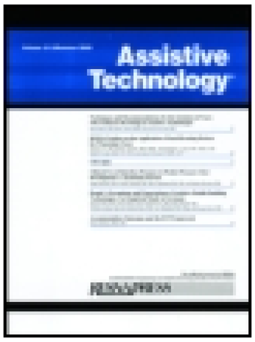

Assistive Technology

The Official Journal of RESNA

ISSN: 1040-0435 (Print) 1949-3614 (Online) Journal homepage: https://www.tandfonline.com/loi/uaty20

Smartphone-based Computer Vision Travelling Aids for blind and visually impaired individuals: A Systematic Review

Andrius Budrionis, Darius Plikynas, Povilas Daniušis & Audrius Indrulionis

To cite this article: Andrius Budrionis, Darius Plikynas, Povilas Daniušis & Audrius Indrulionis (2020): Smartphone-based Computer Vision Travelling Aids for blind and visually impaired individuals: A Systematic Review, Assistive Technology, DOI: 10.1080/10400435.2020.1743381

To link to this article: https://doi.org/10.1080/10400435.2020.1743381

Accepted author version posted online: 24 Mar 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=uaty20

Check for updates

Publisher: Taylor & Francis & RESNA

Journal: *Assistive Technology*

DOI: 10.1080/10400435.2020.1743381

Smartphone-based Computer Vision Travelling Aids for blind and visually impaired individuals: A Systematic Review

Andrius Budrionis[1,2], Darius Plikynas[1], Povilas Daniušis[1], Audrius Indrulionis[1]

[1] *Department of Business Technologies and Entrepreneurship, Vilnius Gediminas Technical University, Vilnius, Lithuania*

[2] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*

Correspondence: Andrius Budrionis (e-mail: Andrius.Budrionis@vgtu.lt).

Accepted Manuscript

Acknowledgments

We want to express our gratitude to our colleagues at Vilnius Gediminas Technical University Lina Pečiūrė, Eglė Jarmolavičiūtė, Marius Gudauskis, and Aurimas Papečkys for their help and insights in preparing this publication. We also want to thank the Association of blind and VI individuals in Lithuania (Lietuvos aklųjų ir silpnaregių sąjunga, LASS) and all participants of VI user survey.

Given the growth in the numbers of visually impaired (VI) people in low-income countries, the development of affordable electronic travel aid (ETA) systems employing devices, sensors, and apps embedded in ordinary smartphones becomes a potentially cost-effective and reasonable all-in-one solution of utmost importance for the VI. This paper offers an overview of recent ETA research prototypes that employ smartphones for assisted orientation and navigation in indoor and outdoor spaces by providing additional information about the surrounding objects. Scientific achievements in the field were systematically reviewed using PRISMA methodology. Comparative meta-analysis showed how various smartphone-based ETA prototypes could assist with better orientation, navigation, and wayfinding in indoor and outdoor environments. The analysis found limited interest among researchers in combining haptic interfaces and computer vision capabilities in smartphone-based ETAs for the blind, few attempts to employ novel state-of-the-art computer vision methods based on deep neural networks, and no evaluations of existing off-the-shelf navigation solutions. These results were contrasted with findings from a survey of blind expert users on their problems in navigating in indoor and outdoor environments. This revealed a major mismatch between user needs and academic development in the field.

Keywords: smartphone device; computer vision techniques; electronic travel aid; obstacle detection; object recognition

Word count: 10106

Introduction

According to the World Health Organization, in 2018, 217 million people worldwide had moderate-to-severe vision impairment, and 36 million were blind. It is estimated that by the year 2020, the number of visually impaired (VI) people will rise to 250 million, while the number of those who are completely blind will reach 75 million (Bourne et al., 2017). It is important to note that almost 90% of blind individuals are living in low-income countries. The majority of people with vision impairment are more than 50 years old. The ageing process and the associated reduction in income have marginalized VI people in developed countries as well. Magnitude, temporal trends, and projections of the global prevalence of blindness and visual impairment can be found in the exhaustive systematic reviews and meta-analyses such as (Bourne et al., 2017; Fricke et al., 2018).

The observed growing numbers of VI people with low incomes give impetus for the development of considerably cheaper assistive devices. This can be achieved by employing widely available smartphones as multifunctional and multisensory mobile devices. These are equipped with a CPU, an operating system, various sensors (GPS, accelerometer, gyroscope, magnetometer, pedometer, and compass) and can run apps for data processing or facilitate continuous wireless data transfer to external servers and cloud platforms for processing. In addition, mobile computing platforms offer standard APIs for general-purpose computing, providing both application developers and users with a level of flexibility that is very conducive to the development and distribution of novel solutions (Csapó et al., 2015).

It is important to note that VI individuals are not so different from the visually able population with regard to smartphone use. In fact, due to their condition, VI individuals are even more inclined to use handheld smartphones for social communication and mobility (making calls, chatting, using social media and many other apps, including GPS navigation, and so on). The screen

Accepted Manuscript

reader interface integrated in modern mobile operating systems is accessible enough for VI people. The number of mobile apps tailored for blind users is also increasing, boosting the use of mobile devices and apps among VI people, and this usage is expected to continue to grow (Griffin-Shirley et al., 2017).

Relatively few studies have been conducted on mobile app use among VI persons. In some preliminary studies, participants rated apps as useful (95.4%) and accessible (91.1%) tools for individuals with visual impairments. More than 90% of middle-aged adults strongly agreed with the practicality of the specifically tailored apps. This shows that VI individuals frequently use apps that are specifically designed to help them accomplish daily tasks. Furthermore, the VI population is generally satisfied with mobile apps and is ready for improvements and new apps (Griffin-Shirley et al., 2017).

Recent advances in computer vision and smartphone devices open up new opportunities, which should motivate the academic community to find novel solutions that combine these evolving technologies to enhance the mobility and general quality of life of VI people. Unfortunately, we have found that this prospective research niche has not yet been well covered in review papers. The only reviews we could identify were several that focused on existing mobile applications for the blind (Csapó et al., 2015; Griffin-Shirley et al., 2017). These findings suggest that electronic travel aids, navigation assistance modules, and text-to-speech applications, as well as virtual audio displays, which combine audio with haptic channels, are becoming integrated into standard mobile devices. Increasingly user-friendly interfaces and new modes of interaction have opened a variety of novel possibilities for the VI (Chessa et al., 2016; Tapu, Mocanu, Bursuc, et al., 2013).

Despite a large number of currently available technological ETA solutions for orientation and navigation, only a few review articles were discovered. Some authors of these articles provide structured information on the technology, functionality, and even rate the solutions. For instance, Tapu et al. provide a summary on wearable devices, grouped by technologies (Tapu et al., 2018). The reviewed systems and devices can be classified into two main groups: sensor-based and video camera-based solutions. Similarly, Islam et al. grouped ETAs into sensor-, computer vision- and smartphone-based solutions, illustrated the groups by a number of selected publications and provided generalized system architectures for every group (Islam et al., 2019). Real and Araujo took a chronological approach and structured their review in a timeline following the development of the navigation tools for the blind from the projects dating 1960s and 1970s to the current days (Real & Araujo, 2019). Elmannai and Elleithy (Elmannai & Elleithy, 2017) also provide a comprehensive analysis of visual assistive technologies. They focus on the vision substitution category divided into three subcategories: electronic travelling aids (ETAs), electronic orientation aids (EOAs), and position locator devices (PLDs). Publications also provide a quantitative evaluation of technological features. Mahida et al. and Plikynas et al. evaluated various wireless technologies and algorithms for indoor positioning solutions (Mahida et al., 2017; Plikynas et al., 2020).

The estimated growth of the VI population (Bourne et al., 2017), the positive user attitudes towards assistive technology (Griffin-Shirley et al., 2017), and advances in computer vision technology form a solid fundament for research and development (R&D) projects in the field. Regardless of the attempts to provide an overview of this rapidly developing field (Csapó et al., 2015; Islam et al., 2019; Jafri et al., 2014; Real & Araujo, 2019), a systematic literature review, focusing on

Accepted Manuscript

smartphone-based ETAs, is still lacking. The main strength of a systematic review is transparent methodology enabling reproducibility of the findings and limiting biases caused by human factors through an objective validity and relevance assessment of each included publication (Haddaway et al., 2018; Haddaway & Pullin, 2014). Therefore, the primary goal of this paper is a systematic review of the latest research related to computer-vision-based and smartphone-based ETAs. The Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) (Moher et al., 2009) methodology was followed to provide a systematic overview of publications indexed in three major databases: PubMed, IEEE Xplore, and the ACM Digital Library (DL).

This systematic review differs explicitly from the existing reviews in the field in terms of unique composition of some key research features like (i) application of standard biomedical systematic review methodology (PRISMA), (ii) investigation of ETA state-of-art computer vision software solutions, using smartphones with built-in cameras, and (iii) user-centred survey of blind ETA expert users, focusing on their needs and expectations regarding ETA functionality. It is important to note that the latter serves as an extended part of our systematic review. It provides essential user-defined ETA evaluation criteria for the reviewed prototypes.

To shed some light on the potential gap between the functionality of recently reported R&D prototypes and actual needs of VI individuals, we performed a survey of user experiences and expectations for ETAs (see the "Materials and Methods" section for details). Results of this survey add an extra dimension to the systematic review process and show to what extent user needs are or are not satisfied by the solutions proposed by the scientific community.

A combination of a systematic review methodology, emphasis on investigating state-of-the-art smartphone-centred computer vision navigation tools for the blind and comparison of academic development in the field (systematic review) to expert user experiences with the existing ETAs demonstrate the novelty of this paper. It summarizes the tendencies and future expectations of high-tech oriented VI individuals regarding ETA functionality and lists user-defined evaluation criteria for modern ETAs.

The article is structured as follows. After a short introduction and justification for this study, the method used for the systematic review is presented. The "Results" section then summarizes the findings of the systematic review of scientific publications. In the subsequent section, these findings are contrasted with the results from a VI expert ETA user survey, which adds an extra dimension to the systematic review results. The paper concludes with a discussion section that elaborates on the findings of the entire study and presents our conclusions.

Materials and Methods

To get a representative overview of the latest achievements in the field, a systematic review of scientific literature was performed. In contrast to other types of review techniques (traditional reviews, meta analyses), the strengths of a systematic review method lie in transparent and reproducible methodology for including all available evidence into the review and objective assessment of validity and relevance of each included study (Haddaway et al., 2018; Haddaway & Pullin, 2014). Due to its popularity and acknowledgement in medical informatics field, PRISMA statement was used to structure the review and reporting of results. PRISMA provides a framework,

Accepted Manuscript

covering entire literature review process from systematic search for publications, selection of relevant papers to result synthesis (Moher et al., 2009).

For scholars from different disciplines to better understand the rationale, benefits, and potential pitfalls of the presented systematic research approach, we (i) summarize the user-oriented usability evaluation methods, (ii) briefly compare and contrast it with other well-known evaluation and review methods, and (iii) present a narrative of the potential limitations.

