Glass detection and recognition based on the fusion of ultrasonic sensor and RGB-D sensor for the visually impaired

Zhiming Huang[a], Kaiwei Wang*[a], Kailun Yang[a], Ruiqi Cheng[a] and Jian Bai[a]
[a]College of Optical Science and Engineering, Zhejiang University, 38# Zheda Road, Hangzhou, Zhejiang, China 310007

ABSTRACT

With the increasing demands of visually impaired people, developing assistive technology to help them travel effectively and safely has been a research hotspot. Red, Green, Blue and Depth (RGB-D) sensor has been widely used to help visually impaired people, but the detection and recognition of glass objects is still a challenge, considering the depth information of glass cannot be obtained correctly. In order to overcome the limitation, we put forward a method to detect glass objects in natural indoor scenes in this paper, which is based on the fusion of ultrasonic sensor and RGB-D sensor on a wearable prototype. Meanwhile, the erroneous depth map of glass object computed by the RGB-D sensor could also be densely recovered. In addition, under some special circumstances, such as facing a mirror or an obstacle within the minimum detectable range of the RGB-D sensor, we use a similar processing method to regain depth information in the invalid area of the original depth map. The experimental results show that the detection range and precision of the RGB-D sensor have been significantly improved with the aid of ultrasonic sensor. The proposed method is proved to be able to detect and recognize common glass obstacles for visually impaired people in real time, which is suitable for real-world indoor navigation assistance.

Keywords: Glass detection and recognition, RGB-D sensor, ultrasonic sensor, sensor fusion, navigation assistance for the visually impaired

1. INTRODUCTION

According to the statistics from the World Health Organization in 2017, around 253 million people live with vision impairment worldwide, of whom about 36 million are blind[1]. Due to some disability to fully perceive the surroundings, visually impaired people face various inconvenience in their daily life. The traditional assistive methods, such as the white cane or guide dog, have inherent limitations during independent navigation. The white cane is the most common tool among the community of visually impaired people, while it only helps to avoid obstacles through physical perception. The trained guide dogs could convey many key information to its owner quickly, but both the expensive price and long-term training cycle hinder the widespread use of guide dogs. It is obvious that neither of the two assistive tools has sufficiently met the actual needs of visually impaired people currently. Therefore, the development of targeted assistive technology becomes especially important.

With the continuous progress of computer vision and sensor technology nowadays, there have been many achievements based on RGB-D sensor to detect pedestrians, traversable areas, traffic signs and other helpful information for visually impaired people[2-5]. RGB-D sensor is known as the visual sensor that could provide not only RGB but also depth information of frames simultaneously. However, it is unable to obtain the correct depth information of glass, because glass materials feature high transmittance and non-texture[6]. Generally, hazardous glass obstacles could not be detected by most of current RGB-D sensor-based wearable systems. Unfortunately, glass materials are ubiquitous in daily life, hence it is very dangerous for visually impaired people. To fill the gap, this paper mainly focuses on the development of glass detection technology for real-world indoor navigation assistance.

* Corresponding author. E-mail: wangkaiwei@zju.edu.cn

As far as we know, there are two popular methods to detect glass in the recent literatures. The first method is to utilize the polarization property of the light by glass. According to the Fresnel formulae[7], the light reflected by the surface of a mirrored smooth object is polarized, which is different from the diffuse reflection on the surface of a rough object. Based on such a difference, researchers use four images captured at four different polarization angles to compute the degree of linear polarization image and angle of linear polarization image[8]. After that, through the combination of color map and depth map, the regions with high degree of polarization usually have the higher probability to be glass objects. However, the polarization camera that acquires four images of different polarization angles simultaneously is very expensive, which limits its wide application in the visually impaired community. In our previous work, we have proposed a low-cost method by using a pRGB-D sensor, which includes a depth camera and two orthogonal polarizers, to efficiently detect water hazards for visually impaired people[9]. However, this method does not achieve satisfactory results on glass detection, because unlike the reflected light from the water surface, the transmitted light of glass tends to be less pronounced in polarization.

