International Conference on Internet Studies,
July 18-19, 2015, Tokyo, Japan

AN OBSTACLE DETECTION SYSTEM USING DEPTH INFORMATION AND REGION GROWING FOR VISUALLY IMPAIRED PEOPLE

Ching-Tang Hsieh
Tamkang University, Taiwan
hsieh@ee.tku.edu.tw

Han-Yi Tsai
Tamkang University, Taiwan
barriego@hotmail.com

Hsieh-Chang Huang
Tamkang University, Taiwan
sanmic@mail.lit.edu.tw

Cheng-Hsiang Yeh
Tamkang University, Taiwan
ychsiang@wda.gov.tw

Li-Ming Chen
Tamkang University, Taiwan
ms0071799@hotmail.com

ABSTRACT

This study proposes an obstacle detection method based on depth information to aid the visually impaired people in avoiding obstacles as they move in an unfamiliar environment. Firstly, we have applied dilation of morphology and erosion of morphology to remove the crushing noise of the depth image and have used the Least Squares Method (LSM) in a quadratic polynomial to approximate floor curves and determine the floor height threshold in the V-disparity. Secondly, we have searched for dramatic changes depth value in accordance with the floor height threshold to find out suspicious stair edge points. Thirdly, we have used the Hough Transform to find out the location of the drop line. In order to strengthen the characteristics of the different objects to overcome the drawbacks of the region growing method, we have applied edge detection to remove the edge. Fourthly, we have used the floor height threshold and features of the ground to remove ground plane. And then our system has used the region growing method to label the tags on different objects. It has analyzed each object to determine whether the object is a stair. Fifthly, if the result is neither up stair nor down stair, we have used K-SVD algorithm to determine whether the object is people. Finally, the system has assisted the users to determine the stairs direction and obstacle distance through a voice prompt by Text To Speech (TTS). Experimental results show that the proposed system has great robustness and convenience.

Keyword: Obstacle Detection, Kinect, Depth Image

1. INTRODUCTION

There are so many visually impaired people relying on the guide cane or guide dogs to move around freely in the world. But, not every visually impaired people can easily pair successfully and obtain guide dogs, they often have to wait for a long time. In addition, visually impaired people have to touch the obstacle with the guide cane before they get the position of the obstacle and avoid it. In recent years, there are a lot of developments and advancements in computer vision. Many scholars have proposed a lot of obstacle detection methods. In [1] Obstacle detection can be classified into three categories: Electronic travel aids (ETAs), electronic orientation aids (EOAs), and position locator devices (PLDs). However, obstacle detection in this paper can be classified into two categories. One is based non-depth information [2] [3] [4] [5] [6], and the other is based on the depth information [7] [8] [9] [10] [11].

For the first category, an object detection algorithm based on edges and motions has been proposed [2]. An obstacle detection algorithm by a single camera was proposed in [3]. This work uses edge detection to segment objects. But these methods needs are without complex texture on the surface of the ground. In [4], an obstacle detection based on saliency map has been proposed. However, the execution of the proposed method required a few obstacles in the environment. An obstacle detection based on a gray-scale image has been proposed in [5]. But the proposed method worked on the gray-scale image, so it is affected easily by illumination. These methods have great robustness under sufficient light, but under insufficient light they loss robustness. Our system uses Kinect directly to capture the depth image, so it can overcome these drawbacks.

For the second category, an obstacle detection algorithm based on U-V disparity map analysis was presented in [7]. In [8], the depth image was obtained by 3D camera. This work combined U-V disparity map to find out the location of the obstacles. An obstacles detection based on a Kinect sensor has been proposed in [9]. This work used a Kinect sensor to obtain color images and depth images. An obstacles detection based on Kinect sensor has been proposed in [9]. This work used Kinect sensor to obtain color images and depth images. In addition, our system compared with [12], this method does not do the processing for the floor, but it segments object directly for calculating standard deviation using object's depth value. To determine whether it is an obstacle by scale of the object's standard deviation. Although this detection method is simple, but smaller objects on the floor are judged to the floor. In contrast, our system filters the floor out before obstacles detection step, so it solves this issue.

