

# Making Cognitive Map of Outdoor Environment

Saburo Tsuji and Shigang Li  
Department of Systems Engineering  
Osaka University  
Toyonaka, Osaka 560, Japan  
tsuji@ce.osaka-u.ac.jp

## Abstract

Exploring an unknown environment and making its model is a key issue for realizing an intelligent robot. This paper presents a framework for representing an outdoor environment along a route as a Cognitive Map. Continuous views by a moving robot are transformed into a path-centered 2(1/2)D scene representation, called *Panoramic Representation*, which the cognitive mapping process takes as input. Range data-based analysis of the *Panoramic Representation*, segments the viewed scene into ground and each object standing upon it. The objects are then arranged into a structure map which provides robust qualitative information on the environment structure. Although range data in the 2(1/2)D representation are non-precise and vary with the camera speed and orientation, the spatial relations and relative sizes of the objects are invariant. The Cognitive Map is described qualitatively in terms of *places* along the route and landmarks found as distinctive objects in each *place*. When the robot moves along the same route, the Cognitive Map guides the navigation; the robot can reliably match the memorized landmarks against objects appearing in its sight.

## 1 Introduction

Exploring an unknown environment and making its model is a key issue for realizing an intelligent robot. A variety of world representations have been proposed since the beginning of research on the intelligent robots. For example, grid or cellular models [Nillson, 69; Kamahampati and Davis, 86; Elefes 87] and a polygonal model [Giralt *et al.*, 84] are used for representing local areas, such as rooms and corridors, in indoor environments and their spatial relations are then represented by a graph structure [Giralt *et al.*, 84] for planning navigation from one room to another.

<sup>1</sup>This work is supported in part by Grant-In-Aid, Ministry of Education, Science and Culture, Japanese Government

From psychological studies, it is widely believed that humans and animals use cognitive maps, acquired while wandering in unknown environments, for inferring spatial relations and determining their actions. A computational theory of cognitive maps has been developed which can explain some of the current findings about cognitive maps in the psychological papers [Yeap 88]. The proposed idea is that the cognitive mapping process should be studied two loosely coupled modules, Raw Cognitive Map and Full Cognitive Map. Under assumption of availability of noise-free precise 2(1/2)D vision data, the Raw Cognitive Map is computed directly from the visual observations, not from high level descriptions what we perceive. The second module, known as the Full Cognitive Map, takes the Raw Cognitive Map as input and produces abstract representations for high level spatial tasks.

The above-mentioned two approaches have studied the mapping process of the indoor environments which are considered much more structured than the outdoor environments. Although autonomous navigation of vehicles has been extensively studied in recent years, not so many reports have been published on autonomous map making of outdoor environments. Navlab group has studied algorithms for building maps of road scenes using an active range and reflectance sensor and for using the maps to transverse a portion of the world already explored [Hebert 89]. Another example is making an elevation map of an unstructured environment by fusing observations using an ERIM sensor at different locations [Kweon and Kanade, 92]. An interesting idea is qualitative navigation utilizing qualitative information of visible landmarks at different areas [Levitt and Lawton, 90]. A qualitative, image-based navigation via homing has been studied also [Hong *et al.*, 91]. These studies give us important ideas and techniques for modeling the outdoor environment, however a more general method is needed to realize a truly intelligent robot working in a real world.

We consider the mapping of outdoor environments cannot be accomplished by simply applying the ideas presented for the indoor environments. *Places*, found as closures with vertical surfaces, are basic elements representing the indoor environment and the global structure is described in a hierarchy representing paths connecting them. Since the geometrical structure of the indoor environment is clear, the corridors are represented as simple

arrows in the graph. Other features of the *places* such as their colors, textures, or objects in them are considered less important.

The outdoor world is much more wide and less structured than the indoor environment, and information on routes connecting *places* is very important for spatial reasoning in it. We claim that two different representations are needed for the outdoor world. One is for representing areas, such as a plaza or a field, as *places*, and the other is for routes connecting them. This paper focuses upon the latter issue and shows both the structural and iconic information is useful for representing the route in the outdoor world.

