A GROUND AND OBSTACLE DETECTION ALGORITHM FOR THE VISUALLY IMPAIRED

Ruiqi Cheng*, Kaiwei Wang*, Kailun Yang*, Xiangdong Zhao*

**State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, China.*
Email: wangkaiwei@zju.edu.cn

Keywords: region growing, image edge, region boundary, aid for the visually impaired.

Abstract

In order to provide the visually impaired with efficient aid, an intelligent aid system to detect ground and obstacle is necessary. In this paper, a new seeded region growing algorithm to detect ground and obstacle is put forward. The algorithm is based on three-dimensional depth image obtained from RGB-D camera and attitude angle obtained from attitude angle sensor. Instead of attaching importance to growing threshold, the Sobel edges of image and the boundaries of region are adequately considered in the algorithm to improve detection accuracy. Seeds are chosen according to the edges of image, and the stop of growing refers to the Sobel edges of image and growing threshold. The ground and obstacle are roughly detected after region growing, however, all of the regions are not intended. Therefore, regions are combined or excluded according to their boundaries. The results of the algorithm tell the user where ground and obstacles are. The experiments in different environment are presented in the paper, and demonstrate that the algorithm achieves qualified accuracy and speed.

1 Introduction

According to WHO, there are approximately 285 million people who are visually impaired in the world [7]. The visually impaired have plenty of difficulties in daily life, because they lack the capability to obtain sufficient visual information. The conventional aid methods for visually impaired, like guide dog and white cane, do not provide the visually impaired with comprehensive information of environment [2]. Therefore, developing an efficient aid system for the visually impaired is of vital importance.

For aiding technology, an accurate and rapid three-dimensional distance measuring technology, which has high refreshing frequency and spatial resolution, is required. Common measurement technologies that are used for the visually impaired aid system include ultrasonic wave sensor, stereo camera and RGB-D camera [2].

Ultrasonic wave measuring technology makes use of time difference between emitting wave and receiving wave to acquire the distance of objects. The technology has been applied in many aid systems for the visually impaired and makes the cost low [2, 9]. The measurement is not disturbed by electromagnetic wave, like visible light and infrared. But the technology has low spatial resolution.

Stereo vision technology takes advantage of the disparity of objects from two separated camera to obtain three-dimensional information. Stereo vision technology has high spatial resolution, high refreshing frequency and large detection range [2]. Other than depth image, colour image acquired from camera may be used to achieve advanced functions, like pattern recognition. However, it is difficult to develop effective stereo matching algorithm to get dense depth image.

RGB-D camera overcomes the defeats of ultrasonic and stereo vision. RGB-D camera acquires three-dimensional depth image and colour image of entire field. RGB-D camera has developed into consumer-level product, like KINECT [10]. Also, RGB-D camera appeals researchers as an effective three-dimensional information acquirement tool [1, 6, 8]. In this paper, KINECT for Windows V2 is utilized, considering its high precision, high resolution and large detection range.

The basic function of aid system for the visually impaired is instructing the user to avoid obstacles and walk on the pathway of ground. Therefore, based on effective distance measurement technology, implementing a rapid and accurate ground and obstacle detection algorithm is a key point for the system.

Ground and obstacle detection algorithms for RGB-D camera have been studied for researchers. Because the depth difference between obstacles and background, edge detection is a method for obstacle detection. Choi et al [1] extract vertical edges of colour image and depth image, and choose common edges, which are obstacles. However, the detection effect of the method is unsatisfied in sophisticated environment. Lee et al [5] put forward an image segmentation algorithm. Edges of depth image are eliminated from the image, and each remaining pixel, as a seed, starts an advance region growing. The set of pixels which have same growing results is defined as an initial region. The initial region, as seed, starts region growing to extend itself. By analysing the depth divergence of growing results, growing results are classified as ground or obstacles. The algorithm implements total segmentation of depth image, but has high time complexity.

