Automatic Extraction of Visual Landmarks for a Mobile Robot under Uncertainty of Vision and Motion*

Inhyuk Moon Jun Miura Yoshiaki Shirai

Dept. Biomedical Engineering
Yonsei University
Wonju, Kangwon 220-071, Korea

Dept. Computer-Controlled Mechanical Systems
Osaka University
Suita, Osaka 565-0871, Japan

Abstract

In this paper, we propose a method to automatically extract stable visual landmarks from visual sonsory data obtained by a mobile robot with stereo vision system so as to adapt to changes of environment. Vertical line segments which are distinct and on planar surfaces are selected as stable features for visual observation because they are expected to be observed reliably from various viewpoints. But the feature positions include uncertainty due to the vision and motion error of the robot. In order to utilize the selected features as landmarks, the feature positions are modeled by probabilistic distribution, and estimated by using the Kalman filter. When the robot moves, it uses several, more reliable landmarks for localization. Experimental results in real scenes show the validity of the proposed method.

1 Introduction

Since positional uncertainty is increased by dead reckoning, it is the most fundamental function for a robot to localize in its environment from internal and external sensory information[1]. In addition, to obtain reliable sensory information is also another important problem.

The landmark-based navigation method is very often used for the mobile robotics field because it takes less cost for sensing and matching to the map. For reliable observation, artificial landmarks such as a sign pattern[2] were used for the mobile robot navigation. It is, however, a reluctant work for us to arrange such artificial landmarks in the environment.

*Proc. 2001 IEEE Int. Conf. on Robotics and Automation, pp. 1188-1193.

Given an obstacle map the robot selects features such as boundary edges of the obstacles as landmarks, and observes the landmarks for localization[3]. But the boundary edge may be *unstable* feature for the visual observation according to the change of viewpoints.

To perform stable observation, natural landmarks such as points and lines[4] in the scene, or regions of high edge density[5] were used as visual landmarks. In off-line learning phase, a set of landmarks extracted from image saves in a map. In on-line execution phase, the robot estimates its current state(position and orientation) by matching appeared features with the map. When the environment is fixed between learning and execution phases, these methods may be useful for the navigation. However, the robot with passive observation by the stationary gaze cannot localize itself when it deviates from the trajectory of the learning phase.

Recently, localizing methods without landmark information were proposed. Using laser range finder or sonar sensor, the robot gathers range data around environments and estimates the current state by Markov localization considering uncertainty[6]. Since these sensors obtain 1D metric data, the robot may be confused when it observes some objects with slender legs such as tables or chairs.

In this paper, we propose a method to automatically extract visual landmarks. Visual features to be detectable from various viewpoints are regarded as *stable visual landmarks*, and the robot on-line selects the stable visual landmarks from visual data obtained by stereo vision system so as to adapt to the change of lighting condition and feature positions. Since features located on a vertical obstacle plane are less affected by the change the background scene than boundary edges[7], the robot extracts *vertical line segments* on the vertical obstacle plane(*reference*

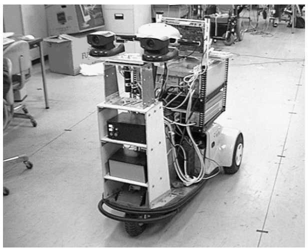

Figure 1: Mobile robot.

plane) as the stable visual landmarks from visual sensory data. The position of extracted features includes uncertainty due to the uncertainty of motion and observation[8]. The robot estimates the probabilistic distribution of the feature positions by matching the obstacle plane belonged the features with the map, and then it registers the features as landmarks in the map.

When the robot moves using the on-line extracted landmarks, it controls gaze on a boundary edge in the obstacle plane belonged the landmarks, and observes the landmarks continuously according to the change of viewpoints. Experimental results using a three-wheeled mobile robot with stereo vision system in Fig. 1 show the validity of the proposed method.

