Proceedings of the 2000 IEEE/RSJ
International Conference on
Intelligent Robots and Systems

A Robot Recognizing Everyday Objects

-- Towards Robot as Autonomous Knowledge Media --

Hideaki Takeda[†‡] Atsushi Ueno[‡] Motoki Saji[‡],
Tsuyoshi Nakano[‡] Kei Miyamato[‡]

[†]**The National Institute of Informatics**
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430, Japan
takeda@nii.ac.jp

[‡]**Nara Institute of Science and Technology**
8916-5, Takayama, Ikoma,
Nara, 630-0101, Japan
ueno@is.aist-nara.ac.jp

http://ai-www.aist-nara.ac.jp/KE/

Abstract

In this paper, we discuss roles of robots as autonomous knowledge media and show our prototyping system of an office work assistant robot based on this approach. We are surrounded by a huge amount of artifacts and information that is making difficult for human to deal with. Robots can help people by gathering and arranging information intelligently instead of people themselves. Our prototyping system called Kappa III is an office work assistant robot that can tell people location of daily goods in office. It firstly looks around to capture images of desks in order to remember what and where such goods. Then it can identify them by cutting them out from the background and categorizing them by color, shape, and figure. People can ask it to take goods in office by either specifying names or features such as color. We also realized discovery of objects three-dimensionally by comparing captured scenes with expected ones.

Keywords: *office robot, knowledge media, object recognition, intelligent environment*

1. Introduction

We are surrounded by an enormous amount of artifacts and information that are increasing year by year. Meantime we are getting frustrated probably because the amount is going to exceed our capacity of cognition. There are various types of research on how an enormous amount of information can be organized and integrated in information management field. Agent is one of the key techniques to enhance human ability because agent is expected to perform actions in information gathering and integration instead of human.

But such techniques are restricted in the information space or network, i.e., they are not able to handle items in the real world. Although our life is engaging to the information world more and more, the basis of our life is still on the real world. So far what we need is techniques to integrate the real and information worlds seamlessly in order to handle the enormous amount of items and facts in the real and information world in a unified way.

In this paper, we discuss how the information and real worlds should be integrated, and propose *knowledgeable environment* approach where robots are intermediate media to bridge human and the real and information worlds.

2. Knowledgeable Environment

Intelligent environment [1] is one of approaches to integrate the real and information world. Intelligent environment aims to create an environment where all objects and actions are recognizable by computers. We emphasize participation of people to integrate these worlds. Integration should be not done with respect to convenience for computers but with respect to human. We therefore propose the idea called "knowledgeable environment" where knowledge has a key role to integrate human, the information and real worlds [2]. Objects in the real and information worlds should be

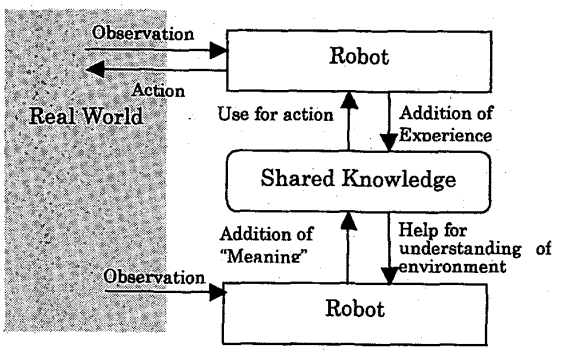


Figure 1: The knowledgeable environment approach

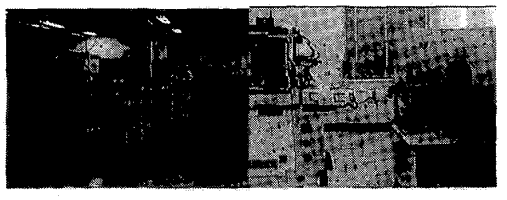


Figure 2: Kappa I

modeled according to human knowledge, and especially active objects like robots have knowledge to determine their actions (Figure 1). We adopt ontologies as basic knowledge shared by human, robots, and objects.

