Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Active-Semantic Localization with a Single Consumer-Grade Camera

Chuho Yi
Division of Electrical and Computer Engineering
Hanyang University
Seoul, Korea
d1uck@hanyang.ac.kr

Il Hong Suh †
College of Information and Communications
Hanyang University, Seoul, Korea
All correspondences should be addressed to
ihsuh@hanyang.ac.kr

Gi Hyun Lim
Division of Electrical and Computer Engineering
Hanyang University
Seoul, Korea
hmetal@hanyang.ac.kr

Byung-Uk Choi
Division of Electrical and Computer Engineering
Hanyang University
Seoul, Korea
buchoi@hanyang.ac.kr

***Abstract*—This study addressed the problem of active localization, which requires massive computation. To solve the problem, we developed abstracted measurements that consist of qualitative metrics estimated by a single camera. These are contextual representations consisting of perceived landmarks and their spatial relations, and they can be shared by humans and robots. Next, we enhanced the Markov localization method to support contextual representations with which a robot's location can be sufficiently estimated. In contrast to passive methodologies, our approach actively uses the greedy technique to select a robot's action and improve localization results. The experiment was carried out in an indoor environment, and results indicate that the proposed active-semantic localization yields more efficient localization.**

***Index Terms*—Contextual map, Active-Semantic Localization, Information Gain**

I. Introduction

Humans do not necessarily require precise quantitative information to perceive their current location or to move to another location. Instead, they remember a few landmarks, such as specific structures or distinct landmarks that delimit the area [18]. Then, they restructure their knowledge based on spatial context and apply this knowledge to the current situation [1][13]. This semantic localization method enables efficient high-level spatial recognition and localization and does not require metrically exact location; instead, it uses but many pieces of spatial information that are accumulated through visual sensors.

In general, metric data gathered using a single camera are inaccurate. Individual metric datasets are extremely unreliable if a robot's location is uncertain [2]. To cope with this problem of unreliable data, we can use contextual information, and the mutual association exploits the geometric relationship among landmarks or objects. Landmarks and objects may be used differently according to the situation, but here we will assume that they are identical.

As discussed above, humans generate a mental 'contextual map' consisting of landmarks and their spatial relationships, as shown in Fig. 1. Based on a specific structure or distinct landmarks, humans restructure their knowledge based on spatial contexts and then reapply the knowledge [1]. Even without exact quantitative data, the accumulation of many fragments of spatial context enables a sufficiently high level of space recognition and localization.

Many robot localization methods have been developed over the last decade based on grid-based maps [3], feature-based maps [4], topological maps [5], or semantic maps [6], or through adaptive selection of various types of maps [7].

Landmarks can be represented in terms of qualitative distance, bearing, and relationships on a topological map; these data are then used to define and use a contextual map. In our proposed representation, the spatial context includes observed objects, robot-to-object distance (r-o; the distance between the robot and a particular object), r-o bearing (the direction from a robot to a particular object), and object-to-object (o-o) relationships.

Thrun et al. conducted research on probabilistic localization by applying the Bayes rule to robot localization [8]. A topolog-

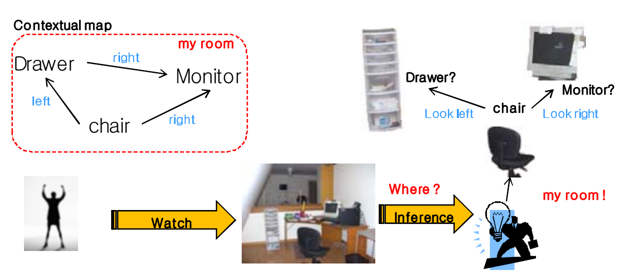

Fig. 1. Conceptual illustration of how a human recognizes a place with contextual information.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

ical map is represented semantically using ontology and relies on semantics to infer navigation [9][10]. Bailey et al. [2] noted that most localization schemes are considered passive in the sense that they did not use robot behaviors, e.g., turning to find a landmark, to collect evidence for localization.