2 Modeling of uncertainty

2.1 Uncertainty of robot motion

The state of the robot, $\boldsymbol{x} = [x\ y\ \theta\ \phi]^T$, consists of the position of the front wheel, $(x,\, y)$, the orientation of the robot, θ, and the viewing direction ϕ. The robot is controlled by input $\boldsymbol{u} = [s\ \lambda\ \psi]^T$ which consists of a moving distance, a steering angle, and a pan angle. The state transition of the robot can be expressed by the following nonlinear equation[10]:

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}(\boldsymbol{x}_t, \boldsymbol{u}_t)\,. \tag{1}$$

By linearizing Eq. (1) using the first-order Taylor series expansion around the mean values, $\hat{\boldsymbol{x}}_t$ and $\hat{\boldsymbol{u}}_t$, we can obtain the prediction uncertainty of $\boldsymbol{x}_{t+1}$ as follows:

$$\boldsymbol{\Sigma}_{x(t+1)} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}_t} \boldsymbol{\Sigma}_{x(t)} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}_t}^T + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{u}_t} \boldsymbol{\Sigma}_{u(t)} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{u}_t}^T\,, \tag{2}$$

where $\boldsymbol{\Sigma}_{u(t)}$ is the covariance matrix of the control input. Assuming that the control errors are Gaussian and independent of each other, $\boldsymbol{\Sigma}_{u(t)}$ is expressed as a diagonal matrix form:

$$\boldsymbol{\Sigma}_{u(t)} = \begin{bmatrix} {\sigma_s}^2 & 0 & 0 \\ 0 & {\sigma_\lambda}^2 & 0 \\ 0 & 0 & {\sigma_\psi}^2 \end{bmatrix}, \tag{3}$$

where ${\sigma_s}^2$ denotes the variance of the odometry measurement error due to slippage, which is considered to be proportional to the distance input s; ${\sigma_\lambda}^2$ and ${\sigma_\psi}^2$ are the variance of control input λ and ψ, respectively. The variance of each error is found from experimental results, and reasonably approximated by Gaussian.

2.2 Uncertainty of vertical line segment

When a point in the 3D coordinates is projected into image, the projected position includes uncertainty due to a quantization error. We model this error as Gaussian distribution with $\sigma_{pixel} = 0.5$[pixel].

Assuming that an observed vertical line segment is straight in the image and that the distribution of each pixel composing the line is independent, the horizontal positional uncertainty of the segment is inversely proportional to the number of pixels[8]. In this paper, we set the standard deviation of the position of the line segment to σ_{pixel} considering calibration error of the stereo camera.

2.3 Uncertainty of projection position

When the robot with state $\boldsymbol{x}$ observes a known point $\boldsymbol{m} = [m_x\ m_y]^T$, it obtains the observed positions in the stereo images, $\boldsymbol{i} = [x_l\ x_r]^T$, using the following non-linear equation[10]:

$$\boldsymbol{i} = \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{m})\,. \tag{4}$$

By linearizing Eq. (4) using the Taylor series expansion around the mean $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{m}}$, the covariance matrix is obtained as follows:

$$\boldsymbol{\Sigma}_i = \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{x}} \boldsymbol{\Sigma}_x \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{x}}^T + \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{m}} \boldsymbol{\Sigma}_m \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{m}}^T\,. \tag{5}$$

In this paper we assume that 3σ area from the mean value is a valid probabilistic area, and define 3σ area as the boundary of uncertainty area. Therefore, the valid region for searching is defined as follows:

$$[u - 3\sigma, u + 3\sigma], \tag{6}$$

where u denotes the mean value.

2.4 Uncertainty of disparity

Disparity is expressed by subtraction of a pair of stereo projected positions x_l and x_r.

$$d = x_l - x_r + \nu, \tag{7}$$

where ν denotes a pixel error with $E[\nu] = 0$ and ${\sigma_\nu}^2 = 2{\sigma_{pixel}}^2$.

From Eqs. (4) (7) the following non-linear equation is obtained:

$$d = D(\boldsymbol{x}, \boldsymbol{m}) + \nu \tag{8}$$

By linearizing Eq. (8), the disparity variance is obtained as follows:

$${\sigma_d}^2 = \frac{\partial D}{\partial \boldsymbol{x}} \boldsymbol{\Sigma}_x \frac{\partial D}{\partial \boldsymbol{x}}^T + \frac{\partial D}{\partial \boldsymbol{m}} \boldsymbol{\Sigma}_m \frac{\partial D}{\partial \boldsymbol{m}}^T + {\sigma_\nu}^2. \tag{9}$$

If the given map has no uncertainty, the covariance $\boldsymbol{\Sigma}_m$ is 0, so we can use simplified equations eliminated the second terms in the right side of Eqs. (5)(9). Given a reference plane valid search area in the disparity image is obtained by using Eqs. (4)(5). And valid disparity data is picked out by using Eqs. (8)(9).