We built the first prototype system called *Kappa I* as multi-agent systems (see Figure 2). Physical objects like robots, a shelf and a door are modeled as agents as well as some information facilities like image processing and mediation (see Figure 3). All agents use a single inter-agent protocol and language, KQML and KIF respectively, and ontologies on object, space, and action.

Typical behavior of the system is as follows. *Watcher* agent observing the room can find needs and compose tasks with respect to the needs. Then mediation agent called *mediator* interprets such a task and decomposes it

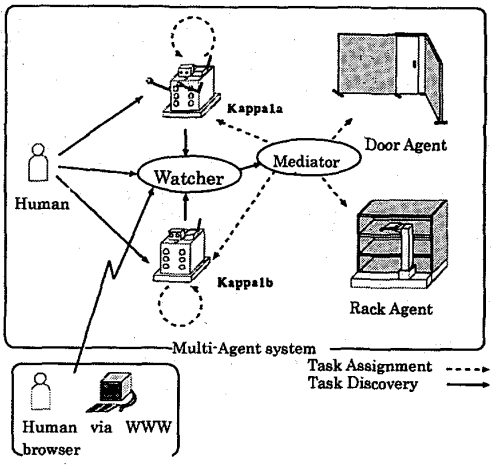


Figure 3: Architecture of Kappa I

into a sequences of smaller tasks each of which can be performed by a "physical" agent by using knowledge on agent ability of performance. Agent ability of performance is described with ontologies for objects, space, and actions that are necessary information to identify phenomena in the real world.

3 A Robot Recognizing Everyday Objects

Due to shared knowledge, the above system can notice and handle registered objects and actions correctly. But knowledge is to be provided, not to be acquired from the environment. The real environment is very dynamic, i.e., objects always come in and out the environment. We should support such knowledge for dynamic environments.

In order to realize acquisition of dynamic knowledge bridging human and the real environment, we implemented a new prototype system for the

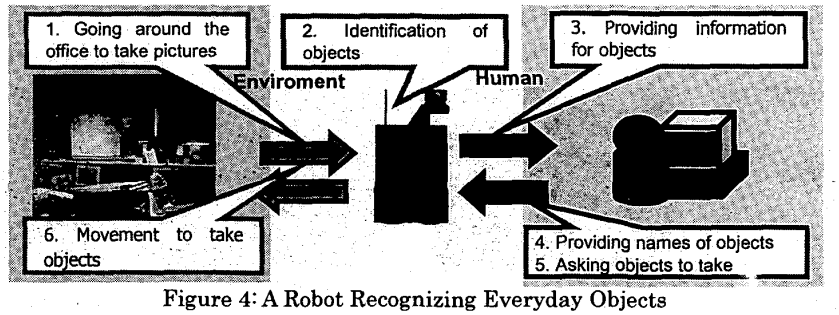


Figure 4: A Robot Recognizing Everyday Objects

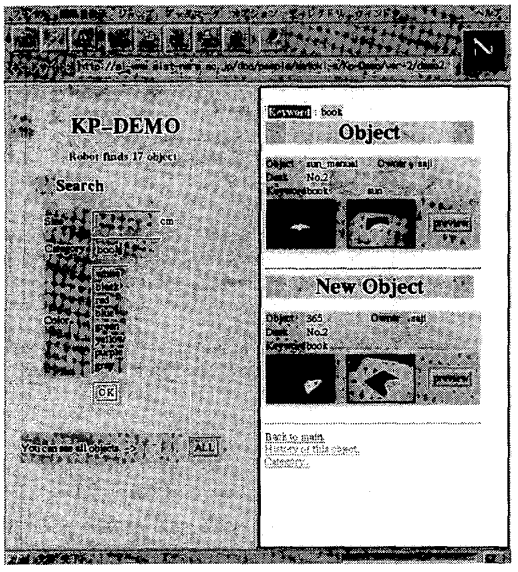


Figure 5: Interface for human in Kappa III

knowledgeable environment approach, i.e., a robot recognizing everyday objects (Figure 4). The system can acquire knowledge for objects in the real world in cooperation with a robot and people. A robot can go around the office and take snapshots of objects in it. The robot recognizes physical properties of objects and categorizes them according to such physical properties. People who want some object can browse objects that are collected and categorized by the robot and specify one of them. They can also tell names of objects to the system in order to specify objects easily. After specifying objects, the robot can go to the location where the specified object is located to take it. By repeating this process, the system can accumulate knowledge for objects.

