Long-range Traversability Awareness and Low-lying Obstacle Negotiation with RealSense for the Visually Impaired

Kailun Yang, Kaiwei Wang*, Shufei Lin, Jian Bai
College of Optical Science and Engineering
Zhejiang University
Hangzhou, China
e-mail: wangkaiwei@zju.edu.cn

Luis M. Bergasa, Roberto Arroyo
Department of Electronics
University of Alcalá
Madrid, Spain
e-mail: luism.bergasa@uah.es

ABSTRACT
The application of RGB-D sensors provides a revolutionary force in the research field of computer vision, where traversable area detection and obstacle avoidance are the fundamental topics to aid visually impaired people. However, the detection using RGB-D sensors is limited by the sparse depth map and the narrow field of view, which hampers longer traversability awareness and lower obstacle negotiation. This paper proposes an effective approach to combine a pair of wearable smart glasses and a waist-worn pathfinder to assist the visually impaired. Traversable lines are generated to feedback the visually impaired through real-time stereo sound after real-world coordinates adjustment with a RealSense R200 (RGB-D sensor) and an IMU integrated in the pair of smart glasses. Low-lying obstacles are detected by using another RGB-D sensor of RealSense, the RS410, as the waist-worn pathfinder to provide early warning. Indoor and outdoor detection results as well as a field test demonstrate the usefulness and effectiveness of our approach.

CCS Concepts
• **Human-centered computing → Ubiquitous and mobile computing → Ubiquitous and mobile computing theory, concepts and paradigms→Ambient intelligence**

Keywords
Traversability awareness, obstacle avoidance, visually impaired, RGB-D, RealSense.

1. INTRODUCTION
According to the World Health Organization, 285 million people are estimated to be visually impaired and 39 million are blind around the world [1]. Compared with the white cane, the guide dog, and the adaptive mobility device (AMD), RGB-D sensor can provide more information to the visually impaired. Thereby, the research community has been spurred to apply RGB-D sensors in wearable assistive systems for prosthetic vision [2], head-mounted glasses [3-8], chest-attached cameras [9-11] and waist-worn prototypes [12]. These approaches based on RGB-D sensors, including light-coding sensors and stereo cameras, have provided proof-of-concepts to help the visually impaired to avoid obstacles and navigate traversable areas. However, it is critical that the limitations of RGB-D sensors be highlighted and pinpointed to garner essential information for the sensor combination [13]. For instance, RGB-D sensor fails to deliver dense 3D information at long distances and low-lying obstacles are not easily detected within limited field of view.

Light-coding sensors, such as Microsoft Kinect and Asus Xtion, consist of an IR laser projector which emits structured near-IR patterns of speckles into the scene and an IR image sensor captures the speckles. The distortions of speckles are deciphered and the depth map is calculated through triangulation algorithms. However, the speckles at long distances are too dark to be sensed with limited projecting power. For this reason, approaches with light-coding sensors were feasible to assist navigation in indoor environments without direct sunlight. Zöllner et al. [3] used the Kinect to help visually impaired people to find their pathway while optical markers were tracked within the buildings. Lee and Medioni [4] incorporated traversability analysis and visual odometry to help in steering away from obstacles during dynamic path planning. Hicks et al. [5] built a real-time display by highlighting short-range objects with Xtion to improve situational awareness for the partially sighted. Aladrén et al. [10] combined range information with image intensities to extend the depth image based ground segmentation. Cheng et al. [14] put forward an algorithm to detect ground with Kinect based on seeded region growing and generated a pathway touch image.

