An Architecture of Object Recognition System for Various Images Based on Multi-Agent

Keiji Yanai
Department of Computer Science,
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, JAPAN
yanai@cs.uec.ac.jp

Koichiro Deguchi
Department of Mathematical Engineering and
Information Physics, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN
deguchi@meip7.t.u-tokyo.ac.jp

Abstract

An image understanding system for real world images which has an ability to recognize various kinds of images is proposed. We propose a multi-agent architecture to integrate object recognition modules for individual target objects. In our method, recognized results by different agents are fused not only on the evaluations by each modules themselves but also on relations of object locations, sizes and so on. This is carried out autonomously between the agents concerned, and the most reliable result is selected after the arbitration between them. We implemented an experimental system on a parallel computer, and achieved recognition for both indoor and outdoor images.

1. Introduction

An image understanding system for the real world must be able to recognize various kinds of images. We have been developing the recognition system that recognizes artificial and natural objects in a single image[4]. Here, "recognition" means to obtain a category name of an object, for example "desk" or "chair". To achieve this image understanding, we have to utilize not only shapes of objects but also relations between them as a recognition clue.

To realize such recognition we propose an architecture of an assembly of agents. In our architecture each agent consists of a recognition module to recognize single kind of object and a communication module to communicate with other agents.

We describe the system architecture and experimental results on the system implemented on highly parallel computer AP1000+.

The conventional image understanding systems based on multi-agent architecture aimed at integration of multiple algorithms[2] or flexible use of relations among objects[1]. They needed consideration of interaction among agents one by one and complicated description of relations among objects. Therefore they usually restricted the target image. To construct such general recognition system, their architecture is not effective. In our architecture, we only prepare recognition modules for each single kind of object independently. Adding cooperation mechanism, that is communication module, to integrate them, we can implement the system easily.

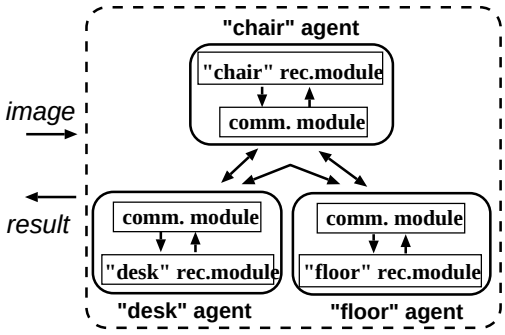

Figure 1. Basic structure of the proposed system. It is constructed as assembly of agents, and each agent consists of a recognition module and a communication module.

2. Basic concept of our system

In our architecture the system is constructed as an assembly of agents that recognized objects from an image independently. By adding agents it enables to recognize different kinds of images. In this architecture one agent consists of a recognition module and a communication module. Each recognition module recognizes only one kind of object. Each communication module communicates with other agents using relational knowledge between objects (Fig.1).

Each agent runs in order to recognize as many objects as possible in the image. Then sometimes conflicts occur among agents, that is, two or more agents find different objects in the same region on the image. Then they negotiate with each other to resolve the conflict.

Each agent also makes use of the information of object candidates generated by other agent. For example, when "a car" is recognized, there possibly exists "road" under it. Such cooperative process is managed by communication modules, while recognition modules deal with input images directly.

3. Recognition module and Communication module

A **recognition module** recognizes target objects as regions in an input image, and generates an object candidate. Our architecture do not assume special implementation for recognition module.

In principle the recognition module recognize objects by extracting common features to various view points and various instances from an image. For example, if there are a parallelogram and four vertical lines below it in the image, they are regarded as "a desk".

When recognition module finds regions of object candidates, it scores evaluation value on it, then, sends the information about the region and the evaluation value to the communication module.

A **communication module** carries out cooperations among agents. It keeps consistency to other agents to check candidates generated by the recognition module. If a conflict between agents occurs, it negotiates with the communication module of agents concerning the conflict.

Every communication module knows the candidates of other objects by exchanging the informations with other modules. If it has knowledge about relation between the candidates of other objects and its own object, it sometimes picks up recognition clues to recognize its target objects. Then, it sends them to the recognition module. "Relational knowledge" is a description about relation generally expected between two object. A communication module has only knowledge related to own target objects.

4. Recognition flow

The processing flow of all the modules is message-driven. Each module waits for a message from other modules. If it receives a message, it responds the message. We describe detail flow of messages in case of the example of Fig.2.

(1) An input image is sent to the recognition modules of all the agents. Then each recognition module starts recognition by the request from the communication module.

(2) Every time a recognition module finds region, it sends the information of the region and its evaluation value of the object candidate to the communication module.

