Dynamic Crosswalk Scene Understanding for the Visually Impaired

Shishun Tian, Minghuo Zheng, Wenbin Zou, Xia Li, Lu Zhang

Dynamic Crosswalk Scene Understanding for the Visually Impaired

Shishun Tian, Minghuo Zheng, Wenbin Zou, Xia Li, and Lu Zhang

***Abstract*— Independent mobility poses a great challenge to the visually impaired individuals. This paper proposes a novel system to understand dynamic crosswalk scenes, which detects the key objects, such as crosswalk, vehicle, and pedestrian, and identifies pedestrian traffic light status. The indication of where and when to cross the road is provided to the visually impaired based on the crosswalk scene understanding. Our proposed system is implemented on a head-mounted mobile device (SensingAI G1) equipped with an Intel RealSense camera and a cellphone, and provides surrounding scene information to visually impaired individuals through audio signal. To validate the performance of the proposed system, we propose a crosswalk scene understanding dataset which contains three sub-datasets: a pedestrian traffic light dataset with 7447 images, a dataset of key objects on the crossroad with 1006 images and a crosswalk dataset with 3336 images. Extensive experiments demonstrated that the proposed system was robust and outperformed the state-of-the-art approaches. The experiment conducted with the visually impaired subjects shows that the system is practical useful.**

***Index Terms*— The visually impaired, blind navigation, crosswalk detection, pedestrian traffic light recognition, object detection.**

I. Introduction

ACCORDING to the World Health Organization [1], there are approximately 2.2 billion visually impaired people and 45 million blind people in the world. The independence and security of daily outdoor travel are crucial and necessary for them. However, lacking the capability to sense ambient environments effectively, the visually impaired feel inconvenient and always encounter various dangers, especially when crossing the roads. Therefore, navigation and obstacle avoidance are two fundamental tasks in blind assistant. Global Positioning System (GPS) is free of charge but the precision heavily depends on satellite geometry, signal blockage, atmospheric conditions, and receiver design quality, which is not stable enough for blind navigation. Besides, the blind track is also widely used to help the visually impaired walk easily in urban areas but it would be blocked by various obstacles. Regarding obstacle avoidance, white cane and guide dog are typically used to help visually impaired people walk outside. However, the perception range of white cane is quite limited and the guide dog is not affordable for everybody. Therefore, even with these resources, mobility autonomy in outdoor environments is still a challenge for the visually impaired.

To this end, many researchers have targeted at the independent outdoor travel. Some solutions have been proposed to help the visually impaired people move independently. For instance, obstacle detection systems [2]–[4] based on non-vision sensors (ultrasonic sensor, infrared sensor, LIDAR) and vision sensors (RGB camera, RGB-D camera), are proposed to avoid nearby obstacles. Based on the information from various non-vision sensors, the non-vision sensor based systems localize the obstacle and avoid it by calculating the distance between the users and obstacles. Although these systems can achieve high localization precision, the category of obstacle remains unknown. Differently, the vision sensor based systems use computer vision techniques to recognize the obstacles and thus provide more precise obstacle information to the users. Crosswalk guidance systems [5], [15], [29] which consist of crosswalk detection and pedestrian traffic light status discrimination are presented to tackle the challenge of road crossing. These systems apply computer vision algorithms to provide more precise navigation information by crosswalk detection and traffic light detection. However, current algorithms cannot obtain satisfactory results.

In this paper, we propose a crosswalk scene understanding system, which detects the location and measures the distance of key objects on the crossroad. The crosswalk detection helps the visually impaired to adjust their position when crossing the road. The presented pedestrian traffic light recognition module improves the detection accuracy by distinguishing pedestrian traffic light (PTL) and vehicle traffic light (VTL). It also reduces false drop probabilities if the traffic light is extinguished. Then the recognized and analyzed results are used to tell the blind through voice guidance in real-time.

Our main contributions are summarized as follows:

1) We propose a crosswalk scene understanding dataset, named SensingAI dataset, which is composed of

Manuscript received February 4, 2021; revised April 6, 2021 and June 7, 2021; accepted July 2, 2021. Date of publication July 12, 2021; date of current version July 29, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61771321, Grant 61871273, and Grant 61872429; in part by the Key Project of DEGP under Grant 2018KCXTD027; in part by the Natural Science Foundation of Guangdong Province, China, under Grant 2020A1515010959; in part by the Natural Science Foundation of Shenzhen under Grant JCYJ20200109105832261; and in part by the Interdisciplinary Innovation Team of Shenzhen University. *(Shishun Tian and Minghuo Zheng contributed equally to this work.) (Corresponding author: Wenbin Zou.)*

Shishun Tian, Minghuo Zheng, Wenbin Zou, and Xia Li are with the Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Electronics and Information Engineering, Institute of Artificial Intelligence and Advanced Communication, Shenzhen University, Shenzhen 518060, China (e-mail: stian@szu.edu.cn; 1910433086@email.szu.edu.cn; wzou@szu.edu.cn; lixia@szu.edu.cn).

Lu Zhang is with the National Institute of Applied Sciences of Rennes (INSA de Rennes), 35700 Rennes, France, and also with the Institut d'Électronique et des Technologies du numéRique (IETR), 35708 Rennes, France (e-mail: lu.ge@insa-rennes.fr).

Digital Object Identifier 10.1109/TNSRE.2021.3096379

three sub-datasets: a pedestrian traffic light dataset with 7447 images, a dataset of key objects on the crossroad with 1006 images, and a crosswalk dataset with 3336 images.[1]

2) A novel crosswalk detection method is proposed to indicate the visually impaired where to cross the road.
3) A pedestrian traffic light recognition module is presented to distinguish PTL and VTL automatically.
4) A depth image based distance measurement is proposed to achieve better understanding of the crosswalk scene.

The rest of this paper is organized as follows: Section 2 provides the related works on blind navigation, crosswalk detection and pedestrian traffic light detection. Then the proposed method including crosswalk detection, pedestrian traffic light recognition and distance measurement is illustrated in detail in Section 3. The experimental results and analysis are shown in Section 4. The final conclusion is drawn in Section 5.

II. Related Work

A. Blind Navigation

Nowadays, the indoor and outdoor navigation for the visually impaired are still widely studied worldwide. In literature, many blind navigation equipments and guidance systems are proposed from the perspective of obstacle detection and route planning. The non-vision sensors such as ultrasonic sensor [21], infrared sensor [26], LIDAR [17] are widely used in obstacle detection systems [2]–[4], [18]. These systems analyze the collected information and compute the distance, orientation of nearby obstacles by a micro-controller. Then the indications are provided to the visually impaired by an audio signal. These systems only provide the distance, orientation and size of obstacles, lacking the category information. They perform well on large obstacles, but the accuracy of small obstacles needs to be improved.