Some research has applied active localization. Fox et al. applied a greedy technique to robot localization [11]. Recently, Kümmerle et al. proposed a method for determining range-laser sensor pan/tilt angles for use in orientating an action for multi-level outdoor surface maps. To reduce computation, each particle was applied to QT-clustering [12].

These approaches used range sensors to measure metric data and focused on building accurate metric maps. However, semantic maps require symbols that are associated with visual features using a single camera [13]. Our previous study focused on metric distances from a robot to an observed landmark, using a single camera to derive spatial contexts and approximate quantitative metrics. Spatial contexts and approximated quantitative metrics enable simple estimation of robot location with comparisons among a few variances. The image processing method has immense potential.

However, our previous approach was problematic in that we could only use information about landmarks and their spatial relationships. It is possible that continuous input will not be available, which could result in potential robot kidnapping problems. Therefore, we developed an active-semantic localization to solve input problems and improve localization efficiency through action selection.

We developed a method based on active-semantic localization using spatial relationships among landmarks. To solve the complication problem, we used measurements of landmarks and their spatial relationships. Exploring an unknown environment during localization may only require the measurement of a few spatial relationships. We conducted extensive localization experiments in an indoor environment; the results indicate that our active-semantic localization approach based on information gathering is valid.

II. Monte Carlo Localization in Contextual Map

A. Contextual Map

We used contextual representation and the Bayesian model to represent spatial relationships between objects (landmarks)[13]. Basically, most data are represented semantically by means of ontology, which uses inference to ensure that only sound and complete data are asserted and propagated. Noisy sensor data such as false positives and true negatives can be filtered using the relations and rules in logical reasoning. Many cases of false-positive data involve illogical properties, e.g., a misclassified object floating in the air by itself or penetrating other objects or walls. These cases can be evaluated logically using axiomatic rules. In addition, a robot can know what will be seen from the next step, which enables robots to predict and to pay attention to an a priori contextual map.

Object recognition is a fundamental factor in contextual representation. In general, an object can be recognized visually by measuring the similarity between its features and those

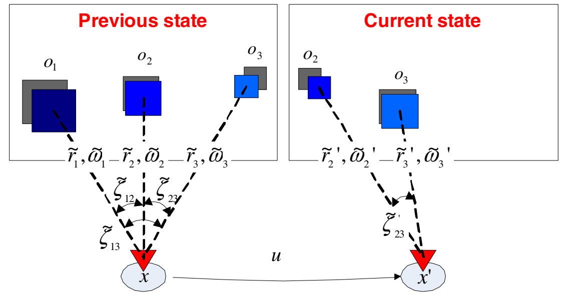

Fig. 2. Metric relationships between a robot and objects in image sequences

of corresponding object models. Our method uses feature transform features that are known to be invariant to image scale and rotation [14][15].

Fig. 2 shows changes in metric relationships between a robot and observed objects according to the transition in robot location from x to x'. Parameters $\tilde{r}$, $\tilde{\omega}$, and $\tilde{\zeta}$ denote estimated metric distance of an object relative to a robot, metric bearing of an object relative to a robot, and bearings among objects in the robot coordinates, respectively, measured by a single camera. In general, metric data collected using a single camera are inaccurate, and individual metric data quantities are extremely unreliable if the robot location is uncertain. Thus, for contextual representation, metric data are linked to symbols of spatial relationships, according to given conditions. However, mutual association exploits the geometric relationship between objects (landmarks) [2].

Table I shows a contextual representation using symbols for all spatial contexts of the objects in Fig. 2. Our robot localization application finds the robot position using only these types of contextual representations with qualitative metric data.