In our literature review, we found that most of the user-oriented usability evaluation methods can be categorized into the following groups:

a) *Think-aloud protocol* (performed during testing, specifically while executing planned tasks, participants express their thoughts on the application; disadvantage: the environment is not natural to the user);

b) *Remote usability testing* (the experimenter does not directly observe the users while they use the application, although activity may be recorded for subsequent viewing; disadvantage: additional software is necessary to observe the participants from a distance);

c) *Focus groups* (a moderator guides a group of users in a discussion of the application; disadvantages: 1) moderators and leaders of the groups may be biased and lead to the inaccurate information, 2) the data collected tend to have low validity because of the unstructured character of the discussions);

d) *Interviews* (the users are interviewed to discover their experiences and expectations; disadvantages: 1) difficult to conduct remotely, 2) does not address the usability issue of efficiency);

e) *Cognitive walkthrough* (a team of evaluators walks users through the application, highlighting usability issues through the use of a paper prototype or a working prototype; disadvantages: 1) does not address user satisfaction or efficiency, 2) the designer may not behave like the average user when using the application);

f) *Pluralistic walkthrough* (a team of users, usability engineers, and product developers review the usability of the prototype of the application; disadvantage: does not address the usability issue of efficiency);

g) *Surveys* (targets a sample population with a questionnaire to elicit relevant opinions; disadvantages: 1) representative surveys cost time and money, 2) surveys are suboptimal for addressing the usability issue of efficiency);

h) *Field experiments* (unlike lab experiments, these are conducted in real-world settings and entail randomly assigning subjects (or other sampling units) to either treatment or control groups to test claims of causal relationships; disadvantages: 1) interference between subjects, 2) the local setting may not represent that of the population of interest);

i) *Observational studies* (can detect signs of the benefits and risks of prototype usage in the general population—the results of observational studies can be highly similar to those reported by similarly conducted randomized controlled trials; disadvantages: 1) additional causal factors

Accepted Manuscript

may have evaded observation/recordation, 2) recorded or unrecorded factors may be correlated, which may yield incorrect conclusions).

As is evident, these usability evaluation methods differ considerably, rendering them difficult to compare. Attempting objective review and comparison of research papers with such different self-evaluation methods poses an almost insurmountable challenge.

While searching for the most suitable method of review, we investigated the main types of literature reviews: evaluative, exploratory, instrumental, systematic, and integrative. Essentially, we were looking for a literature review method that focused on a research question in an attempt to identify, appraise, select, and synthesize all research evidence and arguments relevant to that question. In our opinion, a systematic review method was the optimal choice. Primarily, its meta-analysis aspect is well suited to effectively combine the various self-evaluation data reported in all selected studies to yield a more reliable review result. Moreover, we employed an integrative literature review approach to generate new knowledge on our topic through the process of review, critique, and synthesis of the literature under investigation.

Among systematic review methods, we chose PRISMA because it is internationally recognized and is a rigorous evidence-based minimum set of items for reporting in systematic reviews and meta-analyses of research papers (Moher et al., 2009, 2015). To make the meta-analyses more objective, we organized a semi-structured survey of blind experts; this survey yielded end-user evaluation of the main criteria that we could use for the estimation of the reviewed prototypes.

For comparison, however, consider a brief overview of other systematic review methods. First, a scoping review attempts to search for concepts, maps the language that surrounds them, and adjusts the search method iteratively (Pollock et al., 2017). It may be regarded as a preliminary stage before a PRISMA systematic review.

Second, living systematic reviews are a relatively new type of up-to-date semi-automated online summaries of research that are updated as new research becomes available (Elliott et al., 2014). Unfortunately, to our knowledge, no such living systematic reviews have been produced in the area of ETA for BVI people.

Third, Cochrane reviews comprise six types, and they compose a database of systematic reviews and meta-analyses that summarize and interpret the results of medical research (Silva et al., 2014). Cochrane is a group of over 37,000 specialists in healthcare who systematically review randomized trials of the effects of prevention, treatments, and rehabilitation interventions as well as health systems. When appropriate, they also include the results of other types of research. Although Cochrane systematic reviews are well known in the medical field, we do not find the use of such reviews to be well suited to the technological area of ETA research for BVI people.

Fourth, the quasi-standard for systematic review in the social sciences is based on the procedures proposed by the Campbell Collaboration, which is one of several groups that promote evidence-based policy in the social sciences (The Campbell Collaboration, 2014). It is a sister initiative of Cochrane, but for the social sciences. Thus, because of its purely social orientation, we did not find that it met our systematic review needs.

Accepted Manuscript

On the basis of the analysis of the weaknesses and strengths of the aforementioned systematic review methods, we chose to employ the universal, repeatable, and robust PRISMA method to summarize the review, evaluation, and meta-analysis of selected prototypes in the ETA research domain for BVI.

However, the PRISMA method also has weaknesses; the following is a list of key concerns:

1. For PRISMA to work well, it should build on a protocol that describes the rationale, hypothesis, and planned methods of the review. However, few reviews have reported whether a protocol exists (Moher et al., 2015). Detailed protocols can facilitate the understanding and appraisal of the PRISMA method.
2. Published PRISMA reviews are often biased, out of date, and excessively long (Roberts et al., 2015).
3. Some poor PRISMA research results were described by Altman (Altman, 1994): "much poor research arises because researchers feel compelled for career reasons to carry out research that they are ill equipped to perform, and nobody stops them."
4. Methodological limitations of meta-analysis have also been noted. Standardized meta-analysis methods continue to be unavailable to researchers, despite the necessity for standardized optimal methodological steps (Giang et al., 2019).
5. Another concern is that the methods used to conduct a systematic review are sometimes changed once researchers identify the available trials they are going to include (Page et al., 2014).
6. Significant publication bias is likely, with only “positive” results or those perceived to be favorable being published. A recent systematic review of industry sponsorship and research outcomes concluded that “sponsorship of drug and device studies by the manufacturing company leads to more favorable efficacy results and conclusions than sponsorship by other sources” and that the existence of an industry bias cannot be explained by standard “risk of bias” assessments (Lundh et al., 2017).

Despite the weaknesses mentioned earlier, the PRISMA systematic review method has strengths that make it a superior option to other review methods. Specifically, a) it is based on a protocol that describes the rationale, hypothesis, and planned methods of the review; b) because of its methodological rigor, it stands as a reference standard for synthesizing evidence in healthcare; and c) it has reduced arbitrariness and bias in decision-making with respect to extracting and including papers from primary research.

Three major research databases (PubMed, IEEE Xplore, and ACM DL) were queried using a combination of keywords: “visually impaired”, “blind”, “navigation”, “video”, “computer vision”, and “app”. The exact query used was as follows: ((“visually impaired” OR blind) AND navigation AND video) OR ((“visually impaired” OR blind) AND “computer vision”) OR ((“visually impaired” OR blind) AND video AND app)). The search was performed on the 5th of May, 2018, and covered a publication period of approximately five years (01 January 2013 to 05 May 2018). Older publications were not considered due to the rapid development of computer vision algorithms at a speed that can quickly

Accepted Manuscript

render previously published solutions irrelevant. Results were processed in the Zotero reference manager and inserted into an MS Excel spreadsheet for further analysis.

The dataset was first cleaned to remove duplicates. This resulted in 418 unique research publications included in our literature screening process. Based on a screening of titles and abstracts, 335 irrelevant papers were excluded from the review. Full-text analysis excluded an additional 68 papers. The main reasons for excluding publications after full-text analysis were as follows:

- No details were provided on implementation, and only a theoretical description of the proposed solution was given (n = 20).
- Did not meet our hardware assumptions, meaning the solution was not smartphone-centric, did not use video or images from the camera, or did not perform video processing (for instance, if raw video was forwarded to a human assistant for interpretation) (n = 32).
- Was focused on a very specific problem and did not provide any generalizability (for instance, a pothole detector) (n = 16).

Eventually, 15 research publications remained and were included in this systematic review. Our paper selection process is visualized in the PRISMA flow diagram in Figure 1.

Results

Assumptions we made before starting the review process shaped the results of this study. We emphasized smartphone-centred, computer vision-based solutions, which are analysed in detail in this review. All the included publications presented research prototypes; experimental activities using commercially available solutions for the VI were not identified.

Due to the varying maturity of the projects represented in the included publications, different evaluation and reporting practices, a strategy of highlighting the identified trends was selected for summarizing the results. Considering the pace computer vision technologies evolve, no individual choices made in selected publications are discussed in detail in this section. Instead major focus is paid to emphasizing the general directions this field is taking.

Purpose of the solution and functionality

Describing the surrounding environment through a set of video-processing algorithms is a complex task, incorporating a high level of uncertainty. Lighting conditions, movement, transparent and reflective objects, and region-specific aspects (e.g., the appearance of a rather standard object, such as a bus stop, can vary greatly in various regions) are only a few of the challenges that must be addressed by such computer vision systems. On a more general level, indoor conditions are different from those found in outdoor scenarios, therefore, more than half of the papers (n = 8) chose to focus on either indoor-only or outdoor-only tools, while the others (n = 7) aimed for universal solutions.

Navigation indoors is a well-known use case where traditional navigation systems based on GPS sensors fail. Five papers selected an indoor scenario as the main focus and addressed the challenges using various approaches. To estimate position and direction of movement, computer vision functionality was supplemented by data from motion sensors (Ko, 2013; Ko & Kim, 2017;

Accepted Manuscript

Rituerto et al., 2016). Such sensors were sufficient to track the location of a VI user within a static map of a building, while providing guidance towards the goal. However, motion sensors have their limitations – they require maps and continuous tracking of movements to localize the user. The use of fiducial markers to improve localization was proposed in three papers (two of which originated from the same project). Locations of interest were QR coded (Ko, 2013; Ko & Kim, 2017), simplifying the computer vision and localization tasks to the recognition of the QR codes. Pure computer vision solutions for indoor environments were discussed in two publications, demonstrating the feasibility of such systems (Elloumi et al., 2013; Garcia & Nahapetian, 2015).

The number of identified publications that focused exclusively on outdoor environments was slightly lower (n = 3). These papers presented three approaches to addressing problems related to navigation outdoors: 1) a pure computer vision system (Tapu et al., 2017); 2) an integration of computer vision techniques and GPS sensors (Zheng & Weng, 2016); and 3) a combination of computer vision, social networking, and geo data (Li et al., 2017). While the first two solutions were rather traditional, Zheng et al. presented a more advanced approach, supplementing online information (live video and GPS data) with offline data recorded by other users and tagged by Amazon's Mechanical Turk[1] service. Aiming to minimize the need for local computational resources required to process a live video stream by shifting the work from the smartphone to the cloud, the authors showed promising results in terms of accuracy and frame rate of operation, with a frame rate on average eight times better than in conventional solutions (Li et al., 2017).

