



# A SURVEY ON RECENT ADVANCES IN AI AND VISION-BASED METHODS FOR HELPING AND GUIDING VISUALLY IMPAIRED PEOPLE

Hélène Walle, Barthélemy Serres, Venturini Gilles, Cyril de Runz

## ► To cite this version:

Hélène Walle, Barthélemy Serres, Venturini Gilles, Cyril de Runz. A SURVEY ON RECENT ADVANCES IN AI AND VISION-BASED METHODS FOR HELPING AND GUIDING VISUALLY IMPAIRED PEOPLE. Applied Sciences, 2022, 12 (5), pp.2308. 10.3390/app12052308 . hal-03610027

**HAL Id: hal-03610027**

**<https://hal.science/hal-03610027>**

Submitted on 16 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Article

# A SURVEY ON RECENT ADVANCES IN AI AND VISION-BASED METHODS FOR HELPING AND GUIDING VISUALLY IMPAIRED PEOPLE

Hélène Walle <sup>1</sup>, Cyril De Runz <sup>1</sup> , Barthélémy Serres <sup>2</sup> and Gilles Venturini <sup>1</sup> ,

<sup>1</sup> LIFAT, University of Tours; {helene.walle,cyril.derunz, gilles.venturini}@univ-tours.fr

<sup>2</sup> CETU ILIAD3, University of Tours; barthelemy.serres@univ-tours.fr

\* Correspondence: cyril.derunz@univ-tours.fr

**Abstract:** We present in this paper a state of the art and an analysis of the recent research work and achievements performed in the domain of AI-based and vision-based systems for helping blind and visually impaired people (BVIP). We start by highlighting the recent and tremendous importance that AI has acquired with the use of convolutional neural networks (CNN) and with their ability to efficiently solve image classification tasks. We remind also that VIP have great expectations about AI-based system as a possible way to ease the perception of their environment and to improve their everyday life. Then we set the scope of our survey: we concentrate our investigations on the use of CNN or related methods in a vision-based system for helping VIP. We analyze the existing surveys and we study the current work (a selection of 30 case studies) along several dimensions such as acquired data, the learned models and the Human-Computer interfaces. We compare the different approaches and we conclude by analyzing the future trends in this domain.

**Keywords:** blind and visually impaired people; assistive technologies; artificial intelligence

**Citation:** Walle, H.; De Runz, C.; Serres, B.; Venturini G. A survey on recent advances in AI and vision-based methods for helping and guiding visually impaired people. *Journal Not Specified* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

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

**Copyright:** © 2022 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

According to the World Health Organization 285 million people suffer from important sight loss (39 million blind and 246 million with low vision), and the figures will keep rising up as the population grows older. Assisting blind and visually impaired persons in their everyday life has been a long time research topic, traveling being a particular concern. Traditionally white canes and guide dogs have been acting as walking assistants, but recent advances in deep learning and computer vision technologies have widely broaden the spectrum of possibilities.

Despite being a classical topic, investigated for decades, research teams are continually innovating and lightning hope for a future where vision disability would not be anymore a constant struggle for those affected by it. From the radars in the mid 20th century to the latest AI emerging nowadays, assistive technologies have used an exceptionally diverse set of technologies in designing tools for blind and visually impaired people. The bloom of new, more performing algorithms are opening the path for tomorrow's developments, becoming indeed essential aspects to focus our survey on.

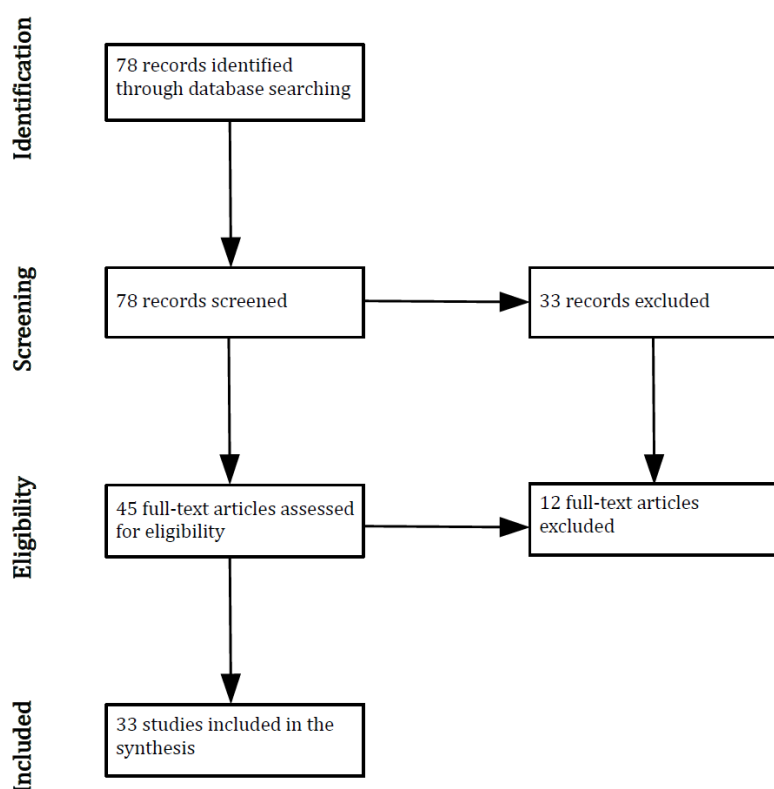
Beyond leading to important innovations, the numerous researches conducted over the last years have brought clearer classifications of the developed assistive tools that is now widely used to define them. Assistive technologies aiming at easing travel are often divided in three parts: Electronic Travel Aids, Electronic Orientation Aids, and Position Locator Devices respectively described by Elmannai et al. [1] as "devices that gather information about the surrounding environment and transfer it to the user", as "devices that provide pedestrians with directions in unfamiliar places" and as "devices that determine the precise position of its holder". Tapu et al. [2] proposes another classification, depending in the kind of skills involved, between perceptual tools that

38 replace vision (images and distances) by other sensory stimuli like acoustic or haptic  
39 signals, and conceptual ones that develop orientation strategies (spatial model or surface  
40 mapping) to represent the environment and prepare to unpredictable situations during  
41 navigation.