Stereo cameras, such as PointGrey Bumblebee and Stereolabs ZED, estimate depth map through stereo matching of images from two lenses. Points on one image are correlated to another image and depth is calculated via disparity, which is the shift between a point on one image and another image. However, since depth error proportionally increases with the increase of the depth, stereo cameras are prone to be unreliable in the distance. In this sense, approaches with stereo cameras are mainly focused on highly-textured outdoor environments. Martinez and Ruiz [15] presented an approach for aerial obstacle detection with a Bumblebee stereo camera, which introduced a stabilization algorithm to maintain the floor of the local map continuously aligned with the horizontal plane. Rodríguez et al. [9] estimated ground plane based on random sampling and filtering techniques, and used a polar grid representation to account for the potential obstacles with the commercial Bumblebee camera. Miksik et al. [6] presented an augmented reality system which comprises a pair of optical see-through glasses with RGB-IR stereo pairs and a handheld laser pointer, which allows the partially sighted to witness real-time 3D reconstructions, while using the pointer to draw onto real-world scenarios to interactively refine the depth map and semantically segment objects of interest. Wang et al. [11] introduced a system with a ZED stereo camera and built on the work of Stixel World [16] for free space parsing. Rizzo et al. [13] built an assistive sensor fusion platform by combining a stereoscopic camera and a solid-state LiDAR in order to remedy the inconsistency of depth information for object detection.

In this paper, a sensor combination framework to assist visually impaired people in terms of traversability awareness and obstacle negotiation is presented. Rather than sparse original depth maps, dense 3D point clouds are acquired by optimizing the preset configuration of RealSense R200 in the pair of wearable smart glasses. Small segments are removed from the depth map before generating traversable lines with real-world coordinate adjustment and height segmentation by utilizing attitude angles acquired with the IMU sensor. By incorporating RealSense RS410 as the waist-worn pathfinder, we segment the low-lying obstacles from traversable areas under the Manhattan World [17] assumption.

The paper is organized as follows. In Section 2, an overview about the assisting system is introduced including the 3D information acquisition and the device for human-computer interaction. In Section 3, the approach designed for long-range traversability awareness and low-lying obstacle negotiation is described. In Section 4, both indoor and outdoor detection results as well as a field test are presented and discussed. Finally, in Section 5, conclusions are drawn and future work is expected.

2. SYSTEM OVERVIEW

As depicted in Figure 1, the system architecture is composed of a pair of wearable smart glasses, a waist-worn pathfinder, a portable processor and a bone-conduction headset. There are two processing threads, which continuously receive images from the sensors and complete the detection at different framerates. The first processing thread receives depth images from RealSense R200, which is quite suitable for integrating in a pair of wearable glasses for navigational assistance thanks not only to its environmental adaptability, but also its small size and light weight [7]. This thread adjusts the 3D point cloud coordinates and segments the ground areas strictly with height information. Thereby, this thread is able to generate the non-semantic stereo sound feedback in real time as it is computationally efficient. The non-semantic stereo sound feedback is faster (24Hz) but the semantic voice warnings provide more explicit information for the visually impaired. Comparably, the second thread receives depth images from the waist-worn RealSense RS410, which combines active speckle projecting and passive stereo matching for dense 3D reconstruction in both indoor and outdoor environments. In addition, the pathfinder is tilted and closer to the floor than the pair of glasses. In this regard, we could estimate the normal vector direction of the floor under the Manhattan World assumption. After that, the low-lying obstacles are segmented from the traversable areas. This thread is not running in real time (7Hz), but it will produce early warnings in advance to assist the visually impaired in terms of low obstacle avoidance.