In this paper, a new ground and obstacle detection algorithm, which based on depth camera, is put forward. Besides, the paper gives several output forms of ground and obstacles information. Section 2 presents the main flow of detection algorithm. In section 3, the acquirement and process of three-dimensional depth data is presented. In section 4, the ground and obstacle detection algorithm is elaborated. In section 5, output forms of ground and obstacles information are presented. In section 6, experiment results of detection algorithm are discussed. Finally, the conclusion of detection algorithm is presented in section 7.

2 Algorithm flow

The detection algorithm is based on depth image acquired from depth camera and attitude angle acquired from attitude angle sensor, and finally gives the direction of pathway, as well as the position and size of obstacles. The flow chart of the algorithm is shown in Figure 1.

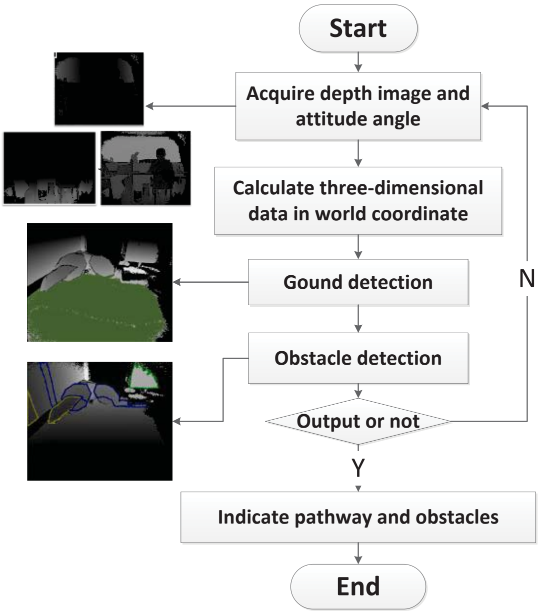


Figure 1: The flow chart of the algorithm.

The main steps of algorithm include acquiring depth image and attitude angle data, calculating three-dimensional depth data in world coordinate, spatially filtering depth image, ground detection and obstacle detection, and output of pathway and obstacles. Ground or obstacle detection includes seeded region growing aided by image edges and regions post-process aided by region boundaries.

3 Depth data acquirement

3.1 Depth data in camera coordinate

RGB-D camera, like KINECT that used in the paper, obtains colour image and three-dimensional depth image at the same time. KINECT is composed of infrared laser, diffracting grating and RGB camera [4], and is based on Light Coding technology [10].

The depth data in the raw image is in camera coordinate, which is represented by (Cx, Cy, Cz). The camera coordinate takes the centre of the plane where the depth is zero as the origin, and takes the opposite direction of optical axis of camera as the positive direction of Z axis, and takes upward direction in the imaging plane of camera as the positive direction of Y axis. The X axis is decided by right-hand rule. The grey value of every pixel in depth image represents the depth value of object in the environment. The objects out of the range and the objects that do not reflect laser are set to zero value in depth image.

The depth data is accurate to one millimetre. Error test is executed in different distance, and the relative errors are less than 1.5% in measurement range, as shown in Figure 2.

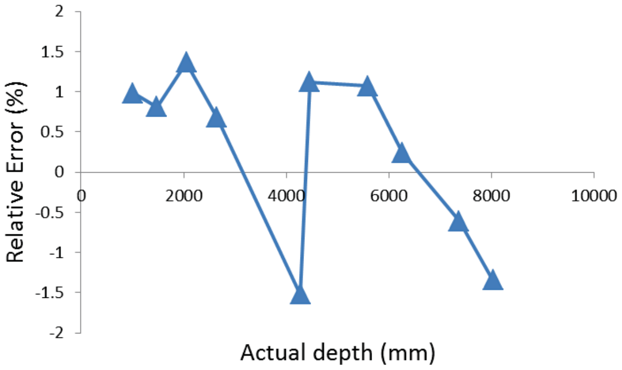


Figure 2: The relative error of measured depth.

The linearity of data detection is high, because R^2 of fitting line is 0.9996, as shown in Figure 3.