With the development of vision sensors, more and more vision sensors, such as RGB sensors [15], [39], [42] and RGB-D sensors [5], [24], [25], [40], [41], are used in blind navigation systems. These systems combine vision information with GPS and white cane. They further detect the category of obstacles and achieve more accurate navigation, compared with the non-vision sensor based systems. Traditional digital image processing algorithms such as image filtering, image denoising, edge detection and Hough Transform, deep learning algorithms such as object detection and semantic segmentation, are used in these systems to help understand the surrounding scenes. For instance, Aladrén *et al.* [24] proposed blind navigation system based on floor segmentation. They detected and classified the main structural elements of the scene, providing the user with obstacle-free paths. Yang *et al.* [15] proposed a real-time semantic segmentation based navigation assistance system. It exploits semantic segmentation technology to help visually impaired perceive traversable areas, stairs, etc. Similarly, some crosswalk guidance systems [5]–[9], [15], [29] which consist of crosswalk detection and PTL status discrimination are presented to tackle the challenge of crossing the roads. To improve the accuracy and robustness of detection, some related algorithms about crosswalk detection and traffic light recognition have been proposed. We will introduce them in the next section.

[1]The dataset will be released on our website. https://ivas.szu.edu.cn

B. Crosswalk Detection

Crossing the road is a great challenge for the visually impaired. An effective crosswalk detection method can accurately indicate the visually impaired where to cross the road. Se [43] first proposed a pedestrian crosswalk detection by grouping lines and checking for concurrency using the vanishing point constraint. But it holds a high computational complexity. Huang and Lin [11] presented an improved method of zebra crossing detection based on bipolarity, similar work in [16], [35]. Image blocks with different sizes, which are set artificially, are used in bipolarity segmentation to detect the areas of alternating black and white stripes. Good performance can be achieved on normal stripe regions, but there still exists certain room for the improvement of accuracy on small stripe regions. Cheng *et al.* [16] proposed a crosswalk detection method based on adaptive thresholding binarization and candidate consistency analysis. But the results can be affected by background information such as buildings and vehicles. Wei *et al.* [36] and Asami *et al.* [37] applied template matching for candidate detection. Limited scenarios are detected by their methods, and the accuracy remains to be promoted. Wang and Tian [10] proposed a zebra crossing recognition based Canny edge detection [12] and Hough Transform. Similarly, Chen *et al.* [13] proposed a zebra crossing recognition method based on Sobel edge extraction and Hough Transform. These Hough Transform based algorithms achieved a good trade-off between accuracy and speed, performed well on most scenarios.

C. Pedestrian Traffic Light Detection

Pedestrian traffic light recognition indicates the visually impaired when to cross the road. With the development of computer vision algorithms, several efforts have been dedicated to pedestrian traffic light detection. Volodymyr Ivanchenko *et al.* [30] proposed a real-time walk light detection based on horizon searching and image matching. They searched for the PTL in the image strip near the horizon. Then high scored candidates were extracted using template convolution. Masectti *et al.* [31] and Roters *et al.* [38] applied range filter and color filter to group contiguous pixels as candidates. Then PTL is extracted by similarity evaluation and feature extraction. The accuracy and robustness of these schemes need to be further guaranteed.

Besides, machine learning based algorithms are widely deployed in recent years. Cheng *et al.* [32] developed color segmentation algorithm to generate candidates. Then Histogram of Oriented Gradient (HOG) descriptor and Support Vector Machine (SVM) are extracted to recognize the candidates. Similarly, Adaboost [33], YOLOv2 [34] and Faster R-CNN [34] algorithms are also used for candidate extraction and classification. These methods showed high accuracy when there exists a PTL in the image. However, VTL and extinguished traffic light are usually recognized as a PTL mistakenly.

In this paper, we propose a crosswalk scene understanding system, which comprises an HSV (Hue, Saturation, Value) color space based crosswalk detection, a pedestrian traffic light recognition and a distance measurement algorithm. These proposed methods are more accurate and more robust than the methods introduced above.

Fig. 1. The examples of crosswalk dataset. (a) negative samples; (b) positive samples, which are respectively complete, broken, waterlogged and badly-lighted crosswalk from left to right.

Fig. 2. The head-mounted assistive device for the blind (SensingAI G1): an RGB-D camera, a bone conduction headset, and a blind assistant app integrates the proposed system and GPS navigation.

III. Crosswalk Scene Understanding System

We propose a crosswalk scene understanding dataset for further study on this topic. It consists of three subdatasets: a pedestrian traffic light dataset, a key object dataset and a crosswalk dataset. First, our pedestrian traffic light dataset contains 36 scenes and 2 labels: red light and green light. It is based on two sources. One source is the public dataset of Germany [44], the others are our manual dataset of China (account for approximately 92%). The manual dataset is captured by Honor-8 mobile phone with 1080 × 1920 resolution. The entire dataset, including manually labeled ground truth data, is composed of three parts: training set (2982 images), validation set (1278 images) and testing set (3187 images). In the testing set, the 2832 positive samples are PTL, the 825 negative samples include VTL, extinguished traffic light and non-PTL. Second, the dataset of key objects on the crossroad contains 4 labels: red light, green light, pedestrian and vehicle. It is based on our manual pedestrian traffic light dataset. Third, our crosswalk dataset, which is for the purpose of testing, is established in Fig. 1. The crosswalk dataset contains 3336 images, most of which are captured in 1080 × 1920 resolution. The 2486 positive samples (Fig. 1(b)) include complete, broken, waterlogged and badly-lighted crosswalks, and the 850 negative samples (Fig. 1(a)) without crosswalk.

