MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

8—11 A Multi-resolution Image Understanding System Based on Multi-agent Architecture for High-resolution Images

Keiji Yanai *
Department of Computer Science,
The University of Electro-Communications

Koichiro Deguchi †
Graduate School of Information Sciences,
Tohoku University

Abstract

Recently a high-resolution image that has more than one million pixels is available easily. However, such an image requires much processing time and memory for image understanding. In this paper, we propose an integrated system of multi-resolution analysis and multi-agent-based image understanding system for high-resolution images. We implemented an experimental system for images of indoor scenes.

1 Introduction

Recently a high-resolution image that has more than one million pixels is available easily due to high-performance digital cameras. However, in researches of image understanding, since such a high-resolution image requires much time and memory to process, the image is usually reduced to a lower-resolution image that has ten thousands of pixels. It may throw away significant information included in the high-resolution image. The feasible practical solution is to exploit multi-resolution analysis for high-resolution images, where we use a low-resolution image and recognize rough structure of scene at first, and use only the needed parts of a higher-resolution image later.

Multi-resolution analysis was originated as works of image processing such as edge detection and region segmentation in 1980s, and later it was applied to image recognition systems. In works by Z.Li[1] and by C.L.Tan[2], an image pyramid was introduced by reducing resolution of a high-resolution input image in several steps. Then, first the rough structures were extracted from low-resolution image, and next significant parts of higher-resolution images were selected and processed based on the rough structure. However, the systems they implemented are very restricted, because of lower ability of computers of those days. Nowadays since computers have made rapid progress, we can realize more large-scale and complicated system.

Here, we introduce multi-resolution analysis to prevent the loss of the high-resolution information to the multi-agent-based image understanding system[3][4]. Our multi-agent-based architecture for an image understanding, MORE (Multi-agent architecture for Object REcognition), is suitable for large-scale and complex recognition system due to its flexible and extensible architecture.

Almost conventional object recognition systems with multi-resolution analysis aimed at recognizing a single object. On the other hand, the objective of our system is to recognize multiple objects in a single image of real-world scene including complex occlusions. In our research, the "recognition" means to obtain a category name of the object, such as "desk" and "chair", from real world scene.

In this paper, we describe design and implementation of a multi-agent-based system employing multi-resolution analysis. The system recognize objects that couldn't be recognized in a low-resolution image without much extra time and memory. We implemented an experimental system for indoor images on PC cluster system.

2 Introducing Multi-resolution Analysis

Generally in image recognition systems unless more than a certain quantity of image features such as line and regions are detected in the initial stage, it is impossible to generate object candidates. Then, our and almost other image understanding systems have mechanism of "re-recognition". It means that the system tries again to recognize objects that were not found in the initial stage again. However, there has not been established method of "re-recognition", so many systems exploit such ad hoc methods as changing parameters or thresholds in low-level image processing algorithms. Object candidates only whose small parts are seen in an input image cannot be recognized by those methods.

In our system, we use a lower-resolution image in the initial stage, and in the "re-recognition" stage we select proper resolution from the image pyramid and process only needed part of the image. An image pyramid is constructed by reducing resolution of an image in several steps (Figure 1). We call images in the pyramid an image of level 0, 1, 2, ..., respectively, in order from the original image.

In the re-recognition stage, the system estimates regions where undetected object candidates are expected to exist using already detected candidates

*Address: 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, JAPAN. E-mail: yanai@cs.uec.ac.jp

†Address: Aramaki-aza-Aoba 01, Aoba-ku, Sendai 980-8579, JAPAN. E-mail: kodeg@fractal.is.tohoku.ac.jp

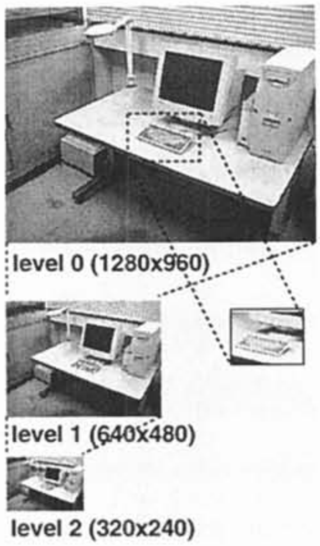

Figure 1: Image pyramid for multi-resolution analysis.

and "relational knowledge". It selects resolution of an image so that number of pixels of that region is less than a certain threshold value. In short, if region of interest is determined, resolution level of image which we use for recognition is determined. We call this resolution level selected for a certain region "**proper level**" for the region.