3 Extracting stable features

3.1 Detection of reference plane

The area-based stereo matching is used for detecting the reference plane. As a result, disparity image is obtained using SAD(Sum of Absolute Differences) to calculate correlation between a pair of stereo images.

First, we obtain the disparity image by using SAD. Given a reference plane, we set a boundary edge of the reference plane to $\boldsymbol{m}$ in Eq. (4), and obtain the mean and variance of the boundary edge position by Eqs. (4)(5). The valid disparity data is also picked out by using Eqs. (6)(8)(9).

The disparity data is transformed to a set of positions in the robot workspace $X-Y$ using the following equation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (x_l - 2B)/d - B \\ 2Bf/d \end{bmatrix}, \tag{10}$$

where B means a half of distance between two focus of stereo cameras. Let the reference plane be $y = a + bx$ in $X - Y$, and fitted line parameter $l = [a\ b]^T$ and its covariance matrix Σ_l can be calculated by using a least squares method.

Figure 2: Matched segments by using the segment- and area-based stereo matching, and detected boundary edge positions.

3.2 Detection of boundary edge

The boundary edge position is obtained from valid disparity data corresponding to the given reference plane. Assuming that the boundary edge is a vertical line, we first set a position in the search region obtained from section 2.3 to a virtual mean position of the boundary edge. Then, we calculate differences between the virtual mean position and each boundary pixel of the planar region. As we regard the difference as an error of the virtual boundary position, the variance of the virtual mean position can be obtained. This process is applied to the hole search region in Eq. (6). Last, we can find the boundary position with the minimum variance.

3.3 Detection of features

In this paper vertical line segments which are distinct and on a vertical plane serve as the stable landmarks. First, vertical line segments are detected using Sobel operator, and then filtered out by comparing with thresholds of length and gradient. Using the segment-based stereo matching by DP[9], the vertical line segments are matched. Fig. 2 shows the matching results when the partition and white board of the distance of 4[m] and 6[m] from the robot are given. The boundary edges are located at $(0, 4000)$[mm] and $(-550, 6000)$[mm], respectively. White vertical lines in the left image are matched segments, and long vertical lines in the right image are detected boundary edge positions from disparity image. The horizontal lines on the top and bottom of edge positions show the uncertainty of edge positions. Gray regions in the right image are disparity data corresponding to the given reference plane. From the result of two processing, false matched segments are filtered out by comparing disparity values between matched segments and the disparity image.

3.4 Extraction of stable visual landmarks

Since the boundary edges are regarded as unstable features for visual observation according to the change of viewpoints, we filter out features nearby the boundary edges from the landmark candidates.

Then, we calculate the length similarity between observed stereo pairs using Eq. (11). If the similarity $\gamma(\ell_l, \ell_l)$ is less than the threshold, the pair is also filtered out from the stable landmarks.

$$\gamma(\ell_l, \ell_l) = \frac{min(\ell_l, \ell_r)}{max(\ell_l, \ell_r)} \qquad (11)$$

4 Modeling of visual landmarks

4.1 Estimation of feature position using the Kalman filter

The feature position obtained by stereo matching may be deviated from the reference plane $y = a + bx$ by uncertainties of vision and motion (see Fig. 3). Assuming that the feature is primarily located on the reference plane, we can express the positional relationship as the following non-linear equation:

$$y_c - (a + b * x_c) = Q(\boldsymbol{p}, \boldsymbol{l}) = 0\,, \qquad (12)$$

where $\boldsymbol{p} = [x_c\ y_c]^T$ denotes x, y position of feature $\boldsymbol{p}$ in the camera coordinates, and $\boldsymbol{l}$ is line parameter $[a\ b]^T$. By linearizing Eq. (12) a new liner observation equation is obtained.