As depicted in Fig. 2, SensingAI G1 is a lightweight, low-power consumption, low-cost and portable head-mounted assistive device for the blind with only 89g weight. It is composed of an RGB-D camera, a bone conduction headset, and a blind assistant APP. The android APP runs with 15-17 FPS and integrates the proposed system and GPS navigation which is generally used by the blind people. The RGB-D camera perceives the surroundings, the cellphone processes the camera's data into navigational information, and the bone conduction headset makes it possible for the blind to collect navigational information and surrounding sound. The RGB-D camera (Intel RealSense D415) includes a depth camera with an additional inertial measurement unit which measures linear accelerations and angular velocities. It has

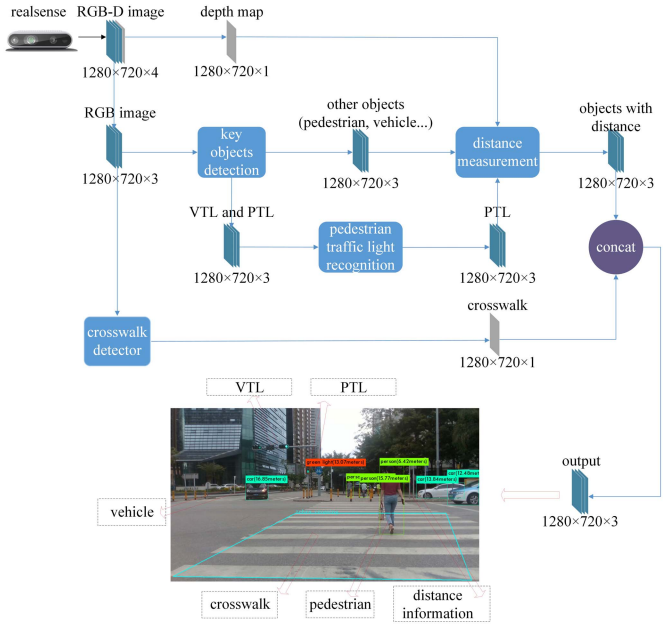

Fig. 3. The crosswalk scene understanding system. The crosswalk, pedestrian, vehicle and PTL are detected. The distance of key objects(pedestrian, vehicle, PTL...) are measured. VTL:vehicle traffic light; PTL:pedestrian traffic light.

a resolution of 1280 × 720 pixels and a 69.4° × 42.5° field of view, with a possible range from 0.1m to 10m (actually up to 60m). Its small dimension (99 × 20 × 23mm^3) makes it suitable for the head-mounted device. The data collected in front of the user are sent to the cellphone (Snapdragon 865 CPU, 8GB + 128 GB memory) for processing.

The framework of the proposed system is shown in Fig. 3, it enhances the white cane's function by detecting key information on the crossroad. It is composed of crosswalk detector, key objects detector (YOLOv4 [14]), pedestrian traffic light recognition and distance measurement modules. The crosswalk detection based on HSV color space indicates the visually impaired where the crossroad is and provides the correct

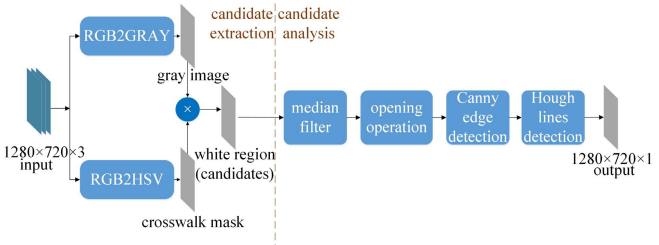

Fig. 4. The framework of crosswalk detector.

straight walking direction. Then key objects such as pedestrian, vehicle, PTL and VTL are detected by a YOLOv4-based key objects detector. In this paper, the aspect ratios of anchors in the first feature map of YOLOv4 are modified to 1, 1/2, 1/3 to detect more small objects (*e.g.* traffic lights). The pedestrian traffic light recognition further distinguishes PTL from VTL. It guides the visually impaired whether it is time to cross the crosswalk. The depth image based distance measurement is combined with object detection to measure the distance of key objects on the crossroad. A warning signal will be provided when any fast-approaching pedestrian or vehicle appears within 3 meters. Finally, the recognized and analyzed results are used to notify the blind via Bluetooth in real-time.

A. Crosswalk Detector

A crosswalk is composed of several bright white stripes and dark background, so parallel straight edges of crosswalk stripes can be recognized as a kind of feature of crosswalks. The bright stripes alternate with a dark background in crosswalks, so the periodic gray value distribution is also a feature of crosswalks. In this paper, a novel crosswalk detection algorithm is proposed based on these features, as shown in Fig. 4.

Candidate extraction and analysis are two components that constitute the detector. The former extracts candidates based on the feature of bright white stripes, the latter analysis candidates using the feature of parallel straight edges.

1) Candidate Extraction: As shown in Fig. 5, the input image (Fig. 5(a)) captured by RGB camera, is transformed into an HSV image (Fig. 5(b)) since the HSV color space can better represent the color information of an image than RGB color space. At the same time, the corresponding gray image (Fig. 5(c)) is also obtained to represent the texture information. According to the feature of white stripes and black background, we extract the white region in the HSV image as a crosswalk mask (Fig. 5(d), a binary image). Then the white region (regarded as candidates) is generated by multiplying the crosswalk mask to gray image. Although most of the noises on the image are filtered in this way, some noises (Fig. 5(e)) caused by buildings, sky, vehicle still exist. These noises are filtered and the crosswalk region is detected in candidates by candidate analysis.

2) Candidate Analysis: In this component, median filtering and morphology opening operation are used to remove small noises, as is shown in Fig. 5(f). According to the feature of parallel straight edges, in Fig. 5(g) and Fig. 5(h), Canny edge detection and Hough Transform algorithm are applied to detect crosswalk's edges and lines respectively. Then we have the

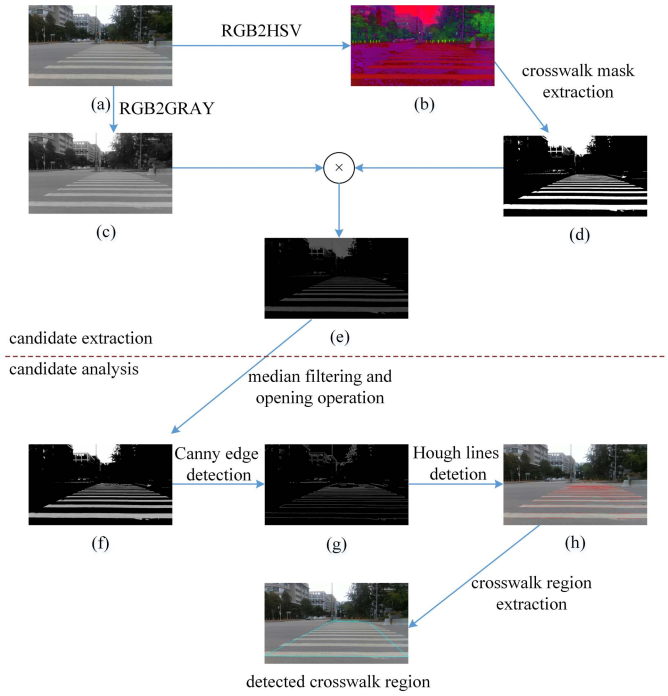

Fig. 5. Diagram of candidate extraction and candidate analysis: (a) input image; (b) HSV image; (c) gray image; (d) crosswalk mask; (e) white region(candidates); (f) result of median filter; (g) result of Canny edge detection; (h) result of Hough transform.

