Qualitative Map Learning Based on Co-visibility of Objects

Takehisa Yairi and Koichi Hori
University of Tokyo
4-6-1 Komaba Meguro-ku Tokyo, Japan
yairi@ai.rcast.u-tokyo.ac.jp

Abstract

This paper proposes a unique map learning method for mobile robots based on the *co-visibility* information of objects i.e., the information on whether two objects are visible at the same time or not from the current position. This method first estimates *empirical* distances among the objects using a simple heuristics - "a pair of objects observed at the same time more frequently is likely to be located more closely together". Then it computes all the coordinates of the objects by multidimensional scaling (MDS) technique. In the latter part of this paper, it is shown that the proposed method is able to learn *qualitatively* very accurate maps though it uses only such primitive information, and that it is robust against some kinds of object recognition errors.

1 Introduction

Map learning problem of autonomous mobile robots has been a central issue in the field of artificial intelligence as well as robotics, because it contains several important aspects of intelligence such as recognition of environment and acquisition of internal representations. In fact, a variety of map building methods have been developed for a wide range of robots, tasks and environments so far. These methods are traditionally classified [Kortenkamp *et a/.*, 1998; Murphy, 2000] in terms of the way of map *representation* into *metric* map building methods[Moravec and Elfes, 1985; Uhlmann *et al,* 1997], *topological* methods[Mataric, 1992; Zimmer, 1996; Shatkay and Kaelbling, 1997], and hybrids of them[Thrun et al., 1998].

On the other hand, when we turn our attention to a new trend of robot navigation called *qualitative navigation[Levitt* and Lawton, 1990; Schlieder, 1993], we can see there is another way of qualitative map representation that is different from both the metric and topological representations. A most important point of the maps used in this qualitative navigation is that they are not required to be accurate in a *metric* sense as long as they correctly preserve the qualitative spatial relationships (such as circular ordering) of the objects or landmarks in the actual environment.

A challenging problem in this paradigm is to construct autonomously such qualitative maps from *qualitative* observation information robots obtain. A representative approach to this qualitative map learning problem is [Sogo *et al,* 2001], in which qualitative information of "how object positions are classified to two sets with respect to arbitrary straight lines" is used to construct a map by propagating "three point" constraints.

In this paper, we also propose a map learning method for mobile robots based on qualitative observation information. It uses the information of "co-visibility" or whether two objects are visible or not at the same time from robot's positions, which is much more primitive than the information used in [Sogo *et al,* 2001] and the ordering information[Schlieder, 1993].

In this method, co-visibility of two objects is translated into an empirical distance based on a simple heuristics "closely located objects are likely to be seen simultaneously more often than distant objects and vice versa" or "temporal and spatial proximities are approximately equivalent". Then, positions of all objects are calculated by well-known *multidimensional scaling* (MDS) technique.

A noteworthy feature of this method from a practical viewpoint is that it does not require robots localize their own positions while mapping. Moreover, it is shown that it can learn qualitatively accurate maps without higher level infonmation such as ordering, and is robust against observation errors. We also discuss the validity of the heuristics above from several viewpoints.

2 Problem Definition

2.1 Assumptions on Environment and Robot

we consider a map building task by a mobile robot, in which the robot estimates the positions of objects in the environment by repeated explorations and observations (Figure 1). More specifically, we make the following assumptions about the environment and robot.

Environment

The environment is a closed area containing a finite number of objects. Each object is assigned a unique ID. In addition, it is assumed that all objects are about the same size.

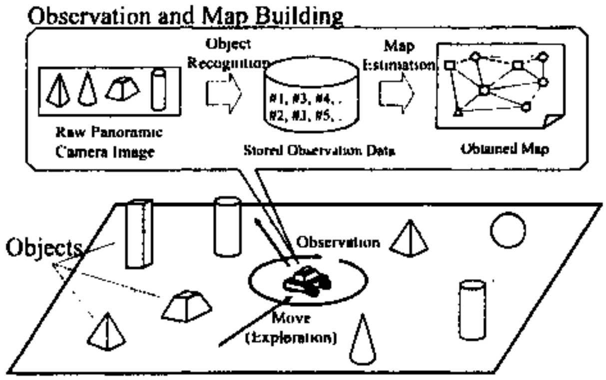

Figure 1: Assumed map building task of a mobile robot (exploration, observation and map estimation)