It is well known that computer analysis of outdoor scenes is very difficult and often fails in acquiring reliable scene information. The outdoor scenes contains objects with a variety of shapes and, moreover, some of them are more irregular in shape than those in the man-made indoor environment. Iconic properties such as brightness and color in the outdoor scenes vary significantly with changes in weather. Noise-free range data provide useful information to find the environment structure [Yeap 88], but we cannot expect precise 2(1/2)D vision data are available in the outdoor environment, in particular for far objects. This paper investigates ideas and methods for analyzing the structure and iconic features of a wide outdoor world from continuous observations by a moving perceiver.

## 2 Representing and Memorizing Outdoor Environment

### 2.1 Representation of Outdoor Environment

The environment is a large scale space, a space which is too large to be perceived in a view at a point and, therefore, the perceiver should move about it in order to develop a Cognitive Map. There are two issues for the map making. One is reactive planning the next route for exploring the unknown environment. The other is interpretation of the consecutive motion/observations of the perceiver into the Cognitive Map. This paper discusses mainly on the latter issue by solving the problem "Build a Cognitive Map of an outdoor environment from observations by a robot moving along a route."

Let us consider how a human represents the route he/she has moved along. The following is some examples.

- [1] I walked the street about 1 km and .....
- [2] I turned to the right at the third corner and .....
- [3] I walked along an avenue and entered in a shopping district .....
- [4] I turned to the left at the corner next to an old Gothic church .....

The first example is a geometrical representation of the route by metric. The second is another geometrical representation in terms of its topology instead of the metric. In the third example, the patterns along the route are used as descriptors and the fourth uses a landmark, an old Gothic church, to represent the route. Note

that all of these representations are symbolic. Humans, however, use also iconic representations to memorize the scenes along the route. We sometimes say "I am sure I have been here before because I memorize this scene". We cannot tell the detail of the memorized scene before viewing it again, but we can match an incoming scene against the memorized representation and find their differences. This indicates that humans use both symbolic and iconic representations of the environment.

### 2.2 Panoramic Representation

This paper presents a new representation of scenes along the route a robot has passed. Most map making systems have been built in the framework of the conventional *static* and *passive* computer vision. A perceiver observes the environment at a rather small number of different locations, makes ego-centered 2(1/2)D representations of the space and then integrate them into a set of consistent Raw Cognitive Maps [Yeap 88]. Since real range data, for example those acquired by stereo, are noisy, methods for fusing the noisy 2(1/2)D data into a more reliable spatial representation have been studied [Ayache and Faugeras, 87]. It is more natural that the perceiver continuously views the environment while moving around and interprets the visual data, by using the motion and temporal context, into a compact yet informative representation of the environment. Because of a huge amount of information content in the temporal image sequence, we propose to convert the vision-motion data into an intermediate path-centered 2(1/2)D representation, called *Panoramic Representation*, which contains almost all information in the continuous observations [Zheng and Tsuji, 90].

Suppose a camera moving along a smooth path at a constant speed with its optical axis orthogonal to the path, as shown in Figure 1. A record of images taken continuously by the camera contains highly redundant information of the environment. If we set a vertical slit with a single pixel width and arranges 1D images through the slit along the time axis, we get a wide view of the environment, called a *Panoramic View*, with a much less amount of information [Zheng and Tsuji, 90]. We can estimate the distance to each object in the scene

