An Intelligent Depth-Based Obstacle Detection System for Visually-Impaired Aid Applications

Chia-Hsiang Lee, Yu-Chi Su and Liang-Gee Chen
DSP/IC Design Lab.,
Graduate Institute of Electronics Engineering
National Taiwan University, Taipei, Taiwan
Email: {ade,steffi,lgchen}@video.ee.ntu.edu.tw

Abstract— **In this paper, we present a robust depth-based obstacle detection system in computer vision. The system aims to assist the visually-impaired in detecting obstacles with distance information for safety. With analysis of the depth map, segmentation and noise elimination are adopted to distinguish different objects according to the related depth information. Obstacle extraction mechanism is proposed to capture obstacles by various object proprieties revealing in the depth map. The proposed system can also be applied to emerging vision-based mobile applications, such as robots, intelligent vehicle navigation, and dynamic surveillance systems. Experimental results demonstrate the proposed system achieves high accuracy. In the indoor environment, the average detection rate is above 96.1%. Even in the outdoor environment or in complete darkness, 93.7% detection rate is achieved on average.**

I. INTRODUCTION

With the advances of computer vision algorithms and the evolution of hardware technologies, intelligent vision systems have attracted increasing attentions in the age. Many mobile applications, including intelligent vehicle navigation, robotic vision, augmented reality, visually-impaired electronic aids, and dynamic surveillance systems, have the vision ability to fulfill specific tasks. Owing to the maturity of 3D-sensors these years, depth-based systems have been adopted to solve many traditional computer vision problems, such as object tracking [1], face recognition [2] and feature extraction [3]. Omar Arif used the 3D-sensor to track and segment different objects in disparity layers [1]. Simon Meers proposed the depth-based face recognition with feature point extraction [2]. Peter Gemeiner presented a corner filtering scheme combining both the intensity and depth image of a TOF camera [3].

For mobile vision applications, obstacle detection and collision avoidance is an important or even basic feature to support useful information to the user for safety. Recently, many works have been proposed to address this issue. These works can be divided into two categories: (1) non-depth-based obstacle detection and (2) depth-based obstacle detection approaches. For the first category, Long Chen proposed the saliency map obstacle detection [4]. This work extracted the visual attention region as an obstacle under certain thresholds. However, it could only be used in environments with few obstacles because the saliency map must be constructed with coherent characteristics. Zhang Yankun introduced an obstacle detection approach by using a single view image [5]. Based on edge detection for object segmentation, the assumption of the work is that the road surface is a plane without texture. For the second category, Cosmin D. Pantilie used motion to detect obstacles for automotive applications [6]. According to his work, the pose of the stereo camera should be fixed in a certain angle and the environment would be restricted to the road. Pangyu Jeong presented a stereo-based real-time obstacle and road plane detector with adapting mechanism for various environments [7].

In this work, we propose a robust depth-based obstacle detection and avoidance framework based on computer vision technologies. Our system aims to assist the visually-impaired in obstacle detection for safety. Compared with previous works [5] [6] [7], which focus on automotive or robot applications, with assumptions of fix height of camera positions, the proposed method enables free viewpoint of camera moving, which is more reasonable and further meet the specification of a head-mounted visually-impaired electronic aids. With analysis of different depth layers and object proprieties in the depth map, obstacles can be correctly detected and the corresponding distances from the user can be estimated. Since for many real mobile applications, only obstacles appearing within a small distance are potential threats to the user, we set the alarm range as 1-2 meters in this paper. Only obstacles exist in the alarm range will be reported to remind the user. The proposed system could still work successfully even in darkness and outdoor environments. More than 10000 frames of each sequence taken in different environments are used to test the robustness of the proposed work. The experimental results prove that the proposed system achieves 96.1% and 93.7% detection rate on average in the indoor and outdoor environments, respectively.