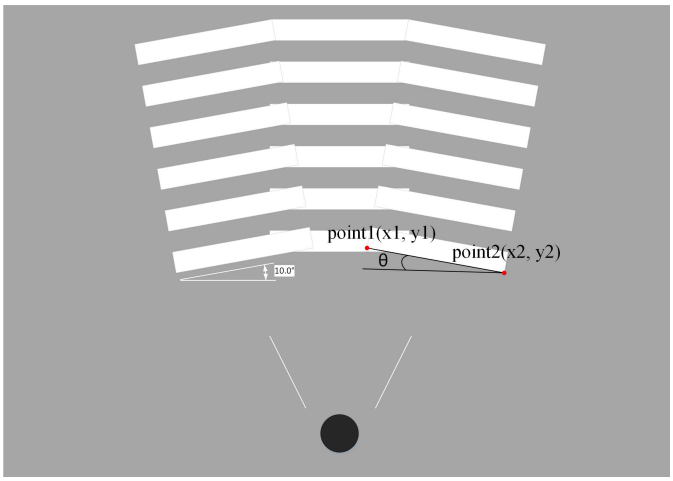

Fig. 6. The sketch of θ, which ranges in [0, 10°].

endpoint coordinates of all lines. In Fig. 6, for a detected line with two endpoints $(x1, y1)$ and $(x2, y2)$, θ represents the absolute angle between the detected line and the horizontal line, which is defined by (1):

$$\theta = arctan(\frac{|x1 - x2|}{|y1 - y2|}). \tag{1}$$

We follow the following three conditions to extract the crosswalk region:

1) The minimum length of a line is 1/4 of the image's width. If the length of a line doesn't satisfy this condition, it will be eliminated in the following computation because it is far away from the visually impaired.

2) The range of θ is set to[0, 10°]. If the angle of a detected crosswalk strip is larger than 10°, it is not a correct straight walking direction for crossing the road.
3) The minimum number of lines which both satisfy condition 1 and condition 2 simultaneously is set to 10. Since a crosswalk usually comprises multiple strips, the lines would not be crosswalk if their number is smaller than the minimum number.

The lines selected by above-mentioned conditions are clustered into a set, which is regarded as the crosswalk region. In Fig. 5, we highlight the crosswalk region on the output image with green lines. It is worth noting that the edge of the crosswalk is outlined with a trapezoid rather than a simple rectangle.

Crosswalk detection is an effective way to find where to cross the road. It indicates whether the visually impaired arrives at a crossroad and provides the correct straight walking direction. The information about when to cross the road is provided by pedestrian traffic light recognition.

B. Pedestrian Traffic Light Recognition

The pedestrian lights are used at intersections to notify the pedestrians when to cross the street. The pedestrian traffic light detection provides the visually impaired with the status of pedestrian signals. In this paper, candidates for PTL are detected by a YOLOv4 based key objects detector adapted for localizing and recognizing PTL. The object detector provides the boundaries (x, y, w, h) and objectness confidence score that quantifies the classification confidence and location of an object.

The candidates, including the PTL, VTL and the extinguished traffic light, are detected by the key object detector. However, only PTL is needed for blind navigation in this paper. Pedestrian traffic light recognition is proposed to prune the unqualified ones, as shown in Fig. 7. The same as the proposed crosswalk detector, first, the candidate (Fig. 7(b)) is extracted from RGB image (Fig. 7(a)) by the object dector. Then it is converted to gray image (Fig. 7(c)) and HSV image (Fig. 7(d)) to obtain the texture and color information. Next, a traffic light mask (Fig. 7(e)), containing both red and green regions, is extracted from HSV image. Then the shape of traffic light (Fig. 7(f)) is generated by multiplying the traffic light mask to gray image. The minimum bounding rectangle (MBR) of the shape of traffic light is highlighted with yellow dotted lines in Fig. 7(f). The qualified candidates are selected by the criteria:

$$AspectRatio = \frac{h}{w}. \tag{2}$$

where h and w indicate the height and width of the candidate region respectively.

AspectRatio denotes the aspect ratio of MBR. Fig. 8 collects various types of traffic light, including the vehicle traffic lights and the pedestrian ones, together with their aspect ratios. The arrow-shaped traffic light and round-shaped traffic light are VTL, whose AspectRatio approximately equals 1. The human-shaped traffic light is PTL, whose AspectRatio is larger than 1.5. The geometrical properties of same shape traffic light devices are similar, because of their similar production standards. Therefore, the threshold of AspectRatio is set to 1.35.

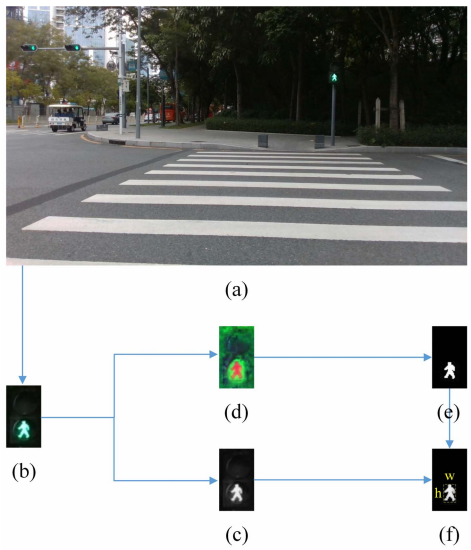

Fig. 7. Diagram of pedestrian traffic light recognition: (a) RGB image; (b) candidate; (c) gray image; (d) HSV image; (e) traffic light mask; (f) the shape of traffic light.

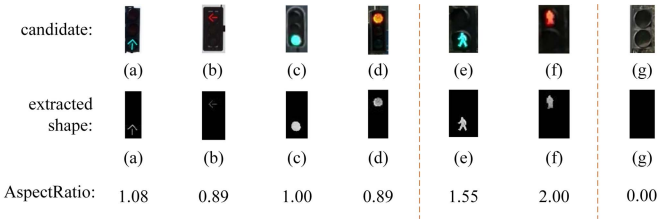

Fig. 8. The AspectRatio of different shapes of traffic light: (a-d) VTL; (e-f) PTL; (g) extinguished traffic light.

Unqualified candidates are pruned in this way: The confidence score of a detected candidate is set to 0 if its AspectRatio is smaller than the threshold.

The traffic light recognition improves the accuracy through extracting qualified candidates detected by the key objects detector. It distinguishes PTL from VTL and extinguished traffic light by analyzing their geometrical properties. Three traffic light statuses, "red", "green" and "flashing" are provided to notify the visually impaired when to cross the street through audio signals.