Here, the spatial context includes distance, bearing, and relationships. The r-o distance context, denoted by s^r, is the level of distance of the object from a robot. Each distance context is represented by one of a set of distance symbols: $s^r = \{nearby, near, far\}$. The r-o bearing context, denoted by $s^\omega = \{$ *front*, *left front*, *left*, *left rear*, *rear*, *right rear*, *right*, *right front*$\}$, is the bearing of the object relative to the robot. The o-o relationship context, denoted by $s^\zeta = \{$ *left far*, *left near*, *left nearby*, *right nearby*, *right near*, *right far*$\}$, is

TABLE I

Contextual representation including all of spatial contexts of object in Figure 2.

state	semantic representation
previous state	nearby(o_1, Robot), left front(o_1, Robot), right near(o_1, o_2), right far(o_1, o_3), far(o_2, Robot), front(o_2, Robot), left near(o_2, o_1), right near(o_2, o_3), far(o_3, Robot), right front(o_3, Robot), left far(o_3, o_1), left near(o_3, o_2)
current state	near(o_2, Robot), left front(o_2,Robot), right far(o_2, o_3) nearby(o_3, Robot), right front(o_3, Robot) left far(o_3, o_2)

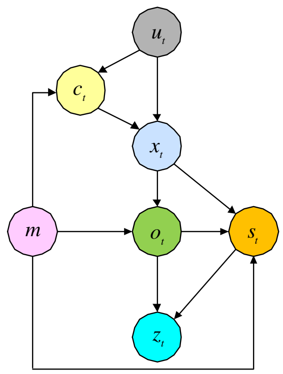

Fig. 3. Graphical model of Bayesian localization.

the relationship among objects.

B. Monte Carlo Localization

Fig. 3 is a graphical model of Bayesian localization. The robot is given a map, and its goal is to determine its location relative to this map given perceptions of the environment and its own movements.

In Fig. 3, the robot's location is denoted by $x = [v_{(i)}\ e_{(i,j)}\ \rho_{(i,j)}\ \theta_{(i,j)}]^T$ where $v_{(i)}$ and $e_{(i,j)}$ are the topological node and edge, respectively. $\rho_{(i,j)}$ and $\theta_{(i,j)}$ denote metric distance and bearing in edge $e_{(i,j)}$, respectively. The object is denoted by o. This metric-topological framework enables the localization to be bounded globally, the map size to increase monotonically in dimensionality, and the location to be calculated locally between two nodes. A set of semantics for spatial contexts of an object, containing distance, bearing, and relationship, is denoted by $s = \{s^r, s^\omega, s^\zeta\}$. A set of semantics for spatial contexts of the robot, containing distance and bearing, is denoted by $c = \{c^\rho, c^\theta\}$. Features extracted from the image are represented by z. The map is denoted by m.

We developed a topological-semantic distance map that consists of spatial contexts of both an object and the robot. In this map, nodes are one component. In a general topological map, nodes act as a standard and contain spatial object context data. In addition, in our proposed contextual representation, a robot's spatial context can yield an approximate distance and bearing from one assigned node to another. The approximate qualitative distance can be described as the node-to-node (n-n) distance context and the qualitative bearing as the node-to-node (n-n) bearing context.

A probabilistic robot location represents beliefs through conditional probability distribution. We can denote belief about a state x_t by $Bel(x_t)$; the posterior $Bel(x_t)$ can be obtained analogously using the Bayesian rule and the Markov assumption. In particular, we have:

$$\begin{aligned} &Bel(x_t) \\ &= \eta \cdot p(s_t, o_t, z_t|x_t, m) \int p(x_t|x_{t-1}, c_t, u_t) Bel(x_{t-1}) dx_{t-1} \end{aligned} \tag{1}$$

where the probability $p(s_t, o_t, z_t|x_t, m)$ is the contextual measurement, the probability $p(x_t|x_{t-1}, c_t, u_t)$ is the state transition, and $Bel(x_{t-1})$ is the belief at the time $t-1$, respectively. $\int p(x_t|x_{t-1}, c_t, u_t)Bel(x_{t-1})dx_{t-1}$ is the prediction model. The probabilistic localization model is divided into two parts: the measurement model and the prediction model, which correspond to two terms on the right-hand side of (1). The measurement model uses the contextual representation and is the main focus of this research. Context data are uncertain, so they should be approximated with a stochastic distribution. This section focuses on the measurement model and contexts of objects. Here, we assume that sensors are uncertain; thus, as fewer contexts are available, a lower distribution is approximated. The reverse would also be true. We can calculate the location posterior using the contextual representation described in Section II. A.

