assistive devices. They point out some of the critical challenges of the solutions like low-light detection, prediction of dynamic obstacles, large size of the prototypes and high costs for creating them. However, their paper lacks an in-depth description of the technologies and does not consider the availability of user evaluations in the studies.

Kuriakose et al. [12] reviewed different modalities for navigation of VI/blind people in different environments. They assessed multimodal systems and their benefits in comparison with unimodal approaches. They also discussed the pros and cons of different modalities like tactile, visual, aural and haptic in navigation systems for VI/blind users. In another review by the same authors about tools for navigation of VI/blind users [13], they suggested some

recommendations to improve these solutions. Their suggestions are: A robust real-time object detection, availability of multiple options for feedback, reducing learning time for the user, portability, privacy, avoiding social stigma and not overloading the user with information.

In a similar study by Romlay et al. [14], the maturity level of various electronic travel aids (ETA) in 70 studies is discussed. However, only one (Guido [15], a smart walker for the blind) of the products mentioned was successful in pre-production unit sales in the USA. Most of the other assistive solutions in their review were tested in a simulated environment without real subjects or constructed hardware.

Finally, Khan and Khusro [16] conducted a review about smartphone-based assistive solutions for VI/blind. They discussed the challenges and issues of different kinds of solutions like vision substitution-based, sensor-based, sonar-based and augmented reality-based interventions. The discussed issues are mostly about the usability of the solutions, for example, inconsistency in interface elements, difficulty of text entry or modification, device incompatibility, illogical order of items and inadequate mapping of feedback with UIs, etc. Additionally, the importance of user-centered design (UCD) for the development of assistive solutions is highlighted, because in this approach the system design begins with the explicit understanding of the user needs, and the user is involved in the process of development and evaluation.

Existing reviews have demonstrated that the availability of assistive solutions for VI/blind users in the market is limited. Moreover, the current available solutions have noticeable problems from the usability point of view.

In our research, we concentrate on aspects of assistive solutions for the VI/blind that have not been thoroughly investigated before. Our focus is on the recent improvements in computer vision methods for scene understanding and the impact they have on assistive solutions for the blind. We also examine the existence of user evaluations for the proposed solutions, as we consider this a critical component in the development of assistive solutions.

3 Systematic mapping study

Following Khan and Khusro [16] we wanted our research to focus on user needs and requirements. Ntakolia et al. [8] provides a list of user requirements after a thorough literature review on studies about requirements elicitation for assistive solutions for VI/blind people. The requirements considered in our research are listed in Appendix 5. Obtaining information from the surroundings is considered as one of the most important requirements for the assistive solutions for VI/blind users, because it helps them to create a more accurate mental map of the environment [17]. Therefore, the

capability of computer vision technologies to analyze and understand a scene has been considered as the core of our mapping study.

The method used for this systematic mapping study is proposed by Petersen et al. [18]. A systematic mapping study's goal is to obtain a detailed overview of a research subject, provide a review of existing literature, identify research gaps and gather evidence for possible research directions [19]. Consequently, the goal of this research is to gather relevant publications that are related to computer vision-driven scene understanding for VI/blind users, and present an overview of the status quo to find the research gaps.

A number of research questions were defined to specify the objective as follows:

1. What are the current computer vision solutions for scene understanding?
2. How are computer vision methods used to assist blind users with their daily activities?
3. How have proposed solutions been evaluated?

After defining the research questions, the process of collecting papers began.

3.1 Search strategy

The major terms used for performing the paper search were "Computer vision," "Visual Impairment" and "Accessibility." Later, by combining these three terms and using similar keywords, the papers were collected. The list of keywords used in the search is listed in Table 1. The scope of the research considered the papers that were published after January of 2017 in journals, academic conferences, workshops and academic books. Web pages, non-academic publications and patents were excluded from the research scope. The search strategy is outlined in Table 2. The reason

Table 1 Search strings

Major terms	Alternative terms
Visual impairment	(Blind people OR visual* impair*) AND
Computer vision	(Computer vision OR visual computing OR object recognition OR image processing OR computational perception) AND
Accessibility	(Accessibility OR scene understanding OR spatial information OR mobility OR assist*)

Table 2 Search strategy

Academic databases	Google Scholar IEEEExplore ACM Digital Library Scopus
Target Items	Conference papers Workshop papers Journal papers Excerpts of academic books
Search applied to	Title Abstract Keywords
Language	English written papers
Publication period	From January 2017 up to June 2021

for choosing the papers that have been published after 2017 was that computer vision methods for object detection have improved drastically since that year and brought significant improvements in the scene understanding research for the VI/blind.

Figure 1 shows the summary of the selection process. The initial number of search results retrieved from all the databases was very large (around 27,000 for papers published after 2017). In order to select suitable papers for the research, the first 50 papers in an ordering based on their relevance and publishing date (most recent) were analyzed according to the exclusion criteria. If more than 10 papers between those 50 were related to the topic, the next 50 papers were also analyzed. This process continued until there were less than 10 papers found related to the research goals. In the application of the exclusion criteria (Table 3), the abstracts, introductions and conclusions of the papers were considered as the main source. In some cases, other parts of the papers were also read to obtain a better comprehension. After removing the duplicates (992 results) and applying the exclusion criteria, 180 papers were left for full text reading, out of which 105 were ultimately useful for the review.

The reason for excluding a large number of papers was that we were focused on the studies that provided a useful and tangible solution. This means that we skipped the papers which were about frameworks or solutions that missed implementation (e.g., prototype, proof of concept (POC) or simulation).

3.2 Data extraction

For the papers that were chosen to be read completely, we needed a template for data extraction so that the comparison and tracking of the information in the papers would be easier. According to the research questions, a number of

Table 3 Paper selection strategy

Exclusion criteria

1. Papers that are not relevant to the primary research goals. This is obtained after reading the abstract, introduction and conclusion of the papers.
3. Papers published in non-academic web pages, personal blogs or patents
4. Papers without providing a solution for obstacle detection and/or object recognition
5. Solutions that miss implementation
6. Not written in English

Fig. 1 Paper selection process**Table 4** Analysis categories

Scene understanding—RQ1	Object recognition Obstacle detection Depth detection Algorithms used Hardware used
Assistance services—RQ2	Type of assistance Modality
Evaluation—RQ3	Technical evaluation User testing evaluation

categories were defined to compare existing solutions. The three major categories are “Scene Understanding,” “Assistance services” and “Evaluation.” “Scene Understanding” is related to the first research question and is focused on the level of the perception that the system has from the surrounding environment of the user. The “Assistance services” category is related to the second research question, and defines how the understanding of the outside world by the system is going to assist VI/blind users. Finally, the “Evaluation” category collects information about the way in which the solution was evaluated.

In Table 4 the main categories and subcategories are listed.

3.2.1 Scene understanding

Object recognition This is one of the most important features of the assistive solutions for scene understanding. There are different methods for recognizing an object using computer vision and each of them has its advantages and disadvantages. In this category, we check the availability of object recognition in each solution.

Obstacle detection Obstacle detection is another primary feature that must be included in order to warn users and avoid obstacles in the environment. In this category, we analyze the approach of each solution for detecting obstacles, either using sensors and/or cameras.

Depth detection This is one of the most challenging aspects in the development of assistive solutions for scene understanding. Estimating in real time the 3D location of objects in the physical world using computer vision is a

complicated task. In this category, we check the availability of depth detection in the solutions.

Algorithms used There are different algorithms that can be used for object recognition, obstacle detection or depth detection. The algorithms used in each solution are analyzed in this category

Hardware used This category focuses on the hardware components that were used in each solution. Different researchers made use of various kinds of devices depending on the budget and purpose of their project.

3.2.2 Assistance services

Type of assistance In this category, we summarize how each solution provides assistance for the user and in which kind of tasks they can help them.

Modality This category considers the methods used in the system for interacting with the users. For example, whether they provide text to speech, a pitch sound, speech recognition, haptic feedback, etc.

3.2.3 Evaluation

Technical evaluation This kind of evaluation is focused on analyzing the solutions from an objective point of view in order to test the performance of the algorithms used. The metrics used for evaluation and the environment in which the evaluation was performed are reviewed.

User testing Testing a system with end users is essential. In this category, we describe if and how the user testing of each solution was performed.

4 Mapping results

The process of gathering data from papers was based on the categories defined in Table 4. Figure 2 shows the distribution of papers published in each year. The number of papers published in 2020 that meet our research goal is almost four times higher than that of 2017. We did not include the number of relevant papers published in 2021 in Fig. 2 because our mapping did not cover the whole year. The constant increase in the number of published papers proves that the topic has been getting more attention during the last years. This is mainly because of the improvements in deep learning algorithms, computer vision cloud services and mobile devices.

4.1 Scene understanding

Scene understanding for the visually impaired/blind has some differences with the classical approach of scene understanding. It is very important that the process of analyzing

Fig. 2 Paper distribution

and perceiving the environment by the system occurs in a level that can be beneficial for the VI/blind users. For instance, it is crucial that the algorithms have sufficient swiftness and accuracy in detecting/recognizing obstacles and objects to give prompt feedback to the user when it is necessary. Additionally, semantic understanding of the environment by the system and finding the relations between different objects in a scene are important to give a comprehensible description of the environment to a user that has no access to visual cues.

After comparing the various research approaches, we came to the conclusion that in the early years, most of the solutions were focused on obstacle detection or image enhancement techniques. Image processing was used in order to make the images perceivable for the visually impaired people. For instance, in [20–22] researchers used techniques like contrast enhancement, image mapping and magnification. Their aim was to increase the visibility of the important features of an image (e.g., edges). However, these methods had some limitations. For example, the algorithms added too much noise to the image or amplified the contrast of some parts of the image that were not necessary for scene understanding. Lately, object recognition, which is a more efficient method for scene understanding, has become more popular thanks to technological advancements (Fig. 3). The percentage of solutions with object recognition has been increasing in the last years, as shown in Fig. 4.

4.1.1 Obstacle detection

For detecting obstacles in the environment, there are usually two different approaches: distance sensors (e.g., ultrasonic, LiDAR and infrared (IR) triangulation) or camera-based (e.g., monocular or RGB-D cameras) techniques. In some cases, the combination of both techniques is used for better accuracy. Detecting the exact distance of an object/obstacle

Fig. 3 Solution approaches during the past 4.5 years

Fig. 4 Frequency of object recognition in the solutions

from the user is one of the complications of creating assistive solutions for blind people. This is because the 3D location of an object in the real world is often inferred from a 2D image taken by a monocular camera. Ultrasonic sensors, point clouds, RGB-D images (taken by stereo vision cameras) or mathematical estimations are the solutions that have been used for tackling this problem. In this section, different approaches are discussed.

Camera-based techniques: Researchers in [23, 24] applied Stixel-World [25], which is a method that is mostly used in autonomous cars, to help VI/blind people navigate in an environment. Stixels algorithm provides environmental awareness based on the depth images provided by an RGB-D camera. RGB-D images are captured using cameras that work in association with sensors for distance detection. Stixels segment objects in the image around the user in vertical regions according to their depth disparity in the environment. Afterward, using object recognition techniques, Stixels semantically categorize objects in the scene.

