Distributed Vision System: A Perceptual Information Infrastructure for Robot Navigation

Hiroshi Ishiguro
Department of Information Science, Kyoto University
Sakyo-ku, Kyoto 606-01, Japan
E-mail: ishiguro@kuis.kyoto-u.ac.jp

Abstract

This paper proposes a *Distributed Vision System* as a *Perceptual Information Infrastructure* for robot navigation in a dynamically changing world. The distributed vision system, consisting of vision agents connected with a computer network, monitors the environment, maintains the environment models, and actively provides various information for the robots by organizing communication between the vision agents. In addition to conceptual discussions and fundamental issues, this paper provides a prototype of the distributed vision system for navigating mobile robots.

1 Introduction

Many researchers are tackling to develop autonomous intelligent mobile robots which behave in a real world in robotics and artificial intelligence. For limited environments such as offices and factories, several types of mobile robots have been developed. However, it is still hard to realize autonomous robots behaving in dynamically changing real worlds such as an outdoor environment. To develop such robots which can adapt to the dynamic worlds is the original purpose of robotics and artificial intelligence.

Attention control

As discussed in technical papers on *Active Vision* [Ballard 89], the main reason lies in attention control to select viewing points according to various events relating to the robot. Two kinds of the attention control exist; one is called *Temporal Attention Control* and the other is *Spatial Attention Control.* If the robot has a single vision, it needs to change its gazing direction in a time slicing manner to simultaneously execute several vision tasks. The control of gazing direction is *Temporal Attention Control.* For example, the robot has to detect free regions even while gazing at the targets. We, human, solve this complex temporal attention control with its sophisticated mechanisms of memory and prediction. Further, the vision fixed on the robot body sometimes cannot provide proper information for the vision tasks. For example, when a robot estimates collisions with a moving obstacle, the side view in which both the robot itself and the obstacle are observed may be more proper than the view from the robot. This view point selection is called *Spatial Attention Control.*

Difficulties in autonomous robots

To realize the attention control is difficult with current technologies for autonomous robots. The following reasons can be considered.

- Active vision systems need a flexible body for acquiring proper visual information like a human. However, vision systems of previous mobile robots are fixed on the mobile platforms and it is generally difficult to build mobile robots which can acquire visual information from arbitrary viewing points in a 3D space.
- An ideal robot builds environment models by itself and uses them for executing commands from humans operators. However, to build a consistent model for a wide dynamic environment and maintain it is basically difficult for a single robot. We, humans, sometimes need helps of other persons to acquire information on the environment.

One of the promising research directions to solve the above-mentioned problems is to develop an infrastructure which provides sufficient information for the robots. This paper discusses such an infrastructure. The infrastructure in this paper differs from the infrastructure for mobile robots which move in factories. Our purpose is not to develop systems which support individual functions of the robots such as guide lines and landmarks for locomotion, but to develop a *Perceptual Information Infrastructure* (PI^2) which actively provides various information for real world agents, such as robots and humans. That is,

> **The PI^2 monitors the environment, maintains the dynamic environment models, and provides information for the real world agents**

As the PI^2, this paper proposes a *Distributed Vision System* (DVS). The DVS consists of multiple cameras which have own computing resource and communication

links with others. The camera, called *Vision Agent* (VA), provides visual information required by the vision-guided mobile robots. The VAs at various locations provide sufficient visual information for attention control of the robots, and they maintain dynamic environment models by organizing communication between them. Of course, we can use other sensors in the infrastructure. The camera, however, is the most compact and low cost passive sensor to acquire various kinds of information and many interesting vision research issues are still remained. Another purpose of this research approach is to deal with the issues and develop real applications of computer vision through the DVS.

In addition to conceptual discussions and research issues, a prototype of the DVS and experimental results using it are shown. The author has confirmed the DVS can robustly navigate the mobile robot in a complex real world.

Related works

Recently novel research approaches using distributed sensors and robots have been proposed in robotics. For example, the *Robotic Room* proposed by Mizoguchi and others [Mizoguchi 96] support human activities with sensors and robots embedded in a room. Their interests are to design mechanisms and develop sensor system for executing well-defined local tasks. On the other hand, the purpose of this paper is to propose a flexible sensor system utilized by various kinds of robotics systems as an information infrastructure.