The second method is based on sensor fusion, which is also widely used in computer vision. Ye et al.[10] proposed a method to reconstruct 3D scenes that contain piece-wise planar transparent objects by integrating a sonar with the Microsoft Kinect camera. They achieved better results on various real-world scenes, but their system is unable to run in real time and lacks the precise depth information of glass objects. Yang et al.[11] introduced a sensor fusion simultaneous localization and mapping (SLAM) method based on the data from sixteen sonar sensors and a laser scanner equipped on a robot. However, their main contribution is to propose the mirror tracking framework to avoid obstacles for robot, not for people. Certainly, there are other methods to detect glass objects in the research line of 3D reconstruction[12,13], but all of them are not suitable for real-time assistive technology for visually impaired people, which represents a challenging and so far largely unexplored topic.

In this paper, we propose a method based on the fusion of ultrasonic sensor and RGB-D sensor to detect glass objects for visually impaired people. The designed device is a 3D printed frame with a depth camera and two ultrasonic sensors attached to it. The detection system will judge whether there are glass obstacles in each frame after the analysis of the depth data from ultrasonic sensors and depth camera. Subsequently, we use the region growing algorithm to determine the location of glass and the correct depth value to recover the sparse depth map. Besides, our device is also able to recover the invalid area of the sparse depth map under some special circumstances, such as facing a mirror or an obstacle within the minimum detectable range of the depth camera. The experimental results show that the method accurately detects the location and distance of glass object in front of the wearer. And compared with the glass detection methods described in the previous literatures, our method performs better in accuracy and has wider applied scenarios in real time. What is more, this proved that the method could be applied in real-world indoor navigation assistance for visually impaired people.

The rest of the paper is organized as follows: Section 2 presents our device design and data preprocessing. Section 3 introduces the detailed principle of our method and Section 4 presents the experiment results and analysis. Conclusions of this work are given and future research is expected in Section 5.

2. SYSTEM OVERVIEW

As it shown in Figure 1, we present our wearable device that incorporates an Intel RealSense R200 depth camera[14] and two ultrasonic range sensors HC-SR04 in this section.

The depth camera Intel RealSense R200 includes two infrared cameras distributed at the left and right side respectively, a color camera, an infrared laser projector, a module connector and an imaging application specific integrated circuit (ASIC). The depth camera can acquire both color and infrared data stream simultaneously and the original depth map is computed from the disparity (pixel shift) between the two separated infrared cameras by triangulation with the help of the laser projector. As reported by the official researchers[14], the accurate detection distance is about from 500mm to 3500mm indoor and can reach to about 10000mm outdoors.

The ultrasonic distance measuring module HC-SR04, which includes an ultrasonic transmitter, an ultrasonic receiver and a control circuit, is operating at 40 kHz and emits a narrow beam whose width is less than 15 degrees. Notably, the detection range is about from 20mm to 4500mm, which is enough for visually impaired people to find and avoid obstacles. As auxiliary sensors, the primary role of them is to provide correct depth information of the objects invisible to RGB-D sensor, such as glass windows and mirrors.

Although there have already been lots of research and applications on the fusion technology of ultrasonic sensor and RGB-D sensor, we have applied it to detect glass objects for visually impaired people for the first time. A simple wearable frame made by 3D-printing technology is used to fix the two sensors together. The two ultrasonic sensors, which could output reliable depth values in real time, are set up on the left and right side on the depth camera respectively to detect corresponding regions and the depth camera outputs color and depth data for processing. They all connect to a computer through USB port to transmit data. The more details about how the system works will be discussed in section 3.

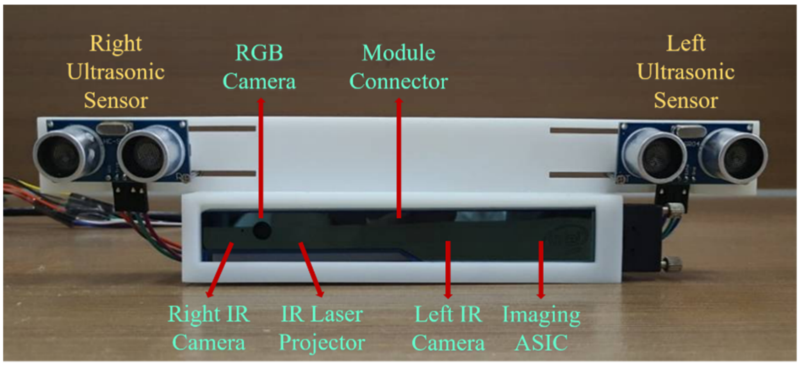


Figure 1. An Intel RealSense R200 depth camera and two ultrasonic sensors HC-SR04 that are installed in the printed wearable frame constitute our prototype.

