Visual SLAM with Line and Corner Features

Woo Yeon Jeong
School of Radio Science and Communications
Hongik University
72-1, Sangsu-dong, Mapo-gu Seoul, 121-791, Korea
wota@cvlab.snu.ac.kr

Kyoung Mu Lee
School of Electrical Engineering, ASRI
Seoul National University
151-600, Seoul, Korea
kyoungmu@snu.ac.kr

***Abstract*— We propose a new vision-based SLAM (Simultaneous Localization and Mapping) technique using both line and corner features as landmarks in the scene. The proposed SLAM algorithm uses an Extended Kalman Filter based framework to localize and reconstruct 3D line and corner landmarks at the same time and in real time. It provides more accurate localization and map building results than conventional corner feature only-based techniques. Moreover, the reconstructed 3D line landmarks enhance the performance of the robot relocation when robot's pose remains uncertain with corner information only. Experimental results show that the hybrid landmark based SLAM, using lines and corners, produces better performance than corner only one's.**

I. Introduction

The Simultaneous Localization and Mapping (SLAM) problem is one of the most important problems in mobile robot research, and have been the central research topic in the robotics society for several decades [1][2] [3][6][7]. The purpose of SLAM is to minimize the localization and mapping error simultaneously, and it has been proved that the only constrain for the SLAM convergence is the perfect data association [2].

For the data association methods, there have been a lot of experiments using various kinds of sensors, and it has been shown that range sensor based SLAM techniques using laser or sonar work well in real environment in both indoor [7] and outdoor [6]. However, due to the high cost, speed, accuracy and safety problems, these active sensors-based SLAM methods have limitations in practical applications. Moreover, since these sensors usually provide not enough unary information of landmarks, lots of multiple measurements should be combined to solve the relocation problem. To the contrary, vision sensors have a lot of advantages in comparison. Usually, the vision sensor is very low cost, and can obtain a huge amount of information from a single shot of passive measurement. Due to the recent rapid development of CPU technology, many complex vision algorithms as well as large data can be handled efficiently in real time. Therefore, so far there have been a lot of researches applying computer vision techniques to robot localization problem [1][3] [8][9][10][11]. However, until now, most of vision-based SLAM techniques have employed point or corner-like landmarks with varying degree of success [1][3][11][13].

In this paper, we propose a new visual SLAM technique that employs line and corner features in a hybrid fashion. Line feature-based data association technique as well as the combining method of line and corner features in SLAM formulation is developed for a ceiling-vision based robot system that consists of a single camera pointing upward direction. Note that there are several advantages of using line landmarks compared to corner or point landmarks. First, they are robust to noise, because a line is composed of many points, the noise on a point usually does not affect the position and orientation of the line substantially. Second, a line can be measured from very wide range of viewing position, so the distorted map-building problem due to the finite FOV (Field of View) can be resolved by using long line landmarks which can be measured from everywhere. Finally, it provides more accurate angular orientation estimation. The proposed algorithm first distinguishes horizontal and vertical lines in an input image, and sequentially builds a 3D line feature map with appropriately estimated initial condition. Other type of visual feature like corners [13] can be combined, and this hybrid-type based SLAM shows better performance in localization and relocation than that of single type of landmark.

This paper is organized as follows. After discussing related works in the following section, we describe the proposed EKF-based SLAM formulation of our system. Section IV presents the visual line feature based data association technique and the estimation method of initial condition of lines. In section V, we describe how to increase relocation performance by adding line matching algorithm to the existing corner based relocation algorithm. Finally in section VI, we present experimental result which compare map-building results between line-based, corner-based and hybrid type based system.

II. Related Works

Recently, in order to overcome the drawbacks of using active range sensors, some works have been proposed to use vision sensors for localization and mapping.

Jogan et al. [8] proposed an appearance based localization method using omni-directional camera. Fast appearance matching was carried out through the eigen-space of trained images. Image map was made manually and to cope with conventional problems in appearance matching which are rotation and occlusion problem, they used Fourier transform of polar mapping and robust- PCA technique. Similarly, Kosecka et. al. [9] built a topological map based robot localization system. SIFT matching technique is used for each frontal view image matching and they enhanced the localization performance using spatial relationship of each image location modeled by Hidden Markov Model.

Wolf et. al. [10] suggested image retrieval based localization technique by using Monte-Carlo localization method. An image location database was made using active sensor localization system. Their matching technique could retrieve target image in spite of large camera motion by minimizing the location uncertainty from multiple measurement through Monte-Carlo method.