42 Assistive systems need three modules to help blind and visually impaired people  
43 (later referred to as BVIP in this paper). The first is the navigation module, or wayfinding,  
44 defined by Kandalan and Namuduri [3] as "designing the set of efficient movements  
45 required to reach the desired destination, which benefits from the knowledge of the user's  
46 initial location and constant update of the user's orientation." This module should ideally  
47 provide the following services: path and surface description, selection of the optimal  
48 path taking several criteria into account (user's preference, avoid hazardous zones, etc),  
49 an accident-free navigation as much as possible and in a reasonable amount of time.  
50 Ideally it must also work: indoor and outdoor, in various light conditions (night/day,  
51 sun/rain, etc), whether the location has already been met or not, and perform real-time  
52 analysis without sacrificing its robustness and accuracy. The second module performs  
53 object detection with two purposes. First it allows to avoid the obstacles that might  
54 cause harm to the user and warn him/her in a helpful way. It must be able to detect  
55 static and dynamic obstacles, as well as their location (preferably ahead), their nature  
56 and estimate their distance to provide timely feedback. Object detection will also be  
57 able to provide a scene description whenever the user asks, in order to let the BVIP  
58 to have a good understanding of his/her environment and build cognitive maps of a  
59 place. The last module is the human-machine interface that encompasses the tools that  
60 will be manipulated by the user. It is comprised of several devices meant to acquire  
61 data, process it and finally return the information to the user. A lot of combinations are  
62 possible and they must be chosen according to several criteria like algorithmic methods,  
63 user's preferences or wearability of devices for example.

64 Several surveys have been conducted in the last years about assistive technologies  
65 for BVIP, and the range of technologies available that could be used to develop them.  
66 Some of those articles have been read when preparing this paper, in order to understand  
67 the different ways in which this subject has been addressed in the past. In their articles  
68 Bhowmick and Hazarika [4], and Khan et al. [5] provided a detailed description of  
69 their research methodologies, highlighting the connections between the different fields  
70 pertaining to assistive technologies for BVIP. They spotted the most frequent keywords  
71 found in the papers they analyzed and their common sources, and stressed out the  
72 trends of research in this domain. Some papers concentrated on the technological  
73 aspects of navigation and object detection. Zhao et al. [6] and Jiao et al. [7] proposed a  
74 survey on object detection with deep learning, its history, its possible techniques and its  
75 current trends. Ignatov et al. [8] and Leo et al. [9] also reviewed recent deep learning  
76 techniques for processing images; Ignatov et al. concentrating on their use in assistive  
77 technologies and Leo et al. on the hardware and frameworks allowing to run them on  
78 Android smartphones. El-Zahraa El-Taher et al. [10] and Kandalan et Namuduri [3]  
79 described several techniques for constructing navigation systems, the tasks that need  
80 to be performed, the hardware available and the possible interfaces. They chose two  
81 different scopes: urban travel for El-Zahraa El-Taher et al. and indoor navigation for  
82 Kandalan and Namuduri. Several papers [1,2,11–14] provided an analysis of different  
83 systems aiming at assisting BVIP in their everyday moves. They did so through several  
84 analysis dimensions such as hardware components and techniques used (sensor or  
85 computer vision based systems), or types of interfaces, proposing more or less detailed  
86 description. Some also described current or future trends and challenges, or even gave  
87 recommendations for future development. It can be noticed that Plikynas et al. [13] and  
88 Tapu et al. [2] associated blind experts to conduct their surveys, and benefit from an end  
89 user's point of view on the proposed assistive devices.

90 This paper has tried to adopt an original approach. First only the most recent papers  
91 were analyzed to build this survey in order to highlight the most recent advances in



**Figure 1.** PRISMA selection diagram that explains the main steps of our survey methodology (see text for explanations).

a constantly evolving research field. The whole development process was taken into account to understand all the aspects involved when helping BVIP, not concentrating on a single AI technologies or assistive function. Those aspects were highlighted separately to get a detailed overview of possibilities in this field, with statistics on the most common and relevant choices among the researchers.

The main contributions of this paper are:

- Bring a state-of-the-art survey based on the latest publications and updates,
- Provide a detailed comparison of several possible configurations of assistive tools for BVIP,
- Emphasize on AI techniques, especially the CNN methods.

The rest of this paper is organized as follow: Section 2 presents the research methodology. Sections 3, 4 and 5 provides papers' analysis respectively on human-machine interface, AI technologies and testing methods. Finally Section 6 ends with a conclusion and some perspectives highlighting achievements and difficulties to solve.

## 2. Method

The survey was conducted with the following search methodology (see Figure 1): the first step was determining the scope of the survey. The selected articles should be:

1. on assistive devices,
2. preferably using Deep Learning technologies,
3. aiming at fulfilling navigation, obstacle detection and/or object recognition tasks,
4. designed for BVIP.

It was also decided to favor the most recent articles, ideally ranging from 2017 until now, to keep only the latest technological advances in AI methods. Articles' quality was particularly taken into account. All publications were assessed through the following tools:

- 117 • Qualis from the Computing Institute of the Federal University of Mato Grosso in  
118 Brasil <sup>1</sup>, with venues ranked A1 or A2,
- 119 • Core Conference Portal <sup>2</sup>, with conferences ranked A or B,
- 120 • Scimago Journal and Country Ranks <sup>3</sup>, with journals or conferences ranked Q1 or  
121 Q2.

122 The articles were found using Google Scholar browser and scanning through scien-  
123 tific databases such as IEEE Xplore, Elsevier ScienceDirect, ACM Digital Library and  
124 PubMed. The search was made using combination of the following keywords:

- 125 • "assistive technologies" or "assistive devices", "navigation",
- 126 • "obstacle detection" and/or "object recognition",
- 127 • "artificial intelligence", "deep learning" or "computer vision",
- 128 • "blind" and/or "visually impaired people".

129 The 78 identified articles were first filtered by reading their abstract to estimate  
130 the correspondence with the chosen topics, thus excluding the unsuitable ones (33). A  
131 few more articles (12) were taken out the list after a quick reading, mostly due to being  
132 technologically inadequate with the subject of this survey. In this paper it was decided  
133 to concentrate on external assistive devices and to set aside the research studies based  
134 on vision replacement by medical bionic prosthetic systems like the project described by  
135 Ge et al. in [15]. 30 case studies, from 33 research papers, were finally incorporated in  
136 the analysis: papers [16–18] and [19,20] were considered as a sole entity, those papers  
137 being part of the same research study. Some more generalist articles were included to  
138 highlight the important steps and challenges when building assistive tools for BVIP.