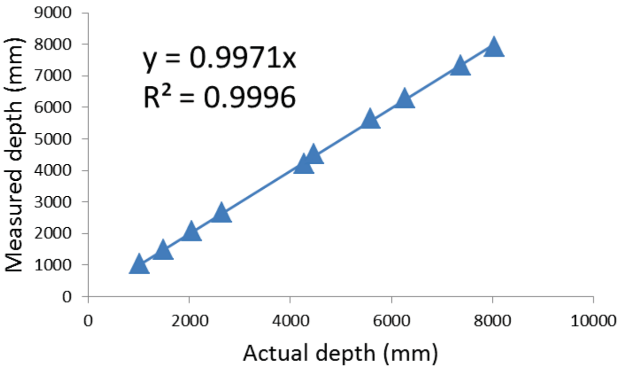


Figure 3: The linearity of depth measurement.

In Figure 2 and 3, the actual depth is measured by portable laser range finder. Therefore, the precision of KINECT is close to that of the range finder, and depth detection satisfies the meet of aid system for the visually impaired.

3.2 Depth data in world coordinate

The depth data in camera coordinate is influenced by attitude angle of depth camera. When the camera is in any attitude, a horizontal ground may not be presented as horizontal in raw depth image. Therefore, raw depth image acquired from depth camera should be combined with attitude angle data to obtain depth data in world coordinate, which is independent on the attitude angle of camera.

The attitude angle sensor is composed of 3-axis gyroscope and 3-axis accelerator, and acquires precise attitude angle with the help of geomagnetic sensor [3]. The three-dimensional depth data in world coordinate is calculated according to the following equation.

$$\begin{bmatrix} Wx \\ Wy \\ Wz \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} Cx \\ Cy \\ Cz \end{bmatrix}$$

In the formula, (Cx, Cy, Cz) represents data in camera coordinate, and (Wx, Wy, Wz) represents data in world coordinate. The world coordinate takes the centre of the plane whose depth value is zero as the origin, and takes the projection of Z axis of camera coordinate in horizontal plane as Z axis, and takes upward direction vertical to horizontal plane as the positive direction of Y axis. The X axis is decided by right-hand rule. In the formula, α, β and γ, which are acquired from attitude angle sensor, represent the attitude angles that rotate around X, Y and Z axis.

3.3 Image data filtering

Depth data in world coordinate is used as initial data in seeded growing, but a lot of noises exist in the image. Objects at edges of field, out-of-range objects and the objects of low-reflectivity are zero in depth image. Besides, many noises appear at the edges of the objects in depth image, like Figure 4(a). Therefore, filtering the depth image to reduce noises is essential before region growing.


(a) (b)

Figure 4: (a) raw Z image (b) filtered Z image.

Spatial filtering is applied to fill the isolated zero point and eliminate isolated non-zero point. After the filtering, the noises of image are reduced, like Figure 4(b).

4 Ground and obstacle detection algorithm

In order to offer the visually impaired with efficient aid, the obstacles and pathway in the surrounding should be hinted. Therefore, region growth method aided by image edges and region boundaries is put forward to achieve efficient ground and obstacle detection. The path of ground is selected from the Y image in world coordinate, and the obstacles are detected from the Z image in world coordinate. The algorithm includes growing initialization, region growing, exclusion and combination of growing results.

4.1 Ground detection

It is obvious that the path of ground is under the camera, and the pixels of Y image that has positive Y value are not engaged in the procedure of growing. Sobel operator is used on Y image, and the horizontal and vertical edges of image are extracted. Divide the entire Y image to several equal-size parts, in which a seed is chosen according to the random number. Therefore, the seeds distribute in different location of the image.

The seed is valid to grow when it meets three conditions. Firstly, the seed does not belong to any grown region. Secondly, the seed does not belong to edges of image. Thirdly, the difference between the grey value of seed and zero is larger than growing threshold. If all of the conditions are met, growing starts from the seed Gs, whose grey value is hs. Otherwise, the seed is invalid, and next seed is considered.

