Cognitive Representation and Bayeisan Model of Spatial Object Contexts for Robot Localization

Chuho Yi, Il Hong Suh, Gi Hyun Lim, Seungdo Jeong, and Byung-Uk Choi

Department of Electronics and Computer Engineering, Hanyang University
17 Haengdang-dong, Sungdong-gu, Seoul, 133-791 Korea
{d1uck,ihsuh,hmetal,sdjeong,buchoi}@hanyang.ac.kr

Abstract. This paper proposes a cognitive representation and Bayesian model for spatial relations among objects that can be constructed with perception data acquired by a single consumer-grade camera. We first suggest a cognitive representation to be shared by humans and robots consisting of perceived objects and their spatial relations. We then develop Bayesian models to support our cognitive representation with which the location of a robot can be estimated sufficiently well to allow the robot to navigate in an indoor environment. Based on extensive localization experiments in an indoor environment, we show that our cognitive representation is valid in the sense that the localization accuracy improves whenever new objects and their spatial relations are detected and instantiated.

1 Introduction

Service robots are required to perform complex tasks with frequent human interaction. Suppose you are visiting a university and ask a staff member how to find Professor Suh's office. The person might say, "Go in the front door of the Information Technology Building and follow the first corridor on your right. You will see the washrooms on your left and classrooms on your right. When you reach the end of the corridor, turn left. At that point, you will see his office on your right." This description would enable you to find the correct office without any difficulty.

Humans do not necessarily require precise quantitative information to perceive space in their current location or to move to another location. Instead, they remember a few landmarks, such as specific structures or distinct objects that delimit the space. They then restructure their knowledge based on spatial contexts and apply that knowledge to the current situation [1]. This method may not metrically show the exact location, but as many pieces of spatial context are accumulated, it enables efficient high-level space recognition and localization.

Robot localization has required building an accurate metric map as well as a semantic map for symbolic inference. Moreover, it has required symbol matching between the metric data and the cognitive representation. An accurate sensor is essential in this complicated process. In the cognitive map proposed by Kuipers et al. [2], the semantic structure of space is inferred using a model based on a

M. Köppen et al. (Eds.): ICONIP 2008, Part I, LNCS 5506, pp. 747–754, 2009.

hierarchy of successive environmental information elements obtained while the robot moves around. A global map is derived by integrating global topology information with local metric data.

Many robot localization methods have been developed over the last decade. These can be classified depending on the type of maps they use: grid-based maps [3], feature-based maps [4], topological maps [5], semantic maps [2], and adaptive selection from multiple types of maps [6]. However, these methods of localization and map representation may not be directly applicable to semantic localization.

Humans find locations easily using cognitive information without any metric data. Cognitive representation is also necessary for service robots that interact with humans to take orders and complete tasks. Until now, research has concentrated on robot-centered knowledge that enables a human to interact with robots [7] [8].

Most research works centered on high-level knowledge have generally required identifying objects in a camera image and estimating the distance from the camera to objects to establish relationships among objects[7]. However, these estimates of distance may be inaccurate due to lens distortion and incorrect feature detection and matching, resulting in errors when used in real environments.

To cope with those issues, we suggest a cognitive representation and a Bayesian model for spatial relationships between objects. We show that our representation can be useful in cognitive localization by a mobile robot. The spatial context used in the proposed cognitive representation includes observed objects, a distance context that represents the distance from the robot to a certain object, and a spatial context that describes the relationship among objects. We propose a probabilistic Bayesian model by which localization accuracy can be improved as more elements of the spatial context are accumulated. Finally, we verify the practicality of the proposed methods through a localization experiment in an indoor environment.

2 Spatial Object Relationship

2.1 Object Recognition and Sensing of Object-Related Metrics

Object recognition is a fundamental factor in cognitive representation. In general, an object can be recognized visually by measuring the similarity between its features and those of the corresponding object model. In this section, we use scale-invariant feature transform features that are known to be invariant to image scale and rotation [9][10].

The metric distance from the robot to an observed object is estimated with a single camera to derive a piece of spatial context. After an object is recognized, its height in an image space is measured using a set of corresponding features. Then, the metric distance is estimated using

$$r_e = \left(\frac{1}{n_r}\right) \sum_{i=1}^{n_r} [\left(\frac{h'}{h_i}\right) \cdot c_{r_i}] \tag{1}$$

where n_r is the number of spatial contexts for the distance, h' is the height estimated from the corresponding features, and h_i and c_{r_i} are the height and distance of the spatial context, respectively, which are pre-computed and saved in the object model. Here, i indicates the corresponding context index. The variables h' and h_i are quantities represented by the number of pixels.