Golledge et al. decomposed wayfinding process into three steps: representing current location, progressing on a target route and computing a novel route (Golledge et al., 1996). The majority of the identified papers (n = 8) focused on the first step of the wayfinding process (representing current location), providing object recognition and scene description functionality to the user. The other papers (n = 7) supplemented the first step of this process (representing current location) by some navigation functionality, representing the second step of wayfinding (progressing on a target route). No papers progressed to the third step defined by Golledge et al. where information collected along the route is used for computing novel routes in the surroundings (Table 1).

Both object recognition/scene description-focused and navigation-focused papers were able to recognize only a very limited set of objects. The number of object classes reported varied between four and five, with the exception of one paper that reported the ability to differentiate among "1,000 kinds of objects" with no further details (Bai et al., 2017). The most common objects recognized by the proposed solutions were vehicles, bicycles, pedestrians, and static obstacles (Mocanu et al., 2016; Tapu et al., 2017; Tapu, Mocanu, Bursuc, et al., 2013; Tapu, Mocanu, & Zaharia, 2013) in an outdoor scenario and doors, corridors, halls, and junctions (Ko, 2013; Ko & Kim, 2017) indoors. Depending on the implementation, performance measures for differentiating between moving objects (vehicle, bicycle, and pedestrian) ranged from 0.82 to 0.95 in terms of precision and from 0.69 to 0.96 in terms of recall. Measures were similar for detecting and recognizing static obstacles

[1] https://www.mturk.com

Accepted Manuscript

(precision = 0.9–0.93, recall = 0.79–0.95) (Mocanu et al., 2016; Tapu et al., 2017; Tapu, Mocanu, Bursuc, et al., 2013).

Input/Output

Due to the specifics of the inclusion criteria of this review, all selected publications used live camera feeds as input for the proposed tools. The use of QR-coded locations was proposed for indoor navigation (Ko, 2013; Ko & Kim, 2017) in combination with motion sensors and building maps (Ko, 2013; Ko & Kim, 2017; Rituerto et al., 2016). Mocanu et al. suggested the use of ultrasonic sensors, which work together with computer vision algorithms to enable accurate distance estimation to detected objects (Mocanu et al., 2016). Distance estimation to obstacles is also feasible using stereo cameras, described in two of the selected publications (Bai et al., 2017; Bułat & Głowacz, 2016). A typical output of all these systems was an audio signal, often communicated to the user through bone-conduction headphones. Eleven out of 15 publications used text-to-speech APIs, communicating directions in a selected time interval (typically every 1.5 – 2 seconds). Some solutions operated using a very limited set (3 - 6) of commands (Sharma et al., 2016; Zheng & Weng, 2016). Only one publication considered using beeping of various frequencies accompanied by text-to-speech to communicate the output of the system to the user (Ko & Kim, 2017). Details on the acoustic feedback were missing in 4 out of 15 included papers. Dasila et al. proposed using special binaural sound techniques in addition to the text-to-speech interface to enhance the user experience and immerse the user into a 3D audio surround-based representation of the environment (Dasila et al., 2017).

Data processing and performance

Our inclusion criteria made a smartphone a central component of all included papers. This component was responsible for both video capture and partial or full processing. The options for processing the live video stream could be easily divided into local (data was processed using the computational power of the smartphone) and remote (video was transmitted to a more powerful device or cloud infrastructure for processing). The majority of the solutions included in this review selected a local data-processing approach (n = 10), while the other five processed data remotely (two used a dedicated laptop computer (Rituerto et al., 2016; Tapu et al., 2017), while three transmitted the video stream to the cloud infrastructure for processing (Bai et al., 2017; Dasila et al., 2017; Sharma et al., 2016)).

Local video processing was implemented on a variety of Android and iOS smartphones. Solutions reached a performance level ranging from 5 frames per second (fps) (Zheng & Weng, 2016) to 10 fps (Ko & Kim, 2017; Tapu et al., 2017; Tapu, Mocanu, Bursuc, et al., 2013) in evaluation experiments (based on five papers). The choice of algorithms used in data-processing pipelines varied greatly. The most popular choices, resulting in highest performance (in terms of fps), were scale-invariant feature transform (SIFT) (Lowe, 2004) and speeded up robust features (SURF) (Bay et al., 2006).

In this review, the category of remote video processing encompasses all initiatives in which data processing was performed outside the smartphone. Therefore, the two publications which used

Accepted Manuscript

a user-carried laptop as the main source of processing power were also included in this group (Rituerto et al., 2016; Tapu et al., 2017). These two publications reported markedly different levels of performance, ranging from 2 fps to 20 fps, in the evaluation experiments. Two publications counted solely on the cloud infrastructure for live video stream processing, however, no performance measures were reported (Bai et al., 2017; Dasila et al., 2017).

A hybrid approach for data processing was proposed by Li et al., leveraging both internal computational resources on the smartphone device and cloud processing (Li et al., 2017). The solution maintained a database of previously visited geotagged and manually labelled locations collected by all users, which were used to accelerate the processing when a specific user appeared in a previously visited location. Tests demonstrated that this approach could boost the performance of the general object-detection algorithms eightfold, making them easier to adopt on mobile platforms (Li et al., 2017).

Algorithms used in data processing

Algorithms used in data-processing pipelines often rely on classical two-step composition: the camera image is first processed using engineering-based computer vision techniques (e.g., feature and descriptor extraction), followed by a machine learning model (e.g., a classifier) utilising the output of the first step. Common choices for the first step were feature descriptors (e.g., histogram of oriented gradients (HOG) (Dalal & Triggs, 2005), SIFT (Lowe, 2004), and SURF (Bay et al., 2006)). For example, two publications used HOG for classifying detected obstacles (Tapu, Mocanu, Bursuc, et al., 2013; Tapu, Mocanu, & Zaharia, 2013), two used SURF descriptors to recognize visual situations and construct a guidance system (Ko, 2013; Ko & Kim, 2017), and one proposed a system relying on features from accelerated segment test (FAST) (Rosten et al., 2010) descriptors for obstacle handling (Mocanu et al., 2016). Support vector machines (SVMs) (Mocanu et al., 2016; Tapu, Mocanu, Bursuc, et al., 2013; Tapu, Mocanu, & Zaharia, 2013) and template matching (Ko, 2013; Ko & Kim, 2017) were identified as popular machine learning methods for the second step. These methods are fast to execute even in embedded computing platforms, which makes them attractive in low-computational-resource scenarios.

Another class of algorithms used in the identified publications counted on direct computer vision approaches without an adaptive machine learning component. For example, Garcia and Nahapetian [12] tried to detect a corridor by using Canny edges followed by a Hough transform (Garcia & Nahapetian, 2015; Gonzalez & Woods, 2008), and Elloumi et al. proposed a camera localisation algorithm based on orthogonal vanishing points (Elloumi et al., 2013). Both approaches were intended to work only in indoor environments by design, and tests conducted by the authors were limited to a single environment, which makes practical evaluation of the results rather preliminary.

It is important to note, that only a few studies (i.e., (Sharma et al., 2016; Tapu et al., 2017; Zheng & Weng, 2016)) investigated computer vision methods based on deep neural networks (DNNs), which perform feature extraction and task modelling (e.g., classification or recognition) steps in a single differentiable neural structure. The parameters of both steps are estimated from the data, optimising a single cost function that corresponds not only to optimisation of the task modelling component but also to feature optimisation. In most cases, neural networks are known to achieve

Accepted Manuscript

better and more robust results compared to approaches that rely on the aforementioned engineering-based feature extractors (LeCun et al., 2015; Schmidhuber, 2015; Voulodimos et al., 2018).

Some authors noticed an important parallel with computer vision-based autonomous robot localisation and mapping methods (e.g., simultaneous localisation and mapping (SLAM) and position estimation via particle filtering) (Bai et al., 2017; Elloumi et al., 2013; Rituerto et al., 2016). This approach provides another useful and under-explored resource for developing assistive systems for VI individuals.

Evaluation of the proposed solutions

Evaluation of proposed solutions is an important part of the research and innovation process. It demonstrates technical feasibility, performance, user acceptance, and impact of the proposed system. Almost all publications (n = 13) included in this review presented some degree of evaluation, while in two publications (Elloumi et al., 2013; Tapu, Mocanu, & Zaharia, 2013) it was either missing or proved only the technical feasibility of the solution.

Evaluation using video/images recorded beforehand (n = 7) was almost as common as user-based evaluation methods (n = 6). Four publications combined both video/image-based and user-based evaluations. Video/image-based evaluations focused primarily on the performance of the computer vision algorithms used in the system. None of the publications used standardised video/image libraries, which are common in benchmarking computer vision algorithms. Instead, materials for the evaluations were collected by the project teams (Bułat & Głowacz, 2016; Garcia & Nahapetian, 2015; Li et al., 2017; Sharma et al., 2016; Tapu et al., 2017; Tapu, Mocanu, Bursuc, et al., 2013; Tapu, Mocanu, & Zaharia, 2013).

User-based evaluation was present in six out of 15 publications, some of which included blind participants (Rituerto et al., 2016). Most of the evaluation procedures were limited in scope, including only three or four participants (Ko, 2013; Rituerto et al., 2016), while one study reported the inclusion of 21 users (Mocanu et al., 2016). Reporting of results varied among the studies, making it impossible to establish any comparisons. Measures ranged from user satisfaction (Ko & Kim, 2017) to accuracy of the algorithms in training datasets (Bai et al., 2017).

Summary of the findings

The findings of this systematic review are summarised in Table 1.

How well are the needs of actual users addressed in the scientific publications?

Sections above followed the PRISMA systematic review method for summarizing scientific evidence on smartphone-centred ETA research [31]. The results highlight the latest R&D achievements and trends. In the reminder of the paper we extend the PRISMA method by scoring the reviewed publications according to the requirements for ETA solutions defined in our expert-user survey. Such scoring complements the systematic review, adding a new dimension and enabling us to evaluate the

Accepted Manuscript

included publications according to user-defined compliance criteria. Such estimates reveal actual market needs in contrast to the solutions and functionalities prioritised by the academic community.

User requirements were collected through a semi-structured online survey, completed by N = 25 blind expert users, having 10+ years of experience (or active interest) in using ETAs for the blind. These mostly EU located blind experts do represent blind individuals with a cutting-edge interest in the newest ETA gadgets and ability to use novel technologies[2]. They, however, cannot represent a majority of the worldwide blind population. In this regard, our survey of 25 blind experts is solely dedicated to a) highlighting tendencies of modern high-tech oriented blind people needs regarding ETA functionality, and b) clarifying ETA evaluation criteria defined by blind experts. Thus, our blind experts' survey is specialized and should not be interpreted as a large-scale representative survey.

The questionnaire used when collecting the blind expert feedback is provided in the supplemental file (see article's page in the journal website). This questionnaire was designed together with two VI individuals after a series of focus group meetings and discussions. Final version of the questionnaire was evaluated by collaborating VI persons before starting the survey. Please note that 10 out of 25 blind experts were interviewed live during the survey. It gave us deeper insights about the criteria for evaluating current ETA R&D prototypes.