The contextual measurement model is based on the joint probability of robot location x, map m, spatial contexts of object s, object o, and extracted feature z, From the graphical model in Fig. 3, the joint probability can be written as:

$$p(s_t, o_t, z_t|x_t, m) = \frac{p(x_t)p(o_t|x_t, m)p(s_t|x_t, o_t, m)p(z_t|s_t, o_t)}{p(x_t, m)} \tag{2}$$

Here, we assume that probability $p(x_t)$ is the same as $p(x_t, m)$ because robot position x_t is located on map m. Applying Bayes' law to the contextual measurement model from (2):

$$\begin{aligned} &p(s_t, o_t, z_t|x_t, m) \\ &\quad = p(o_t|x_t, m)p(s_t|x_t, o_t, m)\frac{p(z_t)p(o_t|z_t)p(s_t|o_t, z_t)}{p(o_t)p(s_t|o_t)} \end{aligned} \tag{3}$$

We assume that probability $p(z_t)$ has a uniform distribution. Therefore, we can get the distribution from the measurement model as:

$$\begin{aligned} &p(s_t, o_t, z_t|x_t, m) \\ &\quad = \eta \cdot p(o_t|z_t)p(s_t|o_t, z_t)p(o_t|x_t, m)p(s_t|x_t, o_t, m) \end{aligned} \tag{4}$$

where η is the normalization constant. $p(o_t|z_t)$ is a term related to object recognition; it is evaluated based on similarity between extracted features of observed objects and features of the corresponding object models.

We used a supervised approach to build object models. Training data consisted of images, each of which contained only one object. These images were captured at every known reference distance. The object model consisted of features for object recognition, the distance from a camera to the corresponding object, and the height of the object in pixels.

$p(s_t|o_t, z_t)$ is the likelihood of similarity in spatial context between the estimated metric data for the observed object and spatial contexts. It is evaluated based on similarity between extracted features of observed objects and features of the corresponding object models.

From the results of estimated metric distances and bearings, an object's spatial contexts can be computed as:

$$p(s_t|o_t, z_t) = \{p(s_t^r|o_t, z_t), p(s_t^\omega|o_t, z_t), p(s_t^\zeta|o_t, z_t)\} \quad (5)$$

$$p(s_t^r|o_t, z_t) = \mathbf{N}(r^s - \tilde{r}, {\sigma_r}^2) \quad (6)$$

$$p(s_t^\omega|o_t, z_t) = \mathbf{N}(\omega^s - \tilde{\omega}, {\sigma_\omega}^2) \quad (7)$$

$$p(s_t^\zeta|o_t, z_t) = \mathbf{N}(\zeta^s - \tilde{\zeta}, {\sigma_\zeta}^2) \quad (8)$$

where $\mathbf{N}(\mu, \sigma^2)$ is normally distributed with mean μ and variance σ^2. r^s, ω^s, and ζ^s are trained at every known reference for the spatial contexts of distance, bearing, and relationship, respectively. The variances of σ_r^2, σ_ω^2, and σ_ζ^2 are introduced to reflect the uncertainty of the vision sensor. The further an object is from the robot or other objects, the less accurate the metric distance and bearing will be. More specifically, dividing the spatial context more finely will improve localization performance.

Probability $p(o_t|x_t, m)$ is the likelihood of similarity between observed objects in the current state and those in the previous state, and can be formulated as:

$$p(o_t|x_t, m) = exp(-\|o - o_x\|^2) \quad (9)$$

where o and o_x represent the observed object in the current and previous states, respectively. Probability $p(s_t|x_t, o_t, m)$ is the likelihood of the spatial context.