In addition to the traffic light, the location of other objects such as pedestrian and vehicle, also contribute to the crosswalk scene understanding. Thus, the YOLOv4 based key object detector detects not only the traffic lights but also other key objects, such as vehicle and pedestrian. The distance measurement of these key objects are introduced in the next section.

C. Distance Measurement

The location of other traffic participants are crucial for the understanding of crosswalk scene. For example, it is dangerous to cross the road if there is a vehicle in front of the visually impaired even if the PTL is "green", and it is usually safe to cross if there exist other pedestrians walking on the crosswalk under the condition of "green" PTL.

In this section, we propose a key object distance measurement based on RGB-D imges. The RGB image is used by the YOLOv4 based object detector to extract the key object

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Methods	TP	FN	TN	FP	Pre	Rec	Acc	FPS
Wang's[10]	1891	595	637	213	0.8987	0.7606	0.7577	**36.80**
Huang's[11]	1987	499	713	137	0.9355	0.7992	0.8093	21.8
Michael's[12]	2422	64	273	577	0.8076	0.9743	0.8078	16.3
Chen's[13]	2321	165	706	144	0.9416	0.9336	0.9074	30.2
ours	2464	22	826	24	**0.9903**	**0.9912**	**0.9862**	30.7

TP: true positive; FN: false negative; TN: true negative; FP: false positive; Pre: precision; Rec: recall; Acc: accuracy; FPS: frame per second.

boundaries (x, y, w, h), based on which the distance of these key objects are calculated:

$$distance = min(D[i][j]), (i \in [x, x+w], j \in [y, y+h]). \quad (3)$$

where $D[i][j]$ is the depth map captured by the RealSense camera. The nearest distance is regarded as the distance between an object and the visually impaired.

The distance measurement together with crosswalk detection, key objects detection and pedestrian traffic light recognition, constitute our crosswalk scene system. SensingAI G1 indicates the visually impaired where and when to cross the road. When the traffic light status is "green", the locations of nearby vehicle and pedestrian are provided to the visually impaired to help identify traffic situations.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

We conducted extensive experiments on crosswalk detection, pedestrian traffic light recognition and distance measurement. The experiments were implemented on a cellphone with Qualcomm Snapdragon 865 CPU, 8 GB + 128 GB memory. We carried out experiments on our proposed SensingAI dataset. which has been introduced in Section III. The crosswalk detection method was compared with four state-of-the-art methods. Ablation study was conducted in order to verify the efficiency of each module in our crosswalk detection method. The results validated the high accuracy and efficiency of our proposed method. We also verified the improvements of modifying anchors in training phase and using pedestrian traffic light recognition module in prediction. Experimental results verified the promotion of anchor modification and the pedestrian traffic light recognition module to traffic light detection. In addition, in order to obtain the feedback of the visually impaired, we conduct a test with the visually impaired in which 4 blind person participate to test SensingAI G1.

A. Crosswalk Detection

We carried out experiments on our crosswalk dataset. Our method was compared with four crosswalk detection algorithms: Wang's [10], Huang's [11], Michael's [12] and Chen's [13]. In TABLE I, the experimental results show that the proposed method achieves the best among all the tested methods. Although Wang's method runs slightly faster than ours, it is significantly outperformed by our method in terms of the Pre, Rec, Acc. The qualitative comparison of these crosswalk detection algorithms is shown in Fig. 9.

We also conduct ablation study to validate the efficiency of each module in the proposed method. The experimental results are shown in TABLE II. The TOHSV means extracting candidates on HSV image. The median filtering (MED) is added based on the gray level transformation (TOGRAY), and the morphological opening operation (OPEN) is added based on the edge detection. Except for the necessary module TOGRAY, the performance is improved by introducing the modules: TOHSV, MED and OPEN. As a result, our main modules for crosswalk detection includes the following: TOGRAY, TOHSV, MED and OPEN.

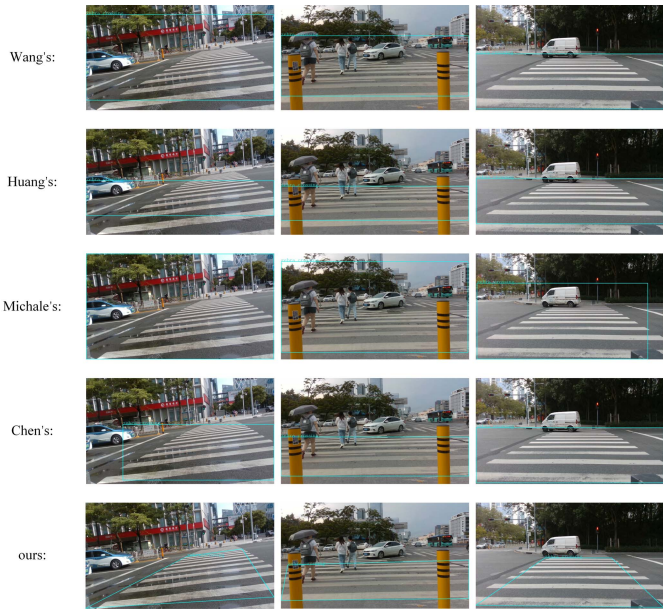

Fig. 9. The qualitative comparison of crosswalk detection algorithms in different scenarios.

TABLE II
THE RESULTS OF ABLATION STUDY

TOGRAY	TOHSV	MED	OPEN	Pre	Rec	Acc	FPS
✓		✓	✓	0.9254	**0.9932**	0.9353	28.1
✓	✓		✓	0.9722	0.9409	0.9359	27.2
✓	✓	✓		0.9734	0.9276	0.9272	28.0
✓	✓	✓	✓	**0.9903**	0.9912	**0.9862**	**30.7**

TOGRAY: converting RGB image to gray image; TOHSV: converting RGB image to HSV image and extracting candidates; MED: median filtering; OPEN: morphological opening operation; Pre: precision; Rec: recall; Acc :accuracy.

It can be observed from TABLE II that the TOHSV proposed in our method improves the precision and accuracy obviously. The precision increases by 6.49% and the accuracy increases by 5.09% when the TOHSV module is added. The reason of this improvement is that the TOHSV extracts crosswalk from the input image efficiently. Most of the background noises are removed by TOHSV. The MED and OPEN improve the performance slightly. They are both used to get cleaner edges and crosswalk stripes by removing high-frequency noises. However, they have different roles. The MED is mainly used to filter impulse noise, the OPEN is mainly used to fill up cavities, which always exist on the candidate extraction results of broken, waterlogged and shadowed crosswalk.

B. Pedestrian Traffic Light Recognition

The experiments are conducted on our traffic light dataset. The improvements of modifying anchors in training phase and the pedestrian traffic light recognition module used in prediction are shown in TABLE III.

