ICIC Express Letters
Volume **16**, Number **12**, December **2022**

ICIC International ©2022 ISSN 1881-803X
pp. **1351–1358**

CONTEXT AWARE NAVIGATION SYSTEM FOR ASSISTING BLIND PEOPLE

Satria Hanafi Rusli[1,*], Suryadiputra Liawatimena[1,2] and Agung Trisetyarso[3]

[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science
[2]Computer Engineering Department, Faculty of Engineering
[3]Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ suryadi; atrisetyarso }@binus.edu; s.liawatimena.id@ieee.org
*Corresponding author: satria.rusli@binus.ac.id

Received December 2021; accepted March 2022

Abstract. *In order to assist VI (Visually Impaired people or person) with their navigation task, many ATs (Assistive Technologies) utilizing CV (Computer Vision) have been developed. However, they often come with many disadvantages such as being bulky, not user friendly, high maintenance, depending on other services to function, and expensive. They often only measure the model accuracy. They rarely utilize visual signs despite their advantages. We propose an independent, low-cost, efficient, portable, and user friendly AT system embedded with deep-learning model based on YOLOv4-tiny (You Only Look Once) converted and quantized into TensorFlow Lite (TFLite) model working on Deep Simple Online and Real-time Tracking (DeepSORT) working with sign recognition to help with their navigation problem in their daily life. The performance of the proposed system is technically acceptable. However, improvements on the physical design and information output method for better user experience are needed.*
Keywords: Assistive technology, Computer vision, Sign recognition, Visual impairment

1. **Introduction.** AI (Artificial Intelligence) is a created machine intelligence to do what usually requires human intellect [1]. ML (Machine Learning) is an AI form where the machine, like humans, through an algorithm, can increase its efficiency by learning through a process of trial and error to gain experience and finish its tasks without being explicitly programmed to finish them [2]. Deep learning based on DNN (Deep Neural Network) capability to perform representation learning [3] is a method to achieve ML's goal. DNN is a multi-hidden-layer neural network (a network consisting of artificial nodes based on human neural tissue design) [4].

CNN (Convolutional Neural Network) is a specialized DNN that excels in learning data in grid topology form such as images [5] by using specialized linear operation (convolution) [6]. YOLO is a CNN-based general-purpose object detection model by Redmon et al. that combines the convolution, activation function, and pooling layers in CNN together to simplify the process which is often praised for its simplicity, reliability, accuracy, speed, and convenience to trade off between accuracy and speed without re-training the model [7]. The last version he worked on was YOLOv3, and YOLOv4 was developed by Bochkovskiy in 2020 [8].

TensorFlow is an open source library developed by Google Brain for ML and numerical computation, including CV. It is used by many large international corporations such as Nvidia, AMD, and Intel. It creates a data flow graph implemented with super-fast

DOI: 10.24507/icicel.16.12.1351

C++ and user-friendly Python programming language [9]. They also released TFLite for lightweight devices.

There are more than 2 billion VI globally [10]. Their hardest daily activities are finding literature and navigating [11]. Sign is something that represents something else [12]. There are many types of visual signs that help us navigate safely such as landmarks, traffic signs, and symbols on a map [13]. Signs are often and best an easily observable, recognizable and distinguishable symbol/object through their distinct color and forms that makes them a great visual object detection target, and simultaneously increases the model accuracy when detecting them. Obviously, a VI cannot benefit from visual signs. However, a trained CV model can help recognize the visual signs for them. There is a lack of CV-based AT that utilizes visual signs. While most of them focused on improving the detection model or method, our solution focuses on the objects (visual signs).

To ease their navigation problem and tackle previous solutions' problems, we propose an independent, low-cost, efficient, portable, and user friendly AT system using an embedded system [14] with deep-learning model based on quantized pre-trained YOLOv4-tiny model converted into TFLite model working on DeepSORT [15] tracking method while taking advantages of visual signs. The performance and user experience of the proposed system are expected to have comparable results with previous works. The results of this study are expected to emphasize the importance of user friendliness in ATs and the benefits from utilizing visual signs in CV.