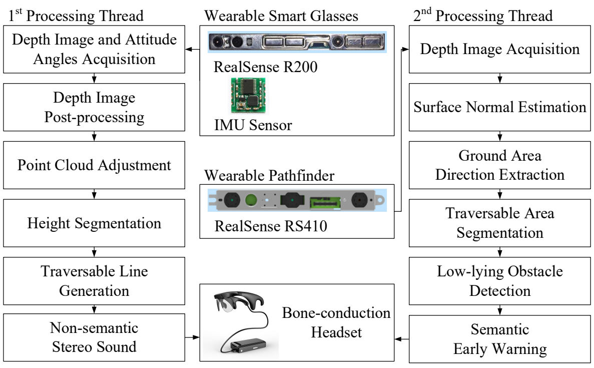


Figure 1. System architecture

As shown in Figure 2, the pair of wearable smart glasses is integrated with the RGB-D sensor RealSense R200, the IMU sensor MPU6050 and the bone-conduction headphone. The waist-worn pathfinder is composed of a RGB-D sensor RealSense RS410. The bone-conduction headphone transfers sound from both processing threads to the visually impaired. As anybody knows, visually impaired people use the surrounding environmental sounds to orientate themselves. Thereby, the prototype is not only wearable but also ears-free, because the bone-conducting interface would not block the ears of the users from hearing environmental sounds.

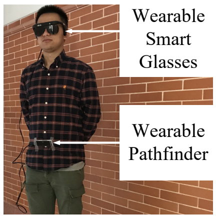


Figure 2. The wearable prototype

3. APPROACH

3.1 Long-range traversability

In order to provide long-range traversability awareness for the visually impaired, we use a different preset configuration with respect to the original depth image of R200. Since the stereo sensor generally requires good trading off between density and accuracy, a dense depth map is preferred to assist the visually impaired so as not to leave out potential obstacles. As shown in Figure 3(b), the depth information of the floor is sparse a few meters away, which hampers longer traversable area awareness. Comparably, by controlling how aggressive the algorithm is at discarding matches, the dense depth maps are acquired as shown in Figure 3. Since a lot of noises and mismatched pixels exist in the depth image, we use a fast and effective method for de-nosing by eliminating the small segments in the depth image which was previously presented in [7].

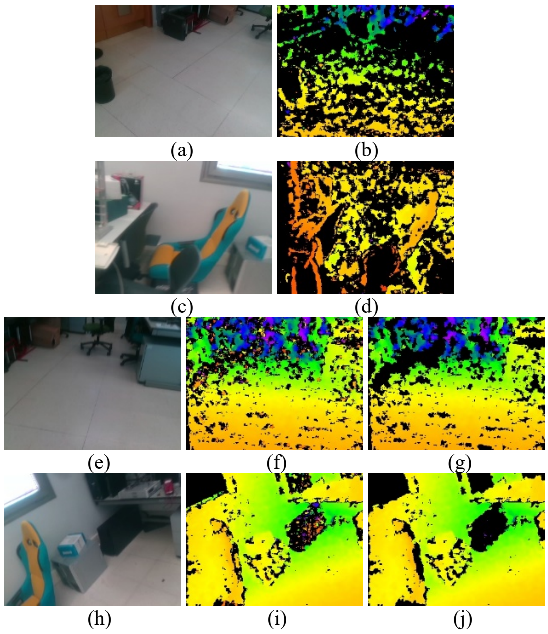


Figure 3. Color images and depth maps from the wearable smart glasses: (a)(c)(e)(h)Color images; (b)(d) Original depth maps; (f)(i) Depth maps with dense preset configuration; (g)(j) Denoised depth maps.

With the help of the attitude sensor, X, Y and Z coordinates in the camera coordinate system are adjusted to world coordinates. Assuming that a point in the camera coordinate system is (X,Y,Z) and the attitude angles acquired from the attitude sensor are (a,b,c), the point (X,Y,Z) rotates about the x-axis by $\alpha = a$, rotates about the y-axis by $\beta = b$ and rotates about z-axis by γ =

c . Multiplying the point (X,Y,Z) by the rotation matrix given by (3), the point (X_w,Y_w,Z_w) in world coordinates is obtained.

Aiming to perceive traversability and generate the stereo sound, a simple but effective technique is proposed. First, 3D coordinates of the point cloud are calculated. Given the depth Z of a pixel (u,v) in the depth image, the calibrated focal length f , the principal point (u_0,v_0) , the 3D point (X,Y,Z) in the camera coordinate system can be derived by (1)(2).