3.1 Architecture of Kappa III system

We implemented the system called *Kappa III* to realize the knowledgeable environment in which knowledge of everyday objects can be accumulated by cooperation with robot and people.

There are two basic modes in the system. The first mode is the *patrolling* mode, and the other is the *mission* mode. The patrolling mode is to be executed in every certain period to update information for objects. In this mode, the robot goes around the office and takes pictures of objects on desks. Then it processes these pictures to extract information for objects on the desks.

The mission mode is an interactive mode with people to take objects according to human requests. A user asks the system to take some object by using WWW browser (Figure 5). Specification of an object is either by typing its name, or choosing some attributes like color and size, or browsing object images and selecting one. In the browsing process, the user can tell names of unnamed objects in the system. Then such names are used for specification by name thereafter. Then the robot moves to the location where the specified object is located.

3.2 Acquisition of physical attributes of objects

We here describe how the system can acquire physical information from collected images. Since the collected images by the patrolling mode are snapshots of desk tops, there needs to cut object shapes form the background and restore shape and size in order to extract objects from images. Furthermore we should identify objects in the different images because a single object can be observed many times after iteration of patrolling modes. We need categorization of objects cut from images. We call an object cut from an image an *object instance* that is a snapshot of a physical object and categories of object instances an *object class* that corresponds a physical object. The procedure of acquisition of physical attributes of objects consists of the following five steps.

(1) Cutting an object shape from the background

Cutting an object shape from the background is done

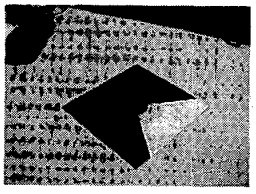


Figure 6:
A captured image

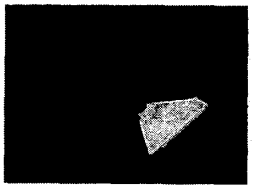


Figure 7:
Cutting the object

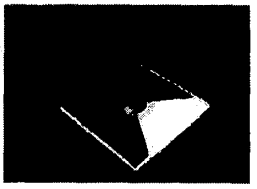


Figure 8:
Simplification of color

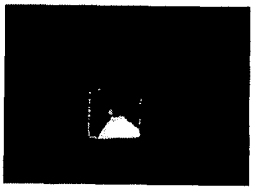
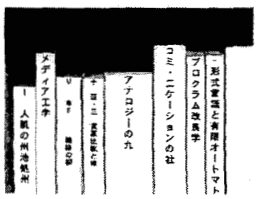


Figure 9:
Perspective Translation

(a) A Captured Image (b) Character Area



(c) OCR results

Figure 10: Extraction of Characters

with edge detection. It is relatively easy because we can assume that objects are located on the desk that is a single color (Figure 6 and 7).

(2) Acquisition of color features

We simplify color of objects into eight colors. One reason is for human to be able to specify color names, and the other is to reduce changes of colors according to changes of light circumstance (Figure 8).

(3) Restoration of shape and size

Then we restore shape and size of objects by perspective translation. It is an easy task because the height and angle of the camera on the robot and the height of desks are given (Figure 9).

(4) Normalization of object images

We re-locate object images in desktop images in order to compare object images from different attitude and location on desks. We call these images object instances.

(5) Identification of objects

Finally we find categories of object instances that are different snapshots of the same objects. We call these categories object classes. We categorize object instances according to their size, shape, and colors. When a new object instance is introduced, we first compare it with registered instances by size, shape, and color. If we find a similar instance for it, it is categorized into the category that the similar instance belongs to. Otherwise, a new category is created for it. Comparison is done with color matching of every pixel.

4. Extraction of Character Information

As one of extension of the above system, we tried to extract characters in the environment in order to obtain more various types of information in the environment. There are many characters where we live and work, i.e., documents, book titles, posters, guide panels, and so on, and they are informative for us. We implemented a prototype system to obtain character information from bookshelves and posters on walls.

