2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

SLAM with Visual Plane: Extracting Vertical Plane by Fusing Stereo Vision and Ultrasonic Sensor for Indoor Environment

Sunghwan Ahn, Kyongmin Lee, Wan Kyun Chung and Sang-Rok Oh

Abstract— **This paper presents an algorithm for visual SLAM based on a visual plane, a reliable grouping of salient visual features along sonar line features. The grouping of visual features improves data association and reduces the number of landmarks against individual visual features. To accomplish this, we propose three techniques: 1) selection of visual features which are invariant to image changes in indoor environment and suitable candidates for the visual plane, 2) extraction of sonar line features with current sensor data, which filters out uncertain outliers efficiently and 3) a scheme on grouping visual features with respect to sonar line features and maintaining database of the extracted visual planes for reliable data association. We integrate above three techniques into one framework and propose a SLAM algorithm for the visual planes. Experimental results in two types of real home environment show that the algorithm can successfully be executed with no human intervention.**

I. INTRODUCTION

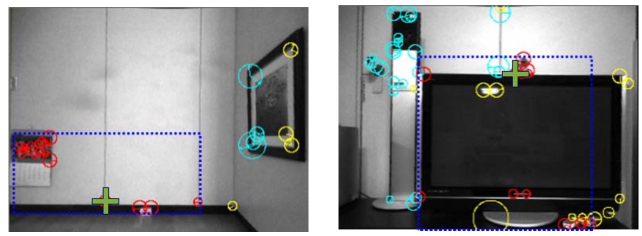
Reliable data association is one of the most essential parts of practical SLAM (Simultaneous Localization And Mapping) in indoor environment. Solutions with vision have made good progress in practical SLAM on account of its affluent information and high cost-performance. Especially, visual SLAM using salient visual features such as SIFT (Scale Invariant Feature Transform) [1], invariant to image variations, has achieved outstanding results [2]–[5].

In another aspect to visual SLAM, an object-based method [6] increases performance by clustering the visual features via object recognition with object model database. A visual object, a set of visual features, represents a physical object as a landmark in the map. This reduces the number of landmarks, making SLAM more computationally feasible and improves data association process, being more reliable than individual matching between a large number of similar visual features. And, additionally, looking for certain objects in the database allows the robot to filter out moving objects and helps it in dynamic environment where people may be moving around the robot.

However the visual object scheme deviates somewhat from the fundamental SLAM assumption of unknown environment. Although it still estimates the position of landmarks and robot, having a set of object database in advance means that it is restricted to known environment. For fully automated SLAM, it has to group visual features without pre-constructed database.

S. Ahn, K. Lee and W. K. Chung are with Robotics & Bio-Mechatronics Lab., Department of Mechanical Engineering, Pohang University of Science & Technology (POSTECH), Pohang, Korea. {ash,lekomin,wkchung}@postech.ac.kr

S. Oh is with Institute for Information Technology Advancement (IITA), Daejeon, Korea. sroh@iita.re.kr



(a) (b)

Fig. 1. Examples of visual plane extracted on (a) a wall and (b) electronics (red circles : clustered visual features, blue rectangles : boundaries of extracted visual planes)

To the authors' knowledge, extraction of physical objects by segmenting out background region from images without any database is a difficult problem, especially in a cluttered environment. For that reason, we focus on extracting planar elements from images, which can be used in SLAM instead of visual objects as clustering constraints.

To segment an image area belonging to a plane, homography, a transformation between perspective views of the plane in different images has been used [7], [8]. And a method to extract quadratic plane relied upon perceptual grouping of edge segments [9]. However they could be prone to detect wrong planes when planes are segmented in a cluttered environment. To resolve this problem, some approaches have used a range sensor because it improves accuracy of metric information of each pixel in images. Using simultaneously vision and laser range finder to cluster 3-D points into planar structures are presented in [10], [11]. Despite of great accuracy, it does not satisfy economical efficiency for practical indoor SLAM due to cost problem of laser range finder.

In this paper, a visual plane is developed as a vertical plane that consists of salient visual features located on walls or planes of furniture and electronics in indoor environment as shown in Fig. 1. To extract the visual plane, we propose a scheme of reliable grouping of visual features with a help of line features of current sonar data. For that purpose, we adopt salient visual features which are suitable for visual matching in home environment [6]. And the sonar line feature that can effectively disregard problematic specular reflection and multi-path echo of sonar sensors is proposed.