On the other hand, researchers in [26] used Apple's ARKit 2 [27] framework to find the 3D location of obstacles

using planes and point clouds. ARKit 2 can detect the vertical and horizontal planes. Therefore, it can differentiate the ground from the other planes that could be potential obstacles. Additionally, point clouds (which are a set of points that represent the objects' salient contours in an image) can be used to detect obstacles that have more complicated shapes. These high level computer vision features of augmented reality enable the possibility of converting 2D points provided by phone cameras to 3D position in order to estimate the approximate position of the obstacles.

It is also possible to mathematically calculate the distance of an object in front of the camera. Lin et al. [28] used a method proposed by Davison et al. [29] for object distance estimation (Equation 1).

$$Z = \frac{f * k_v * h}{v - v_0} \quad (1)$$

In this formula, f is the camera focus, k_v is the pixel density (pixel/meter), h is the camera height from the ground, v_0 is the center coordinate of the formed image, v is the distance from the camera to the target object's ground coordinates, and Z is the distance between the target object and camera location. In [28], they claimed that this method has a high accuracy in distance detection and can detect an object's distance up to more than 10 meters.

The solution in [30] is based on another mathematical method that uses depth images and fuzzy control logic for the approximate measurement of obstacles' distance. This solution divides the frame into three parts (right, left and center), categorizes the location of obstacles in three different categories and provides audio feedback for the user according to them. If the user faces any obstacles, the system makes decisions in order to avoid them based on 18 different fuzzy navigation rules that depend on the location and distance of the obstacles.

Besides the mentioned techniques, it is also possible to estimate the depth from monocular images using deep learning techniques. Recently, this approach has received more attention due to the rapid advancements of deep learning methods. Facil et al. [31] built a depth prediction network that provides a depth map from a single RGB image. Their predictions work with images taken with different camera models. Various kinds of NNs (Neural Networks), like CNNs [32] and RNNs [33], have been implemented showing the effectiveness of monocular depth estimation. In [34], a CNN network is used for calculating the distance of the obstacles. Their method works more accurate than some devices like Kinect, according to their comparisons. We did not come across many assistive solutions that make use of these techniques, but they will surely be applied more in the future development of assistive solutions for the blind since it is a cost-effective and at the same time robust approach.

Distance sensor-based techniques: Using sensors has been a more common approach for obstacle detection in comparison with camera-based techniques. Among the different kinds of sensors, ultrasonic is very popular. This is because of their accuracy, low cost, low power consumption and ease of use [35].

Ultrasonic sensors have a transmitter that generates sound waves with a frequency that is too high for human ears to hear. Then, the receiver of the sensor waits for the rebound of the sound and, based on that, the distance with the obstacle will be calculated. These sensors are better at detecting transparent objects compared to light-based sensors or radars. As an example, Bharatia et al. [36] used ultrasonic sensors for the detection of knee level and low-lying obstacles. One of the drawbacks of using sensors is that they cover a short range and can only detect close obstacles, which makes them more suitable for indoor environments.

Combined distance sensor and camera-based techniques: Lately some researchers have combined these two approaches. Hakim and Fadhil [35] made use of ultrasonic sensors and RGB-D cameras for obstacle detection. Their electronic travel aid (ETA) processes the data received from an RGB-D camera using a Raspberry pi 3 B+ which has ultrasonic sensors attached to it for distance detection. The combination of these two approaches provides a more accurate obstacle detection.

Appendix 1 contains the detail of the different obstacle detection techniques that were used in the reviewed papers.

4.2 Object recognition

During the past years, the use of deep neural networks (DNN) for object detection, especially the latest CNN models such as ResNet [37] or GoogLeNets [38], has considerably extended the potential of computer vision for developing assistive solutions for the VI/blind. They have a notably superior performance that makes real-time object recognition more achievable in comparison with shallow networks such as AlexNet [39]. In addition, CNNs can learn high level semantic features from the input data automatically, optimize multiple tasks simultaneously such as bounding box regression and image classification, and solve some of the classical challenges of computer vision [40].

There are different ways to implement an object recognition solution. One common approach is to execute the process of recognition remotely on cloud services like Google Cloud Vision, Microsoft Azure Computer Vision, Amazon Rekognition [41], etc. These services are already trained with huge datasets that enable improved performance. Bharatia et al. [36] proposed a mobile system that uses Google Cloud Vision to recognize objects, texts and faces. There are also solutions that provide their own cloud computing algorithms for image processing. For instance,

researchers in [42] made a remote object detector using an improved version of ResNet [43] network.

These services mostly have a Representational State Transfer (REST) Application Programming Interface (API) to handle communication between the client and the server. Companies that provide these services usually calculate the costs based on the number of requests sent to the server by the client.

Local image processing is another approach used in many solutions which undertakes the computations related to the object recognition on the client side. Nevertheless, this approach is usually confined to a limited number of objects due to hardware limitations. Dosi et al. [44] made an android app for VI people that locally recognizes objects. They used MobileNets [45] for object recognition which is a neural network for mobile and embedded vision applications. Single Shot Detector (SSD) [46] with MobileNets architecture is another popular algorithm which is used in a considerable number of papers for object recognition and can bring fast and efficient results. SSD can detect multiple objects in an image by taking a single shot. Visual Geometry Group (VGG16) [47], which is a convolutional neural network model, is the base network of the SSD algorithm, followed by a multi-scale feature layer for object category and bounding box predictions. SSD generates anchor boxes in various sizes and predicts objects based on their size. Larger objects are detected by deeper network features and smaller ones are detected by the shallower networks. The inference time of the SSD512 method is 22 milliseconds with about 76.8% Minimum Average Precision (mAP) on Pascal VOC2007 dataset of images, which shows its competence and swiftness in object detection [46].

YOLO [48] is a CNN-based object detection technique and uses Darknet which is an open source network framework written in C and CUDA. YOLO divides an image into $S \times S$ grids and generates B bounding boxes for each of them. Afterward, it predicts the probability of classes for objects and their corresponding bounding boxes. In [48], YOLO VGG-16 has an inference time of 47 milliseconds with 66.4% mAP on Pascal VOC2007 dataset of images, which is close to the SSD performance.

Different versions (YOLO v3 (Tiny) [49], YOLO v2 [50], YOLO 9000 [51]) were used in different researches. The accuracy and number of objects that could be detected varies for each version. For instance, TinyYolo is made for mobile devices and is able to recognize a lower number of objects compared with the other versions. YOLO and SSD are very popular because they are achieving a balance between accuracy and speed.

Figure 5 shows the distribution of object recognition algorithms in the reviewed solutions. As it is shown in the figure, YOLO is the most popular solution for object recognition. The remaining “Other Neural Networks” algorithms which

Fig. 5 Distribution of cloud and local object recognition services

are used in the solutions are various local object recognition methods that are using Inception-v3 [52], stochastic gradient descent (SGD) algorithms on Keras [43, 53], OpenCV [54] functions [55] or Computer Vision System Toolbox of MATLAB [56, 57], to name a few. Appendix 2 contains the object detection methods used in the reviewed solutions.

It is important to mention that the quality of the input images sent to these algorithms can affect their performance noticeably. For example, images taken at night or in low light conditions can have a high noise level or distortion that can reduce the accuracy of object detection algorithms. To overcome this problem, there are different methods. For instance, [35, 58, 59] used a Gaussian filter that blurs the image in order to remove the noise and unnecessary details in the images before sending them to the algorithm. Researchers in [42] used the stereo-image quality assessment (SIQA) approach to collect images with higher quality before sending them as input to the system. They used a disparity map and the traditional approach of Hough transform for image evaluation. IQA evaluation methods are divided into two main categories, deep learning methods such as CNNs [60], and traditional feature extraction methods that help detecting different parts of the desired objects in an image [61]. These two methods are used to make a score prediction for the image quality assessment.

5 Assistance services

It is crucial to consider how the information regarding the environment obtained by sensors, cameras and algorithms is transferred to the user. The assistance provided by these solutions should be swift, accurate and easily understandable for VI/blind people. The assistive solutions reviewed in this paper help users in different tasks relating to their daily life.

The modality of these assistive solutions is mostly based on audio and tactile feedback. Researchers in [50, 51, 62, 63] used binaural audio for scene description. This brings a sense of audio-based augmented reality to the user. Users can hear and feel the approximate 3D location of the object/obstacle based on the audio they hear using normal/bone conduction headphones. In these solutions, the device/app names all the objects in the scene or a specific object that the user is looking for. Moreover, some solutions provide vibrotactile feedback. Saurav and Niketa Gandhi [64] used two servo motors for vibrating feedback so that when there is an obstacle on the left, the left side of the user vibrates and when there is an obstacle on the right, the right side vibrates. In another research [26], the user is warned about the obstacles with an audio beep. The pitch of the sound changes based on the size and proximity of the object.

There are solutions that carry out the scene description for a specific purpose. These solutions might be limited to certain tasks but the overall performance is better because the scope is constrained. Researchers in [43] detect stop lights and crosswalks to help users with crossing streets.

Many solutions provide navigation assistance for users. For instance, [62] provides a navigation service that guides the user through a pre-scanned environment. A virtual assistant repeatedly states “follow me” and, based on the intensity and the direction of the sound in the headset, the user navigates through the environment.

Besides the scene understanding feature, emergency calls are provided in some solutions which can be very useful for the users. Suresh et al. [65] implemented a speech recognition module that can get orders from users to make an emergency call to predefined users or the closest emergency center based on the location of the user. GPS was used for tracking the live location of the user. Researchers in [36, 64, 66] undertook a similar approach by placing a call button on the assistance device.

Arakeri et al. [67] proposed a text recognition module in their solution that reads the text that is in front of the user. They used Google Vision API [68] for this purpose. The same functionality has been provided in [69] using Tesseract OCR library [70] and in [71] using Microsoft’s Computer Vision API [72]. However, the usage of this feature can be confusing for the user since it is not possible for a VI/blind user to distinguish the exact position of a text in the real environment.

Rahman et al. [73] have a face detection module in their solution besides object recognition. They used Multi-task Cascaded Neural Networking (MTCNN) that can detect faces with a 70–100% accuracy.

The kinds of assistance provided in the revised solutions are listed in Appendix 3.

5.1 Context of use: the ICF framework

In order to come up with novel and useful solutions for any kind of disability, it would be necessary to understand the contexts in which they could be applied and the scope of limitations that a disabled person faces. The International Classification of Functioning, Disability and Health (ICF) framework [74] provides a standard language for defining different kinds of disabilities. In the framework, limitations of disabled people in their activities are divided into different categories. In our research, we tried to assess the usefulness of the proposed solutions to VI/blind users by mapping their contributions with the different tasks in this framework. In the case of scene understanding for VI/blind people, the main relevant categories are as follows:

- *Mobility* This is mainly about moving the body and going from one position to another. According to the framework, it includes tasks like “walking and moving,” “changing and maintaining body position” and “moving and handling objects.”
- *Self-care* This includes tasks like washing oneself, caring for body parts, toileting, eating, dressing, drinking and looking after one’s health.
- *Domestic life* It covers daily tasks related to acquisition of necessities, household tasks, caring for household objects and assisting others.
- *Interpersonal relations and relationships* This category consists of general interpersonal interactions and particular interpersonal relationship challenges that a disabled person may have with other people.

