

Vision-Based Assistive Systems for Visually Impaired People: A Review



Sandeep Mandia, Ashok Kumar, Karan Verma,
and Jitendra Kumar Deegwal

Abstract This research reviews the current state of vision-based assistive solutions for the visually impaired (VI). The paper focuses primarily on camera-based assistive system solutions. We focused the review on vision-based assistive solutions proposed for VI people. The sensors, image processing algorithms, and wireless communication protocols employed in the survey have been summarised. Acoustic output devices were used in addition to cameras, Radio Frequency Identification (RFID), and Global Positioning System (GPS). Vision-based assistive solutions have evolved from traditional image processing techniques to machine learning to deep learning for assistance for VI users. Wi-Fi and Bluetooth devices are the most common wireless technologies used by vision-based assistive systems. The literature does not adequately leverage the optimization of deep learning models for edge devices.

Keywords Visually impaired people · Vision-based assistive systems · Wireless technologies

1 Introduction

A significant population in the world is visually impaired. As per World Health Organization (WHO) world report on vision published in 2019, at least 2.2 billion people are visually impaired [1]. It becomes challenging for the person with a vision disability to do their chores. The vision-impaired condition makes the person's life

S. Mandia (✉) · A. Kumar
Department of Electronics and Communication Engineering, Government Mahila Engineering College, Ajmer 305002, India
e-mail: smandia20@gmail.com

K. Verma
Department of Computer Science and Engineering, National Institute of Technology, Delhi 110040, India

J. K. Deegwal
Department of Electronics Instrumentation and Control Engineering, Government Engineering College, Ajmer 305025, India

dependent on a caregiver. That is very expensive and difficult in this fast-moving world.

According to a study, visually impaired people face falls, traffic-related injuries, and occupational injuries [2]. People with reduced visual acuity are 1.7 times more prone to fall and 1.9 times more prone to multiple falls than those with full sight. A hip fracture is between 1.3 to 1.9 times more likely for persons with visual impairment than an average person. As per another study [3], 15% of people with vision disability collide obstacles every month on average, and 40% of people with vision disability fall every year because they hit obstacles. In particular, aerial obstacles, such as awnings, tree branches, and similar objects typically have no projection on the ground or floor [4]. Visually impaired people are generally facing two types of dangers. One is a collision with aerial obstacles in the front and the second fall. Addressing these problems can prevent mishappening.

Traditionally white stick and guide dogs are used to provide guidance when visually impaired people go out independently. The aerial obstacle cannot be localized using the white stick or guide dog. The solution to these problems is to have an assistance system that gives information regarding aerial or ground obstacle well in advance to the visually impaired person so that s/he can protect self from the obstacle. There is much scope in the improvement of the assistance system for visually impaired people. There have been various smart assistant systems proposed by researchers from different parts of the globe that address ground obstacle avoidance. There are solutions in which addresses both aerial and ground obstacle avoidances have.

A general approach towards vision-based assistive systems includes processing of camera input using image processing-based algorithms. The processed output and other outputs from other sensors are used for decision-making; based on this, the VI user receives feedback. Figure 1 shows a generalized approach to the vision-based assistive system for VI users. The processing includes feature extraction from the frames capture by the camera. The decision making utilizes a basic thresholding technique to a sophisticated machine learning or deep learning-based approach. The VI user receives feedback based on the decision made through various means.

As the camera mimics the task performed by the human eye, vision-based assistive systems are most suitable for assistance to VI people. The advancement in the

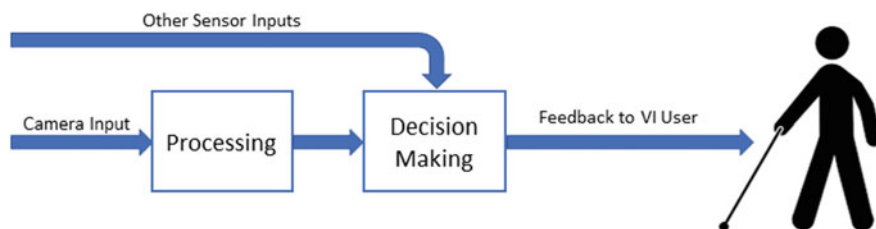


Fig. 1 A general approach to the vision-based assistive system

development of algorithms and extensive use of deep learning in computer vision further makes it a promising candidate for the solution.

