International Journal of Control, Automation, and Systems (2009) 7(1):97-104 http://www.springer.com/12555
DOI 10.1007/s12555-009-0112-0

# Mobile Robot Localization based on Effective Combination of Vision and Range Sensors

**Yong-Ju Lee, Byung-Doo Yim, and Jae-Bok Song***

**Abstract:** Most localization algorithms are either range-based or vision-based, but the use of only one type of sensor cannot often ensure successful localization. This paper proposes a particle filter-based localization method that combines the range information obtained from a low-cost IR scanner with the SIFT-based visual information obtained from a monocular camera to robustly estimate the robot pose. The rough estimation of the robot pose by the range sensor can be compensated by the visual information given by the camera and the slow visual object recognition can be overcome by the frequent updates of the range information. Although the bandwidths of the two sensors are different, they can be synchronized by using the encoder information of the mobile robot. Therefore, all data from both sensors are used to estimate the robot pose without time delay and the samples used for estimating the robot pose converge faster than those from either range-based or vision-based localization. This paper also suggests a method for evaluating the state of localization based on the normalized probability of a vision sensor model. Various experiments show that the proposed algorithm can reliably estimate the robot pose in various indoor environments and can recover the robot pose upon incorrect localization.

**Keywords:** Mobile robot localization, sensor fusion, sensor model, vision-based navigation.

## 1. INTRODUCTION

Localization is a method that uses an environmental map and the information from sensors mounted on a robot to estimate the pose (position and orientation) of a robot. Localization is fundamental and important for successful navigation of the robot because most functions the robot offers require the robot pose relative to the environment.

A laser scanner has been extensively used for pose estimation of a robot [1,2]. It can successfully estimate the robot pose even in a large space, but its high cost prevents it from being widely used for the service robots at home and office environments. Other range sensors such as sonar sensors or IR scanners can substitute for the laser scanner, but they suffer from inaccuracies. Furthermore, when only the range sensors are used, it is difficult to accurately estimate the robot pose when the robot is traveling in a symmetrical environment such as a hallway.

On the other hand, a vision sensor usually provides more information than a range sensor with good performance at a relatively low cost. Therefore, localization using vision sensors has drawn much attention in recent years [3,4]. However, most algorithms for vision-based localization suffer from shortcomings in that their implementation takes longer than those for range-based localization because the extraction of feature information from vision sensors is more sophisticated than that from range sensors.

Localization using relatively cheap sensors is important from a practical point of view, but localization with low-cost sensors seldom provides good performance due to measurement inaccuracies in various environments. Neither the range-based nor the vision-based scheme alone can overcome these sensor limitations; therefore, a combination of various sensors should be implemented to overcome the shortcomings of each sensor.

Several researchers have combined range and vision sensors to overcome the limitations of each individual sensor. The localization scheme that combines Markov localization and Kalman filter localization using polyhedrons extracted by vision sensors and corner features extracted by sonar sensors was proposed in [5]. Another scheme that combines the 3D depth information from a stereo camera with the range information from a sonar sensor in a particle filter was proposed in [6]. However, these existing schemes could not be practically used because they employed an expensive stereo camera. Moreover, most of them use simple features, which limit their applications.

This paper proposes a localization algorithm based on

Manuscript received June 16, 2007; revised March 31, 2008; accepted September 18, 2008. Recommended by Editorial Board member Sooyong Lee under the direction of Editor Hyun Seok Yang. This research was conducted by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy of Korea.

Yong-Ju Lee and Jae-Bok Song are with the School of Mechanical Eng., Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (e-mails: {yongju_lee, jbsong}@korea.ac.kr).

Byung-Doo Yim is with the Dept. of Mechatronics, Korea University, Seoul, Korea (e-mail: ybd1021@korea.ac.kr).

* Corresponding author.

the combination of the range information from a low-cost IR scanner and the visual information from a monocular camera. The proposed localization scheme is mainly based on the Monte Carlo Localization (MCL) algorithm, the most widely used in particle filters [7,8]. Compared to a laser scanner, an IR scanner provides poor range accuracy but it is much cheaper than a laser scanner, making an IR scanner more practical than a laser rangefinder [9]. Objects are recognized by a vision sensor using the SIFT algorithm [10]. Relatively poor range accuracy from an IR scanner can be compensated by vision-based localization, and slow visual object recognition can be overcome by the frequent updates of the range information.

One problem involved in the combination of the range and visual data is the different processing times of these two types of data. Range data are updated faster than visual data because object recognition based on the SIFT algorithm requires a long computation time, especially when the object has many features. In this paper, the data from the two sensors are synchronized using the encoder information. Therefore, both types of sensor data can be used to update the robot pose without time delay.

