Panoramic-Environmental Description as Robots' Visual Short-Term Memory

著者	近野 敦
journal or publication title	IEEE International Conference on Robotics and Automation, 1998. Proceedings
volume	1998
number	4
page range	3253-3258
year	1998
URL	http://hdl.handle.net/10097/46652

doi: 10.1109/ROBOT.1998.680940

Proceedings of the 1998 IEEE
International Conference on Robotics & Automation
Leuven, Belgium • May 1998

Panoramic-Environmental Description as Robots' Visual Short-Term Memory

Kentaro KAYAMA Koichi NAGASHIMA Atsushi KONNO Masayuki INABA
Hirochika INOUE
Dept. of Mechano-Informatics, The Univ. of Tokyo,
7-3-1 Hongo Bunkyo-ku, Tokyo, 113 Japan

Abstract

In this paper, we propose an environmental description for robots to memorizes temporality and spatiality of wide-range environment. The description is region-based and the environment is described as a mosaic of regions. The description consists of three components: Panoramic Labelled Image *that is a segmented color panoramic image,* Region Database *that has spatial and temporal information about each region, and* Regions Relation Network *that describes adjacency state between regions. The description is constantly updated in the background of other visual processing. The description enables robots to detect changes in the environment, to grasp their characteristics, and to memorize them. Moreover, we implement the environmental description on the humanoid robot* Saika *and show how it expands the robot's behavior.*

1 Introduction

When robots work in the dynamic real world such as human's living environment, grasping of wide-range circumstances is needed. If the robots worked in the human's living environment without caring about the surroundings, it would be quite dangerous.

Grasping wide-range circumstances is also helpful for robots to start tasks quickly and to carry them out smoothly. For example, when a robot looks for an object which the robot has already seen and is now not in the robot's field of vision, the memory of wide-range image saves the robot the trouble of searching for the object by moving its neck or eyes again. It is also indispensable for object recognition to consider spatial context such as adjacency and occlusion.

Moreover, it is essential for robots to take account of time-variance of environment for the following reasons:

- to read an intention of human by watching, it is necessary for the robots to pay attention to the sequences of human action,
- watching the time-variance of the state of objects will clarify the properties of objects and the relation between objects (e.g. a part of object A is hidden by object B).
- observing environmental changes caused by robot's action enable the robot to estimate whether the action has successfully ended or not.

As a method of memorizing wide-range environment, Zheng and Tsuji proposed panoramic representation [1]. In this method, the surrounding images are pasted in a cylindrical image. The image is used to localize the mobile robots themselves. Herman and Kanade incrementally composed a bird's-eye view picture by integrating the features extracted from multiple images [2]. These methods are, however, to construct static environmental description and take no account of time-variant dynamic environment. Moreover, these methods need a behavior exclusively for making environmental description and robots cannot carry out other tasks during this procedure.

In this article, we propose wide-range environmental description which makes it easy for robots to grasp spatial structure of wide-range environment and temporal changes in circumstances. The description is region-based and the environment is described as a mosaic of regions. Each region will have various spatial and temporal information. When memory is exhausted, older or less important information is deleted. The proposed environmental description corresponds to human visual short-term memory which is called working memory or visuospatial sketch pad [3] in cognitive psychology.

The proposed environmental description has the following advantages compared with the previously proposed description such as [1, 2]:

1. It is not necessary to control cameras exclusively for making this description. Hence it can be made in background of other visual processing.
2. The description translates the spatial structure and its temporal changes in wide-range environment into forms convenient for various tasks.
3. It makes it easy to do high level process such as environment understanding.

We discuss the components of the description in section 2. In section 3, the method of making the description is explained.

To demonstrate the effectiveness of description, a memory based behavior, termed the delayed response in cognitive psychology, is carried out on the humanoid *Saika* [4]. The delayed response is claimed to be closely related to short-term memory. Experimental results are presented in section 4.

2 Environmental description using *Panoramic Labelled Image*

2.1 Environmental description based on regions

The core of the environmental description is a segmented color image we named *Panoramic Labelled Image*. The *Panoramic Labelled Image* is made as follows:

1. capture camera images at various moments,
2. segment the images by color information and label the segmented regions,
3. combine these small labelled images to form a panoramic image (see upper-right region of Figure 1),
4. iterate the procedures 1–3.

The reasons for using labelled image instead of raw image or edge image are;

- It makes it easier for the robots to check regions' correspondence and detect changes in environment than when using raw image or edge image.
- Using color information increases robustness of processing.
- It makes it easier to record and utilize the correspondence between regions and objects, and the relations between regions.
- It makes it easier to reconstruct description when a robot moves.