Several vision systems which utilize multiple cameras has been reported, especially, in multimedia. Moezzi and others proposed the concept of *Immersive Video* and developed a vision system using precisely calibrated cameras for building a precise geometrical model of an outdoor environment. Pinhanez and Bobick [Pinhanez 96] developed a system which dynamically selects cameras providing proper views for broadcasting a TV show. We, however, consider a demerit of the systems is to use calibrated cameras and geometrical models of the world. Geometrical representations of environments obtained by the calibrated cameras lack robustness and flexibility of the systems. In order to solve the problems, this paper proposes an alternative approach for modeling dynamic environments, which dynamically and locally estimates the camera parameters and directly represents robot tasks.

In distributed artificial intelligence, several fundamental works dealing with systems using multiple sensors have been reported. Lesser [Lesser 83] proposed a *Distributed Vehicle Monitoring Testbet* (DVMT) as an example of distributed sensing problems, and Durfee [Durfee 91] proposed *Partial Global Planning* which is a planning method for globally analyzing signals provided by multiple signal processing agents. The DVS, basically, can be considered as a kind of the distributed sensing systems, such as the DVMT, but deals with vision sensors and communicates with robots. And further, the purpose in the DVS is not to globally analyze the signals, but to navigate mobile robots with local information by representing the navigation tasks in the VA network.

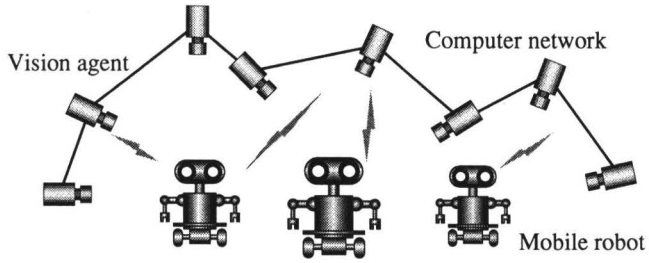


Figure 1: Distributed Vision System

2 Distributed Vision System

2.1 Concept of distributed vision

In order to simultaneously execute the vision tasks, an autonomous robot needs to change its visual attention. The robot, generally, has a single vision sensor and a single body, therefore the robot needs to make complex plans to execute the vision tasks with the single vision sensor. *Active Vision* proposed by Ballard [Ballard 89] is a research direction to solve the complex planning problem with active camera motions bring proper visual information and to enable real-time and robust information processing. That is, the active vision system needs a flexible body to acquire the proper visual information like a human. However, the vision system of previous mobile robots is fixed on the mobile base and it is generally difficult to build autonomous robots which can acquire visual information from arbitrary viewing points in a 3D space.

Our idea to solve the problem is to use many VAs embedded in the environment and connected them with a computer network (See Fig. 1). Each VA independently observes events in the local environment and communicates with other VAs through the computer network. Since the VAs do not have any constraints in the mechanism like autonomous robots, we can install a sufficient number of VAs according to tasks, and the robots can acquire necessary visual information from various viewing points. As a new concept to generalize the idea, the author proposes *Distributed Vision* that multiple vision agents embedded in an environment recognize dynamic events by communicating each other. In the distributed vision, the attention control problems are dealt as dynamic organization problems of communication between the vision agents.

The DVS is not a standard computer network. It is an extended computer network which bridges between physical worlds and virtual worlds building in the computer network. Current computer networks transmit only data, such as images and characters. However, as the services and functions of the computer networks are extended, more efficient and intelligent communication between computers are required. The author calls such

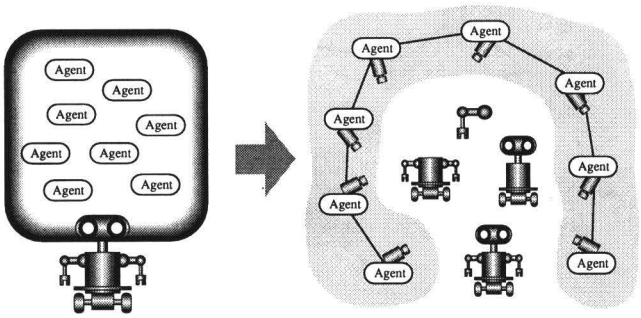


Figure 2: From an autonomous robot to robots integrated with environments

a future computer network *Perceptual Information Infrastructure* (PI^2). The PI^2 observes physical worlds, maintains dynamic models in the computer network and supports robots and humans. The DVS is one example of the PI^2.