Another issue of localization is the evaluation of the state of localization. Detection and recovery from localization failure is essential for achieving reliable and robust navigation. Random samples are used in MCL to cope with the kidnapped robot problem and localization failure [11], but if the number of random samples is not sufficient, the detection of localization failure would take a very long time. This paper proposes a scheme that evaluates the state of localization based on the normalized probability of a vision sensor model. If localization failure is detected, then the proposed scheme can recover the robot pose using a recognized object.

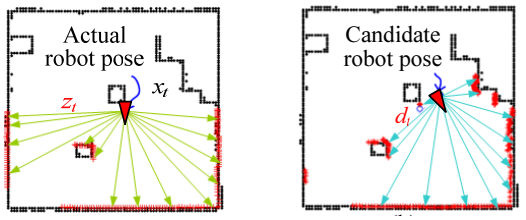
The remainder of this paper is organized as follows. Section 2 presents a range sensor model and a vision sensor model. Section 3 presents the fusion of the two sensor models for MCL. Section 4 presents a scheme for the detection and recovery from localization failure. Finally, section 5 presents the conclusions.

## 2. SENSOR MODELS

In this research, the range and vision sensors are combined to improve localization of a mobile robot. Instead of a laser scanner, which is very accurate but expensive, an IR scanner is used as the main range sensor. The IR scanner generates 121 range values with a resolution of 1.8°. An inexpensive monocular camera is also employed as the main vision sensor instead of a stereo camera, which can provide range information but which is expensive. Objects are recognized by the well-known SIFT algorithm that extracts the visual features. A sensor model for each sensor is required for the probability update of samples (i.e., candidates for the robot pose) in MCL.

### 2.1. Range sensor model

In the range sensor model, the probability of samples




(a) Measured by the actual sensor. (b) Computed from candidate robot pose.

Fig. 1. Range data.

is updated according to the difference between the range data measured by the IR scanner and those computed from the sample pose on the map, as shown in Fig. 1. That is, if the robot pose at time $t$ is denoted as $x_t$, the probability of sample $i$ ($i$ = 1, …, $N$) is updated by

$$p_{ir}^{(i)}(z_t \mid x_t) = \frac{1}{\sum_{k=1}^{k_t}(z_t(k) - d_t^{(i)}(k))^2}, \quad (1)$$

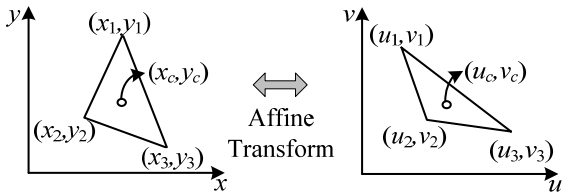
where $z_t(k)$ represents the $k$-th value of the range data measured at time $t$ ($k$ = 1, .., 121), and $d_t^{(i)}(k)$ is the $k$-th value of the range data computed from sample $i$ on the map. Although the IR scanner provides a total of 121 range data, only the data of ranges less than 4m ($k_t$ is the total number of such data) are used in (1) because the data of ranges exceeding 4m were found to be incorrect.

### 2.2. Vision sensor model

In the vision sensor model, the probability is updated according to the difference between the measured range and angle to the recognized object and those computed from the sample poses on the map. A center point of an object is used to calculate the range and angle of the recognized object. The center of an object is a centroid of keypoints extracted from the object by the SIFT algorithm. The following affine transform, which calculates the geometrical relations between the recognized object and the object stored in the database, is used to extract the accurate center point [12].

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (2)$$

where the vector $[t_x, t_y]^T$ is associated with the translation and the parameter $m_j$ ($j$ = 1, .., 4) with the 3D rotation. The vector $[x_i, y_i]^T$ is the keypoint relative to the image stored in the database, and $[u_i, v_i]^T$ is the keypoint extracted in the camera image, as shown in Fig. 2.




(a) Obiect in database. (b) Recognized object.

Fig. 2. Example of affine transform.

Equation (2) is rearranged to compute the 6 parameters as follows:

$$\begin{bmatrix} x_1 & y_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & y_1 & 0 & 1 \\ x_2 & y_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_2 & y_2 & 0 & 1 \\ x_3 & y_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_3 & y_3 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \end{bmatrix}. \quad (3)$$

The parameters can be computed by inversion of the 6 by 6 matrix if only 3 pairs of matched keypoints are given, as shown in Fig. 2. For more than 3 pairs, a pseudo inverse matrix is used to yield the 6 parameters. Once the parameters are identified, the center point of the recognized object, ($u_c$, $v_c$), can be computed by (2) from that of the object in the database, ($x_c$, $y_c$).