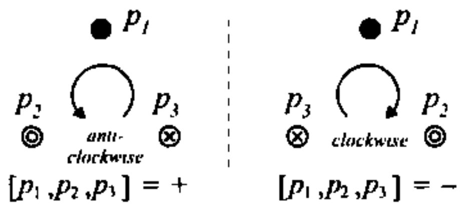

Figure 2: Definition of triangle orientation

Observation

The robot obtains a panoramic camera image at each observation point, and extracts a list of *recognized* or *visible* objects by some image processing and recognition technique. These lists of visible objects are accumulated over time and used to build a map later.

Exploration

The robot explores the environment with some randomness, avoiding collisions with objects.

2.2 Evaluation Criterion for Qualitative Maps

Though a constructed map is represented in the form of numeric coordinates of the objects on 2-D plane as ordinary metric maps, its goodness is measured by the correctness of the *qualitative* spatial relationships rather than by the *metric* accuracy.

To evaluate the qualitative correctness of the obtained maps, we employ the notion of *triangle orientation[Schlieder,* 1993; Sogo *et al.,* 2001] or counter-clockwise order of three points. In short, triangle orientation of three points $[p_1 p_2 p_3]$ in 2-D plane is defined as + when the order p_1, p_2, p_3 yields a counter-clockwise turn, and — otherwise (Figure 2).

When there are *N* objects in the environment, the number of all possible triangles formed by them becomes ${}_N C_3$. So we define the *orientation error* Err_{ori} of a constructed map as the percentage of triangles with wrong orientations, compared with the *real* configuration of the objects, i.e.,

$$Err_{ori} \equiv \frac{\text{(Number of triangles with wrong orientation)}}{{}_N C_3}$$

We use this Err_{ori} as an evaluation criterion of constructed maps in the later simulation study.

2.3 Co-visibility and Distance Between Objects

As previously mentioned, we use a heuristics "a pair of objects observed simultaneously more frequently is likely to be located more closely together" to estimate the distance between them. Though it is hard to prove the validity of this heuristics strictly in general cases, we consider it is approximately appropriate for the following reason.

First, we assume that an object becomes difficult to identify as the distance from the robot increases, because the image size of the object becomes smaller and the chance of occlusion increases.

Given this assumption, the total area size of the region where the robot can observe two objects simultaneously decreases monotonically according to the distance between the objects. For example, consider the case the probability that an object can be observed from the robot at a distance of *r* is $P_1(r) = e^{-a \cdot r^2}$ (Figure 3). If two object A and B are located at $(-\frac{l}{2}, 0)$ and $(\frac{l}{2}, 0)$ respectively, the probability that both objects are visible from an arbitrary robot position X (x,y) becomes

$$P_2(l, x, y) = P_1(\overline{AX}) \cdot P_1(\overline{BX}) = e^{-\frac{al^2}{2}} \cdot e^{-2a(x^2+y^2)}$$

Then, the'expectcd size of area where both objects are visible simultaneously is

$$S_{covis}(l) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_2(l, x, y) dx dy = \frac{\pi}{2a} e^{-\frac{al^2}{2}}$$

In the simplest case that the robot is located at any place in the environment with equal probability, the probability that the two objects are co-visible is expected be proportional to $S_{covis}(l)$, which is a monotonic decreasing function of l.

In section 4, we will discuss this "co-visibility and distance" heuristics in a more general form: "temporal and spatial proximities are approximately equivalent".

3 Proposed Method

3.1 Outline of *CoviMap*

Based on the assumptions above, we propose *CoviMap* - a map learning method based on the co-visibility of objects. In this method, the mobile robot obtains approximate positions of objects based on the co-visibility frequencies of them, which are updated repeatedly by explorations and observations. The outline of *CoviMap* is described as below:

1. The robot repeats the following behavior steps and updates the number of observations of each object (n_i), and the number of simultaneous observations of every pair of objects ($n_{i,j}$).
   (a) The robot moves to the next observation position, avoiding collisions with the objects.