In robotics, the PI^2 enables robust and flexible robotic systems by offering necessary information and develops a new research area. As shown in Fig. 2, a previous autonomous robot consists of a mechanical body and software agents. That is, the intelligence produced by the software agents. On the other hand, the intelligent information processing of robots supported by the PI^2 is done by agents embedded in the environment. Development of *Robots Integrated with Environments* is an important research direction for realizing useful robotic systems.

2.2 Design policies for the DVS

The VAs are designed based on the following idea:

> **Tasks of robots are closely related to local environments.**

For example, when a mobile robot executes a task of approaching a target, the task is closely related to a local area where the target locates. This idea allows to give VAs specific knowledge for recognizing the local environment, therefore each VA has a simple but robust information processing capability.

More concretely, the VAs can easily detect dynamic events since they are fixed in the environment. A vision-guided mobile robot of which camera is fixed on the body has to move for exploring the environment, therefore there exist a difficult problem to recognize the environment through the moving camera. On the other hand, the VA in the DVS easily analyzes the image data and detects moving objects by constructing the background image for the fixed viewing point.

All of the VAs, basically, have the following common visual functions:

- Detecting moving obstacles by constructing the background image and comparing with it.
- Tracking detected obstacles by a template matching method.
- Identifying mobile robots based on given models.
- Finding relations between moving objects and static objects in the images.

The DVS, which does not keep the precise camera positions for robustness and flexibility, autonomously and locally calibrates the camera parameters with local coordinate systems according to demand (the detail is discussed in Section 4.3). That is, the VAs iterate to establish representation frames for communicating with other agents.

The VAs identifies objects with the motions observed in the images in addition to the visual features since they can provide reliable motion information from the fixed viewing points. The author considers the DVS can solve the correspondence problem more robustly and flexibility than the previous vision systems.

The DVS organizes communication between VAs in order to execute given tasks. The design policy that a VA executes particular subtasks in the local environment allows to solve the organization problem in a hierarchical manner. That is, global tasks given to the DVS, generally, can be decomposed into the subtasks and the VAs execute them. However, the subtasks often need to be simultaneously executed and the combinations often change according to various situations. Therefore, the VAs should be globally and locally organized to execute the global tasks. The organization of VAs is the most important research issue of the DVS.

3 Fundamental issues

3.1 Communication between VAs

Remarkable difference of the DVS with previous computer systems is the DVS has two kinds of communication. In addition to communication with a computer network, VAs in the DVS communicate by observing common events. When two VAs which have own local internal representations simultaneously observe a robot from different viewing points, they may synchronously update their local internal representations. The VAs share symbolic and non-symbolic information through the computer network and the observations, respectively. It is an important research issue how to establish sophisticated and flexible communication links through the two types of communication. The author is especially interested in the non-symbolic communication which is difficult to deal with in previous frame works.

3.2 Dynamic environment model

The robust detection of dynamic events enables to hierarchically represent the environment. We, basically, consider static environment models should be generated from dynamic environment models representing the dynamic events. The dynamic environment models give meanings to static objects represented in the static models. For example, a gray region in the images, which is a road in the outdoor environment, is defined as a region where the robot can move. The DVS which can

easily detect the dynamic events is a promising system for realizing the hierarchical environment models.

3.3 Organization of VAs

The DVS needs to organize the VAs for acquiring the dynamic environment models. Let us imagine a DVS navigating a mobile robot. In order to avoid moving obstacles and detect free regions, the mobile robot needs visual information provided by VAs locating around it, and in order to go toward a destination, it also needs information about subgoals from VAs locating along the robot path. That is, the VAs should be locally and globally organized in order to provide proper information for the robot navigation. In the organization process, the DVS represents given tasks by organizing the VAs. For realizing the organization, it is necessary to develop new methods which deal with the total process including image understanding by the VAs, task understanding, and task execution through the symbolic and non-symbolic communication links.

3.4 Distributed model

The dynamic environment models are not shared by all VAs, but distributed over VAs. Robots access to the dynamic environment models through dynamic organizations of the VAs. For example, when a robot avoids a moving obstacle, the DVS continuously organizes the VAs located around it and navigates it. Further, if troubles occur in a VA, other VAs take the place of the VA. To distribute the models in the VA network is important for realizing flexibility and robustness of the DVS .