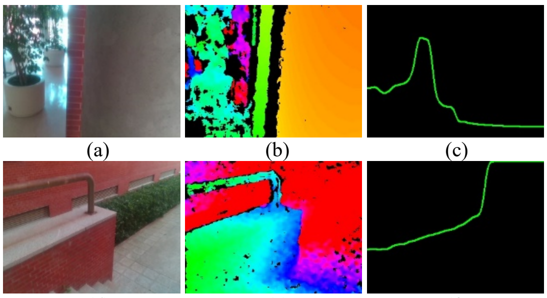
$$X=(u-u_0)/f\times Z \quad (1)$$

$$Y=(v-v_0)/f\times Z \quad (2)$$

With the help of the attitude sensor, X , Y and Z coordinates in the camera coordinate system are adjusted to world coordinates. Assuming that a point in the camera coordinate system is (X,Y,Z) and the attitude angles acquired from the attitude sensor are (a,b,c) , the point (X,Y,Z) rotates about the x-axis by $\alpha=a$, rotates about the y-axis by $\beta=b$ and rotates about z-axis by $\gamma=c$. Multiplying the point (X,Y,Z) by the rotation matrix given by (3), the point (X_w,Y_w,Z_w) in world coordinates is obtained.

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} X \\ Y \\ W \end{bmatrix} \quad (3)$$

After the point cloud adjustment, the height segmentation is implemented to remove the ground area from the depth image by strictly eliminating the 3D points whose vertical coordinate exceeds the height threshold, which is set according to the user height. A traversable line to represent the traversable distances in different directions is proposed for generating the stereo sound feedback. For each image column, we locate the nearest traversable pixel with the minimum depth value and set the traversable distance on the line. A Gaussian filter is applied to smooth this line to reduce noise before generating the feedback to the visually impaired. After the filtering stage, the green traversable lines are obtained as shown in Figure 4.



(a) (b) (c)

(d) (e) (f)

Figure 4. (a)(d) Color images; (b)(e) Depth maps; (c)(f) Traversable lines.

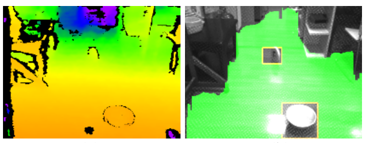
This paper uses the non-semantic audio interface presented in our previous work [7-8] to transfer the traversable directions to the visually impaired by synthesizing stereo sound from the traversable line. The directions of traversable areas are differentiated not only by the different sound sources associated to them in the virtual 3D space and the directions of stereo sound, but also by the musical instruments, whose timbre differs from each other. The generation of the stereo sound follows the rules below to guide the user to take the prioritized direction to detour around hazardous regions:

- Divide the traversable line into five sections which correspond to the five different musical instruments. We only use five instruments to make sure that it is easy to understand and would not sound confusing. Five instruments, including trumpet, piano, gong, violin and xylophone, produce sounds simultaneously.
- Each direction of traversability is represented by a musical instrument in the virtual 3D space. For each musical instrument, the higher the average of height is in the corresponding section of the traversable line, the louder the sound from the instrument will be. At the end, the stereo sound would guide the user to navigate the traversable direction.

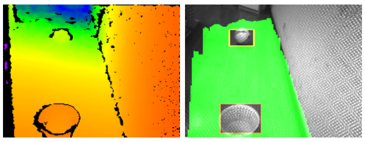
3.2 Low-lying obstacle negotiation

Since the processing thread of the long-range traversability awareness removes the ground area using a strict height segmentation to get real-time feedback, it is not sensitive to the low-lying obstacles on the ground. In addition, as the vertical field angle of the sensor is limited, small objects are not easily detected by the wearable glasses.