Furthermore, in order to apply the visual plane to SLAM, a management method for a database of the extracted visual planes is also presented: registering a new plane to the

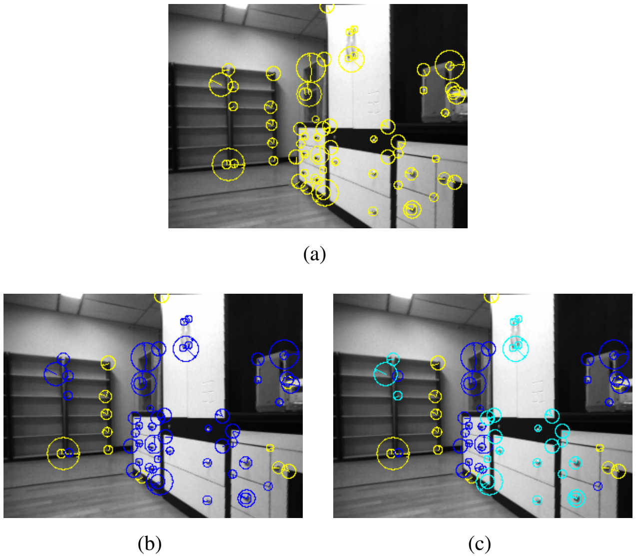


Fig. 2. Sequence of (a) Extracting multi-scale Harris corner and SIFT descriptor (yellow circles), (b) Selecting features that have metric information (blue circles) and (c) Choosing stable features (cyan circles)

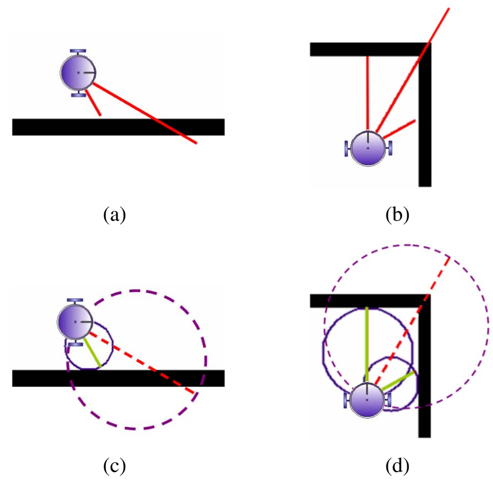


Fig. 3. Problematic sensor errors result from (a) specular reflection and (b) multi-path echo. Both can be disregarded by the definition of bubble circle: (c) specular reflection case and (d) multi-path echo case. (green solid line : correct reading, blue slid circle : bubble circle, red dashed line : false reading, purple dashed circle : non-bubble circle

database and updating the database with the re-observed plane to increase quantity of visual information. It can help to increase the consistency of SLAM and guarantee feasible computation by means of reliable data association.

The proposed scheme has several properties. Firstly, it is a multi-sensor SLAM to fuse stable metric data of sonar line features and excellent discriminating ability for data association of salient visual features. Secondly, it inherits the advantages of the aforementioned visual object [6] and even human intervention is unnecessary. Finally, by generating a group of visual features autonomously as the robot moves in indoor environment, the system can be more applicable to SLAM in unknown environment.

This paper is organized as follows. In Section II, the details of the essential components of the visual plane are described. Section III presents a way of generating reliable groups of visual features and maintaining the database of visual planes. Then section IV shows experimental results of the proposed method implemented with EKF-SLAM in home environment and conclusion follows.

II. PREREQUISITES FOR VISUAL PLANE

A. Visual features

1) Extracting visual features: In this paper, salient visual feature in images is obtained by combining advantages of multi-scale Harris corner and SIFT descriptor (Fig. 2(a)). Multi-scale Harris corner is chosen as a detector because it is suitable for detecting invariant corner-like features of natural objects in indoor environment. After extracting the detector, we use a descriptor that contains local characteristic around the detector by the same way of SIFT descriptor. Our method offers repeatability and stability in wide baseline matching [6].

2) Selecting features with metric information: We use a stereo camera because each visual feature needs metric information instantaneously to be clustered along sonar line features. After extracting visual features, visual features having reliable depth range (for our Bumblebee camera, $z \leq 4.0m$) are chosen from stereo camera (Fig. 2(b)).

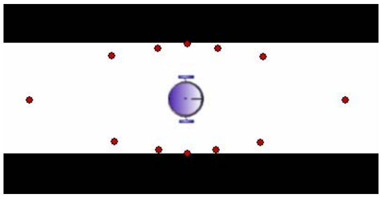
3) Choosing stable features: Unstable features placed on boundaries of a planar element can adversely affect formation of visual plane. They can be eliminated by investigating depth discontinuity, which is feasible with a stereo vision. A statistical method can be used to detect such unstable features as follows. Firstly, a small sampling window ($0.3m \times 0.3m$) at each corresponding feature location is set. Then the standard deviation from depth map of the stereo vision within the window is calculated. Finally, the feature whose resulting standard deviation violates the predetermined condition ($0.0m < \sigma_z \leq 0.2m$) can be removed (Fig. 2(c)). This process eliminates most of features which are not located on the real planar element with spurious disparity values.

B. Sonar line features

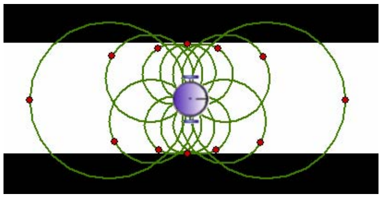
1) Filtering specular reflection of sonar data: Specular reflection and multi-pass echo disturb range accuracy of sonar data frequently. However, fortunately, they occur under specific conditions. Specular reflection is shown up when the incident angle of ultrasonic wave from sonar sensors is larger than a certain angle (Fig.3(a)) and multi-pass echo usually appears at corners (Fig.3(b)). In both cases, namely, the problematic sensor readings have peculiarly long ranges than their neighbor ones.

