Feature Recognition and Obstacle Detection for Drive Assistance in Indoor Environments

Chunhui Zheng
Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
czh23@uclive.ac.nz

Richard Green
Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
richard.green@canterbury.ac.nz

***Abstract*—The paper presents a robust indoor feature recognition and vision-based obstacle detection algorithm. A method is proposed using the fusion of colour features, edge map, range information and motion analysis to help the system interpret visual cues. The system is able to detect ground plane, drop-offs, stairs, open doors and obstacles, and is able to provide motion information. The results showing accurate indoor feature recognition and accurate distances to various detected indoor features, suggest that this proposed colour/edge/motion/depth approach would be useful as a navigation aid through doorways and hallways.**

Keywords-obstacle detection; feature recognition; autonomous navigation; structured light camera; Kinect

I. INTRODUCTION

Autonomous navigation for vehicles and mobile robots has been extensively researched in last two decades. Most autonomous navigation systems are based on different types of sensors such as infrared, sonar, laser range finders and visual sensors to provide obstacle detection, path planning and other navigation tasks [1] [2]. Vision-based navigation has received significant attention as more information can be retrieved by images than other types of sensors.

Obstacle and hazard detection is an essential task for path planning and other complex navigation tasks. When driving in an indoor environment, various potential hazards are present, including steps, upward stairs, walls, furniture, people, doorways and ramps. Table I lists several kinds of hazards in indoor environments [3].

This paper proposes a theoretical basis for vision-based obstacle detection, open-door detection, stairs and drop-off detection and motion calculation to support robust drive assistance in indoor environments. The proposed method exploits the fusion of colour features, edge map, range information and motion analysis to effectively analyse visual cues.

TABLE I. COMMON HAZARDS IN INDOOR ENVIRONMENTS

Hazards		**Examples**
Negative obstacles	Drop-offs	Downward stairs, steps
Positive obstacles	Static	Walls, furniture
	Dynamic	People, doors
	Transparent	Glass doors, glass walls
Overhangs		Table tops
Inclines		Wheelchair ramps
Narrow regions		Doorways, elevators

Depth values in our previous works were calculated by using a Bumblebee2 stereo camera [4]. However, a commercial low-cost structured light camera Kinect was launched by Microsoft, which provides more reliable depth measurements under a larger variety of indoor conditions than the Bumblebee stereo camera. Therefore, the Kinect camera is now being used in this research instead of the Bumblebee2 stereo camera for depth value generation.

The paper is structured as follows. Section 2 reports a brief overview of previous works for obstacle detection, and some background information of the Kinect camera. Section 3 explains the methodology to conduct this research. Section 4 presents the results of the experiments.This paper concludes with the limitations of the proposed system and an outline of future research.

II. BACKGROUND

A. *Obstacle Detection Overview*

Various vision-based obstacle detection approaches have been proposed in the literature, which can be classified into one of the following three categories: 1) *knowledge-based*, 2) *motion-based*, and 3) *depth-sensor-based* method.

1) Knowledge-based method: Exploits a prior knowledge of the obstacle. Features such as colour, edge, shape, textures and so on are used to detect obstacles [5] [6] [7]. Objects with different appearance from the ground may be classified as obstacles. Knowledge-based obstacle detection methods can be efficient in simple environments without clutter, but can easily fail when the background environment contains many similar colours, different textures and many geometric lines and edges [8]. Moreover, previous work on knowledge-based obstacle detection shows that this kind of approach is unable to detect overhanging obstacles such as the edge of a table.

2) Motion-based method: Motion information is extracted from successive images, and it can be calculated from optical flow. Optical flow describes the motion with vectors at feature points in a captured image, and it can be used for calculating the time to contact with a surface. Some motion-based obstacle detection algorithms recover the depth information of the environment [9] [10] [11]. Others extract 2D information instead of 3D reconstruction [12] [13]. Motion-based obstacle detection has several weaknesses. Sun et al. [8] enumerated three factors that can affect the computation of motion information. First, significant pixel displacement between successive images by fast movement of the camera can cause errors to optical flow calculation. Second, lack of texture in the images can lead to unreliable motion information. Third, shocks and vibration of the camera can also influence the accuracy of motion information. Moreover, motion-based obstacle detection cannot be used to detect static obstacles with no motion.