Growing starts from a central pixel G, whose grey value is h. Gs is taken as the first G in the growing of current seed. One of the four-connected neighbours is Gi, whose grey value is hi. Whether Gi belongs to G's region depends on the following four growing conditions.
Condition 1: Gi is not located at Sobel edge of image.
Condition 2: Gi does not belong to any other regions.
Condition 3: Gi is not visited during the growing course of the current seed.
Condition 4: $|hs-hi|<\delta$, where δ is growing threshold.

According to the four conditions, following situations are considered.
(1) If all of four conditions are true, flag Gi as the same region with G.
(2) If the fourth condition is false and others are true, flag Gi as boundary pixel.
(3) If the second and third conditions are true and the first condition is false, flag Gi as sharp boundary pixel.
(4) If the third condition is true and the second condition is false, flag Gi as neighbour boundary pixel.
(5) When Gi is not located at the rim of image, flag Gi as boundary pixel.
(6) Do not react for any other situations.

If situation (1) is met, Gi is qualified for the region grown from G. Each qualified neighbour pixel is put into the stack. When all of G's four-connected neighbour pixels have been visited, pop Gi out from the stack and let Gi be the central pixel and repeat the above operation. When the stack is empty, one seed growing finishes.

After the growing course, several regions have appeared on the image. However, all of the regions do not belong to what we want – the pathway of ground. Meanwhile, the entire ground may be grown to several parts rather than one part. Therefore, the operation after seed growing is critical for accurate results.

Some regions which have few pixels are excluded. After that, neighbouring regions that have similar height may be combined together. Through above operations, small regions which may be produced by noises are eliminated, and ground is more intact.

But the regions which are not ground will also appear in results. In order to exclude the non-ground results of growing, boundaries of regions are applied. The two following conditions are considered.
(1) Consider the ratio E = m/n of every region, where m represents the number of sharp boundary pixels and n represents the number of boundary pixels. If E is larger than the set threshold, the region is viewed as ground and is retained. Otherwise, the region should be excluded, because the region does not belong to ground.
(2) Consider the maximum and minimum value of Z in every region. If the difference between the maximum and minimum value of Z in a region is larger than the set threshold, the region should be excluded, because the region does not belong to ground. Otherwise, the region tends to be retained.

The last exclusion is more crucial, in that it excludes the non-ground results and decides the accuracy of ground detection. After all of the operations to the results of growing, the remaining regions in depth image are horizontal plane. Set the regions that are near to the bottom rim of image as ground. In section 5, the detailed pathway instructions are given.

4.2 Obstacles detection

The same method with ground detection is applied in obstacles detection. Of course, some details in ground detection are changed to adapt obstacles detection. Different from ground detection, obstacle detection use Z image as initial data of growing. Apply the same method to Z image to extract its Sobel edge. Unlike selecting seeds at random in ground growing, seeds in obstacles growing are selected according to the Sobel edges. Select several evenly distributed rows in depth image. In each row, seeds are selected at the middle point of two Sobel edges pixels.

To avoid duplicate detection in obstacle growing, the ground results of ground detection are eliminated in Z image. And obstacles grow in the remained part of Z image. The four conditions in section 4.1 are also considered, but the six situations are substituted by the followings:
(1) If all of four conditions are true, flag Gi as the same region with G.
(2) If the fourth condition is false and others are true, flag Gi as boundary pixel. Especially, when |h-hi|<δ is satisfied, flag Gi as continuous boundary pixel.
(3) If the second and third conditions are true and the first condition is false, flag Gi as boundary pixel.
(4) If the third condition is true and the second condition is false, flag Gi as neighbour boundary pixel. Especially, when h and hi satisfy |h-hi|<δ, flag Gi as continuous boundary pixel.
(5) Do not react for any other situations.

After growing, several regions have appeared on the image. Like the growing results of ground, all of the regions do not belong to obstacles. Meanwhile, the entire obstacle may be grown to several parts rather than one part. According to the number of pixels in region, some regions which have few pixels are excluded. After that, neighbouring regions that have similar height with each other may be combined together.