Lowe [3] introduced triclops-vision system that used their own wide baseline matching technique; SIFT (Scale Invariant Feature Transform) [4]. SIFT finds scale and rotation invariant position in an image, and next, compare each feature using their own robust feature descriptor which is partially invariant to small translation, affine and illumination changes. 3D SIFT features are obtained from stereo camera that is constrained by SIFT scale and rotation. Unlike basic EKF-SLAM, It maintains robot pose and landmark position separately.

Recently, Davison [1] proposed a vision-based real-time SLAM, called Mono-SLAM, which employs only a single camera without any additional camera motion information. It increases localization accuracy by integrating camera velocity into optimization target variables. The original scale of structure cannot be obtained from single camera only 3D reconstruction, one can get only its' relative scale. For this reason Mono-SLAM needs an initial manual calibration process when it starts. Also this does not provide relocation algorithm, so it is not possible to reuse 3D map constructed previously.

Meltzer et al. [11] proposed simple extended version of single-camera SLAM. In each landmark, all kind of view image is stored and each image group is compressed by KPCA to improve searching speed. While in sequential map building, data association is performed through Lucas-Kanade tracker and landmark matching of relocation is achieved by PCA projection. More recently, Jeong and Lee proposed a fast and accurate EKF-based SLAM system called CV (Ceiling Vision)-SLAM [13] that used a ceiling vision which consists of a single camera pointing upward direction, and employed corner features as landmarks.

III. EKF based SLAM Framework

In this paper, we employ an Extended Kalman Filter (EKF) based framework for the SLAM using line and corner features. While in sequential process, EKF estimate the mean and covariance of the robot pose and landmark positions successfully in the condition that noises can be modeled by Gaussian and successful data association is guaranteed.

A. State Vector and Covariance Matrix

In an EKF base framework, the optimization target vector and its covariance matrix are partitioned into robot pose and each landmarks type. The diagonal terms of a covariance matrix represent the uncertainty of the robot pose and landmark positions, and the off-diagonal terms mean the correlations among the landmark positions and the robot pose [2]. As the measurements are gradually progressed with correct data association and an appropriately estimated initial self-covariance, the uncertainties decrease monotonically. In the case of using three kinds of landmarks including corner, vertical and horizontal lines, the state vector and its covariance matrix are divided into four partitions as follows:

$$\bar{x} = \begin{bmatrix} \bar{r} \\ \bar{l}^c \\ \bar{l}^h \\ \bar{l}^v \end{bmatrix}, \mathbf{P} = \begin{bmatrix} \mathbf{P}_{\bar{r}\bar{r}} & \mathbf{P}_{\bar{r}\bar{l}^c} & \mathbf{P}_{\bar{r}\bar{l}^h} & \mathbf{P}_{\bar{r}\bar{l}^v} \\ \mathbf{P}_{\bar{l}^c\bar{r}} & \mathbf{P}_{\bar{l}^c\bar{l}^c} & \mathbf{P}_{\bar{l}^c\bar{l}^h} & \mathbf{P}_{\bar{l}^c\bar{l}^v} \\ \mathbf{P}_{\bar{l}^h\bar{r}} & \mathbf{P}_{\bar{l}^h\bar{l}^c} & \mathbf{P}_{\bar{l}^h\bar{l}^h} & \mathbf{P}_{\bar{l}^h\bar{l}^v} \\ \mathbf{P}_{\bar{l}^v\bar{r}} & \mathbf{P}_{\bar{l}^v\bar{l}^c} & \mathbf{P}_{\bar{l}^v\bar{l}^h} & \mathbf{P}_{\bar{l}^v\bar{l}^v} \end{bmatrix}, \tag{1}$$

$$\bar{l}^c = \begin{bmatrix} \bar{l}^c_0 \\ \vdots \\ \bar{l}^c_{N_c} \end{bmatrix}, \bar{l}^h = \begin{bmatrix} \bar{l}^h_0 \\ \vdots \\ \bar{l}^h_{N_h} \end{bmatrix}, \bar{l}^v = \begin{bmatrix} \bar{l}^v_0 \\ \vdots \\ \bar{l}^v_{N_v} \end{bmatrix}, \tag{2}$$