The remainder of the paper is organized as follows: Sect. 2 discusses the literature review of the existing vision-based solutions for visually impaired people based on sensors, processing techniques, and wireless communication techniques. Section 3 concludes this paper with future directions for the said problem.

2 Investigation of AI-Based Vision Assistive System for VI People

A mobile camera-based solution for visually impaired people is reported [5] for indoor Fig. 2(a). The pre-defined paths were marked with colour tapes, and the mobile camera was used to track the path. Extended Kalman Filter (EKF) and Weighted Moving Average Filter (WMA) are used to overcome optical flow errors. Arianna is a framework for determining a safe walking path in interior environments. The solution is based on a video camera incorporated inside a smartphone device at the hardware level. The user feedback is positive. Vibration patterns are used to transfer information. A series of interest locations, denoted by arrows, are used to design the walking path. QR codes can be scanned, or a path on the floor can be followed.

It [6] introduced a new marker-based technique called mobile vision (MV). The technology is incorporated on a smartphone device in an indoor context and uses special colour markers Fig. 2(b). The user is directed via red, green, and blue colour markers to locate interesting sites, such as restrooms, elevators, or exits. Feedback messages are delivered via text-to-speech transcripts.

The Smart Vision navigation framework is presented in [7], which combines GPS, Wi-Fi localization with GIS (Geographic Information System) [8], passive RFID tags, and computer vision algorithms for outdoor scenarios. The system is not intended to replace the white cane but rather to supplement it by alerting the visually impaired (VI) user to impending dangers. Here, a database with prospective objects of interest (e.g., elevator, welcome desk, plants, cash machine, and telephone booth) is being created Fig. 2(c). The reference images stored a priori are sought among video frames captured by the camera during the test. The approach, however, is extremely sensitive to camera movement and is strongly reliant on the amount of the training sample. Furthermore, it experiences adaptability issues, since for a bigger dataset with various objects of interest, the computational time increments fundamentally.

A deterrent location and order strategy totally coordinated on a standard cell phone is presented in [9] that is more, reached out in [10] Fig. 2(e). The framework is intended to work with the VI client route in both indoor and open-air conditions. In [9], creators propose distinguishing the block's area by extricating interest focuses that are followed between progressive casings utilizing the standard Lucas-Kanade

