The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Robust Ground Plane Detection for Obstacle Avoidance of Mobile Robots Using a Monocular Camera

Chia-How Lin, Sin-Yi Jiang, Yueh-Ju Pu and Kai-Tai Song, IEEE Member

Abstract—**This paper presents a vision-based obstacle avoidance design using a monocular camera onboard a mobile robot. An image processing procedure is developed to estimate distances between the robot and obstacles based-on inverse perspective transformation (IPT) in image plane. A robust image processing solution is proposed to detect and segment navigatable ground plane area within the camera view. The proposed method integrate robust feature matching with adaptive color segmentation for plane estimation and tracking to cope with variations in illumination and camera view. After IPT and ground region segmentation, a result similar to the occupancy grid map is obtained for mobile robot obstacle avoidance and navigation. Practical experimental results of a wheeled mobile robot show that the proposed imaging system successfully estimates distance of objects and avoid obstacles in an indoor environment.**

I. Introduction

Detection and localization of obstacles are essential for mobile robot to navigate in an environment. Ultrasonic sensors and laser scanners are most used for environment detection and obstacle avoidance of mobile robots. However, ultrasonic sensors are relatively slow and suffer from spectral reflections and poor angular resolution. Laser scanners can provide much better resolution with higher scanning speed, but they are quite expensive for practical applications. Beside these drawbacks, most range-based systems provide only distance information of the environment, making it difficult to distinguish between obstacles and objects, with which the robot needs to interact.

On the other hand, vision-based systems provide rich information from the environment. They have become promising alternatives considering availability of low-cost image sensors and high-performance processors. Among various vision-based approaches, omni-directional and stereo vision systems have been widely used for environment detection, due to their wide view angles and the ability to estimate depth information. Their hardware structure, however, raise both the price and system complexity. Monocular vision-based approaches, on the other hand, do not require special camera configuration and are more suitable for various low-cost robotic applications.

Manuscript received March 10, 2010. This work was supported in part by the National Science Council of Taiwan, ROC under grant NSC 95-2218-E-009-178 and 95-2218-E-009-008 and the Ministry of Economic Affairs under grant 97-EC-17-A-04-S1-054.

C.H. Lin, S.Y. Jiang and K.T. Song are with Institute of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, ROC. (e-mail:{ jotarun.ece87@; sinch.ece97g ;ktsong@mail.}nctu.edu.tw). Yueh-Ju Pu is with Mechanical and Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C. (e-mail: YJPu@itri.org.tw)

Most monocular vision-based approaches to obstacle detection need priory knowledge of the scene or obstacles. For instance, color cue is commonly used to segment obstacles and non-obstacles in a structured environment [1]. In [2], moving object tracking techniques are proposed to find obstacles which have significant difference in motion. However, these approaches can only be applied to limited scenes. Optical flow, on the other hand, is a more general approach, which can be used to estimate time-to-collision of obstacles, or the environmental structure from motion [3, 4]. The computational load of optical flow is rather high, but the robustness against disturbance is still questionable. Recently, some researchers have proposed to use machine learning approaches to obtain depth or planar structure with monocular cues from single view [5,6]. Their results are quite impressive in many cases, but the performance in natural environments is still unpredictable.

Due to the difficulty in generalized obstacle detection, many researchers propose to resolve the problem of floor or ground detection based on planar property of road or indoor environment [7-12]. Liang *et al.* [7] proposed to use reciprocal-polar (RP) image rectification and ground plane segmentation by sinusoidal model fitting in RP-space to segment the ground plane from a mobile robot's visual field of view. They then measure the height of off-ground plane features above the mobile robot's ground plane. Zhou *et al.* [8, 9] derived constraints that the homography of the ground plane must satisfy, and then used these constraints to design algorithms for detecting the ground plane by assuming that the camera is fixed on the robot platform and can at most rotate horizontally. These methods, however, can only distinguish whether feature points are located on the ground or not. This information is insufficient for obstacle avoidance.