where $\bar{r} = [r_x, r_y, r_\theta]^T$ is the robot pose, $\bar{l}^c_i = [l^c_x, l^c_y, l^c_z]^T$ is the i-th corner landmark position, $\bar{l}^h_i = [l^h_r, l^h_\theta, l^h_z]^T$ represents the i-th horizontal line, $\bar{l}^v_i = [l^v_x, l^v_y]^T$ represents the i-th vertical line, and $[l^h_r, l^h_\theta, l^h_z]^T$ denote the distance, angle and height of a horizontal line, and $[l^v_x, l^v_y]^T$ denote the position of a vertical line in the X-Y (ground) plane, respectively. And, N_c, N_h and N_v are the number of corner, horizontal line and vertical line landmarks, respectively. Note that the end points of vertical and horizontal lines are not used in our EKF formulation, since these positions are often unstable and sometimes not visible in an image. Instead, these information are maintained separately for robot relocation. In the case of using line landmarks only, $\bar{l}^c$ terms in $\bar{x}$ and the corresponding rows and columns in $\mathbf{P}$ have to be deleted.

B. Measurement Model

The measurement model in EKF is the projection function that maps the landmarks in 3D space to the image plane. So, each type of landmark is projected to the measurement space through its own measurement model. Fig. 1 shows the measurement dimensions for each type of landmark, where (a) (b) and (c) represent the projected corner point, horizontal line and vertical line, respectively. The corresponding projection functions for the three features are described as follows.

(a) Corner model:

$$\begin{bmatrix} z^c_r \\ z^c_\theta \end{bmatrix} = \begin{bmatrix} \sqrt{(l^c_x - r_x)^2 + (l^c_y - r_y)^2} \times \frac{f}{l^c_z} \\ \tan^{-1}\frac{l^c_y - r_y}{l^c_x - r_x} - (r_\theta - 90) \end{bmatrix}, \tag{3}$$

(b) Horizontal line model:

$$\begin{bmatrix} z^h_r \\ z^h_\theta \end{bmatrix} = \begin{bmatrix} \sqrt{dx^2 + dy^2} \times \frac{f}{l^h_z} \\ \frac{\tan^{-1} dy}{dy - (r_\theta - 90)} \end{bmatrix} \tag{4}$$

$$dx = (-r_x \cos l^h_\theta - r_y \sin l^h_\theta + l^h_r) \cos l^h_\theta \tag{5}$$

$$dy = (-r_x \cos l^h_\theta - r_y \sin l^h_\theta + l^h_r) \sin l^h_\theta \tag{6}$$

(c) Vertical line model:

$$z^v_\theta = \tan^{-1}\frac{l^v_y - r_y}{l^v_x - r_x} - (r_\theta - 90) \tag{7}$$

where f is the focal length of a camera, $[z^c_r, z^c_\theta]^T$ and $[z^h_r, z^h_\theta]^T$ denote the polar coordinates of a projected corner and a

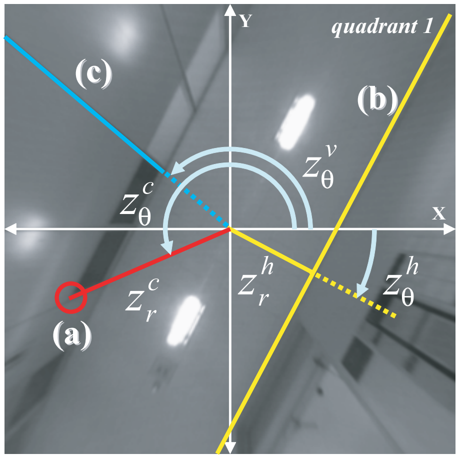


Fig. 1. Measurement dimension of each landmark; (a) corner, (b) horizontal line, and (c) vertical line

horizontal line in an image, respectively. And z_θ^v represents the polar angle of a vertical in the image coordinates as shown in Fig. 1.

IV. Line based Data Association

It has been known that the data association is the critical problem for the successful SLAM operation in real environments. And when using visual data, the data association problem is to make correspondences across multi view images. In this work, for the corner landmarks, we use the same data association algorithm described in [13], and for line landmarks, following line detection and tracking technique is applied.