This study is organized as follows. In Section 2, we describe the strengths and weaknesses from several related works. From it, we will build the proposed system's prototype using the most suitable components and detection model (Section 3). After that, we will conduct simple tests and a field test experiment (Section 4). Finally, we will evaluate our proposed system's (especially sign recognition) performance and effectiveness in helping VI's navigation task (Section 5) and draw a conclusion (Section 6).

2. **Related Works.**

Some of the previous works relied on external services (not independent like [16, 17]) which limit their utility, practicality, and effective range. It also increases the complexity and operating cost of their proposed system. Some previous works (often ones that utilize image segmentation [18, 19, 20]) utilize many input/output devices at once, making their solution impractical, uncomfortable, complex, heavy, rigid, and fragile like [21, 22, 23]. The smartphone-based solutions [24, 25] are compact, simple, and easy to replace but often too lightweight (shakes often), having low battery life, low computational power, easy to overheat, and easy to lose. Related works that utilize signs [26, 27] are often quick and accurate, but also not reliable, compact, and uncomfortable to use as wearable AT because of lack of user-based evaluation. Other works [28, 29, 30] suffer the same difficulties as mentioned above.

Less than half of the reviewed solutions conduct user-based evaluation. ATs are used often in a disabled person's daily life, and should consider the users' comfort, troubles that could impede the device's performance and dependability (ease of maintenance, troubleshooting, and replacement). The proposed system focuses on simplifying previous designs, maximizing user comfort, and utilizing the fact that signs are much faster and easier to perceive (focusing on detected object more than object detection method). We propose a simple, compact, independent, low-cost, efficient, lightweight, easy to use, easy to learn, comfortable, wearable, and user-friendly system. The proposed system will utilize the most efficient processing configuration and robust budget edge computing hardware using fastest object detection method and positioning using a simple grid method to detect objects' type, distance, and location relative to the user.

3. Proposed System.

3.1. **Detection model selection.** The proposed architecture will utilize the pre-trained YOLOv4-tiny weights that has been trained on MS COCO dataset 2017, converted to TFLite model after its extraordinary result on our model comparison test with COCO dataset on RPi4B (Raspberry Pi 4B) (See Table 1), and quantized to reduce the model size and accelerate the model inference speed by changing its floating-32bits into integer-8bits. Finally, the converted and quantized detection model will be implemented in DeepSORT tracking method.

TABLE 1. Detection model comparison test result

No.	Model name	Speed (FPS)	mAP (%)
1	**YOLOv4-tiny**	**1.5**	**79.8**
2	Faster R-CNN	0.4	69.9
3	MobileNet SSD	0.7	71.2
4	R-FCN	0.3	82.1

3.2. **Components and workflow.** Figure 1 shows the proposed system's components and workflow. The camera component captures images of objects and signs around the user to the Embedded System component to process using the method mentioned in Section 3.1. The Embedded System is powered by a normal power bank with 10000 mAh capacity. Finally, it outputs the information verbally to the user through a wireless bluetooth bone-conducting earphone by playing back a prerecorded .wav audio file of the detected object/sign's type, distance and position calculated through the object's bounding box (See Figure 2). The bone-conducting earphone ensures that the system's audio output will not