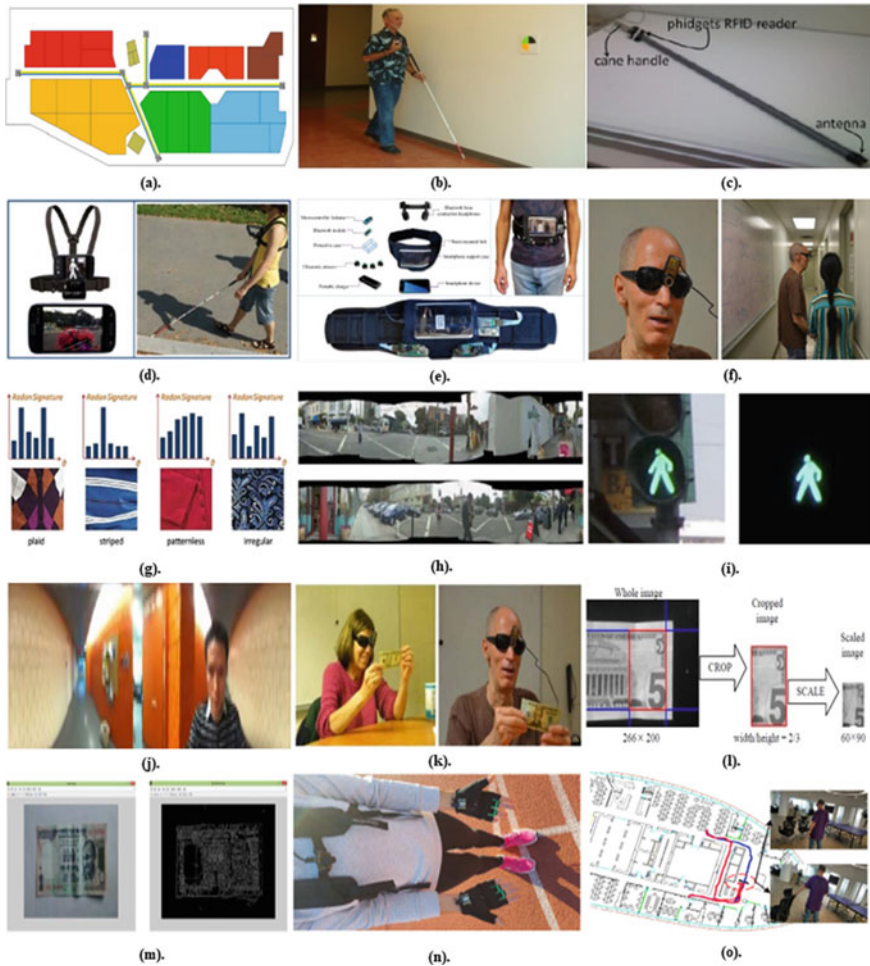


Fig. 2 Vision-based assistive systems reported for VI users [5–21]

algorithm Fig. 2(d). The object’s movement is recognized from the camera development with the assistance of many homographic changes grouped by applying the Random sample consensus (RANSAC) calculation [11]. The identified objects are additionally arranged by joining the Histogram of Oriented Gradients (HOG) descriptor into a Bag of Visual Words (BOW) portrayal. Even though the framework generally returns great outcomes, it can’t identify huge, level designs or effectively gauge the distance between the VI client and a check. In [10], the creators proposed addressing the previously mentioned constraints by coordinating ultrasonic sensors inside the framework. The methodology shows promising outcomes; however, it demonstrates to be delicate when various moving obstructions are available in the scene.

In [12], a computer vision-based way-finding technology is suggested that helps independent access to indoor but unfamiliar locations Fig. 2(f). The system consists of a camera, microphone, computer, and Bluetooth earpieces on the hardware level. The framework uses a geometric design mixed with a corner and edge recognition method to detect doors, elevators, and cabinets. The system may then discriminate between foreground and background objects using an optical character recognition approach. A Canny edge detector is utilized for the detection of doors and Optical Character Recognition for text classification.

A system proposed in [13] includes a module for identifying textiles Fig. 2(g). Four clothing textures can be distinguished using a new Radon Signature descriptor: Plaid, striped, patternless, and irregular and eleven clothing colours. Although both modules were created with individuals with disabilities in mind, no studies or tests with actual VI users have been conducted too far. Furthermore, the framework is incapable of handling object occlusion or operating in real-time.

Developments in the Crosswatch system for providing guidance to visually impaired travellers at traffic intersection is discussed, and also new functionalities are described [14] Fig. 2(h). The panoramic image processing was used for the analysis of the crossroad view. The VI user captured the panoramic image of the viewpoint by rotating the camera for 360°. Another traffic light recognizer is proposed to detect traffic light signals for VI users [15] Fig. 2(i). The Active Optical Unit (AOU) is extracted from the image captured, and based on the AOU, the distance between VI users and the traffic light is calculated.

ShopMobile II has been proposed for supermarket grocery shopping for VI users [16]. The navigation is based on the barcode scanner on the products in the supermarket. The barcode localization and decoding are done by using computer vision algorithms. The localization of the barcode is based on the number of zero to one and one to zero transitions on two horizontal lines on the image.

Molina et al. proposed the use of visual nouns for VI user navigation in both indoor and outdoor situations in [17]. The system generates mosaic images, which are then used to help the VI navigate around streets and corridors. Signage, visual text, and visual icons are considered visual nouns. However, a number of open conditions must be met for the system to be beneficial to VI people: (1) development of an appropriate human-machine interface; (2) integration into a wearable assistive device; and (3) development of an acoustic or haptic interface.

Another system for VI people is proposed utilizing a smartphone camera to capture panoramic images and a Graphic Processing Unit (GPU) server to extract features from an image or a short video [18] Fig. 2(j). The modelling of images was done by converting the image into the HSI colour model and then taking the projection of H, S, I and gradients to calculate the Omni-projection. The Fast Fourier Transform (FFT) of the normalized projection curves was taken. In the query stage, the frame is processed the same and compared with all the modelled images. The closest matching frame is obtained using phase curves of the Omni-directional images. The use of multi-core CPUs or GPUs is proposed for enhancing computational speed.