For pixel level obstacles representation, warping or inverse perspective transform (IPT) technique has been widely used in both stereo-based and monocular-based approaches [10-14]. This technique uses ground plane homography to warp one of the images. Only pixels correspond to points on the ground plane will match in both images. This discrepancy can then be detected by image subtraction and thresholding. For monocular-based cases, one needs to estimate the homography between monocular image sequences. Yamaguchi *et al*.[12] estimate relevant parameters for the camera motion as well as the 3D road plane from correspondence points between successive images. The road region is determined by estimating the ground homography matrix and warping the image of previous frames. Although subtraction is a simple and effective method to detect pixel-wise floor regions, it still suffers from some limitations [13]. For instance, due to the nature of subtraction method, textureless obstacles cannot be detected since no discrepancy can be observed even under erroneous projection. Similarly, part of the obstacles may be projected into the floor region and cause some floor areas to be recognized as obstacles.

This paper aims to find a more robust method to detect ground plane and estimate range information for mobile robot navigation control. The main concept is to integrate multiple visual cues to avoid limitations in prior works and achieve robustness against illumination and camera view changes. First, Speeded Up Robust Features (SURF) [15] is adopted for feature extraction and matching to increase both robustness and speed in homography estimation step. Further, a key-frame selection criterion is proposed to guarantee the estimation is proper and avoid the virtual plane problem caused by zero/near-zero camera translations. To observe ground region pixel-wise, we propose an adaptive self-segmentation and tracking algorithm to find the ground region according to both the feature points and color segments. After IPT and ground region segmentation, results similar to an occupancy grid map can be obtained. The robot can then use distance information to avoid obstacles. The constructed map itself can also be used for many other robot control purposes.

The rest of this paper is organized as follows. Section II presents the design concept and architecture of the overall system. The proposed ground feature classification algorithms are described in Section III. Section IV describes the ground region segmentation method. Several interesting experimental results are presented in Section V. The contribution of the paper in summarized in Section VI.

II. Proposed System Architecture

The basic idea of estimating distance using monocular image sequence is to apply IPT on those images. In general, IPT maps three-dimensional points to image pixels if these points all belong to a known plane. We assume that the ground is planar, the robot moves parallel to the ground, and the camera is fixed on the robot. The distance between the camera and the ground plane is then fixed. The IPT of the ground (height =0) can therefore be determined beforehand, since it is also fixed. By applying IPT to the acquired images, the camera can be virtually rotated to the downward-looking pose. The projected images can then provide the relative position of each pixel on the ground in world coordinate [10, 11]. Off-ground objects, on the other hand, will be distorted after the transformation. However, the transformation itself is insufficient to determine whether pixels belong to ground. Therefore, the key idea of the proposed system is to robustly segment the ground region from monocular sequence via multiple visual clues.

Fig. 1 illustrates the architecture of the proposed system. Two parallel procedures will be launched simultaneously after the image preprocessing step. In the ground feature detection phase, the system extracts all the features in current image at time step *t*, and matches them to features observed in previous N frames, According to the correspondence and pre-determined ground features, ground homography and features on the ground in the newly acquired image can be found. The system then uses the homography matrices among consecutive frames to classify all other features as on the ground or not. Meanwhile, the images are segmented into patches according to color homogeneity. These patches can be classified into ground or off-ground region by comparing