The *Bubble circle* (BC) is motivated by those characteristics of sensor errors and the procedure is as follows.

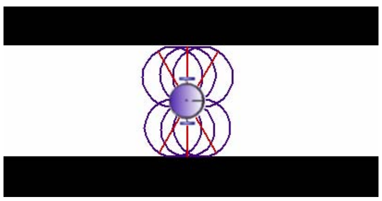
i) We present sonar sensor readings based on centerline model [12] as points (Fig.4(a)).



(a)



(b)



(c)

Fig. 4. Procedure to obtain bubble circles: (a) Replacing 12 sonar readings with points via center line model, (b) Generating 12 circles formed by 12 sonar range points and the robot and (c) Bubble Circles (BCs)

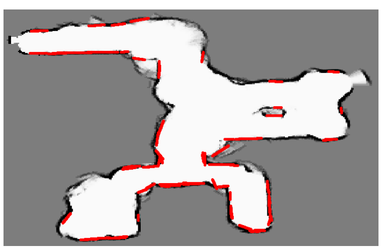


Fig. 5. Experimental result of the proposed sonar line features (red lines) in indoor environment (Fig. 15)

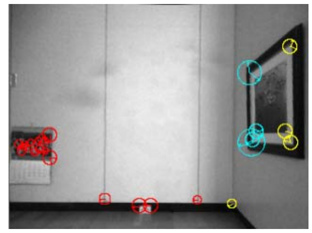
ii) If we have 2 different points, a circle whose diameter is a length between the two points can be formed. Accordingly, 12 circles can be formed using 12 points and the robot (Fig. 4(b)).

iii) We define the bubble circle: a circle that does not contain any other points inside it (Fig. 4(c)).

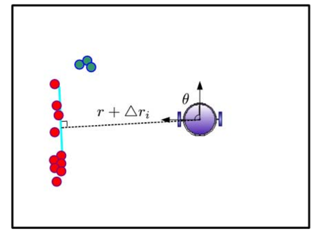
By the definition of BC, it can discriminate peculiarly long sensor readings from normal neighbor ones. For example, in Fig. 3(c) and Fig. 3(d), since the purple dashed circles contain their neighbor points, the peculiarly long readings cannot be BCs. Therefore, troublesome sensor readings of specular reflection and multi-path echo can be removed by BCs.

To filter out the errors, we choose sonar sensor readings that used to form BCs and then acquire line features by using them.

2) Extracting line features: Sonar sensor has wide range of aperture angle. When a line element is shown up in front of sonar ring, the property leads that three adjacent sonar sensors have similar range and the middle one among them has the smallest range value.



(a)



(b)

Fig. 6. An example of grouping visual features along a sonar line feature: (a) an image (red circles : grouped visual features) and (b) a diagram of the grouping (red points : grouped visual features, blue points : ruled-out visual features, cyan line : one of the candidates for a visual plane parallel to a sonar line feature)

After obtaining the filtered raw sonar data of BC, we can choose the sonar data that meet the above conditions. Then a line feature can be extracted from the three adjacent sonar data as follows:

i) choosing three points based on a centerline sensor model [12],

ii) applying least square line fitting with these points and

iii) obtaining the length of the line feature from the aperture angle of the sonar sensor.

The result of line feature detection is shown in Fig. 5. It shows that the line features (red lines) lie along the boundaries of the given environment and the proposed method works well.

III. SLAM WITH VISUAL PLANE

A. Extracting visual plane

Every visual feature has its own 3-D distance acquired from stereo camera system while sonar line feature is expressed in 2-D space. For that reason, to group the visual features, we assume that a vertical plane is located on the line feature and there can exist candidates for visual plane parallel to the vertical plane.

In section II-B, a line feature whose range r and angle θ are presented with respect to current mobile coordinate can be determined. The equation of a set of the candidates is

$$\begin{bmatrix}\cos\theta & \sin\theta & 0\end{bmatrix}\begin{bmatrix}x & y & z\end{bmatrix}^T = r+\triangle r_i \qquad (1)$$

where $\triangle r_i$ is an offset distance for each candidate i. Then the visual features are classified into corresponding plane candidates according to the distance from the visual features to the candidates. If there is a plane which possesses the maximum number of visual features, at least ten points close to the plane, then it becomes a good fit of the visual plane. An example of extracting visual plane is shown in Fig. 6.

B. Adding new landmark to database

When the robot detects a new visual plane, it has to be augmented to state vector as a landmark for SLAM. One arbitrary visual feature of the new plane is selected as a representative point and its 3-D distance with respect to the current mobile coordinate is acquired. Then it works in SLAM as a point landmark like visual object [6].

And information about the new visual plane is also registered at a database for data association as the robot navigates the environment. The components of the database are as follows: for each visual feature within the plane,

i) position with respect to the image coordinate
ii) scale, orientation, descriptor vector and
iii) relative 3-D distance to the representative point.