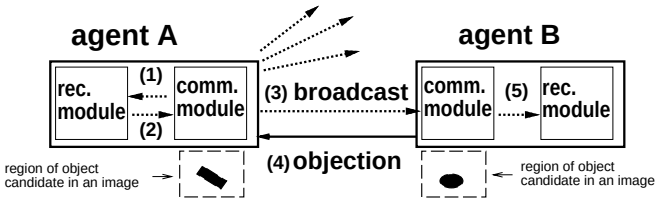

Figure 2. Flow of messages. (1) Initial recognition request. (2) Information of an object candidate. (3) Broadcasting information of an object candidate. (4) Objection message. (5) Conditional recognition request.

(3) The communication module broadcasts it for all other agents.

(4) Other agents examine if the broadcast is consistent with own object candidates. If not, the agent sends back an objection message. Then, conflict resolution is carried out between the communication modules concerned.

(5) When some relations exist between the broadcasted object candidate and an object of the receiving agent, the communication module of the receiving agent estimates region where it is expected for target objects or scale of object. The communication module sends "conditional recognition request" to the recognition module.

In addition to this, our architecture has "revival mechanism". It revives an object candidate that was canceled once, when evaluation about relation is changed and the result of comparison becomes invalid. By this mechanism the system always keep object candidates that are consistent with each other.

If the modules of all the agents are in the state of waiting for a message and there is no message on communication lines, the whole recognition of the system have completed.

5. Cooperation among agents

There are two kinds of cooperations among agents.

- Conflict resolution between object candidates
- Recognition using relational knowledge for other object candidates generated by the other agent

5.1. Relational knowledge

Each communication module has knowledges about relations generally expected between its own object and

Table 1. Examples of relational knowledge.

```
desk on floor
desk taller_than chair
ws on desk
book on desk
```

Table 2. evaluation value about shape for "chair".

value	detected parts
5	sitting surface, seat back, stable legs
4	sitting surface, stable legs
3	sitting surface, seat back
2	sitting surface with exact shape
1	sitting surface with rough shape

other objects about relative location and scale between two objects. They are called "relational knowledges".

Relational knowledges are represented by triplets of "source object's name", "relation name" and "destination object's name". We show some examples in Table 1.

5.2. Conflict resolution

"Conflict" means that there exists the region where two or more object candidates are assigned. When a conflict occurs, the concerning agents compare evaluation about object candidates and cancel one candidate. The evaluation for candidate is done by combination of evaluation value about shape and relation. In case of conflict between three or more objects, the agents resolve it between two of them.

If a compatible relation holds between two objects, one of the objects is regarded as hiding the other object. For example, if regions of "a desk" and "a book" are overlapping and a relation knowledge "`book on desk`" is applicable between them in the image, compatible relation holds. When compatible relation holds, no objective message is sent.

5.2.1. Evaluation about shape

Evaluation about shape is made for generated candidates by each recognition module. Evaluation value is score in five steps. The standard of the evaluation depends on each recognition module. We only define a rough standard for the value about shape.

We adopted such five step evaluation, because each recognition module has different standard of the evaluation and strict estimation is meaningless.

For example, we show the standard of the scoring for "a desk" in Table 2. In principle, the system evaluates the value by examining the kinds of detected component parts.

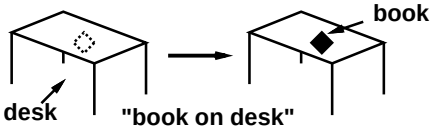

Figure 3. Recognition using relational knowledge. "Book" agent searches "book" candidates using information of "desk" candidate and relational knowledge of "`book on desk`".

5.2.2. Evaluation about relation

Each communication module makes evaluation about relation by examining the number of relations realized between its own and other objects. For example, when the communication module of "chair" agent has relational knowledge "`chair on floor`", it examines if the region of "floor" candidate exists under the region of its own candidates in the image. If "floor" exists under "chair" in the image, relation "`on`" is considered to be realized.

The evaluation about relation is estimated by checking the relational knowledge one by one. The value is the summation of weights of applicable relations.

5.2.3. Comparing conflicting candidates

Two conflicting candidates are compared by their evaluation values about shape at first. If difference between two values is more than certain value, smaller one is canceled. Otherwise, values about relation are compared, then smaller one is canceled. If both differences are small, the decision is made by comparing sizes of their regions.

5.3. Recognition using relational knowledge

Information on object candidates is broadcasted for reporting it to all the other agents. At this time by integrating received information on candidates and relational knowledge, the agents sometimes can expect the region where its own objects exist. For example, in Fig.3, "desk" agent generates an object candidate, and "book" agent gets the information. Using relational knowledge "`book on desk`" it estimates that a square region over "desk" region is "book".