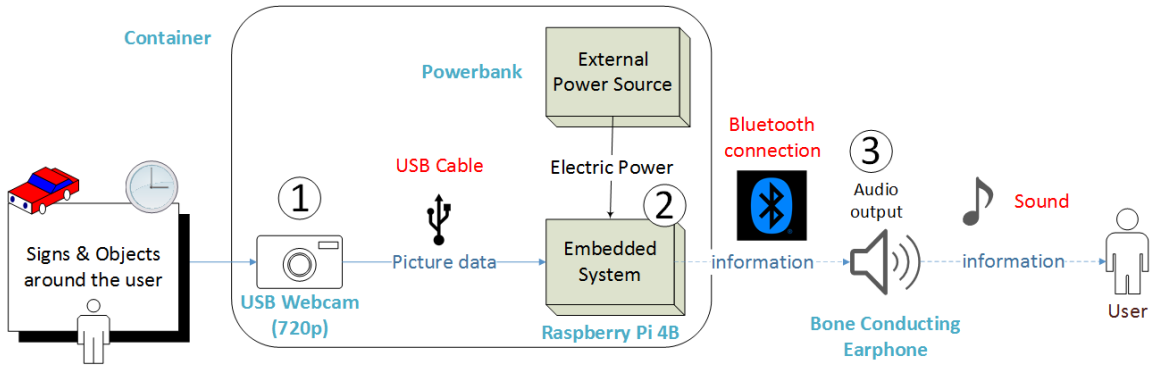

FIGURE 1. The proposed system's components and workflow

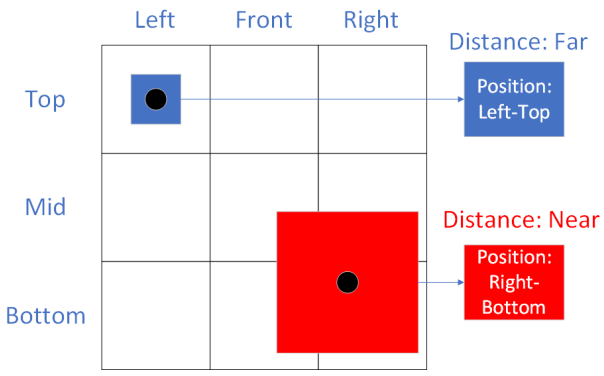

FIGURE 2. Visualization of distance and position calculation method

obstruct other sounds. The prerecorded audio playback files are in .wav format for its lossless, uncompressed, and lightweight properties (perfect for looping purposes).

4. Experiment.

4.1. **Simple measurements.** In order to obtain several quantitative measurements from the prototype, we created several tests for the system.

- Precision: Measure the model's precision using 100 images obtained from Google Image Search for car, clock, stop sign, person, and computer mouse class, and then measure its capability to detect transparent objects with 50 images of transparent wine glass and cup from Google Image Search.
- Speed: Measure the model's frame rate in FPS (Frame per Second) from log file after the field test.
- Power usage: See remaining power left on the previously fully-charged power supply component after the field test.
- Weight: Measure the weight for each component with a digital scale.
- Cost: Calculate the price for each component (excluding tax and shipping cost).

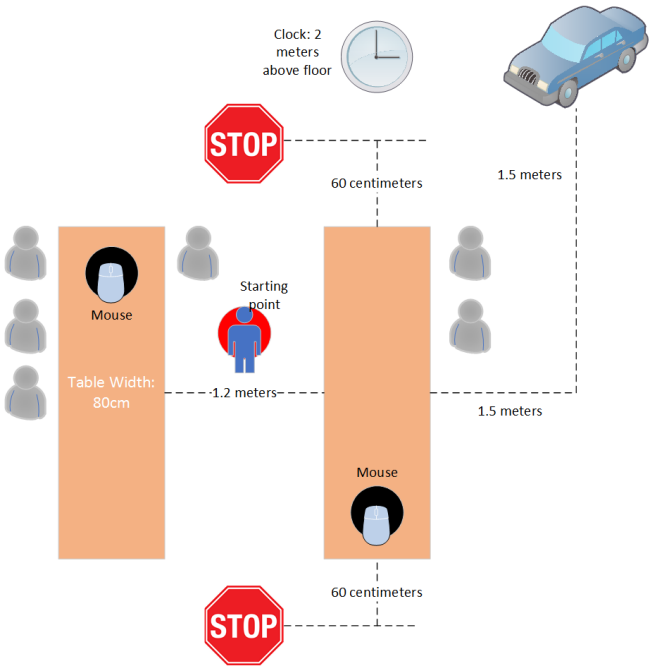

FIGURE 3. Field test map

4.2. **Field test.** The test field is an elevated, half open, mass dining hall (See Figure 3). We put stop signs to limit their movement area to prevent them from falling. Each participant is given a task to start in the red circle, retrieve an object (normal common mouse computer, black circle), and give that object to any other person while not bypassing the stop signs in 5 minutes. The participants are fully blind adult males (19-27 years old) who are blind since birth/childhood with varying body types (obese, overweight, and normal). During the field test, each participant's performance will be measured. After completing field test, each participant will be given a verbal questionnaire about the system's performance and then reply with a Likert scale from 1 (very disagree) to 6 (very agree) for the following factors:

- Positive Factors
  - Balanced Weight: Does it feel too heavy/lightweight?
  - Accurate: How do you feel about its accuracy?