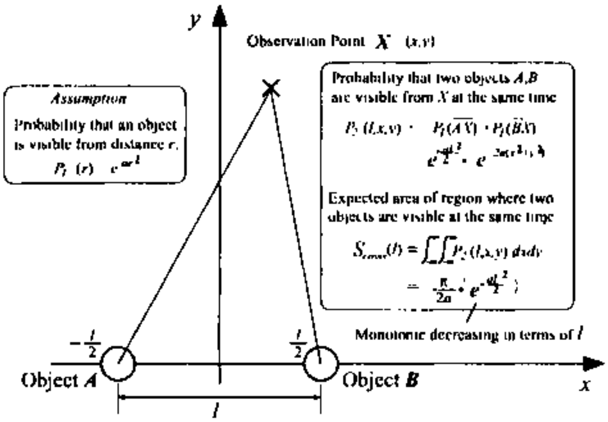

Figure 3: Relationship between co-visibility and distance of two objects in a simplified environment

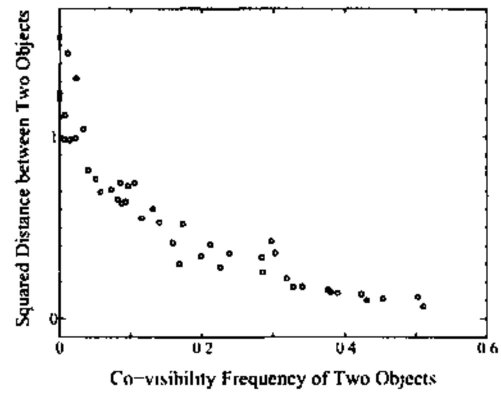

Figure 4: Relationship between squared distance $d_{i,j}^2$ and co-visibility frequency f_{ij} of objects in the simulation environment

(b) It obtains a list of *visible* objects L_o from the panoramic image captured at the current position, and updates n_i and n_{ij} as bellow:

$$n_i \leftarrow n_i + 1 \quad \text{for each object } i \text{ in } \mathcal{L}_o$$
$$n_{i,j} \leftarrow n_{i,j} + 1 \quad \text{for each pair of } i, j \text{ in } \mathcal{L}_o$$

2. After a specified number of steps, *co-visibility frequencies* $f_{i,j}$ is computed for each pair of objects based on n_i, n_j, and $n_{i,j}$. Then *empirical distance* $d_{i,j}'^2$ of each pair is computed from $f_{i,j}$.
3. The robot obtains the estimated positions of all objects $(\hat{x}_1, \cdots, \hat{x}_N)$ by applying *Multi-Dimensional Scaling* (MDS) to the distance matrix *D* whose (i, *j)* element is $d_{i,j}'^2$.

In the remaining of this section, we explain the second and third parts in the above procedure.

3.2 Computation of Co-visibility Frequency and Empirical Distance

Co-visibility frequency $f_{i,j}$ between two objects is defined as follows:

$$f_{i,j} = \frac{n_{i,j}}{n_i + n_j - n_{i,j}}$$

$f_{i,j}$ stands for the conditional probability that two objects are visible at the same time, given that at least one of them is visible. It takes a value between 0 and 1. This definition of $f_{i,j}$ is also known as Jaccard's coefficient.

With the definition of co-visibility frequency $f_{i,j}$, the heuristics introduced in 2.3 can be rewritten as "distance between two objects $d_{i,j}$ *monotonically decreases* as $f_{i,j}$ increases". Figure 4 (scattered points) illustrates the actual relationship between the *real* (squared) distance and co-visibility frequency $f_{i,j}$ in the simulation environment in section 4. The result indicates that the heuristics is approximately valid.

Therefore, we introduce a notion of *empirical (squared) distance* $d_{i,j}'^2$ between an arbitrary pair of objects, which is defined by some monotonic decreasing non-negative function ϕ, i.e.,

$$d_{i,j}'^2 = \phi(f_{i,j})$$

We also define an *empirical distance matrix D* whose (i, j) element equals to $d_{i,j}'^2$. *D* is a symmetric matrix (i.e., $d_{i,j}'^2 = d_{j,i}'^2$), and its diagonal components ($d_{i,i}'^2$) equal to zero.