139 The analysis has been made following three main themes, divided in the following  
140 sections:

- 141 • First, human-machine interface (data acquisition and processing, feedback trans-  
142 missions);
- 143 • Second, Artificial Intelligence techniques (scope, algorithms, datasets, training  
144 techniques);
- 145 • Third, testing methods (context, end-users participation).

146 For each part, the results have been summed up in tables with their percentage of  
147 occurrence.

### 148 3. Human-Machine Interfaces for data acquisition and user feedback

149 In this section, we consider the interface between the BVIP and the AI-based  
150 assistive systems in a broad sense: it encompasses the data acquisition techniques, the  
151 data processing approach and the feedback to the user. Several aspects must be taken  
152 into account when designing such human-machine interface. First some general aspects  
153 need to be considered:

- 154 • The overall interface must be robust and reliable, given that a failure of the system  
155 could potentially be very harmful (even physically) to the users, and capable of  
156 functioning over important time duration (effective energy consumption manag-  
157 ing);
- 158 • The interface must be comfortable to wear, unobtrusive and discrete to avoid the  
159 stigmatization of the carrier;
- 160 • The hardware components must be easily accessible and the total cost must stay  
161 affordable to the majority of the public;
- 162 • The system must also be user-friendly and require minimum training from users, a  
163 too complex interface often resulting in people's misusing or giving up the device.

<sup>1</sup> <https://qualis.ic.ufmt.br/>

<sup>2</sup> <http://portal.core.edu.au/conf-ranks/>

<sup>3</sup> <https://www.scimagojr.com/journalrank.php>

### 164 3.1. Data acquisition and processing

165 The acquisition and processing hardware must be adapted to the deep learning  
 166 techniques chosen to perform the tasks (choice of sensors and cameras types, remote or  
 167 wearable processors). Furthermore, the acquisition step, as in many computer vision  
 168 research or application, is crucial. In the case of embedded assistive technologies, this  
 169 data acquisition can be even more difficult because the user is often a part of the system  
 170 (the user holds or wears the acquisition devices, while moving) and the conditions of  
 171 acquisition cannot be easily controlled (day, night, rain, etc).

#### 172 3.1.1. Type of acquisition interface

Acquisition device	% of papers	Ref. to papers
smart glasses	30	[19,21–28]
smart cane	6.7	[29,30]
smartphone	36.7	[31–36] [18,37–39]
other wearable device	26.7	[40–47]

Table 1: The type of acquisition devices used in each paper

173 Several types of solutions are possible when designing the acquisition part of the  
 174 interfaces (see Table 1), each having its own pros and cons. Smart glasses are frequently  
 175 chosen as interface type because of their many advantages like being able to carry several  
 176 acquisition tools or adopting the eyes' points of view. They are also easy to wear by  
 177 users and can be found quite easily (several models with built-in sensors, cameras and  
 178 headsets are currently available). Their main drawback is their high cost, making them  
 179 unaffordable to many people. Acquisition devices can also be mounted on white cane,  
 180 with the main advantage of using a tool that is already used by most BVIP. The problem  
 181 is that the electronic tools on the canes tend to make them quite heavy and uncomfortable  
 182 to carry. The most common type of interface in the analyzed articles is the smartphone, a  
 183 tool that is both very easy to find and already used by a majority of people. Smartphones  
 184 are getting more and more technologically advanced and some models are now well  
 185 equipped in sensors and high resolution cameras. Despite these advances, many older  
 186 or cheaper models do not hold sufficient data acquisition tools and replacing them is  
 187 sometimes a problem for some users. Some research teams have decided to create their  
 188 own wearable interface to deal with the several issues mentioned above. For example,  
 189 Chen et al.[40] designed an acquisition interface worn on a headband, while Wang et al.  
 190 [41] and Malek et al. [46] opted for small bags worn across the chest.

#### 191 3.1.2. Data acquisition tools

192 The types of sensors that acquire the data must be chosen carefully, taking several  
 193 factors into account (see Table 2). A varied set of sensors will increase a system's  
 194 accuracy, which is crucial in obstacles and objects detection tasks. On the other hand,  
 195 a huge amount of diverse data need to be processed and fused, thus increasing the  
 196 computational cost and time of achievement of a task. When designing their systems,  
 197 researchers must find the right balance between those points, depending on the purpose  
 198 they chose to aim at. One of the most common solution is to use smartphones equipped  
 199 with RGB-D or monocular cameras and position sensors (accelerometer, gyroscope  
 200 and magnetometer), as shown in examples [24–26,30,39,43]. Another solution is the  
 201 introduction of Infra-Red or laser sensors in the acquisition system, a choice made in  
 202 articles [29,44,46]. For outdoor navigation, GPS is almost always chosen as navigational  
 203 support, being easily accessible, cheap and with a very wide coverage.



Acquisition sensors	% of papers	Ref. to papers
monocular camera(s)	33.3	[25,33–35,44] [18,36,37,39,45]
stereo vision (two cameras)	10	[23,28,47]
RGB-D camera	43.3	[21,22,24,30,40–42] [19,26,27,38,43,46]
wide angle camera	6.7	[19,30]
GPS	10	[26,40,47]
position sensors	33.3	[24,29–31,40] [25,26,39,43,47]
IR/laser	20	[19,28,29,43,44,46]

Table 2: Data acquisition tools. By "position sensors" we consider techniques and sensors such as inertial measurement units, odometry, or accelerometer/gyroscope/magnetometer from smartphone.

Type of device	% of papers	Ref. to papers
smartphone	23.3	[26,31,34,37–39] [45]*
tablet, laptop, etc	53.3	[21–24,30,33,40,41] [19,27,35,36,42,43,46] [29]*
remote server	23.3	[16,25,28,44,47] [29,45]*

Table 3: The devices used for data processing. In some papers (mentioned with a "\*"), a hybrid client-server architecture is used (smartphone/laptop + remote server).

### 3.1.3. Types of processors

Another choice to be made when designing the systems is through what ways the data is going to be processed and analyzed (see Table 3). Several papers have chosen a smartphone as sole processor, thus having limited treatment and energy capacities but a cheap and easily accessible solution. An other option is to use a remote server to analyze the data, a solution with high computation power but with important risks of failures due to connection issues, especially when navigating indoor. Among the papers analyzed in this survey, the most common method was the addition of another wearable device, such as a tablet or a laptop, to act as the processor. This solution offers better computational power than the smartphone, and no connection issues; although this option may cause a potential loss of comfort for users when carrying the system.