2.2 Components of the environmental description

As shown in Figure 1, the proposed environmental description consists of:

- *Panoramic Labelled Image* that covers all visible field of a robot,
- *Region Database* that has spatial and temporal information about each region,
- *Regions Relation Network* whose nodes and edges represent each region and adjacency state respectively.

Panoramic Labelled Image can be regarded as a wide-range image which is segmented by color information. The *Panoramic Labelled Image* is continuously modified based on newly obtained camera images. *Region Database* and *Regions Relation Network* are also modified in process. Since a series of processing need not control the cameras for itself, it can run as a background job of other vision processes.

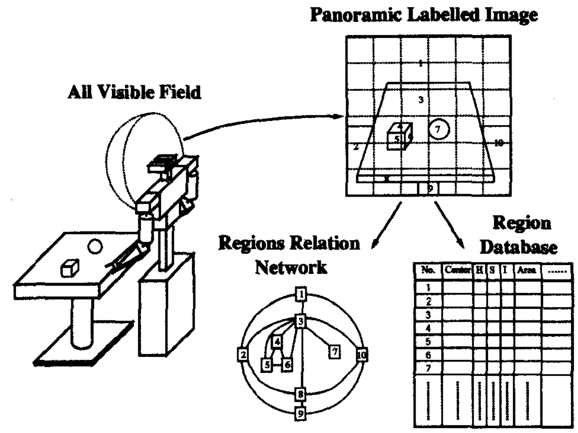


Figure 1: Environmental Description using *Panoramic Labelled Image*

Region Database contains spatial information of each region such as color, size, shape and depth, and temporal information which is a history of appearance, growth, reduction, and disappearance of each region. Moreover, if there is a region to which the correspondence of an object is made clear, its information is also recorded.

Regions Relation Network mainly represents adjacency relation between regions (see lower-middle region in Figure 1). When the relations of the adjacent regions, such that some adjacent regions belong to the same object, becomes clear, the relations are added as characteristics of edges. In addition, spatial relations such as which region is nearer to a robot than the other region are described.

2.3 Advantages of the environmental description

The environmental description is in a medium level of description between raw image (low level description) and fully symbolized description (high level description). The originality and usability of the proposed environmental description is summarized in the following four points:

- The *Panoramic Labelled Image* is modified and refined using many images, therefore, the uncertainty of the segmentation and the partiality of information decrease while robustness against modeling error increases.
- The *Panoramic Labelled Image*, *Region Database* and *Regions Relation Network* clearly describe the surroundings in a simple form, which makes it easier to perform high-level processing such as object search.
- Since the description memorize history of temporal changes in environment, it is easy to evaluate how regions overlap each other and whether the adjacent regions belong to the same object.

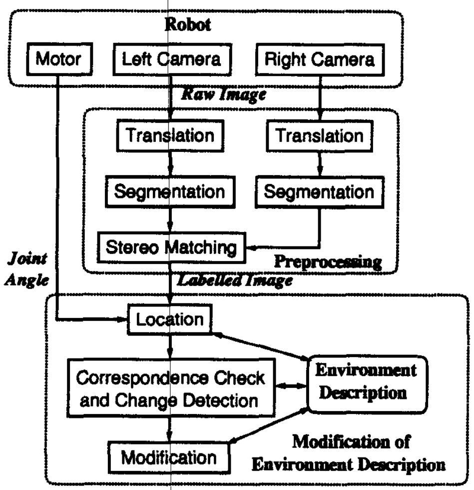


Figure 2: Procedure of making the environmental description using *Panoramic Labelled Image*

- One of the aims of the description is to artificially realize the visual short-term memory. The method does not require the active control of cameras to build the environmental description. Therefore, the process of making the environmental description is done in the background.

3 Making of environmental description

We discuss how to make the environmental description in this section. Figure 2 shows the flow of processes.

3.1 Preprocessing

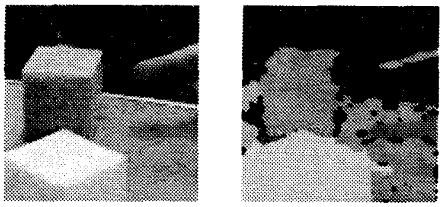
First, the images inputted from the cameras are segmented. The segmentation procedure is:

1. translate inputted image of RGB format to HSI format,
2. regard the parts of the image whose intensity is lower than 50 as *black* regions,
3. regard the parts of the image whose saturation is lower than 15 as *white* regions,
4. segment the remaining part to one-color regions.