The relative distance of visual features is required to update the position of every feature in the plane and supplement features that the database does not include, when the plane is re-observed.

A relative state of the j^{th} visual plane, $\mathbf{X}_{rel,j}$ with respect to the representative point $\mathbf{x}_{t,j}$ is presented as

$$\begin{aligned}\mathbf{X}_{rel,j} &= \begin{bmatrix}\mathbf{x}_{1t,j}^T & \mathbf{x}_{2t,j}^T & \cdots & \mathbf{x}_{nt,j}^T\end{bmatrix}^T, \\ \mathbf{x}_{it,j} &= \mathbf{x}_{i,j} - \mathbf{x}_{t,j}\end{aligned} \tag{2}$$

where $\mathbf{X}_{VP,j} = [\mathbf{x}_{1,j}^T, \mathbf{x}_{2,j}^T, \cdots, \mathbf{x}_{n,j}^T]^T$ is 3-D distance of the n visual features of the j^{th} plane with respect to the current robot coordinate. And a corresponding covariance, $\mathbf{P}_{rel,j}$ can be determined as

$$\mathbf{P}_{rel,j} = \mathbf{R}_{i,j} + \mathbf{R}_{t,j} \tag{3}$$

where $\mathbf{R}_{\cdot,j}$ is a measurement covariance of each visual feature of the j^{th} plane. The measurement covariance is computed by noise parameters σ_u, σ_v, σ_d for pixel uncertainties in image position u, v and disparity d as follows:

$$\mathbf{R}_{\cdot,j} = \nabla\mathbf{g}_{\cdot,j}\begin{bmatrix}\sigma_u^2 & 0 & 0\\ 0 & \sigma_v^2 & 0\\ 0 & 0 & \sigma_d^2\end{bmatrix}\nabla\mathbf{g}_{\cdot,j}^T \tag{4}$$

where $\nabla\mathbf{g}_{\cdot,j}$ is Jacobian of

$$\mathbf{g}_{\cdot,j} = \begin{bmatrix}z_{\cdot,j}\\ x_{\cdot,j}\\ y_{\cdot,j}\end{bmatrix} = \begin{bmatrix}fB/d_{\cdot,j}\\ u_{\cdot,j}z_{\cdot,j}/f\\ v_{\cdot,j}z_{\cdot,j}/f\end{bmatrix} = \begin{bmatrix}fB/d_{\cdot,j}\\ u_{\cdot,j}B/d_{\cdot,j}\\ v_{\cdot,j}B/d_{\cdot,j}\end{bmatrix} \tag{5}$$

with respect to u, v, d. f and B are the camera focal length and base line, respectively.

C. Reinforcement of database

When the robot observes the visual plane again, by means of the same strategy as an object recognition in [6], the robot can match an observed plane to the database. Basically, we use the RANSAC clustering based on descriptor vectors of visual features to recognize robustly the plane by retrieving the previously registered database.

As mentioned earlier, the old database has relative 3-D distance with respect to the position of the representative point. After an observed visual plane is matched to the old database, the relative state of the old database can be updated via the matched (re-observed) points and then the unmatched (new) features can be inserted into the old database by using a relation to the matched points. Consequently, it can lead to increase chance of reliable data association and result in an improvement of the SLAM estimation. Fig. 7 shows a schematic diagram of reinforcement of the database.

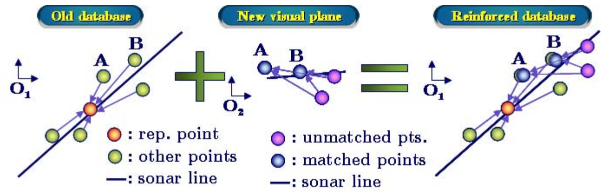


Fig. 7. A diagram for updating re-observed visual features and adding new visual features to the old database

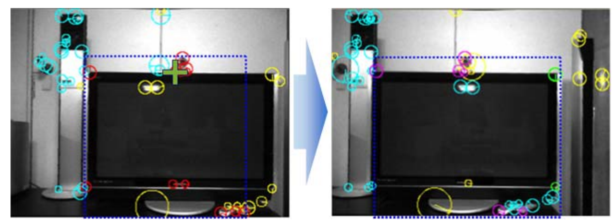


Fig. 8. Reinforcement of database (left: old database, right: updated database with added new features (green circles))

1) Update re-observed visual features: In order to update the relative state of each visual feature in the database, Kalman update [13] is used.

3-D distance of visual features in an observed plane after data association with the j^{th} plane is presented as follows:

$$\mathbf{M} = \begin{bmatrix}\mathbf{x}_{m_1}^T & \cdots & \mathbf{x}_{m_k}^T & \mathbf{x}_{u_1}^T & \cdots & \mathbf{x}_{u_l}^T\end{bmatrix}^T \tag{6}$$

where $\mathbf{x}_{m_1}, \cdots, \mathbf{x}_{m_k}$ are matched to the visual features $\mathbf{x}_{\varphi(m_1),j}, \cdots, \mathbf{x}_{\varphi(m_k),j}$ in $\mathbf{X}_{VP,j}$ respectively, $\varphi(\cdot)$ is data association describing the mapping between indices of visual features and $\mathbf{x}_{u_1}, \cdots, \mathbf{x}_{u_l}$ are unmatched features.