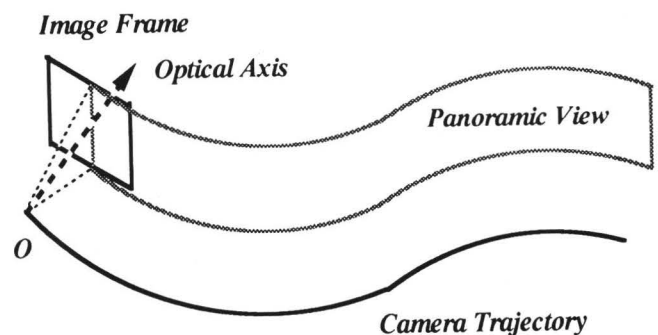


Figure 1: Formation of *Panoramic View*

from their image velocities moving across the slit. Mathematical foundation of the depth estimation from paths with different shapes is given in [Zheng and Tsuji, 92]. By integrating the *Panoramic View*, depth estimates, path geometry and camera velocity, we get *Panoramic Representation*, a 2(1/2)D representation of scenes along the route. Note that this representation is *path-centered*, while the conventional 2(1/2)D representation used for making the cognitive map is *ego-centered*. *Panoramic Representation* is qualitative in the sense that it provides non-precise information on the environment geometry, because the estimates vary with changes in camera speed and orientation, but it contains almost all information acquired by the vision sensor through the navigation. Thus, we consider it provides effective information for making Cognitive Maps. A simple and straightforward application of the *Panoramic Representation* is to memorize 2D iconic properties viewed from the route. If the robot moves along the same route again, it can match a *Panoramic View* generated from new incoming scenes against that memorized before, by a DP (Dynamic Programming) based method. Robust matching is possible in spite of significant changes in the actual path and camera velocity between the two moves [Zheng and Tsuji, 96].

### 3 Qualitative Route Representation by Cognitive Map

Cognitive science claims a key function of the human visual memory is selectivity. If a human walks along a new route, what he/she memorizes on the route is probably the most remarkable and impressive objects or scenes in certain ranges. Usually, we call them landmarks.

The method described in the preceding section memorizes all views, although data are compressed, along the route, and as a result, works only for short routes; it wastes too much memory and computing power. We, thereby, explore making of a qualitative map, which is considered as a Cognitive Map, of the route. Since humans can recognize objects selected as landmarks, a compact symbolic representation of route such as "Turn right at the corner next to an old Gothic church" is possible. Conventional computer vision systems, however, have difficulty in recognizing a wide variety of objects along the route. Thus, we study an easier approach to autonomous making of a qualitative map. At present, our system does not find identities of objects, such as a church or a bridge, but examines 3D and iconic properties in *Panoramic Representation*, finds objects which are distinctive in some attributes, and uses them as landmarks.

Figure 2 displays the concept how the Cognitive Map is yielded from the *Panoramic Representation*. The *Panoramic View* is replaced with a compact representation by the distinctive objects selected as landmarks. Qualitative motion representation in terms of turn directions and fuzzy distance descriptions, such as *very long* or *short*, is used instead of quantitative records of linear and angular velocities. The numerical map geometry is converted into a topological route representation.

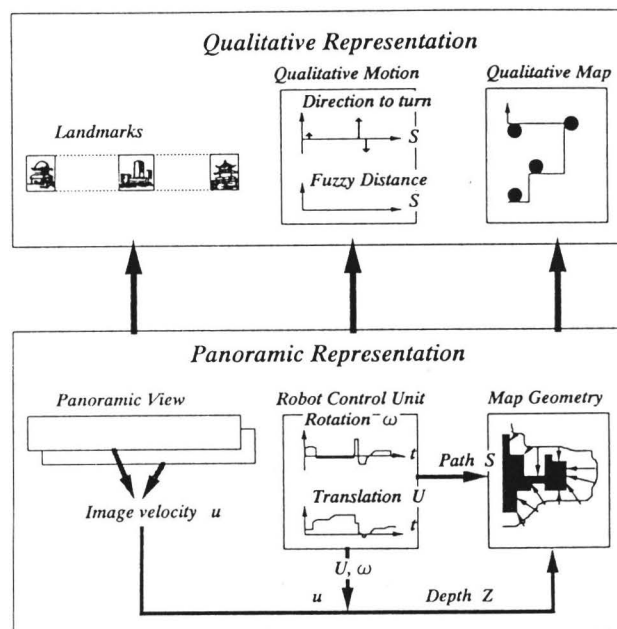


Figure 2: Cognitive Map generated from *Panoramic Representation*

#### 3.1 How Landmarks Are Found ?

What are the distinctive patterns along the route ? They depend upon the environment. Objects distinctive in some attributes, such as height, shape or color, in one area may be very popular in others. The robot, thus, should examine distribution of image/scene features along the route and select distinctive ones in each part of the route. If an object or pattern in a *Panoramic View* is distinctive in some attributes, we register it as a landmark in the area.