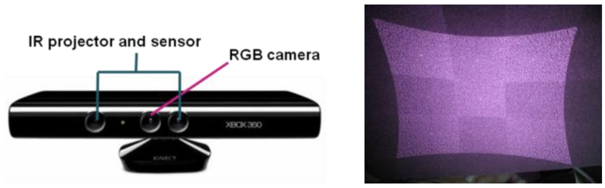
3) Depth-sensor-based method: Stereo vision cameras and structured light cameras can recover the depth information in images. Many depth-sensor-based obstacle detection methods are based on ground plane estimation [14] [15]. Compared with motion-based methods, depth-sensor-based obstacle detection has the advantage of depth information being derived without prior knowledge of the scene, and it is more accurate and less sensitive to the environmental changes. However, depth sensors alone cannot provide motion information of the camera or other dynamic objects.

In order to avoid problems with one single method, current vision-based obstacle detection approaches often combine multiple algorithms [3] [16].

B. Structured Light Camera

The basic process of a structured light camera is to project a known light pattern onto a scene. The structured light pattern appears distorted after it reaches the objects in the scene. The depth information of those objects can be retrieved according to the scale of distortion.

The Kinect camera combines information from a standard RGB camera with an infrared based depth sensor as shown in Figure 1(a). A structured infrared pattern of dots in Figure 1(b) is projected on to the environment and viewed by an infrared camera. The level of distortion in this pattern is used to calculate the distance from the camera to objects.



(a) (b)

Figure 1. Kinect camera and its IR pattern. (a) Kinect camera. (b) Structured infrared pattern projected by Kinect camera.

Table II shows the comparison between the PointGrey Bumblebee 2 stereo camera and the Microsoft Kinect camera [17]. The Kinect camera has a reasonable resolution, a good working range and a low price. Moreover, the depth calculation of the stereo camera is performed on the host machine and so requires more computational cost, while the Kinect camera calculates the depth value directly using its built-in processor.

TABLE II. COMPARISON BETWEEN BUMBLEBEE STEREO CAMERA AND KINECT CAMERA

	Bumblebee2 camera	**Kinect camera**
Frame size	1280×960	640×480
Maximum frame rate	15 fps	30 fps
Working range	0.5-4.5m	1.2-3.5m
Market price	$NZ 3000	$NZ 220

Images in Figure 2 are results from the Bumbleebee2 stereo camera, illustrating the difficulty obtaining depth information in non-textured areas. By comparison, the Kinect camera projects an infrared light pattern to cover the full scene and so retrieves depth information for almost every pixel as shown in Figure 3.

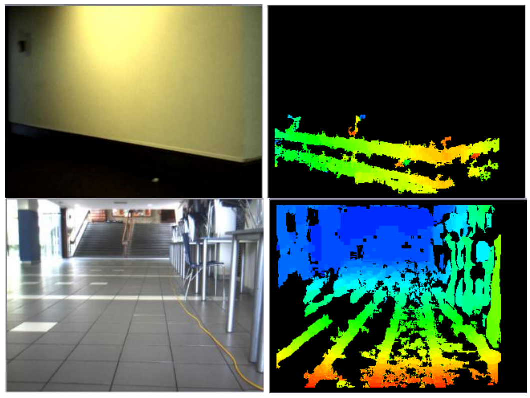


Figure 2. Depth image generated by Bumblebee2 stereo camera.

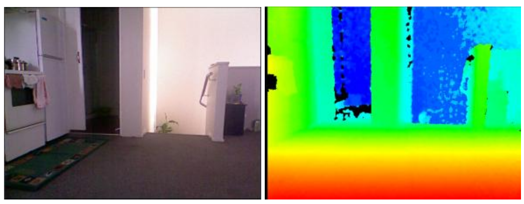


Figure 3. Depth image generated by Microsoft Kinect camera.

III. METHODS

This paper proposes a method that combines colour, edge, motion information and depth information to enable effectively analysing visual cues.