This paper is organized as follows. In the next section, we refer to our proposed method. Section 3 reports the experimental results.

2. Proposed Method

2.1. Noise reduction

Because of Kinect hardware limitations, the depth image may break. In order to make the depth image more complete, we apply some simple morphology processing. In this paper, we use erosion and expansion to repair the black broken areas. From Fig. 2, we can see that the processing depth images are better than original depth images.

2.2 Ground detection

For detection needs of subsequent steps, the captured depth image first has to remove noise and project it into V disparity map. The Y-axis height of V disparity map corresponds to the Y-axis height of the depth images, so the vertical length of image represents the high of the actual object in the image.

Closer to right side of depth map means that the distance between the object and the sensor is farther. The higher pixel value in V disparity map presents bigger object in the image. The normalization equation of the cumulative amount of the depth is shown in the following equation.

$$\text{Depth cumulative value} = \frac{cumulative\ value}{Max\ cumulative\ value} * 255 \tag{1}$$

According to [11], a floor is a rising curve in V disparity map. We use LSM for the curve equation, as shown in Eq. (2).

$$\mathrm{a}y^2 + by + c = d \tag{2}$$

where a, b, and c respectively stand for the parameters of the equation, y is the image height, d is the horizontal axis values(0 to 255) in V-disparity image. However, a floor is not only a simple line in the V-disparity map. Because the same height pixel in depth map maybe has different depth value, it causes that the curve becomes the strip. To overcome this problem, we calculate the quadratic equation
for next step. The quadratic equation is shown as Equation (3).

$$\mathrm{TH1} = \mathrm{ay}^2 + \mathrm{by} + \mathrm{c} - \mathrm{offset} = \mathrm{d} - \mathrm{offset} \tag{3}$$

where TH1 is the shifted threshold according to the floor height. The floor height threshold value indicates a height of depth image and the minimum value cannot be less than TH1. The appropriate offset value is 35, which is an experience value. When the floor curve in V disparity map has not been offset, the part of floor is determined to an obstacle by system as Fig. 1. Then the floor curve in V disparity map has been offset, the floor is detected ground correctly as Fig. 2.

2.3. Removal of the edge

In the depth image, the depth represents the distance of the objects and the sensor. According to depth variations, we can know whether these obstacles are the same or not. Depth variations of the same object are usually not too intense. In the different objects, the relationship of the distance causes depth value varies strongly. In this paper, in order to strengthen the characteristics of different objects, we remove the strong edge. And there are so many edge detection methods, as Roberts, Prewitt, Sobel, Laplace and Canny. In this paper, we detect edge by the following Equation (4). The process result is shown in Fig. 3.

$$\mathrm{P(x)} = \begin{cases} 0 \quad , & \text{if } \sum\limits_{\mathrm{x_i} \in \mathrm{x_n}} |\mathrm{P(x)} - \mathrm{P(x_i)}| \geq \mathrm{TH3} \\ \text{unchange}, & \text{others} \end{cases} \tag{4}$$

2.4 Downstairs detection

In this section, the method of searching and recording the points that have acute variations from noiseless image is proposed. In our study, the pixel values are bigger than setting threshold (50) and we define them as acute variation. Next, we apply floor height threshold (TH2) to filter out possible points as shown in Fig. 4(a). After that, Hough Transform technique transforms the filtered points into a horizontal line, as shown in Fig. 4(b).

(a) Real scene (b) The floor curve in V disparity map (c) The result of no offset

Figure 1. No offset

(a) Real scene (b) The floor curve in V disparity map (c) the result of offset value= 20

Figure 2. Offset value = 20

(a) Noiseless image (b) Processing result (c) Noiseless image (d) Processing result

Figure 3. Removal of the edge

(a) Suspicious points of downstairs depth map (b) The results of the Hough transform

Figure 4. Results of downstairs detection

2.5. Removal of the ground

We apply [14] to search for the inner and outer contours, we can find out regions and sizes of the different suspicious planes. And then remove the plane by equation (5), which has a large area. We use the least squares method (LSM) in a quadratic polynomial to approximate floor curves and determine the floor height threshold in the V-disparity.