The 39-question survey covers demographics, sight-related aspects, and experiences with existing ETA systems. The survey was completed in February 2019. Demographics of the participants are presented in Table 2.

The sample of the expert users is characterized by a mean age of 33 years and includes mostly employed (52% fully employed, 20% partially employed) individuals, residing in big cities (84%). Sixty percent of the participants had higher education, and the group had an average professional work experience of 11 years (Table 2). The experts were identified based on recommendations from associations of blind and VI individuals and, therefore, their demographics may not be representative for the general population of the VI. The level of expertise was determined by an online questionnaire.

To uncover the essential ETA features, we employed open-ended hierarchical questions. Each respondent was asked to list up to five key problems in navigating indoors and outdoors in order of decreasing importance (a score of 5 indicates the highest importance, while 1 is the lowest). Items populating these lists were not provided in the survey and had to be filled in by the participants forming a hierarchy of preferences for every participant. The 20 most common problems of highest importance were extracted (Table 3, Table 4). Based on the importance scores, we calculated averages for each criterion (i.e., each identified problem). To adjust for overestimation or underestimation of some criteria, average values were normalized by multiplying them by weight

[2] In a few pages of this publication, we only provide a glimpse of our survey results. The main body of the survey will be published in a dedicated article. In total, 78 VI (blind) individuals participated in this study, only 25 of them qualified as experts.

Accepted Manuscript

coefficients, which took into consideration 1) the proportion of respondents who picked that criterion and 2) the proportion of the sum of importance scores S_i for the selected criterion i with respect to the sum of all scores for all criteria. This procedure reduced the bias in the estimates. Then, the weighting rate W_i of each criterion i was calculated using Equation 1.

$$W_i = Mean_i * \frac{n_i}{N} \frac{S_i}{\sum_j S_j}, \quad (1)$$

where $Mean_i$ denotes the mean importance score for criterion i; $\sum_j S_j$ denotes sum of scores of all included criteria; n_i the number of respondents who picked criterion i, and N is the total number of respondents (N = 25).

Navigation criteria for outdoors and indoors were scored separately. The 10 criteria with highest weighting rates W_i for each location are listed in Table 3 and Table 4. It is important to notice that Tables 3 and 4 provide estimates for just the top 10 criteria. However, Eq. 1 uses estimates from all criteria, including those which are not included in the tables.

Criteria listed in Table 3 and Table 4 were used to evaluate the publications included in the systematic review. To assess how well each article and the associated prototype (Table 1) corresponds to the actual needs of the end users (see Table 3 and Table 4), we calculated the sum of weighting rates W_i multiplied by binary labels (0 if a prototype does not address the criterion and 1 if a prototype addresses the criterion). This way, for each article and associated prototype, we obtained a quantitative estimate of compliance that takes into account a weighted set of user-defined evaluation criteria. We evaluated and compared articles using both outdoor and indoor sets of user-defined evaluation criteria (see Figure 2).

Figure 2 indicates that user-defined needs and expectations for outdoor navigation were addressed best in publications (Bai et al., 2017; Li et al., 2017; Tapu et al., 2017), while they were poorly addressed in the articles (Dasila et al., 2017; Elloumi et al., 2013; Garcia & Nahapetian, 2015; Ko, 2013; Ko & Kim, 2017; Rituerto et al., 2016; Tapu, Mocanu, & Zaharia, 2013). Indoor navigation criteria were addressed best in the articles (Bułat & Głowacz, 2016; Elloumi et al., 2013; Garcia & Nahapetian, 2015; Ko, 2013; Ko & Kim, 2017; Rituerto et al., 2016; Sharma et al., 2016), while publications (Bai et al., 2017; Dasila et al., 2017; Li et al., 2017; Tapu et al., 2017; Tapu, Mocanu, Bursuc, et al., 2013; Tapu, Mocanu, & Zaharia, 2013; Zheng & Weng, 2016) addressed them poorly. No publications addressed both outdoor and indoor evaluation criteria well. Two publications ((Dasila et al., 2017; Tapu, Mocanu, & Zaharia, 2013)) addressed both outdoor and indoor criteria poorly. To highlight the gap between the focus of academic community (systematic review) and actual user needs (survey of expert users), we estimated the differences between the two. For each criterion, all labels (0 if a prototype does not address the criterion and 1 if a prototype addresses the criterion) assigned for the articles were summed up and divided by the total number of all assigned labels for all criteria. This way we estimated the weighted importance of each criterion from the perspective of scientific publications. Experts' estimates were obtained from Table 3 and Table 4, using percentage expressions of the weighting rate W_i. Finally, we calculated the difference between experts' estimates and scientific publication's estimates (%) (see Figure 3 and Figure 4). A high positive difference indicates a high underestimation of the importance of the criterion by the

Accepted Manuscript

academic community compared with the experts' estimate. A high negative difference indicates a high overestimation of the importance of the criterion by the scientific community compared with the experts' estimate. Small differences indicate similar estimates on the importance of the corresponding evaluation criterion (Figure 3).

Figure 3 shows a clear mismatch between the expectations and needs of end users and the assumptions of scientific community regarding the importance of evaluation criteria for outdoor navigation ETAs'. The weighted importance of expert-defined criteria gradually diminishes from #1 to #10 (Figure 3, blue columns). However, in scientific publications an opposite trend may be identified (Figure 3, orange columns). This discrepancy means that scientific publications place more value on less important criteria and disregard the most important ones identified by users. Criteria #1 to #6 (Pedestrian crossings, Finding the elevator, Reading numbers in the bank, etc.) are highly underestimated by the researchers, while criteria #6 (-26.82%, Finding objects (e.g. shops, hotels, WC, etc.)) and #10 (-32.95%, Unexpected obstacles on the passages, etc.) are considerably overestimated by the researchers (Figure 3 and Table 3). This observation is alarming and shows a significant mismatch between user needs and the efforts of the academic community.

We analysed the set of indoor evaluation criteria in a similar manner. Differences between the experts' and researchers' priorities are depicted in Figure 4.

The analysis of the estimated differences highlights another important observation with practical implications for indoor ETA development. There is a major mismatch between expectations and needs of end users and the focus of scientific community regarding the importance of criterion #1 (Finding room by number) (see Table 4). The priority of this criterion is greatly underestimated by researchers, while it is of major importance to users (64.2%, Figure 4). The values of other criteria are overestimated by researchers: for instance, criteria #5 (-14.78%, Recognition of objects) and #7 (-17.10%, Detection of obstacles) (Figure 4).

Conclusions and Discussion

The findings of this review show a relatively strong focus of the academic community worldwide on developing computer vision-based travelling aids for the blind. The accuracy, robustness, and efficiency of computer vision algorithms are improving, fuelled by the novelties in neural network-based models and the increasing availability of computational resources. However, in the identified research applications, this class of algorithms is adopted only to a limited extent (Tapu et al., 2017; Zheng & Weng, 2016), hinting that academic community may struggle to utilise novel state-of-the-art image/video-processing algorithms. The performance of these novel algorithms is highly dependent on both quantity and quality of training data. While general purpose computer vision datasets exist, specialized training data tailored to the needs of the VI are not yet publicly available. Lack of benchmark datasets slow down the adoption of the state-of-the-art computer vision algorithms and hinders objective comparison and evaluation of the developed solutions.

Involvement of the end-users in the design and development process could be perceived as a more qualitative approach to the aforementioned benchmarking problem. Domains such as medical informatics were dealing with formalizing experiments with users and have developed several frameworks for structuring and reporting user trials in a more objective manner. Model for

Accepted Manuscript

Assessment of Telemedicine (MAST) (Kidholm et al., 2012) and Statement on Reporting of Evaluation Studies in Health Informatics (STARE-HI) (Talmon et al., 2009) are good examples of evaluation and reporting frameworks that could improve the quality and reproducibility of user-based experiments.

In the reminder of this section we discuss the findings of this work in a light of relevant research and new technology frontiers. To generalize the findings, we present them in a more structural way, motivating prospective ETA research addressing the limitations of existing publications, and provide insights for potential improvements.

Limited use of haptic interfaces

While there is adoption of the latest computer vision algorithms in developing travelling aids for the blind, interfaces to communicate the findings of these systems to the VI user are limited to the audio channel (based on the papers included in this review). We observed, however, that the majority of the selected papers did not delve deeper into the development of the auditory interface design (with the exception of paper (Dasila et al., 2017)), even though the main sensory information channel for the blind is auditory (Csapó & Wersényi, 2013). Similarly, we did not find novel, experimentally verifiable considerations suggesting new trends in “tactification” or “haptification” of visual information (Maclean & Enriquez, 2003). Moreover, we did not identify any attempts to design interfaces for controlling the proposed devices that would be tailored according to the needs of VI individuals. This implies that the selected research papers are mostly technology-centric (driven by the application of mobile technologies) instead of being user-centric (developing novel interfaces according to the needs of blind individuals). These trends, however, have to be considered with caution. They were identified in a rather specific subset of publications, putting emphasis on computer vision-based ETAs.

Audio feedback is relatively simple to implement; however, it may not be the most efficient way to convey information to the user. Alternative mechanisms in the form of vibrating bracelets (Scheggi et al., 2014), gloves with micro-motors (Advani et al., 2017; Poggi & Mattoccia, 2016), vibrating belts (H.-C. Wang et al., 2017) and other means of tactile feedback (Peiris et al., 2016) have been reported, demonstrating the feasibility of enhancing acoustic feedback. Even though the combination of computer vision input and haptic output was not identified in the selected publications, both are quickly evolving fields that may bring novelties to the design and development of ETA systems for the blind. After reviewing haptic assistive technologies for audition and vision sensory disabilities, Sorgini et al. pointed out lacking acceptance of haptic interfaces among users and suggested focusing research effort on miniaturized, low-cost haptic interfaces, integrating with personal devices, such as smartphones (Sorgini et al., 2018). A variety of such interfaces have already been suggested to support navigation (Csapó et al., 2015; Meier et al., 2015; Sorgini et al., 2018). However, these systems have not coupled haptic feedback with computer vision algorithms and, therefore, were not analysed in detail in this review.

No evaluation of commercial applications

Commercial tools designed to meet the needs of the VI have existed for many years in the form of standalone devices (such as Trekker, designed and manufactured by HumanWare and launched in

Accepted Manuscript

2003) and smartphone applications (including the previously mentioned BlindSquare and TapTapSee, both released in 2012). Regardless of the availability of such applications, this review did not identify any research projects using existing applications. Instead of reusing, in many cases researchers focused on creating tools, delivering functionality similar to that already available in the market.

The fact that scientific community is not taking the available tools into consideration is alarming. Research on existing tools is an important process in identifying weaknesses and limitations and contributing to the development of improved tools that better meet the requirements of VI individuals. The available tools rarely cover all user requirements and need to be used in combination to deliver necessary functionality. Systems of such tools have high potential for improving quality of life for VI individuals, however, research is lacking on how such tools should be integrated into an ecosystem to maximise gains for users.