In Figure 1, the image pyramid consists of three images the resolutions of which are 1280x960, 640x480 and 320x240, respectively. At first, the system analyzes the image of maximum level, that is, level 2, and extracts floor, a desk and a display on the desk. Next, it analyzes the region of the desk candidate extracted from level 0 and its peripheral region in the image of level 1 to examines where if there are some objects on the desk. If the region where an object is expected to exist is detected, it analyzes the region using the image of level 0. Then, a keyboard in front of the display can be detected.

3 Recognition Strategy

Target image of our system is an image of scene that consists of artifacts, for example, laboratory scene and PC room scene. The system recognizes each single object and its relations to other objects in an image.

For recognition of each single object, first, the system extracts line segments and regions by conventional image processing algorithms, for example, Canny edge detector, Hough transformation, region growing segmentation method, snake and so on. Next, it searches groups of line segments and regions corresponding to each element of a structure model of an object. It fits the model to the group of line segments and regions extracted from the image. The system estimates 3D structure of an object by

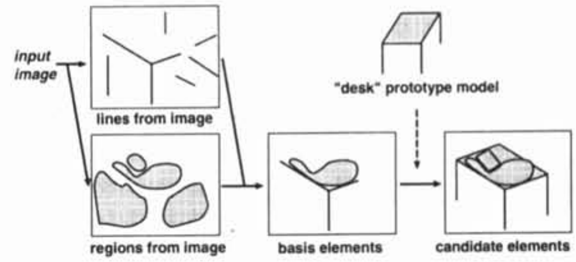

Figure 2: Flow of estimating an object candidate of "desk".

fitting a 3D structure model to the image qualitatively (Figure 2). We compute confidence value of a candidate as a weighted sum of the ratio of fitting model to segments and regions extracted from the image.

Next, it checks "supporting relation" between objects using spatial relations among fitted models. "Supporting relation" describes that which object supports which one. All objects are supported by other objects in the real world due to the gravity of the earth. We provided such physical knowledge to the system. Using this knowledge, the system expects existence of an undetected object under a detected object in the re-recognition stage. Details of model fitting and supporting relation are described in [4].

The system has "relational knowledges". They are descriptions about relative relation generally expected between two objects.It is used for computing confidence value of relation and expecting the region where own target object exists with high possibility in the re-recognition stage.

4 Overview of the System

We designed the system based on "MORE" architecture we proposed in [3]. It is multi-agent-based architecture and constructed as an assembly of agents that recognize objects from an image separately. It enables to recognize various different kinds of objects by adding agents. In our system, one agent consists of **a recognition module**(RM), **a communication module**(CM) and **candidate objects**(CO) (Figure 3). In addition, the system has **a feature extraction module**(FE) that makes an image pyramid and extracts straight edges and regions by requests from RM. The processing flow among all the modules is message-driven.

RM recognizes only one kind of target objects by sending FE requests of extracting image feature. CM carries out cooperation among agents. It checks supporting relations to candidates generated by other agents and resolves conflict among the agents. Using supporting relation and relational knowledges every CM has, it estimates the region where own target object exists with high possibility for "re-recognition". CO doesn't exist before starting of the recognition. It is generated by CM, every time a new object candidate is found.

4.1 Flow of Recognition

The processing flow among all modules is message-driven. We describe the detail flow of messages and recognition requests in the case shown in Figure 3.

(a) Initial Recognition At first, CM sends RM **initial recognition request**, and RM initiates *initial recognition* for the whole region of the highest level image with minimum resolution (Figure 3(1)). If no candidate is detected, RM executes *initial recognition* for one more lower level image again.