II. PROPOSED FRAMEWORK

Initially, a 3D-sensor is used to provide depth images as well as color images. There are two stages to be performed. The first stage is object segmentation. Edge detection on the depth image is used to find the discontinuous depth values in real-world environments. Then depth layers of the depth image are analyzed to distinguish different objects. We regard the pixels with similar depth values in neighboring locations as the same object. However, the segmented images are noisy. To overcome this problem, noise elimination is performed with information of detected edges and pixel numbers. The second

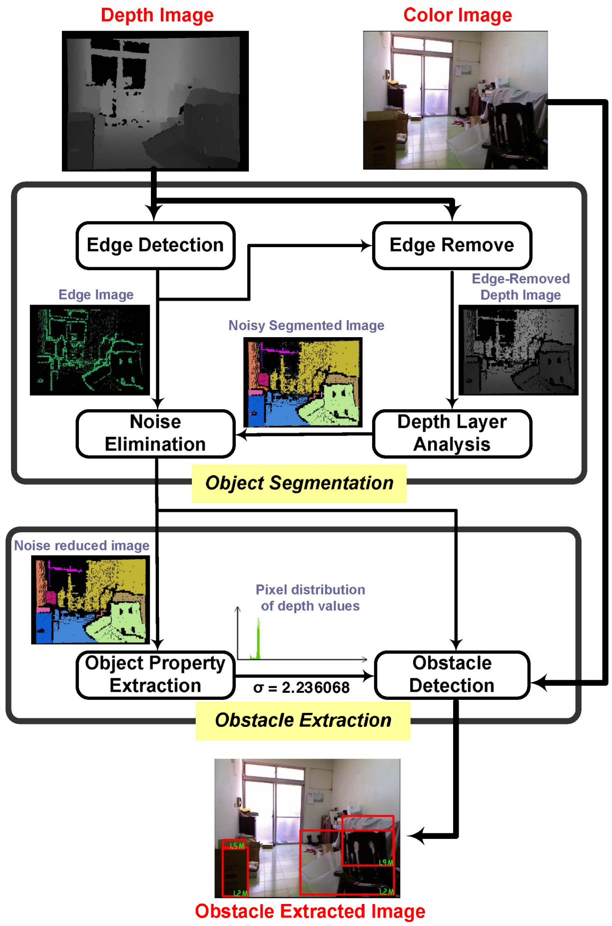


Fig. 1. System flow diagram

stage is obstacle extraction. Properties of detected objects are extracted to determine which ones belong to obstacles. Once obstacles are confirmed, they would be labeled with distance information far from the user. The system flow diagram is shown in Fig. 1.

A. 3D-sensor

The Time-of-Flight (TOF) camera has been widely used as the 3D-sensor recently. It provides each pixel with depth information in intensity values in real-time. In spite of physical limitations, it is reliable in general conditions. In this paper, we use TOF camera to generate depth images and color images. The 640x480 resolution video inputs are tested in this work.

B. Object Segmentation

The purpose of the first stage is to segment each depth image into several regions, which represent different objects. The discontinuous depth values in the depth image measured by the TOF camera form the boundaries of objects in the real environment. Therefore, extracting points with discontinuous depth values for edge detection is the first step. Let $I(x)$ is the depth value of pixel x. We define $m(x_c, x_i)$ as a binary variable, which represents whether the difference of depth values between the center x_c and its neighboring pixel x_i is large than threshold T_V or not. The representation is shown as follows.

$$m(x_c, x_i) = \begin{cases} 1 & \text{if } |I(x_c) - I(x_i)| > T_V, \\ 0 & \text{others,} \end{cases} \tag{1}$$

To find out pixels on the object boundaries, we define $E(x_c)$ as a binary value, representing whether x_c locates on one edge or not. Let $N(x_c)$ be the set of neighboring pixels of the center pixel x_c. For the center pixel x_c, once the number of its neighboring pixels, whose $m(x_c, x_i)$ is equal to 1, is larger than a given threshold T_N, the value of $E(x_c)$ would be set to 1. The equation is shown in Eq. 2.

$$E(x) = \begin{cases} 1 & \text{if } \sum\limits_{x_i \in N(x_c)} m(x_c, x_i) > T_N, \\ 0 & \text{others,} \end{cases} \tag{2}$$

After all edges in the depth map are found, we eliminate them so that the depth image could be segmented more accurately. The next step in object segmentation is depth layer analysis. Compared with Watershed and Histogram-Based algorithm in segmentation, the concept of Region-Growing [8] is a more suitable method to separate objects in different depth layers and spatial locations. This algorithm spreads several seeds in the image plane and merges their neighbors iteratively. However, this algorithm is seed-dependent and it would fail if the initial seeds not ideally spread on objects. To overcome this defect, we modified the approach. Let every pixel initially be set as a seed, denoted as s_x. These seeds would be merged into the same union A_i as their neighboring seeds $N(s_x)$ if the difference of their depth values is smaller than a threshold T. Then, multiple unions are generated. Let $A_{i(s_x)}$ represents the seed set which contains the same union index $i(s_x)$. For each iteration, because of the scan order for processing the region growing step, a seed may not be merged within the union containing its neighbors with the nearest depth value. After one iteration, some seeds of $N(s_x)$ may have depth values with difference smaller than T but still do not belong to $A_{i(s_x)}$. Thus, we define the set of seeds as W. The representation is shown below.