Limited use of state-of-the-art computer vision algorithms

The conducted systematic literature review revealed that smartphone-based computer vision tools for blind individuals often rely on rather outdated image/video-processing methods (e.g., SURF (Bay et al., 2006) and SIFT (Lowe, 2004), among others) that do not reflect current state of the art. Such approaches are not sufficiently efficient and lack robustness in real-world conditions, which is essential in this particular use case. On the contrary, computer vision methods based on DNNs are far less limited in this sense (LeCun et al., 2015; Schmidhuber, 2015; Voulodimos et al., 2018). For example, DNNs can be trained to perform object classification (Szegedy et al., 2016), specific object detection (Dai et al., 2016; Huang et al., 2017), face recognition (Amos et al., 2016), scene description (Liu et al., 2017), wayfinding (OhnBar et al., 2018), obstacle detection (Pinard et al., 2017), and other potentially useful tasks for the blind. For certain problems (e.g., object classification or detection), DNNs may reach accuracy comparable to human decisions in real-world conditions (He et al., 2015). To construct an efficient DNN-based model, however, high-quality training datasets are essential [27], [29]. Although quite large general object detection (Lin et al., 2014) or classification (Deng et al., 2009) datasets are publicly available, they do not include some object classes that may be important for the blind. For instance, neither ImageNet (Deng et al., 2009) (classification) nor COCO (Lin et al., 2014) (object detection) datasets include “corridor”, “stairs”, “elevator”, and other potentially useful objects, especially for indoor navigation scenarios. Construction and publication of a comprehensive, high-quality dataset tailored according to the needs of the VI would contribute significantly to further development of computer vision-based systems for the blind.

Training DNNs is a computationally intensive task, requiring specific hardware (e.g., high-end graphic processing units, or GPUs). However, highly specialised neural networks can often be replaced by simpler, specially designed DNN components, (for instance, mobilenet (Sandler et al., 2018), peleenet (R. J. Wang et al., 2018), and NASNet (Zoph et al., 2018)). These models achieve close to real-time inference and provide comparable performance to the original large, specialised models even on low-computational-power devices such as smartphones. Moreover, cloud technologies and the high availability of 4G networks allow for the adoption of even larger and more complex DNN-based models that cannot be efficiently deployed on mobile handsets. The emerging 5G networking will boost the effectiveness of DNN-based cloud-hosted applications in scenarios where both mobile network speed and latency are critical.

Accepted Manuscript

An important feature of DNN-based methods is their capability to process multi-sensor data (for instance, fused image, GPS, and inertial measurement unit data) (Clark et al., 2017; Zhang et al., 2018). These data sources are relatively easily accessible (especially in a smartphone-based system scenario) and can potentially increase the accuracy and robustness of the entire system.

Development of smartphone-based assistive technologies for visually impaired users, especially relying on locally implemented computer vision algorithms, inevitably includes power consumption optimization issues, since computer vision algorithms usually are computationally demanding. Papers included in this review, however, did not provide any information about this feature. It may be explained by the fact that power consumption characteristics are highly dependent not only on the algorithms, but also on their implementation and hardware. It is necessary to note that the reviewed papers mostly report early results of their prototypes R&D stage when many relevant design and practical implementation details are yet to be considered. However, neglecting power consumption estimates can be regarded as a definite drawback, as it can cause severe implications and constraints in the later implementation stages.

In recent years, various libraries (e.g. QNNPACK[3]) and low-cost hardware solutions (e.g. Coral.ai[4] (~2 Watts), Intel Movidius[5] (~1 Watt)) were proposed for facilitating the use of modern computer vision algorithms (e.g. deep neural network-based object detection) on low power mobile devices. Hence, additional 1-2 Watts can be regarded as estimated power consumption increase for hardware-accelerated computer vision algorithm implementations.

Evaluation of the proposed solutions

Evaluation of the proposed solutions was identified as a weak point in many of the proposed solutions. When evaluating technical feasibility and performance, lack of using standard computer vision benchmark datasets was observed. Relatively small datasets, collected by the project teams were utilized, making it impossible to compare the performance between various solutions. It may have been caused by the lack of open computer vision datasets tailored to developing solutions for the VI. Generic benchmark datasets, such as ImageNet (Deng et al., 2009) (classification) and COCO (Lin et al., 2014) (object detection) could potentially be used to some extent. However, these datasets lack objects of vital importance to the VI, for example corridors, stairs, elevators. A collaborative effort for creating a high-quality benchmark dataset for developing and evaluating computer vision solutions for the VI is required.

User-based evaluation presents a common challenge in the academic community. Only six out of 15 included papers had some user evaluation present, many limited in scope. The involvement of VI individuals to research projects is challenging and often suffer from selection bias. Moreover,

---

[3] Web link https://engineering.fb.com/ml-applications/qnnpack/

[4] Web link https://coral.ai

[5] Web link https://www.movidius.com

Accepted Manuscript

lack of standardized evaluation methods and potential reporting bias limit the representativity of these experiments.

Among the reviewed papers, we did not find any ETA related R&D prototypes thoroughly tested in a representative sample of blind users. Thus, it may be reasonable to assume that most of the published research papers report results of early development stages, when prototypes are not yet ready for the full-scale testing. Instead, some researchers report limited-scope testing of their prototypes' functionality without organizing time-consuming large-scale population surveys. Such surveys in many cases are more economically and technically feasible for the commercially available releases (beta versions), when users' feedback helps to improve the characteristics of the final product. Performing end-user evaluation too late in the development process brings a risk of disconnection from the actual user needs. Prototypes optimised and tested in the lab often fail to deliver the same performance in real-life settings. Publications also show that some prototypes are tested by the researchers developing the solutions (for example blindfolded) without consulting VI users. Such tests are likely to overestimate the performance comparing to end-user testing. Besides, the review highlighted a narrow-scope testing problem, when performance of a specific feature is reported without consideration of the integral functionality and performance of the system.

The primary focus of our systematic review is on new and ongoing technological developments and trends, which in most cases are still in the process of being developed or made, they are not necessarily fully implemented systems at this point. Therefore, it is an inescapable fact that in most of the selected R&D cases, full testing of research prototypes, which are still in the process of development, is not feasible. Of course, if the current R&D stage results in successful implementation, full demographic testing will be undertaken. Nevertheless, knowledge of such ongoing technological developments is highly valuable for other researchers; it helps to direct their research efforts and estimate future implications. Therefore, the main inclusion criteria do not restrict the selection to only fully implemented ETA systems that have been tested with representative demographic samples of BVI persons. However, we emphasize that even in the early development stages, researchers should undertake thorough end-user evaluations and testing of prototypes. In addition to previously mentioned advantages, such undertakings also save time and effort in the final development stages.

While user testing was weakly represented in the included papers, good examples of including more than five VI individuals in the evaluation process exist in literature (Ahmetovic et al., 2011; Fusco & Coughlan, 2018; Manduchi, 2012; Neat et al., 2019). Many publications report evaluation based on two-three subjects (Fusco et al., 2014; Ivanchenko et al., 2008; Manduchi et al., 2010; Schauerte et al., 2012; Shen & Coughlan, 2012). Due to the limited length of covered publication period and selection of research databases, the aforementioned publications were not included in this review and are only used to discuss our findings. Naturally we noticed more publications reporting user testing originating from accessibility, human factors, human-computer interfaces-oriented publication channels. Technology-focused publishers put more emphasis on technology advances and push user evaluation to the second plan. Paper inclusion criteria made this review to favour technology- over usability-oriented publications, making sufficient technical details essential to evaluate the proposed solutions. Naturally, publications mostly focused on user experiments received less attention. Even though the trends identified in this review may be influenced by the choice of research databases and paper inclusion criteria, they communicate an

Accepted Manuscript

important finding – involvement of end-users in testing assistive technology is insufficient and should be improved, especially in technology-oriented publications.

Safety and reliability of the developed ETA prototypes is of vital importance for the well-being of the VI users. It is essential to take safety and reliability concerns into consideration from the initial stages of the research project. Therefore, researchers and developers should take methodological approaches more common to medical field, such as literature reviews, surveys, evaluations, performance and user testing, into account. For instance, PRISMA systematic review method, employed in this paper, serves well in medical research and also in other domains to obtain a representative overview of the latest achievements in the field (Moher et al., 2009). In addition, standardized testing and evaluation of user experience is advisable as an integral part of development and implementation phases. It enables performance comparison of various prototypes using common standards.

Compliance with user needs and experiences

The results of our systematic review find relatively strong interest of academic community in addressing the needs of VI individuals through smartphone-based ETAs. However, our analysis shows that academic initiatives are often disconnected from the needs of end users. For instance, it is noteworthy to observe that the majority of VI experts have chosen “finding room by number” as the most critical problem in indoor navigation (Table 4), but this task has not even been mentioned as important in most of the selected publications. The importance of some criteria was highly underestimated, and others were overestimated in the reviewed articles (Figure 3 and Figure 4).

Some mismatches between the foci of users and academic community may be explained by the specialized nature of the prototypes. Even though the participants in the survey clearly indicated the need for a convenient, all-in-one solution for navigating both indoors and outdoors, researchers tended to separate these environments. This, however, may not be surprising: from a technical point of view, outdoor and indoor environments are rather different, requiring a different set of techniques to provide the necessary functionality. This is also visible in the two almost non-overlapping sets of problems the VI encounter while navigating outdoors (Table 3) and indoors (Table 4) as identified by expert users.

While evaluating the selected articles and their corresponding prototypes, we noticed several limitations. For instance, some authors do not perform validity and robustness testing of the proposed prototypes in various outdoor and indoor conditions or do not elaborate on the details of their testing results. Prototypes are often tested in lab conditions without involving actual users. Moreover, the majority of the papers did not address navigation, object-recognition, and obstacle-detection problems in various realistic conditions (e.g., bad weather or lack of illumination).

Limitations

The results reported in this systematic review should be interpreted keeping the following limitations in mind.

Accepted Manuscript

This review is relatively restricted in scope. It fails to provide a comprehensive overview of the ETAs developed for VI persons. Instead, it analyses a rather specific subset of articles representing smartphone-centred, computer vision-based tools for the blind. These inclusion criteria rendered many of the publications identified in the search irrelevant. Some publications that selected computational platforms other than the smartphone in the reported prototypical implementations may have been discarded even though they may have the potential for becoming smartphone-centric in their more mature project stages.

The search for publications was performed in three major research databases (Pubmed, IEEE Xplore, and ACM DL) using a limited set of keywords and limiting publication dates (1 January 2013 – 5 May 2018). The choice of databases and search keywords may have left some important publications out of this review. Only publications in English were included.

Some of the methodological limitations were inherited from PRISMA statement that was used to structure the review process. PRISMA was developed for medicine and health sciences and is adopted in other fields to a limited extent. Medical terminology and lack of guidance for performing the review may have limited its adoption in other fields (Haddaway et al., 2018). Publication search process in PRISMA counts on reproducible queries in major research databases, potentially missing publication channels that are not indexed there. While selectively including publication channels could be an option, it is difficult to produce an exhaustive list of journals, relevant for a specific field. Manual selection of publication channels increases subjectivity and reduces transparency of the paper search.