In order to exclude the results that do not belong to obstacle, like wall beside the user, the ratio of continuous boundary pixels out of all boundary pixels are considered. Consider the ratio E' = m'/n' of every region, where m' represents the number of continuous boundary pixels and n' represents the number of boundary pixels. If E' is less than threshold, the region is viewed as obstacle and is retained. Otherwise, the region should be excluded.

The ratio of continuous boundary is used in exclusion, because boundaries of most regions that are not obstacles are mostly continuous boundary. The ratio of sharp boundary number does not fit for the results exclusion of obstacle detection, because the obstacles whose boundaries do not vary sharply in grey value may be excluded. After all of the steps, the remaining regions in depth image are obstacles.

5 Ground and obstacles prompts

Based on the calibrated RGB image, ground and obstacles detection results are painted, and Figure 5 shows a detection result.

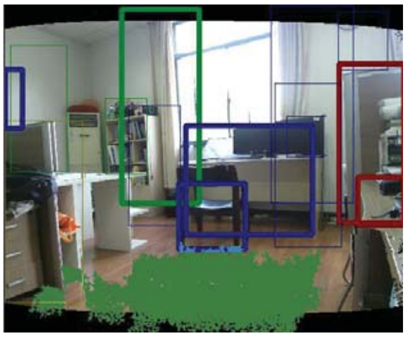


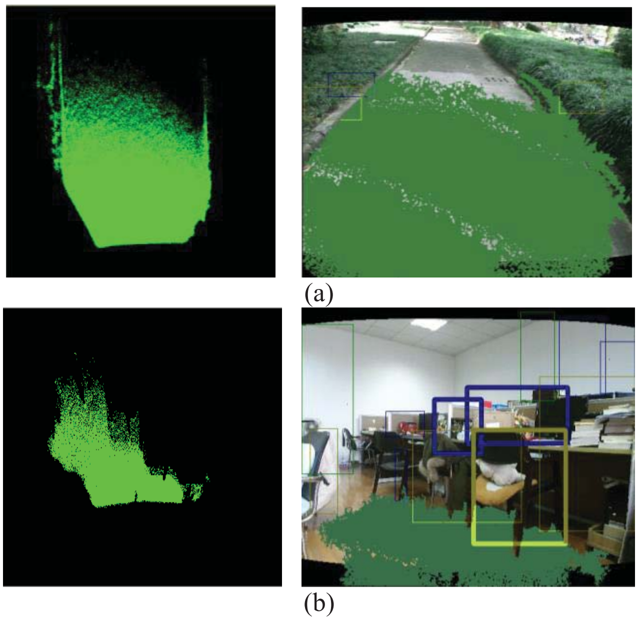
Figure 5: A detection result under certain circumstance.

Ground is filled with pseudo-colour. Red represents near ground, and blue represents far ground. Obstacles are contoured with rectangle. The obstacles located within 1 metre far from the user are labelled with red, and the obstacles located from 1 metre to 2 metres far are labelled with yellow, and the obstacles located from 2 metres to 4 metres far are labelled with yellow, and the obstacles located from 4 metres to 6 metres far are labelled with green. Thick lines are used to label obstacles whose centre's X value is from negative one metre to positive one metre, and thin lines are used to label other obstacles.

5.1 Pathway touch image

To inform the visually impaired with pathway in front of him or her explicitly, pathway touch image is an effective interactive manner. The touch image is a binary image, where one-value pixels (green in image) represent walkable pathway and zero-value pixels mean impassable. The origin of world coordinate is located at the centre of the bottom rim of touch image, and the horizontal axis of touch image represents X axis of world coordinate, and the vertical axis of touch image represents Z axis of world coordinate. The touch image presents ground within 4 metres far in Z axis and 4 metres wide in X axis. Figure 6 shows a touch image under some circumstances.

Each pixel of detected ground is drawn on the touch image according to its Z and X value in world coordinate. Moreover, obstacles are also reflected in touch image. Ground behind the obstacles is eliminated from touch image. For example, in Figure 6 (b), ground under the chair is not pathway, and should be eliminated from pathway touch image.


(a)

(b)

Figure 6: Pathway touch images under some circumstances.