  - Comfort: Is the system comfortable enough to use?
  - Low-Cost: Is the system affordable at IDR2,500,000?
  - Easy to Learn: Is the system easy to learn?
  - Easy to Use: Is the system easy to use?
- Negative Factors
  - Overwhelming: Does the amount of information conveyed overwhelm the user?
  - Annoying: Does it annoy the user and their daily activity? Is it too hot/noisy?

4.3. **Experiment result.** See Table 2 for precision test result. The proposed system also passed the transparent object detection test perfectly. The overclocked RPi4B reached 2-5 FPS (depending on the number of tracked objects). It drains around 45% of the entire external power supply capacity after around 4 hours of continuous use in field test. In Table 3, the Embedded System component seems expensive because it is bundled with heat sink case, power charger, HDMI cables, screwdriver, and 32GB SD card. The weight of the ES also includes the heat sink case.

Table 2. Precision test result

No.	Object class	Precision	MS COCO 2017 dataset object count
1	Mouse	.69	851
2	Person	.76	89202
3	Car	.79	15084
4	Clock	.84	2275
5	**Stop sign**	**.97**	**686**

Table 3. The prototype's components' weight and price

No.	Component part	Mass (gram)	Price (IDR)
1	External power supply	228	165,000.00
2	Embedded system	46	1,900,000.00
3	Camera	80	89,000.00
4	Audio output	63	265,000.00
Total		**417**	**2,419,000.00**

Table 4. The field test participants' performance

No.	System shake	Task success	Clear time (minute)	Object grab mistake	Bypassing the stop sign	See other objects
1	2	0	5	2	0	1
2	1	0	5	1	0	1
3	1	1	3	1	0	2
4	1	1	1	0	0	2
5	1	1	1	0	0	2
Total	**6**	**3**	**15**	**4**	**0**	**8**
Avg	**1.2**	**0.6**	**3**	**0.8**	**0**	**1.6**

Two participants failed the task. The third participant made a mistake and grabbed the wrong object, but then grabbed the correct one immediately after. No participant bypassed the area as they quickly identified the stop sign information from the system. Each participant detected car parked near the field test but only three last participants detected the clock on the wall (See Table 4).

All participants agree that the system is easy to learn, easy to use and does not annoy their daily activities and hearing at all. However, they agree that the system is not accurate

Table 5. Questionnaire result

No.	Comfort	Low-Cost	Easy to learn	Easy to use	Accurate	Balanced weight	Annoying	Overwhelming
1	4	2	3	3	1	2	3	5
2	1	1	5	5	2	1	3	5
3	2	2	4	5	3	2	3	5
4	6	6	6	6	1	6	1	6
5	6	6	6	6	1	6	1	1
Sum	**19**	**17**	**24**	**25**	**8**	**17**	**11**	**22**
Avg	**3.8**	**3.4**	**4.8**	**5**	**1.6**	**3.4**	**2.2**	**4.4**

enough and the audio feedback is overwhelming. They also have mixed feelings about the comfort, physical weight balance, and affordability (See Table 5).

When asked for additional comments, the first two participants commented on how their sweat makes the container bag and the strap material itchy and uncomfortable. Meanwhile, other participants asked about how easy it is to clean, troubleshoot, replace parts of, and asked about other features the proposed system might have in the future. They also asked about what possible features they can expect in the future and where the proposed system might not work on, such as underwater, in the mountain, or in darkness. All participants asked if the system could be cheaper.

5. **Discussion.** The stop sign has a higher precision score despite its low training image count. This gets better as unlike most object, obstruction of another object is not a problem for the sign recognition task as signs are usually placed in unobstructed places from views. As for the frame rate, despite its low number due to RPi4B low computational power, it is still fast enough for real-time detection.

No participant bypassed the stop sign during the field test, indicating that visual signs are easily recognizable with simple object detection and tracking methods. Device shake only occurs when the participant is bowing down to grab an object on the table. This is because of the gravity force that keeps the system to stay in an upright position as the system is hung on the participant's neck. Three participants successfully finished the given task and all participants successfully identified the parked car near the test field. However, the first two participants could not detect the clock on the wall. This could be caused by the camera angle and the system position that changes due to the first two participants' body shape and/or being shorter than the next three participants. This and the device shake problem can be solved by implementing more adjustable straps.