Paper inclusion process is based on strict criteria that are defined beforehand and may be limited to rather specific questions. Critical attitude towards this approach was expressed by Haddaway et al., pointing out that these criteria minimize selection bias, rather than limitations to validity (Haddaway et al., 2018). Publications meeting the inclusion criteria do not necessarily represent the top-rated papers in the field. Instead, they meet formal requirements defined by the authors aiming to answer study-specific questions. This approach has its pros and cons and its acceptance may vary depending on the discipline. While formalized paper inclusion is transparent and reproducible, important publications may be discarded due to minor discrepancies. To be able to compare various ETA solutions in this review, inclusion criteria put emphasis on the presence of technical description in the manuscripts. Naturally, papers mostly focusing on user testing and lacking technical details may have been left out. Findings of this review have to be considered having inclusion and exclusion criteria in mind.

Our expert-user survey has its own weaknesses. Small sample size and a predominance of participants residing in the EU region introduce biases and limit the representativeness of the findings. Predominance of EU residents in the survey may come as a contradiction to the fact that the majority of VI individuals live in low-income developing countries (Bourne et al., 2017). Underrepresentation of low-income countries in the survey hints a limited adoption of advanced ETAs by the residents in these regions due to availability, economic and other factors. This survey aimed to highlight trends and identify user-defined evaluation criteria for ETA solutions, rather than provide statistically significant findings. Therefore, the distribution of participants is skewed towards active ETA user naturally living in higher income countries.

Accepted Manuscript

Funding

This project has received funding from European Regional Development Fund (project No 01.2.2-LMT-K-718-01-0060) under grant agreement with the Research Council of Lithuania (LMTLT).

Conflicts of Interest

None declared.

References

Advani, S., Zientara, P., Shukla, N., Okafor, I., Irick, K., Sampson, J., Datta, S., & Narayanan, V. (2017). A Multitask Grocery Assist System for the Visually Impaired: Smart glasses, gloves, and shopping carts provide auditory and tactile feedback. *IEEE Consumer Electronics Magazine*, *6*(1), 73–81. https://doi.org/10.1109/MCE.2016.2614422

Ahmetovic, D., Bernareggi, C., & Mascetti, S. (2011). Zebralocalizer: Identification and localization of pedestrian crossings. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, 275. https://doi.org/10.1145/2037373.2037415

Altman, D. G. (1994). The scandal of poor medical research. *BMJ*, *308*(6924), 283–284. https://doi.org/10.1136/bmj.308.6924.283

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications*. CMU-CS-16-118, CMU School of Computer Science.

Bai, J., Liu, D., Su, G., & Fu, Z. (2017). A Cloud and Vision-based Navigation System Used for Blind People. *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies*, 22:1–22:6. https://doi.org/10.1145/3080845.3080867

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *Computer Vision – ECCV 2006* (pp. 404–417). Springer Berlin Heidelberg.

Accepted Manuscript

Bourne, R. R. A., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Resnikoff, S., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y., Taylor, H. R., & Vision Loss Expert Group. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *The Lancet. Global Health*, *5*(9), e888–e897. https://doi.org/10.1016/S2214-109X(17)30293-0

Bułat, J., & Głowacz, A. (2016). Vision-based navigation assistance for visually impaired individuals using general purpose mobile devices. *2016 International Conference on Signals and Electronic Systems (ICSES)*, 189–194. https://doi.org/10.1109/ICSES.2016.7593849

Chessa, M., Noceti, N., Odone, F., Solari, F., Sosa-García, J., & Zini, L. (2016). An integrated artificial vision framework for assisting visually impaired users. *Computer Vision and Image Understanding*, *149*, 209–228. https://doi.org/10.1016/j.cviu.2015.11.007

Clark, R., Wang, S., Wen, H., Markham, A., & Trigoni, N. (2017). VINet: Visual-inertial Odometry As a Sequence-to-sequence Learning Problem. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 3995–4001. http://dl.acm.org/citation.cfm?id=3298023.3298149

Csapó, Á., & Wersényi, G. (2013). Overview of Auditory Representations in Human-machine Interfaces. *ACM Comput. Surv.*, *46*(2), 19:1–19:23. https://doi.org/10.1145/2543581.2543586

Csapó, Á., Wersényi, G., Nagy, H., & Stockman, T. (2015). A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research. *Journal on Multimodal User Interfaces*, *9*(4), 275–286. https://doi.org/10.1007/s12193-015-0182-7

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object Detection via Region-based Fully Convolutional Networks. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 379–387. http://dl.acm.org/citation.cfm?id=3157096.3157139

Accepted Manuscript

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *1*, 886–893. https://doi.org/10.1109/CVPR.2005.177

Dasila, R. S., Trivedi, M., Soni, S., Senthil, M., & Narendran, M. (2017). Real time environment perception for visually impaired. *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, 168–172. https://doi.org/10.1109/TIAR.2017.8273709

Deng, J., Dong, W., Socher, R., Li, L., & and. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., & Gruen, R. L. (2014). Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. *PLoS Medicine*, *11*(2), e1001603. https://doi.org/10.1371/journal.pmed.1001603

Elloumi, W., Guissous, K., Chetouani, A., Canals, R., Leconge, R., Emile, B., & Treuillet, S. (2013). Indoor navigation assistance with a Smartphone camera based on vanishing points. *International Conference on Indoor Positioning and Indoor Navigation*, 1–9. https://doi.org/10.1109/IPIN.2013.6817911

Elmannai, W., & Elleithy, K. (2017). Sensor-Based Assistive Devices for Visually-Impaired People: Current Status, Challenges, and Future Directions. *Sensors (Basel, Switzerland)*, *17*(3). https://doi.org/10.3390/s17030565

Fricke, T. R., Tahhan, N., Resnikoff, S., Papas, E., Burnett, A., Ho, S. M., Naduvilath, T., & Naidoo, K. S. (2018). Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling. *Ophthalmology*, *125*(10), 1492–1499. https://doi.org/10.1016/j.ophtha.2018.04.013

Fusco, G., & Coughlan, J. M. (2018). Indoor Localization Using Computer Vision and Visual-Inertial Odometry. In K. Miesenberger & G. Kouroupetroglou (Eds.), *Computers Helping People with*

Accepted Manuscript

Special Needs (Vol. 10897, pp. 86–93). Springer International Publishing. https://doi.org/10.1007/978-3-319-94274-2_13

Fusco, G., Shen, H., Murali, V., & Coughlan, J. M. (2014). Determining a Blind Pedestrian's Location and Orientation at Traffic Intersections. In K. Miesenberger, D. Fels, D. Archambault, P. Peňáz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (Vol. 8547, pp. 427–432). Springer International Publishing. https://doi.org/10.1007/978-3-319-08596-8_65

Garcia, G., & Nahapetian, A. (2015). Wearable Computing for Image-based Indoor Navigation of the Visually Impaired. *Proceedings of the Conference on Wireless Health*, 17:1–17:6. https://doi.org/10.1145/2811780.2811959

Giang, H. T. N., Ahmed, A. M., Fala, R. Y., Khattab, M. M., Othman, M. H. A., Abdelrahman, S. A. M., Thao, L. P., Gabl, A. E. A. E., Elrashedy, S. A., Lee, P. N., Hirayama, K., Salem, H., & Huy, N. T. (2019). Methodological steps used by authors of systematic reviews and meta-analyses of clinical trials: A cross-sectional study. *BMC Medical Research Methodology*, *19*(1). https://doi.org/10.1186/s12874-019-0780-2

Golledge, R. G., Klatzky, R. L., & Loomis, J. M. (1996). Cognitive Mapping and Wayfinding by Adults Without Vision. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 215–246). Springer Netherlands. https://doi.org/10.1007/978-0-585-33485-1_10

Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall.

Griffin-Shirley, N., Banda, D., M. Ajuwon, P., Cheon, J., Lee, J., Ran Park, H., & N. Lyngdoh, S. (2017). *A Survey on the Use of Mobile Applications for People who Are Visually Impaired*. *111*. https://doi.org/10.1177/0145482X1711100402

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*(1). https://doi.org/10.1186/s13750-018-0121-7

Accepted Manuscript

Haddaway, N. R., & Pullin, A. S. (2014). The Policy Role of Systematic Reviews: Past, Present and Future. *Springer Science Reviews*, *2*(1–2), 179–183. https://doi.org/10.1007/s40362-014-0023-1

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., & Murphy, K. (2017). Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3296–3297. https://doi.org/10.1109/CVPR.2017.351

Islam, M. M., Sadi, M. S., Zamli, K. Z., & Ahmed, M. M. (2019). Developing Walking Assistants for Visually Impaired People: A Review. *IEEE Sensors Journal*, 1–1. https://doi.org/10.1109/JSEN.2018.2890423

Ivanchenko, V., Coughlan, J., & Shen, H. (2008). Crosswatch: A Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections. In K. Miesenberger, J. Klaus, W. Zagler, & A. Karshmer (Eds.), *Computers Helping People with Special Needs* (Vol. 5105, pp. 1122–1128). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70540-6_168

Jafri, R., Ali, S. A., Arabnia, H. R., & Fatima, S. (2014). Computer vision-based object recognition for the visually impaired in an indoors environment: A survey. *The Visual Computer*, *30*(11), 1197–1222. https://doi.org/10.1007/s00371-013-0886-1

Kidholm, K., Ekeland, A. G., Jensen, L. K., Rasmussen, J., Pedersen, C. D., Bowes, A., Flottorp, S. A., & Bech, M. (2012). A Model for Assessment of Telemedicine Applications: MAST. *International Journal of Technology Assessment in Health Care*, *28*(01), 44–51. https://doi.org/10.1017/S0266462311000638

Accepted Manuscript

Ko, E. (2013). Blind Guidance System Using Situation Information and Activity-based Instruction. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 80:1–80:2. https://doi.org/10.1145/2513383.2517037

Ko, E., & Kim, E. Y. (2017). A Vision-Based Wayfinding System for Visually Impaired People Using Situation Awareness and Activity-Based Instructions. *Sensors (Basel, Switzerland)*, *17*(8). https://doi.org/10.3390/s17081882

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Z., Rahman, M., Robucci, R., & Banerjee, N. (2017). PreSight: Enabling Real-Time Detection of Accessibility Problems on Sidewalks. *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 1–9. https://doi.org/10.1109/SAHCN.2017.7964930

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (Vol. 8693, pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, C., Mao, J., Sha, F., & Yuille, A. (2017). Attention Correctness in Neural Image Captioning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence,* 4176–4182. http://arxiv.org/abs/1605.09553

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.MR000033.pub3

Accepted Manuscript

Maclean, K., & Enriquez, M. (2003). Perceptual design of haptic icons. *In Proceedings of Eurohaptics*, 351–363.