Smartphone position	% of papers	Ref. to papers
in hand	20	[31–33,37–39]
worn	13.3	[18,34–36]

Table 4: How the smartphone is hold by the users. Not all systems have been implemented on a smartphone.

When acquiring data, smartphones can be carried in two ways: in hand or worn on the body with a specific outfit (see Table 4). Despite being natural, keeping one hand busy with the phone may cause annoyance in every day life activities, not to mention the risk of accidentally dropping it, or the risk of being robbed. Wearing the device on the body may be more comfortable but the localization must be chosen carefully to avoid social stigma: Tapu et al. [35] and Sato et al. [16] opted for a system worn on a belt, while Neugebauer et al. [34] chose a specific headset to place the smartphone on top of the head.

### 3.2. Feedback

The design of the feedback module must be carefully chosen to avoid difficult user's experiences. Feedback is defined by El-Zahraa El-Taher et al. [10] as "the means used by the system to convey information to the blind and visually impaired people." To be efficient the navigation instructions must be delivered quickly and clearly. They must be adapted to the difficulty of the task performed (for example, scene description delivers more complex information than turn left or right instructions), but also to the type of environment encountered (level of noise). This module must respect the user's knowledge (choice of symbolic representations that will be easy to understand) and avoid sensory overload (information's prioritization). Developers must also make sure that the interface delivering the feedback will not have a negative impact on other senses and communication capabilities (by favoring bone-conducting headphones rather than traditional ones for example).

#### 3.2.1. Type of feedback

Type of feedback	% of papers	Ref. to papers
speech	70 %	[21,23,25,29,30,32,40] [26,33–36,42,43] [16,20,27,38,45–47]
<i>in combination with other types</i>	20 %	[16,27,30,32,38,45]
vibrations	13.3 %	[16,30,41,45]
sonification	13.3 %	[22,24,32,38]
tactile	10 %	[27,39,41]

Table 5: General methods and techniques for providing feedback to users. Some tools have only been tested online and their types of feedback remain theoretical.

Feedback is essential in navigation tasks, especially when helping BVIP (see Table 5). A wrong choice may result in hazardous, and potentially accident-prone situations for the user. The most popular choice is audio feedback through speech instructions or sonification guidance (use of sounds or music to depict the different elements present in a given scene). Tactile interface, like a Braille display [41], or haptic systems, such as a vibrating smartphone [16,39], are among the other possibilities. The most effective solution seems to be a combination of several feedback interfaces to adapt to the different situations encountered by the users. For example, audio feedback might not be the most adequate answer in noisy environments, but the best solution when many information need to be delivered, like in scene descriptions tasks. Combination of feedback types is illustrated by Bauer et al. [45] where navigational instructions (turn left or right) are delivered by vibrating smartwatches that can also perform an audio scene description; or by Li et al. [30] with two vibrating motors on a cane for turning left and right and an Android smartphone with text-to-audio software to provide more precise instructions.

#### 3.2.2. Feedback conveyors

Audio-feedback devices	% of papers	Ref. to papers
earphones/headsets	46.7	[23,25,26,29,34,40,42] [20,27,28,35,43,46,47]
bone-conducting earphones	16.7	[16,21,22,24,36]
phone/tablet	13.3	[30,33,37,38]

Table 6: Devices for audio feedback

Several tools are available to deliver audio feedback (see Table 6). In some cases the instructions can be transmitted directly by the smartphone, a method quite incompatible with the respect of users' privacy. The use of headsets tends to be privileged by the



researchers: traditional ones are cheap and easily accessible, but they often disrupt the sense of hearing on which BVIP rely heavily to understand their environment. Another solution is the bone-conducting earphones, that allow to deliver the instructions without covering the ears, therefore not impairing sounds' perception and possibility to communicate with other people.

The devices for the other types of feedback are as follow: motors on a cane [30], vibrating belt and Braille interface [41], two smartwatches [45], and smartphone vibrations [16,39].

The users can also interact with the interface, either to choose a mode (wayfinding or scene description), or just to switch the system on or off. In most papers it is done with the smartphone, thanks to the built-in voice controllers (iOS' Voice Control or Voice Over, Android's Voice Access or TalkBack) to set up the system. In [41] instructions are given through a Braille tablet, or with smartwatches in [45].

#### 4. Artificial Intelligence Techniques

As research on AI and Computer Vision keep expanding, more and more solutions are being available to develop assistive tools for BVIP. The technical choices are motivated by several factors and it is crucial for all the needed features to be previously known for the engineering process to be carried out smoothly. One of the aspect that will influence the technical choice is whether the system will be working indoor, outdoor or both, some tools having a limited usage range (like the GPS, only accessible outside). Development teams must also take into account the kind of tasks that will be performed by the system: sole wayfinding, additional scene description; and the available data sources and computational capabilities, both depending on the interface design. Some techniques, like RFID tags and BLE beacons, require installation of materials previous to the deployment of a system in a specific site, thus highly increasing the costs of usage and maintenance. After deciding which algorithm models will be employed, researchers must finally set up the training procedure with the appropriate datasets and data treatment techniques.

##### 4.1. Scope of system

scope of systems	% of papers	Ref. to papers
indoor	30	[16,23,29–32,41,44,46]
outdoor	10	[21,28,33]
both	43.3	[22,24–27,36,40,42,45,48] [20,39,47]

Table 7: Scope of assistive systems. Some studies do not clearly specify their targeted scope.

Navigation systems can be designed to cover either indoor or outdoor navigation, thus only partially meeting the every day needs of BVIP (see Table 7). Almost half of the systems analyzed in this survey potentially proposes solution for both situations (although this double coverage has often only been tested in one situation so far).