A method for finding distinctive iconic properties in the *Panoramic Views* has been presented [Zheng et. al, 91]. Dividing a wide *Panoramic View* into regions with almost homogeneous attributes, we map all parameters of image/scene attributes, such as color, edge density, minimum range and maximum height, of each region onto a multi-parameter space along the path. We analyze the space to evaluate which attribute or combination of attributes is remarkable and how wide it is distinctive from others, and select the landmarks. The results, however, highly rely on success in finding and/or measuring such attributes. In the outdoor scenes, color attributes such as hue and saturation significantly change with the spectral distribution of illuminant. Determining shape properties, such as areas and perimeters of texture elements, in natural scenes by computer is often unreliable. Thus, we need a more robust method for finding the landmarks.

#### 3.2 Environment-Structure-Based Cognitive Map

This paper presents a new idea of examining the spatial information in the *Panoramic Representation* as a ro-

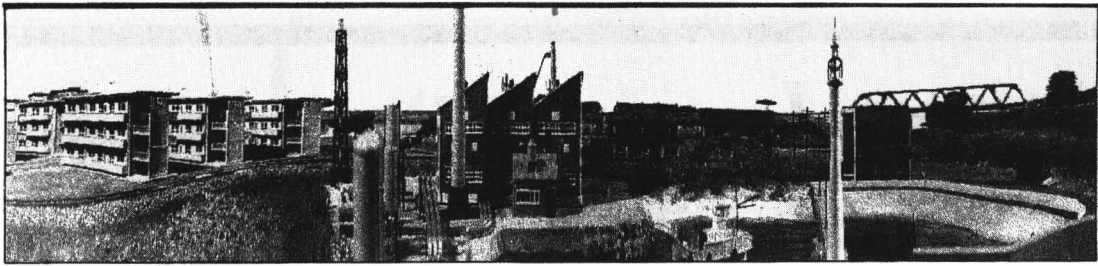


Figure 3: Panoramic view generated from an image sequence with 1016 frames

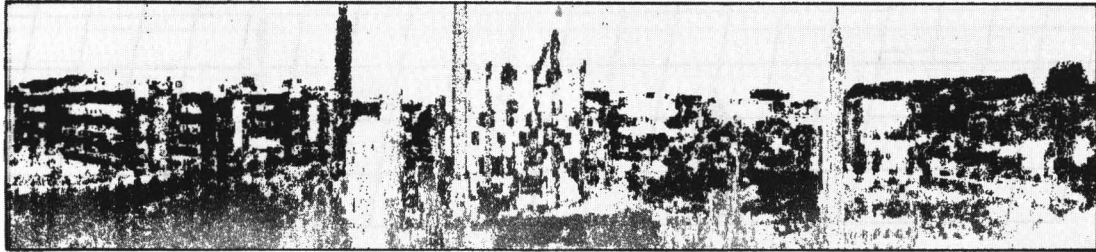


Figure 4: Disparity map

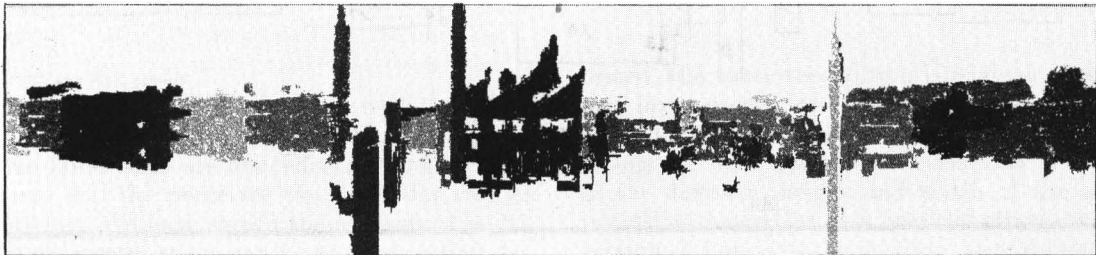


Figure 5: Objects determined by color and range

bust and qualitative cue. Although the range estimates by the moving camera are non-precise and vary with the camera velocity, the spatial and size relations between objects are invariant. The environment structure is characterized by ground and objects standing upon it. The objects are segmented from the ground by their heights, and then, they are segmented each other by their ranges. The results are arranged into a structure map that represents their spatial relationships as well as their iconic properties. By examining the structural map, we can find the environment structure and distinctive objects used as landmarks.