We analyzed the compatibility of the above mentioned ICF categories with the solutions provided by different researchers. Mobility is the most explored category, with researches in [62, 75, 76] helping VI/blind with navigating in different environments (indoor/outdoor). They combine various methods like GPS, obstacle detection and object detection to help the VI/blind in tasks like “walking and moving” from one location to another. Moreover, [57] defines two user stories that are compatible with the “Domestic Life” and “Self-care” categories. In the first user story, the blind user receives assistance for detecting the right kind of pasta at home, which helps her with cooking and eating. In the second user story, a man wants to buy a specific kind of biscuits and using the assistive device he can find them in the store. Their solution provides feedback about the differences in objects that have the same tactile appearance.

By comparing the categories of the ICF framework with the kinds of assistance provided in the revised solutions, we noticed that their scope is generally not well defined in terms of the context of use. For instance, in papers like [50, 65, 75] object detection could potentially cover some tasks

in “Self-care” or “Domestic life.” However, these specific use cases are not mentioned. Furthermore, researchers in [73, 77, 78] provide face detection, a service that could be related to the “Interpersonal Relations and Relationships” category, but the final purpose is not clear. It appears that the researchers have focused their efforts more in proving the technical feasibility and performance of the solutions than in demonstrating how the solutions can help VI/blind people in their daily life. This fact becomes more evident when we analyze the way in which the solutions have been evaluated.

6 Evaluation

Based on our review, the evaluation process of an assistive solution should tackle two main aspects. One is the technical evaluation and validity assessment of the system from a technical point of view, and the other is the testing of the system with the target end users to evaluate the performance and usefulness of the solutions.

Despite the fact that technical evaluation is a matter of importance, user testing is equally essential. This is because the ultimate goal of any assistive solution is to be useful for VI/blind people. Unfortunately, user testing is neglected in a considerable number of papers. Figure 6 shows the percentage of papers that only undertook the technical evaluation, and the ones that performed both.

6.1 Technical evaluation

In the development process of any system, it is crucial to test and measure its performance with objective metrics. In the case of assistive solutions, it is essential to measure the accuracy and efficiency of algorithms in the detection and recognition of obstacles and objects. It is common to evaluate the performance of object detection algorithms using the calculation of precision, recall and mAP (the mean of average precision calculated out of precision and recall metrics).

Fig. 6 Evaluation types

A model with high precision returns more correctly predicted results than irrelevant ones and a high recall means that the model returns most of the relevant results. In other words, precision is a measure for quality, while recall is a measure for quantity. Some papers, such as [59] and [26], measured the accuracy of their approach based on recall and precision. Others, like [43, 76, 79], used mAP to measure the effectiveness and accuracy of the algorithm they used.

Additionally, other works [80–82] tried to compare their solution with other state-of-the-art solutions based on the accuracy, number of detected objects, kind of recognition (object, text, face, obstacle, etc.), average distance, convenience and so on, in order to assess the functionality of their solution and ascertain the competence of their approach.

In technical evaluation, researchers usually test their system by simulating a user scenario. For instance, Wang et al. [24] used a prerecorded video to test their Stixels model.

6.2 User testing

The most important aspect in the evaluation of a solution for VI/blind people is to test it with the target users. This is because assistive solutions are for a target group that is different from average users. A person without the disability cannot evaluate the solution properly, given that, due to the variation in sensory input, they do not possess the same mental models of the environment and qualities of embodied experience as people that genuinely have visual impairment. Sadly, this critical point is neglected in a noticeable number of research projects. Up to 27.5% of the papers that were included in this review were testing their solutions with blindfolded/sighted users. Figure 7 summarizes the visual perception status of the testers in evaluations.

Nonetheless, there are some works that report testing with VI/blind users. Wang et al. [83] performed 100 hours of testing for the navigation module of their solution with simple and complex paths. Afterward, they asked the 5

blind testers to fill in questionnaires regarding the comfortability and effectiveness of the system. In another research [26], the solution was tested with 13 participants that were VI/blind. They conducted 15 minutes test sessions and asked users to fill in a Likert-like scale questionnaire. Furthermore, [84] reports a detailed evaluation and found that there is a significant difference between early blind and late blind testers, and that the former group could perform better. The authors also came to the conclusion that vibrotactile cues are less efficient in comparison with auditory cues for detection in the central region of the environment.

The research of Guerreiro et al. [85], which presents a navigation robot for the VI/blind, also includes a satisfactory user study. They evaluated their solution with 10 blind participants. The tasks for testing are well designed and explained in the paper. After the testing process, user feedback was obtained about confidence, safety and trust in the solution using questionnaires.

The results obtained from questionnaires are subjective and contain personal opinions of the users. In some cases, personal opinions can be considered unreliable when they are obtained from blindfolded users instead of blind/VI users. For example, in [23] researchers used the NASA-TLX [86] evaluation questionnaire that analyzes tasks based on Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. They noticed that users reported unexpected low scores on physical demand. This is because they were blindfolded and the tasks appeared to be more frustrating for someone with normal vision in comparison with a VI/blind user.

The main methods used for testing the solutions were as follows:

- Surveys: Asking a series of questions from the users, usually with Likert-type scales, to obtain their feedback;
- Think-aloud protocol: Users share their opinion about the solution while performing the test tasks;
- Controlled environment testing: Testing the solution in a laboratory environment that was designed by the researchers and observing the user's behavior to detect the advantages and problems of the prototype;
- Field experiments: Testing the solution in real-world settings with the subjects. In this type of testing, users might make unexpected decisions which help to find out the scenarios that were not considered by the researchers;
- Remote usability testing: Users are not directly observed while using the assistive solution. Data are gathered and then later analyzed by the researchers;
- Interviews: Users are interviewed to share their opinion about the experience of using the solution and its pros and cons.

Fig. 7 Evaluation approach

Each of these methods have their drawbacks. According to [87] representative surveys cost money and time, the think-aloud protocol is not accurate because the environment is not natural to the user and the tasks are usually performed in a controlled environment; field experiments may not represent the correct population; remote testing needs additional tools for collecting data; and interviews do not sufficiently cover usability issues. Additionally, controlled environment testing might not consider some factors that exist in the real environment which may affect the user's experience.

Appendix 4 details how user testing was performed in the selected papers. The ones, that are not included, only had performed a technical evaluation.

7 Conclusions and discussion

In this systematic mapping study, a selection of published papers during the past four and a half years related to computer vision-driven scene understanding for VI/blind people were reviewed. They provide various assistive services for scene understanding, like obstacle detection and object recognition. Obstacle detection can be performed using sensors, cameras or both. Sensors can detect obstacles in shorter distances in comparison with cameras, but they are easier to use and can perform better in some cases. Therefore, some researchers prefer to combine these two approaches.

For object recognition, there are two main ways to process the data. Some of them undertake the process of computing remotely and others locally. Cloud services and remote servers can perform heavy computations and recognize a wider range of objects. However, they require to be constantly connected to the network that manages the computing. This can also cause security problems that should be taken into consideration. Local computing does not have those problems but it offers limited capabilities.

We should highlight that technologies in this field have improved a lot in the past few years. Computer vision is thriving with the appearance of object recognition methods like YOLO, and more advanced solutions have been developed recently that provide higher accuracy and performance.

However, despite the fact that these solutions help users overcome social barriers and give them more independence in life, computer vision models rely on camera input which could threaten the privacy of users and surrounding people. One of the greatest risks is that the collected data get misused, especially in solutions that rely on remote servers for computation instead of the users' devices. Some studies show that there is a trade-off between the provided services and the privacy costs, and that some users are willing to accept the privacy costs in exchange for the service they receive [88]. Lee et al. [89] conducted a study about the social acceptance of assistive solutions

for the blind from the perspective of both blind users and bystanders. They concluded that a considerable amount of people in the society are still not very comfortable to be exposed to these devices, specially if they include a camera. Their results indicate that a thorough evaluation in a real environment is needed to evaluate the social acceptance of assistive solutions and the needs of people who are exposed to the technology. The study of Akter et al. [90] shares a similar point of view. Both sighted and non-sighted users in their study were concerned about their privacy and the accuracy of the information provided by the assistive solution.

Another important issue that came up in some of the reviewed solutions is the exclusion of VI/blind people in the evaluation process, an aspect that may be critical for the adoption of the technology.

Verza et al. [91] suggest that some of the reasons for abandoning assistive technologies are neglecting users' opinions in the development process, inefficiency of devices and insufficient training of the user. Furthermore, a research by Phillips and Zhao [92] identified four factors related to the abandonment of assistive devices. They noted that change in user needs and priorities in time is one of the main factors in device abandonment. For instance, according to [93], some changes in VI users, like worsening eye condition due to macular degeneration, can imply a significant change in user needs. Other abandonment reasons mentioned include not considering users' opinion, ease of device procurement and poor device performance.

Additionally, testing of the systems should be performed in a real-world scenario to assess if they are usable out of the laboratory's controlled environment. The final goal of an assistive solution is to improve the lives of disabled end users. However, if the usability or performance of the system is not properly tested in the real context of use, it can end up causing negative impacts on the target end user. Our study suggests that researchers in this field should pay more attention to the VI/blind user needs and the applicability of the solutions. Many of the papers reviewed in our systematic mapping had insufficient/no data regarding the target context of use for the proposed solutions. Consequently, it was not possible to assess the compatibility level of their solutions with the ICF categories and their specific tasks, such as dressing, eating, drinking, preparing meals, acquisition of necessities and so on. There are solutions that prove to be able to detect persons, objects or obstacles, but their expected benefits in the end users' life are vague. This situation raises important generalizability concerns, given that a solution just tested in a toy or simulated scenario might not have the same effectiveness in different use cases or scenarios. The cost and effort associated with the adaptation of a specific solution for its application in a different context is generally overlooked.

Finally, the way that information is represented and delivered to the user is crucial. Users should be able to comprehend the information provided by the solution without complications. In most of the solutions, binaural audio, vibrotactile feedback and basic audio instructions were used for this purpose. However, these might not be the best approaches. For example, binaural audio can approximately indicate the location of an object or an obstacle, but still it is not very accurate.

All mentioned issues may have contributed to the abandonment of existing commercial assistive solutions, leading to the unavailability of them in the market. Paying more attention to the ICF framework, user requirement studies and the UCD methodology would be desirable for developing more usable, useful and better accepted assistive solutions. Besides that, learning from some of the success stories of products like WeWalk [9] that was designed by a visually impaired person known as Kursat Ceylan, who was familiar with the needs of blind people, can lead us to the development of more useful assistive devices.

This study highlighted the main challenges in this rapidly evolving field and compared different approaches of computer vision-based assistive solutions in recent years. The results of this systematic mapping can be valuable for researchers who want to develop more useful and effective assistive solutions for VI/blind people in the future.