##### 4.2. Machine or Deep Learning algorithms

A lot of methods are currently available for developing tools based on navigation or object detection tasks (see Table 8). In the field of navigation and wayfinding the most popular methods in this survey are:

- SLAM (Simultaneous Localization And Mapping) algorithms are used to construct maps of the encountered environment and localize the user within it. Several methods were experimented in the analyzed papers: semantic visual SLAM (ORB SLAM) in [40], 2-STEP Graph SLAM in [29], VSLAM in [26] and ORB-SLAM2 in [19].

method	% to papers	Ref. to papers
SLAM	16.7	[19,26,29,38,40]
Encoder/Decoder	16.7	[22,28,40,46,48]
RANSAC	16.7	[23,24,26,29,47]
A*	16.7	[19,25,26,29,30]
Kalman's filter	16.7	[23,25,26,30,39]
YOLO	16.7	[32,33,35,42,45]
VGG	13.3	[25,28,36,44]
Inception	10	[37–39]
specific algorithm	20	[28,30,41,42,47,48]

Table 8: most frequent methods

- RANSAC (RANDOM Sample Consensus) is a method for detecting (and eliminating) outliers, and was developed to solve the Location Determination Problem (extracting feature points and localize them on a projection).
- A\* is a search algorithm for wayfinding that uses heuristics to determine the path with the smallest cost.
- Kalman filter algorithm (Linear Quadratic Estimation) is a method to estimate unknown variables from the observation of a series of measurements that can be employed for many tasks like pose estimation [23], obstacle motion estimation [30] or error reduction [39].

Other architectures are available for dealing with object recognition tasks, mostly used in the papers to detect obstacles and to describe a scene's content if needed:

- YOLO (You Only Look Once) is a CNN designed for real-time object detection (it recognizes what objects are present in a scene, and where) created in 2015. Several updated versions are currently available (YOLOv1-v3).
- VGG is deep CNN architecture derivating from AlexNet, developed and already trained (on ImageNet dataset) by Oxford University's Visual Geometry Group. It is mainly designed for image classification and object recognition tasks and adapted to the transfer learning method. Two versions are available: VGG16 (16 convolutional layers) and VGG19 (19 layers).
- Inception is a CNN classifier developed by Google (and named after the movie) that serves for analysis of images and object detection. Among the several versions that have been released, only Inceptionv2 [39] and v3 [37,38] are used in the papers. Others are: Inceptionv1 (aka GoogLeNet), Inceptionv4 and InceptionResNetv1 and v2 (hybrid Inception and ResNet architecture).

To perform the complex tasks needed to build navigation and object recognition assistive tools, the Encoder/Decoder architecture is also a popular choice as shown in [22,28,40,46,48]. Of course numerous algorithms and architectures are possible when developing navigation and object detection systems, and only the most frequent (among the analyzed sample) have been described in this survey. Despite the wide range of possibilities currently available, some papers have opted for the development of specific algorithms tailored to their objectives of research. They are used to perform specific tasks of the overall development like list construction and object detection in [30], or object extraction and obstacle avoidance in [47].

#### 4.3. Choices of Datasets

A wide range of datasets are currently available, the most popular in the described papers being ImageNet, followed by COCO and PASCAL ensembles (see Table 9). Those datasets are quite general and do not always fit the requirements of the systems developed. They are often used for pre-training the models, before using specific datasets created specifically for the project. This last option has been chosen in [28,33,35–37,39,44].

Datasets	% of papers	Ref. to papers
specific	36.7	[25,27,28,33,35–39,44,46]
ImageNet	30	[21,22,28,33,35–37,44,45]
PASCAL-VOC	10	[22,33,45]
COCO	13.3	[22,26,32,39]

Table 9: The most frequent image datasets

#### 336 4.4. Data Processing Methods

ML techniques	% of papers	Ref. to papers
data augmentation	6.7	[22,42]
transfer learning	16.7	[25,36–39]

Table 10: Techniques used for model training.

337 Often, when designing systems for a specific task, finding or creating the right  
 338 dataset turns out to be an important obstacle (see Table 10). To train the models efficiently,  
 339 researchers need to use special methods to overcome this problem. The most frequent one  
 340 within the papers analyzed is transfer learning: the system is entirely pre-trained with  
 341 an already available general dataset (like ImageNet), then the last layer(s) is(are) trained  
 342 with a smaller, more specific one. The other method used here is data augmentation.  
 343 New synthetic images are created from the original dataset with several techniques like  
 344 random flipping, cropping, scaling, rotation and color jittering for [22], and rotation,  
 345 skewing, mirroring, flipping, brightness and noise levels in [42].

#### 346 4.5. Type of Model Training

Type of training	% of papers	Ref. to papers
incremental	16.7	[35–38,48]
offline	46.7	[21,22,25,31–33,42] [26–28,36,37,39,45]

Table 11: Incrementality of learning procedures.

347 Two types of methods are possible when training a model (see Table 11). Most  
 348 of the papers opted for an offline training: the model is trained once with pre-defined  
 349 dataset(s) and then deployed. However 5 papers [35–38,48] chose incremental learning:  
 350 the model keeps being trained with data acquired by the users themselves, in order to  
 351 obtain an increasing accuracy and deal more easily with complex situations uncovered  
 352 by the datasets.

#### 353 4.6. Solving Challenges

354 Despite their promising results, these systems still have to be improved in order to  
 355 be massively deployed on smartphones and small devices. As Bertheliet et al. [49] state  
 356 in their survey "deep learning based methods have achieved state-of-the-art performance  
 357 in many applications such as face recognition, semantic segmentation, object detection,  
 358 etc [...] but] to run these applications on embedded devices the deep models need to be  
 359 less-parametrized in size and time efficient." To address these issues, several compression  
 360 techniques for CNNs are being currently studied:

- 361 • Pruning, which consists in removing unused parameters of a network while still  
 362 achieving state-of-the-art results;
- 363 • Quantization, that approximates a neural network by reducing floating-point num-  
 364 bers' precision, with higher risks of error and lower accuracy;

- Hash methods that convert original features into low-dimensional hash codes, regrouping data according to similarity and avoiding redundancy. Some hashing systems are already available like HashedNets by NVIDIA;
- Knowledge distillation, which is based on the process of transferring knowledge from a deep neural network to a shallow one, while keeping the same efficiency and learning capacities.

Compression methods has been widely investigated in the past years, and are already proposed as part of several frameworks such as TensorflowLite or Apple's Core ML.

Another way of reducing computational needs is by optimizing the system's architecture itself. Several methods to reduce the cost of convolutional operations have been successfully tested like the conception of modules (with specifically organized and sized layers) or the use of separable convolutional layers. However the most promising results have emerged in the field of Neural Architecture Search (NAS) that aims to develop self-organized structures that would automatically shape their design to fit the targeted tasks (Neural gas methods, Neuroevolution, Network morphism, Supergraphs).

Incremental learning is another crucial challenge for the research on assistive technologies with CNNs. Its implementation on embedded devices represents an important step for BVIP as it is going to allow them to customize their system to detect personal belongings or recognize friends, colleagues and family members. Many research teams are currently developing solutions, and proposals are being made like the ones described by Luo et al. [50] in their recent survey. Indeed incrementality is even more difficult to implement on embedded devices because of their limited computational power.