Our Cognitive Map is described qualitatively in terms of a set of *places* along the route and landmarks found as distinctive objects in each *place*. When the robot moves along the same route, the Cognitive Map guides the navigation; the robot can reliably match the memorized landmarks against objects appearing in its sight. The next section describes in more detail how the structure map is made and used.

## 4 Structure Map

### 4.1 Disparity Map

Ranges of feature points in the *Panoramic View* can be estimated from their image velocities [Zheng and Tsuji, 92]. We set a pair of vertical slits symmetric to the image center. Feature points appearing in the left slit are tracked in the consecutive frames until they arrive at the other slit by a simple correlation method and estimate their disparities from the mean horizontal velocities. We get a disparity map by assigning the measured disparities to the corresponding locations in the *Panoramic View*. Figure 3 is a 1016 by 240 *Panoramic View* generated from an image sequence of 1016 frames taken while the camera is moving linearly. Figure 4 shows the gray-value-encoded disparity map obtained from the image sequence; a darker point is farther from the camera path.

The first step of the structure analysis is to segment objects, both man-made and natural objects, from the ground (including lower areas like a pond). We estimate heights of all feature points in the disparity map and classify them into two categories, ground and object, by their heights. A simple one-pass cluster analysis seg-

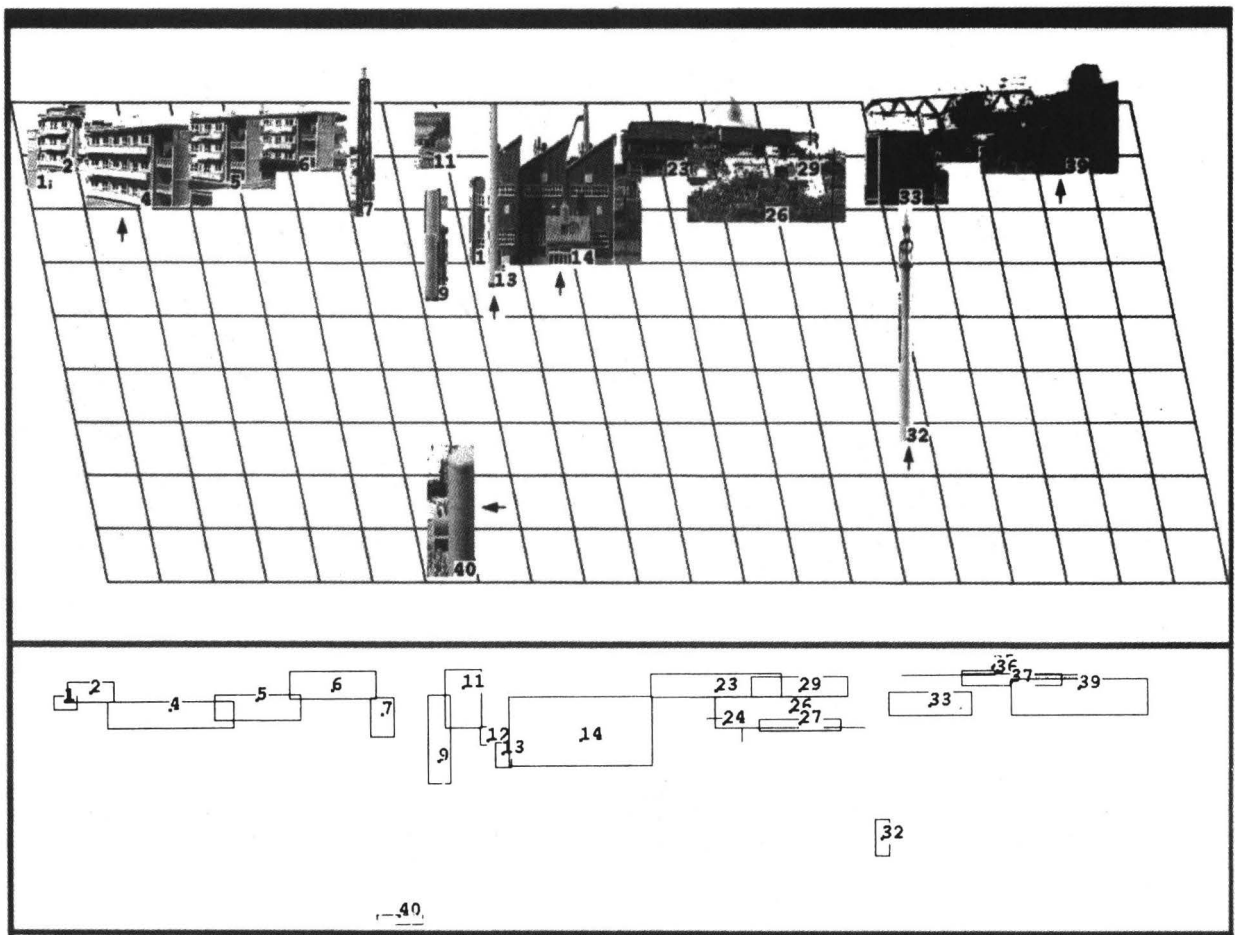


Figure 6: A perspective view of the structure map yielded from *Panoramic Viewshown* in Figure 3 (above) and the object arrangement in it (bottom). Arrows indicate objects selected as landmarks.

ments these object points into each object [Li and Tsuji, 92].

Since the disparity map contains only feature points, it is rather sparse. We determine each object in the *Panoramic View* by fusing the segmented results by color with those by range. Figure 5 shows the fusion results; different brightness indicates objects segmented as different [Li and Tsuji, 92].