5.2 Pathway and obstacles instruction

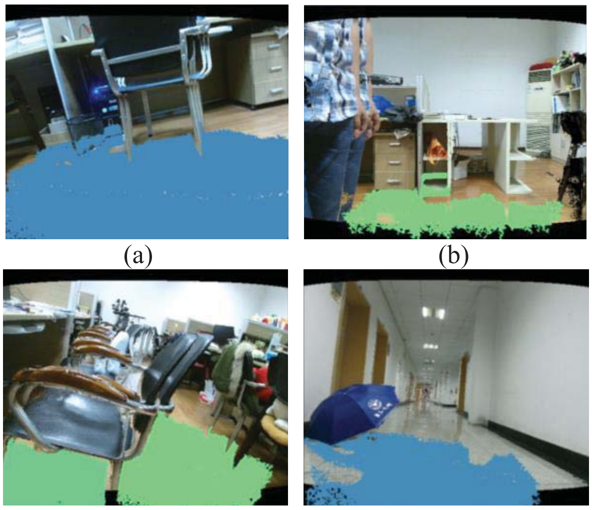
In order to give voice or vibration instruction of pathway, touch image is processed further. According to the number of ground pixels in the different partitions of touch image, walkable direction is decided.

Taking the origin of touch image as the centre, the Z axis is rotated around the centre for ±5 degree, ±15 degree, ±25 degree and ±35 degree to form eight borderlines which divide touch image into seven partitions, as shown in Figure 8. Every partition represents a practical direction of pathway. If the number of ground pixels in the partition is larger than a threshold, the direction is defined walkable.

6 Analysis of detection results

In the section, ground and obstacles detection results under different circumstances, which include daytime and night, indoor and outdoor, are presented.

Under different circumstances, some ground detection results is shown in Figure 7. From Figure 7 (a) and (b), the grounds of different height are correctly detected in complicated indoor environment, and the chair, the wall or other stuffs are not detected as ground. Locating the camera at attitude angle, the detection results of ground is shown in Figure 7 (c) and (d). The grounds are correctly detected. Meanwhile, curved surfaces like umbrella and other plane like walls are not detected as ground.


(a) (b)

(c) (d)

Figure 7: Some ground detection results, in which the camera is located at different height and attitude angle.

According to the experiments indoor and outdoor, the correct ground detection rate, that the ratio of correct ground detection results to all of the results, is up to 95%.

The obstacles detection results are shown in Figure 8, where obstacles in different shape and location are correctly detected. Curved surface obstacle, like the umbrella in (c), and plane obstacle, like the wall in (d), are both detected as obstacle. Moreover, near obstacles in (a) and (b), as well as distant obstacles in (e) and (f), are correctly detected.

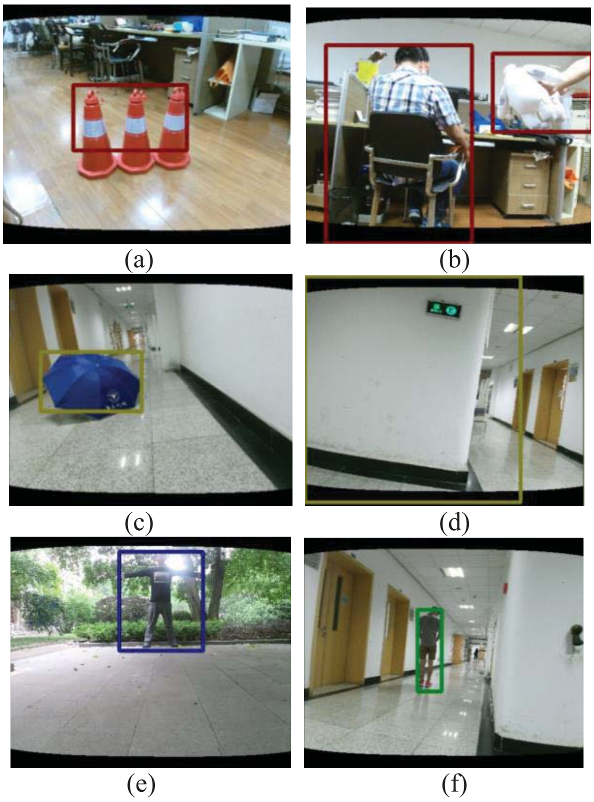


Figure 8: Some obstacle detection results. The roadblocks in (a) and the human body and the doll in (b) are detected. The umbrella in (c) and the wall in (d) are detected, and the distant human body in (e) and (f) are also detected.