(b) Generating Object Candidates When a new candidate is detected at RM, its information is sent to CM (2). After checking supporting relation, CM broadcasts it and generates CO (3)(4).

(c) Receiving Information of Candidates If another CM receiving information of a new candidate founds conflict, the CM informs occurrence of conflict to CO concerned with conflict (5).

(d) Conflict Resolution COs concerned with the conflict carry out conflict resolution by comparing each confidence value of a candidate and relation (6). The winning CO remains, and the other CO sends CM **modification request** (7)(8). By *modification request* RM re-recognizes the region of the losing candidate in the proper level image to modify its own region lest conflict occurs. If modification fails, the candidate is canceled. Details of conflict resolution and modification is described in [4].

(e) Renewal of Object Candidates If a new generated candidate doesn't make conflict, **renewal request** is sent from CO to CM, and RM re-recognizes the region of the candidate in the proper level image (9)(10).

(f) Estimation of Candidate Region base on Supporting Relation If CM receives information of a new candidate without supporting relation with any other candidates, CM sends RM **supporting request** and RM searches a supportable candidate object for the unsupported candidate in the region under the candidate in the proper level image (11).

If CM has relational knowledge related to a new candidate sent from other CM, CM sends RM **to-be-supported request** and RM examine supportable region of the new candidate object to detect its own new candidate.

(g) Re-recognition for Vacant Region If all modules of all the agents are in the state of waiting for a message and there is no message on communication lines, the system enters the final recognition stage. In the final recognition, if there are regions where no candidates detected in an image, **Recognition request for vacant regions** is sent (12). Re-recognition for these regions is carried out.

After the final recognition, the whole recognition of the system completes.

There are these six kinds of recognition request as stated above. One is initial recognition request, and others belong to re-recognition request. When each request is issued, in according to size of region to be recognized CM selects proper resolution of the image, that is, proper level.

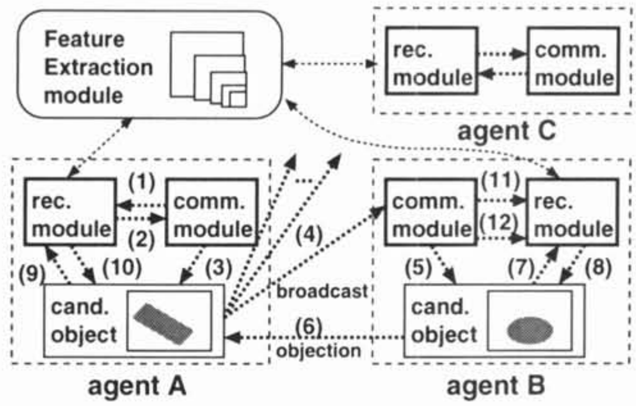

Figure 3: System structure and flow of messages. (1) **initial recognition request** (2) information of an new object candidate (3) generation of a candidate object (4) broadcasting information of a new candidate (5) notifying a new candidate of another agent (6) objection message (7) **modification request** (8) information of a modified candidate (9) **renewal request** (10) information of a renewed candidate (11) **supporting request** or **to-be-supported request** (12) **recognition request for vacant regions**

5 Recognition Results

We have implemented an experimental system for relatively simple indoor images with eight agents ("desk", "chair", "wall", "floor", "book", "cup", "pen", and "work station (WS)") on PC cluster system that consists of eight PCs (Intel Celeron 450MHz) using the PVM library. In this system, each agent is implemented on each one PC.

A sample indoor image (1280x960) in Figure 4 includes a "book", a "pen" and a "WS" are on a "desk". In the experiment by single-resolution recognition for the reduced image (320x240), only a "desk", "floor" and a "WS" were recognized correctly, but a "pen" and a "book" were not recognized. On the other hand, in multi-resolution analysis we used an image pyramid consisting 5 level images from level 0 to 4. Reduction ratio was 0.7, and size of the maximum-level image (level 4) was 308x231. In the experiment, at first in initial recognition for level-4 image a "desk", "floor" and a "WS" candidates were generated. Next, by relational knowledges of `on(book,desk)` and `on(pen,desk)`, "book" agent and "pen" agent initiated re-recognition for the region of "desk" and its peripheral region in level-3 image by *to-be-supported request*, and generated a "book" and "pen" candidates, respectively (Figure 6). After that, both candidates were re-

Figure 4: Indoor sample image.