The angle of the recognized object can be obtained by Eq. (4) and Fig. 3(a). This angle at time $t$, denoted as $\theta_t^{obj}$, is given by

$$\theta_t^{obj} = \tan^{-1}\left( \frac{(w_{image}/2) - u_c}{(w_{image}/2)} \cdot \tan(\theta_{fov}/2) \right), \quad (4)$$

where ($x_r$, $y_r$) of Fig. 3(a) represents the robot frame, $u_c$ is the coordinate of the center point relative to the image frame, $w_{image}$ is the number of pixels (e.g., 320 pixels) of the image plane in the $u$ axis of the image frame, and $\theta_{fov}$ is the camera's field of view.

In this research, the IR scanner is used to provide the distance to the recognized object. The position and heading of the IR scanner is identical to those of the camera. The scanning range of the IR scanner always includes the field of view of the camera, as shown in Fig. 3(b). Therefore, the range data of the IR scanner corresponding to this angle is used as the distance to the center point of the recognized object. If the angle of the recognized object lies between two adjacent range data of the IR scanner, the distance to the recognized object is calculated by interpolation.

In the vision sensor model shown in Fig. 4, the distance ($d_t^{obj}$) and the angle ($\theta_t^{obj}$) of the recognized object are compared with the range and angle of the object computed at the pose of sample $i$ on the map ($d_t^{(i)}$ and $\theta_t^{(i)}$). Based on the difference between the measured and the computed results, the probabilities are updated as follows:

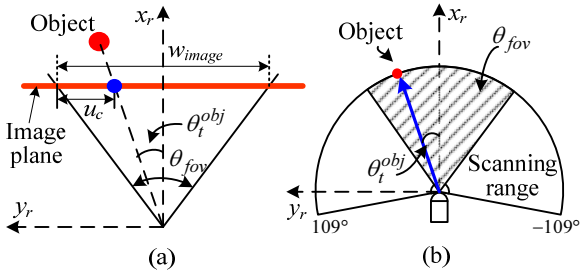



Fig. 3. Extraction of visual feature; (a) a center point of the object in the image, (b) detection range of IR scanner.

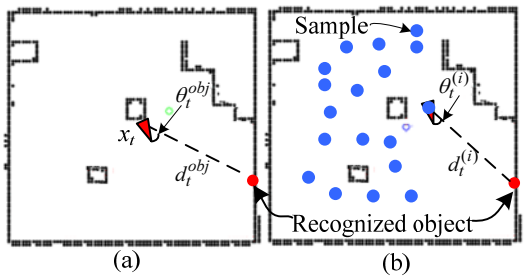



Fig. 4. Range and relative angle to the recognized object; (a) measured from the robot, and (b) computed from sample $i$.

$$p_d^{(i)}(z_t \mid x_t) = \eta_r \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{1}{2}\frac{(d_t^{obj} - d_t^{(i)})^2}{\sigma_d^2}\right), \quad (5)$$

$$p_\theta^{(i)}(z_t \mid x_t) = \eta_\theta \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{1}{2}\frac{(\theta_t^{obj} - \theta_t^{(i)})^2}{\sigma_\theta^2}\right), \quad (6)$$

where $p_d^{(i)}(z_t \mid x_t)$ and $p_\theta^{(i)}(z_t \mid x_t)$ are the probabilities associated with the range and angle, and $\eta_r$ and $\eta_\theta$ are the normalizing constants for the range and angle, respectively. Each sensor model has a Gaussian distribution with a mean of $d_t^{obj}$ and $\theta_t^{obj}$, and a variance of $\sigma_d^2$ and $\sigma_\theta^2$, respectively. For each sensor model, the overall vision sensor model is given by

$$p_v^{(i)}(z_t \mid x_t) = p_d^{(i)}(z_t \mid x_t) \times p_\theta^{(i)}(z_t \mid x_t). \quad (7)$$

## 3. COMBINATION OF RANGE AND VISION SENSORS

If there are many objects with visual features in the environment, vision-based localization can generally give a more effective localization performance than range-based localization. However, vision-based localization alone is not sufficient to provide satisfactory localization performance in most environments because only a small number of objects can be used as visual features in normal indoor environments. Thus, if the recognized objects cannot be found at the current robot pose, only the range sensor model is used to update the probability of the samples as follows:

$$p(z_t \mid x_t) = p_{ir}(z_t \mid x_t), \quad (8)$$

where $p_{ir}(z_t|x_t)$ is the range sensor model given by (1). If the vision sensor recognizes any object, the range sensor model and the vision sensor model are combined to update the probability of samples as shown below.

$$p(z_t \mid x_t) = p_v(z_t \mid x_t)\, p_{ir}(z_t \mid x_t), \quad (9)$$

where $p_v(z_t|x_t)$ is the vision sensor model given by (7).