Figure 1 illustrates the ratio relationship between a recognized object and its corresponding object model. Since it can be exactly calculated only for an object directly in front of the camera, we assume that the metric distance is given as a Gaussian random variable corresponding to the imperfect alignment of object and camera. This random variable is normally distributed with mean r^e and variance $(\frac{r^e}{c^{qr}})$.

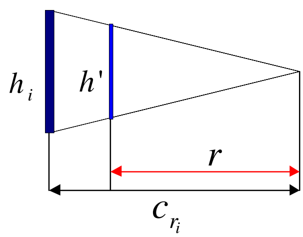


Fig. 1. Relationship of the ratio between the recognized object and corresponding object model

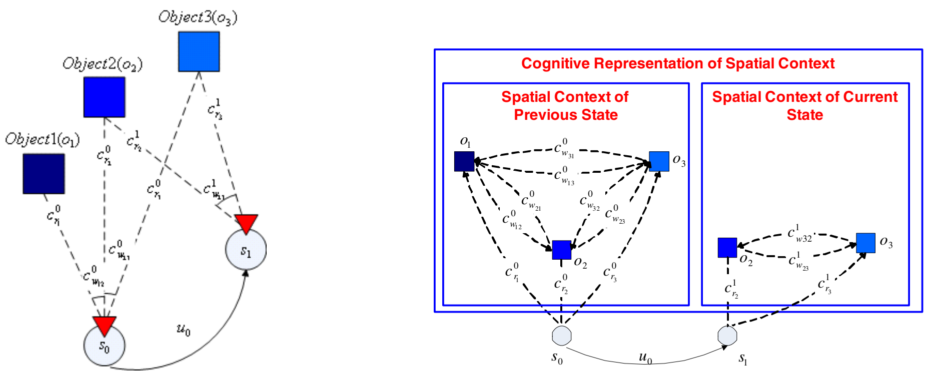
The metric bearing between objects is estimated using the horizontal position of each object and the optical angle of the camera. Previous work [11] has shown that this is simple but inaccurate. To account for this uncertainty, we handle the bearing with a probabilistic model as we did for distance. It is also a Gaussian random variable with mean ω^e and variance $(\frac{\omega^e}{c^{q\omega}})$.

2.2 Cognitive Representation

Figure 2(a) shows a change of metric relations between the robot and observed objects according to the robot's displacement from previous state (location) s_0 to a state in the real world s_1. The r and ω denote the metric distance of object relative to the robot and the bearing between objects in the robot coordinates, respectively. The subscripts indicate the indexes of observed objects. In general, the metric data quantities are inaccurate. Thus, for cognitive representation, the metric data are linked to the symbol of the spatial relationship appropriate to the given condition.

Figure 2(b) illustrates the cognitive representation consisting of observed objects and their respective spatial symbols. Here, the spatial context includes the distance context and the bearing context. The distance context denoted by c_r is a distance level of object from the robot. Each distance context is represented with one of a set of distance symbols, i.e., $c_r = \{nearby, near, far\}$. The bearing context denoted by c_ω is a bearing of one object relative to another.

Table 1 shows the cognitive representation including all the spatial relationships for the example with symbols. The robot localization scheme we developed



(a) (b)

Fig. 2. (a) Spatial relationship between robot and objects in the real world and (b) cognitive representation comprising a set of respective spatial contexts among observed objects and the robot

Table 1. Cognitive representation including all the spatial relationships in Fig. 2

frame	cognitive representation
previous frame	distance(o_1, nearby), bearing(o_1,o_2,right near), bearing(o_1,o_3, right far) distance(o_2, near),bearing(o_2,o_1, left near),bearing(o_2,o_3, right near) distance(o_3, far),bearing(o_3,o_1, left near), bearing(o_3,o_2, left near)
current frame	distance(o_2, far),bearing(o_2, o_3, right near) distance(o_3, near),bearing(o_3, o_2, left near)

finds the position of the robot using only these kinds of cognitive representation without any metric data.

For symbolic representation, the reference distance may vary according to the performance of the camera, the efficiency of the vision algorithms, and the size of the robot. Due to such distance variations, the symbolic representation should be easily adaptable to diverse applications. That is, symbolic cognitive representation provides us with high-level knowledge to be shared among a number of robot applications [7].