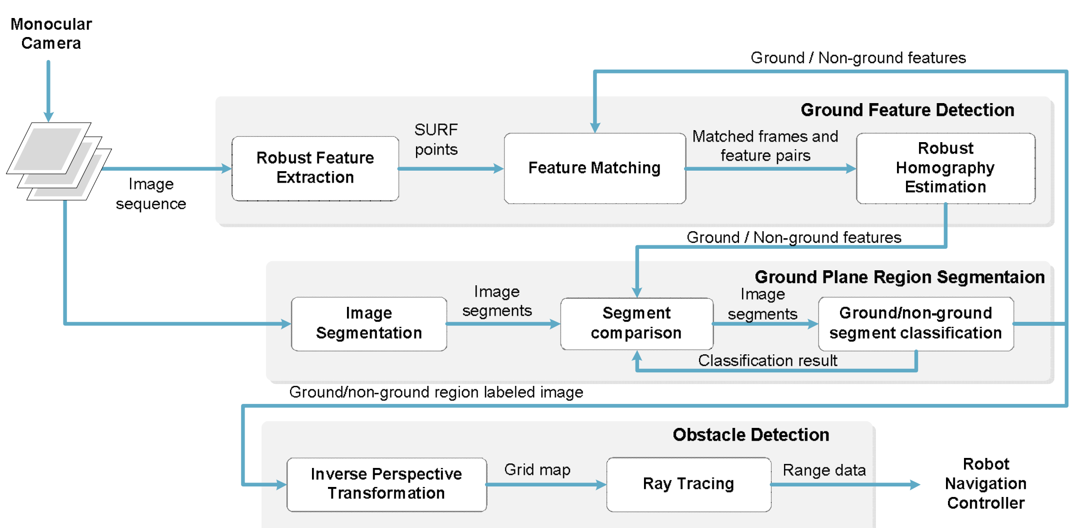


Fig.1. The architecture of the proposed vision-based obstacle avoidance design

with warped images. Once the ground and off-ground areas are isolated, the range information can be obtained by applying IPT and ray tracing. The detailed algorithm is described in following Sections.

III. Robust Ground Feature Detection

A good feature extractor should have invariance to rotation, scale, illumination variation and image noise, which lead to robust corresponding points matching results. The Harris corner detector has been widely used in many previous works [7-10]. Its robustness, however, is limited in practical applications. Recently, there have been many advances in scale/rotation-invariant extractors, such as Scale-Invariant Feature Transform (SIFT)[16]. An array of image gradient histograms is used as a descriptor instead of a raw image patch. However, in many cases, the execution speed of SIFT is not suitable for real-time applications. Speeded Up Robust Features (SURF) is adopted in this designs for its superior execution speed. While the robustness can be preserved, SURF is about three times faster than SIFT in feature extraction step. SURF also provides an indexing value for fast matching purpose. Nevertheless, these methods may still fail under large motion blur or abrupt changes in view angle. The speed of the camera motion is therefore limited in this study.

A. Ground Feature Classification

Once features are found and matched, the next is to determine whether these feature points are on the ground or not. Consider a ground plane projected into two views. With the pinhole camera model, two views of the same plane are related by a unique homography [17]. That is, for a plane $\Pi = [\mathbf{n}^T\ 1]^T$, the rays corresponding to a point p in the image **I** and its corresponding point p' in the image **I'** meet the plane at a point x_Π, as shown in Fig. 2. Therefore, if a set of coplanar points p_i with homogeneous image coordinate $(x_i, y_i, 1)^T$ are found, and their correspondences $\{p_i \leftrightarrow p'_i\}$ in two images are also found, there exist a 3 by 3 homography matrix **H** such that

$$p'_i \rightarrow p_i = \mathbf{H}p'_i \tag{1}$$

$$\mathbf{H} = \alpha\mathbf{K}(\mathbf{R} + \frac{1}{d}\mathbf{t}\mathbf{n}^T)\mathbf{K}^{-1} \tag{2}$$

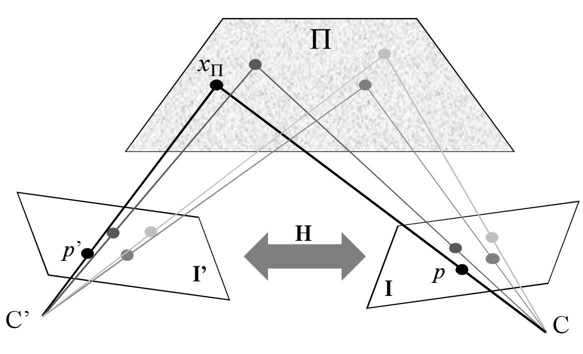


Fig.2. Homography of coplanar points.