It is important that the data from the IR scanner and the vision sensor are combined synchronously. However, in contrast to the relatively fast response of the IR scanner, vision-based object recognition often requires a rather long processing time. Because the robot moves during object recognition, the distance and angle to the object recognized at the beginning cannot be used to update the current robot pose. In other words, the result of object recognition is related to the robot pose at the beginning of object recognition.

For synchronization, the range data obtained from an IR scanner at the beginning of object recognition must be combined with the visual data at the end of object recognition. As the processing time for object recognition increases, several sets of recent range data should be discarded for synchronization with the vision data, as shown in Fig. 5(a). Thus, the overall update rate of the sample probability becomes slow, so the estimation time of the robot pose becomes longer. Consequently, the converged samples of MCL are behind the true robot pose, as shown in Fig. 5(b). In order to cope with this problem, the range data and the vision data are used independently. That is, all the range data are used to update the probability of the samples while object recognition is in process. The range and visual data are combined only when the object recognition is completed and the visual data are available.

Fig. 6 illustrates the delay of object recognition. Suppose that object recognition starts at the robot pose of

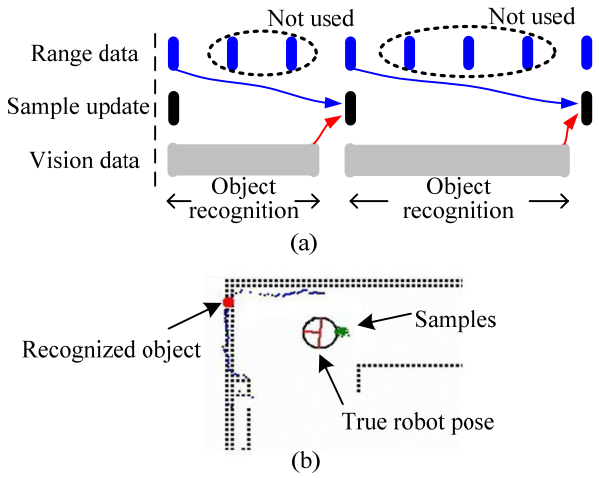



Fig. 5. Sensor fusion with loss of range information.

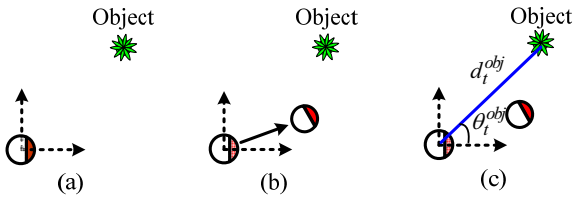



Fig. 6. Relation between robot and object (a) at the beginning of object recognition, (b) during object recognition, and (c) at the end of object recognition.

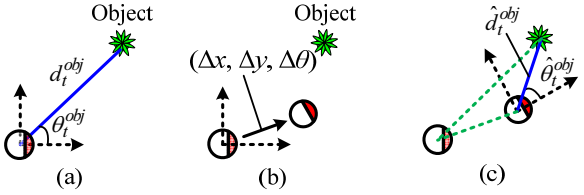



Fig. 7. Compensation of observations by using encoder data accumulated during object recognition.

Fig. 6(a). During object recognition, the robot moves, as shown in Fig. 6(b). In Fig. 6(c), an object is recognized and the distance and angle to the recognized object are obtained. However, the observation $(d_t^{obj}, \theta_t^{obj})$ is actually related to the robot pose of Fig. 6(a). Therefore, this observation should be modified to be applicable to the robot pose of Fig. 6(c). This compensation is performed by the encoder data accumulated during object recognition. It is assumed that the change of the robot pose estimated by the encoder data over a short period of time (about 200~400ms) is relatively accurate.

Fig. 7 illustrates how the encoder data compensates the delay. Fig. 7(a) represents the range and angle of the recognized object at time $t$. The displacement of the robot during object recognition are recorded as ($\Delta x$, $\Delta y$, $\Delta\theta$), as shown in Fig. 7(b). By combining the encoder data with the object data, the range and angle of the recognized object relative to the current robot pose can be estimated, as shown in Fig. 7(c).