It is much more difficult task to obtain characters from images of environment than images of documents because camera resolution is limited and styles of character and sentence are more various. We can solve the first problem by movement of robots and controlling of camera, but cannot solve the second problem yet because it is the problem of OCR and we used a commercial product of OCR in this system.

The process of acquisition of character information consists of two parts, i.e., acquisition of images and extraction of characters from them. The robot moves in front of bookshelves or posters on wall. It first takes an image to cover a wide area of shelves or posters and then takes images that can be processed by the OCR program by controlling the tilt angle and zooming function (Figure 10(a)). The robot repeats the process for every point to observe the specified bookshelves or posters.

Then collected images are processed to extract character information. We first determine character areas by checking distribution of black and white pixels (Figure 10(b)), and then pass these areas to the OCR program to extract character sequences (Figure 10(c)). Furthermore we join character sequences if needed. Character sequences are often divided into some images. Because camera resolution is about five times lower than scanner, images are well zoomed up to match OCR requirements. Joining of character sequences are done by using locations of character areas.

Human can use the acquired information by keyword search. When a user can give some words to the system, the system can find where book or poster containing them is located and show its image. Search of given words is done with ambiguity because recognition rate of OCR is expected low. The system answers five possibilities by varying partial sequences of the given words to search.

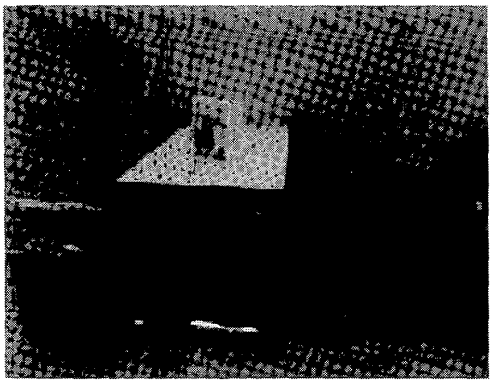


Figure 11: The captured image

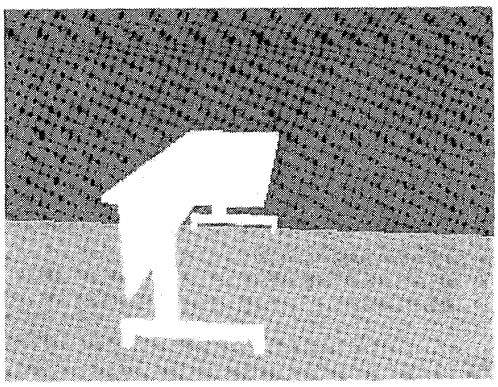


Figure 12: A generated image by 3D space

As a result, although recognition rate is low (26.2% for bookshelves and 68.4% for posters), search results are relatively good (57.3% precision and 59.3% recall for 30 test cases).

5. Discovery of Objects in 3D Space

The other extension to capture information in the environment is discovery of objects in 3D space. In section 3, we aim to capture objects only on desks, but we need to capture other objects in office. In this section, we show a simple approach to capture objects in 3D space. The basic idea is to categorize objects in space into two, i.e., fixed objects and movable objects. Fixed objects are objects that move rarely and can be modeled in advance, and movable objects are objects that move frequently and that we want to detect. By using fixed objects as reference, we can easily detect movable objects form images. When the robot moves around the office, it compares the scene from its camera and scene from 3D virtual space where fixed objects are arranged, and detect new movable objects as difference between two scenes. The basic process is as follows;

(1) Creation of fixed objects

We provide fixed objects as 3D models and arrange them in 3D virtual space (VRML). A 3D model has its actual shape and a certain number of colors as representative color of the actual objects. We also arrange known movable objects as initial situation.

(2) Generation of images by robot camera (Figure 11)

(3) Generation of images by 3D virtual space

We generate a view image from the position of the robot by the 3D virtual space (Figure 12).

(4) Comparison of two images

By comparing two images, we can detect new movable objects as areas of unmatched pixels. Comparison of two images is done as comparison of every corresponding pixel by color. In this compassion, color of pixel in the 3D space is the registered set of color for objects. If a connected area of unmatched pixels exceeds the threshold, we infer a new object. In Figure 13, three areas are identified as unknown objects.