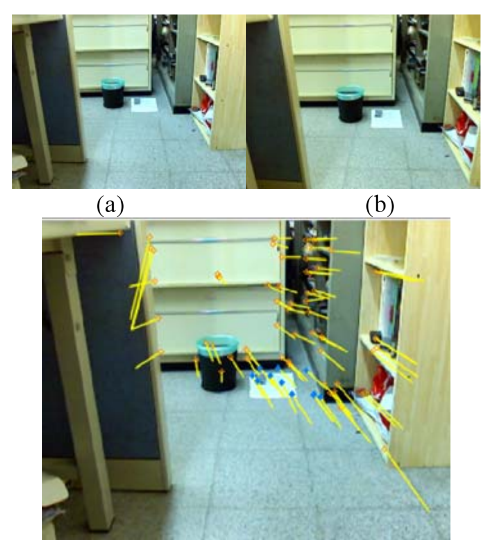
where **K** is the intrinsic camera matrix, **R** is the rotation matrix, **t** is the translation vector, α is the scale factor, and d is the distance between the plane and camera. To determine **H**, four non-degenerated corresponding points are required since each point correspondence provides two independent constraints. In the proposed system, the features on the ground are initially determined, i.e., a subset of p_i is known as on the ground plane. The homography of the ground plane can thus be determined initially. Considering there might be some matching error, RANdom SAmple Consensus (RANSAC) is applied to eliminate outliers and robustly determine **H**. Note that **H** matrix is only relevant to the plane relationship among two frames. Therefore, it will not be diverged after a long run.

The rest of the corresponding points can then be checked if it is on the ground by using the back projection technique,

$$p'_i \in \begin{cases} \text{ground, if } \|p'_i - \mathbf{H}p_i\| < \textit{threshold} \\ \text{off - ground}, \textit{otherwise} \end{cases}. \tag{3}$$

Fig. 3 shows an example of ground feature classification. We observe that there is no corresponding feature on the floor. This is in fact a case when the pattern on the ground floor is not robust enough. However, the goal here is not to find as many features as possible, but to determine **H.** The number of features is therefore not a critical issue.

On homography estimation, in the case of near-zero camera translation, i.e., $\mathbf{t} \cong 0$ in (2), no information on coplanarity can be inferred since the plane normal **n** can be arbitrary [18]. This case must be determined since all the points will be determined as ground. The simplest test is to determine whether all observed points conform to the same



(a) (b)

(c)

Fig.3. An example of ground feature detection. (a) and (b) are images taken in different views. (c) is the classification result of the matched features observed from (a) and (b). The features with blue label indicate features on the ground, while the features with orange label are features off the ground.

homography. This test will fail only when there is just a single scene plane visible, or when camera intrinsic parameters are changed. Both of these cases are rare and can be avoided.

To further improve the robustness, the homography estimation step will not be performed among consecutive frames straight forwardly, but only when the displacement is above a threshold. This step helps to reduce quantization error when the distance between p_i and p'_i is too small.

IV. GROUND SEGMENTATION

The quality of segmentation result depends not only on the algorithm used, but also the targeted application. In our design, the segmentation method should be able to distinct objects from image frames in a fast, robust, unsupervised, and adaptive way under various environments. In the proposed system, an image frame is segmented in HSI (Hue-Saturation-Intensity) color space using multi-scale Mean-shift algorithm.