The uncertainty of the encoder data should be considered in the computation of $(\hat{d}_t^{obj}, \hat{\theta}_t^{obj})$ as follows:

$$\begin{aligned} \sigma_d + \alpha_1 |\Delta d| + \alpha_2 |\Delta\theta| &= \sigma_{\hat{d}}, \\ \sigma_\theta + \alpha_3 |\Delta d| + \alpha_4 |\Delta\theta| &= \sigma_{\hat{\theta}}, \end{aligned} \tag{10}$$

where $\sigma_d$ and $\sigma_\theta$ are the uncertainties of the distance and angle measured by the vision sensor, and $\Delta d$ and $\Delta\theta$ are the translational and rotational motions of the robot during object recognition, respectively. The parameters, $\alpha_1$ and $\alpha_2$ associated with $\sigma_{\hat{d}}$, and $\alpha_3$ and $\alpha_4$ with $\sigma_{\hat{\theta}}$, depend on the characteristics of the robot. In the actual experiments, $\sigma_d$ and $\sigma_\theta$ were set to 0.2m and 3°, respectively. The parameters $\alpha_1$, $\alpha_2$, $\alpha_3$ and $\alpha_4$ were set to 0.1, 0.2, 0.5 and 2, respectively. The uncertainties $\sigma_{\hat{d}}$ and $\sigma_{\hat{\theta}}$, which increase with $\Delta d$ and $\Delta\theta$, are used in (5) and (6) instead of $\sigma_d$ and $\sigma_\theta$. Using this synchronization, all the range data from the IR scanner can be used for sample update, as shown in Fig. 8(a). Localization based on this scheme can be conducted more frequently and efficiently than that by Fig. 5(a). Compared to Fig. 5(b), therefore, the samples can keep track of the true robot pose reasonably well, as shown in Fig. 8(b).

Various experiments were conducted using a robot equipped with an IR scanner (Hokuyo PBS-03JN) and a monocular camera (a web camera). As shown in Fig. 9(a),

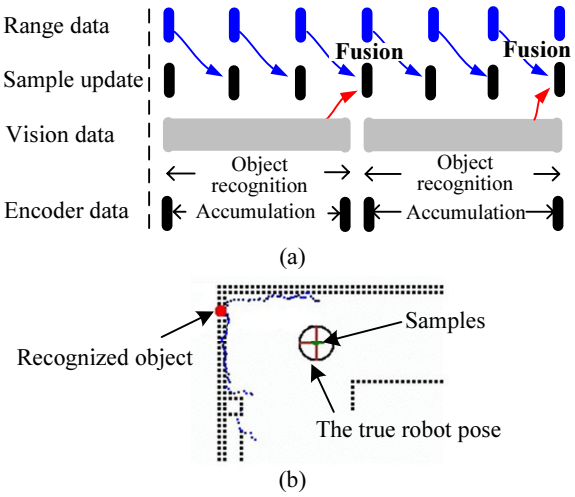



Fig. 8. Sensor fusion without loss of sensor information.

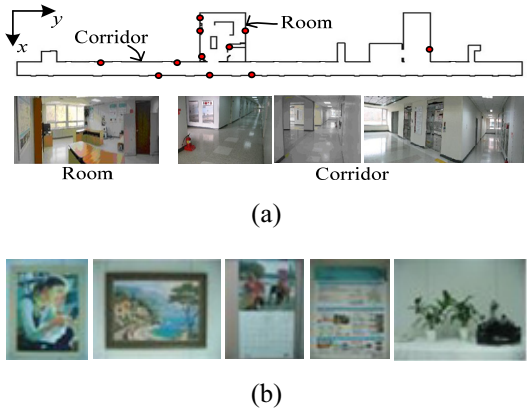
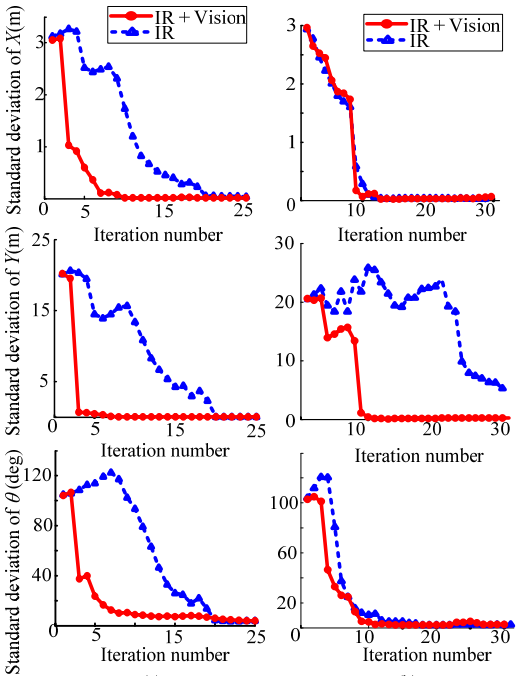