A possible solution to the problem of deciding a suitable combination of function and parameter values for ϕ is to select the one which minimizes the *stress* value of the MDS results described later. Furthermore, if no appropriate empirical distance functions are provided, non-metric MDS (or ordinal MDS) can be used instead.

Another problem is that $d_{i,j}'^2$ cannot be computed properly for the pairs whose co-visibility frequencies are zero, because ϕ becomes some constant value. In this case, *CoviMap* utilizes the *triangular inequality constraint* (TIC) to correct the value of $d_{i,j}'^2$ as follows:

$$d'_{i,j} \leftarrow \min_k (d'_{i,k} + d'_{k,j})$$

3.3 Map Construction Based on Multi-Dimensional Scaling (MDS)

Multi-dimensional scaling (MDS)[Young and Householder, 1938] is a multivariate data analysis technique used to visualize a potentially high-dimensional data structure by mapping it into a relatively low dimensional space[Cox and Cox, 2001]. The purpose of MDS is to find an optimal configuration of objects in a low-dimensional space, when dissimilarities or distances among them are given.

While there are several kinds of MDS methods[Cox and Cox, 2001], *CoviMap* employs *classical scaling*[Young and Householder, 1938] method that is a kind *of metric MDS* to reproduce a map of objects in 2-D plane from the set of empirical distances. In brief, it obtains a set of 2-D coordinates of *N* objects x_i (i = 1, • • ■, *N)* by applying *Young-Householder transformation* and *spectral decomposition* to the *empirical distance matrix D.*

A non-trivial problem of using MDS in CoviMap is that there is a fifty percent chance of obtaining the *mirror image* of the expected map. However, if any correct *triangle orientation* of three objects is given, *CoviMap* can detect the mirror map and get the right map by turning it over.

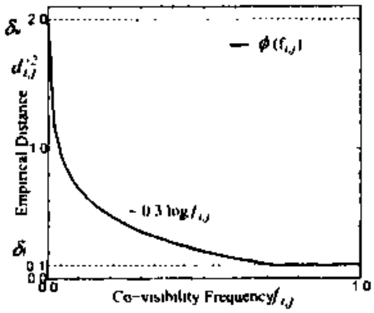

Figure 5: Empirical distance functions $\phi(f_{i,j})$ used in the simulation

4 A Simulation Study

We conducted a simulation study to examine the effectiveness of our method with Cyberbotics[1] WEBOTS simulator (ver.2.0).

4.1 Settings

Environment and Objects

The environment is a square field of 1.5[m] x 1.5[mJ containing 5 - 30 objects. Each object is cylinder-shaped (height: 160[mm], diameter: 48[mm]), and is given a unique ID number like *obO,obJ,*

Observation and Object Recognition

The robot has a camera with a resolution of 160 x 120 pixels, and obtains a panoramic image by rotating and capturing several images at each observation position. We make a simplified assumption that the robot can recognize an object if its image is wider than 10 pixels.

Exploration Strategy

At each observation position, the robot chooses its next moving direction randomly within the range of $\pm\theta_r = 45[\mathrm{deg}]$, and proceeds $l_r = 100[\mathrm{mm}]$ in the direction. In addition, the robot has 8 proximity sensors to detect nearby obstacles (objects and walls) around it, and avoids collisions with them.

Empirical Distance Function

We chose a logarithmic function as the empirical distance function ϕ as below:

$$\phi = \begin{cases} -\alpha \cdot \log_e f_{i,j} & \text{if } \exp\{-\frac{\delta_u}{\alpha}\} \leq f_{i,j} \leq \exp\{-\frac{\delta_l}{\alpha}\} \\ \delta_u & \text{if } 0 \leq f_{i,j} < \exp\{-\frac{\delta_u}{\alpha}\} \\ \delta_l & \text{if } \exp\{-\frac{\delta_l}{\alpha}\} < f_{i,j} \leq 1 \end{cases}$$

where, $\alpha, \delta_l, \delta_u$ are parameters which take positive values. In this simulation, we set them as: $\alpha = 0.3, \delta_l = 0.1, \delta_u = 2.0$, because the stress values of MDS were relatively small with this combination compared with others. Figure 5 shows the shape of this empirical distance function. In addition, *triangular inequality constraint* (TIC) mentioned in 3.2 was used to correct the empirical distances for the object pairs whose co-visibility frequencies were zero.