Hence a following relation can be established by relative information of the matched pair.

$$\begin{aligned}\mathbf{x}_{m_p} - \mathbf{x}_{m_q} &\approx \mathbf{x}_{\varphi(m_p),j} - \mathbf{x}_{\varphi(m_q),j}\\ &= \mathbf{x}_{\varphi(m_p)t,j} - \mathbf{x}_{\varphi(m_q)t,j} \qquad (p < q)\end{aligned} \tag{7}$$

And a relative consistency constraint

$$\mathbf{H}\mathbf{X}_{rel,j} = \mathbf{b} \tag{8}$$

where

$$\mathbf{H} = \begin{bmatrix}0 & \cdots & 1 & \cdots & -1 & \cdots & 0\end{bmatrix} \tag{9}$$
$$\mathbf{b} = \begin{bmatrix}\mathbf{x}_{m_p} - \mathbf{x}_{m_q}\end{bmatrix} \tag{10}$$

is obtained.

By applying the constraint by Kalman update to the relative state $\mathbf{X}_{rel,j}$ and covariance $\mathbf{P}_{rel,j}$, a constrained relative state $\mathbf{X}_{rel,j}^c$ and covariance $\mathbf{P}_{rel,j}^c$ can be updated.

$$\mathbf{X}_{rel,j}^c = \mathbf{X}_{rel,j} + \mathbf{K}(\mathbf{b} - \mathbf{H}\mathbf{X}_{rel,j}) \tag{11}$$
$$\mathbf{P}_{rel,j}^c = \mathbf{P}_{rel,j} - \mathbf{K}\mathbf{H}\mathbf{P}_{rel,j} \tag{12}$$
$$\mathbf{K} = \mathbf{P}_{rel,j}\mathbf{H}^T[\mathbf{H}\mathbf{P}_{rel,j}\mathbf{H}^T + \mathbf{R}]^{-1} \tag{13}$$

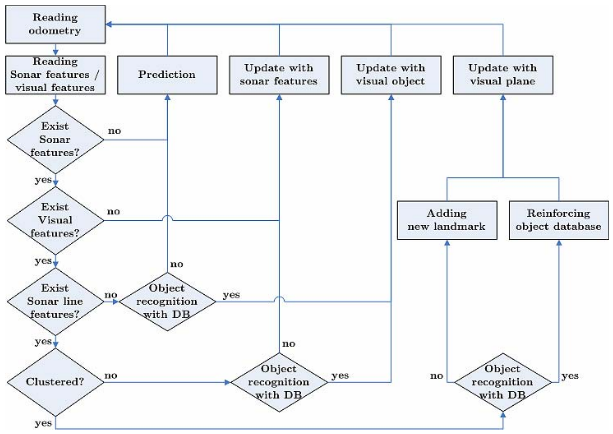


Fig. 9. Overall process of SLAM with visual planes

2) Adding unmatched visual features to old database: Based on the updated visual features, we can insert the unmatched features to the database. In other words, 3-D distance (x_i, y_i, z_i) of the unmatched features can be reconstructed with respect to the mobile coordinate of the old database with a help of the matched features (Fig. 7).

Then row and column position with respect to the image coordinate of old database is updated via the camera model,

$$r_i = v_o + f\frac{y_i}{z_i} \tag{14}$$

$$c_i = u_o + f\frac{x_i}{z_i} \tag{15}$$

where (u_o, v_o) is a camera center point in the image coordinate and f is a focal length of the camera.

The unmatched features can be included in the old database in this way and it can make the database enriched incrementally as the robot moves (Fig. 8).

D. Map management of visual plane

The extracted visual planes are located uniquely over the environment since their salient visual features are invariant to the image variations. However it has a little chance to be overlapped because the constraint applied for the visual features to select the stable ones is so strong that it decreases the number of visual features to match. This results in a failure in plane matching and adds a new plane to the database.

To make up this problem, we adopt a system of fully automatic construction of a panoramic image using SIFT [14]. For unordered visual planes, it can stitch them without human intervention. Additionally, it is robust to changes in camera viewpoints and illumination. The usage examples of the mosaic system will be shown up through the experimental results in the section IV.

E. Overall process

The visual plane of the database can be constructed and updated when there exists a sonar line feature, as aforementioned. However, in case of SLAM estimation, even when the robot misses to detect any line features, it is necessary to extract the visual plane with use of database and return accurate metric information for SLAM update. For the purpose of more frequent update of visual information, the robot tries to do an object recognition procedure [6] with the database of the generated visual planes using only stereo vision.

The overall process of SLAM with visual planes is shown as Fig. 9.