(5) Estimation of attributes of new objects

We estimate attributes of a new object from the figure and location of the unmatched area. Since we just use a single vision system, we cannot estimate 3D shape and location in principle. But we estimate them very roughly. In this system, we assume that every object is a box whose height and depth are equal. With this assumption, 3D shape is estimated from shape of the area. Its location is estimated by inferring which fixed object it is on. After determination of its shape and location, it is put into the 3D virtual space so that the 3D virtual space is updated.

Figure 14 is the result of the estimated objects. A box on the desk and a box beside the desk are well estimated in shape and location but a chair is not correctly estimated because its legs are not recognized.

By repeating this process, the virtual space is always updated. This information is used for robots to plan paths as well as for human to recognize changes of the environment. The difference from sensor-based environment recognition is not to recognize just change of shape of obstructive situation but to recognize addition or deletion of objects so that it is informative

Figure 13: Detection of new objects

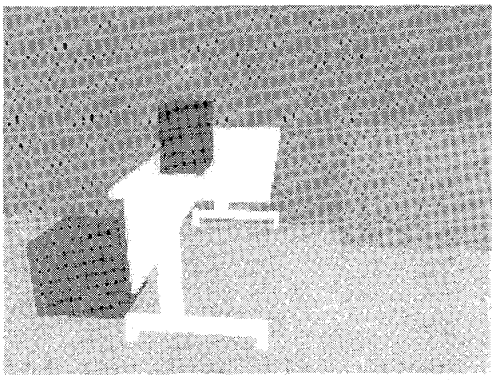


Figure 14: Arrangement of new objects

both for robots and human.

6. Related Work

As we mentioned, there are some proposals and systems for intelligent environments [1]. The difference lies on how human should be involved in the systems. We emphasize involvement of people in the system. We need sufficient information sharing between people and the system as well as sufficient interface between them

For example, Robotic Room [3] is an integrated system with various sensors and robots mainly for diseased people. In this system human is not deeply involved because human is just to be observed and received some services by the system. Intelligent Room [4] is the same approach for human involvement. Jijo-2 project [5] aims to build a secretary robot in office. In this project, the robot is autonomous by recognizing the environment and asking questions to people if needed. In this system, human is expected to have more active roles in the system by regarding that robot and human should be equal partners. But interface for the system and human is restricted to be occasional interaction by voice.

7. Summary and Future Work

We propose the knowledgeable environment to bridge human activities and robot- or computer-based activities. The key issue is to provide sharing information available and usable both for human and computer. Rapid growth of information infrastructure will soon enable to have good information sharing facilities everywhere. Availability alone is not enough, but usability is more important. The key measurement for usability is how much such information is based on knowledge for both partners. For computer and robot side, it should be independent from particular programs and show functionality of these machines. For human side, it should be natural representation. We tried in our system how such information should be and examined that such information is possible and promising.

We just tested our approach with very limited environment. We should discuss how sharable knowledge should be more generally and propose architectures for it.

References

[1] M. Coen (ed.), Intelligent Environments, Technical Report SS-98-02, AAAI, 1998

[2] H. Takeda, N. Kobayashi, Y. Matsubara, and T. Nishida, A knowledge-level approach for building human-machine cooperative environment. In A. Drogoul, M. Tambe, and T. Fukuda (eds.), Collective Robotics, Lecture Notes in Artificial Intelligence 1456, pages 147-161. Springer, 1998.

[3] T. Sato, Y. Nishida, and H. Mizoguchi: "Robotic Room: Symbiosis with human through behavior media," Robotics and Autonomous System, Vol. 18, pp185-194, 1996.

[4] M. Coen: Design Principles for Intelligent Environments. AAAI/IAAI 1998: 547-554

[5] T. Matsui, H. Asoh, I. Hara, Y. Motomura, S. Akaho, T. Kurita, F. Asano, and N. Yamasaki: An office conversant mobile robot that learns by navigation and conversation, Proceedings of the 1997 RWC symposium, pp.59-62, Jan. 29, 1997.