A. Depth Generation

Both depth and RGB images are obtained from the Kinect camera using the driver and library supported by OpenNI [18]. With the Kinect camera, the depth and RGB images are captured by two different cameras from two different viewpoints, and so it is necessary to align depth pixels with the corresponding RGB pixels. The process of image registration for the Kinect camera can be found in [19]. Figure 4 shows the 3D point cloud data generated by the Kinect camera that contains RGB values for each point.

Figure 4. 3D Point cloud.

B. Ground Plane Detection

Ground plane estimation is a prerequisite for obstacle detection. The RANSAC plane fitting [20] is used to find ground plane in the 3D space. The procedure is the following:

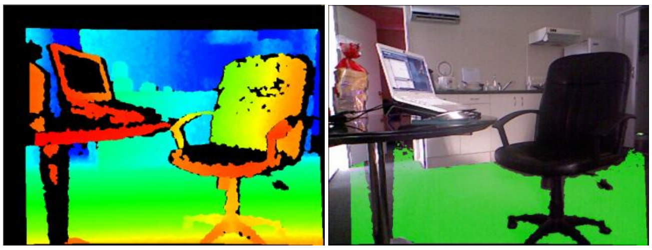
1) Randomly select 300 points within the depth image.

2) From the selected 300 points, randomly choose three points to fit a plane $Ax + By + Cz + D = 0$. Check the remaining 297 points whether fit the estimated plane equation, and calculate the number of fitting points.

3) Repeat step 2 until the number of fitting points is greater than a specified threshold. The ground plane is found.

4) If the number of fitting points is never greater than the threshold, the ground plane is plane with the largest number of fitting points.

Figure 5 shows a result of ground plane detection using the Kinect camera.



(a) (b)

Figure 5. Ground plane detection. (a) Depth map. (b) Result of ground plane detection.

C. Frame of Reference

To obtain a real world representation, it is necessary to know the camera placement. As the ground plane equation in camera view coordinates is already known, the approximate height of the camera above the floor can be estimated. As a result, the 3D position of each pixel is converted from camera view coordinates into world coordinates.

D. Obstacle Detection

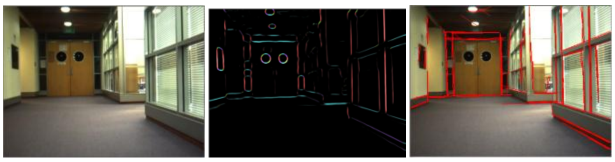
For the obstacle detection, firstly, the ground surface is removed from the image after the ground plane has been observed. The remaining data is divided into connected components in image and depth map space according to their 3D positions in world coordinates. The objects that lie directly in front of the camera could influence the movements. Those pixels outside the range (x_{min}, x_{max}) and (y_{min}, y_{max}) are removed from the image. The distance to each detected obstacle is calculated and compared. Finally, the closest obstacle can be located. Figure 6 shows some results of obstacle detection.

(a) (b)

Figure 6. Detecting the nearest obstacle: (a) locating stairs (b) locating table and chairs

E. Edge and Line Extraction

Reliable line extraction method is essential for doorway, stairs and drop-offs detection. The standard deviation ridge detector [21] is exploited to detect boundaries in the environment for this research. It is a real-time boundary detector that can retrieve more useful data than an edge detector, such as the Canny edge detector. Next, an edge thinning process is carried out to reduce the thick edges to one pixel wide edge elements. Finally, the Hough Transform is used to find straight lines in the edge images (Figure 7).



(a) (b) (c)

Figure 7. Standard deviation ridge detector and the Hough Transform: (a) original image (b) boundaries found by the standard deviation ridge detector (c) straight lines found by the Hough transform

F. *Drop-off and Stairs Detection*

For drop-off recognition, lines are extracted from the image first. Then, depth discontinuities are examined at each horizontal line. If the depth value of the region above the line is much greater than the value below the line, the drop off is found. The stairs recognition is looking for a set of parallel lines; and then checks if the depth of each stair line changes gradually. The numbers in Figure 8 show the distance to the detected stairs or drop-off in meters.

(a) (b)

Figure 8. Drop-offs and stairs detection. (a) Drop-off line detected. (b) Stairs lines detected.