A. Multi-scale Mean-shift Clustering

Mean-shift algorithm [19] is a nonparametric, iterative clustering technique which does not require prior knowledge of the number of clusters and constrain the shape of the clusters. It is suitable for unsupervised color segmentation. Although Mean-shift is nonparametric, the bandwidth of the kernel still needs to be determined. In general, the image is under-segmented when the bandwidth is too large, and over-segmented when the bandwidth is too small. In this work, proper bandwidth is determined dynamically, according to the frequency analysis results of the image. For instance, a clustered image such as crowds often indicates larger energy in high frequency, and requires a smaller bandwidth value. On the contrary, a simple image such as a white wall, will give a larger bandwidth value.

In order to further boost up the speed for real-time applications, images are scaled down beforehand. Small objects may thus be neglected. Therefore, the proposed algorithm estimates the purity of each semi-object with original resolution, and segments again if possible. This method achieves 5-20 times faster with better segmentation quality than directly process the original scale.

Clustering under different color space may bring extremely different results. Experiments show that clustering under HSI color space performs more stable than other color space. However, hue is unstable under low intensity and saturation. Therefore, if a pixel has a saturation value<0.1 or an intensity value <0.1, its hue value will be set to 2, which indicates that this pixel is colorless.

B. Ground Plane Labeling

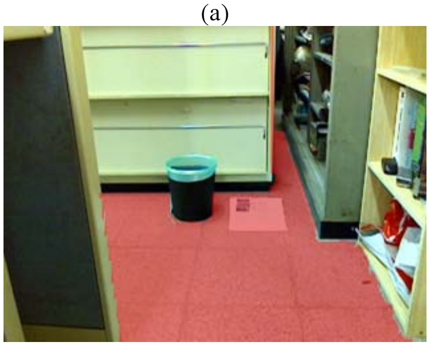
After ground/off-ground features are determined, the system needs to classify the ground region in a pixel-wise manner. Many previous works warp image by using the homography matrix and calculating Sum of Absolute Differences (SAD) between the warped image and current image. However, it is difficult to determine a proper value of threshold since the SAD value is correlated to the environment. Further, homogenous obstacles may be neglected, as mentioned earlier. Therefore, the proposed system determines if the region is on the ground according to the displacement and the distortion of each segment separately. To achieve this, corresponding segments are first found using both SURF features and its color distribution. Similar techniques have been used in stereo vision when estimating disparity maps [20]. Both static and moving obstacles can be observed. Those segments which fail to be matched will be labeled as undetermined. Note that when the robot stops or its motion is a pure rotation, this method cannot be used to determine the ground region.

Fig. 4 shows both the result of segmentation and ground plane labeling based on the images in Fig. 3. Note that in the image a piece of paper on the ground is classified as ground as expected.

V. EXPERIMENTAL RESULTS

The proposed system has been implemented on an experimental mobile robot. As shown in Fig. 5, the robot is equipped with a USB webcam on a two DOF pan-tilt head. An Industrial PC (IPC) is used for image processing and robot control. All the motors are velocity-servoed by using DSP-based motion control boards. In order to make the system more responsive, the acquired image is down sampled to 320×240 pixels. Images are captured at the beginning of each procedure in the vision system. The complete image processing will take 100 to 300ms, depending on the extracted features and the complexity of input images.

(a)



(b)

Fig.4. (a) Color segmentation result and (b) ground region labeling result based on the images in Fig.3.

A. Calibration Experiment of Distance Measurement

In this experiment, the distance measurement was calibrated using the developed system. The experiment was performed with static obstacles on the ground, while the robot is controlled to stand from 1.6m to 6m from the obstacle. The estimated distance is compared with the ground truth values. The IPT homography was initially estimated by 4 known points on the ground within 3m. Table I shows the experimental results. The error is within 3% in 5m range. The reasons of the error are mainly due to the imperfect result of segmentation and the quantization error in IPT. These errors are especially significant when the object is far away and cannot be distinguished due to limited resolution of the image. This accuracy is comparable to Hokuyo laser scanner [21]. Note that the minimal and maximum sensing ranges are both related to the height and the tilt angle of the camera.