Appendix 1: Techniques used for obstacle detection

Article	Distance sensor-based	Camera-based	Detection approach
Akula et al. [94]	✓		Ultrasonic sensor
Arakeri et al. [67]	✓		Ultrasonic sensor
Bharatia et al. [36]	✓		Ultrasonic sensor
Breve and Fischer [95]		✓	A deep learning model for detecting obstacles
Canez et al. [59]		✓	Time To Collision (TTC) technique
Caraiman et al. [69]	✓	✓	3D reconstruction of the sensed environment using IR-based depth sensor and a stereo vision system
Chen et al. [96]		✓	RGB-D camera

Article	Distance sensor-based	Camera-based	Detection approach
Cheng et al. [97]		✓	Intel RealSense D435 depth camera
Dasila et al. [63]		✓	Cloud points
dos Santos et al. [98]	✓		Ultrasonic sensor
Elmannai and Elleithy [30]		✓	Using fuzzy logic
Endo et al. [99]		✓	SLAM
Gandhi and Gandhi [64]		✓	Pixy Cam
Gay et al. [100]		✓	ArUco Marker [101] and OpenCV [54]
Guerreiro et al. [85]	✓	✓	LiDAR sensor and ZED RGB-D camera
Hakim and Fadhil [35]	✓	✓	Otsu's threshold algorithm with ultrasonic sensor and RGB-D camera
Hsieh et al. [34]		✓	Using a CNN network
Huppert et al. [102]		✓	OptiTrack [103] environment with 12 cameras
Hussain et al. [104]		✓	Intel RealSense D435 depth camera
Islam et al. [105]		✓	Not mentioned in detail
Islam et al. [106]	✓		Ultrasonic sensor
Joshi et al. [81]	✓		Ultrasonic sensor
Kayukawa et al. [107]		✓	iPhone 11 pro camera
Kayukawa et al. [108]		✓	ZED RGB-D camera
Kayukawa et al. [109]	✓	✓	LiDAR sensor and RGB-D camera
Khairnar et al. [110]	✓		Ultrasonic sensor
Kim et al. [75]		✓	Distance estimation using a stereo camera
Kommey et al. [51]		✓	Based on the size of bounding boxes
Kuribayashi et al. [111]	✓	✓	LiDAR sensor and iPhone 11 pro camera
Li et al. [43]		✓	ZED RGB-D camera

Article	Distance sensor-based	Camera-based	Detection approach
Li et al. [28]		✓	Mathematical depth calculation
Li et al. [62]	✓	✓	Create 3D map using Infrared sensors
Martinez et al. [23]		✓	Stixels
Megalingam et al. [112]	✓		Ultrasonic and Infrared sensors
Ntakolia et al. [8]		✓	Intel RealSense D435 depth camera
Pehlivan et al. [76]		✓	Mathematical depth calculation
Presti et al. [26]		✓	ARKit 2
Rahman et al. [113]		✓	Mathematical depth calculation
Saha et al. [114]		✓	Sensor fusion technique
Silva and Wimalaratne [115]	✓	✓	Using sonar and vision sensors
Suny et al. [116]	✓		Ultrasonic sensor
Suresh et al. [65]	✓		Ultrasonic Sensors
Tapu et al. [117]		✓	Point extraction method
Tepelea et al. [55]		✓	Mathematical depth calculation
van Erp et al. [118]		✓	RGB-D camera
Vyavahare and Habeeb [66]	✓	✓	Sensors and camera
Wang et al. [83]	✓	✓	Point clouds and sensors
Wang et al. [119]	✓		LiDAR sensor
Wang et al. [24]		✓	RGB-D camera
Zeng et al. [120]		✓	3D Time-of-Flight (ToF) depth camera
Zhang and Ye [58]		✓	SLAM and SIFT
Zhao et al. [121]		✓	RGB-D using ZED stereo camera

Appendix 2: Object detection approach

Object detection method	Paper
YOLO Variations	Liu et al. 2018 [62], Rahman et al. 2019 [73], Kim et al. 2019 [75] Eckert et al. 2018 [50], Duman et al. 2019 [49] Kommey et al. 2019 [51], Lin et al. 2017 [28] Tapu et al. 2017 [122], Zhao et al. 2019 [121] Guerreiro et al. 2019 [85], Iwamura et al. 2020 [123] Yohannes et al. 2020 [124], Martinez et al. 2020 [125] Mandhala et al. 2020 [126], Abraham et al. 2020 [127] Aralikatti et al. 2020 [78], Vaidya et al. 2020 [128] Joshi et al. 2020 [81], Kandoth et al. 2020 [129] Vaidya et al. 2020 [130], Kuribayashi et al. 2021 [111] Shen et al. 2020 [131], Suny et al. 2020 [116] Kayukawa et al. 2020 [108]
SSD Variations	Wang et al. 2018 [132], Suresh et al. 2019 [65], Pehlivan et al. 2019 [76], Gianani et al. 2018 [133], Hussain et al. 2020 [104], Nguyen et al. 2020 [134], Bhole and Dhok 2020 [79], Chen et al. 2020 [135], Khairnar et al. 2020 [110], Shah et al. 2020 [136], Boldu et al. 2020 [137], Tahoun et al. 2019 [138], Akkapusit and Ko 2021 [139]
Microsoft Azure Computer Vision	Baskaran et al. 2019 [140], Saha et al. 2019 [114], Thomas et al. 2019 [71], Vyava-hare and Habeeb 2018 [66]
Google Cloud Vision	Thomas et al. 2019 [71], Silva and Wimalaratne 2017 [115], Dasila et al. 2017 [63], Arakeri et al. 2018 [67], Bharatia et al. 2019 [36]
Clarifi [141]	Thomas et al. 2019 [71]
Cloud Sight API [142]	Thomas et al. 2019 [71]
SGD Algorithm on Keras	Li et al. 2019 [43]
OpenCV	Tepelea et al. 2019 [55]
Computer Vision System Toolbox2 from MATLAB	Sosa-García and Odone 2017 [57]
VGGNet	Afif et al. 2020 [143]
VGG16	Alhichri et al. 2020[77]
AlexNet	Shelton and Ogunfunmi 2020 [144]
TensorFlow API [145]	Suresh et al. 2020 [146], Islam et al. 2020 [106], Eskicioglu et al. 2020 [147]
Quadratic Discriminant	Imtiaz et al. 2020 [148]
MobileNet v2	Cheng et al. 2020 [97]
ResNet	Georgiadis et al. 2019[149]

Object detection method	Paper
Local Binary Pattern Histogram (LBPH)	Sarwar et al. 2021 [150]
Multi-label Support Vector Machines (M-CSVM)	Bazi et al. 2019 [151]
Support Vector Machines (SVM)	Kedia et al. 2017 [152], Zeng et al. 2017 [120]
Inception-v3 [52]	Ahmetovic et al. 2020 [153]
CNNdroid [154]	Awad et al. 2018 [155]
N/A ^a	Stearns and Thieme 2018 [156], Islam et al. 2018 [105], Hudec and Smutny 2018[157]

^aNot available/mentioned

Appendix 3: Assistance services

Kind of assistance	Paper	Year
Obstacle detection	Martinez et al. [23], Caraiman et al. [69], Endo et al. Tapu et al., [99], Tapu et al., [117], Zhang and Ye [58], Srinivasan et al. [158], Malek et al. [80], van Erp et al. [118]	2017
	Wang et al. [24], Gandhi and Gandhi [64], Elmannai and Elleithy[30],	2018
	Hakim and Fadhil [35], Presti et al. [26], Rahman et al. [113], Canez et al. [59], Zhao et al. [121], Saha et al. [114], Guerreiro et al. [85], Li et al. [159], Megalingam et al. [112]	2019
	dos Santos et al. [98], Ntakolia et al. [8], Breve and Fischer [95], Hussain et al. [104], Khairnar et al. [110], Joshi et al. [81], Srinivasan et al. [158], Kandoth et al. [129], Calabrese et al. [82], Hsieh et al. [34], Gay et al. [100], Kayukawa et al. [107], Cheng et al. [97], Suny et al. [116], Islam et al. [106], Kayukawa et al. [108], Akula et al. [94], Kayukawa et al. [109]	2020
	Wang et al. [119], Chen et al. [96], Huppert et al. [102], Kuribayashi et al. [111]	2021

Kind of assistance	Paper	Year
Object Detection	Sosa-García and Odone [57], Lin et al. [28], Wang et al. [83], Silva and Wimalaratne [115], Tapu et al. [122], Dasila et al. [63], Rizzo et al. [160], Malek et al. [80], Kedia et al. [152], Jabnoun et al. [161], Malek et al.[80],	2017
	Wang et al. [132], Jiang et al. [42], Gianani et al. [133], Dosi et al. [44], Eckert et al. [50], Liu et al. [62], Vyavahare and Habeeb [66], Hudec and Smutny [157], Mante and Weiland [84], Islam et al. [105], Arakeri et al. [67], Awad et al.[155]	2018
	Baskaran et al. [140], Rahman et al. [73], Duman et al. [49], Thomas et al. [71], Li et al. [43], Kim et al. [75], Kommey et al. [51], Suresh et al. [65], Tepelea et al. [55], Bharatia et al. [36], Guerreiro et al. [85], Bazi et al. [151], Georgiadis et al. [149], Tahoun et al. [138], Saha et al. [114], Pehlivan et al. [76]	2019
	Afif et al. [143], Ghosh et al. [162], Afif et al. [125], Yohannes et al. [124], Abraham et al. [127], Suresh et al. [146], Vaidya et al. [134], Nguyen et al. [135], Chen et al. [128], Calabrese et al. [82], Shen et al. [131], Islam et al. [106], Shah et al. [136], Eskicioglu et al. [147], Fusco and Coughlan [163], Iwamura et al. [123], Ntakolia et al. [8], Alhichri et al. [77], Shelton and Ogunfunmi [144], Mandhala et al. [126], Hussain et al. [104], Aralikatti et al. [78], Bhole and Dhok [79], Khairnar et al. [110], Joshi et al. [81], Srinivasan et al. [158], Vaidya et al. [130], Imtiaz et al. [148], Cheng et al. [97], Suny et al. [116], Kayukawa et al. [108], Ahmetovic et al. [153], Boldu et al.[137]	2020
	Sarwar et al. [150], Kuribayashi et al. [111], Akkapu-sit and KO [139]	2021