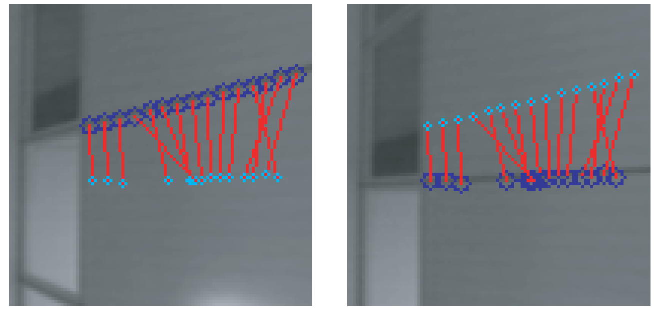
A. Line Detection and Tracking

In our experiment in noisy real environment, the Nebatia-babu edge detector [14] showed better performance than the Canny edge detector [15], but its speed was too slow for real-time processing. So we used the Canny edge detector for the edge detection and applied Nebatia-Babu algorithm for the line linking and fitting. The Lucas-Kanade feature tracker was employed for line matching between frames.

Detected lines in the initial frame are sampled into points with regular intervals first. And then, the tracking result of each sample points can be found in next frame. Since each sampled point on a line does not have strong unary information as the corner points, there might arise the aperture problem while tracking as shown in Fig. 2. Some of correspondence pairs between Fig. 2(a) and Fig. 2(b) are not matched correctly. However, as long as the tracked points are all placed on the same line in the next frame, the line is considered to be matched successfully. So, by checking the co-linearity of the tracked points in the next frame, we can simply determine the line tracking is failed or not.

B. Landmark Registration with Initial Condition

In our work, only the vertical and horizontal lines are considered as line landmarks, due to the parameterization



(a) Current frame (b) Next frame

Fig. 2. Line tracking example: Aperture problem in optical flow does not affect line tracking

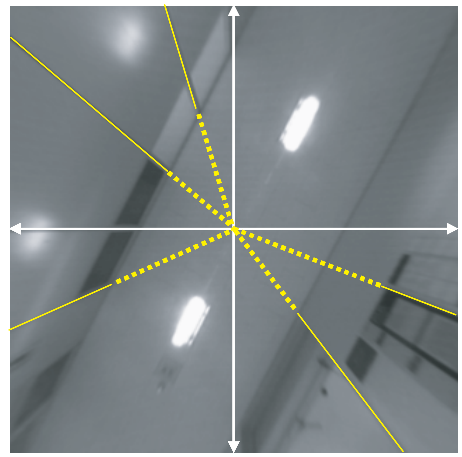


Fig. 3. Vertical line detection: All vertical lines vanish at the optical center

problem of a general line in 3D space. So, we need a preprocessing for extracting the vertical and horizontal lines in an input image. The horizontal lines can be determined by checking the angle variation while the robot is in straight motion. Vertical line detection is simpler. Since the camera is looking upward direction perpendicular to the ground plane, all vertical lines projected in the image plane vanishes to the optical center as shown in Fig. 3. After a few measurements, we can estimate the initial ground positions of vertical lines and the heights of horizontal lines using triangulation. Proper guess of initial covariance is also required for the correct estimation of those parameters through the EKF. As shown in Fig. 4, each landmark uncertainty bound is set to pass the robot position when the landmark is first registered. If the direction of robot motion is appropriate for the estimation of the line parameters and tracking is stable, covariances of lines become gradually decreased and finally, the line parameters are converged.

V. Enhanced Robot Relocation Using Line Matching

Relocation can be accomplished by matching the current image (features) with the map of 3D landmarks obtained

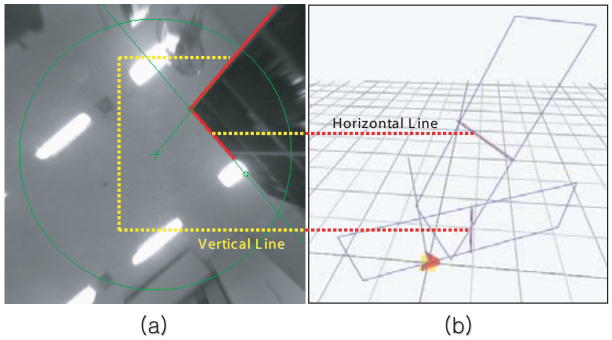


Fig. 4. Example of initial 3D line covariances : (a) line features in an input image, (b) initial rectangular uncertainty bounds for the horizontal and vertical lines

through the map building process. In general, since the line features do not exhibit enough unary information, line feature based relocation techniques often require higher dimensional geometric relational information between lines for matching, and this demands very high computational cost. However, the line feature can be used efficiently for enhancing the corner feature based relocation algorithm. For examples, when the relocation results by the corner feature only are not convincing due to the repeated corner landmarks, line matching can be used for the verification, and it can prevent the drifting or bias of mapping results due to the uneven distribution of features and the finite FOV of a camera. The process of proposed relocation algorithm using both 3D corner and line maps is shown in Fig. 5, and described in the following:

- Step 1: Find robot's candidate poses using the corner based relocation method [13].
- Step 2: For each pose candidates, 2D line image is generated by projecting the 3D line map by the line projection model. At this time, the information of the end positions of each line segment that are maintained separately is incorporated together.
- Step 3: By matching the generated line image and the observed line image, we can find the best robot pose among the candidates.

For the line image matching, we use the generalized Hausdorff measure [12] given by

$$\Phi\left(P, Q^d\right) = \frac{\#\left(P \wedge Q^d\right)}{\#\left(P\right)}, \tag{8}$$

where P and Q are binary images that have to be compared with each other. X^d denotes the dilation of X by a disk of radius d. $\#(S)$ denotes the number of 1s in the binary image S and $\wedge$ denotes the logical and (or the product) of two bitmaps.

VI. Experimental Results

The proposed algorithm was tested with a robot vacuum cleaner as shown in Fig. 6.

The robot is a two wheel based system with maximum speed of 15cm/sec. It is equipped with a CCD camera of 640 $\times$ 480 resolution with a wide angle lens, main processing

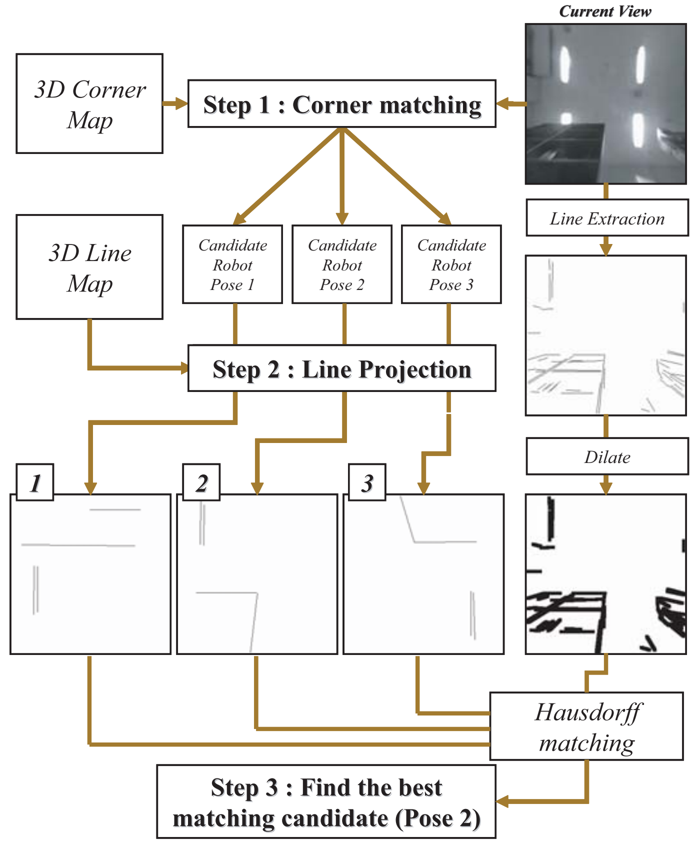


Fig. 5. Proposed relocation algorithm using both 3D corner and line landmarks

Fig. 6. Test robot platform

unit with a Pentium III 1 GHz CPU, IR-Sensors and a frontal bumper sensor for obstacle detection. We have tested the proposed algorithm in real environment as well as in off-line simulation using pre-recorded vision and odometry data set. All experiment was accomplished in real time and all procedures were performed automatically except for the robot's motion commands for wandering.