The likelihood of the spatial context can be computed as:

$$\begin{aligned} & p(s_t|x_t, o_t, m) \\ &= \textstyle\sum_a^N [f_c(s^{r_a} - s_x^r) f_c(s^{\omega_a} - s_x^\omega) \sum_b^N f_c(s^{\zeta_{ab}} - s_x^\zeta)] \end{aligned} \quad (10)$$

where $f_c(s_a, s_x) = exp(-\|s_a - s_x\|^2)$, and s and s_x represent the estimated spatial contexts of an object in the current and previous states, respectively. Each spatial context of an object belongs to one of some number of different distributions. In equation (10), each context is described by a component probability density function, and its mixture of distributions is the probability that an observation comes from this component. Here, a mixture of three normal distributions with different means may result in a density with three spatial contexts of an object, which is not modeled by standard parametric distributions.

In equation (1), the last term on the right-hand side is an update term. The control model u represents simple motion data related to a state of transition as follows:

$$\begin{bmatrix} v'_{(i)} \\ e'_{(i,j)} \\ \rho'_{(i,j)} \\ \theta'_{(i,j)} \end{bmatrix} = \begin{bmatrix} v_{(i)} \\ e_{(i,j)} \\ \rho_{(i,j)} \\ \theta_{(i,j)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta\rho \cdot \cos(\theta_{(i,j)} + \Delta\theta - \theta^c_{(i,j)}) \\ \Delta\theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \varepsilon_\rho \\ \varepsilon_\theta \end{bmatrix} \quad (11)$$

where $\Delta\rho$ and $\Delta\theta$ are robot motions of bearing and movement, respectively, and ε_ρ and ε_θ are zero-mean error variables of motion and bearing.

The localization posterior is calculated using equation (1), but the modification is not applied directly to our Bayesian model because equation (1) is not a predictable term. To solve this problem, we can modify (1) in the localization model to a particle filter, a type of recursive Bayesian estimation [8][16], to manage the complicated computations.

In the particle filter, samples of a posterior distribution are called particles and can be denoted by:

$$\chi_t := x_t^{[1]}, x_t^{[2]}, ..., x_t^{[Y]} \quad (12)$$

Each particle $x_t^{[y]}$ (with $1 \leq y \leq Y$) is a concrete instantiation of the state at time t. For each particle $x_t^{[y]}$, the so-called importance factors are used to incorporate the measurement into the particle set. Therefore, importance is the probability of the measurement under the particle, given by:

$$w_t^{[y]} = p(o_t|z_t) p(s_t|o_t, z_t) p(o_t|x_t^{[y]}, m) p(s_t|x_t^{[y]}, o_t, m) \quad (13)$$

By resampling particles with probability proportional to $w_t^{[y]}$, we can see that resulting particles are indeed distributed according to the product of the proposal and importance weights $w_t^{[y]}$.

$$Bel(x_t) = \eta' \cdot w_t^{[y]} \sum_y^Y p(x_t^{[y]}|x_{t-1}^{[y]}, c_t, u_t, m) Bel(x_{t-1}) \quad (14)$$

III. Active-Semantic Localization

To achieve active semantic localization, we can apply the greedy technique set out by Fox et al. [11]. First, we assume that a robot can execute a discrete set of actions A at a given time t. Each set of actions can be represented by one of a set of distance and bearing symbols, for example: $A = \{move\ forward\ 1\ step, \cdots, move\ forward\ K\ step,$ $move\ backward\ 1\ step, \cdots, move\ backward\ K\ step,$ $turn\ to\ left\ foward,$ $turn\ to\ left,$ $turn\ to\ left\ rear,$ $turn\ to\ right\ forward,$ $turn\ to\ right,$ $turn\ to\ right\ rear\}$. The benefit of a sensing action $a \in A$ can be determined by considering the uncertainty of the posterior $p(x'|o, s, a)$. The uncertainty of the location estimate can be calculated using the following entropy equation:

$$H_p(x) = -\int_x \textit{\textbf{bel}}(x) \mathbf{log} \textit{\textbf{bel}}(x) dx \quad (15)$$