Accordingly, this paper aims to integrate a waist-worn pathfinder to better assist the visually impaired in terms of low obstacle avoidance. As the RS410 is closer to the ground and it is pointed toward the forward path, it could be assumed that the ground area covers a large portion of the depth image, which is a variant of the Manhattan-world assumption. In the first place, the surface normal vectors of an image grid are estimated using a least square method. As the ambiguity on the sign of the normal vector is not analytically solvable, each image block has two normal vectors. Conventionally, only the normal vector pointing upward is considered. Thereupon, a histogram of normal directions over a unit hemisphere is computed, since the dominant axe rather than signed directions is extracted in this step. Here, the histogram is subdivided into 1000 bins and the original depth image with the resolution 640×480 is segmented into 64×48 image blocks. Based on this assumption, the normal vector direction of the floor is set with the average of the normal vectors within the largest bin of the histogram. Subsequently, with the point cloud in the largest bin and the normal vector direction of the floor, we could determine the attitude angle of the waist-worn sensor and the height difference between the sensor and the floor. To warn against collisions, we not only implement the method presented in our previous work [8] to segment the first visible relevant obstacle in the positive direction of depth and traversable area using dynamic programming, a Mean-Shift algorithm is also introduced to produce bounding windows, as shown in Figure 5, to detect the low-lying obstacles with distinctively different normal vectors and aggregated depth pixels. Here, we use the cosine to measure the difference between normal vectors of image bins and the ground area. Thereupon, semantic early warning such as “please mind your step” is generated.



(a) (b)



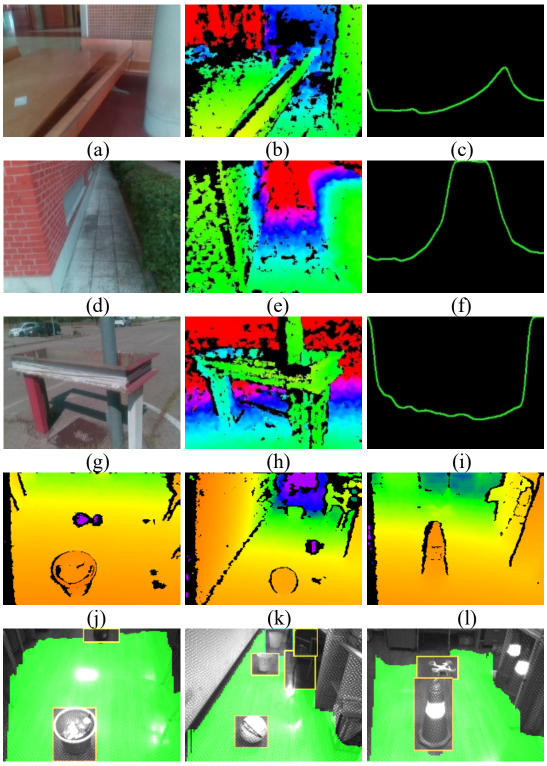
(c) (d)

Figure 5. Low-lying obstacle detection and traversable area segmentation: (a)(c) Depth maps in fake color; (b)(d) Low-lying obstacle detection results by producing bounding windows and traversable area segmentation results by producing the green mask to overlay on the IR images.

4. EXPERIMENTS

In this section, the algorithm is verified with detection results tested on indoor and outdoor scenarios in public spaces around the Yuquan Campus at Zhejiang University and the Polytechnic School at University of Alcalá. The performance of the proposed approach is also evaluated with a field test in the real-world environment.

Figure 6 shows a number of traversability awareness and obstacle detection results in indoor and outdoor environments. As depicted in Figure 6(a-i), the approach correctly generates traversable lines, which would suggest the prioritized directions to guide the visually impaired to avoid close obstacles and lead the users to navigate the traversable directions with the stereo sound. As shown in Figure 6(j-p), the low-lying obstacles which would impede the safety of navigation are correctly segmented from the floor including garbage cans, chairs, football, traffic cones and unmanned aerial vehicle.



(m) (n) (p)

Figure 6. Detection results: (a)(d)(g) Color images of scenarios in both indoor and outdoor environments; (b)(e)(h)(j-l) Depth maps acquired with the smart glasses and the waist-worn pathfinder; (c)(f)(i)Traversable lines; (m-p) Traversable area and low-lying obstacles segmentation results.