4.2 Experiment 1 : Varying Number of Objects

First, we examined the basic map building capabilities of *CoviMap* with varying number of objects in the environment

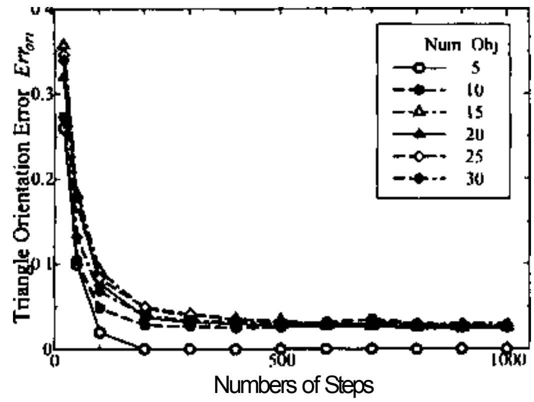

Figure 6: (Experiment 1) Change of average map errors comparison among various numbers of objects

Steps	5 obj.	10	15	20	25	30
20	26.0	27.3	35.7	32.0	34.7	34.0
100	2.00	4.83	9.28	7.77	8.37	6.84
500	0.00	2.50	3.34	2.72	3.24	3.00
1000	0.00	2.50	2.64	2.46	2.64	2.97

Table 1: Change of average map errors [%] with 6 different numbers of objects

from 5 to 30. In each case, the robot repeated the observation and exploration steps for 1000 times, and built a map eachtime the number of steps reached one of specified numbers (20,50,100,...,1000). For each case, we prepared 5 layouts of objects which arc randomly generated and averaged the orientation errors.

Figure 6 and Table 1 show how the average orientation error Err_{ori} changes according to the number of steps for each number of objects. As can be seen from this, Err_{ori} converges to a narrow range (approximately between 2.5[%] and 3.0[%]) after several hundreds of steps, when the number of objects varies from 10 to 30. This means that *CoviMap* is not affected drastically by the increase in the number of objects.

Figure 7 illustrates an example of real configuration which contains 30 objects, while Figure 8 illustrates a map constructed by *CoviMap* after 500 steps. In this case, out of total 4060 triangles, the number of triangles whose orientations are inconsistent between the real map and the constructed map is 98 $(Err_{ori} = 2.4[\%])$. Interestingly, if we consider only the triangles whose largest angles are smaller than 170[deg] in the real map (the number of such triangles is 3835), the number of inconsistent triangles falls to 27 $(Err_{ori} = 0.70[\%])$. This means our method is able to acquire very accurate qualitative layouts of the objects, unless the objects are almost on a line. Dotted lines in the figures represent *Delaunay graphs* of the configurations. A comparison of them also tells that the constructed map is *qualitatively* very accurate.

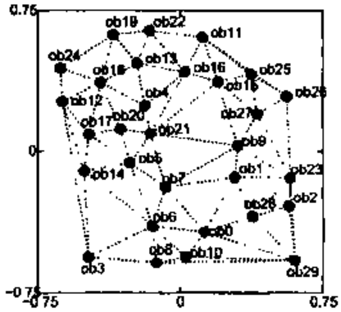

Figure 7: (Experiment 1) An example of object configuration with 30 objects, dotted lines are edges *of Delaunay graph*

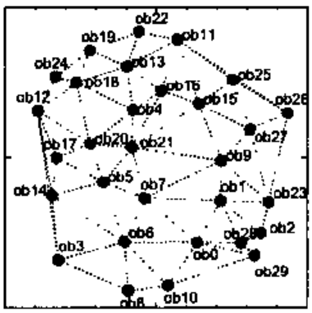

Figure 8: (Experiment 1) Constructed map after 500 steps $(Err_{ori} = 2.4\%$ for all triangles)

4.3 Experiment 2 : Robustness Against Object Recognition Errors

In the previous experiment, it was assumed that there are no errors in the object recognition process. In the real environment, however, it is almost inevitable to fail in recognizing objects occasionally, due to various uncertainties. Therefore, we examined how the recognition errors affect the quality of maps constructed by *CoviMap.*

More specifically, we considered two kinds of recognition errors - *non-recognition* and *mis-recognition.* These errors were artificially generated as follows:

Non-recognition : Randomly chosen elements in L_o are removed

Mis-recognition : Randomly chosen elements in L_0 are replaced with other object IDs.