HSI image format consists of Hue, Saturation, and Intensity. Although various translation schemes from RGB space to HSI space have been proposed, but we adopted the translation based on the hexcone model [5].

For segmentation we apply recursive thresholding [6], and for automatic thresholding we apply the method proposed by Otsu [7].

An example of the segmentation is shown in Figure 3. Figure 3 (a) shows the inputted image from the camera, while Figure 3(b) shows the segmented image. Black parts are discarded by segmentation process as noise or fragmentary regions.



(a) raw image (b) segmented image

Figure 3: Segmentation

Figure 4: An example of the *Panoramic Labelled Image*

After the segmentation, the image from right and left cameras are compared, and if there are corresponding regions, the disparity is calculated.

3.2 Modification of *Panoramic Labelled Image* and detection of changes in environment

An example of *Panoramic Labelled Image* is shown in Figure 4. Black part represents noise or fragmentary regions mentioned above or a part which a robot still has not seen.

Process of modification is concretely described in this section.

First, using joint angles of the head, it is determined to which part of *Panoramic Labelled Image* newly inputted image should correspond.

Secondly, the system searches for correspondences between regions in newly inputted image and those in *Panoramic Labelled Image*. As to the regions in the newly inputted image which do not have correspondence with the regions in the *Panoramic Labelled Image*, there are following two possibilities:

- uncertainty in segmentation,
- temporal changes in environment.

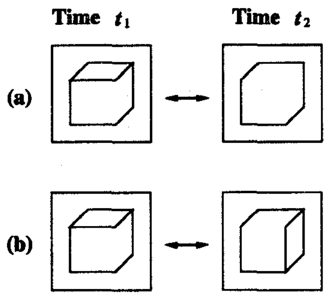


Figure 5: Uncertainty in segmentation: (a) appearance/disappearance of boundary, (b) difference in boundary

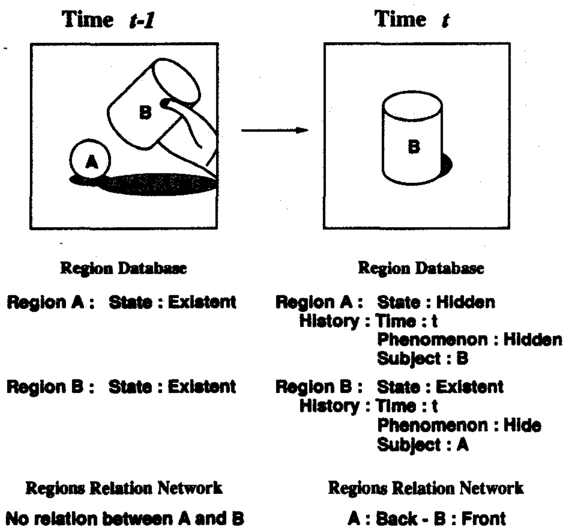


Figure 6: Description of change history in environment and spatial relation

Differentiation between these two is crucial.

Representative phenomena caused by uncertainty are shown in Figure 5. Figure 5(a) shows a phenomenon that an object is observed as single region at one time and two regions at another time. Figure 5(b) shows a phenomenon that the boundary of two regions changes. In these cases, representation in *Panoramic Labelled Image* is decided by vote of plural inputted images.

For regions where it is inferred that changes in environment have occurred, the characteristics of the changes such as appearance, growth, reduction, and disappearance are evaluated. Moreover, whether a region hides other regions or not is inferred from the depths of the regions.

3.3 Description of environment change history in *Region Database*

In modification of *Panoramic Labelled Image*, the history of changes is recorded in *Region Database* if any. If the number of record exceeds a specified number, the older record will be deleted. This elimination of older record corresponds to human oblivion.

Contents of change history recorded in the *Region Database* are the following information:

- quantitative information such as the coordinates of the center and the size,
- qualitative information such as appearance, growth, reduction, and disappearance of regions,
- the region concerned in the change.

Here, let us assume that the change in environment occurs as shown in Figure 6. In the period between times $t-1$ to t, region A has disappeared and region B, which is nearer to the robot than region A, appeared at the same point. Then the state of region A in *Region Database* is changed from "existent" to "hidden". Change history is also added to *Region Database.* To the column of region A, the information that region A was hidden by region B at time t is added. To region B, the information that region B hid region A at time t is added. Moreover, quantitative information such as the coordinates of the center and the size of each region is renewed.