## 5. Testing Methods

There are several ways of assessing an assistive system. First the testing phase can be theoretical (virtual scenarios performed on computer) and/or practical (pre-determined paths or real-life environments). Systems can be assessed with objective performance metrics evaluating for instance accuracy, time of response, best choice of path. Subjective evaluations can also be conducted to measure the level of acceptance and usefulness of the system for BVIP, with for instance wearability, appropriateness of feedback or integrability in everyday life.

### 5.1. Types of Tests

method	% of papers	Ref. to papers
only simulation	26.7	[28,31,32,35,39,44,46,48]
only on field	20	[16,20,24,34,41,43]
both	53.5	[21–23,25,26,29,30,33,40,42] [27,36–38,45,47]

Table 12: Method of testing

Several phases of testing can be observed in the papers analyzed (see Table 12). Usually the system is first experimented theoretically on computer to check if its accuracy meets the standards, before passing to practical tests in a given environment (either in pre-set paths or buildings, or unknown real-life conditions). Some studies, probably less advanced in their development, only tested their proposed system online, while others skipped the theoretical phase to deploy their system directly in real conditions.

### 5.2. End-Users Testing

The majority of analyzed systems were tested directly by blind or visually impaired users, which is important to check whether the needs of the end users have been met or not (see Table 13). A third of papers were only tested by seeing people, blindfolded or

Tests with BVIP	% of papers	Ref. to papers
yes	56 %	[21–25,30,33,41,42] [16,20,26,27,34,36–38]
no	34 %	[28,29,31,39,40,44–46,48] [47]

Table 13: Tests with BVIP or not (in general blind folded people).

not, leaving uncertainty about the adequation between the proposed solution and the highlighted issues.

### 5.3. Testing panel

Nb of testers	% of papers	Ref. to papers
<5	10	[25,30,33]
[5 - 15]	26.7	[21,22,24,34,36–38,41]
>15	16.7	[16,23,26,27,42]

Table 14: Number of of VIP testers.

The number of visually impaired testers varies greatly between papers, depending on the possibilities of recruitment or time the researchers had to conduct their testing phase (see Table 14). Gathering enough end users to test the system is important when developing an assistive tool to check if all the requirements are met. Furthermore their unique feedback will allow to correct several aspects of the system, such as interface or wearability, in order to make it as easy and comfortable to use as possible.

## 6. Conclusions and Perspectives

### 6.1. Achievements

In this paper we have studied the recent advances in the field of AI techniques for developing assistive technologies for BVIP. This survey has been made by assessing among the most recent case studies in the field with a well-defined research methodology. Several aspects of the proposed systems have been analyzed, including their human-machine interface (acquisition and processing hardware, feedback types and conveyors), the AI techniques chosen and the ways in which they have been assessed.

Helping disabled people, like BVIP, has been a major research topic for a long time and the burst of technologies in our everyday life has dramatically accelerated the development of that field. Although lacking functions, several smartphone apps are already available for BVIP and are being adopted by many to ease their everyday activities. Some examples are:

- *Lookout* (Android) and *SeeingAI* (iOS) to identify objects or people,
- *TapTapSee* and *VizWiz* (iOS/Android) to identify elements in pictures taken by the user,
- *Ariadne GPS*, *Microsoft Soundscape* and *Blindsquare* (iOS) that are GPS apps that provide customized environment descriptions,
- *Evelity* (iOS/Android) to navigate in equipped buildings thanks to a specific GPS system,
- *MyMoveo* (iOS/Android) to activate connected crosswalks and spot important elements (like entrance doors).

Benefiting from the recent innovations of AI, assistive technologies have developed their possibilities while increasing their robustness and efficiency. In a very close future, it will be possible to propose safe and reliable wayfinding systems to BVIP, as well as detailed scene description and customized object recognition. Devices are also getting cheaper and more comfortable to wear, especially smartphones that are now equipped

with many sensors, such as cameras, gyroscopes, accelerometers, and endowed with enough computing capacities to run deep or machine learning algorithms.

## 6.2. Limitations and challenges

Despite varied and interesting proposals of robust and accurate systems, several challenges and perspectives must be studied and addressed. They are mainly related to the following points (as further discussed in this section):

- most of the reviewed systems fail to become really operational, and reasons for this must be analyzed,
- human factors are very important for BVIP, and they must be carefully understood and taken into account, raising challenges for user-centered AI approaches,
- the use of embedded devices is mandatory but it raises important constraints on AI-based techniques, due to limited storage and computational resources, and to higher power consumption. In addition, incrementality becomes important for BVIP, as a way to adapt the object recognition step to specific objects, including personal ones,
- in comparison to other general fields, the available image datasets or pre-trained CNNs specific to BVIP are not widely accessible (or even defined). The adaptation of such models to specific conditions (cultural, dynamic, etc) is difficult.

In the majority of the studied papers, presented prototypes are not yet ready to be deployed. This difficulty to achieve the last step of development is caused by several factors that will have to be solved, the lack of complete and available solutions representing an additional difficulty for BVIP. Firstly, despite the continuing innovations in the field of AI and computer vision, researchers are still facing serious issues when developing navigation tools for BVIP. Many situations can be challenging for navigation systems:

- obstacles situated at ground or head levels are still difficult to detect, and those are indeed very important for BVIP,
- many elements present in urban environments are geographically or culturally dependent (New York yellow cabs are quite different from London black ones),
- appropriate scene description is a quite subjective notion (what information is really important depends on the user and the current context).

These lacks are due to the huge complexity of the problems, but also to the fact that some research fields have been more developed than others in the last years, and not for the specific case of BVIP. For example object detection and outdoor localization are much more advanced today than safe street crossings or environment mapping. In the past, salient object detection has been a major issue when developing assistive technologies for BVIP, that is now on the edge to be overcome thanks to recent innovations, such as the one described by Ji et al. in [51]. In addition, it might seem to be a minor point, but one should notice that most existing CNNs available for reuse are trained with images of objects or scenes that are not specific to BVIP. Many relevant objects or situations (for BVIP) are missing from training sets. Building such training sets and their specific CNNs and making them available to the scientific (and developers) community would certainly help to expand AI for BVIP.