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{p}_t + \boldsymbol{v}_t\,, \qquad (13)$$

where

$$\boldsymbol{y}_t = -Q(\tilde{\boldsymbol{p}}_t, \tilde{\boldsymbol{l}}_t) + \frac{\partial Q}{\partial \boldsymbol{p}_t}\tilde{\boldsymbol{p}}_t$$

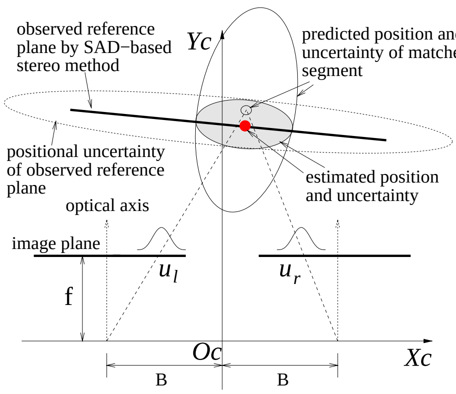

Figure 3: Combining two observation results.

$$\boldsymbol{H}_t = \frac{\partial Q}{\partial \boldsymbol{p}_t}$$

$$\boldsymbol{v}_t = \frac{\partial Q}{\partial \boldsymbol{l}_t}(\boldsymbol{l}_t - \tilde{\boldsymbol{l}}_t)$$

By using Kalman filter[11] the feature position on the reference plane and its uncertainty are estimated (see Fig. 3).

4.2 Modeling of stable visual landmarks

If a feature position on the reference plane is estimated, we can localize the position in the world coordinates system by matching the boundary edge position with the map. When the positions of boundary edges and extracted landmarks are estimated at $\boldsymbol{p}_b$ and $\boldsymbol{p}_f$ in the camera coordinates, respectively, the distance d_p is obtained, and d_p is also unchanged in the world coordinates(see Fig 4). Given a reference plane with the two boundary edges, $\boldsymbol{m}_1 = [x_1\ y_1]^T$ and $\boldsymbol{m}_2 = [x_2\ y_2]^T$, the landmark position $\boldsymbol{p}_{xy} = [x\ y]^T$ in the workspace $X - Y$ can be obtained by the following non-linear equation:

$$\begin{aligned} \boldsymbol{p}_{xy} &= \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} \frac{d_p}{d_m}(x_2 - x_1) \\ \frac{d_p}{d_m}(y_2 - y_1) \end{bmatrix} \\ &= \boldsymbol{M}(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{p}_b, \boldsymbol{p}_f), \end{aligned} \qquad (14)$$

where d_m denotes width of the given reference plane. By linearizing Eq. (14) the covariance matrix $\boldsymbol{\Sigma}_{p_{xy}}$ is obtained. If $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ have no uncertainty, $\boldsymbol{\Sigma}_{p_{xy}}$ can be expressed as follows:

$$\boldsymbol{\Sigma}_{p_{xy}} = \frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}_b}\boldsymbol{\Sigma}_{p_b}\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}_b}^T + \frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}_f}\boldsymbol{\Sigma}_{p_f}\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}_f}^T, \qquad (15)$$

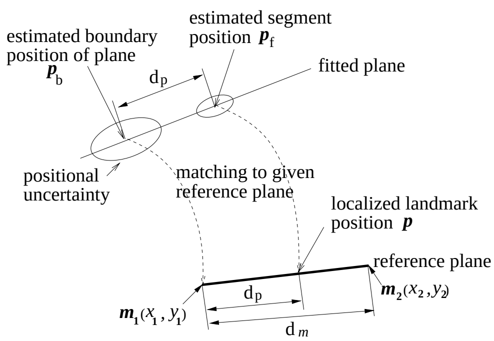

Figure 4: Localizing extracted landmarks in the map by matching the observed reference plane with the map.