## 4.2 Structure Map

Using the segmented results, we build a map representing the environment structure, called a structure map, by arranging the objects along the path which is aligned horizontal in the map. Figure 6 displays a perspective view of the structure map generated from Figure 3. Forty objects are found in the scene and each is assigned an identity number. We use the disparity as the distance from the path, rather than the estimated Euclidean distance, for setting an equal scale in estimation reliability. A top view of the structure map is shown at the bottom of Figure 6. Visible points on each object exist within the corresponding rectangle in the top view.

The structure map contains the spatial information

such as distances of objects from and their arrangements along the path. It also provides both the geometry of each object, such as its height and width, and its orthogonal view which provides color and shape. Note that although the disparities and the widths of the objects vary with the change of the perceiver's motion, their relative distances, relative sizes and order appearing along the path are invariant.

## 4.3 Analysis of Structure Map

The structure map provides useful information for making the Cognitive Map. One application is to divide the visible space into *places* and build a structured world representation, similarly as the cognitive map of the indoor environment. Although the range estimates by our method are imprecise, simple analysis of the 2D arrangement shown in Figure 7 can classify the objects into two categories, far and near. The group of the far objects are considered as the boundaries of the visible space, which correspond to the closure surfaces in the indoor environment. The space can be segmented further into three *places* by the three groups, 1-7, 9-29 and 33-39 of the boundary objects, and the two objects 32 and 40

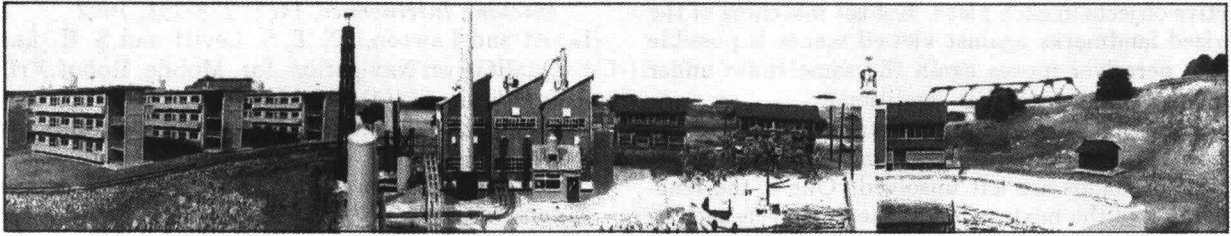


Figure 7: A *Panoramic View* acquired from observations from the same route of Figure 3 under different camera speed and illumination