G. *Open door Detection*

Since the Kinect camera provides more reliable depth measurements than a stereo camera for close range indoors, a more useful depth discontinuity map can be generated as shown in Figure 9 (b). The white pixels show the positions where the depth discontinuity is generated by a value greater than 0.5 meters. The Hough Transform is then used to find straight lines in the depth discontinuity map. In a depth discontinuity map, when there are two vertical lines connected with a horizontal line at top, such an area is normally represented as an open door. Using these depth discontinuity maps, the height and width of a typical open door can be calculated.

(a) (b) (c)

Figure 9. Open door detection. (a) Depth map. (b) Depth discontinuity map. (c) Detected straight lines in depth discontinuity map.

H. *Motion Calculation*

The motion information of the camera is calculated using optical flow to show the direction of camera movement along a hallway. Feature points are selected for each frame by detecting corners. Kanade-Lucas optical flow is used to track these feature points. Flow vectors are calculated by locating the positions of each feature point in the previous and current frame. However, some outliers may also be generated as shown in Figure 10 (a). In order to prune outliers, statistical measures, such as lower, upper and inter quartile values are used. As a result, outliers with uncharacteristic magnitude and slope are eliminated (Figure 10 (b)).

For each flow vector, the 3D positions of each flow point in previous frames and the current frame are retrieved by integrating with depth information. The optical flow vectors are divided into 9 regions as shown in Figure 10 (b). The average distance from the camera to each region is calculated separately. These distances are then used to calculate time to impact and 6 DOF pose movement of the camera.

(a) (b)

Figure 10. Optical flow vectors. (a) Raw optical flow vectors generated by the moving camera. (b) Optical flow vectors after outlier removal.

IV. Results

A. *Evaluation of Indoor Feature Recognition*

The proposed algorithms for drop-offs, stairs and open-door recognition are evaluated. Four video data sets with different indoor environments were collected to evaluate the indoor feature detection. Table III shows the results of this indoor feature detection. It is clear that both the drop-off and stairs recognition are accurate, but the open door recognition only has 69% true positive results.

TABLE III. Evaluation of Indoor Feature Detection

	True Positive	**False Positive**	**False Negative**
Drop-offs detection	94%	15%	6%
Stairs detection	96%	8%	4%
Open-door detection	69%	-	31%

B. *Evaluation of Distance Measurement*

The accuracy of measuring distances to the detected indoor features is also evaluated where Figure 11 graphs results of these distance measurements. It shows that by using Bumblebee stereo camera, within 5 meters, the distance offset is within 0.3 meters (6%). However, the accuracy is not improved by using the Kinect camera.

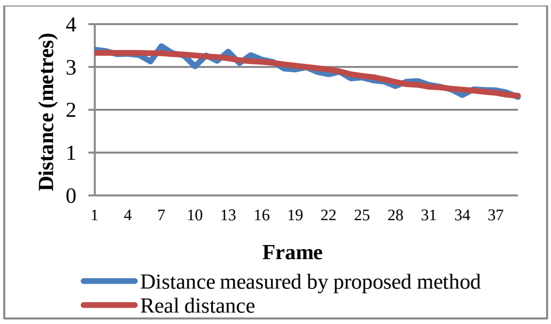


Figure 11. Results of measuring distance to the detected indoor features.

Several factors influence the accuracy of distance measurements. First, our results depend on the depth measurements provided by depth sensor. For example, the accuracy of depth measurement by Kinect camera can be influenced by other IR sources such as direct sunlight (although this research is intended for indoor use).. Second, reflective or transparent surfaces can introduce significant depth errors using both types of cameras. Third, line extraction methods may divide an actual corridor line into several line segments which could potentially cause false feature recognition.

V. CONCLUSION AND FUTURE WORKS

In conclusion, the project proposed a method combines different visual cues to detect indoor features, avoid obstacles and calculate movements. The results suggest that the system is able to detect ground plane, open doors, drop-offs and stairs. Generally, the results of indoor feature recognition and distance measurement to the detected indoor features are accurate.