B. Experiment of Classification Rate Using a Synthetic Environment

In order to test the proposed system under different environments, a synthetic room is used to examine the classification rate with ground truth. In each trail of the experiment, the path of the robot is randomly generated to observe 10 consecutive views, as shown in Fig. 6. The texture of the environment also changes randomly in each trail. The experiment after 100 trails shows that the false detection rate of features is lower than 1%. The experiment also shows that false detection area of ground regions are within 3%.

C. Experiment of Mobile Robot Obstacle Avoidance

The aim of this experiment is to test if the proposed system can be used to replace a laser scanner and accomplish a simple autonomous navigation task. In this experiment, an autonomous navigation scheme of authors' previous work was adopted [22]. The navigation task is achieved by behavior fusion of two navigation behaviors, namely obstacle avoidance and goal seeking. The behavior of goal seeking is treated as trying to move towards the direction of the target. The obstacle avoidance behavior is to move along the direction without colliding into any obstacles on the way. The main challenge of the proposed vision-based system is that the original laser scanner provides a view angle of at least 180-degree-wide., which the camera, however, can provide 60-degree-wide at most, due to its field of view. Therefore, the robot might not avoid obstacles appeared in the blind zone.

TABLE I
EXPERIMENTAL RESULT FOR DISTANCE ESTIMATION OF OBSTACLE

Ground truth(m)	Estimated distance(pixels)	Estimated distance(m)	Error(m)	Percentage error (%)
1.6	45	1.63	0.03	1.94
2	100	1.97	0.03	1.41
2.4	133	2.42	0.02	0.66
2.8	158	2.87	0.07	2.57
3.2	176	3.26	0.06	2.01
3.6	193	3.68	0.08	2.32
4	206	4.04	0.04	0.92
4.4	220	4.45	0.05	1.09
4.8	232	4.83	0.03	0.55
5.2	238	5.02	0.18	3.38
5.6	243	5.19	0.41	7.25
6	247	5.33	0.67	11.1

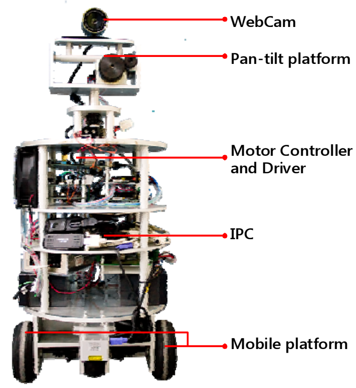
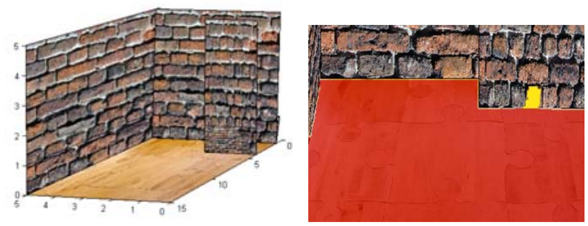


Fig.5. The mobile robot used in the experiments



(a) ((b)

Fig.6. The synthetic room experiment. (a) The structure of the environment. (b) A classification result. The red zone indicates the region correctly labeled as ground, while the yellow region indicates false detection.

The experimental results are shown in Fig. 7 and Fig. 8. The recorded trajectory in Fig. 7 shows that the robot successfully avoided several obstacles and reached the destination. Various obstacles with different colors and sizes were detected and avoided. The robot turned to the left after it detected the tool boxes, as shown in Fig.7 (c) and Fig.8 (c). Later, the robot turned right since it detected the chairs and desks, as shown in Fig.7 (d) and Fig.8 (d). Video clips and supplemental materials of these experiments can be found in http://isci.cn.nctu.edu.tw/Video/IROS2010/.