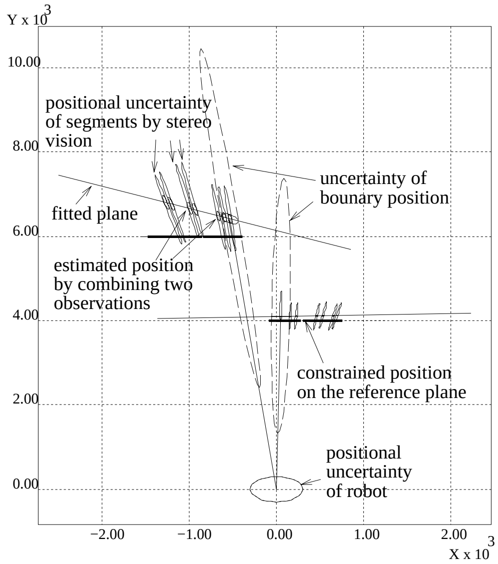

Figure 5: Positional uncertainty of the extracted landmarks in Fig. 2.

where $\boldsymbol{\Sigma}_{p_b}$ and $\boldsymbol{\Sigma}_{p_f}$ denote the positional uncertainty of the boundary and landmarks, respectively.

From Eq. (15) we can see that the covariance matrix of localized landmarks, $\boldsymbol{\Sigma}_{p_{xy}}$, is composed of the uncertainties of the boundary position and landmark, and the rank of covariance matrix is 1. As a result, the uncertainty of the extracted landmarks is constrained on the reference plane.

Fig. 5 shows the positional uncertainty of the extracted landmarks in Fig. 2. By combining the segment position and the fitted plane using the Kalman filter explained in section 4.1, the positional uncertainty was reduced.

5 Experimental results

We conduct experiments in our lab shown in Fig. 6 given the partition and white board as the reference planes. The robot motion is performed by the following rules: While processing the observed information, the robot moves to the next viewpoint without stopping. If the robot meets one of the following three cases: 1. it has no extracted landmarks, 2. the distance to the attentive boundary edge of the reference plane is nearer than d_{th}, and 3. the gaze angle is larger than ϕ_{th}, it stops its motion and extracts a set of new landmarks. Then, it moves to next viewpoints using the extracted landmarks, repeatedly.

In this experiments we set the threshold values of distance and gaze angle to $d_{th} = 1.5$[m], $\phi_{th} = 25$[deg], respectively. As the result the robot stopped at the third viewpoint at the distance of 2[m] from the initial position in order to perform the landmark selection process.

We test two cases. At first the robot moves using the on-line extracted landmarks under a normal condition environment. Next it moves in the same environment but the illumination condition and some features are changed, e.g., light turned off and positions of some features are changed.

Fig. 7 shows selected boundary edges and landmarks when the robot observes the partition and white board as the reference plane at initial position and after moving 2[m] under normal lighting condition, and Fig. 8 shows the motion result. The robot estimates the system state repeatedly by combining the observed landmark information. The ellipses drawn by thin line show the estimated robot uncertainties.

Fig. 9 also shows on-line extracted landmarks in the same environment but several lights are turned off and some features are changed. Fig. 10 shows the motion result under the changed environment.

From these results we can see that the robot can estimate its state successfully, and that the on-line extracted landmarks are useful for the visual observation under various conditions. Therefore, the proposed method is feasible for visual navigation.

6 Conclusion

In this paper, we proposed a method for a vision-based mobile robot to extract visual landmarks automatically under uncertainty.

As the robot on-line extracts the stable visual landmarks, it can perform stable visual observations and

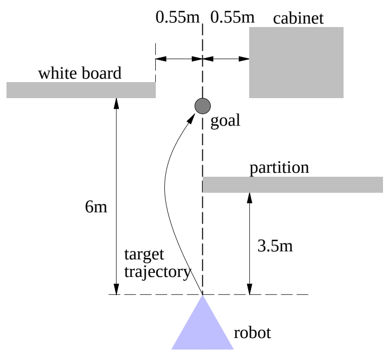

Figure 6: Indoor experimental environment.

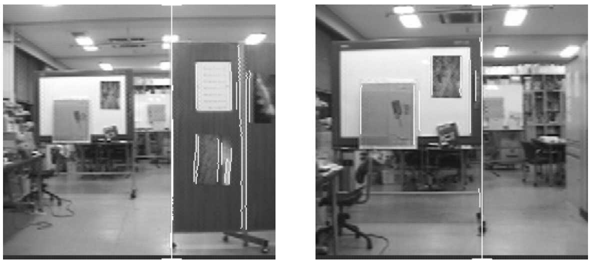

Figure 7: On-line extracted landmarks at the initial position and after 2m motion under normal lighting condition, respectively.