$$W = \{s_x | \exists s \in N(s_x) : [I(s) - I(s_x) < T] \Lambda [i(s) \neq i(s_x)]\} \tag{3}$$

In our approach, every seed in W would be merged again with the region that has the most nearest mean of depth values. The equation could be represented as follows.

$$s_x' = \underset{s \in N(s_x)}{\arg\min} \left| I(s_x) - \frac{\sum\limits_{s_n \in A_{i(s)}} I(s_n)}{\text{card}(A_{i(s)})} \right|, \tag{4}$$

Where the selected seed s_x' is a neighboring seed to s_x within the region that has the most nearest mean value of depth for all pixels inside. We iteratively renew the union index $i(s_x)$ until $W \subset \emptyset$.

The segmented image, however, is noisy because of the imperfect depth image measured by the TOF camera. Fortunately,

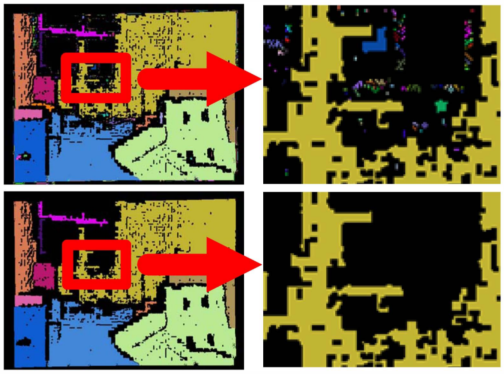
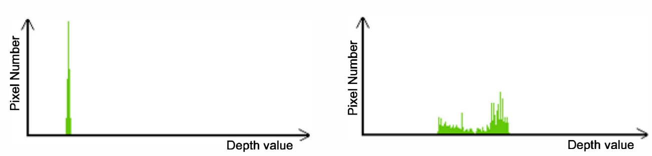


Fig. 2. Segmented images. The left frame of each pair is a complete image and the right frame of each pair is the ROI. The image pairs on the first row shows the noisy-segmented result. The second row image pairs presents the noise-reduced segmentation result.

noise usually has certain characteristics such as small number of pixels and sparse locations. The small regions with few pixel numbers and sparse locations near the edges usually represent the noise, which would be eliminated to guarantee the correct segmentation. Fig. 2 show the comparison between a noisy segmented frame and a noise-reduced frame.

C. Obstacle Extraction

The purpose of this stage is to extract obstacles by object proprieties, such as standard deviation and the mean of depth values. The term ”obstacle” is defined as moving and standing objects those appear within a certain alarm range, which is set as 1-2 meters far from the user in this paper. The alarm range could be adjusted according to different environments. Therefore, whenever obstacles occur within the alarm range, they will be captured and labeled with the estimated distance from the user.



(a) (b)

Fig. 3. Pixel distribution of depth values on the chair (a) and the floor (b).

Intuitively, obstacles can be extracted among the detected objects by calculating the median or mean of depth values. This simple method is indeed able to extract all obstacles, however, it will cause high false-alarm rate when detecting the floor, which has low mean, is not what we aim to capture. In order to solve this problem, the standard deviation of each object should be extracted to distinguish whether it is floor or not. Since pixel distribution of depth values of the floor is scattered, its standard deviation would be larger than other objects. Fig. 3 shows the pixel distribution of depth values on the chair (Fig. 3-a) and floor (Fig. 3-b). Then, the threshold of standard deviation is established to remove the floor. Therefore, false alarm rate could be successfully reduced.

III. Experimental results

Fig. 4. Obstacle detection within 1.5m (left) and 2m (right) alarm range.

Fig. 5. Indoor results. Robustness test in the house within 2m alarm range.