Fig. 7. shows an example of the convergence of 3D line landmarks. Because all kind of single camera measurements are nothing but to be bearing only, each landmark has uncertain parameters that have to be estimated from multiple measurements. As the robot moved on, the uncertainties of

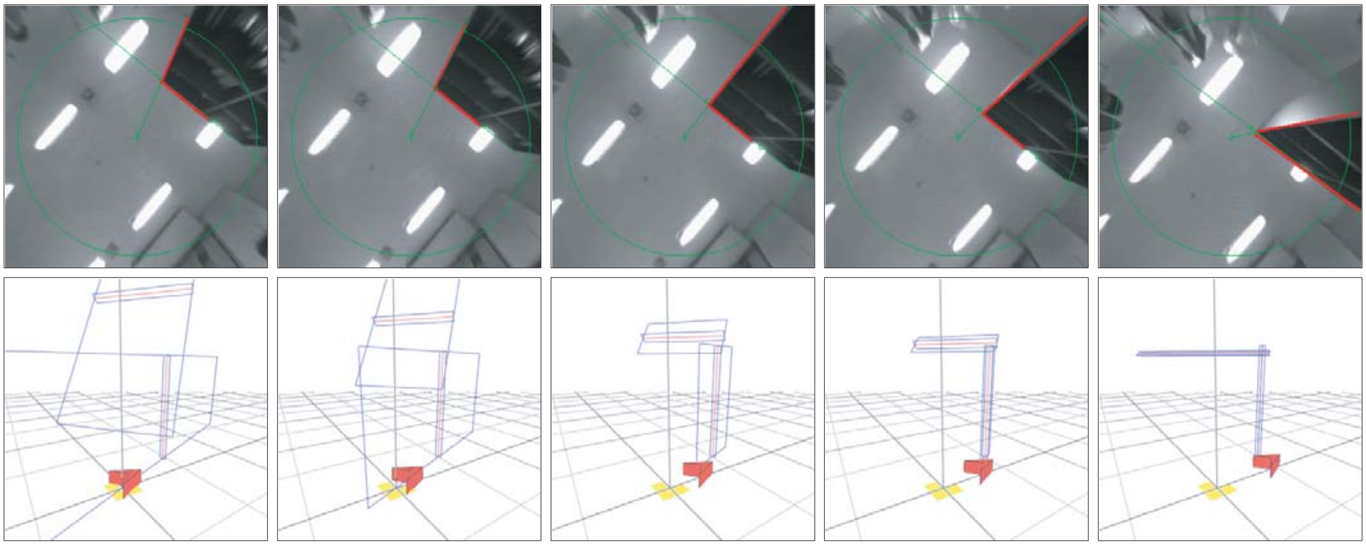


Fig. 7. Convergence of vertical and horizontal 3D lines by gradual measurements of 2D lines

landmarks, the heights of horizontal lines and the positions of vertical lines, were monotonically decreased.

We have evaluated the performance of map building accuracies of the corner-based, line-based and hybrid landmark-based SLAM, and the results are shown in Fig. 8. The test environment was a long straight corridor as shown in Fig. 8(a). In Fig. 8(b)-(d), Each grid size is 1 by 1 meter and each map is made by the same off-line recorded vision and odometry data. View point of each map (b), (c) and (d) is perpendicular to ground plane without perspective effect. The ellipsoids in (b) and (d) shows the uncertainties of corner landmarks, and the cross marks in (c) and (d) represent the position uncertainties of vertical lines. Two parallel and horizontal lines in map (c) and (d) are the ceiling edges shown in (a). The arrow in each map denotes the robot pose, and the line connected to the arrow shows the trajectory of the robot motion. The reconstructed map (features on the corridor walls) in Fig. 8(b) is somewhat skewed to upper direction due to the uneven distribution of corner landmarks. In contrast, the results in Fig. 8(c) and (d) show this distortion is corrected by employing line features. We note that a slightly better reconstruction was obtained by the hybrid landmark-based SLAM than the line-based one. Fig. 9 shows other views of the hybrid landmark-based SLAM result. Although since we don't have the ground truth map of the test environment, we cannot provide the quantitative comparison, the results demonstrate qualitatively that the line-feature and hybrid landmark-based SLAM techniques apparently produced better performance than corner based one.

VII. Conclusion and Future Works

We proposed a new line based data association technique for ceiling vision. It gradually estimates uncertain line parameters in real time. Through the experiments, we showed that the proposed algorithm is more accurate than conventional corner-based SLAM in certain environment like a long corridor. And we also showed that the proposed algorithm can be combined with corner-based SLAM and it enhances the accuracy and robustness of the map-building and localization results.

References

[1] A.J. Davison. "Real-time simultaneous localization and mapping with a single camera," In Proc. ICCV 2003.