VI. CONCLUSION

A monocular vision-based robot navigation system has been developed by combining IPT and ground region segmentation. A local grid map can be built with single camera for robot navigation. Experimental results show that the system has adequate accuracy of 3% in 5 m range of distance measurement of immediate obstacles. The proposed system has been successfully tested for mobile robot navigation in an indoor environment. In the future, we will work on reducing the system execution time in order to use high resolution images. The combination of monocular vision and pan-tilt platform will be investigated to solve the problem of limited camera view angle. The proposed system will be extended to provide a simultaneous localization and mapping for mobile robot navigation.

REFERENCES

[1] S. Lenser and M. Veloso, "Visual Sonar Fast Obstacle Avoidance Using Monocular Vision," In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, Nevada, USA, 2003, pp. 886- 891.

[2] C. Chen, C. Cheng, D. Page, A. Koschan and M. Abidi, "A Moving Object Tracked by a Mobile Robot with Real-Time Obstacles Avoidance Capacity," in *Proc. of the 18th International Conference on Pattern Recognition*, Washington, D.C., USA, 2006, pp.1091-1094.

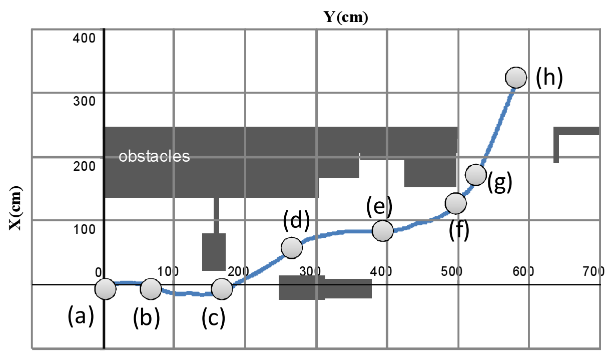


Fig.7. The recorded trajectory of the navigation experiment. Label (a)-(g) represent the position of the robot in Fig. 8(a)-(g).

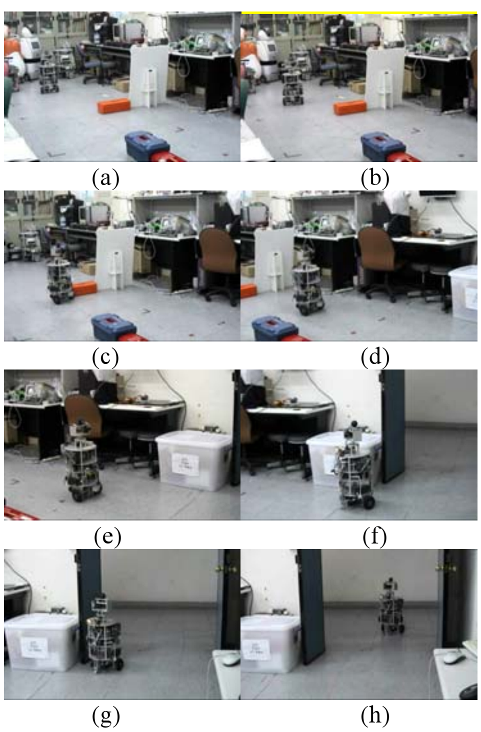


Fig.8. Snapshots from the navigation experiment. (a) The robot started to navigate. (b) The robot was avoiding obstacles. (c-f) The robot was avoiding obstacles while heading towards its goal. (g) The robot reached the goal position.

[3] N. Ohnishi and A. Imiya, "Dominant Plane Detection from Optical Flow for Robot Navigation," *Pattern Recognition Letter*, vol. 27, pp. 1009–1021, 2006.

[4] Y. Kim and H. Kim, "Layered Ground Floor Detection for Vision Based Mobile Robot Navigation," in *Proc. of IEEE Int. Conf. on Robotics and Automation*, New Orleans, Louisiana, USA, 2004, pp. 13-18.