3 Mobile Robot Localization

3.1 Probabilistic Localization Model

We use a Bayesian approach, which partially brings out the probabilistic Bayesian model that was proposed by Ranganathan et al. [12], to estimate the robot location with a cognitive representation for spatial object contexts. The following notation is used to explain the probabilistic model. The robot location and object are denoted by $S = [x \quad y \quad \theta]^T$ and O, respectively. $C_r = \{c_{r_{1:N}}\}$ denotes the spatial context of distance, which is an $N \times 1$ matrix. $C_{\omega_{1:N}} = \{c_{r_{\omega_{1:N,1:N}}}\}$ is the spatial context of bearing, which is an $N \times N$ matrix. Thus, a set of cognitive representations for spatial object contexts containing distance and bearing is denoted by $C = \{C_r, C_{\omega_{1:N}}\}$, where N is the number of observed objects in the corresponding state. We represent the features extracted from the image with $Z = \{z_{1:n}\}$, where n is the number of features. A set of object model features is denoted by $A = \{a_{1:m}\}$, where m is the number of object features. The numbers of spatial contexts of distance and bearing are n_r and n_q, respectively.

The localization model is described using the cognitive representation. This is the principal focus of this work. Spatial contexts are uncertain data, so they should be approximated with stochastic distributions. The focuses of this section are the location model and the object contexts. We assume that sensors are uncertain, so less context results in a wider distribution. The location posterior is calculated using the cognitive representation described in section 2.

In our research, inaccurate metric data caused a large error in estimating the robot location. We model the localization with a probabilistic approach to overcome this problem.

To estimate the robot location, the localization denoted by $p(S|O, C, Z, A)$ can be factored as

$$p(S|O, C, Z, A) \propto p(O|S, Z, A)p(C|S, O, Z, A)p(S) \tag{2}$$

where $p(O|S, Z, A)$ is the object likelihood that there is a similarity between observed objects in the current state and the ones in the previous state. This is formulated as

$$p(O|S, Z, A) = exp(-\|O - O_S\|^2) \tag{3}$$

where O and O_S represent observed object in the current state and in the previous state, respectively. The $p(C|S, O, Z, A)$ is the spatial object context likelihood. The $p(s)$ is a prior location that is initialized with a uniform distribution at initial states, and then estimated with the location distribution of the previous state.

The spatial object context likelihood is computed as

$$p(C|S, O, Z, A) = \prod_{j=0}^{N}\prod_{i=0}^{N}\left[exp\left(\frac{-(c_{r_i} - r_i^e)^2}{2\left(\frac{r_i^e}{c_{qr}}\right)^2}\right) \cdot exp\left(\frac{-(c_{\omega_{ij}} - \omega_{ij}^e)^2}{2\left(\frac{\omega_{ij}^e}{c_{q\omega}}\right)^2}\right)\right] \tag{4}$$

where c_{qr} and $c_{q\omega}$are the range of the spatial context of distance and bearing, respectively. The denominator terms $\frac{r_i^e}{c_{qr}}$ and $\frac{\omega_{ij}^e}{c_{q\omega}}$ in Eq. (4) are variances used to reflect uncertainty, where r_i^e and ω_{ij}^e are the estimated metric distance and bearing. The further object is from the robot or other objects, the more inaccurate the metric distance and bearing will be. Dividing the spatial object context more finely will improve the localization performance.

3.2 Recursive Bayesian Model

The localization posterior can be calculated with Eq. (2). This is inaccurate because it uses the spatial context containing a high uncertainty. To improve that, we modified the probabilistic Bayesian model to a recursive Bayesian model. However, the modification cannot be applied directly to our Bayesian model because Eq. (4) is not a predictable term. To address this, we modified the location model in Eq. (2) to

$$p(s_k|O_k, C_k, Z_k, A_k) \propto p(O_k|s_k, Z_k, A_k)p(C_k|s_k, O_k, Z_k, A_k)\sum_i \phi^i_{k-1}p(s_k|s^i_{k-1}) \tag{5}$$

where the last term on the right side is an update term. $p(s_k|s_{k-1})$ and ϕ_{k-1} are the state transition and weighted particles, respectively. Spatial objects contexts only vary when motion occurs. Thus, in our research, the state transition takes place in increments based on index k.

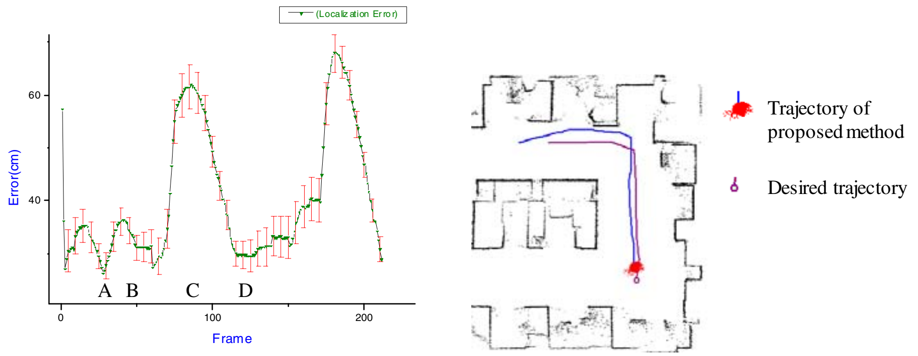
Finally, the localization posterior was estimated with the multiplication recursive Bayesian model and object context in Eq. (5). In this work, we used a particle filter that is a kind of recursive Bayesian estimation [13][14] to manage the complicated computation.