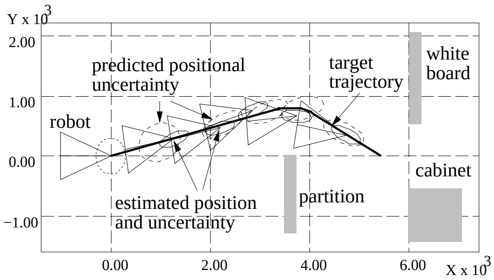

Figure 8: Motion result using on-line extracted landmarks under normal lighting condition.

successfully achieve the given task. From experimental results we can see that the robot can perform a robust and efficient navigation by combining two methods that are the on-line selection of viewpoints[10] and the extraction of landmarks proposed in this paper.

Currently we assume that the environment is stationary and known in advance. It is a future work to extend this study to the dynamic environment including dynamic obstacles.

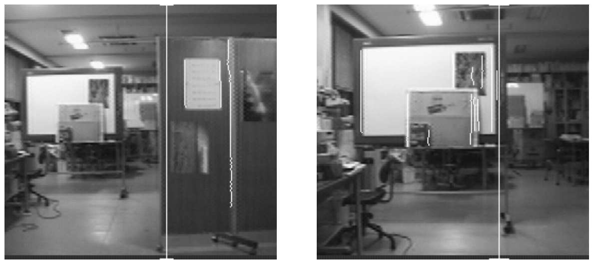

Figure 9: On-line extracted landmarks but several lights are turned off and some features are changed.

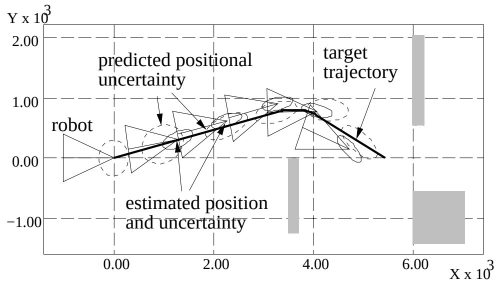

Figure 10: Motion result under the changed environment.

References

[1] I.J. Cox, "Blanche-an experiment in guidance and navigation of an autonomous robot vehicle," *IEEE Trans. on Robotics and Automation*, Vol. 7, No. 2, pp. 193-204, 1991.

[2] C. Becker, J. Salas, K. Tokusei, and J.C. Latombe, "Reliable Navigation Using Landmarks," in *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 401-406, 1995.

[3] R. Talluri and J.K. Aggarwal, "Mobile Robot Self-Location Using Model-Image Feature Correspondence," *IEEE Trans. on Robotics and Automation*, vol. 12, No. 1, pp. 63-77, 1996.

[4] G. Hager, D. Kriegman, E. Yeh, and C. Rasmussen, "Image-based Prediction of Landmark Features for Mobile Robot Navigation," in *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 1040-1046, 1997.

[5] R. Sim and G. Dudek, "Mobile Robot Localization from Learned Landmarks," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1060-1065, 1998.

[6] D. Fox and W. Burgard, "Active Markov Localization for Mobile Robots," *Robotics and Autonomous Systems*, Vol. 25, pp. 195-207, 1999.

[7] J.J. Little, J. Lu, and D.R. Murray, "Selecting Stable Image Features for Robot Localization Using Stereo," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1072-1077, 1998.

[8] J. Miura and Y. Shirai, "Vision and Motion Planning for a Mobile Robot under Uncertainty," *The Int. Journal of Robotics Research*, Vol. 16, No. 6, pp. 806-825, 1997.

[9] S.H. Lee and J.J. Leou, "A Dynamic Programming Approach to Line Segment Matching in Stereo Vision," *Pattern Recognition*, Vol. 28, No. 8, pp. 961-986, 1994.

[10] I. Moon, J. Miura, and Y. Shirai, "On-line Viewpoint and Motion Planning for Efficient Visual Navigation under Uncertainty," *Robotics and Autonomous Systems*, Vol. 28, pp. 237-248, 1999.

[11] P.S. Maybeck, "The Kalman Filter:An Introduction to Concepts," in: I.J. Cox and G.T. Wilfong, eds., *Autonomous Robot Vehicles*, Springer-Verlag, pp. 194-204, 1990.