[5] J. Michels, A. Saxena and A.Y. Ng, "High Speed Obstacle Avoidance using Monocular Vision and Reinforcement Learning," in *Proceedings of the Twenty-first International Conference on Machine Learning (ICML)*, Bonn, Germany, 2005, pp. 593-600.

[6] B. Micusik, H.Wildenauer and M. Vincze, "Towards Detection of Orthogonal Planes in Monocular Images of Indoor Environments," in *Proc. of IEEE Int. Conf. on Robotics and Automation*, Los Angeles, USA, 2008, pp. 999-1004.

[7] B. Liang, N. Pears and Z. Chen, "Affine Heigh Landscapes for Monocular Mobile Robot Obstacle Avoidance", in *Proc. of the 8th Int. Conf. on Intelligent Autonomous Systems*, Amsterdam, The Netherlands, 2004, pp.863-872.

[8] J. Zhou and B. Li, "Homography-based Ground Detection for a Mobile Robot Platform Using a Single Camera," in *Proc. of the 2006 IEEE International Conference on Robotics and Automation,* Orlando, Florida, 2006 pp, 4100-4105.

[9] J. Zhou and B. Li, "Robust Ground Plane Detection with Normalized Homography in Monocular Sequences from a Robot Platform," in *Proc. of the 2006 IEEE International Conference on Image Processing*, Atlanta, GA, USA, 2006, pp. 3017-3020.

[10] N.Simond and M. Parent, "Obstacle Detection from IPM and Super-Homography," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems,* San Diego, CA, 2007, pp.4283-4288,

[11] S. Wybo, R. Bendahan, S. Bougnoux, C. Vestri, F. Abad and T. Kakinami, "Improving Backing-Up Manoeuvre Safety with Vision-Based Movement Detection," *Intelligent Transport Systems, IET*,vol., no. 2, pp.150-158,June 2007.

[12] K. Yamaguchi, A. Watanabe, T. Naito, and Y. Ninomiya, "Road Region Estimation Using a Sequence of Monocular Images," in *Proc. of the 19th international Conference on Pattern Recognition,* Tampa, Florida, USA, 2008, pp.1-4.

[13] E. Fazl-Ersi, and J.K. Tsotsos, "Region Classification for Robust Floor Detection in Indoor Environments," in *Proc. of the 6th International Conference on Image Analysis and Recognition (ICIAR)*, Halifax, Canada, 2009, pp. 717-726

[14] F. Bonin-Font and A. Ortiz, "Building a Qualitative Local Occupancy Grid in a New Vision-Based Reactive Navigation Strategy for Mobile Robots," in *Proc. of IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Palma de Mallorca, Spain, 2009, pp. 1511-1514

[15] H. Bay, A, Ess, T. Tuytelaars and L.V. Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol.110, pp. 346-359, 2008.

[16] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp.91-110, 2004.

[17] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision, Cambridge University Press*, New York, NY, 2003.

[18] D. Comaniciu, V. Ramesh and P. Meer, "The Variable Bandwidth Mean-shift and Data-Driven Scale Selection," in *Proc. of Eighth Int'l Conf. on Computer Vision*, Vancouver, British Columbia, Canada, July 2001, pp. 438-445.

[19] O. Kähler and J. Denzler, "Detecting Coplanar Feature Points in Handheld Image Sequences," in *Proc.* of Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, Spain , 2007, pp. 447-452.

[20] L. Hong and G. Chen, "Segment-Based Stereo Matching Using Graph Cuts," in *Proc.* of *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, D.C., USA, 2004, pp.74-78.

[21] H. Kawata, A. Ohya, S. Yuta, W. Santosh and T.Mori, "Development of Ultra-Small Lightweight Optical Range Sensor System," *in Proc. of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, Edmonton, Canada, 2005, pp. 1078-1083

[22] K.T. Song and J.Y. Lin, "Behavior Fusion of Robot Navigation Using a Fuzzy Neural Network," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics* , Taipei, Taiwan, 2006, pp. 4910-4915.