[2] G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, and M. Csorba. "A solution to the simultaneous localization and map building (SLAM) problem," IEEE Trans. on Robotics and Automation, vol. 17, pp. 229-241, 2001.

[3] Stephen Se, David G. Lowe, James J. Little. "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," IJRR vol. 21, pp. 735-758, 2002.

[4] David G. Lowe. "Distinctive image features from scale-invariant key-points," IJCV vol. 2, pp. 91-110, 2004.

[5] K. Mikolajczyk and C. Schmid. "An affine invariant interest point detector," In Proc. ECCV 2002.

[6] C.-C. Wang. And C. Thorpe. "Simultaneous Localization Mapping with Detection and Tracking of Moving Objects," In Proc. ICRA 2002.

[7] M. Montemerlo and S. Thrun. "Simultaneous Localization and Mapping with Unknown Data Association Using Fast SLAM," In Proc. ICRA 2003.

[8] M. Jorgan and A Leonardis, "Robust localization using an omni directional appearance based subspace model of environment," Robotics and Autonomous Systems, vol. 45, pp. 51-72, 2003.

[9] J. Kosecka and F. Li, "Vision Based Topological Markov Localization," In Proc. ICRA 2004.

[10] J. Wolf, W. Burgard, H. Burkhardt. "Robust vision-based localization by combining an image-retrieval system with Monte Carlo localization," In Trans. on Robotics, vol. 21, pp. 208-216, 2005.

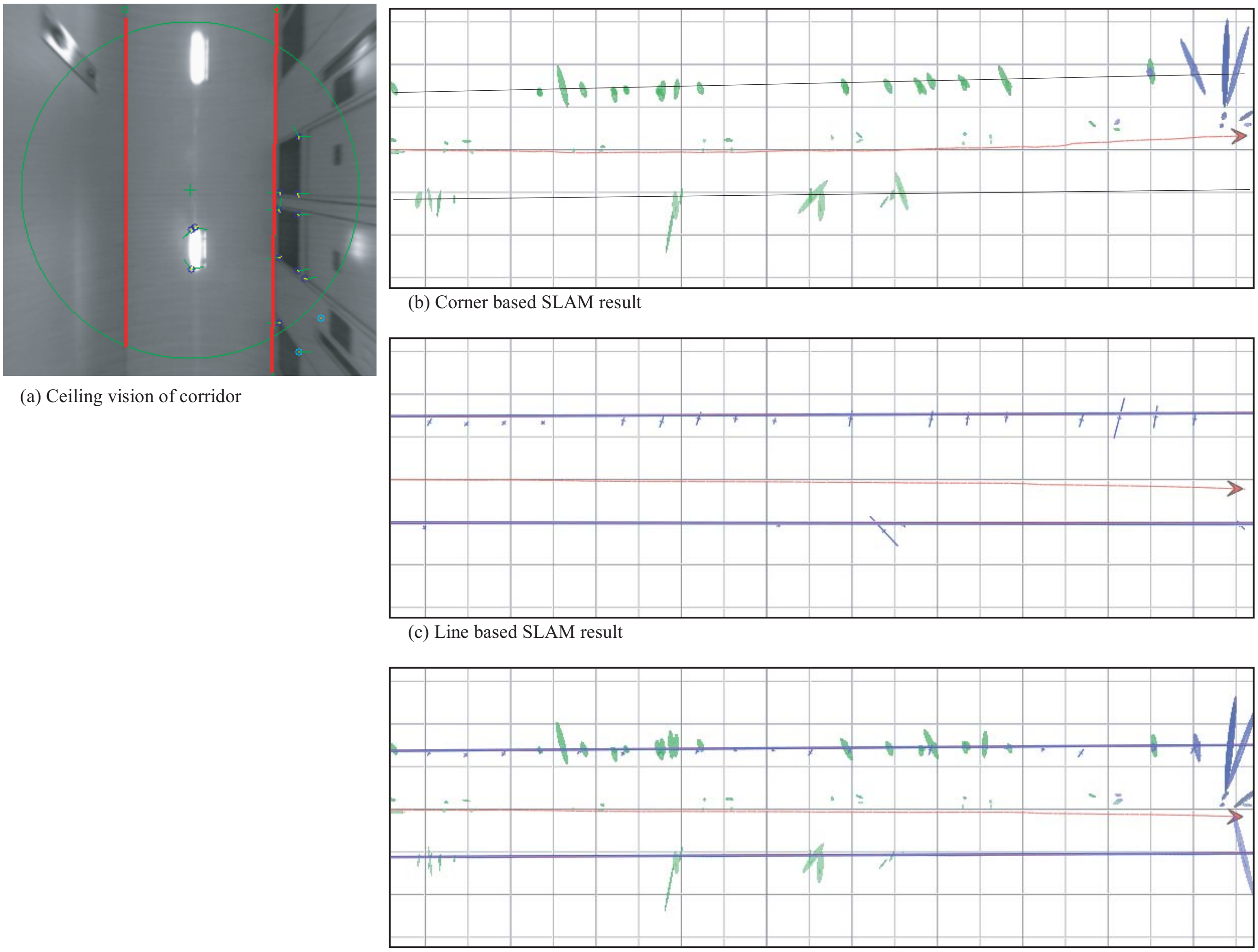
[11] J. Meltzer, R. Gupta, M-H. Yang, S. Soatto. "Simultaneous Localization and Mapping using Multiple View Feature Descriptors," In Proc. IROS 2004.