Participants agree that the system is easy to learn and use as all they need to do is to turn the system on and listen to the information given. They do not agree that it annoys them from doing daily activities. They do not feel annoyed by the audio output component because bone-conduction earphone does not obstruct other sound around them. The participants did not have any noise or overheat problem, showing that a good heat sink has a great role in user comfort and results in low score on Annoying factor.

They have mixed feelings about the system's physical weight balance. Some complained that it is too heavy, which may be caused by their fitness based on their BMI. It results in low score on the system's balance weight factor. They give low scores on Accurate factor and high scores on Overwhelming factor with a mixed feeling on Comfort factor and Low-Cost factor. The first three participants give low scores on every positive aspect and high scores on how overwhelming it is as the first two cannot solve such a simple task and how inaccurate it is. As a result, they do not think the system worth IDR2,500,000 as it is now. They commented that the output and the physical design needs improvement to justify its price.

The main problem stems from high feedback latency from listening to the audio output. When it had finished conveying information, the participant had already moved and made the information obsolete. It gets worse as the system classifies the users' hands as people. They complained on the overwhelming obsolete information from the system and must move slowly. This latency problem becomes the bottleneck of the system speed performance, making it uncomfortable to use and reduces their score on the Accurate factor. This problem can be solved by accelerating the audio playback speed and/or adding a new output system. Further testing is required.

In each participant's additional comments, the participants who gave bad scores commented on how the bag should be made from a more comfortable material. This could be caused by Indonesia's hot tropical climate especially in the dry season. Meanwhile, the participants who gave good scores showed genuine interest in using the proposed system in their daily life and are curious on what to expect next. Their concern about the proposed system's waterproof capability can be addressed by making the proposed system's container waterproof and airtight to protect the components inside. Further consultation with an expert is required.

6. **Conclusions.** We propose an independent, low-cost, efficient, portable, and user friendly system architecture that utilizes CV to aid VI users with their navigation. The proposed system is accurate and fast enough for real-time tracking, although better performance is preferable. Findings indicate that sign recognition helps the system in user navigation greatly. The proposed system is compact enough to wear around, is independent from Internet services or a separate processing unit, is efficient in utilizing the available resources, easy to learn, and easy to use. However, the physical design needs improvement, and there is a huge latency when conveying information to the user, making the system not feel as comfortable, lightweight, and user-friendly as expected.

Future works include re-designing the physical system placement and physical design (by consulting an expert), replace the RPi4B with a newer, cheaper, and faster edge computing device as the embedded system component, replace the detection and tracking method with newer, faster, and more accurate method, implement a faster and/or adding more feedback output medium aside from audio output in order to send more information faster to the user and reduce the information transmission latency, and add new functions to the proposed system.

Acknowledgment. This work is supported by Research and Technology Transfer Office, Bina Nusantara University as a part of Bina Nusantara University's International Research Grant contract number: No. 026/VR.RTT/IV/2020 of PIB52 and contract date: 6 April 2020.

REFERENCES

[1] G. C. Allen, *Understanding AI Technology*, https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf, Accessed on 2020-05-07.
[2] University of Helsinki, *Elements of AI*, https://course.elementsofai.com/, Accessed on 2020-05-07.
[3] Y. Bengio, A. Courville and P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.8, pp.1798-1828, 2013.
[4] A. Krenker, J. Bester and A. Kos, Introduction to the artificial neural networks, *Artificial Neural Networks: Methodological Advances and Biomedical Applications*, pp.1-18, 2011.
[5] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.
[6] R. Stureborg, *Conv Net for Dummies*, https://towardsdatascience.com/conv-nets-for-dummies-a-bottom-up-approach-c1b754fb14d6, Accessed on 2020-05-07.
[7] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You only look once: Unified, real-time object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.779-788, 2016.