Mahida, P. T., Shahrestani, S., & Cheung, H. (2017). Localization techniques in indoor navigation system for visually impaired people. *2017 17th International Symposium on Communications and Information Technologies (ISCIT)*, 1–6. https://doi.org/10.1109/ISCIT.2017.8261229

Manduchi, R. (2012). Mobile Vision as Assistive Technology for the Blind: An Experimental Study. In K. Miesenberger, A. Karshmer, P. Penaz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (Vol. 7383, pp. 9–16). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_2

Manduchi, R., Kurniawan, S., & Bagherinia, H. (2010). Blind guidance using mobile computer vision: A usability study. *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '10*, 241. https://doi.org/10.1145/1878803.1878851

Meier, A., Matthies, D. J. C., Urban, B., & Wettach, R. (2015). Exploring vibrotactile feedback on the body and foot for the purpose of pedestrian navigation. *Proceedings of the 2nd International Workshop on Sensor-Based Activity Recognition and Interaction - WOAR '15*, 1–11. https://doi.org/10.1145/2790044.2790051

Mocanu, B., Tapu, R., & Zaharia, T. (2016). When Ultrasonic Sensors and Computer Vision Join Forces for Efficient Obstacle Detection and Recognition. *Sensors (Basel, Switzerland)*, *16*(11). https://doi.org/10.3390/s16111807

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1). https://doi.org/10.1186/2046-4053-4-1

Accepted Manuscript

Neat, L., Peng, R., Qin, S., & Manduchi, R. (2019). Scene text access: A comparison of mobile OCR modalities for blind users. *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*, 197–207. https://doi.org/10.1145/3301275.3302271

OhnBar, E., Kitani, K., & Asakawa, C. (2018). Personalized dynamics models for adaptive assistive navigation systems. *Conference on Robot Learning*, 16–39.

Page, M. J., McKenzie, J. E., Kirkham, J., Dwan, K., Kramer, S., Green, S., & Forbes, A. (2014). Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.MR000035.pub2

Peiris, H., Kulasekara, C., Wijesinghe, H., Kothalawala, B., Walgampaya, N., & Kasthurirathna, D. (2016). EyeVista: An assistive wearable device for visually impaired sprint athletes. *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, 1–6. https://doi.org/10.1109/ICIAFS.2016.7946558

Pinard, C., Chevalley, L., Manzanera, A., & Filliat, D. (2017). End-to-end depth from motion with stabilized monocular videos. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *4*.

Plikynas, D., Žvironas, A., Budrionis, A., & Gudauskis, M. (2020). Indoor Navigation Systems for Visually Impaired Persons: Mapping the Features of Existing Technologies to User Needs. *Sensors*, *20*(3), 636. https://doi.org/10.3390/s20030636

Poggi, M., & Mattoccia, S. (2016). A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning. *2016 IEEE Symposium on Computers and Communication (ISCC)*, 208–213. https://doi.org/10.1109/ISCC.2016.7543741

Pollock, A., Campbell, P., Struthers, C., Synnot, A., Nunn, J., Hill, S., Goodare, H., Watts, C., & Morley, R. (2017). Stakeholder involvement in systematic reviews: A protocol for a systematic review of methods, outcomes and effects. *Research Involvement and Engagement*, *3*(1). https://doi.org/10.1186/s40900-017-0060-4

Accepted Manuscript

Real, S., & Araujo, A. (2019). Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems. *Sensors*, *19*(15), 3404. https://doi.org/10.3390/s19153404

Rituerto, A., Fusco, G., & Coughlan, J. M. (2016). Towards a Sign-Based Indoor Navigation System for People with Visual Impairments. *ASSETS. ACM Conference on Assistive Technologies*, *2016*, 287–288. https://doi.org/10.1145/2982142.2982202

Roberts, I., Ker, K., Edwards, P., Beecher, D., Manno, D., & Sydenham, E. (2015). The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ*, *350*(jun02 17), h2463–h2463. https://doi.org/10.1136/bmj.h2463

Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(1), 105–119. https://doi.org/10.1109/TPAMI.2008.275

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Schauerte, B., Martinez, M., Constantinescu, A., & Stiefelhagen, R. (2012). An Assistive Vision System for the Blind That Helps Find Lost Things. In K. Miesenberger, A. Karshmer, P. Penaz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (Vol. 7383, pp. 566–572). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_83

Scheggi, S., Talarico, A., & Prattichizzo, D. (2014). A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. *22nd Mediterranean Conference on Control and Automation*, 20–23. https://doi.org/10.1109/MED.2014.6961320

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Accepted Manuscript

Sharma, T., Apoorva, J. H. M., Lakshmanan, R., Gogia, P., & Kondapaka, M. (2016). NAVI: Navigation aid for the visually impaired. *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 971–976. https://doi.org/10.1109/CCAA.2016.7813856

Shen, H., & Coughlan, J. M. (2012). Towards a Real-Time System for Finding and Reading Signs for Visually Impaired Users. In K. Miesenberger, A. Karshmer, P. Penaz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (Vol. 7383, pp. 41–47). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_7

Silva, V., Grande, A. J., Carvalho, A. P. V. de, Martimbianco, A. L. C., & Riera, R. (2014). Overview of systematic reviews—A new type of study. Part II. *Sao Paulo Medical Journal*, *133*(3), 206–217. https://doi.org/10.1590/1516-3180.2013.8150015

Sorgini, F., Caliò, R., Carrozza, M. C., & Oddo, C. M. (2018). Haptic-assistive technologies for audition and vision sensory disabilities. *Disability and Rehabilitation: Assistive Technology*, *13*(4), 394–421. https://doi.org/10.1080/17483107.2017.1385100

Szegedy, C., Ioffe, S., & Vanhoucke, V. (2016). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*.

Talmon, J., Ammenwerth, E., Brender, J., de Keizer, N., Nykänen, P., & Rigby, M. (2009). STARE-HI--Statement on reporting of evaluation studies in Health Informatics. *International Journal of Medical Informatics*, *78*(1), 1–9. https://doi.org/10.1016/j.ijmedinf.2008.09.002

Tapu, R., Mocanu, B., Bursuc, A., & Zaharia, T. (2013). A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People. *2013 IEEE International Conference on Computer Vision Workshops*, 444–451. https://doi.org/10.1109/ICCVW.2013.65

Tapu, R., Mocanu, B., & Zaharia, T. (2017). DEEP-SEE: Joint Object Detection, Tracking and Recognition with Application to Visually Impaired Navigational Assistance. *Sensors (Basel, Switzerland)*, *Deep Learning for Computer Vision: A Brief Review*(11). https://doi.org/10.3390/s17112473

Accepted Manuscript

Tapu, R., Mocanu, B., & Zaharia, T. (2018). Wearable assistive devices for visually impaired: A state of the art survey. *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2018.10.031

Tapu, R., Mocanu, B., & Zaharia, T. (2013). A computer vision system that ensure the autonomous navigation of blind people. *2013 E-Health and Bioengineering Conference (EHB)*, 1–4. https://doi.org/10.1109/EHB.2013.6707267

The Campbell Collaboration. (2014). *Campbell Collaboration Systematic Reviews: Policies and Guidelines*. The Campbell Collaboration. https://doi.org/10.4073/cpg.2016.1

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Wang, H.-C., Katzschmann, R. K., Teng, S., Araki, B., Giarre, L., & Rus, D. (2017). Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 6533–6540. https://doi.org/10.1109/ICRA.2017.7989772

Wang, R. J., Li, X., & Ling, C. X. (2018). Pelee: A Real-Time Object Detection System on Mobile Devices. *ArXiv:1804.06882 [Cs]*. http://arxiv.org/abs/1804.06882

Zhang, L., Zhou, T., & Lian, B. (2018). Integrated IMU with Faster R-CNN Aided Visual Measurements from IP Cameras for Indoor Positioning. *Sensors*, *18*(9), 3134. https://doi.org/10.3390/s18093134

Zheng, Z., & Weng, J. (2016). Mobile Device Based Outdoor Navigation with On-Line Learning Neural Network: A Comparison with Convolutional Neural Network. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 11–18. https://doi.org/10.1109/CVPRW.2016.9

Accepted Manuscript

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8697–8710. https://doi.org/10.1109/CVPR.2018.00907

Accepted Manuscript

Table 1. Summary of the findings

Publication ID	Purpose/functionality	Step in wayfinding process (Golledge et al., 1996)	Input/output	Data processing	Evaluation
(Tapu, Mocanu, Bursuc, et al., 2013)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video/?	Local	Video/image - based
(Ko & Kim, 2017)	Indoor navigation	Representing current location and progressing on a target route	Video + motion sensors/audio	Local	Video/image- and user-based
(Rituerto et al., 2016)	Indoor navigation	Representing current location and progressing on a target route	Video + environment map + motion sensors/audio	Remote	User-based
(Ko, 2013)	Indoor navigation	Representing current location and progressing on a target route	Video + motion sensors/audio	Local	User-based
(Garcia & Nahapetian, 2015)	Indoor object detection, recognition, avoidance, and navigation	Representing current location and progressing on a target route	Video/audio	Local	Video/image - based
(Elloumi et al., 2013)	Indoor object detection, recognition, and avoidance	Representing current location	Video/?	Local	Proof-of-concept
(Tapu et al., 2017)	Outdoor object detection, recognition, and avoidance	Representing current location	Video/audio	Remote	Video/image-based
(Zheng & Weng, 2016)	Indoor object detection, recognition, avoidance, and navigation	Representing current location and progressing on a target route	Video/audio	Local	Video/image- and user-based
(Li et al., 2017)	Indoor object detection, recognition, avoidance, and navigation	Representing current location and progressing on a target route	Video, images, geo data/?	Local/ remote	Video/image - based
(Bai et al., 2017)	Indoor/outdoor object detection, recognition, avoidance, and navigation	Representing current location and progressing on a target route	Video and audio/audio	Remote	Video/image- and user-based

Accepted Manuscript

(Mocanu et al., 2016)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video + ultrasonic sensors/audio	Local	Video/image and user-based
(Tapu, Mocanu, & Zaharia, 2013)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video/?	Local	Video/image - based
(Bułat & Głowacz, 2016)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video/audio	Local	Video/image - based
(Dasila et al., 2017)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video, images/audio, binaural special sound	Remote	Proof-of-concept
(Sharma et al., 2016)	Indoor/outdoor object detection, recognition, and avoidance	Representing current location	Video/audio	Remote	Lab-based

Question marks in the table indicate than a certain criterion was not explicitly mentioned in the selected publication

Accepted Manuscript

Table 2. Semi-structured survey: demographic profile of the survey participants.