Also, the human factors are very important in such AI applications. There is often a gap between end-users' needs and developers' decisions. As shown in this paper, BVIP are not systematically implicated in the final testing phase, and even more rarely in the engineering process. Some important requirements often remain unfulfilled, thus leading the users into giving up the developed systems. Including BVIP as well as health professionals in the development process of AI-based systems is certainly mandatory to increase the chance of those systems to become operational. More generally, the human factors involved in such specific AI applications play a crucial role, are part of the user-centered AI approaches. Taking into account such factors raises challenging



issues. For instance, the design of human-machine interfaces does not always take into account the users' relation to their environment. For example, many BVIP do not wish to be walking around with their smartphones in their hand, fearing of letting them fall or being assaulted and robbed. Also, scene description methods should get additional attention for BVIP who mostly rely on audio information.

Running deep learning systems on small devices has become a central questions for researchers worldwide, as the demand for always more efficient, sophisticated and easily accessible AI applications grows up. Beyond the search on neural networks' compression and optimization, as described in Section 4.6, several developments have been made about more adequate hardware units to host those systems. The subject of customizing deep learning methods to fit their targeted tasks and types of hardware has been widely studied on the last years, especially the use of Graphics Processing Units (GPUs), Field Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs), thus highlighting their pros and cons. The developments made in the last years have been analyzed and summarized by several authors: Seng et al. [52] for FPGAs, Moolchandani et al. [53] for ASICs and Ang et al. [54] for GPUs.

Other aspects of deep learning models increase the difficulty of such applications: for instance, an incremental learning model is important, as it can allow the user to teach the AI with new objects to recognize. This can concern personal objects that belong to the user, objects that are culturally dependent, or even people (like family members) or pets. Incrementality raises the problem of dynamic learning, which becomes even more complex on an embedded device with limited resources.

Finally, in the future, intelligent assistive technologies for BVIP will have to interact with connected areas, such as smart cities. Integrating people with disability and their assistants in the development of tomorrow's environments is a question that need to be taken into account by researchers. A few proposals are already available like the ones proposed by Chang et al [55] in their recent article.

**Author Contributions:** "Conceptualization, H.W., C.d.R., B.S. and G.V.; methodology, H.W., C.d.R., and G.V.; validation, H.W., C.d.R., B.S. and G.V.; formal analysis, H.W.; investigation, H.W.; data curation, H.W.; writing—original draft preparation, H.W. and G.V.; writing—review and editing, C.d.R., B.S. and G.V.; supervision, C.d.R. and G.V.; project administration, C.d.R.; funding acquisition, C.d.R. and G.V. All authors have read and agreed to the published version of the manuscript.", please turn to the [CRediT taxonomy](#) for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

**Funding:** This research received no external funding.

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

## References

1. Elmannai, W.; Elleithy, K. Sensor-based assistive devices for visually-impaired people: current status, challenges, and future directions. *Sensors* **2017**, *17*, 565.
2. Tapu, R.; Mocanu, B.; Zaharia, T. Wearable assistive devices for visually impaired: A state of the art survey. *Pattern Recognition Letters* **2020**, *137*, 37–52.
3. Kandalan, R.N.; Namuduri, K. Techniques for Constructing Indoor Navigation Systems for the Visually Impaired: A Review. *IEEE Transactions on Human-Machine Systems* **2020**, pp. 492–506.
4. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces* **2017**, *11*, 149–172.
5. Khan, S.; Nazir, S.; Khan, H.U. Analysis of Navigation Assistants for Blind and Visually Impaired People: A Systematic Review. *IEEE Access* **2021**, pp. 26712–26734.
6. Zhao, Z.Q.; Zheng, P.; Xu, S.t.; Wu, X. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems* **2019**, *30*, 3212–3232.
7. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A survey of deep learning-based object detection. *IEEE Access* **2019**, *7*, 128837–128868.
8. Ignatov, A.; Timofte, R.; Chou, W.; Wang, K.; Wu, M.; Hartley, T.; Van Gool, L. AI benchmark: Running deep neural networks on android smartphones. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Springer, 2018, pp. 288–314.