Kind of assistance	Paper	Year	Article	Testing type	Subjects	Number of subjects	VI/ Blind testing
Navigation	Endo et al. 2017 [99]	2017					
	Li et al. 2018 [62], Elmannai and Elleithy [30]	2018					
	Rahman et al. 2019 [113], Li et al. 2019 [159], Bazi et al. 2019 [151]	2019	Akkapusi and Ko [139]	Controlled environment testing, Survey (Likert-like scale questions), Interview	Blind-folded	20 (14 male and 6 female)	✓
	Khairnar et al. [110], Suresh et al. [146], Alhichri et al. [77], Suresh et al. [146], Gay et al. [100], Kayukawa et al. [108], Eskicioglu et al. [147], Fusco and Coughlan [163]	2020	Awad et al. [155]	Controlled environment testing, Survey (Likert-like scale questions)	Visually impaired	10	✓
	Chen et al. [96], Wang et al. [119]	2021	Boldu et al. [137]	Controlled environment testing, Survey (Likert-like scale questions)	Blind	9 (8 male and 2 female)	✓
Face detection	Kedia et al. 2017 [152]	2017					
	Hudec and Smutny 2018 [157]	2018	Canez et al. [59]	Controlled environment testing	N/A ^a	N/A	✓
	Rahman et al. 2019 [73], Aralikatti et al. 2020 [78], Alhichri et al. 2020 [77], Shah et al. 2020 [136], Ahmed et al. 2020 [164]	2019, 2020	Caraiman et al. [69]	Personal interview and survey	Visually impaired and blind-folded	24 (12 VI and 12 Sighted)	✓
	Sarwar et al. 2021 [150], Sarwar et al. 2021 [150]	2021	dos Santos et al. [98]	Field experiments and Survey	Blind-folded, Blind	41 (10 blind and 31 blind-folded)	✓
	Vyavahare and Habeeb 2018 [66], Gandhi and Gandhi 2018 [64]	2018					
Emergency calls	Suresh et al. 2019 [65], Bharatia et al. 2019 [36]	2019	Elmannai and Elleithy [30]	Field experiments (outdoor and indoor environments)	Blind-folded	1	✗
	Caraiman et al. 2017 [69]	2017					
Text recognition and reading	Arakeri et al. 2018 [67]	2018	Eskicioglu et al. [147]	Controlled environment testing	Blind and Sighted	5 (3 sighted and 2 blind)	✓
	Thomas et al. 2019 [71]	2019					
	Abraham et al. 2020 [127], Srinivasan et al. 2020 [158]	2020					
	Li et al. 2019 [43]	2019	Fusco and Coughlan [163]	Controlled environment testing	Blind	6	✓
Stop light detection and passing crossings	Yohannes et al. 2020 [124], Chen et al. 2020 [135]	2020					
	Suresh et al. 2019 [65], Bharatia et al. 2019 [36]	2019	Gandhi and Gandhi [64]	Controlled environment testing	N/A	N/A	N/A
Live location tracking			Gay et al. [100]	Controlled environment testing	deaf-blindness	5	✓
			Ghosh et al. [162]	Controlled environment testing, Survey (Likert-like scale questions)	Blind-folded	2	✗
			Guerreiro et al. [85]	Field experiments and Survey	Blind	10 blind	✓
			Huppert et al. [102]	Controlled environment testing	Blind and visually impaired	10 (4 female, 6 male)	✓
			Hussain et al. [104]	Controlled environment testing	N/A	2	✗

Appendix 4: Evaluation approaches

Article	Testing type	Subjects	Number of subjects	VI/ Blind testing
Ahmed et al. [164]	Controlled environment testing and Survey (NASA TLX form)	Blind and visually impaired	13	✓

Article	Testing type	Subjects	Number of subjects	VI/ Blind testing	Article	Testing type	Subjects	Number of subjects	VI/ Blind testing
Iwamura et al. [123]	Controlled environment testing, Interview and Survey (Likert-like scale questions)	Visually impaired	7	✓	Martinez et al. [23]	Surveys (NASA TLX form)	Blind-folded	6 (2 female and 4 male)	✗
Jiang et al. [42]	Survey (Likert-like scale questions)	Blind and Sighted	2 groups (number of participants is not mentioned)	✓	Ntakolia et al. [8]	Controlled environment testing, Survey (Likert-like scale questions)	Blind-folded	10	✗
Joshi et al. [81]	Field experiments (outdoor and indoor environments)	N/A	36 (20 VI and 16 blind-folded)	✓	Presti et al. [26]	Think-aloud protocol and Survey (Likert-like scale questions including System Usability Scale (SUS))	Blind/ visually impaired	13 (7 male and 6 female)	✓
Kayukawa et al. [107]	Controlled environment testing, Survey (Likert-like scale questions)	Blind	6	✓	Rahman et al. [113]	Field experiments	N/A (testers are called actors)	N/A	N/A
Kayukawa et al. [108]	Controlled environment testing, Survey (Likert-like scale questions)	Blind (4 male and 2 female)	6	✓	Saha et al. [114]	Controlled environment testing	Blind	13	✓
Kayukawa et al. [109]	Controlled environment testing, Field experiments and Survey (Likert-like scale questions)	Blind	14	✓	Shen et al. [131]	Controlled environment testing	Blind and Blind-folded	6 (3 blind and 3 blind-folded)	✓
Kim et al. [75]	Controlled environment testing	Blind-folded	1	✗	Sosa-García and Odone [57]	Survey (Likert-like scale questions) and Field experiments	Blind-folded, blind and visually impaired	8 (4 blind-folded, 4 blind (1 congenitally blind and 3 low vision) 2 women and 6 men	✓
Kuribayashi et al. [111]	Survey (Likert-like scale questions), Interview	Blind-folded, Blind	6	✓	Stearns and Thieme 2018 [156]	Controlled environment testing	Blind-folded	5	✗
Li et al. [159]	Controlled environment testing	Blind-folded, Blind	N/A	✓	Suresh et al. [65]	Controlled environment testing	Blind	3	✓
Lin et al. [28]	Field experiments and Survey (Likert-like scale questions)	N/A	4	N/A	Tapu et al. [117]	Field experiments	Visually impaired	N/A	✓
Liu et al. [62]	Remote usability testing (recording user's behavior and analyzing it later) and Think-aloud protocol	Blind	7	✓	van Erp et al. [118]	Controlled environment testing, Survey (Likert-like scale questions)	Visually impaired	5	✓
Mante and Weiland [84]	Controlled environment testing (ANOVA analysis) and Interview	Early blind and late blind	12 (5 early blind, 7 late blind)	✓	Wang et al. [83]	Survey (Likert-like scale questions)	Blind	15 (9 congenitally blind)	✓
					Wang et al. [24]	Survey (Likert-like scale questions)	Blind-folded	6 (3 female, 3 male)	✗

Article	Testing type	Subjects	Number of subjects	VI/ Blind testing
Zhang and Ye [58]	Survey (Likert-like scale questions) and Controlled environment testing (indoor environment)	Blind-folded	7	✗
Zhao et al. [121]	Controlled environment testing	Blind-folded	1	✗

^aNot available/mentioned

Appendix 5: VI/blind user requirements for scene understanding

Requirements

1. Audio descriptions of high quality
2. Accuracy of directions and information
3. Alerts for unexpected events
4. Real-time performance for detection/recognition tasks
5. Ease of use, natural/intuitive user interface, acceptable by a broad user population, including senior citizens
6. A simple training procedure, potentially scalable to new objects and personalization
7. Tolerance to viewpoint variations
8. Tolerance to illumination variations
9. Tolerance to blur, motion blur, out of focus, and occlusions
10. Providing information on location, guidance and navigation
11. Providing information on the system operation
12. Selection of the device operation mode offline-online
13. Personalization by giving the ability to the users to define their own level of disability
14. Minimize the dangers and errors by preventing consequences of incidental or unintentional activity
15. Sharing information for accompanying contents of surroundings (coffee shops, hotels, hospitals, etc.)
16. A camera for detecting obstacles for also obstacle avoidance (moving and static objects/obstacles' shape, location, moving speed, etc.)
17. Recognizing the color of clothes
18. An affordable price
19. Early alert for obstacles, especially in a waist level
20. Position restore actions when the user gets lost
21. Notification of uneven floor surfaces such as loose street tiles, puddles or other small holes
22. Systems should reliably provide relevant information when needed, while also considering information accuracy
23. The types of obstacles that are communicated to the user should be restricted to those that are unexpected. This is, especially important to limit information overload and reduce system complexity

Requirements

24. Different contexts may require different types of user interaction. Environments with many obstacles may require different types of notifications (i.e., more frequent, closer in range)
25. A better detection of horizontal objects, ground and small objects as well
26. Smaller and more efficient device
27. Obstacle recognition after detection (information in the output that would allow the user to distinguish between different types of obstacles)
28. A strong enough vibration signal to indicate an imminent collision
29. Detection of moving obstacles and small objects
30. Accurate voice and language recognition
31. Short sentences to be used as input configuration commands to the assistive mobile application
32. Combination of audio and touch model: Audio navigation commands and vibration alerts for early obstacle warning (2 m distance) with crescent frequency
33. Directions to avoid obstacles with vibration signals or audio guidance
34. Body wearable product (the preferred position is the waist)
35. System run locally on the device without the need for internet connection
36. Light product with small size
37. A clear description of the indoor place to create a mental map of it
38. Triggering/Sharpening the visually impaired user's environmental sensing
39. System should cooperate with human helper that guide the visual impaired