$$\mathrm{Ground}(x,y) = \begin{cases} 0 \;, P(x,y) - P(x,y-10) \geq 2 \\ \quad \wedge \bigcap_{i=0}^{9} P(x,y-i) - P(x,y-i-1) \geq 0 \\ \quad \wedge\, P(x,y) > TH1 \\ \text{Keep pixel value}, \qquad \text{others} \end{cases} \tag{5}$$

2.6. Labeling

The sensing range of Kinect is 0.8 to 4.0 meters. When the range is farther than the max distance, it is not able to distinguish the distance. So, we need to remove the far information. In order to measure distances accurately, we retain the distances information within 3 meters. Afterwards, we label different tags on different objects. The general labeling methods have 8 connected component labeling and region growing. But the tag harmonization of connected component labeling needs a lot of iterations, because of the complex shape of the connected area.

$$(i,j) = \begin{cases} (i,j), \text{if } [\,(P(i-1,j-1) = 0) \\ \qquad \wedge (P(\;i\;, j-1) = 0 \\ \qquad \wedge (P(i+1, j-1) = 0 \\ \qquad \wedge (P(i-1,\; j\;) = 0 \\ \qquad \wedge (P(\;i\;,\; j\;) \neq 0)\,] \\ \text{not seed}\;, \qquad \text{others} \end{cases} \tag{6}$$

Here, S(i,j) represents the seed coordinates. P(i,j) represents the pixel value at the coordinates (i,j). In order to make the system more efficient, we use Connected

Component Region Growing [13]. The initialization of traditional region growing has to sprinkle some seeds in the image. If the distribution of the sprinkled seeds is not appropriate, it causes the growth result be imperfect. So the choice of the initial position of the seeds has made some improvements in the proposed system. We exploit the information of the object edges. Because the previous step removes the edge information of the object, each object is isolated by black color. We use the following equation (6) and the mask of initial seed as shown in Fig. 5 to select the coordinates of initial seeds and utilize these coordinates to execute region growing. It is able to ensure that each object has an initial seed, and the growth of the place would not be repeated treatment. Therefore, we propose system to reduce the amount of computation.

2.7. Upstairs detection

Next, our system analyzes each tagged objects individually to determine whether the object is upstairs by depth value changing. The upstairs' depth value has a hierarchical characteristic, which is from top to down and from large to small. When the obstacle meets the above characteristics, it is marked as upstairs. The detection results are shown in Fig. 6.

2.8. Object detection by K-SVD method

Sparse representations of signals have received a lot of attention in recent years [14]. In this step, we used K-SVD method [15] to classify objects people or other objects, when the result of analysis is neither up stair nor down stair. As in dictionary learning, a window of the same size is sided into many non-overlapping patches. Each patch is then classified by comparing the three reconstruction errors Ei, i=1, 2 resulting from the people or others dictionaries:

$$E_i(\mathrm{x}, D_i) \cong \|\mathrm{x} - D_i\alpha * (\mathrm{x}, D_i)\|_2^2,$$
where
$$\alpha(\mathrm{x}, \mathrm{D}) = \min_{\alpha \in R^k} \|\mathrm{x} - \mathrm{D}\alpha\|_2^2, \mathrm{s.t.}\ \|\alpha\|^0 < L \tag{7}$$

where $\mathrm{D_i}$, i=1, 2 indicates the trained people dictionary, the other objects dictionary, respectively. If $\mathrm{E_1}$ is smaller than $\mathrm{E_2}$. Thus, the patch should be classified as people.

2.9. Labeling object and inform user

Our system labels objects out with rectangle. It shows information of detected objects on the image, and the distance of the obstacle or staircase. Finally, our system applies Text-To-Speech (TTS) software [16]. When the obstacle appears in front of the user, the system informs user distance of the obstacle and obstacle category with voice; when the system detects the stairs, it provides the direction and distance of the stairs to the user for safety. This voice alarm is very short that focuses on the most concise information of the closest obstacle.