Variable	Value	Ratio of Respondents, %
Age (years)	< 20	4%
	20 - 25	20%
	25 - 30	28%
	30 - 35	12%
	35 - 40	16%
	40 - 45	4%
	45 - 50	4%
	50 - 55	0%
	55 - 60	8%
	> 60	4%
Professional work experience (years)	None	8%
	<5	44%
	5 - 10	16%
	10 - 20	20%
	>20	12%
Resident in	Big city (>100000)	88%
	Town (<100000)	12%
Education	Primary	8%
	Secondary	32%
	University/College	60%
Employment	Fully employed	52%
	Partially employed	20%
	Unemployed	28%
Region of residence	EU	88%
	USA	8%
	India	4%
Marital status	Married	24%
	Single	64%
	Divorced/Widowed	12%
Yearly Income (€)	<6000	28%
	6000 - 12000	40%
	12000 - 24000	24%
	>24000	8%

Accepted Manuscript

Table 3. A sorted list of the criteria defined by the participants in the expert survey answering the question "Please, list up to 5 biggest problems (in diminishing order) you experience when orientating/navigating outdoors?". Here n_i indicates the number of respondents who picked each criterion.

Rank	**Experts' criteria: problems navigating outdoors**	**n_i**	**$Mean_i$**	**Sum of importance scores**	**Weighting rate (W_i)**
1	Pedestrian crossings	13	3.38	44	0.2466
2	Knowing what is nearby, above, etc.	10	4.00	40	0.2038
3	Snow, ice, rain, and other bad weather conditions	10	3.60	36	0.1651
4	Traffic lights	9	3.22	29	0.1071
5	Finding the safest route (traffic, etc.)	7	3.86	27	0.0929
6	Finding objects (e.g., shops, hotels, WC, etc.)	8	3.13	25	0.0796
7	Lack of certain landmarks (such as pav ement edges, etc.)	5	4.20	21	0.0562
8	Finding bus or other stops	5	3.20	16	0.0326
9	Accurate distance/time to a selected location	4	3.00	12	0.0183
10	Unexpected obstacles on the passages, etc.	5	2.40	12	0.0183

Accepted Manuscript

Table 4. A sorted list of the criteria defined by the participants in the expert survey answering the question “Please list up to 5 biggest problems (in diminishing order) you experience when orientating/navigating indoors (e.g. public places, at home, etc.)”. Here n_i indicates the number of respondents who picked each criterion.

Rank	**Experts' criteria: problems navigating indoors**	**n_i**	**$Mean_i$**	**Sum of importance scores**	**Weighting rate (W_i)**
1	Finding room by number	15	4.07	61	0.642
2	Finding elevator	6	3.67	22	0.083
3	Reading numbers in the bank	5	3.60	18	0.056
4	Finding stairs	6	3.00	18	0.056
5	Recognising objects	3	4.67	14	0.034
6	Finding entrances	5	2.60	13	0.029
7	Detecting obstacles	4	3.25	13	0.029
8	Finding exits	6	1.83	11	0.021
9	Identifying steps and trip hazards	2	5.00	10	0.017
10	Navigating in large open spaces	2	4.50	9	0.014

Accepted Manuscript

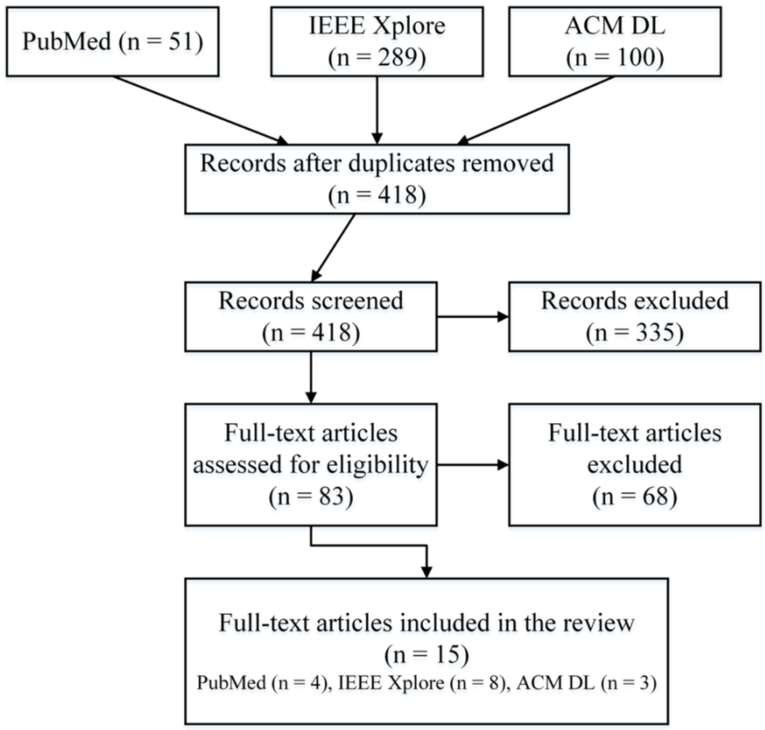

Figure 1. PRISMA flow diagram.

Accepted Manuscript

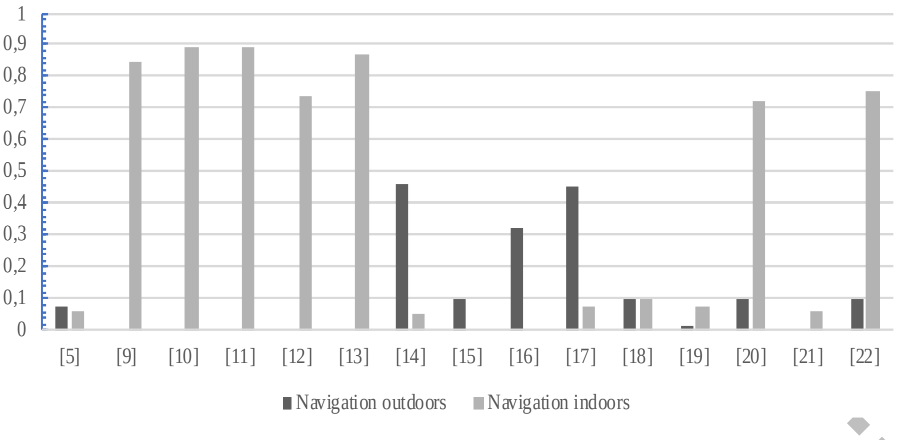

Figure 2. Evaluation of how well articles and associated prototypes address expert-defined criteria for outdoor and indoor ETA solutions (black and grey columns, respectively).

Accepted Manuscript

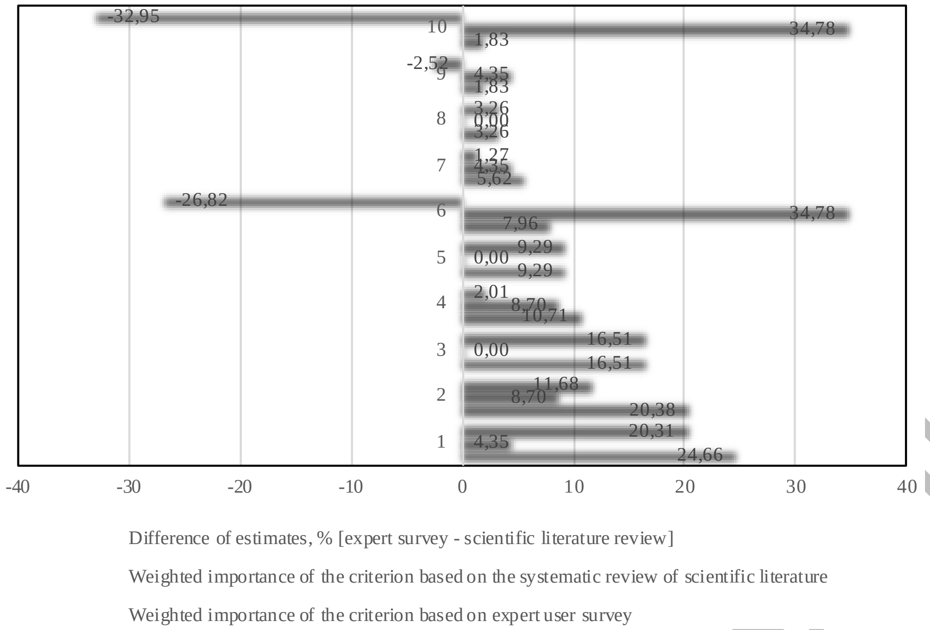

Figure 3. The importance (%) of the outdoor criteria (vertical numbers 1–10 indicate criteria) as estimated by the experts (blue columns) and researchers (orange columns). Differences (%) between the two estimates are represented by grey columns. Numbers along the columns indicate the importance of the corresponding estimates in %.

Accepted Manuscript

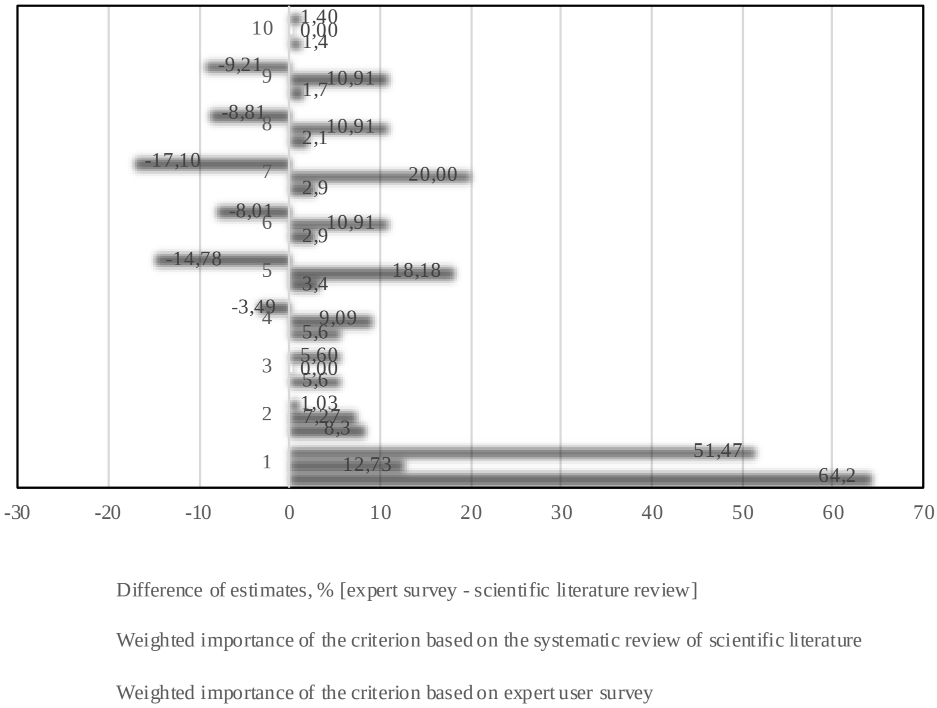

Figure 4. The importance (%) of the indoor criteria (vertical numbers 1–10 indicate criteria) as estimated by the experts (blue columns) and researchers (orange columns). The differences (%) between the two estimates are represented by the grey columns. Numbers along the columns indicate the importance of the corresponding estimates in %.

Accepted Manuscript