Acknowledgements Special thanks to Gofore Spain SL for funding this research. This work was carried out in the context of an industrial PhD, which is a collaboration between Universidad Politécnica de Madrid and Gofore Spain SL.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Pascolini, D., Mariotti, S.P.: Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.* **96**(5), 614–618 (2012)
- Aileni, R.M., Suci, G., Suci, V., Pasca, S., Ciurea, J.: *Smart Systems to Improve the Mobility of People with Visual Impairment Through IoM and IoMT*, pp. 65–84. Springer, Cham (2020). ISBN 78-3-030-16450-8
- Munger, R.J.Y.B., Hilkes, R.G., Perron, M., Sohi, N.: Apparatus and method for a dynamic “region of interest” in a display system, April 11 (2017). US Patent 9,618,748
- Microsoft. Seeing AI App from Microsoft, (2017). <https://www.microsoft.com/en-us/ai/seeing-ai>
- Aira. Aira Tech Corp. your life, your schedule, right now, (2018). <https://aira.io/>
- Chanana, P., Paul, R., Balakrishnan, M., Rao, P.V.M.: Assistive technology solutions for aiding travel of pedestrians with visual impairment. *J. Rehabil. Assist. Technol. Eng.* **4**, 1–11 (2017)
- Maidenbaum, S., Hanassy, S., Abboud, S., Buchs, G., Chebat, D.-R., Levy-Tzedek, S., Amedi, A.: The “eyecane” a new electronic travel aid for the blind: technology, behavior & swift learning. *Restor. Neurol. Neurosci.* **32**(6), 813–824 (2014)
- Ntakolia, C., Dimas, G., Iakovidis, D.K.: User-centered system design for assisted navigation of visually impaired individuals in outdoor cultural environments. *Universal Access in the Information Society*, pp. 1–26, (2020)
- WeWalk. Smart cane for visually impaired and blind people, (2019). <https://wewalk.io/en/>
- Qinghui, T., Malik, M.Y., Hong, Y., Park, J.: A real-time localization system using RFID for visually impaired. [arXiv:1109.1879](https://arxiv.org/abs/1109.1879), (2011)
- Kaur, P., Garg, R.: Camera and sensors-based assistive devices for visually impaired persons: a systematic review. *Int. J. Sci. Technol. Res.* **8**(8), 622–641 (2019) ISSN 22778616
- Kuriakose, B., Shrestha, R., Sandnes, F.E.: Multimodal navigation systems for users with visual impairments—a review and analysis. *Multimodal Technol. Interact.* **4**(4), 1–19 (2020). ISSN 24144088. <https://doi.org/10.3390/mti4040073>
- Kuriakose, B., Shrestha, R., Sandnes, F.E.: Tools and technologies for blind and visually impaired navigation support: a review. *IETE Tech. Rev.* 1–16 (2020). ISSN 09745971. <https://doi.org/10.1080/02564602.2020.1819893>
- Romlay, M.R.M., Toha, S.F., Ibrahim, A.M., Venkat, I.: Methodologies and evaluation of electronic travel aids for the visually impaired people: a review. *Bull. Electr. Eng. Inform.* **10**(3), 1747–1758 (2021). ISSN 23029285. <https://doi.org/10.11591/eei.v10i3.3055>
- Lacey, G.J., Rodriguez-Losada, D.: The evolution of Guido. *IEEE Robot. Autom. Mag.* **15**(4), 75–83 (2008). <https://doi.org/10.1109/MRA.2008.929924>
- Khan, A., Khuro, S.: An insight into smartphone-based assistive solutions for visually impaired and blind people—issues, challenges and opportunities. *Univers. Access Inf. Soc.* **19**, 1–25 (2020). <https://doi.org/10.1007/s10209-020-00733-8>
- Majerova, H.: The aspects of spatial cognitive mapping in persons with visual impairment. *Procedia Social Behav. Sci.* **174**, 3278–3284, (2015). ISSN 18770428. <https://doi.org/10.1016/j.sbspro.2015.01.994>
- Petersen, K., Vakkalanka, S., Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: an update. *Inf. Softw. Technol.* **64**, 1–18 (2015)
- Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), vol. 12, pp. 1–10 (2008)
- Peli, E., Arend, L.E., Jr., Timberlake, G.T.: Computerized image enhancement for visually impaired persons: new technology, new possibilities. *J. Vis. Impair. Blindness* **80**, 849–854 (1986)
- Peli, Eli, Goldstein, Robert B., Young, George M., Trempe, Clement L., Buzney, Sheldon M.: Image enhancement for the visually impaired. Simulations and experimental results. *Investig. Ophthalmol. Vis. Sci.* **32**(8), 2337–2350 (1991)
- Peli, Eli, Tamar: Image enhancement for the visually impaired. *Opt. Eng.* **23**(1), 230147 (1984)
- Martinez, M., Roitberg, A., Koester, D., Stiefelhausen, R., Schauerte, B.: Using technology developed for autonomous cars to help navigate blind people. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 1424–1432 (2017)
- Wang, J., Yang, K., Hu, W., Wang, K.: An environmental perception and navigational assistance system for visually impaired persons based on semantic stixels and sound interaction. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1921–1926. IEEE (2018)
- Badino, H., Franke, U., Pfeiffer, D.: The stixel world—a compact medium level representation of the 3d-world. In: Joint Pattern Recognition Symposium, pp. 51–60. Springer (2009)
- Presti, G., Ahmetovic, D., Ducci, M., Bernareggi, C., Ludovico, L., Baratè, A., Avanzini, F., Mascetti, S.: Watchout: obstacle sonification for people with visual impairment or blindness. In: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, pp. 402–413 (2019)
- Apple. Augmented reality apple developer, (2018). <https://www.apple.com/newsroom/2018/06/apple-unveils-arkit-2/>
- Lin, B.-S., Lee, C.-C., Chiang, P.-Y.: Simple smartphone-based guiding system for visually impaired people. *Sensors* **17**(6), 1371 (2017)
- Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: Monoslam: real-time single camera slam. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 1052–1067 (2007)
- Elmannai, W.M., Elleithy, K.M.: A novel obstacle avoidance system for guiding the visually impaired through the use of fuzzy control logic. In: 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 1–9. IEEE (2018)
- Facil, J.M., Ummenhofer, B., Zhou, H., Montesano, L., Brox, T., Civera, J.: Cam-convs: camera-aware multi-scale convolutions for single-view depth. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11826–11835 (2019)
- Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N.: Deeper depth prediction with fully convolutional residual networks. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 239–248 (2016). <https://doi.org/10.1109/3DV.2016.32>
- Wang, R., Pizer, S.M., Frahm, J.-M.: Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5555–5564 (2019)
- Hsieh, Y.-Z., Lin, S.-S., Fu-Xiong, X.: Development of a wearable guide device based on convolutional neural network for blind or visually impaired persons. *Multimedia Tools Appl.* **79**(39), 29473–29491 (2020)
- Hakim, H., Fadhil, A.: Navigation system for visually impaired people based on RGB-D camera and ultrasonic sensor. In: Proceedings of the International Conference on Information and Communication Technology, ICICT '19, pp. 172–177. Association for Computing Machinery, New York (2019). ISBN 9781450366434
- Bharatia, D., Ambawane, P., Rane, P.: Smart electronic stick for visually impaired using android application and Google's cloud

- vision. In: 2019 Global Conference for Advancement in Technology (GCAT), pp. 1–6. IEEE (2019)
37. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
38. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
39. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
40. Zhao, Z., Zheng, P., Xu, S., Wu, X.: Object detection with deep learning: a review. *IEEE Trans. Neural Netw. Learn. Syst.* **30**(11), 3212–3232 (2019)
41. Amazon. Rekognition, *. <https://aws.amazon.com/rekognition/>
42. Jiang, B., Yang, J., Lv, Z., Song, H.: Wearable vision assistance system based on binocular sensors for visually impaired users. *IEEE Internet Things J.* **6**(2), 1375–1383 (2018)
43. Li, X., Cui, H., Rizzo, J.-R., Wong, E., Fang, Y.: Cross-safe: a computer vision-based approach to make all intersection-related pedestrian signals accessible for the visually impaired. In: Science and Information Conference, pp. 132–146. Springer (2019)
44. Dosi, S., Sambare, S., Singh, S., Lokhande, N., Garware, B.: Android application for object recognition using neural networks for the visually impaired. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1–6. IEEE (2018)
45. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017)
46. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: Single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37. Springer (2016)
47. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, pp. 1–14 (2015)
48. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
49. Duman, S., Elewi, A., Yetgin, Z.: Design and implementation of an embedded real-time system for guiding visually impaired individuals. In: 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–5. IEEE (2019)
50. Eckert, M., Blex, M., Friedrich, C.M. et al.: Object detection featuring 3D audio localization for Microsoft Hololens. In: Proceedings of 11th International Joint Conference on Biomedical Engineering Systems and Technologies, vol. 5, pp. 555–561 (2018)
51. Kommey, B., Herrman, K., Addo, E.O.: A smart vision based navigation aid for the visually impaired. *Asian J. Res. Comput. Sci.* **4**(3), 1–8 (2019)
52. Inception-v3, *. <https://cloud.google.com/tpu/docs/inception-v3-advanced>
53. Keras, *. <https://keras.io>
54. Opencv, *. <https://opencv.org>
55. Tepelea, L., Buciu, I., Grava, C., Gavrilut, I., Gacsádi, A.: A vision module for visually impaired people by using raspberry pi platform. In: 2019 15th International Conference on Engineering of Modern Electric Systems (EMES), pp. 209–212. IEEE (2019)
56. Matlab. Computer vision system toolbox2, *. <https://www.mathworks.com/products/matlab.html>
57. Sosa-García, J., Francesca, O.: “Hands on” visual recognition for visually impaired users. *ACM Trans. Access. Comput. (TACCESS)* **10**(3), 1–30 (2017)
58. Zhang, H., Ye, C.: An indoor wayfinding system based on geometric features aided graph slam for the visually impaired. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(9), 1592–1604 (2017)
59. Canez, A.V., Sartori, J., Barwaldt, R., Rodrigues, R.N.: Collision detection with monocular vision for assisting in mobility of visually impaired people. In: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), pp. 269–274. IEEE (2019)
60. Zhang, W., Qu, C., Ma, L., Guan, J., Huang, R.: Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. *Pattern Recognit.* **59**, 176–187 (2016). ISSN 0031-3203. <https://doi.org/10.1016/j.patcog.2016.01.034>. Compositional Models and Structured Learning for Visual Recognition
61. Liu, Y., Yang, J., Meng, Q., Lv, Z., Song, Z., Gao, Z.: Stereoscopic image quality assessment method based on binocular combination saliency model. *Signal Process.* **125**, 237–248 (2016). ISSN 0165-1684. <https://doi.org/10.1016/j.sigpro.2016.01.019>
62. Liu, Y., Stiles, N.R.B., Meister, M.: Augmented reality powers a cognitive assistant for the blind. *eLife* **7**, e37841 (2018)
63. Dasila, R.S., Trivedi, M., Soni, S., Senthil, M., Narendran, M.: Real time environment perception for visually impaired. In: 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), pp. 168–172. IEEE (2017)
64. Gandhi, S., Gandhi, N.: A CMUcam5 computer vision based arduino wearable navigation system for the visually impaired. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1768–1774. IEEE (2018)
65. Suresh, A., Arora, C., Laha, D., Gaba, D., Bhambri, S.: Intelligent smart glass for visually impaired using deep learning machine vision techniques and robot operating system (ROS). In: International Conference on Robot Intelligence Technology and Applications, pp. 99–112. Springer (2019)
66. Vyavahare, P., Habeeb, S.: Assistant for visually impaired using computer vision. In: 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES), pp. 1–7. IEEE (2018)
67. Arakeri, M.P., Keerthana, N.S., Madhura, M., Sankar, A., Munnavar, T.: Assistive technology for the visually impaired using computer vision. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1725–1730. IEEE (2018)
68. Google. Google vision api, (2017). <https://cloud.google.com/vision>
69. Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszotarski, D., Botezatu, N., Herghelegiu, P., Moldoveanu, F., Strumillo, P., Moldoveanu, A.: Computer vision for the visually impaired: the sound of vision system. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1480–1489 (2017)
70. Google. Tesseract an optical character recognition (OCR) engine, (2015). <https://opensource.google/projects/tesseract>
71. Thomas, M., et al.: iSee: artificial intelligence based android application for visually impaired people. *J. Gujarat Res. Soc.* **21**(6), 200–208 (2019)
72. Microsoft. Azure computer vision API, n.d. <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>
73. Rahman, F., Ritun, I.J., Farhin, N., Uddin, J.: An assistive model for visually impaired people using YOLO and MTCNN. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, ICCSP 19, pp. 225–230. Association for Computing Machinery, New York (2019). ISBN 9781450366182