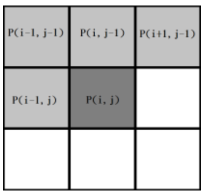
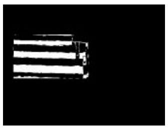


Figure 5. Mask of initial seed



(a) Satisfied conditions of a suspicious plane

(b) Upstairs detection image

Figure 6. Upstairs detection

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

Figure 7. Indoor and Outdoor obstacle detection

3. Experiment Results

Microsoft Kinect sensor is used as a tool to capture images and experimental platform is Windows 7. Programming language is Visual C++ 2010 with OpenCV 2.3 running on notebook with Intel(R) Core(TM) i5-3210M CPU@2.5GHz 8G 64 bits. The image resolution is 640x480 and the depth image capture rate is 30 frames per second. The sensing range is 0.8 to 4.0 meters.

The experiments are tested in indoor environment under sufficient light, indoor environment under insufficient light and outdoor environment under sufficient light as Fig. 7.

Acknowledgments: This work was supported by the National Science Council under grant number MOST 103-2632-E-032 -001-MY3 and MOST 103-2410-H-032 -052.

References

[1] Dakopoulos, D.; Bourbakis, N.G, “Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey”, IEEE Transactions on Systems, vol. 40, no. 1, pp. 25-35, 2010.

[2] Dwivedi, M. ; Ran Li ; Chong Sun ; Kummert, A, “A Real-Time Rear View Camera Based Obstacle Detection”, 12th IEEE International Conference on Intelligent Transportation Systems, pp. 1-6, 2009.

[3] Yankun Zhang ; Hong Chuyang; Weyrich N., “A single camera based rear obstacle detection system”, 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 485 – 490, 2011.

[4] Chen L.; Guo B.L.; Sun W., “Obstacle Detection System for Visually Impaired People Based on Stereo Vision”, 2010 Fourth International Conference on Genetic and Evolutionary Computing, pp.723-726, 2010.

[5] Ying Jie; Song Yanbin, “Obstacle Detection of a Novel Travel Aid for Visual Impaired People”, 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp. 362-364, 2012.

[6] Luis A. Guerrero; Francisco Vasquez; Sergio F. Ochoa, “An Indoor Navigation System for the Visually Impaired”, vol. 12, no. 6, pp.8236-8258, 2012.

[7] Zhang Min.; et al., An Obstacle Detection Algorithm Based on U-V Disparity Map Analysis, 2010 IEEE International Conference on Information Theory and Information Security (ICITIS), pp. 763-766, 2010.

[8] Yuan Gao; et al., “Obstacle Detection with 3D Camera Using U-V Disparity”, 2011 7th
International Workshop on Systems, Signal Processing and their Applications (WOSSPA),
Tipaza , pp. 239-242, 2011.

[9] Jinwook Choi; et al., “Rear Obstacle Detection System Based on Depth From Kinect”, 2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp.98-101, 2011.

[10] Donghun Kim; Kwangtaek Kim; Sangyoun Lee, “Stereo Camera Based Virtual Cane System with Identifiable Distance Tactile Feedback for the Blind”, Sensors 2014; vol.14, no.6, pp.10412-10431.

[11] Alberto Rodríguez; J. Javier Yebes; Pablo F. Alcantarilla; Luis M. Bergasa; Javier Almazán ; Andrés Cela, “Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback”, Sensors 2012, vol.12, no.12, pp.17476-17496.

[12] C. H. Lee; Y. C. Su; L. G. Chen, “An intelligent depth-based obstacle detection system for visually-impaired aid applications”, WIAMIS, pp.1-4, 2012.
[13] R. Adams; L. Bischof Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 6, pp. 641-647, 1994.
[14] R. Gribonval, M. Nielsen, “Sparse approximations in signal and image processing, Signal Processing”, vol. 86, no. 3, pp.415-416, 2006.
[15] M. Aharon ; M. Elad ; A.M. Bruckstein ; “The KSVD: an algorithm for designing of overcomplete dictionaries for sparse representations,” IEEE Transactions on Signal Processing, vol. 54, no.11, pp. 4311-4322, 2006.
[16] TTS: Available online: http://msdn.microsoft.com/en-us/library/ms723627(v=vs.85).aspx (accessed on 25 October 2014)