We conduct a field test in a corridor about 2.5 wide and 38m deep with many umbrellas popped as shown in Figure 7. This field test is organized by KR-VISION [18], which involved 10 visually impaired volunteers, who have experienced the learning stage to adapt to the prototype. It is worth mentioning that a portable PC with Intel Atom x5 Z8500 processor and 2GB memory is chosen as the computing platform. As a comparison task, the volunteers are required to walk from one end of the corridor to the other one with the stereo sound feedback of traversability awareness first and then complete the route again with the combination of the stereo sound and the warning of low-lying obstacles. Traversing time and collision times are recorded for each participant. The timer starts when a participant was sent to the start region and stops when the participant completed the route. The collisions include collisions with obstacles and walls. As shown in Figure 7, we can see that when navigating with feedback from both traversability awareness and low-lying obstacle detection, not only collision times but also the traversing time are significantly less than without the warning of low obstacles. It is the detection of traversable areas and obstacles that endows participants to take the prioritized navigable direction and pay attention to the potential hazards near the feet. The results suggest the safety and robustness of the navigation has been enhanced with the combination of long-range traversability awareness and low-lying obstacle negotiation.

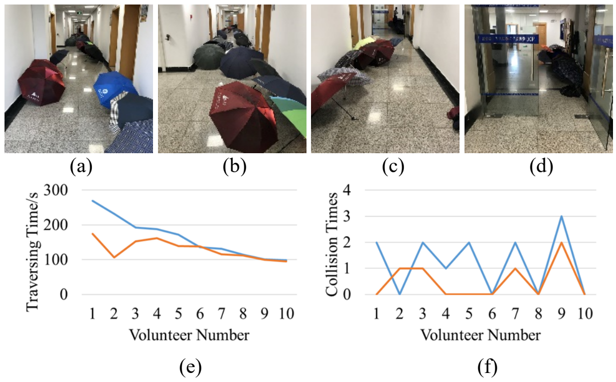


Figure 7. Field test: (a-d) test scenario; (e)(f) experiment data sorted in descending orders of the traversing time while the blue lines are collected with traversability awareness and the orange lines are collected with traversability awareness and low-lying obstacle detection.

5. CONCLUSIONS AND FUTURE WORK

RGB-D sensors are a ubiquitous choice to assist the navigation for the visually impaired. However, most solutions are confined to the sparse depth information within limited field of view. The main contribution of this paper is the implementation of a wearable multi-thread assistance system with sensor combination, which enhances the safety and robustness of navigation for the visually impaired. The novelty of our method is in the description of RGB-D images as traversable lines and in the segmentation of the obstacles in two levels: floor, visible obstacles and low-lying ones, which are transferred to stereo sound and semantic warnings. Indoor and outdoor empirical evidences as well as a field test demonstrated its usefulness and reliability.

In future works, we aim to improve our navigation assistance approach to reach a higher level of perception and offer more independence to the visually impaired. Specifically, we look forward to including real-time semantic segmentation [19] to unify the terrain awareness.

6. ACKNOWLEDGMENTS

This work has been partially funded by the Zhejiang Provincial Public Fund through the project of visual assistance technology for the blind based on 3D terrain sensor (No. 2016C33136) and cofunded by State Key Laboratory of Modern Optical Instrumentation.

This work has been partially funded by the Spanish MINECO through the SmartElderlyCar project (TRA2015-70501-C2-1-R) and from the RoboCity2030-III-CM project (Robótica aplicada a la mejora de la calidad de vida de los ciudadanos, fase III; S2013/MIT-2748), funded by Programas de actividades I+D (CAM) and cofunded by EU Structural Funds.

7. REFERENCES

[1] World Health Organization. 2014. Visual impairment and blindness. http://www.who.int/mediacentre/factsheets/fs282/en/.

[2] Li, W. H. 2013. Wearable computer vision systems for a cortical visual prosthesis. In Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on (pp. 428-435). IEEE.