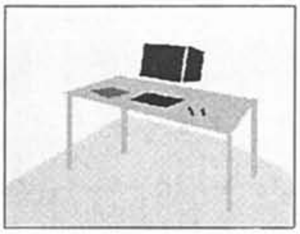

Figure 5: Recognition result using only multi-resolution images.

Figure 6: The region of the desk-top.

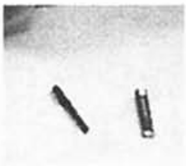

Figure 7: The region of the pen candidate.

Figure 8: Complex indoor sample image.

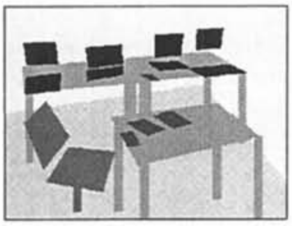

Figure 9: Recognition result for Figure 8.

recognized in level-0 image by *renewal request* (Figure 7). Finally, we got a result shown in Figure 5. While this process, conflict between a keyboard part of a WS and a book occurred, and some parts of a bookshelf of left hand were recognized as a book and a pen candidate. However, they were canceled finally by conflict resolution and checking "supporting relation".

We show execution times in case of single-resolution for three sizes of the image and multi-resolution for the sample image (Figure 4, 1280x960) in Table 1. With 640x480 image all objects were detected, but the execution time was about five times as long as one with 320x240 image. The system employing multi-resolution analysis could recognize all objects, and its execution time was only about twice as long as one with 320x240 image.

We show another sample of an indoor image in Figure 8 and its recognition result in Figure 9. In the experiment with 320x240 image the system couldn't recognize four WSs on the back desks, but such a complex image could be recognized by using 5 level images.

Table 2 shows results for 20 images including various indoor images from a simple image like sample image no.1 (Figure 4) to a relatively complex image like sample image no.2 (Figure 4). In single-

Table 1: Execution time

resolution	time[sec.]
320x240	8.3
640x480	38.3
1280x960	× (out of memory error)
multi-resolution	19.6

Table 2: Results for 20 images.

	almost correct	half correct	almost incorrect
single-resolution(320x240)	5	8	7
multi-resolution(5 level)	12	3	5

resolution recognition only 5 images were interpreted correctly, but in multi-resolution recognition 12 images were interpreted correctly. On the other hand, both numbers of almost incorrect results were almost same, because these images were too complex to extract significant image features in the initial recognition stage and efficient re-recognition could-n't be initiated.

6 Conclusion

In this paper, we applied multi-resolution analysis to multi-agent-based image recognition system we proposed and realized the system that can use a high-resolution image effectively without much extra processing time.

For future work, we plan to construct a recognition module with higher ability and more effective cooperation mechanism to make up for shortage of ability of individual recognition modules.

Acknowledgments

A part of this work was supported by a grant from the Inamori Foundation.

References

[1] Z. Li and L. Uhr. Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down process. *IEEE Trans. on System, Man and Cybernetics*, 17(2):250–263, 1987.

[2] C. L. Tan and W. N. Martin. An analysis of a distributed multiresolution vision system. *Pattern Recognition*, 22(3):257–265, 1989.

[3] K. Yanai and K. Deguchi. An architecture of object recognition system for various images based on multi-agent. In *Proc. of 14th International Conference of Pattern Recognition*, volume 1, pages 278–281, 1998.

[4] K. Yanai and K. Deguchi. Recognition of indoor images employing qualitative model fitting and supporting relation between objects. In *Proc. of 15th International Conference of Pattern Recognition*, volume 1, pages 964–967, 2000.