Fig. 9. (a) global map of experimental environment, and (b) objects used as visual landmarks.

the experimental environment is 15m x 80m and consists of a long hallway and several doors. The grid size of the grid map is 10cm. Fig. 9(b) illustrates the objects used as visual landmarks for localization, and their positions are shown as red dots in Fig. 9(a).

Eleven visual landmarks were used in the experiments for localization. The experiments were carried out in a room and a hallway. When the robot navigated through a hallway, the range information from the IR scanner is not sufficient for successful localization, which often led to slow convergence of the samples because of symmetry. Five thousand samples were initially distributed throughout the entire environment and these samples were converged to a small local area by continuously updating the probability of samples using the sensor information. After the deviations of the sample distribution became smaller than a pre-determined threshold, the robot pose was estimated.

If the robot was placed in a room at the beginning of MCL, the sample deviations converge to zero and thus the estimated robot pose can keep track of the actual pose reasonably well because the range information from




(a) Room. (b) Hallway.

Fig. 10. Deviation of sample pose in $x$, $y$ axis and $\theta$ in localization.

the IR scanner covers the room sufficiently. Furthermore, if the vision data are combined with the IR scanner data, the sample deviations converge to zero more rapidly than with only the range data, as shown in Fig. 10(a). In the combination-based localization, samples quickly converge when objects are recognized. As shown in Fig. 10, the deviation of distribution of samples with respect to the $y$-axis (i.e., along the hallway) is much larger than that with respect to the $x$-axis because the length of the environment in the $y$-axis is much longer than that in the $x$-axis.

If a robot is located in a hallway at the beginning of MCL, localization performance is generally worse than if a robot is located in a room at the beginning of MCL. First, few geometric features can be collected by the range sensor because the geometric information in the hallway is symmetric. Furthermore, the IR scanner has a relatively short sensing range and the sample deviations associated with the $y$-axis do not converge satisfactorily; sometimes they diverge, as shown in Fig. 10(b). Even in this case, however, if the visual information is added for localization, the samples converge faster and the localization succeeds.

In the proposed method of sensor fusion, the probability update of samples becomes more efficient because all the data from the range and vision sensors can be utilized. From the experiments, objects are robustly recognized, but the average time for object recognition is about 500ms. However, the robot can move with an average speed of 0.3m/s and the pose estimation error is less than ±35cm and ±8° through synchronization.

## 4. RECOVERY FROM LOCALIZATION FAILURE

If the differences between the given environment map and the sensor information become large in the localization of a mobile robot, the samples used in MCL can either converge to wrong places or diverge, thus resulting in localization failure. Other sources of localization failures are caused by use of inaccurate encoder data resulting from slippage during navigation and the kidnapped robot problem. In case of localization failure, it is important to quickly detect this failure and recover from it for safe and reliable navigation.

To evaluate the state of localization, the probabilistic vision sensor model, (5) and (6) of Section 2, are adopted in this research. The probability associated with the vision sensor model depends on the uncertainties $\sigma_{\hat{d}}$ and $\sigma_{\hat{\theta}}$ given by (10). It is difficult to select a fixed probability threshold to determine whether localization is successful or not because these uncertainties change at every period of MCL. Therefore, a new criterion to evaluate the state of localization is proposed in this research.

Fig. 11(a) and (c) depict the probability distribution associated with the distance to the visual object. Both distributions are Gaussian with a mean of 2m, but their standard deviations (representing the uncertainty) are set to 0.2m for Fig. 11(a) and 0.4m for Fig. 11(c). Fig. 11(e) and (g) depict the probability distribution associated with the angle to the object. Both are Gaussian with a mean of 0°, but their standard deviations are 3° for Fig. 11(e) and 5° for Fig. 11(g). Figs. 11(b), (d), (f) and (h) represent the probability distributions normalized by their respective maximum probability (indicated as a red dot in Figs. 11(a), (c), (e) and (g)).