A robust banknote recognition system is proposed based on computer vision for blind people [19] Fig. 2(k). The banknote dataset was collected in various circumstances and labelled with note values. The Speeded Up Robust Features (SURF) have been utilized for matching for banknotes. The authors claim to have achieved 100% true recognition and 0% false recognition rate. Similarly, another smartphone-based USA currency note recognition system was proposed [20] Fig. 2(l). The system utilizes the Principal Component Analysis (PCA) based image recognition method, Eigenfaces, to recognize currency notes. The authors have achieved a 99.8% of accuracy rate and 7 frames per second processing speed. The processing was done on a Grayscale image that was converted from an RGB image [20]. Another mobile application based Indian currency note recognition system is proposed [21] Fig. 2(m). A median filter and histogram equalization are utilized for noise removal and image enhancement. Morphological operations are performed for feature extraction, and these features are used for currency note matching and recognition.

A vision-based system is proposed for VI users during walking, jogging, and running [22] Fig. 2(n). The system utilizes image processing for lines and lane detection on the road in the outside environment. The system uses a camera and haptic gloves for feedback. The haptic gloves were fitted with vibration motors, and the commands to the VI user were encoded in the form of sequences of vibration in the haptic gloves. The extraction is done by using probabilistic Hough Line Transform.

The Charge Coupled Device (CCD) camera-based assistance gadgets have been more convenient and comfortable to manage than sensory systems. However, when estimating the real distance between the VI user and a detected obstacle, these solutions have low accuracy. Any monocular system has the drawback of being unable to determine the global object scale only on a frame. The concern is exacerbated when dealing with outdoor environments since scale drifts between map sections and their projected motion vectors are more common [23].

The E-vision system is proposed for VI users for three distinct daily activities: Supermarket visit, public administration Building visit, and outdoor walk [24] Fig. 2(o). The system exploits the classification and Optical Character Recognition (OCR) for supermarket visits, OCR and object detection & face and emotion recognition for administrative building visits, and face recognition & text-to-speech conversion techniques for outdoor environments.

A Convolutional Neural Network (CNN) based wearable travel system for VI users for indoor and outdoor environments has been proposed [25]. The system is capable of providing environment perception and navigation for VI users. The system utilizes an Inertial Measurement Unit (IMU) for the acquisition of the altitude angle of the camera. A smartphone is utilized for position acquisition, navigation, object detection, and acoustic feedback to the VI user. A lightweight CNN based PeleeNet [26] object detection model trained on the MS COCO dataset has been used in the system. Another similar deep learning based wearable assistive system for VI users to enhance the environment perception has been proposed [27]. The CNN-based segmentation and obstacle avoidance system is proposed that utilizes CPU and GPU computation power for real-time performance. The smartphone is utilized for touch

Table 1 Summary of vision-based assistive systems for VI users

Ref. No.	Sensors	Image processing techniques	Wireless communication	Remarks
[5]	Mobile camera	Optical Flow with EKF and WMA	–	Pre-defined area for movement
[6]	Mobile Camera, Acoustic signal	Distance-based one-class-classifier	–	Manual installation of markers
[7]	Camera, RFID, GPS	SURF (Speeded Up Robust Features)	Wi-Fi	It covers both indoor and outdoor
[9]	Camera	Lucas-Kanade Algo, Bag of Visual Words (BoVW)	–	It covers both indoor and outdoor
[10]	Camera, Ultrasonic sensors, Microphones	Support Vector Machine with Aggregated Descriptors	Bluetooth	Use of two sensors for detection
[12]	Camera, Microphone	Canny Edge Detector, OCR	Bluetooth	It covers only indoor environments with few detections
[13]	Camera, Microphone	Wavelet sub-band, Random Signature Descriptor	Bluetooth	For clothing pattern detection, not for movement assistance
[14]	Camera, Microphone, GPS, GIS	Panoramic image processing	–	Covers only outdoor environment
[15]	Camera, Accelerometer, Gyroscope	AOU detection	–	Specific to traffic light application in the outdoor environment
[16]	Mobile camera	Computer vision algorithm	–	Shopping assistance
[17]	Camera	Computer vision algorithm	–	It covers both indoor and outdoor
[18]	Smartphone camera, omnidirectional lens (GoPano)	HSI model, Omni-projection curves, FFT	–	Addresses only indoor environment
[19]	Camera	SURF	–	Confined to banknote detection
[20]	Smartphone camera, Acoustic output	PCA based eigenfaces algo	–	Confined to banknote detection

(continued)

Table 1 (continued)