The ideal action would allow a robot to determine its position with a high certainty. Therefore, the information gain $I_b(a)$ of an action a to change the robot location is defined by:

$$I_b(a) = H_p(x) - H_b(x'|o, s, a) \quad (16)$$

where $H_b(x'|o, s, a)$ defines the entropy of new location x' based on new evidences o and s obtained by execution of action a. In general, we do not know what measurement a robot will obtain after it moves to a new location according to action a. . Therefore, it is preferable to consider the expected entropy by integrating all possible measurements o, s:

$$I_b(a) = H_p(x) - E_{o,s}[H_b(x'|o, s, a)] \tag{17}$$

where $E_{o,s}[H_b(x'|o, s, a)]$ defines the expected entropy after the integration of measurements of recognized objects and spatial contexts obtained after execution of action a. Consequently, we can consider the conditional entropy of the action, with measurements integrated out:

$$\begin{aligned} &E_{o,s}[H_b(x'|o, s, a)] \\ &= \int \sum_s \sum_o [H_b(x'|o, s, a)p(o, s|x')p(x'|a, x)]dx \qquad (18) \\ &= \int \sum_s \sum_o [H_b(x'|o, s, a)p(o|x')p(s|x', o)p(x'|a, x)]dx \end{aligned}$$

Calculating the expected entropy would require a great deal of computation. To reduce the required computation time, we can simulate a subset calculated using particle filtering, because particles are typically located in a small number of areas with high probability.

$$\begin{aligned} &E_{o,s}[H_b(x'|o, s, a)] \\ &= \sum_{x^{[Y]}} \sum_s \sum_o [H_b(x'|o, s, a)p(o|x')p(s|x', o)p(x'|a, x^{[y]})] \end{aligned} \tag{19}$$

Next, action $\hat{a}$ can be selected out of the action set A to maximize information gain as follows:

$$\hat{a} = \underset{a \in A}{\textbf{argmax}} \; \{\alpha(H_p(x) - E_{o,s}[H_b(x'|o, s, a)]\} \tag{20}$$

The expected entropy $E_{o,s}[H_b(x'|o, s, a)]$ can be determined by calculating objects and their spatial contexts in the given contextual map.

In general, the information gain approach may require substantial time, to the point that it becomes impractical. Our proposed method uses measurements of objects and their spatial contexts. The number of potential cameras obtained at a specific location is $3^N \times 8^N \times 6^N$, where N is the number of objects and 3^N, 8^N, and 6^N are spatial contexts of qualitative distance, bearing, and relationships, respectively.

IV. Experimental Results

To evaluate our proposed method, a Pioneer 3 AT robot carrying a single consumer-grade camera (Logitech QuickCam Pro 4000) was driven around a $14m \times 26.5m$ indoor environment. The camera observed 16 objects (trained landmarks) during its travel. Distinctive objects such as doorplates, fireplugs, bulletin boards, panel boards, etc. were used for object recognition.

To derive a fragment of the spatial context, a single camera was used to estimate the distance from the robot to an observed object. After the object was recognized, its height in the image space was measured using a set of corresponding features, and then a metric distance was estimated using visual pattern recognition software (ViPR) [17].