3. METHODOLOGY

3.1 Transparent glass detection

As large transparent object, for example the glass door, sometimes it is even difficult for human eyesight to distinguish whether the door is open or not without specific references like doorknob. When computers run into the problem, their "eyes" are usually traditional image sensors, such as CCD and CMOS, which are difficult to see these "invisible" glass objects. Therefore, we determine to integrate ultrasonic sensors to help computers perceive glass like human hands to understand the surrounding environment, since ultrasonic wave would not spread through glass material completely or be largely absorbed like light wave.

The propagation path of ultrasonic wave emitted by HC-SR04 can be approximately viewed as a wide ray beam, because of its high frequency and good directivity. For this reason, we could find out the corresponding detection areas of the left and the right ultrasonic sensors on the depth map respectively by simple yet effective calibration. This calibration method greatly simplifies the processing procedure and is proved to be effective by the detection results.

In our method, we divide the depth map into right and left halves and analyze different depth data of the two sides simultaneously in every frame of the camera. The left and the right parts are considered separately, but the detection results will be fused as the final output. To make the following explanations clear, we adopt the left side to describe our method, since the right side follows just the same situation. During the data acquisition procedure, the left ultrasonic depth value (*LUD*) that is the average of 5 continuous valid ultrasonic data and the left camera depth value (*LCD*) that is the average depth value of the left calibration area, in which the 0 and outliers value have been removed, are computed in every frame. Afterwards, the system will make a judgement that whether there is possible glass object in the left side depending on the following three main judging conditions.

(1) Whether the value of *LCD* minus *LUD* is greater than $\delta 1$.

(2) Whether *LCD* is greater than $\delta 2$.

(3) Whether *LUD* is between 100mm and 4500mm.

Where δi (i=1,2) is two thresholds defined by experimental experience. In actual test, we set $\delta 1$, $\delta 2$ to 700mm, 2000mm respectively. If an affirmative judgment is given by the system in the frame, a variable called *GLASS* (The default value is 0) that represents the confidence level of the existence of glass will be added one, otherwise it will be reduced by one.

Once the variable *GLASS* reaches more than 3 in one frame, the system will determine the presence of glass object and output relevant detection information on the color map. To prevent the confidence level from becoming too large or too small on the subsequent frame, we limit the variable *GLASS* between 0 and 5. The specific detection steps are shown as the flow chart in Figure 2.

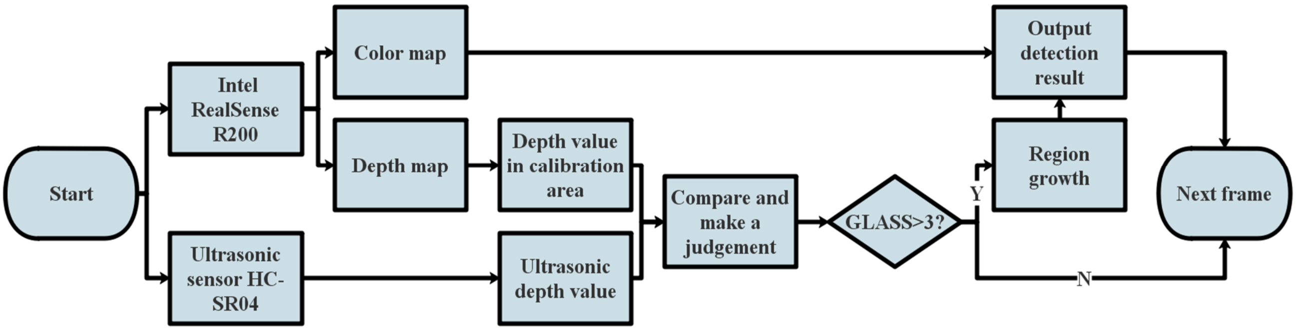
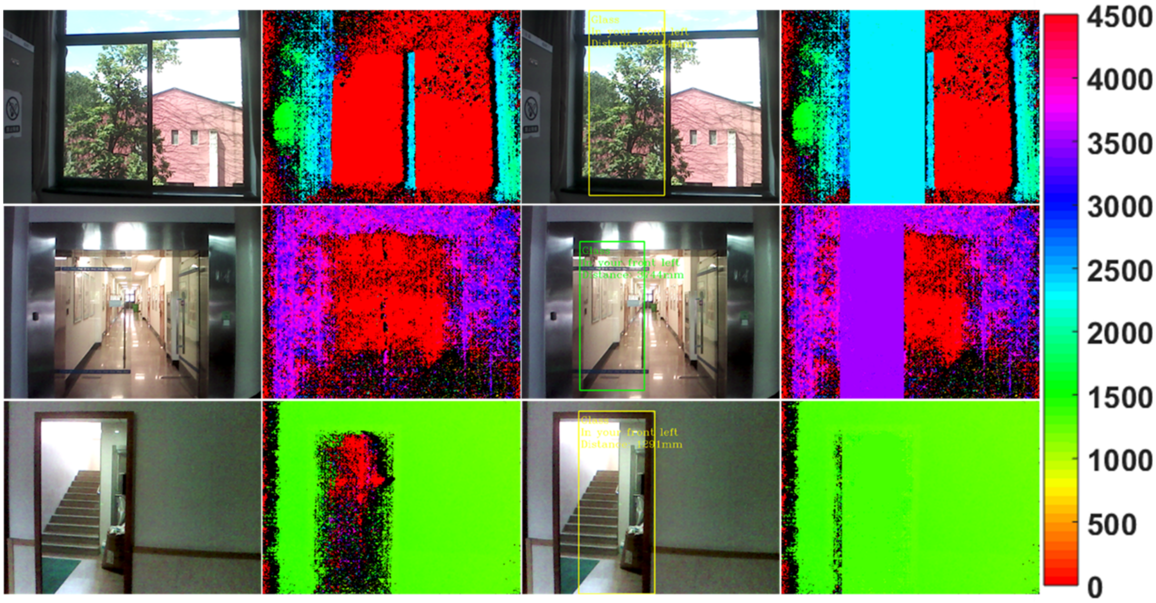