Future work will focus on dynamic obstacle detection using optical flow and depth information. The relative speed of dynamic objects with the moving camera with time-to-impact will be more accurately estimated.

Current limitations due to large illumination variation over 24 hours and camera shaking will be reduced by fusion with other types of sensors.

REFERENCES

[1] G. D. Castillo, S. Skaar, A. Cardenas, and L. Fehr, "A sonar approach to obstacle detection for a vision-based autonomous wheelchair," Robotics and Autonomous Systems, vol. 54(12), pp. 967-981, 2006.

[2] H. Surmann, A. Nuchter, and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," Robotics and Autonomous Systems, vol. 45(3-4), pp. 181-198, 2003.

[3] A. Murarka, M. Sridharan, and B. Kuipers, "Detecting obstacles and drop-offs using stereo and motion cues for safe local motion," IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2008, pp. 702-708.

[4] C. Zheng and R. Green, "Vision-based autonomous navigation in indoor environments," 25th International Conference of Image and Vision Computing New Zealand (IVCNZ), New Zealand, Queenstown 2010.

[5] I. Ulrich and I. Nourbakhsh, "Appearance-based obstacle detection with monocular color vision," Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 866 – 871, 2000.

[6] N. Tada, K. Murata, T. Saitoh, T. Osaki, and R. Konishi, "Monocular vision based indoor mobile robot," The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2008), pp. 41-44, 2008.

[7] C. N. Viet and I. Marshall, "An efficient obstacle detection algorithm using colour and texture," Proceedings of World Academy of Science Engineering and Technology, vol. 60, pp. 123-128, 2009.

[8] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(5), pp. 694-711, 2006.

[9] C. Demonceaux, A. Potelle, and D. Kachi-Akkouche, "Obstacle detection in a road scene based on motion analysis," IEEE Transactions on Vehicular Technology, vol. 53(6), pp. 1649-1656, 2004.

[10] T. Naito, T. Ito, and Y. Kaneda, "The obstacle detection method using optical flow estimation at the edge image," 2007 IEEE Intelligent Vehicles Symposium, pp. 817-822, 2007.

[11] X. Wang, K. Ban, and K. Ishii, "Estimation of mobile robot ego-motion and obstacle depth detection by using optical flow," IEEE International Conference on Systems, Man, and Cybernetics, pp. 1770-1775, 2009.

[12] M. Sarcinelli-Filho, H. J. A. Schneebeli, and E. M. O. Caldeira, "Using optical flow to control mobile robot navigation," Proceedings of the 15th World Congress on Automatic Control, 2002.

[13] Y. Shen, X. Du, and J. Liu, "Monocular vision based obstacle detection for robot navigation in unstructured environment," Lecture Notes in Computer Science, Advances in Neural Networks, ISNN 2007, vol. 4491, pp. 714-722, 2007.

[14] D. Burschkal, S. Lee, and G. Hager, "Stereo-based obstacle avoidance in indoor environments with active sensor re-calibration," IEEE International Conference on Robotics and Automation, vol. 2, pp. 2066-2072, 2002.

[15] R. Cucchiara, E. Perini, and G. Pistoni, "Efficient stereo vision for obstacle detection and AGV navigation," 14th International Conference on Image Analysis and Processing, pp. 291-296, 2007.

[16] S. Fazli, H. M. Dehnavi, and P. Moallem, "A robust obstacle detection method in highly textured environments using stereo vision," Second International Conference on Machine Vision (IEEE), pp. 97-100, 2009.

[17] F. Pece, J. Kautz, and T. Weyrich, "Three depth-camera technologies compared," First BEAMING Workshop 2011, Barcelona, 2011.

[18] OpenNI. Retrieved from http://www.openni.org.

[19] N. Burrus, "Kinect calibration," Retrieved from http://nicolas.burrus.name/index.php/Research/KinectCalibration.

[20] Q. Yu, H. Araujo, H. Wang, "A stereovision method for obstacle detection and tracking in non-flat urban environments", Autonomous Robots, vol. 19(2), pp. 141-157, 2005.

[21] R. Hidayat and R. Green, "Real-time texture boundary detection from ridges in the standard deviation space," British Machine Vision Conference 2009.