74. ICF. International classification of functioning, disability and health framework, *. <https://apps.who.int/classifications/icfbrowser/>
75. Kim, J.-H., Kim, S.-K., Lee, T.-M., Lim, Y.-J., Lim, J.: Smart glasses using deep learning and stereo camera. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), pp. 294–295. IEEE (2019)
76. Pehlivan, S., Unay, M., Akan, A.: Designing an obstacle detection and alerting system for visually impaired people on sidewalks. In: 2019 Medical Technologies Congress (TIPTEKNO), pp. 1–4. IEEE, (2019)
77. Alhichri, H., Bazi, Y., Alajlan, N.: Assisting the visually impaired in multi-object scene description using OWA-based fusion of CNN models. *Arab. J. Sci. Eng.* **45**(12), 10511–10527 (2020)
78. Aralikatti, A., Appalla, J., Kushal, S., Naveen, G.S., Lokesh, S., Jayasri, B.S.: Real-time object detection and face recognition system to assist the visually impaired. *J. Phys. Conf. Ser.* **1706**, 012149 (2020)
79. Bhole, S., Dhok, A.: Deep learning based object detection and recognition framework for the visually-impaired. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pp. 725–728 (2020)
80. Malek, S., Melgani, F., Mekhalif, M.L., Bazi, Y.: Real-time indoor scene description for the visually impaired using autoencoder fusion strategies with visible cameras. *Sensors* **17**(11), (2017). ISSN 1424-8220. <https://doi.org/10.3390/s17112641>
81. Joshi, R.C., Yadav, S., Dutta, M.K., Travieso-Gonzalez, C.M.: Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people. *Entropy* **22**(9), (2020). ISSN 1099-4300. <https://doi.org/10.3390/e22090941>
82. Calabrese, B., Velázquez, R., Del-Valle-Soto, C., de Fazio, R., Giannoccaro, N.I., Visconti, P.: Solar-powered deep learning-based recognition system of daily used objects and human faces for assistance of the visually impaired. *Energies* **13**(22), (2020). ISSN 1996-1073. <https://doi.org/10.3390/en13226104>
83. Wang, H.-C., Katzschmann, R.K., Teng, S., Araki, B., Giarre, L., Rus, D.: Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 6533–6540. IEEE (2017)
84. Mante, N., Weiland, J.D.: Visually impaired users can locate and grasp objects under the guidance of computer vision and non-visual feedback. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1–4. IEEE (2018)
85. Guerreiro, J., Sato, D., Asakawa, S., Dong, H., Kitani, K.M., Asakawa, C.: Cabot: designing and evaluating an autonomous navigation robot for blind people. In: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '19, pp. 68–82. Association for Computing Machinery, New York (2019). ISBN 9781450366762
86. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, Volume 52 of *Advances in Psychology*, pp. 139–183. North-Holland, (1988). [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
87. Budrionis, A., Plikynas, D., Daniušis, P., Indrulionis, A.: Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. *Assistive Technology*, 0435, (2020). ISSN 19493614. <https://doi.org/10.1080/10400435.2020.1743381>
88. Townsend, D., Knoefel, F., Gouban, R.: Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4749–4752. IEEE (2011)
89. Lee, K., Sato, D., Asakawa, S., Kacorri, H., Asakawa, C.: Pedestrian detection with wearable cameras for the blind: a two-way perspective. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, pp. 1–12. Association for Computing Machinery, New York (2020). ISBN 9781450367080. <https://doi.org/10.1145/3313831.3376398>
90. Akter, T., Ahmed, T., Kapadia, A., Swaminathan, S.M.: Privacy considerations of the visually impaired with camera based assistive technologies: misrepresentation, impropriety, and fairness. In: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '20. Association for Computing Machinery, New York (2020). ISBN 9781450371032. <https://doi.org/10.1145/3373625.3417003>
91. Verza, R., Lopes Carvalho, M.L., Battaglia, M.A., Messmer Uccelli, M.: An interdisciplinary approach to evaluating the need for assistive technology reduces equipment abandonment. *Multiple Scler. J.* **12**(1), 88–93 (2006)
92. Phillips, B., Zhao, H.: Predictors of assistive technology abandonment. *Assist. Technol.* **5**(1), 36–45 (1993). ISSN 19493614
93. Petrie, H., Carmien, S., Lewis, A.: Assistive Technology Abandonment: Research Realities and Potentials, vol. 10897. Springer, LNCS (2018)9783319942735. https://doi.org/10.1007/978-3-319-94274-2_77
94. Akula, R., Sai, B.R., Jaswitha, K., Kumar, M.S., Yamini, V.: Efficient obstacle detection and guidance system for the blind (haptic shoe). In: Satapathy, S.C., Srujan Raju, K., Shyamala, K., Rama Krishna, D., Favorskaya, M.N. (eds.) *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, pp. 266–271. Springer International Publishing, Cham (2020). ISBN 978-3-030-24318-0
95. Breve, F., Fischer, C.N.: Visually impaired aid using convolutional neural networks, transfer learning, and particle competition and cooperation. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, (2020)
96. Chen, Z., Liu, X., Kojima, M., Huang, Q., Arai, T.: A wearable navigation device for visually impaired people based on the real-time semantic visual slam system. *Sensors* **21**(4), (2021). ISSN 1424-8220. <https://doi.org/10.3390/s21041536>
97. Cheng, R., Wang, K., Bai, J., Zhijie, X.: Unifying visual localization and scene recognition for people with visual impairment. *IEEE Access* **8**, 64284–64296 (2020). <https://doi.org/10.1109/ACCESS.2020.2984718>
98. dos Santos, A.D.P., Medola, F.O., Cinelli, M.J., Ramirez, A.R.G., Sandnes, F.E.: Are electronic white canes better than traditional canes? A comparative study with blind and blindfolded participants. *Univers. Access Inf. Soc.* **20**, 93–103 (2020)
99. Endo, Y., Sato, K., Yamashita, A., Matsubayashi, K.: Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM. In: 2017 International Conference on Biometrics and Kansei Engineering (ICBAKE), pp. 158–162. IEEE (2017)
100. Gay, J., Umfaher, M., Theil, A., Buchweitz, L., Lindell, E., Guo, L., Persson, N.-K., Korn, O.: Keep your distance: a playful haptic navigation wearable for individuals with deafblindness. In: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '20. Association for Computing Machinery, New York (2020). ISBN 9781450371032. <https://doi.org/10.1145/3373625.3418048>
101. Aruco: a minimal library for augmented reality applications based on OpenCV, (2012). <https://www.uco.es/investiga/grupos/ava/node/26>
102. Huppert, F., Hoelzl, G., Kranz, M.: Guidecopter-a precise drone-based haptic guidance interface for blind or visually impaired

- people. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2021)
103. Optitrack, *. <https://optitrack.com>
 104. Hussain, S.S., Durrani, D., Khan, A.A., Atta, R., Ahmed, L.: In-door obstacle detection and avoidance system for visually impaired people. In: 2020 IEEE Global Humanitarian Technology Conference (GHTC), pp. 1–7 (2020)
 105. Towhidul Islam, S.M., Woldegebriel, B., Ashok, A.: Taxseeme: a taxi administering system for the visually impaired. In: 2018 IEEE Vehicular Networking Conference (VNC), pp. 1–2. IEEE (2018)
 106. Islam, M.T., Ahmad, M., Bappy, A.S.: Microprocessor-based smart blind glass system for visually impaired people. In: Uddin, M.S., Bansal, J.C. (eds.) Proceedings of International Joint Conference on Computational Intelligence, pp. 151–161. Springer Singapore, Singapore (2020). ISBN 978-981-13-7564-4
 107. Kayukawa, S., Takagi, H., Guerreiro, J., Morishima, S., Asakawa, C.: Smartphone-based assistance for blind people to stand in lines. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20, pp. 1–8. Association for Computing Machinery, New York (2020). ISBN 9781450368193. <https://doi.org/10.1145/3334480.3382954>
 108. Kayukawa, S., Ishihara, T., Takagi, H., Morishima, S., Asakawa, C.: Blindpilot: a robotic local navigation system that leads blind people to a landmark object. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA '20, pp. 1–9. Association for Computing Machinery, New York (2020). ISBN 9781450368193. <https://doi.org/10.1145/3334480.3382925>
 109. Kayukawa, S., Ishihara, T., Takagi, H., Morishima, S., Asakawa, C.: Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians. In: Proceedings of ACM Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 3 (2020). <https://doi.org/10.1145/3411825>
 110. Khairnar, D.P., Karad, R.B., Kapse, A., Kale, G., Jadhav, P.: Partha: a visually impaired assistance system. In: 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), pp. 32–37 (2020). <https://doi.org/10.1109/CSCITA47329.2020.9137791>
 111. Kuribayashi, M., Kayukawa, S., Takagi, H., Asakawa, C., Morishima, S.: Linechaser: a smartphone-based navigation system for blind people to stand in lines. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21. Association for Computing Machinery, New York (2021). ISBN 9781450380966. <https://doi.org/10.1145/3411764.3445451>
 112. Megalingam, R.K., Vishnu, S., Sasikumar, V., Sreekumar, S.: Autonomous path guiding robot for visually impaired people. In: Mallick, P.K., Balas, V.E., Bhoi, A.K., Zobaa, A.F. (eds.) Cognitive Informatics and Soft Computing, pp. 257–266. Springer Singapore, Singapore (2019). ISBN 978-981-13-0617-4
 113. ur Rahman, S., Ullah, S., Ullah, S.: A mobile camera based navigation system for visually impaired people. In: Proceedings of the 7th International Conference on Communications and Broadband Networking, pp. 63–66 (2019)
 114. Saha, M., Fiannaca, A.J., Kneisel, M., Cutrell, E., Morris, M.R.: Closing the gap: designing for the last-few-meters wayfinding problem for people with visual impairments. In: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '19, pp. 222–235. Association for Computing Machinery, New York (2019). ISBN 9781450366762
 115. Silva, C.S., Wimalaratne, P.: Towards a grid based sensor fusion for visually impaired navigation using sonar and vision measurements. In: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 784–787. IEEE, (2017)
 116. Suny, S.S., Basak, S., Chowdhury, S.M.M.H.: Virtual vision for blind people using mobile camera and sonar sensors. In: Smys, S., Tavares, J.M.R.S., Balas, V.E., Iliyasu, A.M. (eds.) Computational Vision and Bio-Inspired Computing, pp. 1044–1050. Springer International Publishing, Cham (2020). ISBN 978-3-030-37218-7
 117. Tapu, R., Mocanu, B., Zaharia, T.: A computer vision-based perception system for visually impaired. Multimedia Tools Appl. **76**(9), 11771–11807 (2017)
 118. van Erp, J.B.F., Kroon, L.C.M., Mioch, T., Paul, K.I.: Obstacle detection display for visually impaired: Coding of direction, distance, and height on a vibrotactile waist band. Front. ICT **4**, 23 (2017). ISSN 2297-198X. <https://doi.org/10.3389/fict.2017.00023>
 119. Wang, L., Zhao, J., Zhang, L.: Navdog: robotic navigation guide dog via model predictive control and human-robot modeling. In: Proceedings of the 36th Annual ACM Symposium on Applied Computing, SAC '21, pp. 815–818. Association for Computing Machinery, New York (2021). ISBN 9781450381048. <https://doi.org/10.1145/3412841.3442098>
 120. Zeng, L., Weber, G., Ravyse, I., Simros, M., Van Erp, J., Mioch, T., Conradie, P., Saldien, J.: Range-IT: detection and multimodal presentation of indoor objects for visually impaired people. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2017, pp. 1–6 (2017). <https://doi.org/10.1145/3098279.3125442>
 121. Zhao, Y., Huang, R., Hu, B.: A multi-sensor fusion system for improving indoor mobility of the visually impaired. In: 2019 Chinese Automation Congress (CAC), pp. 2950–2955 (2019)
 122. Tapu, R., Mocanu, B., Zaharia, T.: Seeing without sight-an automatic cognition system dedicated to blind and visually impaired people. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1452–1459 (2017)
 123. Iwamura, M., Inoue, Y., Minatani, K., Kise, K.: Suitable camera and rotation navigation for people with visual impairment on looking for something using object detection technique. In: Miesenberger, K., Manduchi, R., Rodriguez, M.C., Peñáz, P. (eds.) Computers Helping People with Special Needs, pp. 495–509. Springer International Publishing, Cham (2020). ISBN 978-3-030-58796-3
 124. Yohannes, E., Lin, P., Lin, C.Y., Shih, T.K.: Robot eye: automatic object detection and recognition using deep attention network to assist blind people. In: 2020 International Conference on Pervasive Artificial Intelligence (ICPAI), pp. 152–157 (2020)
 125. Afif, M., Ayachi, R., Pissaloux, E., Said, Y., Atri, M.: Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people. Multimedia Tools Appl. **79**(41), 31645–31662 (2020)
 126. Mandhala, V.N., Bhattacharyya, D., Vamsi, B., Thirupathi Rao, N.: Object detection using machine learning for visually impaired people. Int. J. Curr. Res. Rev. **12**(20), 157–167 (2020)
 127. Abraham, L., Mathew, N.S., George, L., Sajan, S.S.: Vision-wearable speech based feedback system for the visually impaired using computer vision. In: 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), pp. 972–976. IEEE (2020)
 128. Vaidya, S., Shah, N., Shah, N., Shankarmani, R.: Real-time object detection for visually challenged people. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 311–316 (2020). <https://doi.org/10.1109/ICICCS48265.2020.9121085>
 129. Kandoth, A., Arya, N.R., Mohan, P.R., Priya, T.V., Geetha, M.: Dhrishti: a visual aiding system for outdoor environment. In: 2020 5th International Conference on Communication and