Figure 2. The detection flow chart in one side of our system.

Subsequently, the system enters the depth map recovery phase, where the region growing algorithm is used in the depth map. The seed point is selected randomly in the calibration area, but its depth value must be approximate to *LCD* to guarantee qualified growth effect. In addition, it is worth noting that when the door opens only one side, we find that the growing region sometimes would go into another side without glass by error. Therefore, we use two masks to control the left and the right growing region respectively. Additionally, in the process of growth, the location of the edge pixel in the growing region would be recorded one by one. Finally, the value *LUD* is used to replace the wrong and invalid pixel value in the recorded locations. At this point, our system would successfully detect the location of the front transparent glass and output a correct sparse depth map. As it shown in Figure 3, in order to display the depth map more intuitively, pseudo color depth map is used here (*Black* represents invalid depth value). From left to right, the images of Figure 3 are the color map, original depth map, detection result and recovered depth map, respectively. The detailed introduction of the output form will be discussed in 3.3.



(a) (b) (c) (d)

Figure 3. Glass detection and recognition process. (a): The color map. (b): The original depth map. (c): The detection result. (d): The recovered depth map.

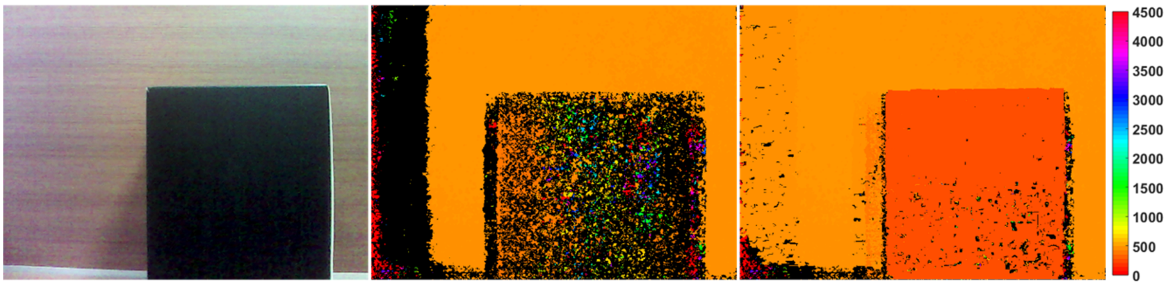
3.2 Extended functionalities

In addition to finding transparent glass objects, we also use our device to do some extension works, including detecting mirror and recovering the depth information of the object in the blind area, which is the minimum detectable range of the depth camera.