3.4 Description about spatial relation

Spatial relation is inferred from changes in environment. If a change in environment are detected, its characteristics is estimated from depths of the concerned regions. For example, the changes in environment as shown in Figure 6 can be interpreted either as that region B hid region A or that region A was lost and hidden region B appeared. Whether interpretation is appropriate is determined by the relationship in depths between the regions.

The decision is reflected in *Regions Relation Network*, such as an edge is a) made b) discarded or c) re-characterized. In case of Figure 6, region B is nearer to the robot than region A, so a new edge with information that region B is in front of region A is made between regions A and B.

4 Experiments

4.1 Implementation of the environmental description

The environmental description is implemented on the humanoid robot *Saika* [4] (Figure 7), which has binocular camera arrangement and two 7-DOF arms. System configuration is illustrated in Figure 8.

Segmentation and labelling processes run on the DSP-based vision system ISHTAR-II [8]. The size of segmented images was set to be 32 pixel square and the segmentation process spent 1/30 second for each inputted image. All preprocessing time including color

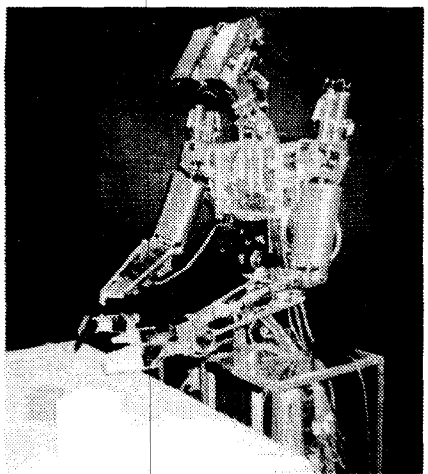


Figure 7: Humanoid robot *Saika*

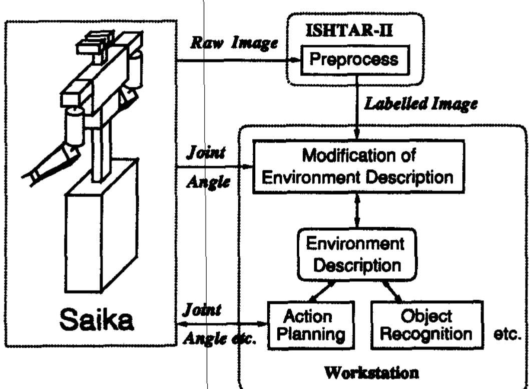


Figure 8: System configuration

translation, stereo matching and transmission to workstation is 4/30 second.

The size of *Panoramic Labelled Image* is 512 pixel square. The modification of environmental description is carried out on Sun SparcStation and spends 1/30 second for every inputted image. These processes are, however, done in pipeline processing, so the interval time of modification is 1/30 second.

4.2 Experiments

We adopted the delayed response as the action which utilizes the description of change history in environment and spatial relation between regions. The action is that, after we show the robot the environment we hide a certain object, then we command the robot to find out and pick up the hidden object. This experiment was performed using the humanoid robot *Saika* [4].

4.3 Procedure of the experiment and process in the robot

The experiment was executed as follows:

1. put some objects on the table, show this environment to the robot and make *Saika* memorize it. This phase corresponds to *cue* in the delayed response.
2. remove a certain object or hide one by another one while *Saika* is looking away. This phase corresponds to *delay* in the delayed response.
3. command *Saika* to find out and pick up the object. This phase corresponds to *response* in the delayed response.

When *Saika* received the last command, without memorizing the environment, *Saika* would have to search all environment by moving *Saika*'s viewpoint little by little. This search, however, takes long time, and if the object is hidden and cannot be seen *Saika* cannot find it.

With the environmental description, *Saika* is able to find out the object in a short time. The finding procedure is:

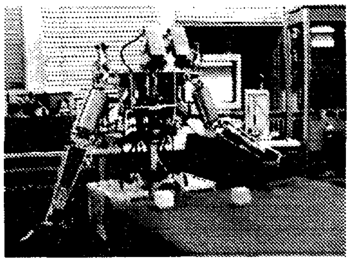
1. *Saika* searches *Region Database* for the regions corresponding to the object, and looks towards the region.
2. *Saika* observes the neighborhood of the region decided in 1, and see if the region still exists or some changes have occurred. In the case when the region still exists, *Saika* checks whether there is any obstacles. In the case when some changes have occurred, *Saika* determines whether the region has disappeared or has been hidden by other regions.