As is shown in TABLE III, the mAP is 0.7905 and the recall is 0.8212 before adding MA and PTLR. They increase to 0.8666 and 0.8372 when the MA module is added. The reason

TABLE III
THE IMPROVEMENTS OF OUR METHODS

MA	PTLR	TL	TP	FP	FN	mAP	Rec	mIOU	FPS
		red light & green light	2618	694	214	0.7905	0.8212	0.6806	**42.8**
✓		red light & green light	2669	411	163	0.8666	**0.8372**	0.7498	42.7
	✓	red light & green light	2566	333	266	0.8851	0.8049	0.7635	33.3
✓	✓	red light & green light	2615	183	217	**0.9346**	0.8203	**0.8106**	33.3

MA: modifying anchors in training; PTLR: using pedestrian traffic light recognition module in prediction; TL: traffic light; TP: true positive; FP: false positive; FN: false negative; mAP: mean average precision; Rec: recall; mIOU: mean intersection over union; FPS: frame per second.

of this improvement is that the shape of anchor boxes we set in the first feature map can better match the object shape, and thus make the loss function easier to descend. When the PTLR is added separately, the mAP increases to 0.8851 since it can distinguish PTL from vehicle traffic and extinguished traffic ones. In our experiments, the detector achieves best performance when using both MA and PTLR. The mAP increases by 14.41% and the mIOU (mean intersection over union) increases by 14%. The FPS declines slightly because the added modules need extra 6 to 7 milliseconds for computation.

A direct comparison with other solutions is unfeasible because it is not possible to test other solutions with our experimental setting due to the unavailability of the implementations of previous solutions. The results of YOLOv4 and our method are shown in Fig. 10. It shows that YOLOv4 (Fig. 10 (a)) recognizes a traffic sign, an extinguished PTL and a VTL as green PTL by mistake, while our proposed method (Fig. 10 (b)) gives more accurate results.

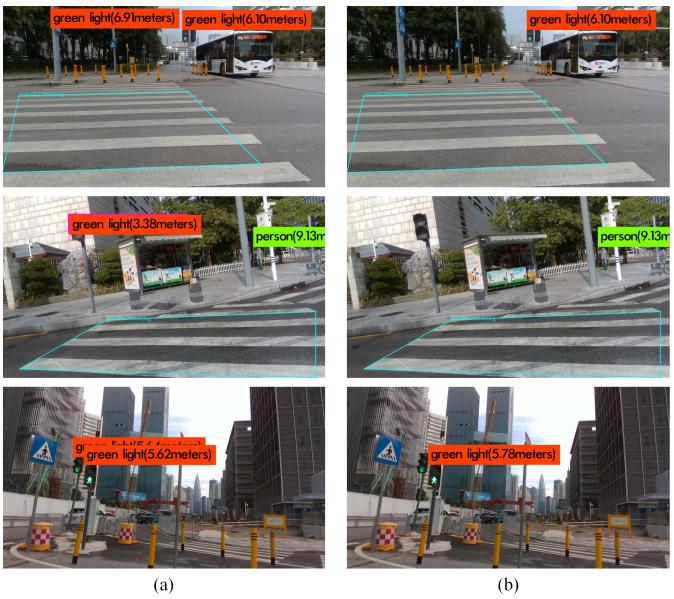

Fig. 10. The detected results by YOLOv4 (a) and our method (b).

C. Distance Measurement

The experiments are conducted to verify the accuracy of distance measurement. In this paper, an HCJYET laser rangefinder, is used to measure the actual distance of objects, as is shown in Fig. 11(a). It has a 600m measuring range, with 0.5m measuring error. The average of three measurement values of the laser rangefinder is regarded as the actual distance. Fig. 11(b) and Fig. 11(c) are measurement cases of rangefinder. In order to test the accuracy of our proposed distance measurement, we collected 18 crosswalk scenes. In each scene, 10 times of tests were conducted and their averages are taken as the final results. The experimental results are shown in Fig. 12, where the green line is the distance measured by rangefinder and the red points are the distances measured by our system. The absolute errors are shown in TABLE IV. The average absolute error ε is 0.35m when the distance is within 5m, it becomes to 0.74m and 1.65m when the distance increases to 10m and 20m respectively. The proposed distance measurement module achieves a total accuracy higher than 90%.

As is shown in Fig. 13, the proposed method provides a dynamic understanding of crosswalk scene for the visually impaired by integrate all the crosswalk detection modules: key objects detection, pedestrian traffic light recognition and distance measurement. In addition, in Fig. 13, the crosswalk detection indicates the visually impaired the location of crosswalk and provides them the correct straight walking direction. The pedestrian traffic light recognition provides more accurate and more robust information about when to cross the street. A warning signal will be provided when a flashing green light

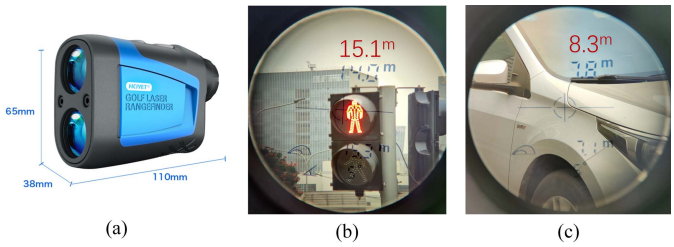

Fig. 11. The laser rangefinder and its use cases. The distances labeled in red are the results measured by our system and the distances marked in dark blue are that measured by the rangefinder which represent the ground truth. (a) The laser rangefinder; (b) The measurement case of PTL; (c) The measurement case of vehicle.

is recognized based on the pedestrian traffic light recognition. A signal will be sent when any pedestrian or vehicle appears in front of the user within 3 meters.

D. Test With the Visually Blind

Four blind volunteers were invited to be human subjects in our experiment. Before the test, all the subjects are asked to have a short training about how to use this system. Then, the subjects were taken to a nearby traffic intersection to test the system. Fig. 14 shows the traffic intersection and the testing routes. Each subject was invited to walk according to the routes. They firstly start walking follow the red path and then go back follow the blue path, and followed by a staff

TABLE IV
THE ABSOLUTE ERRORS OF MEASURED DISTANCES

actual distance (m)	1.5	2.7	3.9	4.4	5.4	6.7	7.3	8.6	9.2	11.5	13.0	13.5	13.7	14.8	17.4	18.0	18.4	18.5
absolute error (m)	0.28	0.25	0.31	0.53	0.31	0.76	1.17	1.00	2.07	1.22	1.76	1.26	2.50	1.62	5.97	4.87	1.59	2.22

TABLE V
WALKING TIME (T1-T8) OF EACH SUBJECT TO PASS THE ZEBRA CROSSING.