In this experiment, we fixed the number of objects to 15 and conducted 5 trials for each of 5 different object configurations.

Result with Non-recognition Error

Figure 9 shows the average *orientation errors* of the constructed maps when the percentage of non-recognition error is set to 0.0 (no errors), 3.0, 10, 20[%] respectively. In an early stage of learning, i.e, with fewer steps, the map error Err_{ori} becomes larger according to the magnitude of the non-recognition error. However, the difference becomes smaller and almost negligible as the number of steps increases.

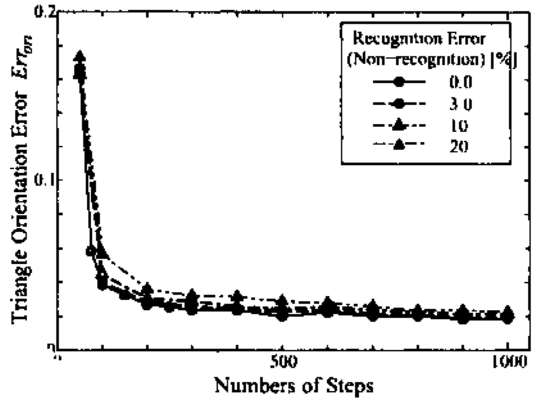

Figure 9: (Experiment 2) Change of average map errors comparison among various *non-recognition* rates

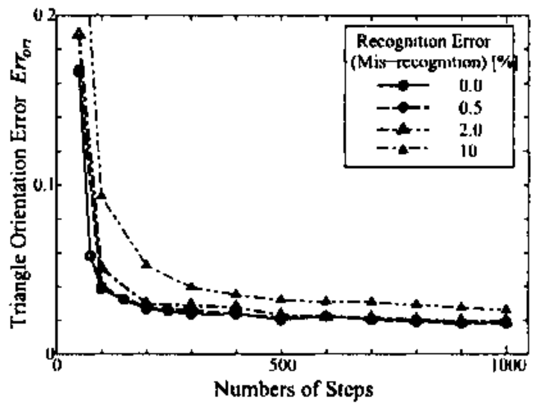

Figure 10: (Experiment 2) Change of average map errors comparison of various *mis-recognition* rates

Result with Mis-recognition Error

Next, Figure 10 shows the results when mis-recognition error is set to 0.0, 0.5, 2, 10[%] respectively. Not surprisingly, the influence of this kind of error on the map error Err_{ori} is relatively larger than the non-recognition error.

In both cases, however, we can conclude *CoviMap* is robust against those recognition errors, in that the map error steadily decreases according to the increase of the number of steps.

5 Related Works

As mentioned before, the central idea of our map building method is a simple heuristics - "a pair of objects observed simultaneously more frequently is likely to be located more closely", which is used to estimate the distances and qualitative configuration of objects from the co-visibility information of them. In this section, we discuss the meaning of this heuristics from other viewpoints, especially, considering several related works in other fields.

First, as *co-visibility* of objects can be regarded as *co-occurrence* of events that objects are visible, a lot of studies on learning of behavior and models based on the *co-occurrence* information have been made. The most basic principle related to this issue is *Hebb's rule* which postulates that if two events co-occur often, the connection of them should be strengthened. Many recently developed *reinforcement learning* methods such as Q-learning[Watkins and Dayan, 1992] are also considered to be based on the *co-occurrence* or *temporal proximity* between behavior and reward events. Interestingly, there is another similarity between reinforcement learning and our *CoviMap.* That is to say, the former learns a complicated behavior system from discrete events *of rewards,* while the latter learns a qualitatively accurate map from discrete events *of co-visibility* among objects.