Mirrors are common in our daily life, while related algorithms about detecting mirrors in the field of computer vision are not so common. Although mirrors are also made by glass material, the situation is a little different from transparent glass, since its total reflection. As a result, the depth value computed by stereo match algorithm is usually about two times as much as the actual value, which is similar to the subjective feeling when we look at a mirror. The mirror detection method employs the fact that mirrors are usually framed. In other word, the boundaries of a mirror are visible to the depth camera, which makes mirror area may seem like a hole from the depth map. This is basically the same as the case of transparent glass. Therefore, the main detection steps we describe above is still applicable here.

The depth map recovery phase is also similar to the pervious description in Section 3.1, so it does not be described repeatedly here. However, it is noteworthy that all the pixel points meet the above conditions have to be bracketed together before the region growing algorithm. Only in this way could the complete mirror area be detected by growing algorithm. The result of mirror detection could be seen in the third line of Figure 3.

The blind area of the depth camera, which is related to the baseline, usually appears at close quarters. Through our actual test on the Intel RealSense R200 depth camera, we find that some wrong depth values begin to appear at about 500mm and when the object is within 300mm from the depth camera, the depth map will become almost invalid as shown in Figure 4 (b). However, the shortest detectable distance of the ultrasonic sensor is about 20mm, which means that our system is able to detect the depth of the object in blind area. Besides, due to the field of view captured by the depth camera in blind area is very small, the objects usually only need to be divided into one or two categories as Figure 4 (a) shows. Thus we only need to use the region growing algorithm to make an simple object classification after median filtering on the color map. According to the result of the classification, the wrong depth of the object in blind area would be replaced by the valid data measured by ultrasonic sensors. Figure 4 (c) shows the recovered depth map of the object in blind area.



(a) (b) (c)

Figure 4. The depth map recovered process of the object in blind area. (a): The color map including two category objects. (b): The depth map including an invalid area with erroneous measurements. (c): The recovered depth map.

3.3 Output form

All the output information will be displayed on the color map after the detection program in every frame and Figure 5 shows an output example of the detection result.

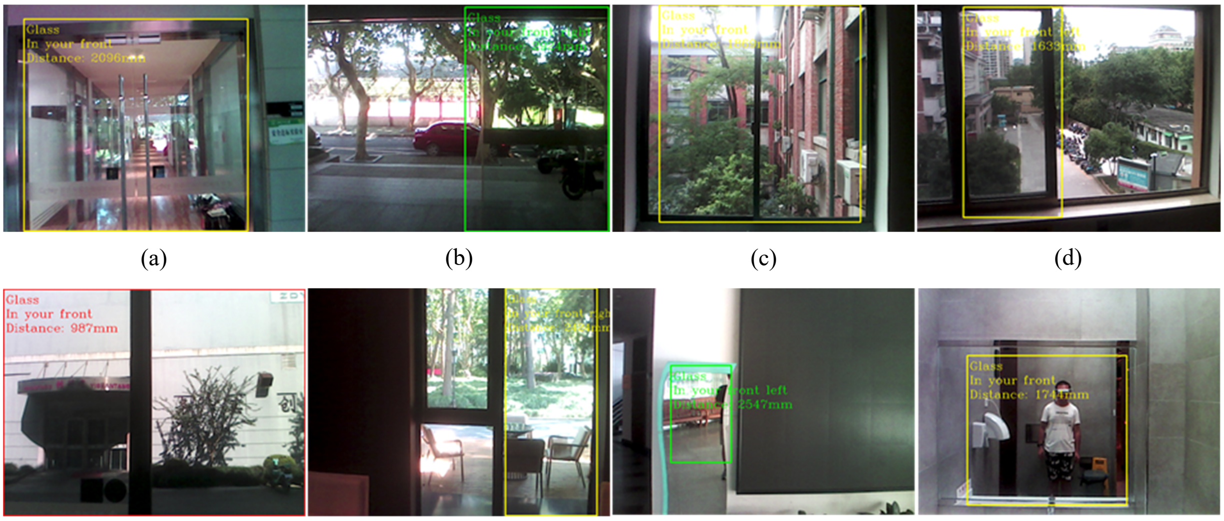
All relevant information, such as the object description, location and distance, will be shown in the color map of the Intel RealSense R200 depth camera. The object description and distance information are displayed with a rectangular box, which represents the location of the glass object. Herein, due to the valid detection distance is about from 500mm to 4500mm, we define three distance regions: warning distance (within1000mm), obstacle avoidance distance (from 1000mm to 2500mm) and safe distance (beyond 2500mm). In different regions, there are different output colors. Similar to the rules of the traffic lights, we use *red* to represent the warning distance, *yellow* to represent the obstacle avoidance distance and *green* to represent the safe distance. Moreover, based on different results on the left and right sides, the location information also includes three different direction descriptions: front right, front left and front ahead (both sides). These direction descriptions are useful for visually impaired people to avoid obstacle accurately. For the mirror detection, the output form is the same as above description.