According to the experiments indoor and outdoor, the algorithm is able to detect most obstacles, and the correct obstacle detection rate is up to 90%.

More detection results are presented in Figure 9. From the figure, the ground and obstacles are correctly detected, and even in night the detection result is valid.

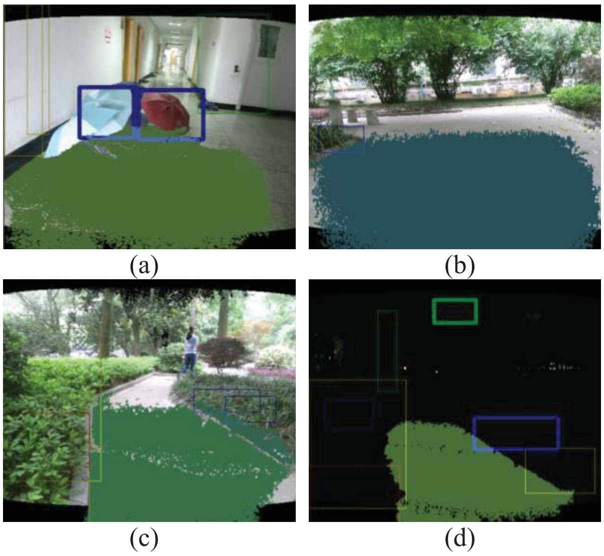


Figure 9: Ground and obstacle detection results.

As for processing speed, on a computer whose CPU is 3.30GHz, the refreshing frequency of algorithm is slightly more than 10 frames per second, in that every frame takes 78 to 94 milliseconds to process. The refreshing frequency makes real-time detection possible, and totally meets the need of aid for the visually impaired.

7 Conclusion

In the paper, a new ground and obstacle detection algorithm for the visually impaired is presented. The algorithm based on depth image acquired from depth camera and attitude angle data. Seeded region growing method aided by image edges and region boundaries is the main content of algorithm. The detection results turn out that the algorithm achieves high correct detection rate of 90% and high refreshing frequency of 10 frames per second.

References

[1] J. Choi, D. Kim, H. Yoo, and K. Sohn, "Rear obstacle detection system based on depth from kinect," in *2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 98-101, (2012).

[2] L. Dunai, I. Lengua Lengua, I. Tortajada, and F. Brusola Simon, "Obstacle detectors for visually impaired people," in *2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, pp. 809-816, (2014).

[3] InvenSense, "MPU-6000 and MPU-6050 Product Specification Revision 3.4," ed, 2013.

[4] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data", *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* **vol**. **3812**, pp. 133-138, (2011).

[5] C.-H. Lee, Y.-C. Su, and L.-G. Chen, "An intelligent depth-based obstacle detection system for visually-impaired aid applications," in *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 1-4, (2012).

[6] M. Maisto, M. Panella, L. Liparulo, and A. Proietti, "An Accurate Algorithm for the Identification of Fingertips Using an RGB-D Camera", *IEEE Journal on Emerging & Selected Topics in Circuits & Systems,* **vol**. **3**, pp. 272-283, (2013).

[7] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010", *British Journal of Ophthalmology,* **vol**. **96**, pp. 614-618, (2012).

[8] W. Shang, W. Jiang, and J. Chu, "A machine vision based navigation system for the blind," in *2011 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, pp. 81-85, (2011).

[9] B. S. Shin and C. S. Lim, "Obstacle Detection and Avoidance System for Visually Impaired People," in *Haptic and Audio Interaction Design*, ed: Springer Berlin Heidelberg, pp. 78-85, (2007).

[10] T. Yu, *Kinect in action*: China Machine Press, 2013.