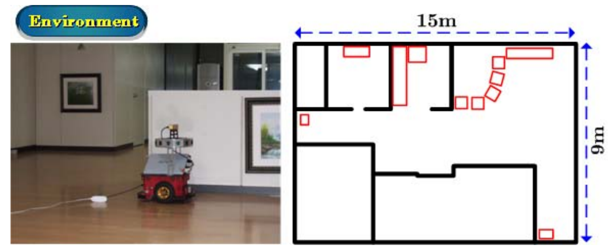
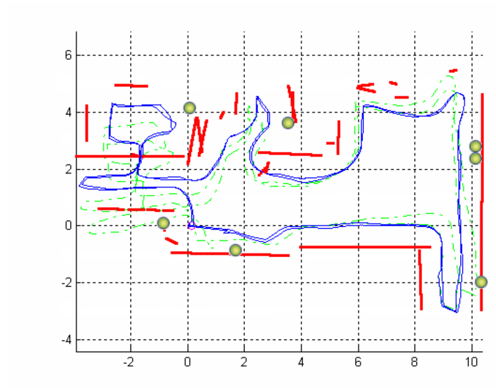
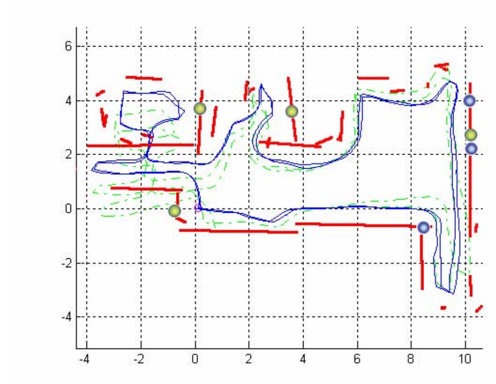


Fig. 10. Home environment I in ETRI



(a)

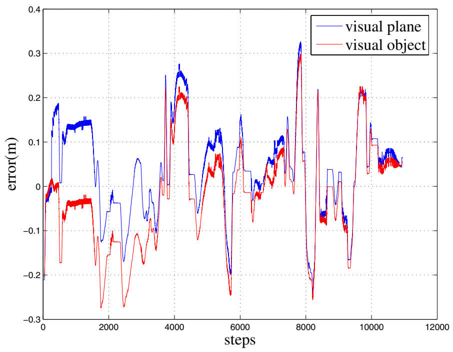


(b)

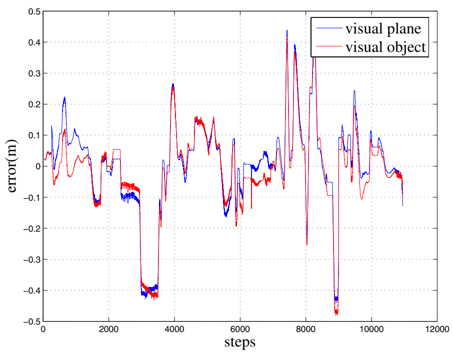
Fig. 11. Comparison of the result of SLAM with (a) visual plane and (b) visual object in home environment I

IV. EXPERIMENTAL RESULTS

To verify the proposed scheme on SLAM using visual planes, experiments in two different kinds of home environment were executed using Pioneer3-DX equipped with 12 Murata piezo-electric sonar sensors and bumblebee stereo camera.



(a)



(b)

Fig. 12. Difference between two estimated loops in (a) x direction and (b) y direction in home environment I

Fig. 13. Extracted visual planes in home environment I

The robot navigated two times with a speed of about $0.2m/s$ along the wall by using sonar sensors to cover all planar elements in the environment. Since the robot moved by a sensor-based navigation (wall following), the former loop that the robot moved almost corresponded to the later one. Therefore, we checked the performance of the proposed method via the degree of the coincidence of the two estimated loops.

A. Home environment I in ETRI

The ETRI environment (Fig. 10) is a home-like environment as a test-bed for indoor navigation, which covers in $15m \times 9m$ area. Moreover, it has several pieces of furniture, electronics and picture frames to reproduce a real home.

The experimental result of the proposed SLAM method with visual planes is shown in Fig. 11(a) and, with the purpose of comparison, a SLAM result using visual objects [6] is also presented (Fig. 11(b)). In these results, green

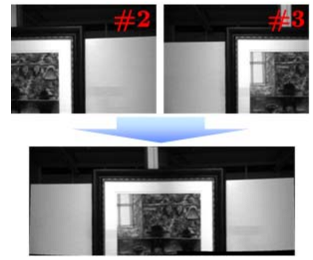


Fig. 14. Adding up the coincident visual planes in home experiment I

dotted line is cumulative odometry path and blue solid line is the estimated path. The resulting SLAM maps have red lines and blue circles that represent line and point features of sonar sensors respectively. And green circles mean visual planes and visual objects for each case.

The difference of two corresponding estimated loops can show the capability of the estimation. The magnitude of the estimated errors for both cases is almost similar (Fig. 12) and the estimated positions of visual planes quite conform to those of visual objects (Fig. 11).

However, comparing the estimated paths around the lower right corner $(9m, -2m)$ of the map (Fig. 11), we were able to check the discrepancy between the two results. Such an improvement in SLAM using visual planes is caused by the characteristic that visual planes have more visual information than visual objects. More information can be gathered by updating and reinforcing the database incrementally as the robot moves. Moreover, we need to keep in mind that visual objects require a pre-constructed database for SLAM but visual planes can generate a database autonomously.