4 Experimental Results

To evaluate the performance of the proposed localization process, we used a Pioneer 3 AT robot carrying a single consumer-grade camera in a 7×6 m indoor environment. The camera captured 252 images as the robot traveled about the test area. Some distinctive objects such as a toy box, a table, a monitor, a drawer and a toy robot were selected for object recognition.

The experimental localization errors are shown in Fig. 3(a). After a quantitative analysis of the results, we learned even though the robot location was initially estimated accurately, errors can increase due to the lack of context obtained during subsequent movement. If the robot observes objects that are far away, the localization error and deviation are relatively high; however, the distribution of robot locations shrinks during subsequent movements as observed objects become closer. If a previously observed object disappears due to robot rotation, the robot deviation increases. In this experiment, minimum and maximum errors were 26.37 and 68.04 centimeters, respectively. The minimum and maximum deviation values were 2.26 and 4.21, respectively.

In Frame A of Fig 3(a), the distribution of robot locations decreased because observed objects were close to the robot. In Frame B, object disappeared due to robot rotation; this caused the distribution to increase. In the case of Frame C, observed objects were far away from the robot. This caused high uncertainty in the spatial context and thus the error of the robot location as well as the distribution increased. However, as shown in Frame D, the error of the robot location and its corresponding distribution decreased again as the robot moved closer to objects.



(a) (b)

Fig. 3. (a) Localization errors when estimated by the spatial object contexts. (b) Localization results show the trajectory of the robot estimated by the proposed spatial object contexts.

In Fig. 3(b), the circle with line (violet line) indicates the robot trajectory measured using a precise laser sensor. The particles (red dots) represent the posterior distribution of the robot localization.

5 Conclusions

In this paper, we propose that a cognitive representation and Bayesian model of observed objects and their spatial contexts be described by symbols. Our proposed method enabled robots to be localized using spatial object contexts and their probabilistic models. Experimental results from tests of the proposed cognitive robot localization method in an indoor environment can be concluded that as the number of contextual clues is increased, the location accuracy is improved in spite of using inaccurate sensors such as a consumer-grade camera.

References

1. Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G., Burgard, W., Kavrakij, L.E., Thrun, S.: Principles of Robot Motion - Theory, Algorithms, and Implementations. MIT-Press, Cambridge (2005)
2. Kuipers, B.: The Spatial Semantic Hierarchy. Artificial Intelligence, 191–233 (2000)
3. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. IEEE Computer 22, 44–57 (1989)
4. Miro, V., Zhou, J., Dissanayake, W.: Towards vision based navigation in large indoor environments. In: Proceeding of IEEE the IROS, pp. 2096–2102 (2006)
5. Beeson, P., Kuipers, B.: Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph. In: Proceeding of IEEE ICRA, pp. 4384–4390 (2005)
6. Tully, S., Moon, H., Morales, D., Kantor, G., Choset, H.: Hybrid Localization using the Hierarchical Atlas. In: Proceeding of IEEE the IROS, pp. 2857–2864 (2007)

7. Suh, I.H., Lim, G.H., Hwang, W., Suh, H., Choi, J.H., Park, Y.T.: Ontology-based Multi-layered Robot Knowledge Framework (OMRKF) for Robot Intelligence. In: Proceeding of the IEEE the IROS (2007)
8. Lim, G.H., Suh, I.H.: Weighted Action-coupled Semantic Network (wASN) for Robot Intelligence. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2008)
9. Lowe, D.G.: Distinctive image features from scale invariant keypoints. Int'l. Journal of Computer Vision 60(2), 91–110 (2004)
10. Ke, Y., Sukthankar, R.: PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In: Proceedings of Computer Vision and Pattern Recognition (2004)
11. Yim, B.D., Hwang, S.Y., Song, J.B.: Mobile Robot Localization based on Fusion of Vision and Range Information. In: CASS 2006 (2006)
12. Ranganathan, A., Dellaert, F.: Semantic Modeling of Places using Object. In: Robotics: Science and Systems (RSS) (2007)
13. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2005)
14. Dellaert, F., Fox, D., Burgard, W., Thrun, S.: Monte Carlo Localization for Mobile Robot. In: Proceeding of the ICRA (1999)