   - **In the case when the region exists and there is no obstacle**

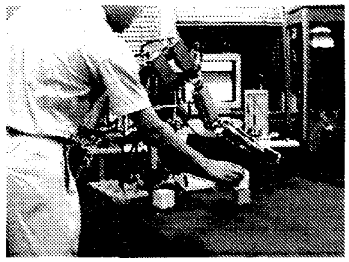
     Saika confirms that the region corresponds to the object and pick it up, so the task is completed.
   - **In the case when there are some obstacles or when the region has been hidden**

     Saika removes obstacles or hiding objects. After that, return to 1.
   - **In the case when the region has disappeared**

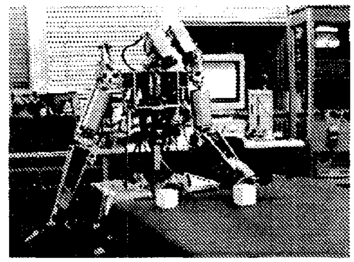
     Saika picks up other candidate. If there doesn't remain other candidate, *Saika* searches all environment again.



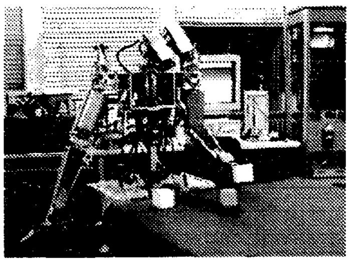
(a) *Saika* observes environment



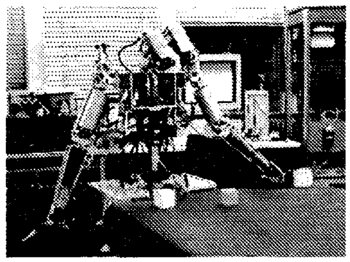
(b) human hides objects



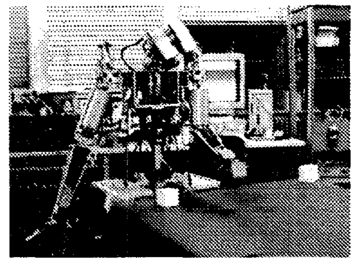
(c) human commands *Saika* to pick up an object



(d) *Saika* removes an obstacle



(e) *Saika* finds out the object



(f) *Saika* picks it up

Figure 9: Scenes from the experiment of the delayed response.

4.4 Results of the experiments

The scenes from the experiment are shown in Figure 9. In Figure 9 (a), *Saika* observes the environment. The human hides objects on the table while *Saika* is looking away ((b)), and commands *Saika* to pick up the object on the *Saika*'s left hand side ((c)). Once *Saika* receives the command, *Saika* looks towards the region where the object was. However, since the objects were hidden by an obstacle, the object cannot be seen. *Saika* infers the existence of the object under the obstacle from the depths of both the obstacle and the object, and removes the obstacle searching for the object ((d)). Since *Saika* successfully finds out the object under the obstacle ((e)), *Saika* picks it up ((f)).

5 Conclusion

We proposed wide-range environmental description which helps robots comprehend the spatial and temporal information of the environment. We also showed that it allows simple descriptions of the change history and spatial relations in environment. Moreover, we realized the behavior of finding out of the hidden object on a humanoid robot on the basis of the description. We make it clear that the variety of behavior of robots increased by the use of short-term memory.

In the next work, we will make use of the description of change history of environment to know whether the action of the robot has succeeded, and to make the robot refine the action automatically.

Acknowledgments

This research has been supported by Grant-in-Aid for Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF96P00801).

References

[1] J. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, Vol. 9, No. 1, pp. 55–76, 1992.

[2] M. Herman and T. Kanade. Incremental reconstruction of 3d scenes from multiple, complex-images. *Artificial Intelligence*, Vol. 30, pp. 289–341, 1986.

[3] A. D. Baddeley. Working memory. *Science*, Vol. 255, pp. 556–559, 1992.

[4] A. Konno et al. Development of a humanoid robot saika. In *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'97)*, pp. 805–810, 1997.

[5] A. R. Smith. Color gamut transformation pairs. *SIGGRAPH-ACM Computer Graphics*, Vol. 12, No. 3, pp. 12–19, 1978.

[6] Y. Ohta, T. Kanade, and T. Sakai. Region segmentation by recursive thresholding. *Computer Graphics and Image Processing*, Vol. 13, No. 3, pp. 222–241, 1980.

[7] N. Otsu. A threshold selection method from gray-level histgrams. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-9, No. 1, pp. 62–66, 1979.

[8] M.Shiohara et al. Real-time optical flow processor ISHTAR. In *Asian Conf. on Computer Vision*, pp. 790–793, 1993.