Ref. No.	Sensors	Image processing techniques	Wireless communication	Remarks
[21]	Smartphone camera, Acoustic output	Median Filter, Morphological operation, Histogram equalization	–	Confined to banknote detection
[22]	Camera, Haptic Glove	Thresholding, Probabilistic Hough Line Transform	Bluetooth Low Energy	Addresses outdoor environment specific to a track
[24]	Camera, Acoustic output	OCR, Object detection, Face & emotion recognition	Wi-Fi	Addresses only indoor environment
[25]	Camera, IMU, Smartphone	PeleeNet, text to Speech, Speech Recogniser	–	Covers both indoor and outdoor with deep learning-based object detection model
[27]	Camera, Acoustic output, smartphone	FuseNet	Bluetooth	Covers both indoor and outdoor with deep learning-based segmentation model

interface to provide environmental information to the VI user. A CNN based FuseNet [28] has been utilized for the segmentation of captures image frames.

Table 1 summarises the literature based on sensors used, image processing techniques used for decision making, and wireless communication techniques used for feedback to the VI user.

3 Conclusion

In this research, a literature review of computer vision-based solutions for visually challenged people is presented. Table 1 shows the survey summary, categorizing the studies based on the sensors used, image processing algorithms used, and communication techniques used by the individual study. This study suggests that standard digital image processing techniques were used in the early days of computer vision-based assistive solutions for VI users, and machine learning and deep learning techniques recently. Wi-Fi and Bluetooth have been used in the majority of studies where wireless communication techniques have been used. Many assistive systems have used a simple camera, and others have used RFID, GPS, GSM, Ultrasonic sensor, sound output, and other technologies along with the camera. Researchers have begun to apply deep learning approaches for assistive solutions for VI users as machine learning, and deep learning techniques have grown with the arrival of good computation powers to machines. However, carrying computational power devices

for vision-based assistive solutions is inconvenient for VI users. The deep learning models may be optimized for edge inference using current optimization techniques. Quantization and layer pruning are part of the deep learning model optimization process. With the optimization of deep learning models for inference on edge devices, vision-based assistive solutions can be upgraded.

References

1. WHO W (2019) World report on vision. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO. World Health Organization
2. Legood R (2002) Are we blind to injuries in the visually impaired? A review of the literature. *Inj Prev* 8:155–160. <https://doi.org/10.1136/ip.8.2.155>
3. Manduchi R, Kurniawan S Watch Your Head, Mind Your Step: Mobility-Related Accidents Experienced by People with Visual Impairment, vol 11
4. Chang W-J, Chen L-B, Chen M-C et al (2020) Design and implementation of an intelligent assistive system for visually impaired people for aerial obstacle avoidance and fall detection. *IEEE Sens J* 20:10199–10210. <https://doi.org/10.1109/JSEN.2020.2990609>
5. Croce D, Giarre L, La Rosa FG, et al (2016) Enhancing tracking performance in a smartphone-based navigation system for visually impaired people. In: 2016 24th mediterranean conference on control and automation (MED), pp 1355–1360. IEEE, Athens
6. Manduchi R (2012) Mobile vision as assistive technology for the blind: an experimental study. In: Miesenberger K, Karshmer A, Penaz P, Zagler W (eds) *Computers Helping People with Special Needs. ICCHP 2012. LNCS*, vol 7383, pp 9–16. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_2
7. Du B, Barroso J (2011) The SmartVision navigation prototype for blind users
8. Kammoun S, Macé MJM, Oriola B, Jouffrais C (2012) Towards a geographic information system facilitating navigation of visually impaired users. In: Miesenberger K, Karshmer A, Penaz P, Zagler W (eds) *Computers Helping People with Special Needs. ICCHP 2012. LNCS*, vol 7383, pp 521–528. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_77
9. Tapu R, Mocanu B, Bursuc A, Zaharia T (2013) A smartphone-based obstacle detection and classification system for assisting visually impaired people. In: 2013 IEEE international conference on computer vision workshops, pp 444–451. IEEE, Sydney, Australia
10. Mocanu B, Tapu R, Zaharia T (2016) When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition. *Sensors* 16:1807. <https://doi.org/10.3390/s16111807>
11. Lee J, Kim G (2007) Robust estimation of camera homography using fuzzy RANSAC. In: Gervasi O, Gavrilova ML (eds) *Computational Science and Its Applications – ICCSA 2007. ICCSA 2007. LNCS*, vol 4705, pp 992–1002. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-74472-6_81
12. Tian Y, Yang X, Yi C, Arditi A (2013) Toward a computer vision-based way-finding aid for blind persons to access unfamiliar indoor environments. *Mach Vis Appl* 24:521–535. <https://doi.org/10.1007/s00138-012-0431-7>
13. Yang X, Yuan S, Tian Y (2014) Assistive clothing pattern recognition for visually impaired people. *IEEE Trans Hum Mach Syst* 44:234–243. <https://doi.org/10.1109/THMS.2014.2302814>
14. Coughlan JM, Shen H (2013) CrossWatch: a system for providing guidance to visually impaired travelers at traffic intersection. *J Assist Technol* 7:131–142. <https://doi.org/10.1108/17549451311328808>