The communication module of each agent examines if it has some relation to received object candidates as soon as it receives. If some useful information is found, communication module sends it "conditional recognition request" to recognition module.

6. Implementation and experiment

We have implemented an experimental system with 9 agents ("desk", "chair", "workstation", "wall",

Figure 4. Sample image 1 (indoor image).

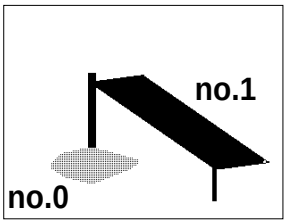

Figure 5. Candidates of "desks".

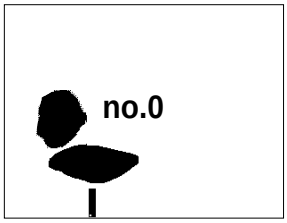

Figure 6. A candidate of "chair".

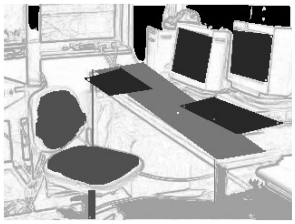

Figure 7. Recognition result (indoor image).

"floor", "road", "car", "sky" and "tree") on parallel computer Fujitsu AP1000+ with 16 processor elements. Each agent that consists of a recognition module and a communication module is implemented on one processor element.

Implementation of each recognition module is different from one to one. For an artificial object, we recognize by paying attention to the part that represents its function[3]. For example, sitting surface is paid attention for "chair" and desk surface is for "desk". For an object having no specific shape like "road" or "floor", we use conventional region segmentation methods.

In first experiment for an image of indoor scene shown in Fig.4 ten object candidates (two "desks" (Fig.5), a "chair" (Fig.6), two "workstations", "floor", "wall", "sky" and "tree") were generated by the respective agents. During this recognition conflict resolution occurred seven times, cancellation occurred five times, and revival occurred twice. Finally six objects of five kinds remained as shown in Fig.7. Total recognition time was 76 seconds.

A conflict occurred between desk candidate no.0 in Fig.5 and chair candidate no.0 in Fig.6. For chair no.0 a sitting surface, a seat back and a leg were found. So the evaluation value about shape was 5. On the other hand, for desk no.0 only a desk surface was found, so the evaluation value was 3. Therefore, desk no.0 was canceled. A conflict occurred between wall candidate and sky candidate, too. Both of the evaluation values about shape were 4, but the evaluation value about relation were 1 for wall because of "`wall higher_than floor`" and 0 for sky. So sky candidate was canceled. There was a region overlapping between desk candidate and workstation candidate. But, due to relation

Figure 8. Sample image 2 (outdoor image).

Figure 9. Recognition result (outdoor image).

knowledge "`ws on desk`", cancellation did not occur and they were regarded as compatible.

In this experiment, recognition using other object candidate and relational knowledge was realized. "Workstation" agent recognized a workstation candidate using the information of a desk candidate and relational knowledge "`ws on desk`". Thus, though the recognition module simply regarded parallelograms as a keyboard or a display, a workstation could be recognized.

In the next experiment for the image of outdoor scene shown in Fig.8 five candidates ("sky", "a car", "road", "tree" and "wall") were generated. Conflict between "wall" and "sky" occurred, but "wall" was canceled by comparing evaluation value about relation. Finally we obtained the result shown in Fig.9.

7. Conclusion and future study

In this paper we described the architecture of image understanding system based on multi-agent architecture. It is constructed as an assembly of agents for individual objects. Each agent has recognition and communication modules. The system recognizes both outdoor and indoor images by cooperation among agents using relational knowledge.

For future works, we will study how to evaluate object candidates, more effective cooperation mechanism and implementation of each recognition module.

References

[1] T. Matsuyama and Y. Ohta eds. Special issue on "parallel, distributed and cooperative image understanding". *Journal of Japanese Society for Artificial Intelligence*, 9(5):629–643, 1994. (in Japanese).

[2] Y. Ohta, M. Watanabe, and Y. Sumi. Approaches to parallel computer vision. *The Institute of Electronics, Information and Communication Engineers(IEICE) Trans.*, E74(2):417–426, 1991.

[3] L. Stark. Functionality in object recognition. *Computer Vision and Image Understanding*, 62(2):145–146, 1995.

[4] K. Yanai and K. Deguchi. Implementation of object recognition system employing multiagent architecture at highly parallel computer. In *Proceedings of the Sixth Parallel Computing Workshop*, pages P2–a, 1996.