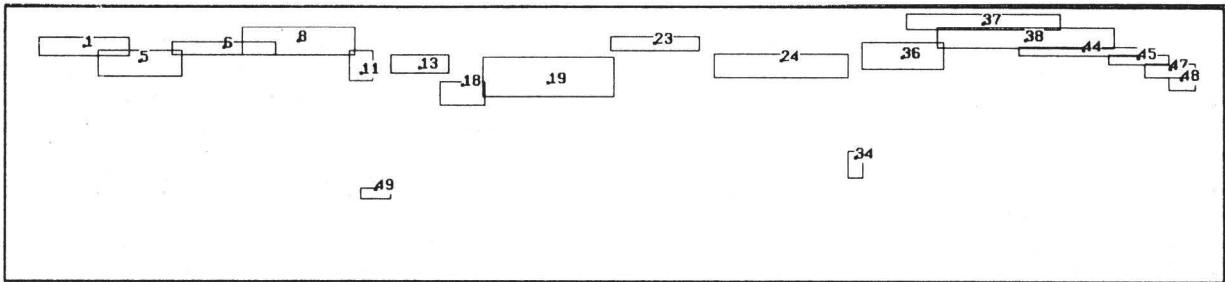


Figure 8: Object arrangement in the structure map yielded from Figure 8

standing close to the path.

As mentioned already, landmarks are useful to guide navigation; the perceiver can locate him/her in the route. We think the landmarks are also effective to give identities of *places* and the perceiver can use it for various spatial reasoning. We have shown the analysis of multi-dimensional parameter space can be applied to find distinctive patterns along the route [Zheng *et. al*, 91]. A similar method can be applied to our case, but this paper presents an alternative method.

Since the space has been segmented into *places*, the objects which segment a *place* from others are used as landmarks. These objects are suitable as landmarks, because they stand close to the path and cannot be occluded by others even if the camera path and direction change significantly. We also select one boundary object as a descriptor of each *place*. Larger and higher objects have a higher priority for the selection because of high reliability in detecting them. Sizes, shapes and iconic properties of these selected objects are compared with all other objects to examine whether they are distinctive from others. At present, we use color as the iconic property. Color similarity between two objects is estimated as the absolute difference sum of hue and saturation histograms. If there exist an object with similar properties, a candidate with a next highest priority is selected and tested. As shown arrows in Figure 6, five objects 4, 13, 32, 39 and 40 are selected as landmarks.

#### 4.4 Route Recognition by Matching Selected Landmarks

Based upon the selected landmarks, the route is recognized by the perceiver when it pursues the same route. Since the object sequence selected as landmarks is mem-

orized, the route recognition is equivalent to search the new input scenes for each landmark.

Although the perceiver moves along the same route, change of its speed causes variations in the estimates of the distance, height and width of the object. New occlusion may occur as a result of change in viewing direction. Change in the spectral distribution of the illuminant influences on the perceived color. Figure 8 shows a *Panoramic View* acquired while the robot moves along the same route as Figure 3, but with a slower speed and different illumination. As results, the widths of objects increase and color shifts toward red. The dark steel tower becomes occluded partly by the grey tank standing close to the path in Figure 7.

Figure 7 is also transformed into a new structure map of which object arrangement is displayed in Figure 8. A model-based matching method, which examines the differences of the distance from the path, shape, and color distributions, succeeds in reliable matching of the memorized landmarks to the corresponding objects in the new structure map. This time, we allow a certain amount of color shift and estimate the color similarity.