Subject	T1(s)	T2(s)	T3(s)	T4(s)	T5(s)	T6(s)	T7(s)	T8(s)	Total time(s)	Average speed (m/s)
S-1	28.3	30	28.4	32.6	31.1	29.5	31.4	30.1	241.4	0.47
S-2	25.9	32.1	27	27.3	33.8	26.2	25.8	27.6	225.7	0.50
S-3	23.6	27.6	23.8	25.9	28.3	24.9	26.8	21.8	202.7	0.57
S-4	25.5	27.7	24.4	30.2	31.9	25.4	29	26.3	220.4	0.51

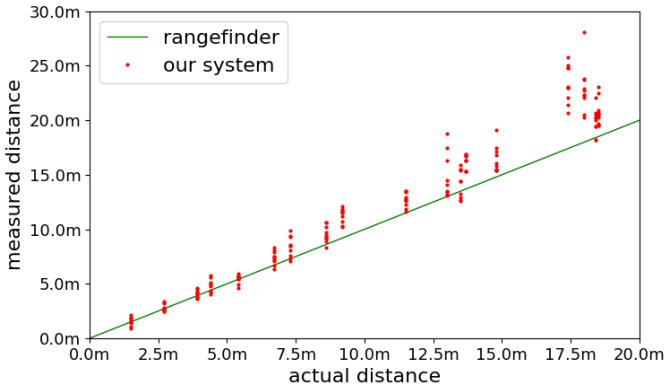

Fig. 12. The results measured by the rangefinder and the proposed system. x-axis: the actual distance; y-axis: the measured distance.

Fig. 13. Some examples of image output detected by the proposed system.

to guarantee their security. The lengths of the zebra crossings (C1, C2, C3, C4) are 13.1m, 15.9m, 11.4m and 16.3m respectively. The zebra crossing C5, C6, C7, C8 corresponds to C4, C3, C2, C1 with opposite walking direction. The walking time (represented by T1-T8) of each subject (S1, S2, S3, S4) to pass each crossing are shown in Table V.

Overall, all the blind subjects can pass the zebra crossing at a speed around 0.5m/s, which is close to the normal walking

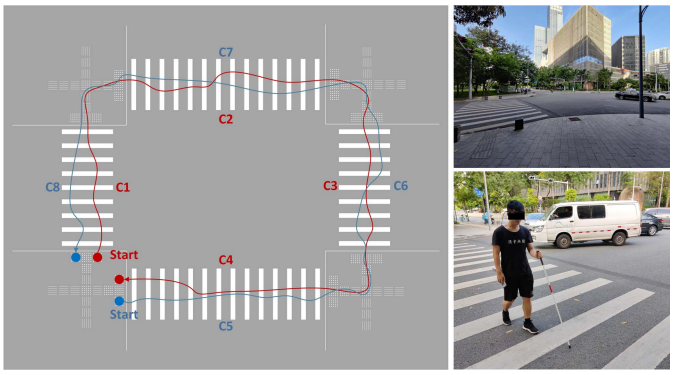

Fig. 14. The testing routes and crossroad scenes.

speed of the visually impaired. Since the traffic light and crosswalk were out of the camera field of view, the voice signal may not be provided by the system when they were walking at the end of the crosswalk. Thus, they need to walk straight for about 1m until they get to the blind track. After the test, all subjects think that the system is portable and accurate, and they also suggest that, in addition to the voice prompt, the vibration pattern could be a good way to inform them of the obstacles and the change of traffic light. For example, we can assign two different vibration mode for obstacles and traffic light change, and increase the vibration frequency when the obstacle gets closer to the subject. We will apply their valuable suggestions in our system in the future. The experimental results show that the system do help the subjects cross the road independently.

V. CONCLUSION

To help the visually impaired walk outdoors safely, in this paper, we propose a system of crosswalk scene understanding which provides them the key information such as crosswalk, PTL, pedestrian and vehicle. It guides the visually impaired where and when to cross the street.

The proposed system consists of four modules. The crosswalk detection algorithm is proposed to locate crosswalk and provide straight walking direction. The pedestrian traffic light recognition is proposed to recognize PTL, VTL and extinguished traffic light. The flashing green light can be recognized based on the pedestrian traffic light recognition. The conducted extensive experiments validated the accuracy, speed and robustness of the proposed method. In the future work, in addition to the voice prompts, the vibration pattern will be considered to inform them of the obstacles and the

change of traffic light. Besides, the "vehicle" and "pedestrian" objects detected as most important obstacles currently, we are planning to propose a new obstacle detection method which can detect any obstacles in the future.

REFERENCES

[1] *World Health Organization*. Accessed: Jan. 7, 2021. [Online]. Available: https://www.who.int/zh/news-room/detail/08-10-2019-who-launches-first-world-report-on-vision

[2] R. V. Jawale, M. V. Kadam, R. S. Gaikawad, and L. S. Kondaka, "Ultrasonic navigation based blind aid for the visually impaired," in *Proc. IEEE Int. Conf. Power, Control, Signals Instrum. Eng. (ICPCSI)*, Sep. 2017, pp. 923–928.

[3] P. Marzec and A. Kos, "Low energy precise navigation system for the blind with infrared sensors," in *Proc. 26th Int. Conf. Mixed Design Integr. Circuits Syst.*, Jun. 2019, pp. 394–397.

[4] M. Owayjan, A. Hayek, H. Nassrallah, and M. Eldor, "Smart assistive navigation system for blind and visually impaired individuals," in *Proc. Int. Conf. Adv. Biomed. Eng. (ICABME)*, Sep. 2015, pp. 162–165.

[5] M. T. Islam, M. Ahmad, and A. S. Bappy, "Development of a microprocessor based smart and safety blind glass system," in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC4ME2)*, Jul. 2019, pp. 1–4.

[6] P. Angin, B. Bhargava, and S. Helal, "A mobile-cloud collaborative traffic lights detector for blind navigation," in *Proc. 11th Int. Conf. Mobile Data Manage.*, 2010, pp. 396–401.

[7] L. Shangguan, Z. Yang, Z. Zhou, X. Zheng, C. Wu, and Y. Liu, "CrossNavi: Enabling real-time crossroad navigation for the blind with commodity phones," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Sep. 2014, pp. 787–798.

[8] M. C. Ghilardi, J. J. Junior, and I. Manssour, "Crosswalk localization from low resolution satellite images to assist visually impaired people," *IEEE Comput. Graph. Appl.*, vol. 38, no. 1, pp. 30–46, Jan./Feb. 2018.