[8] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv Preprint*, arXiv: 2004.10934, 2020.
[9] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., TensorFlow: A system for large-scale machine learning, *The 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI16)*, pp.265-283, 2016.
[10] World Health Organization, *Blindness and Vision Impairment*, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, Accessed on 2020-05-07.
[11] WeCapable, *Daily Life Problems Faced by Blind People*, https://wecapable.com/problems-faced-by-blind-people/, Accessed on 2020-05-07.
[12] Steven Bradley, *An Introduction to Semiotics – Signifier and Signified*, http://vanseodesign.com/web-design/semiotics-signifier-signified/, Accessed on 2020-05-07.
[13] Steven Bradley, *Icon, Index, and Symbol – Three Categories of Signs*, https://vanseodesign.com/web-design/icon-index-symbol/, Accessed on 2020-05-07.
[14] C. Koulamas and M. T. Lazarescu, Real-time embedded systems: Present and future, *Electronics*, vol.7, no.9, 205, https://doi.org/10.3390/electronics7090205, 2018.
[15] N. Wojke, A. Bewley and D. Paulus, Simple online and realtime tracking with a deep association metric, *2017 IEEE International Conference on Image Processing (ICIP)*, pp.3645-3649, 2017.
[16] R. Jiang, Q. Lin and S. Qu, Let blind people see: Real-time visual recognition with results converted to 3D audio, *Report No. 218*, Standord University, Stanford, USA, 2016.
[17] P. J. Duh, Y. C. Sung, L. Y. F. Chiang, Y. J. Chang and K. W. Chen, V-eye: A vision-based navigation system for the visually impaired, *IEEE Trans. Multimedia*, vol.23, pp.1567-1580, 2020.
[18] K. Yang, K. Wang, L. M. Bergasa, E. Romera, W. Hu, D. Sun, J. Sun, R. Cheng, T. Chen and E. López, Unifying terrain awareness for the visually impaired through real-time semantic segmentation, *Sensors*, vol.18, no.5, 1506, 2018.
[19] E. Yohannes, T. K. Shih and C. Y. Lin, Content-aware video analysis to guide visually impaired walking on the street, *International Visual Informatics Conference*, pp.3-13, 2019.
[20] Y. Lin, K. Wang, W. Yi and S. Lian, Deep learning based wearable assistive system for visually impaired people, *Proc. of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, Korea, 2019.
[21] M. L. Mekhalfi, F. Melgani, A. Zeggada, F. G. De Natale, M. A. M. Salem and A. Khamis, Recovering the sight to blind people in indoor environments with smart technologies, *Expert Systems with Applications*, vol.46, pp.129-138, 2016.
[22] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, Smart guiding glasses for visually impaired people in indoor environment, *IEEE Trans. Consumer Electronics*, vol.63, no.3, pp.258-266, 2017.
[23] N. Long, K. Wang, R. Cheng, W. Hu and K. Yang, Unifying obstacle detection, recognition, and fusion based on millimeter wave radar and RGB-depth sensors for the visually impaired, *Review of Scientific Instruments*, vol.90, no.4, 044102, 2019.
[24] R. Tachiquin, R. Velázquez, C. Del-Valle-Soto, C. A. Gutiérrez, M. Carrasco, R. De Fazio, A. Trujillo-León, P. Visconti and F. Vidal-Verdú, Wearable urban mobility assistive device for visually impaired pedestrians using a smartphone and a tactile-foot interface, *Sensors*, vol.21, no.16, 2021.
[25] G. Fusco and J. M. Coughlan, Indoor localization for visually impaired travelers using computer vision on a smartphone, *Proc. of the 17th International Web for All Conference*, pp.1-11, 2020.
[26] Y. Tian, X. Yang, C. Yi and A. Arditi, Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments, *Machine Vision and Applications*, vol.24, no.3, pp.521-535, 2013.
[27] S. Wang, H. Pan, C. Zhang and Y. Tian, RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs, *Journal of Visual Communication and Image Representation*, vol.25, no.2, pp.263-272, 2014.
[28] H. C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré and D. Rus, Enabling independent navigation for visually impaired people through a wearable vision-based feedback system, *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp.6533-6540, 2017.
[29] M. Poggi and S. Mattoccia, A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning, *2016 IEEE Symposium on Computers and Communication (ISCC)*, pp.208-213, 2016.
[30] M. Martinez, K. Yang, A. Constantinescu and R. Stiefelhagen, Helping the blind to get through COVID-19: Social distancing assistant using real-time semantic segmentation on RGB-D video, *Sensors*, vol.20, no.18, 5202, 2020.