The evaluation of the state of localization is illustrated in Fig. 11(b), (d), (f) and (h). The state of localization or quality can be classified into three cases. Case A corresponds to the upper 30% of the normalized probability distribution and it is regarded as successful localization. Case C, the lower 30%, is recognized as localization failure and the remainder, Case B, is classified as a warning. For instance, suppose the computed distance and angle are given as 2.3m and 10° from the sensor model given in Fig. 11(c) and 11(g), which correspond to the normalized probability of 0.75 and 0.13, respectively. In this situation, localization is judged as a failure, because the normalized probability associated with the angle falls into region C even though the probability associated with the range is in region A.

If the normalized probability is less than 0.3 three times in a row, we judge that the result of localization is poor. In addition, if the normalized probability lies between 0.3 and 0.7 ten times in a row, we conclude that localization has failed. Fig. 12 shows the normalized probability calculated using the recognized objects during navigation. The normalized probability is greater than 0.7 during most of the navigation. Sometimes, the normalized probability falls into a warning or localization failure state, but it is not judged as localization failure because it does not continuously happen. In the case of region I in Fig. 12, the state of localization is judged as localization failure because the normalized probability associated with the distance is smaller than 0.7 ten times due to the slippage. In addition, in the case of region II, the state of localization also is judged as localization failure, because both normalized probabilities are under 0.3 three times due to some problems such as kidnapping.

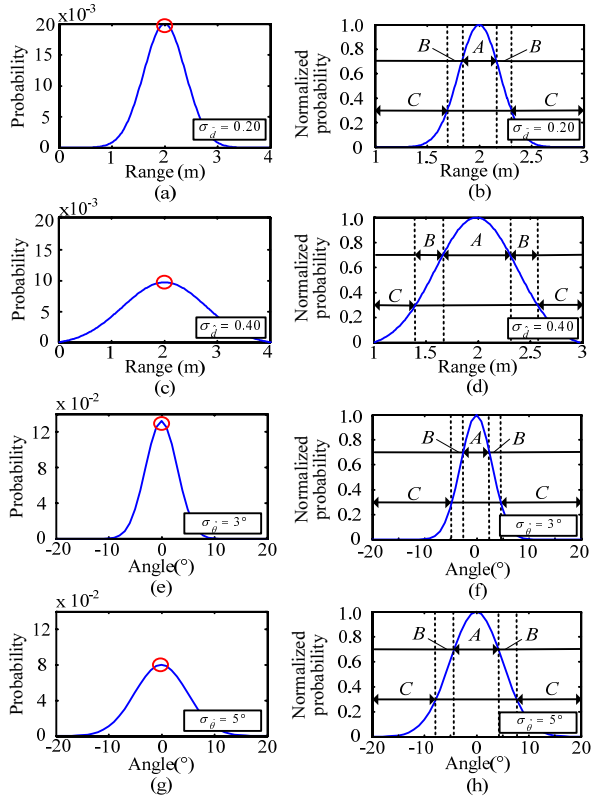



Fig. 11. Decision of the state of localization based on vision with uncertainty: (a), (c), (e) and (g) the probability distribution of a visual feature, (b), (d), (f) and (h) normalized probability and the state of localization, *A*: successful localization, *B*: warning, *C*: localization failure.

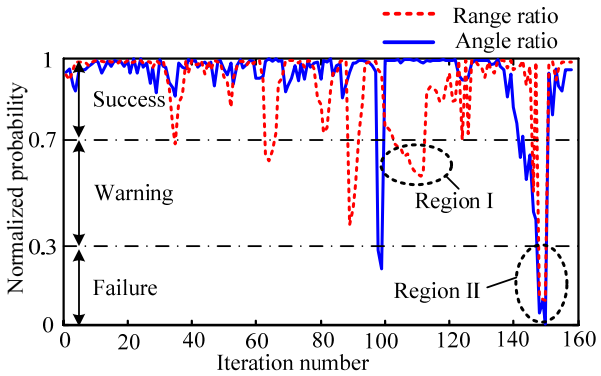



Fig. 12. Normalized probability associated with the range and angle.

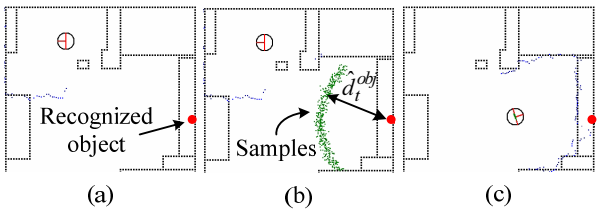



Fig. 13. Recovery from localization failure; (a) detection of localization failure, (b) distribution of samples near recognized visual feature, and (c) recovery of the robot pose.