4 A prototype of the DVS

This section discusses a developed prototype of the DVS [Tanaka 97]. The prototype system briefly deals with the fundamental issues discussed in Section 3, but it does not completely solve them. The issues are carefully dealt with in the feature works.

4.1 Mobile robot navigation

The outline for mobile robot navigation by the DVS is as follows. First, a human operator teaches tasks by manually controlling a robot. The human operator does not directly give task models or behavior models of the robot, but gives examples to the DVS. While the robot moves, each VA tracks it within the visual field with simple image processing functions discussed in section 2.2. Then the DVS decomposes the given example paths into several components which can be maintained by each VA and memorizes then by organizing the VAs. After organizing the VAs, the DVS autonomously navigates the mobile robot while the VAs communicate each other. All of the VAs monitor the robot motions and send messages to other VAs according to the memorized organization patterns for global and local tasks.

4.2 The architecture

A VA consists basic modules and memory modules as shown in Fig. 3. For the basic modules, the

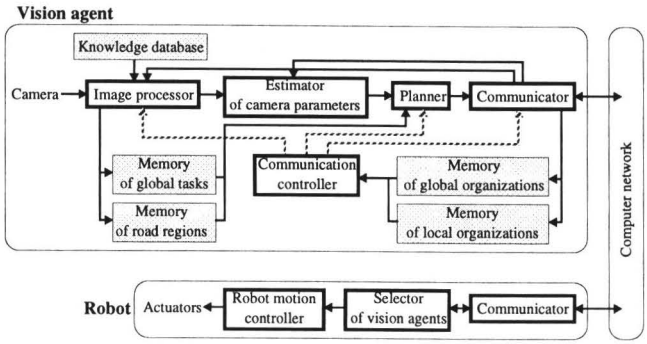


Figure 3: The architecture of the DVS

VA has *Image processor*, *Estimator*(Estimator of camera parameters), *Planner*, *Communicator*, and *Controller*(Communication controller). For the memory modules, it has a knowledge database for image processing, memories to memorize global and local tasks, and memories to maintain relations with other VAs for executing the global and local tasks. In this experimentation, the global task is to navigate toward goals and the local task is to avoid obstacles.

Image processor detects moving robots and tracks them by referring to the knowledge database which stores visual features of robots. *Estimator* receives the results and estimates camera parameters for establishing representation frames for sharing robot motion plans with other VAs. *Planner* plans robot actions based on the estimated camera parameter and sends them to the robot through *Communicator*. The robot corrects the plans, selects proper plans, and executes the plans. The selected plans are sent back to the VAs and memorized. The memorized plans are directly applied in the same situations of the VAs and the robot by *Controller*.

4.3 Global and local organization

For selecting and integrating the robot motion plans from VAs, all plans at a time should be represented with a common representation frame. In the case of the DVS, the robot motion is represented with a $X - Y$ robot path-centered coordinate system. The coordinate transformation from the camera frame of a VA to the robot path-centered coordinate system is represented with two parameters α and β. Since the vision data is very noisy and the obtained plans are simple, the DVS assumes orthographic projection for the obtained image and represents the coordinate transform with only two parameters of camera rotations. As a more sophisticated method, it is possible to use an automatic calibration method proposed by Hosoda and others [Hosoda 94].

Fig. 4 shows data flows between functions for globally and locally organizing VAs. *Estimator* computes α, β and their error estimates $\Delta\alpha$ and $\Delta\beta$ for the coordinate transformation (Estimating and updating camera parameters). *Planner* plans robot motions with the obtained coordinate system (Planning a robot motion).

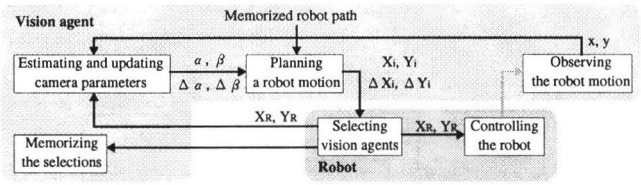


Figure 4: Global and local organization

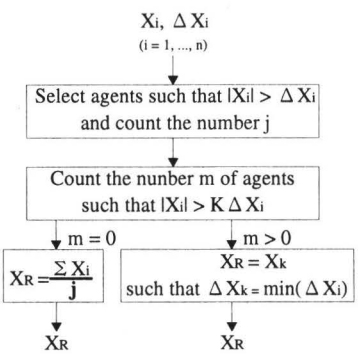


Figure 5: Plan selection for the global task

The plans X_i and Y_i and their error estimates ΔX_i and ΔY_i by agent i are send to the robot and the robot selects proper plans based on the error estimations (Selecting VAs). While the robot executes the integrated plans X_R and Y_R, it sends the plans to the VAs. The VAs memorize the selection results as organization patterns.