9. Leo, M.; Furnari, A.; Medioni, G.G.; Trivedi, M.; Farinella, G.M. Deep learning for assistive computer vision. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Springer, 2018, pp. 3–14.
10. El-Zahraa El-Taher, F.; Taha, A.; Courtney, J.; McKeever, S. A systematic review of urban navigation systems for visually impaired people. *Sensors* **2021**, *21*, 3103.
11. Islam, M.M.; Sadi, M.S.; Zamli, K.Z.; Ahmed, M.M. Developing walking assistants for visually impaired people: A review. *IEEE Sensors Journal* **2019**, *19*, 2814–2828.
12. Real, S.; Araujo, A. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors* **2019**, *19*, 3404.
13. Plikynas, D.; Žvironas, A.; Budrionis, A.; Gudauskis, M. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors* **2020**, *20*, 636.
14. Kuriakose, B.; Shrestha, R.; Sandnes, F.E. Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. *IETE Technical Review* **2020**, pp. 1–16.
15. Ge, C.; Kasabov, N.; Liu, Z.; Yang, J. A spiking neural network model for obstacle avoidance in simulated prosthetic vision. *Information Sciences* **2017**, *399*, 30–42.
16. Sato, D.; Oh, U.; Guerreiro, J.; Ahmetovic, D.; Naito, K.; Takagi, H.; Kitani, K.M.; Asakawa, C. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing (TACCESS)* **2019**, *12*, 1–30.
17. Murata, M.; Ahmetovic, D.; Sato, D.; Takagi, H.; Kitani, K.M.; Asakawa, C. Smartphone-based indoor localization for blind navigation across building complexes. 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, 2018, pp. 1–10.
18. Sato, D.; Oh, U.; Naito, K.; Takagi, H.; Kitani, K.; Asakawa, C. NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2017, pp. 270–279.
19. Bai, J.; Lian, S.; Liu, Z.; Wang, K.; Liu, D. Virtual-blind-road following-based wearable navigation device for blind people. *IEEE Transactions on Consumer Electronics* **2018**, *64*, 136–143.
20. Bai, J.; Lian, S.; Liu, Z.; Wang, K.; Liu, D. Smart guiding glasses for visually impaired people in indoor environment. *IEEE Transactions on Consumer Electronics* **2017**, *63*, 258–266.
21. Lin, S.; Cheng, R.; Wang, K.; Yang, K. Visual localizer: Outdoor localization based on convnet descriptor and global optimization for visually impaired pedestrians. *Sensors* **2018**, *18*, 2476.
22. Yang, K.; Wang, K.; Bergasa, L.M.; Romera, E.; Hu, W.; Sun, D.; Sun, J.; Cheng, R.; Chen, T.; López, E. Unifying terrain awareness for the visually impaired through real-time semantic segmentation. *Sensors* **2018**, *18*, 1506.
23. Simões, W.C.; Silva, Y.M.; Pio, J.L.d.S.; Jazdi, N.; F de Lucena, V. Audio Guide for Visually Impaired People Based on Combination of Stereo Vision and Musical Tones. *Sensors* **2020**, *20*, 151.
24. Hu, W.; Wang, K.; Yang, K.; Cheng, R.; Ye, Y.; Sun, L.; Xu, Z. A comparative study in real-time scene sonification for visually impaired people. *Sensors* **2020**, *20*, 3222.
25. Son, H.; Krishnagiri, D.; Jegannathan, V.S.; Weiland, J. Crosswalk guidance system for the blind. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 3327–3330.
26. Bai, J.; Liu, Z.; Lin, Y.; Li, Y.; Lian, S.; Liu, D. Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics* **2019**, *8*, 697.
27. Lin, Y.; Wang, K.; Yi, W.; Lian, S. Deep learning based wearable assistive system for visually impaired people. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. IEEE, 2019, pp. 2549–2557.
28. Dimas, G.; Diamantis, D.E.; Kalozoumis, P.; Iakovidis, D.K. Uncertainty-Aware Visual Perception System for Outdoor Navigation of the Visually Challenged. *Sensors* **2020**, *20*, 2385.
29. Zhang, H.; Ye, C. An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2017**, *25*, 1592–1604.
30. 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 transactions on mobile computing* **2018**, *18*, 702–714.
31. Mahida, P.; Shahrestani, S.; Cheung, H. Deep Learning-Based Positioning of Visually Impaired People in Indoor Environments. *Sensors* **2020**, *20*, 6238.
32. Yang, G.; Saniie, J. Sight-to-Sound Human-Machine Interface for Guiding and Navigating Visually Impaired People. *IEEE Access* **2020**, *8*, 185416–185428.
33. Lin, B.S.; Lee, C.C.; Chiang, P.Y. Simple smartphone-based guiding system for visually impaired people. *Sensors* **2017**, *17*, 1371.
34. Neugebauer, A.; Rifai, K.; Getzlaff, M.; Wahl, S. Navigation aid for blind persons by visual-to-auditory sensory substitution: A pilot study. *Plos one* **2020**, *15*, e0237344.
35. Tapu, R.; Mocanu, B.; Zaharia, T. DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance. *Sensors* **2017**, *17*, 2473.
36. Mocanu, B.; Tapu, R.; Zaharia, T. Deep-see face: A mobile face recognition system dedicated to visually impaired people. *IEEE Access* **2018**, *6*, 51975–51985.
37. Kacorri, H.; Kitani, K.M.; Bigham, J.P.; Asakawa, C. People with visual impairment training personal object recognizers: Feasibility and challenges. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 5839–5849.

38. Ahmetovic, D.; Sato, D.; Oh, U.; Ishihara, T.; Kitani, K.; Asakawa, C. Recog: Supporting blind people in recognizing personal objects. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–12.
39. Lo Valvo, A.; Croce, D.; Garlisi, D.; Giuliano, F.; Giarré, L.; Tinnirello, I. A Navigation and Augmented Reality System for Visually Impaired People. *Sensors* **2021**, *21*, 3061.
40. 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* **2021**, *21*, 1536.
41. Wang, H.C.; Katschmann, R.K.; Teng, S.; Araki, B.; Giarré, L.; Rus, D. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 6533–6540.
42. 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* **2020**, *22*, 941.
43. Grayson, M.; Thieme, A.; Marques, R.; Massiceti, D.; Cutrell, E.; Morrison, C. A dynamic AI system for extending the capabilities of blind people. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–4.
44. Cornacchia, M.; Kakillioglu, B.; Zheng, Y.; Velipasalar, S. Deep learning-based obstacle detection and classification with portable uncalibrated patterned light. *IEEE Sensors Journal* **2018**, *18*, 8416–8425.
45. Bauer, Z.; Dominguez, A.; Cruz, E.; Gomez-Donoso, F.; Orts-Escolano, S.; Cazorla, M. Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors. *Pattern recognition letters* **2020**, *137*, 27–36.
46. Malek, S.; Melgani, F.; Mekhalfi, M.L.; Bazi, Y. Real-time indoor scene description for the visually impaired using autoencoder fusion strategies with visible cameras. *Sensors* **2017**, *17*, 2641.
47. Elmannai, W.M.; Elleithy, K.M. A highly accurate and reliable data fusion framework for guiding the visually impaired. *IEEE Access* **2018**, *6*, 33029–33054.
48. Wang, L.; Famouri, M.; Wong, A. DepthNet Nano: A Highly Compact Self-Normalizing Neural Network for Monocular Depth Estimation, 2020, [[arXiv:cs.CV/2004.08008](https://arxiv.org/abs/2004.08008)].
49. Bertheliet, A.; Chateau, T.; Duffner, S.; Garcia, C.; Blanc, C. Deep Model Compression and Architecture Optimization for Embedded Systems: A Survey. *Journal of Signal Processing Systems* **2021**, *93*, 863–878.
50. Luo, Y.; Yin, L.; Bai, W.; Mao, K. An Appraisal of Incremental Learning Methods. *Entropy* **2020**, *22*, 1190.
51. Ji, Y.; Zhang, H.; Zhang, Z.; Liu, M. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Information Sciences* **2021**, *546*, 835–857.
52. Seng, K.P.; Lee, P.J.; Ang, L.M. Embedded Intelligence on FPGA: Survey, Applications and Challenges. *Electronics* **2021**, *10*, 895.
53. Moolchandani, D.; Kumar, A.; Sarangi, S.R. Accelerating CNN Inference on ASICs: A Survey. *Journal of Systems Architecture* **2021**, *113*, 101887.
54. Ang, L.M.; Seng, K.P. GPU-Based Embedded Intelligence Architectures and Applications. *Electronics* **2021**, *10*, 952.
55. Chang, I.; Castillo, J.; Montes, H. Technology-Based Social Innovation: Smart City Inclusive System for Hearing Impairment and Visual Disability Citizens. *Sensors* **2022**, *22*, 848.