## 5 Discussions and Conclusions

We have shown a framework of making a Cognitive Map of the outdoor environment along a route. Continuous views while moving along a route yield a path-centered 2(1/2)D scene representation, which is then transformed into a structure map. The structure map provides qualitative information on the spatial arrangement of objects in the viewed space as well as their iconic properties. We segment the space into *places* and find landmarks as their descriptors. The Cognitive Map is described in terms of the qualitative description of route geometry, a

set of *places* along the route and the landmarks found as distinctive objects in each *place*. Robust matching of the memorized landmarks against viewed scenes is possible when the perceiver moves again the same route under different motion and weather conditions.

Our research is at preliminary stage of making the Cognitive Map of the outdoor environment and many important problems are left unsolved. One is the reactive planning of the next action of the perceiver, such as selection of an exploring route or an observation point. We can use the structure map for the planning. For example, the structure map of Figure 6 contains a candidate for a *place*, an area surrounded by objects 23, 24, 27 and 29, but it is too far from the current path to get its detailed information. An intelligent perceiver could plan a route to go to a better observation point and find the structure of the area.

This paper concentrates on the making the route-based Cognitive Map. Area-based Cognitive Map is also important for representing *places*. We have shown an idea of examining the environment structure and determine the *places*. However, the observations along a given route does not provide sufficient information to acquire spatial information of an wide area. A looking-around method by swiveling a camera provides a wide view of such an area [Ishiguro and Tsuji, 92], and we should use both schemes, looking around at a point and viewing while moving along a path, to scan a wide area.

## REFERENCES

- [Ayache and Faugeras, 87] N. Ayache and O. D. Faugeras. Building, registering and fusing noisy visual maps, In *Proceedings of the First International Conference on Computer Vision*, pages 73-82, 1987.
- [Elefes, 87] A. Elefes. Sonar-based real-world mapping and navigation. *IEEE Journal on Robotics and Automation*, RA-3(3):249-265, 1987.
- [Giralt et. al, 84] G. Giralt, R. Chatila and M. Vaisset. An integrated navigation and motion control system for autonomous multisensory mobile robots. M. Brady and R. Paul (eds.) *Robotics Research The First Symposium*, MIT Press, pages 191-214, 1984.
- [Hebert 89] M. Hebert. Building and navigating maps of road scenes using an active sensor. In *Proceedings of 1989 IEEE International Conference on Robotics and Automation* pages 1136-1142, 1989.
- [Hong et. al, 91] Image-based homing. In *Proceedings of 1991 IEEE International Conference on Robotics and Automation* pages 620-625, 1991.
- [Ishiguro et. al, 92] H. Ishiguro, M. Yamamoto, M and S. Tsuji. Omni-directional stereo, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):257-262, 1992.
- [Kambahampati and Davis 86] S. Kambahampati and L. S. Davis. Multiresolution path planning for mobile robots, *IEEE Journal on Robotics and Automation*, RA-2(3):135-145, 1986.
- [Kweon and Kanade, 92] I. S. Kweon and T. Kanade. High resolution terrain map from multiple sensor data, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):278-293, 1992.
- [Levitt and Lawton, 90] T. S. Levitt and S. H. Lawton. Qualitative Navigation for Mobile Robot. *Artificial Intelligence* 44(3):305-360, 1990.
- [Li and Tsuji, 92] S. Li and S. Tsuji. Selecting distinctive features for landmarks, In *Proceedings of 1992 IEEE International Conference on Robotics and Automation* pages 53-59, 1992.
- [Nilsson, 69] N. J. Nilsson. A mobile automaton. In *Proceedings of the First International Joint Conference on Artificial Intelligence*, pages 509-520, 1969.
- [Yeap 88] W. K. Yeap. Toward a computational theory of cognitive maps, *Artificial Intelligence*, 34:297-360, 1988.
- [Zheng and Tsuji, 90] J. Y. Zheng and S. Tsuji. Panoramic representation of scenes for route understanding. In *Proceedings of the Tenth International Conference on Pattern Recognition*, pages 161-167, 1990.
- [Zheng et. al, 91] J. Y. Zheng, M. Barth and S. Tsuji. Autonomous landmark selection for route recognition by a mobile robot. In *Proceedings of 1991 IEEE International Conference on Robotics and Automation* pages 2004-2009, 1991.
- [Zheng and Tsuji, 92] J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9:(1) 55-76, 1992.