Figure 5. Detection result of a representative scenario. The output information includes object description, location and distance.

4. EXPERIMENT

In this section, we first apply the method to real-world indoor scenarios to represent our results on glass detection and recognition, then the rate of false detection (*RFD*) and the rate of missed detection (*RMD*) of the proposed method are assessed for its effectiveness.

4.1 Results Presentation



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 6. Detection results in different indoor scenarios and conditions. (a): A closed glass door. (b): A half-open glass door. (c): A closed glass window. (d): A half-open glass window. (e): A closed glass door within warning distance. (f): A floor window with some chairs behind. (g): A mirror hanging on the wall. (h): A mirror in front of a man.

To test the detection effectiveness of our system, we carried out experiments in different indoor scenarios that include glass. Figure 6 shows the detection results of our system in finding glass objects under different indoor circumstances. From (a), (b), (c) and (d), it can be illustrated that no matter where the location of the glass is, it will always be detected and marked by our system correctly. The (e) represents the detection result in warning distance, which is important for visually impaired people to avoid glass obstacle. In (f), there are some chairs behind the glass window, which means the depth information of this area is the depth of the chairs, not an open area. This is a challenge to distinguish, but our system has successfully overcome it. The (g) and (h) are two detection results of the mirror. There are a part of corridor and a mirror in the left side of (g), their depth values are both larger than the white wall's in the depth map, but our system is able to distinguish them and accurately locate the mirror. And as it shown in (h), a man wearing our device is standing in front of a mirror. Though the texture of the mirror and the gray wall is somewhat similar, the mirror is also detected by comparing the different data of the two sensors. Through the test of all above environment, we have full confidence in the practical application of our system to help visually impaired people find glass objects in the future.

Although the detection effect could meet the basic requirements for practical application, there are still some false detection and missed detection. The main reasons of these problems are poor spatial resolution of the ultrasonic sensor and the glass frames that limit region growth algorithms. For example, the left of (d) is a glass region which should be detected, however the window frame finally influences the detection result. The same situation is shown in (f), but such detected results could also be used to generate effective warnings to visually impaired users, such that they would avoid the hazardous glass obstacles safely.

4.2 Effectiveness Evaluation

Besides the presentation of Figure 6, there is also some effectiveness evaluation that includes some quantitative data to measure the advantages and disadvantages of the method. To better illustrate the advantages of our method, we define two evaluating indicators:

$$RFD = \frac{P_{false}}{P_{all}} \tag{1}$$

$$RMD = \frac{P_{missed}}{P_{all}} \tag{2}$$

Where the P_{false} represents the number of the pictures without glass objects that are detected, the P_{missed} represents the number of the pictures with glass objects that are not detected and the P_{all} is the number of all the tested pictures. Accordingly, we have collected 720 pictures indoor scenarios in 5 different buildings during the day time, including 540 pictures with glass objects in the valid detection distance. In addition, based on the large number of experimental tests we have done, we divide the glass obstacles into three common categories that have been mentioned above: glass door, glass window and mirror. Table 1 shows the *RFD* and *RMD* of the three glass objects.

Table 1. The false detection rate and missed detection rate of different glass categories.

Category	**Glass door**	**Glass window**	**Mirror**
RFD	2.84%	3.53%	4.09%
RMD	8.22%	7.54%	3.71%

Through the analysis of the experimental results and raw data, some possible reasons that cause these detection problems are given as following. For glass door and window detection, we think that some objects on the glass, such as stickers, result in a higher *RMD*. Because the depth of these objects can be calculated correctly by RGB-D sensor, it means that if they happen to be in the calibration area, the system would not make an affirmative judgment in the frame. Therefore, how to eliminate such an influence is also an important part to improve the accuracy of detection. For mirror detection, lower *RMD* is caused by the fact that mirrors are usually fixed on the walls, which means it is easier to satisfy the judging conditions. In addition, we find that most of the missed detection pictures are caused by abnormal ultrasonic data and the poor spatial resolution of the ultrasonic sensor. This influence could be reduced by improving the performance of the hardware, especially ultrasonic sensor.