In our experiment, the TOF camera is used to capture depth images and RGB 3-channels multiple 8-bits color images. A user takes this camera and walks around. The two different alarm ranges, 1.5 and 2 meters, are tested in the same image in Fig. 4. All obstacles were detected successfully in the cluttered environment with light reflection. Fig. 5, Fig. 6 and Fig. 7 show more indoor results in different places. Fig. 8 presents the robustness of the system outdoors for 1 meter alarm range. In all the above environments, obstacles could be extracted. In addition, the proposed system also has good detection performance in dark environment. Fig. 9 shows the successful detection of a bicycle which appears abruptly in darkness. In fact, moving objects, such as bicycles and cars can be considered as moving obstacles. These obstacles also bring potential threats to the user.

The proposed system achieves high detection rate (DR) and low false alarm rate (FAR) indoors and outdoors. DR is defined as the total number of detected obstacles over the total number of existent obstacles in the alarm range. Our system achieves on average 96.1% DR in indoor cases and 93.7% DR in outdoor cases. FAR is defined as the total number of false detected obstacles over the total number of real existent obstacles in the alarm range. The proposed system achieves less than 5.33% FAR in indoor cases and 5.62% FAR in outdoor cases. The results are obtained by testing more than 10000 frames of sequences in all kinds of environments.

Fig. 6. Indoor results. Robustness test in the house within 1.5m alarm range.

Fig. 7. Indoor results. Robustness test in the convenience store within 1.5m alarm range.

Fig. 8. Outdoor results. Robustness test in the campus within 1m alarm range.

Fig. 9. Obstacle detection of bicycle in darkness. The Right frame of each image pair is the depth image and the left frame of each image pair is the RGB color image. The first row and the second row of image pairs are taken in the same place at different timestamps.

IV. CONCLUSION

In this paper, we propose robust depth-based obstacle detection system for visually-impaired aid applications. The proposed system can be used to help the blind prevent from dangers and can also be adopted to many emerging mobile applications. This system achieves on average 96.1% and 93.7% detection rate in the indoor and outdoor environment, respectively. Convincing results are demonstrated in diverse conditions. Environmental limitation is reduced by this framework. Even in darkness, the ability of detection of the proposed system can still work successfully. In conclusion, we establish a practical framework for visually-impaired aid applications in obstacle detection.

REFERENCES

[1] Arif, O.; Daley, W.; Vela, P.; Teizer, J.; Stewart, J.; , "Visual tracking and segmentation using Time-of-Flight sensor," *Image Processing* , 2010 17th IEEE International Conference on , vol., no., pp.2241-2244, 26-29 Sept. 2010

[2] Meers, S.; Ward, K.; , "Face Recognition Using a Time-of-Flight Camera," *Computer Graphics, Imaging and Visualization*, 2009. Sixth International Conference on , vol., no., pp.377-382, 11-14 Aug. 2009

[3] Gemeiner, P.; Jojic, P.; Vincze, M.; , "Selecting good corners for structure and motion recovery using a time-of-flight camera," *Intelligent Robots and Systems*, 2009. IEEE/RSJ International Conference on , vol., no., pp.5711-5716, 10-15 Oct. 2009

[4] Chen, L.; Guo, B.; Sun, W.; , "Obstacle Detection System for Visually Impaired People Based on Stereo Vision," *Genetic and Evolutionary Computing* , 2010 Fourth International Conference on , vol., no., pp.723-726, 13-15 Dec. 2010

[5] Yankun, Z.; Hong, C.; Weyrich, N.; , "A single camera based rear obstacle detection system," *Intelligent Vehicles Symposium*, 2011 IEEE , vol., no., pp.485-490, 5-9 June 2011

[6] Pantilie, C.D.; Bota, S.; Haller, I.; Nedevschi, S.; , "Real-time obstacle detection using dense stereo vision and dense optical flow," *Intelligent Computer Communication and Processing*, 2010 IEEE International Conference on , vol., no., pp.191-196, 26-28 Aug. 2010

[7] Jeong, P.; Nedevschi, S.; , "Obstacle detection based on the hybrid road plane under the weak calibration conditions," *Intelligent Vehicles Symposium*, 2008 IEEE , vol., no., pp.446-451, 4-6 June 2008

[8] Adams, R.; Bischof, L.; , "Seeded region growing," *Pattern Analysis and Machine Intelligence*, IEEE Transactions on , vol.16, no.6, pp.641-647, Jun 1994