The robot selects the proper plans for global as shown in Fig. 5 by referring to the error estimates. For the plans such that $|X_i| > \Delta X_i$, the robot computes the mean value if the error estimate is large ($|X_i| > K\Delta X_i$, K=2), otherwise, it selects a plan which has the minimum error estimate. The algorithm means that if $|X_i| > \Delta X_i$ the plan is incorrect and if $|X_i| > K\Delta X_i$ the plan is roughly correct. For the local task, avoiding obstacles, the robot selects vision agents such that $|X_i| > \Delta X_i$. The VAs generate negative values of X_i and Y_i as the plans to avoid obstacles.

4.4 Experimental setup

Fig. 6 shows a model town and a mobile robot used in the experimentation. The model town, of which reduced scale is 1/12, has been made for representing enough realities of an outdoor environment, such as shadows, textures of trees, lawns and houses. Sixteen VAs have been established in the model town and used for navigating the mobile robot.

Images taken by the VAs are sent to an image encoder which integrates sixteen images into one image (the size of each image is reduced from 512 x 512*pixels* to 128 x 128*pixels*). Then, the image is sent to a color frame grabber and a motion estimation processor which can detect optical flows in real time. The main computer,

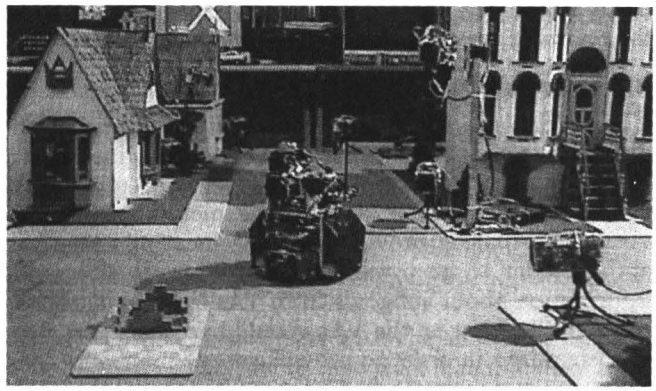
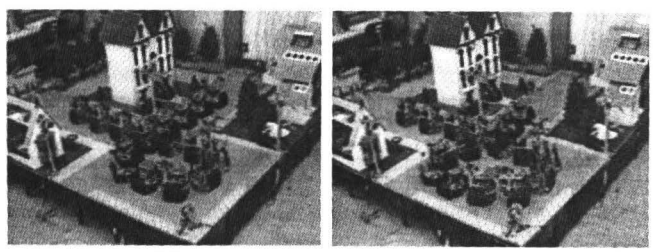


Figure 6: Model town



(a) (b)

Figure 7: An example path and a robot trajectory navigated by the DVS

Sun Spark Station 10, executes the vision functions by using data from the color frame grabber and the motion estimation processor. This system, unfortunately, cannot compute in parallel. The author is currently developing a new parallel computing system using twenty four C40 DSPs.

4.5 Experimental results

Fig. 7(a) shows an example path taught by a human operator in the teaching phase. Fig. 7(b) shows a robot trajectory autonomously navigated by the DVS. Because of simplicity of the image processing, the DVS could robustly navigate the mobile robot in a complex environment.

Fig. 8 shows images taken by VAs in the autonomous navigation phase. The vertical axis and the horizontal axis indicate the time stamp and the ID numbers of the VAs, respectively. The white boxes and the black boxes indicate selected VAs for the global and local tasks, respectively. Here, all VAs which simultaneously observe the robot are locally organized. As shown in Fig. 8, the DVS have dynamically organized the VAs for executing the global and local tasks.