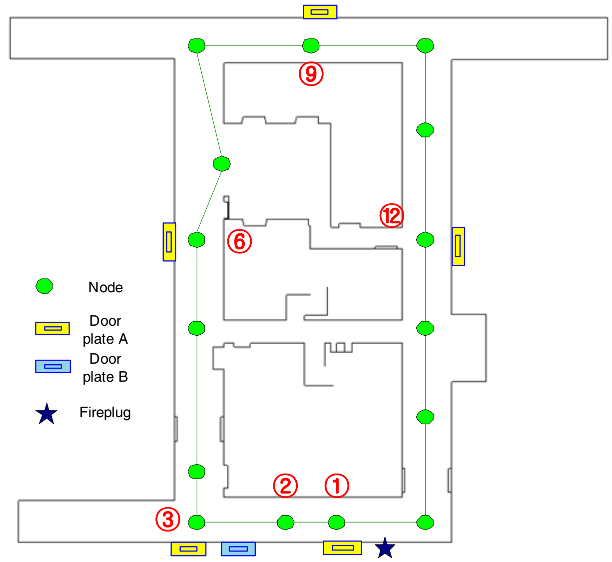

Fig. 4. Landmark (object) based map produced as a result of exploring the indoor environment.

Fig. 4 illustrates the indoor space, with objects and nodes marked on a ground-truth map. The green circles indicate the nodes in the contextual map, and the yellow and blue square boxes represent doorplates of Type A and Type B, respectively. It was difficult in this experiment to ensure recognition of doors using a scale invariant feature transform (SIFT), so we used the doorframe and the nearby doorplate for training. Type A doorplates were positioned to the right of the doorframe, and Type B doorplates were positioned to the left of the doorframe. In this map, Type A doorplates are identified in 6 nodes.

Fig. 5 shows the probabilistic distribution that denotes the robot location reflecting an object (Type A doorplate) and spatial contexts. As shown in Figs. 3 and 4, our active approach using information gain means the robot will prefer to turn left.

Fig. 6 depicts the improved accuracy of the location after the robot recognizes the fireplug in node 1.

V. Conclusion

We developed an active-semantic localization method that solves the problem of general methods requiring substantial time. To solve this problem, we developed abstracted measurements that consist of qualitative metrics. To this end, we developed a contextual representation and used the Bayesian model for robot localization by incorporating spatial contexts among objects, which were described using symbols. Then, we applied the greedy technique based on Monte Carlo methodology to the robot's action selection. We tested the proposed method using experiments conducted in a real indoor environment. Our analysis of action based on measurement revealed that this method enables localization that is more efficient than that of other methods.

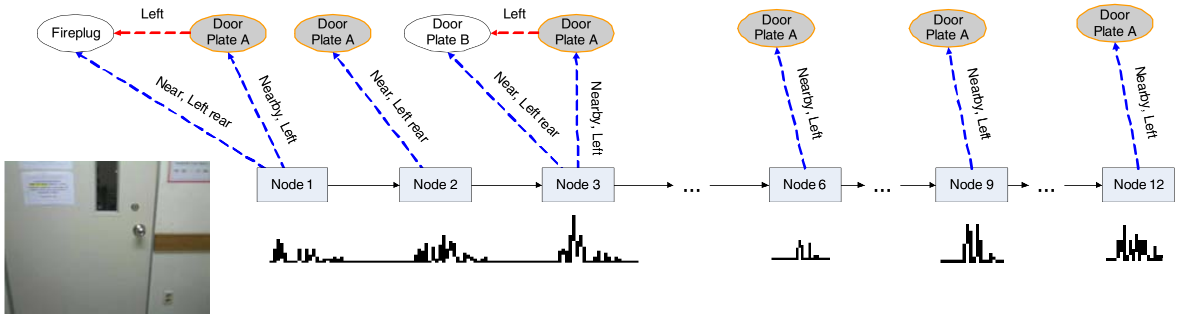

Fig. 5. Belief distribution when Type A doorplates are observed.