15. Mascetti S, Ahmetovic D, Gerino A, Bernareggi C, Busso M, Rizzi A (2016) Supporting pedestrians with visual impairment during road crossing: a mobile application for traffic lights detection. In: Miesenberger K, Bühler C, Penaz P (eds) *Computers Helping People with Special Needs*. ICCHP 2016. LNCS, vol 9759, pp 198–201. Springer, Cham. https://doi.org/10.1007/978-3-319-41267-2_27
16. Kulyukin VA, Kutiyawala A (2010) Demo: shopMobile II: eyes-free supermarket grocery shopping for visually impaired mobile phone users. In: 2010 IEEE computer society conference on computer vision and pattern recognition - workshops, pp 31–32. IEEE, San Francisco
17. Molina E, Zhu Z, Tian Y (2012) Visual nouns for indoor/outdoor navigation. In: Miesenberger K, Karshmer A, Penaz P, Zagler W (eds) *Computers Helping People with Special Needs*. ICCHP 2012. LNCS, vol 7383, pp 33–40. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-31534-3_6
18. Hu F, Zhu Z, Zhang J (2015) Mobile panoramic vision for assisting the blind via indexing and localization. In: Agapito L, Bronstein M, Rother C (eds) *Computer Vision - ECCV 2014 Workshops*. ECCV 2014. LNCS, vol 8927, pp 600–614. Springer, Cham. https://doi.org/10.1007/978-3-319-16199-0_42
19. Hasanuzzaman FM, Yang X, Tian Y (2012) Robust and effective component-based banknote recognition for the blind. *IEEE Trans Syst Man Cybern C* 42:1021–1030. <https://doi.org/10.1109/TSMCC.2011.2178120>
20. Grijalva F, Rodriguez JC, Larco J, Orozco L (2010) Smartphone recognition of the US banknotes' denomination, for visually impaired people. In: 2010 IEEE ANDESCON, pp 1–6. IEEE, Bogota, Colombia
21. Manikandan K, Sumithra T (2015) Currency recognition in mobile application for visually challenged
22. Mancini A, Frontoni E, Zingaretti P (2018) Mechatronic system to help visually impaired users during walking and running. *IEEE Trans Intell Transport Syst* 19:649–660. <https://doi.org/10.1109/TITS.2017.2780621>
23. Tapu R, Mocanu B, Zaharia T (2020) Wearable assistive devices for visually impaired: a state of the art survey. *Pattern Recogn Lett* 137:37–52. <https://doi.org/10.1016/j.patrec.2018.10.031>
24. Kalaganis FP, Miglotzidis P, Georgiadis K et al (2021) Lending an artificial eye: beyond evaluation of CV-based assistive systems for visually impaired people. In: Antona M, Stephanidis C (eds) *Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments*. HCHI 2021. LNCS, vol 12769, pp 385–399. Springer, Cham. https://doi.org/10.1007/978-3-030-78095-1_28
25. Bai J, Liu Z, Lin Y et al (2019) Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics* 8:697. <https://doi.org/10.3390/electronics8060697>
26. Wang RJ, Li X, Ling CX (2019) Pelee: a real-time object detection system on mobile devices. [arXiv:180406882](https://arxiv.org/abs/180406882) [cs]
27. Lin Y, Wang K, Yi W, Lian S (2019) Deep learning based wearable assistive system for visually impaired people. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW), pp 2549–2557. IEEE, Seoul, Korea (South)
28. Hazirbas C, Ma L, Domokos C, Cremers D (2017) FuseNet: incorporating depth into semantic segmentation via fusion-based CNN architecture. In: Lai SH, Lepetit V, Nishino K, Sato Y (eds) *Computer Vision – ACCV 2016*. ACCV 2016. LNCS, vol 10111, pp 213–228. Springer, Cham. https://doi.org/10.1007/978-3-319-54181-5_14