Contrary to SLAM with visual objects, SLAM with visual planes needs more computational time about $0.2sec.$ for each step but it is capable of carrying out in real-time implementation with $1Hz$ sampling of vision data.

Finally, the robot detected distinct 7 visual planes (Fig. 13). Although the second and the third planes are parts of the same picture frame, they were discriminated as different visual planes for the reason that they have not enough common visual features to share. By applying the mosaic algorithm (section III-D) to the two separated planes, they can be merged into one in the proposed algorithm framework autonomously (Fig. 14).

B. Home environment II in a real apartment

We expanded the experiment of the visual plane to more general environment, an real apartment where a family is living (Fig. 15). The robot navigated two bedrooms, a living room and a dining room in a wall following manner and the covered area was $11m \times 8m$. The final SLAM map and the estimated path with visual planes is shown in Fig. 16. The two estimated loops that the robot run were comparatively coincident (Fig. 17) and we could check the performance of the proposed algorithm in real environment.

As a result, the robot generated 18 visual planes over the environment (Fig. 18). After that, the sixth and the seventh planes and the fifteenth and the sixteenth planes put together

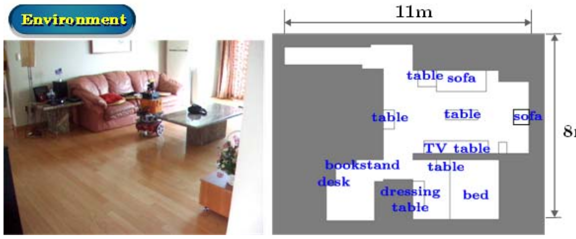


Fig. 15. Home environment II, an apartment

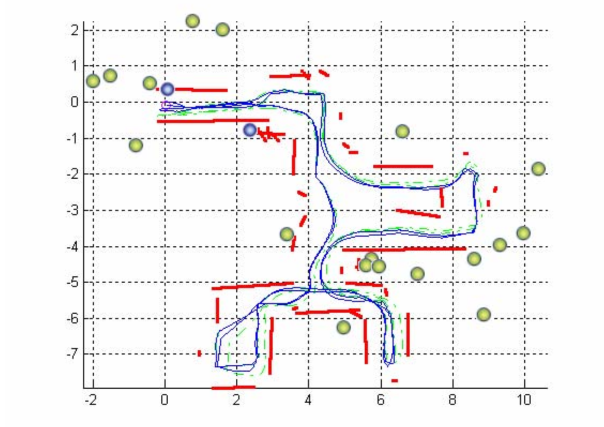
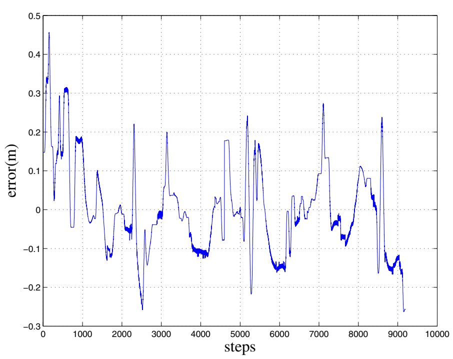
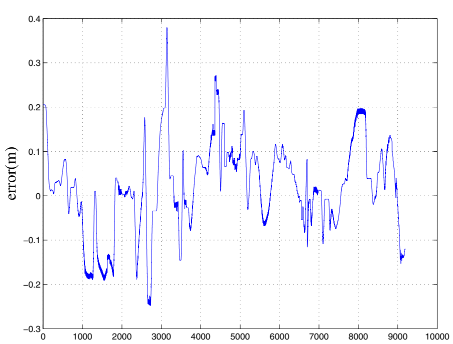


Fig. 16. The result of SLAM with visual plane in home environment II



(a)



(b)

Fig. 17. Difference between two estimated loops in (a) x direction and (b) y direction in home environment II

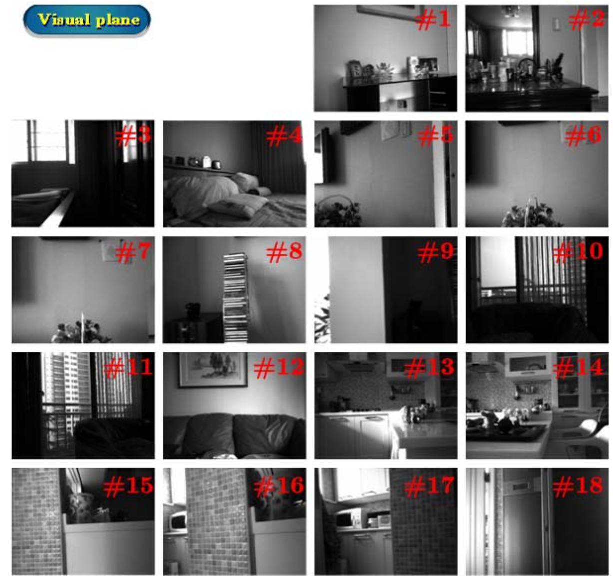
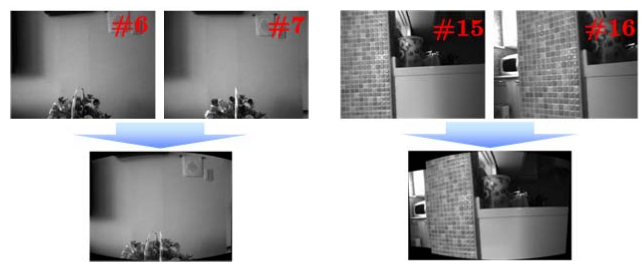