- Electronics Systems (ICCES), pp. 305–310 (2020). <https://doi.org/10.1109/ICCES48766.2020.9137967>
130. Vaidya, S., Shah, N., Shah, N., Shankarmani, R.: Real-time object detection for visually challenged people. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 311–316 (2020). <https://doi.org/10.1109/ICICCS48265.2020.9121085>
 131. Shen, J., Dong, Z., Qin, D., Lin, J., Li, Y.: iVision: an assistive system for the blind based on augmented reality and machine learning. In: Antona, M., Stephanidis, C. (eds.) *Universal Access in Human–Computer Interaction. Design Approaches and Supporting Technologies*, pp. 393–403. Springer International Publishing, Cham (2020). ISBN 978-3-030-49282-3
 132. Wang, L., Patnik, A., Wong, E., Wong, J., Wong, A.: Oliv: an artificial intelligence-powered assistant for object localization for impaired vision. *J. Comput. Vis. Imaging Syst.* **4**(1), 3 (2018)
 133. Gianani, S., Mehta, A., Motwani, T., Shende, R.: Juvo-an aid for the visually impaired. In: 2018 International Conference on Smart City and Emerging Technology (ICSCET), pp. 1–4. IEEE (2018)
 134. Nguyen, H., Nguyen, M., Nguyen, Q., Yang, S., Le, H.: Web-based object detection and sound feedback system for visually impaired people. In: 2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), pp. 1–6 (2020)
 135. Chen, Q., Chen, Y., Zhu, J., De Luca, G., Zhang, M., Guo, Y.: Traffic light and moving object detection for a guide-dog robot. *J. Eng.* **2020**(13), 675–678 (2020). <https://doi.org/10.1049/joe.2019.1137>
 136. Shah, J.A., Raorane, A., Ramani, A., Rami, H., Shekhar, N.: Eyeris: a virtual eye to aid the visually impaired. In: 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), pp. 202–207 (2020). <https://doi.org/10.1109/CSCITA47329.2020.9137777>
 137. Boldu, R., Matthies, D.J.C., Zhang, H., Nanayakkara, S.: Aisee: an assistive wearable device to support visually impaired grocery shoppers. In: *Proceedings of ACM Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4 (2020). <https://doi.org/10.1145/3432196>
 138. Tahoun, N., Awad, A., Bonny, T.: Smart assistant for blind and visually impaired people. In: *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence, ICAAI 2019*, pp. 227–231. Association for Computing Machinery, New York (2019). ISBN 9781450372534. <https://doi.org/10.1145/3369114.3369139>
 139. Akkaputit, P., Ko, I.-Y.: Task-oriented approach to guide visually impaired people during smart device usage. In: 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 28–35, (2021). <https://doi.org/10.1109/BigComp51126.2021.00015>
 140. Baskaran, H., Leng, R.L.M., Rahim, F.A., Rusli, M.E.: Smart vision: assistive device for the visually impaired community using online computer vision service. In: 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 730–734. IEEE (2019)
 141. Clarifi. Computer vision, *. <https://www.clarifai.com>
 142. Cloud Sight. Computer vision, *. <https://cloudsight.ai>
 143. Afif, M., Ayachi, R., Said, Y., Pissaloux, E., Atri, M.: An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation. *Neural Process. Lett.* **51**, 1–15 (2020)
 144. Shelton, A., Ogunfunmi, T.: Developing a deep learning-enabled guide for the visually impaired. In: 2020 IEEE Global Humanitarian Technology Conference (GHTC), pp. 1–8 (2020)
 145. TensorFlow. Object detection, *. <https://www.tensorflow.org>
 146. Shrikesh, S. et al. Vision: android application for the visually impaired. In: 2020 IEEE International Conference for Innovation in Technology (INOCON), pp. 1–6. IEEE (2020)
 147. Eskicioglu, O.C., Ozer, M.S., Rocha, T., Barroso, J.: Safe and sound mobile application: a solution for aid people with visual disabilities' mobility. In: 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion, DSAI 2020, pp. 22–28. Association for Computing Machinery, New York (2020). ISBN 9781450389372. <https://doi.org/10.1145/3439231.3440616>
 148. Imtiaz, M.A., Aziz, S., Zaib, A., Maqsood, A., Khan, M.U., Waseem, A.: Wearable scene classification system for visually impaired individuals. In: 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1–6 (2020). <https://doi.org/10.1109/ICECCE49384.2020.9179439>
 149. Georgiadis, K., Kalaganis, F., Miglotzidis, P., Chatzilari, E., Nikolopoulos, S., Kompatsiaris, I.: A computer vision system supporting blind people—the supermarket case. In: Tzovaras, D., Giakoumis, D., Vincze, M., Argyros, A., (eds.) *Computer Vision Systems*, pp. 305–315. Springer International Publishing, Cham (2019). ISBN 978-3-030-34995-0
 150. Sarwar, M.G., Dey, A., Das, A.: Developing a LBPH-based face recognition system for visually impaired people. In: 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), pp. 286–289, (2021). <https://doi.org/10.1109/CAIDA51941.2021.9425275>
 151. Bazi, Y., Alhichri, H., Alajlan, N., Melgani, F.: Scene description for visually impaired people with multi-label convolutional SVM networks. *Appl. Sci.* **9**(23), (2019). ISSN 2076-3417. <https://doi.org/10.3390/app9235062>
 152. Kedia, R., Yoosuf, K.K., Dedeepya, P., Fazal, M., Arora, C., Balakrishnan, M.: Mavi: an embedded device to assist mobility of visually impaired. In: 2017 30th International Conference on VLSI Design and 2017 16th International Conference on Embedded Systems (VLSID), pp. 213–218 (2017). <https://doi.org/10.1109/VLSID.2017.38>
 153. Ahmetovic, D., Sato, D., Uran, O., Ishihara, T., Kitani, K., Asakawa, C.: ReCog: Supporting Blind People in Recognizing Personal Objects, pp. 1–12. Association for Computing Machinery, New York (2020) 9781450367080
 154. Oskoui, S.S.L., Golestani, H., Hashemi, M., Ghiasi, S.: CNN-droid: GPU-accelerated execution of trained deep convolutional neural networks on android. In: *Proceedings of the 2016 ACM on Multimedia Conference, MM '16*, pp. 1201–1205 (2016)
 155. Awad, M., El Haddad, J., Khneisser, E., Mahmoud, T., Yaacoub, E., Malli, M.: Intelligent eye: a mobile application for assisting blind people. In: 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), pp. 1–6 (2018). <https://doi.org/10.1109/MENACOMM.2018.8371005>
 156. Stearns, L., Thieme, A.: Automated person detection in dynamic scenes to assist people with vision impairments: an initial investigation. In: *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '18*, pp. 391–394. Association for Computing Machinery, New York (2018). ISBN 9781450356503
 157. Hudec, M., Smutny, Z.: Advanced scene recognition system for blind people in household: the use of notification sounds in spatial and social context of blind people. In: *Proceedings of the 2nd International Conference on Computer Science and Application Engineering, CSAE '18*. Association for Computing Machinery, New York (2018). ISBN 9781450365123. <https://doi.org/10.1145/3207677.3278101>
 158. Srinivasan, A.K., Sridharan, S., Sridhar, R.: Object localization and navigation assistant for the visually challenged. In: 2020 Fourth International Conference on Computing Methodologies

- and Communication (ICCMC), pp. 324–328 (2020). <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00061>
159. Li, B., Muñoz, J.P., Rong, X., Chen, Q., Xiao, J., Tian, Y., Arditi, A., Yousuf, M.: Vision-based mobile indoor assistive navigation aid for blind people. *IEEE Trans. Mob. Comput.* **18**(3), 702–714 (2019). <https://doi.org/10.1109/TMC.2018.2842751>
 160. Rizzo, J.-R., Pan, Y., Hudson, T., Wong, E.K., Fang, Y.: Sensor fusion for ecologically valid obstacle identification: Building a comprehensive assistive technology platform for the visually impaired. In: 2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO), pp. 1–5 (2017). <https://doi.org/10.1109/ICMSAO.2017.7934891>
 161. Jabnoun, H., Benzarti, F., Amiri, H.: Visual scene prediction for blind people based on object recognition. In: 2017 14th International Conference on Computer Graphics, Imaging and Visualization, pp. 21–26 (2017). <https://doi.org/10.1109/CGiV.2017.19>
 162. Ghosh, A., Al Mahmud, S.A., Uday, T.I.R., Farid, D.M.: Assistive technology for visually impaired using tensor flow object detection in raspberry pi and coral USB accelerator. In: 2020 IEEE Region 10 Symposium (TENSYP), pp. 186–189 (2020)
 163. Fusco, G., Coughlan, J.M.: Indoor localization for visually impaired travelers using computer vision on a smartphone. In: Proceedings of the 17th International Web for All Conference, W4A '20. Association for Computing Machinery, New York (2020). ISBN 9781450370561. <https://doi.org/10.1145/3371300.3383345>
 164. Ahmed, S., Balasubramanian, H., Stumpf, S., Morrison, C., Sellen, A., Grayson, M.: Investigating the intelligibility of a computer vision system for blind users. In: International Conference on Intelligent User Interfaces, Proceedings IUI, pp. 419–429 (2020). <https://doi.org/10.1145/3377325.3377508>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.