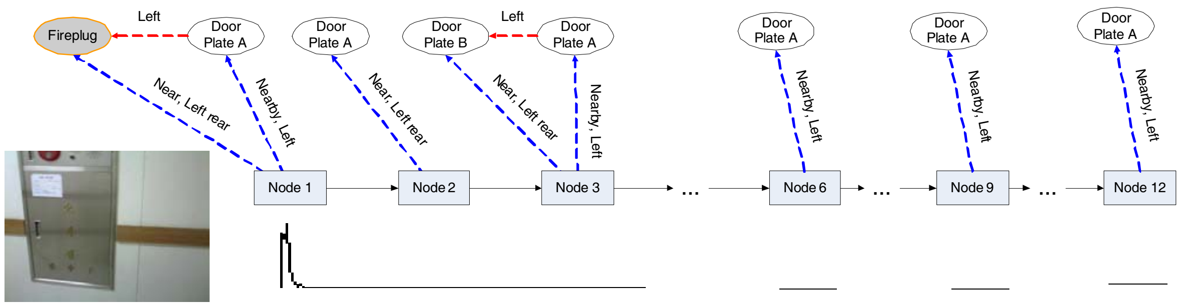

Fig. 6. Belief distribution when the fireplug is observed after turning.

ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program of MKE/KEIT [2008-F-038-01, Development of Context Adaptive Cognition Technology]

REFERENCES

[1] H. Choset, K.M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L.E. Kavrakij, and S. Thrun, *Principles of Robot Motion - Theory, Algorithms, and Implementations*, MIT-Press, 2005.

[2] T. Bailey and H. Durrant-Whyte, "Simultaneous Localization and Mapping (SLAM): Part II," *IEEE Robotics and Automation Magazine*, pp. 108-117, 2006.

[3] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *IEEE Computer, Vol.22*, pp.44-57, 1989.

[4] V Miro, J., Zhou, and W. Dissanayake, "Towards vision based navigation in large indoor environments," *Proceeding of IEEE the IROS*, pp. 2096-2102, 2006.

[5] P. Beeson, and B. Kuipers, "Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph," *Proceeding of IEEE ICRA*, pp. 4384-4390, 2005.

[6] B. Kuipers, "The Spatial Semantic Hierarchy," *Artificial Intelligence*, pp.191-233, 2000.

[7] S. Tully, H. Moon, D. Morales, G. Kantor, and H. Choset, "Hybrid Localization using the Hierarchical Atlas," *Proceeding of IEEE the IROS*, pp. 2857-2864, 2007.

[8] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*, MIT Press, Cambridge, MA, 2005.

[9] B. Krieg-Bruckner, U. Frese, K. Luttich, C. Mandel, T. Mossakowski and R. J. Ross, "Specification of an Ontology for Route Graphs," *Lecture notes in Computer Science*, pp. 390-412, 2005.

[10] J. Bateman and S. Farrar, "Modelling Models of Robot Navigation Using Formal Spatial Ontology," *Lecture notes in Computer Science*, pp. 366-389, 2004.

[11] D. Fox, W. Burgard, and S. Thrun, "Active Markov Localization for Mobile Robots," *Robotics and Autonomous Systems(RAS), 25*, pp. 195-207, 1998.

[12] R. Kummerle, P. Pfaff, R. Triebel, and W. Burgard, "Active Monte Carlo Localization in Outdoor Terrains using Multi-Level Surface Maps," *Journal of Field Robotics*, pp. 346-359, 2008.

[13] C. Yi, I. H. Suh, G. H. Lim, S. Jeong, and B. U. Choi, "Cognitive Representation and Bayeisan Model of Spatial Object Contexts for Robot Localization", *ICONIP*, pp.227-228, 2008.

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

[15] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," *Proceedings of Computer Vision and Pattern Recognition*, 2004.

[16] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte Carlo Localization for Mobile Robot", *Proceeding of the ICRA*, 1999.

[17] M. E. Munich, P. Pirjanian, E. D. Bernardo, L. Goncalves, N. Karlsson, and D. Lowe, "SIFT-ing Through Features with ViPR," *IEEE Robotics and Automation Magazine*, pp. 72-77, 2006.

[18] S. Werner, B. Krieg-Brückner, H. A. Mallot, K. Schweizer and C. F, "Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation," *Informatik*, pp. 41-50, 1997.