[9] S. Ou, H. Park, and J. Lee, "Implementation of an obstacle recognition system for the blind," *Appl. Sci.*, vol. 10, no. 1, p. 282, Dec. 2019.

[10] S. Wang and Y. Tian, "Detecting stairs and pedestrian crosswalks for the blind by RGBD camera," in *Proc. IEEE Int. Conf. Bioinf. Biomed. Workshops*, Oct. 2012, pp. 732–739.

[11] X. Huang and Q. Lin, "An improved method of zebra crossing detection based on bipolarity," *Comput. Appl. Softw.*, vol. 12, no. 34, pp. 202–205, Dec. 2017.

[12] M. Hodlmoser, B. Micusik, and M. Kampel, "Camera auto-calibration using pedestrians and zebra-crossings," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV)*, Nov. 2011, pp. 1697–1704.

[13] N. Chen, F. Hong, and B. Bai, "Zebra crossing recognition method based on edge frature and Hough transform," *J. Zhejiang Univ. Sci. Technol.*, vol. 6, no. 31, pp. 476–483, Aug. 2019.

[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934

[15] K. Yang *et al.*, "Unifying terrain awareness for the visually impaired through real-time semantic segmentation," *Sensors*, vol. 18, no. 5, p. 1506, May 2018.

[16] R. Cheng *et al.*, "Crosswalk navigation for people with visual impairments on a wearable device," *J. Electron. Imag.*, vol. 26, p. 053025, Oct. 2017.

[17] R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 583–593, Mar. 2018.

[18] P. Kwiatkowski, T. Jaeschke, D. Starke, L. Piotrowsky, H. Deis, and N. Pohl, "A concept study for a radar-based navigation device with sector scan antenna for visually impaired people," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Gothenburg, Sweden, May 2017, pp. 1–4.

[19] E. Cardillo *et al.*, "An electromagnetic sensor prototype to assist visually impaired and blind people in autonomous walking," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2568–2576, Mar. 2018.

[20] C. Ton *et al.*, "LIDAR assist spatial sensing for the visually impaired and performance analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1727–1734, Sep. 2018.

[21] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 2, pp. 172–182, Apr. 2018.

[22] S. T. H. Rizvi, M. J. Asif, and H. Ashfaq, "Visual impairment aid using haptic and sound feedback," in *Proc. Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Islamabad, Pakistan, Mar. 2017, pp. 175–178.

[23] S. Sharma, M. Gupta, A. Kumar, M. Tripathi, and M. S. Gaur, "Multiple distance sensors based smart stick for visually impaired people," in *Proc. IEEE 7th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Jan. 2017, pp. 1–5.

[24] A. Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-D sensor with range expansion," *IEEE Syst. J.*, vol. 10, no. 3, pp. 922–932, Sep. 2016.

[25] K. Yang, K. Wang, W. Hu, and J. Bai, "Expanding the detection of traversable area with realsense for the visually impaired," *Sensors*, vol. 10, no. 4, p. 1954, Nov. 2016.

[26] F. Prattico, C. Cera, and F. Petroni, "A new hybrid infrared-ultrasonic electronic travel aids for blind people," *Sens. Actuators A, Phys.*, vol. 201, pp. 363–370, Oct. 2013.

[27] O. Miksik *et al.*, "The semantic paintbrush: Interactive 3D mapping and recognition in large outdoor spaces," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Seoul, South Korea, Apr. 2015, pp. 3317–3326.

[28] B. Ando, "A smart multisensor approach to assist blind people in specific urban navigation tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 6, pp. 592–594, Dec. 2008.

[29] C. Ye and X. Qian, "3-D object recognition of a robotic navigation aid for the visually impaired," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 441–450, Feb. 2018.

[30] V. Ivanchenko, J. Coughlan, and H. Shen, "Real-time walk light detection with a mobile phone," in *Proc. ICCHP*, vol. 6180, 2010, pp. 229–234.

[31] S. Mascetti, D. Ahmetovic, A. Gerino, C. Bernareggi, M. Busso, and A. Rizzi, "Robust traffic lights detection on mobile devices for pedestrians with visual impairment," *Comput. Vis. Image Understand.*, vol. 148, pp. 123–135, Jul. 2016.

[32] R. Cheng, K. Wang, K. Yang, N. Long, J. Bai, and D. Liu, "Real-time pedestrian crossing lights detection algorithm for the visually impaired," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 20651–20671, Aug. 2017.

[33] X.-H. Wu, R. Hu, and Y.-Q. Bao, "Fast vision-based pedestrian traffic light detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 214–215.

[34] R. Ash, D. Ofri, J. Brokman, I. Friedman, and Y. Moshe, "Real-time pedestrian traffic light detection," in *Proc. IEEE Int. Conf. Sci. Electr. Eng. Isr. (ICSEE)*, Dec. 2018, pp. 1–5.

[35] M. S. Uddin and T. Shioyama, "Bipolarity and projective invariant-based zebra-crossing detection for the visually impaired," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2005, p. 22.

[36] Y. Wei, X. Kou, and M. C. Lee, "A new vision and navigation research for a guide-dog robot system in urban system," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, Jul. 2014, pp. 1290–1295.

[37] T. Asami and K. Ohnishi, "Crosswalk location, direction and pedestrian signal state extraction system for assisting the expedition of person with impaired vision," in *Proc. 8th Europe-Asia Congr. Mecatronics*, Nov. 2014, pp. 285–290.

[38] J. Roters, X. Jiang, and K. Rothaus, "Recognition of traffic lights in live video streams on mobile devices," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1497–1511, Oct. 2011.

[39] A. Mancini, E. Frontoni, and P. Zingaretti, "Mechatronic system to help visually impaired users during walking and running," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 649–660, Feb. 2018.

[40] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 258–266, Aug. 2017.

[41] M.-C. Kang, S.-H. Chae, J.-Y. Sun, S.-H. Lee, and S.-J. Ko, "An enhanced obstacle avoidance method for the visually impaired using deformable grid," *IEEE Trans. Consum. Electron.*, vol. 63, no. 2, pp. 169–177, May 2017.

[42] M. L. Mekhalfi, F. Melgani, A. Zeggada, F. G. B. De Natale, M. A.-M. Salem, and A. Khamis, "Recovering the sight to blind people in indoor environments with smart technologies," *Expert Syst. Appl.*, vol. 46, pp. 129–138, Mar. 2016.

[43] S. Se, "Zebra-crossing detection for the partially sighted," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Charleston, SC, USA, Jun. 2000, pp. 211–217.

[44] J. Roters. (2011). *Pedestrian Lights Database*. Accessed: Mar. 28, 2017. [Online]. Available: http://www.uni-muenster.de/PRIA/en/forschung/index.shtml