Fig. 18. Extracted visual planes in home environment II



(a) (b)

Fig. 19. Adding up the coincident visual planes in home environment II for two cases: (a) and (b)

(Fig. 19) by the mosaic algorithm and this increases the quality of the SLAM map.

V. CONCLUSIONS

This paper addressed a SLAM algorithm in indoor environment with visual planes composed of salient visual features. It could preserve advantages of SLAM with visual objects [6] without pre-constructed model database. For that purpose, we proposed three associated schemes. Firstly, selection of robust visual features combining advantages of multi-scale Harris corner and SIFT descriptor was suggested. Secondly, we analyzed and redeemed problems of sonar sensors like specular reflection and multi-path echo phenomena. Then robust line features were extracted by using the filtered sonar data. Finally, we proposed a scheme creating and managing the clusters of the visual features based on the sonar line features.

Experiments were performed in two kinds of home environment to validate the algorithm. The robot registered the visual planes autonomously with wall following navigation. Moreover, the resulting visual planes were located uniquely with each other and performed data association reliably.

Future directions would be to attempt to construct a local map with the proposed robust sonar features and maintain global consistency by using the visual planes. It is expected that the visual planes can play more important role in SLAM.

VI. ACKNOWLEDGMENTS

This research was supported in part by the National Research Laboratory (NRL) Program(M1-0302-00-0040-03-J00-00-024-00) of the Ministry of Science and Technology, by a grant(A1100-0602-0026) of the Ministry of Information and Communication, and by a grant(02-PJ3-PG6-EV04-0003) of Ministry of Health and Welfare, Republic of Korea.

REFERENCES

[1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," in *International Jounal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] S. Se, D. G. Lowe, and J. Little, "Mobile Robot Localization and Mapping with Uncertainty using Scale-invariant Visual Landmarks ," in *International Journal of Robotics Research*, vol. 21, no. 8, pp. 735–758, 2002.

[3] N. Karlsson, E. D. Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich, "The vSLAM Algorithm for Robust Localization and Mapping," in Proc. of *International Conference on Robotics and Automation*, pp. 24–29, 2005.

[4] T. D. Barfoot, "Online Visual Motion Estimation using FastSLAM with SIFT Features," in Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3076–3082, 2005.

[5] P. Elinas, R. Sim and J. J. Little, "σSLAM: Stereo Vision SLAM using the Rao-Blackwellised Particle Filter and a Novel Mixture Proposal Distribution," in Proc. of *International Conference on Robotics and Automation*, pp. 1564–1570, 2006.

[6] S. Ahn, M. Choi, J. Choi and W. K. Chung, "Data Association Using Visual Object Recognition for EKF-SLAM in Home Environment," in Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2588–2594, 2006.

[7] A. Dick, P. Torr and R. Cipolla, "Automatic 3D Modelling of Architecture," in Proc. of *Proc. 11th British Machine Vision Conference*, pp. 372–381, 2000.

[8] J. Jung, T. Kim, O. Kwon, and J. Paik, "Plane Identification and Segmentation from Uncalibrated Stereo based on Evolutionary Approach," in Proc. of *International Signal Processing, Communication Systems*, pp. 80–84, 2004.

[9] J. B. Hayet, F. Lerasle, and M. Devy, "A Visual Landmark Framework for Indoor Mobile Robot Navigation," in Proc. of *International Conference on Robotics and Automation*, pp. 3942–3947, 2002.

[10] D. Ortin, J. M. M. Montiel and A. Zisserman,"Automated Multisensor Polyhedral Model Acquisition," in Proc. of *International Conference on Robotics and Automation*, pp. 1007–1012 , 2003.

[11] H. Andreasson, R. Triebel and W. Burgard, "Improving Plane Extraction from 3D Data by Fusing Laser Data and Vision," in Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2656–2661, 2005.

[12] H. Choset, K. Nagatani, and N. A. Lazar, "The arc-transversal median algorithm: A geometric approach to increasing ultrasonic sensor azimuth accuracy," in *IEEE Transactions on Robotics and Automation*, vol. 19, no. 3, pp. 513–523, 2003.

[13] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, "Estimation with Applications to Tracking and Navigation," John Wiley, 2001.

[14] M. Brown and D. G. Lowe, "Recognising panoramas," in Proc. of *International Conference on Computer Vision*, pp. 1218–1225, 2003.