In conclusion, it could be found from Table 1 that the overall correct detection rate is beyond 90% when we apply our system into real-world indoor scenarios. Thus, it means that our system has ability to effectively help visually impaired people avoid glass obstacles in daily life.

5. CONCLUSIONS

In this paper, a method based on the fusion of sensors for visually impaired people to detect glass obstacles is proposed. Accordingly, we develop effective algorithms to work with our wearable device, which incorporates an RGB-D sensor and two ultrasonic sensors. To our knowledge, there are few related literatures dealing with this problem. The main contribution of our system is to detect glass object and recover the sparse depth map by fusing depth values from different sensors in real time. In addition, we perform some extension work regarding the issues on the mirror detection and the limited detectable range of the depth camera. The experiment results turn out that the method achieves correct detection rate beyond 90%, which qualifies the proposed framework for real-world indoor navigation assistance in visually impaired individuals.

For future work we plan to extend our system with advanced algorithms for object recognition so that visually impaired people could better understand the surrounding environment while avoiding obstacle effectively. Moreover, we will also explore the glass detection method based on polarized light, which is promising to further enhance the framework in terms of spatial resolution and robustness.

REFERENCES

[1] Bourne R. R. A., Flaxman S. R., Braithwaite T, Cicinelli M. V., Das A., Jonas JB, Jonas, J. B., Keeffe J., Kempen J. H., Leasher J., Limburg H., Naidoo K., Pesudovs K., Resnikoff S., Silvester A., Stevens G. A., Tahhan N., Wong T. Y. and Taylor H. R., "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," Lancet Glob Health, 5(9), (2017).

[2] Tian, Y., "RGB-D Sensor-Based Computer Vision Assistive Technology for Visually Impaired Persons," Chapter in book Computer Vision and Machine Learning with RGB-D Sensors, Springer International Publishing, 173-194 (2014).

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

[4] Yang, K., Wang, K., Bergasa, L. M., Romera, E., Hu, W., Sun, D., Sun, J., Cheng, R., Chen, T. and López, E., "Unifying Terrain Awareness for the Visually Impaired through Real-Time Semantic Segmentation," Sensors, 18(5), (2018).

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

[6] Mchenry, K., Ponce, J. and Forsyth, D., "Finding Glass," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on IEEE, 973-979 vol. 2 (2005).

[7] Born, M., Wolf, E. and Hecht, E., "Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light," Pergamon Press, New York, 40-43 (1980).

[8] Wolff, L. B. "Polarization vision: a new sensory approach to image understanding," Image & Vision Computing, 15(2), 81-93 (1997).

[9] Yang, K., Cheng R., Wang K., Hu, W., Huang X. and Bai J., "Detecting Traversable Area and Water Hazards for the Visually Impaired with a pRGB-D Sensor," Sensors, 17(8), (2017).

[10] Ye, M., Zhang, Y., Yang R. and Manocha, D., "3D Reconstruction in the presence of glasses by acoustic and stereo fusion," Computer Vision and Pattern Recognition IEEE, 4885-4893 (2015).

[11] Yang, S. W. and Wang, C. C., "Dealing with laser scanner failure: Mirrors and windows," IEEE International Conference on Robotics and Automation, 3009-3015 (2008).

[12] Roy, N., Newman, P. and Srinivasa, S., "Recognition and Pose Estimation of Rigid Transparent Objects with a Kinect Sensor," MIT Press, 504 (2012).

[13] Mchenry, K. and Ponce, J., "A Geodesic Active Contour Framework for Finding Glass," IEEE Computer Society Conference on Computer Vision & Pattern Recognition, IEEE Computer Society, 1038-1044 (2006).

[14] Keselman, L., Woodfill, J. I., Grunnetjepsen, A. and Bhowmik, A., "Intel(R) RealSense(TM) Stereoscopic Depth Cameras," Computer Vision and Pattern Recognition Workshops IEEE, 1267-1276 (2017).