[12] W.J. Rucklidge, "Locating Objects Using the Hausdorff Distance," In Proc. ICCV 1995.

[13] W. Jeong, K. Lee. "CV-SLAM: A new Ceiling Vision-based SLAM technique," In Proc. IROS 2005.

[14] R. Nevatia and K. R. Babu. "Linear Feature Extraction and Description," Computer Vision Graphics and Image Processing, 13:257.269, 1980.

[15] J. Canny A Computational Approach to Edge Detection, IEEE Tran. on PAMI, Vol 8, No. 6, Nov 1986



(a) Ceiling vision of corridor

(b) Corner based SLAM result

(c) Line based SLAM result

(d) Corner and line based SLAM result

Fig. 8. Comparison of map building and localization results: results by (b) corner, (c) line, and (d) hybrid landmark-based SLAM technique for a test sequence of a long corridor shown in (a)

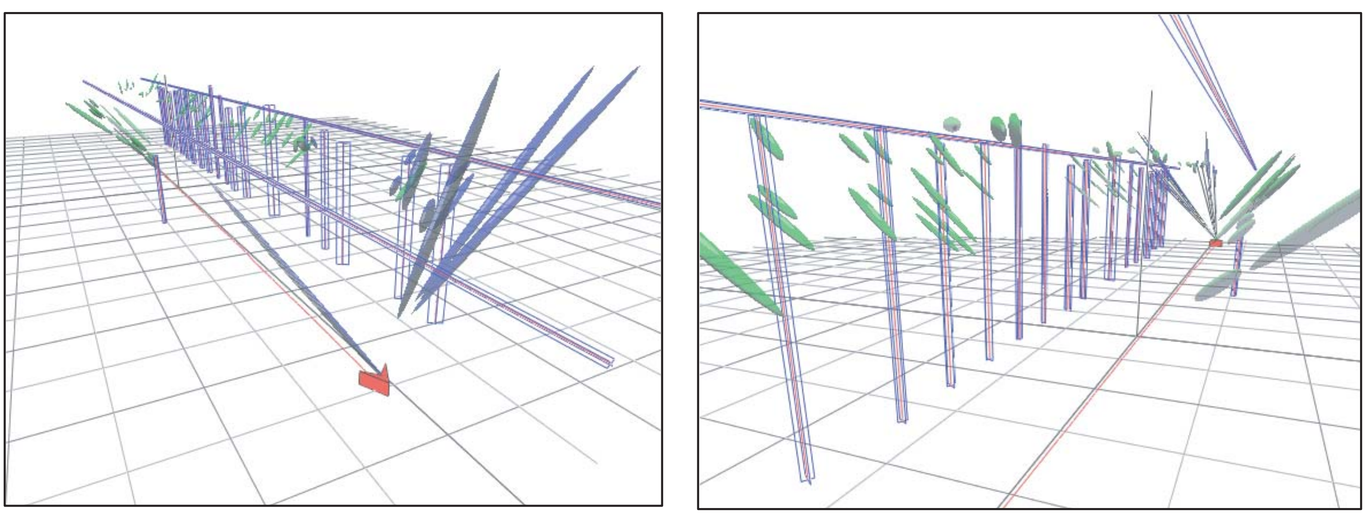


Fig. 9. Perspective views of the map building result by the proposed hybrid-landmark based SLAM