The heuristics above can also be regarded as a special case of more general one - "temporal and spatial proximities are approximately equivalent in many environments". In the research of human spatial memory, it has been reported that humans often rely on the *temporal* proximity rather than the *spatial* proximity to memorize spatial structures of environments[Curiel and Radvansky, 1998]. This might suggest that the principles and results of *CoviMap* presented in this paper are helpful to understand some aspects of human spatial memory and cognitive mapping process, although we doubt that an MDS-like algorithm exists in human brain.

6 Conclusion

In this paper, we proposed a unique map building method for mobile robots named *CoviMap* which is based on the *co-visibility* information of objects in the environment. It can learn *qualitatively* accurate maps from such primitive information using a simple heuristics "temporal and spatial proximities are approximately equivalent" and multidimensional scaling (MDS) technique. It was also shown by the simulation results that *CoviMap* is applicable to large environments containing dozens of objects, and robust against observation errors such as non-recognition and mis-recognition.

A most important theoretical issue to be studied is to validate and generalize the heuristics on the *equivalence of temporal and spatial proximities.* Especially, we need to investigate how our method will be affected if the camera image is not panoramic but more restricted, or the object recognition rate is not isotropic but dependent on the direction. On the other hand, future works from a practical point of view will include experiment with real robots, integration of *CoviMap* with qualitative navigation[Levitt and Lawton, 1990], and extension or replaccment [Faloutsos and Lin, 1995] of MDS.

References

[Cox and Cox, 2001] T. Cox and M. Cox. *Multidimensional Scaling.* Chapman & Hall/Crc, 2001.

[Curiel and Radvansky, 1998] J. M. Curiel and G. A. Radvansky. Mental organization of maps. *Journal of Experimental Psychology,* 24(1):202-214, 1998.

[Faloutsos and Lin, 1995] C. Faloutsos and K. Lin. FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data,* pages 163-174, 1995.

[Kortenkamp et al., 1998] D. Kortenkamp, R. Bonasso, and R. Murphy, editors. *Artificial Intelligence and Mobile Robots : Case Studies of Successful Robot Systems.* MIT Press, 1998.

[Levitt and Lawton, 1990] T. Levitt and D. Lawton. Qualitative navigation for mobile robots. *Artificial Intelligence,* 44(3):305,361 1990.

[Mataric, 1992] M. Mataric. Integration of representation into goal-driven behavior-based robots. *IEEE Transactions on Robotics and Automation,* 8(3):304-312, 1992.

[Moravee and Elfes, 1985] P. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proceedings of the IEEE International Conference on Robotics and Automation,* pages 116-121, 1985.

[Murphy, 2000] R. Murphy. *Introduction To AI Robotics.* MIT Press, 2000.

[Schlieder, 1993] C. Schlieder. Representing visible locations for qualitative navigation. In N. P. Carret'e and M. G. Singh, editors, *Qualitative Reasoning and Decision Technologies,* pages 523-532. 1993.

[Shatkay and Kaelbling, 1997] H. Shatkay and L. Kaelbling. Learning topological maps with weak local odometric information. In *Proc. of IJCAI-97,* pages 920-927, 1997.

[Sogo *et al,* 2001] T. Sogo, H. Ishiguro, and T. Ishida. Acquisition and propagation of spatial constraints based on qualitative information. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23(3):268-278, 2001.

[Thrun et al, 1998] S. Thrun, J. Gutmann, D. Fox, W. Burgard, and B. Kuipers. Integrating topological and metric maps for mobile robot navigation: A statistical approach. In *Proc. of AAAI-98,* pages 989-995, 1998.

[Uhlmannet al. , 1997] J. K. Uhlmann, S. J. Julier, and M. Csorba. Nondivergent simultaneous map-building and localization using covariance intersection. In *Navigation and Control Technologies for Unmanned Svstems II,* pages 2-11.1997.

[Watkins and Dayan, 1992] C. Watkins and P. Dayan. Tchnical note: Q-learning. *Machine Learning,* 8(3-4):279-292, 1992.

[Young and Householder, 1938] G. Young and A. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika,* 3:19-22, 1938.

[Zimmer, 1996] U. R. Zimmer. Robust world-modelling and navigation in a real world. *Neurocomputing,* 13:247-260, 1996.