[3] Zöllner, M., Huber, S., Jetter, H. C., and Reiterer, H. 2011. NAVI–a proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect. In IFIP Conference on Human-Computer Interaction (pp. 584-587). Springer, Berlin, Heidelberg.

[4] Lee, Y. H., and Medioni, G. 2011. RGB-D camera based navigation for the visually impaired. In Proceedings of the RSS.

[5] Hicks, S. L., Wilson, I., Muhammed, L., Worsfold, J., Downes, S. M., and Kennard, C. 2013. A depth-based head-mounted visual display to aid navigation in partially sighted individuals. PloS one, 8(7), e67695.

[6] Miksik, O., Vineet, V., Lidegaard, M., Prasaath, R., Nießner, M., Golodetz, S., Hicks, S. L., Pérez, P., Izadi, S., and Torr, P. H. S. 2015. The semantic paintbrush: Interactive 3d mapping and recognition in large outdoor spaces. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3317-3326). ACM.

[7] Yang, K., Wang, K., Hu, W., and Bai, J. 2016. Expanding the detection of traversable area with RealSense for the visually impaired. Sensors, 16(11), 1954.

[8] Yang, K., Wang, K., Cheng, R., Hu, W., Huang, X., and Bai, J. 2017. Detecting traversable area and water hazards for the visually impaired with a pRGB-D sensor. Sensors, 17(8), 1890.

[9] Rodríguez, A., Yebes, J. J., Alcantarilla, P. F., Bergasa, L. M., Almazán, J., and Cela, A. 2012. Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. Sensors, 12(12), 17476-17496.

[10] Aladren, A., López-Nicolás, G., Puig, L., and Guerrero, J. J. 2016. Navigation assistance for the visually impaired using RGB-D sensor with range expansion. IEEE Systems Journal, 10(3), 922-932.

[11] Wang, H. C., Katzschmann, R. K., Teng, S., Araki, B., Giarré, L., and Rus, D. 2017. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In Robotics and Automation (ICRA), 2017 IEEE International Conference on (pp. 6533-6540). IEEE.

[12] Yang, K., Wang, K., Cheng, R., and Zhu, X. 2016. A new approach of point cloud processing and scene segmentation for guiding the visually impaired. Iet International Conference on Biomedical Image and Signal Processing (pp.1-6). IET.

[13] Rizzo, J. R., Pan, Y., Hudson, T., Wong, E. K., and Fang, Y. 2017. Sensor fusion for ecologically valid obstacle identification: Building a comprehensive assistive technology platform for the visually impaired. In Modeling, Simulation, and Applied Optimization (ICMSAO), 2017 7th International Conference on (pp. 1-5). IEEE.

[14] Cheng, R., Wang, K., Yang, K., and Zhao, X. 2016. A ground and obstacle detection algorithm for the visually impaired. Iet International Conference on Biomedical Image and Signal Processing (pp.1-6). IET.

[15] Martinez, J. M. S., and Ruiz, F. E. 2008. Stereo-based aerial obstacle detection for the visually impaired. In Workshop on Computer Vision Applications for the Visually Impaired.

[16] Badino, H., Franke, U., and Pfeiffer, D. 2009. The stixel world-a compact medium level representation of the 3d-world. In Joint Pattern Recognition Symposium (pp. 51-60). Springer, Berlin, Heidelberg.

[17] Furukawa, Y., Curless, B., Seitz, S. M., and Szeliski, R. 2009. Manhattan-world stereo. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 1422-1429). IEEE.

[18] KR-Vision. 2016. To tackle the challenges for the visually impaired. http://krvision.cn/.

[19] Romera, E., Alvarez, J. M., Bergasa, L. M., and Arroyo, R. 2017. Efficient convnet for real-time semantic segmentation. In Intelligent Vehicles Symposium (IV), 2017 IEEE (pp. 1789-1794). IEEE.