The experimentation shows important aspects of the DVS. The DVS can memorizes the tasks for navigating the robot along a path by organizing the VAs and iterate to select proper VAs for robustly executing the tasks in

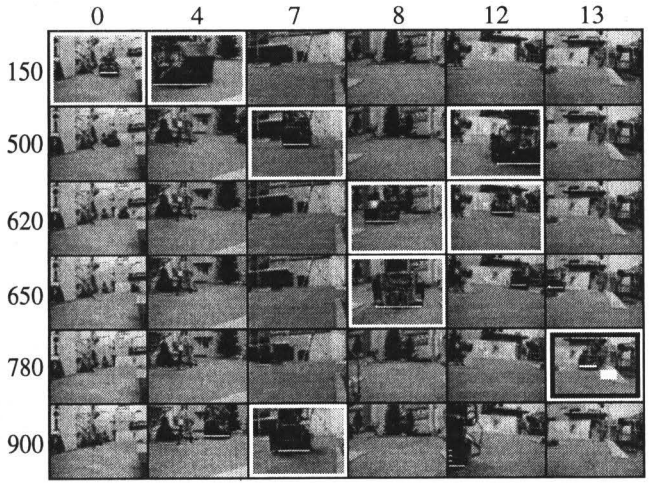


Figure 8: Images taken by VAs

a complex environment. That is, the DVS solves the attention control problems for the autonomous robots discussed in Section 1 with a different but more robust manner.

5 Conclusion and Research Plan Toward PI^2

The DVS is an alternative approach for realizing robust behaviors of intelligent robots. By organizing VAs, the DVS tightly couples robot actions and observations by the VAs. The previous autonomous robot has constraints of the hardware, the size and the number of sensors. On the other hand, the DVS does not have such constants and it is a promising approach for realizing real systems.

The purposes in this research approach are to solve the fundamental problems of the DVS, to develop real systems and to establish key technologies of PI^2. Especially, to develop systems for real applications is important. It will extend possibility of robotics and computer networks and be able to make users realize usefulness of computer vision techniques. The plan of this research is as follows:

1. Develop a DVS for navigating robots with a model town as a testbed (See Section 4).
2. Study the fundamental issues of mobile robot navigation (See Section 3).
3. Develop a DVS for observing and supporting human behaviors.
4. Study the fundamental issues of the human behavior support.
5. Extend the DVS to the PI^2 which supports various real world agents and develop a PI^2 in the university campus.

Results for the fundamental problems can be estimated with traditional evaluation criterion, such as originality and applicability. On the other hand, evaluations for the developed infrastructure is not so easy since the system should be collectively evaluated with various criterion. For the evaluations, the author considers it is important to disclose the details of the system development, which cannot be reported in technical papers, with the world-wide web (See http://www.lab7.kuis.kyoto-u.ac.jp/vision/).

Recent developments of multimedia computing environments have established huge number of cameras and computers in offices and towns. They are expected to be more intelligent systems as the PI^2 discussed in this paper. The PI^2 is a key issue in the next decade.

Acknowledgment

The author would like to thank Prof. Toru Ishida for his stimulating discussions and constructive criticism, and Mr. Goichi Tanaka for his programming work.

References

[Ballard 89] D. H. Ballard, Reference frames for animate vision, Proc. IJCAI, pp. 1635-1641, 1989.

[Durfee 91] E. H. Durfee and V. R. Lesser, Partial global planning: A coordination framework for distributed hypothesis formation, IEEE Trans. SMC, Vol. 21, No. 5, pp. 1167-1183, 1991.

[Hosoda 94] K. Hosoda and M. Asada, Versatile Visual Servoing without Knowledge of True Jacobian, Proc. IROS, pp. 186-193, 1994.

[Lesser 83] V. R. Lesser, and D. D. Corkill, The distributed vehicle monitoring testbed: A tool for investigating distributed problem solving networks, AI Magazine, pp. 15-33, 1983.

[Moezzi 96] S. Moezzi, An emerging Medium: Interactive three-dimensional digital video, Proc. Int. Conf. Multimedia, pp. 358-361, 1996.

[Pinhanez 96] C. S. Pinhanez and A. F. Bobick, Approximate world models: Incorporating qualitative and linguistic information into vision systems, Proc. AAAI, pp. 1116-1123, 1996.

[Mizoguchi 96] H. Mizoguchi, T. Sato and T. Ishikawa, Robotic office room to support office work by human behavior understanding function with networked machines, Proc. ICRA, pp. 2968-2975, 1996.

[Tanaka 97] G. Tanaka, H. Ishiguro and T. Ishida, Mobile robot navigation by distributed vision agents, Proc. ICCIMA, pp. 86-90, 1997.