Fig. 13 shows recovery from localization failure using a recognized object. Once the robot is aware of localization failure, it wanders to collect visual objects while the range-based localization is in progress. If an object is recognized by the vision sensor, it becomes needless to distribute all the samples for MCL on the entire empty area of the environment because the position of the recognized object is known from the map information.

As shown in Fig. 13(b), samples are mainly drawn near the circle with a radius of the measured range and centered at the recognized object. Obviously, this sample distribution is more efficient in localization than the uniform distribution covering the entire environment. Although visual features are sparse in some environments, vision-based recovery of the robot pose is efficient and robust provided that visual features are available.

## 5. CONCLUSIONS

This paper proposed an efficient sensor fusion based localization algorithm, in which an IR scanner and an inexpensive monocular camera are used. From this research the following conclusions have been drawn.

1) The proposed localization algorithm based on sensor combination enabled the samples in MCL to converge to the actual robot pose faster than either range-based or vision-based localization algorithm alone.
2) Although the processing time for object recognition took a long time, the probability of samples was updated in real time through encoder-based synchronization.
3) The proposed algorithm for the evaluation of the state of localization based on the vision sensor model worked well to detect localization failure and recover from it.

Currently, research on the improvement in the accuracy and speed of the object recognition algorithm is under way.

## REFERENCES

[1] Y. J. Lee, T. B. Kwon, and J. B. Song, "SLAM of a Mobile Robot using Thinning-based Topological Information," *International Journal of Control, Automation, and Systems,* vol. 5, no. 5, pp. 577-583, October 2007.

[2] J. Gutmann, T. Weigel, and B. Nebel, "A fast, accurate, and robust method for self-localization in polygonal environments using laser range finders," *Advanced Robotics*, vol. 14, no. 8, pp. 651-668, February 2001.

[3] J. Kosecka and F. Li, "Vision based topological Markov localization," *Proc. of IEEE Int'l Conf. on Robotics and Automation*, pp. 1481-1486, 2004.

[4] X. D. Nguyen, B. J. You, and S. R. Oh, "A simple framework for indoor monocular SLAM," *International Journal of Control, Automation, and Systems,* vol. 6, no. 1, pp. 62-75, February 2008.

[5] W. Shang and D. Sun, "Multi-sensory fusion for mobile robot self-localization," *Proc. of IEEE International Mechatronics and Automation*, pp. 871-876, 2006.

[6] S. Thimpson and S. Kagami, "Stereo Vision and Sonar Sensor Based View Registration for 2.5 Dimensional Map Generation," *Proc. of IEEE Intl. Conf. on Intelligent Robots and Systems*, pp. 3444-3449, 2004.

[7] D. Fox, W. Burgard, F. Dellaert, and S. Thrun, "Monte Carlo Localization: Efficient Position Estimation for Mobile Robots," *Proc. of the Sixteenth National Conference on Artificial Intelligence*, pp. 343-349, 1999.

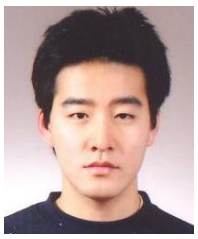
[8] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo Localization for Mobile Robots," *Artificial Intelligence Journal,* vol. 128, no. 1-2, pp. 99-141, May 2001.

[9] M. Alwan, M. B. Wagner, G. Wasson, and P. Sheth, "Characterization of Infrared Range-Finder PBS-03JN for 2-D Mapping," *Proc. of IEEE Int. Conference on Robotics and Automation*, pp. 3936-3941, 2004.

[10] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no 2, pp. 91-110, November 2004.

[11] S. Thrun, W. Burgard and D. Fox, *Probability Robotics*, MIT Press, 2005.

[12] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," *Proc. of Int. Conference on Computer Vision*, pp. 273-280, 2003.



**Yong-Ju Lee** received the B.S. degree in Mechanical Engineering from Korea University in 2004. He is now a Student for Ph.D. of Mechanical Engineering from Korea University. His research interests include mobile robotics.

**Byung-Doo Yim** received the B.S. degree in Control and Instrumentation Engineering from Seoul National University of Technology in 2005. Also, he received the M.S. degree in Mechatroncis Engineering from Korea University in 2007. His research interests include mobile robotics.

**Jae-Bok Song** received the B.S. and M.S. degrees in Mechanical Engineering from Seoul National University in 1983 and 1985, respectively. Also, he received the Ph.D. degree in Mechanical Engineering from MIT in 1992. He is currently a Professor of Mechanical Engineering, Korea University, where he is also the Director of the Intelligent Robotics Laboratory from 1993. His current research interests lie mainly in mobile robotics, safe robot arms, and design/control of intelligent robotic systems.