'93 KACC (1993. 10. 20~22)

PC Networked Parallel Processing System for Figures and Letters

M. Kitazawa and Y. Sakai
Dept. of Mechanical Engineering, Yamaguchi University
Ube, Yamaguchi 755, JAPAN

Abstract

In understanding concepts, there are two aspects; image and language. The point discussed in this paper is things fundamental in finding proper relations between objects in a scene to represent the meaning of the that whole scene properly through experiencing in image and language. It is assumed that one of the objects in a scene has letters as objects inside its contour. As the present system can deal with both figures and letters in a scene, the above assumption makes it easy for the system to infer the context of a scene. Several personal computers on the LAN network are used and they process items in parallel.

1. INTRODUCTION

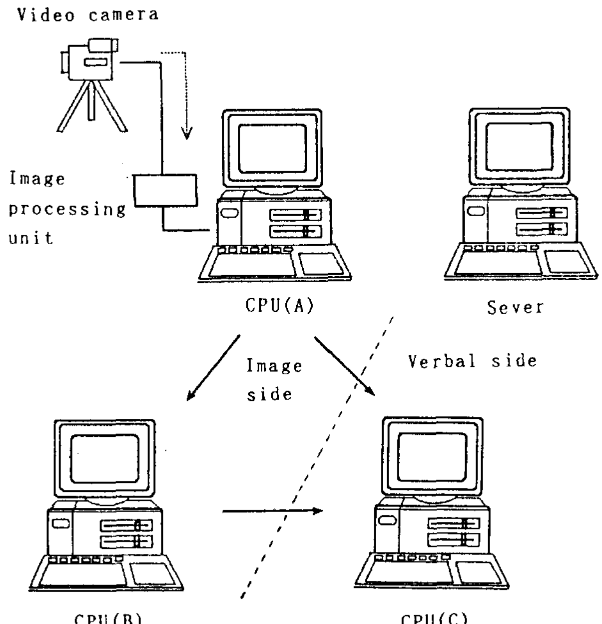
In the previous paper, it was described that the intelligent man-machine interface is essential in decreasing human's load, and that concepts are understood through experiencing on this interface [1, 2, 3]. The system proposed as a model for concept understanding in that paper identifies figures and letters from the imagerial aspect.[2] The system did not use verbal information[3] on the network. The figures identified in that system were not allowed to include objects inside their contours. The present system is so built as to be able to deal with these points. That is, objects processed in this system are figures which include objects and letters inside their contours. And this system also uses arc processing in order to deal with curved lines as part of figures.

Objects in a scene are identified through experiencing images and the proper relations between objects are extracted. From these relations and verbal information, the present system understands the meaning of the objects as a whole. To do so, this system consists of four personal computers as a LAN network and processes objects in parallel in the levels of image and language.

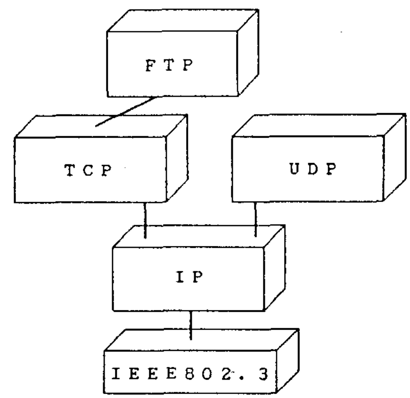
2. OUTLINE OF PARALLEL PROCESSING SYSTEM

Thin Wire Ethernet is employed for connecting four personal computers one of which serves as the server for the rest. Its basic system for concept understanding is described in Reference [2]. As was already mentioned, only the imagerial aspect of concept is dealt with in that system and inputted objects are identified one by one, from its own experience. Then these objects are dealt with in both levels of figures and letters in parallel. In the present system, any inputted object is processed separating it into its contour and objects inside that contour. Those inside objects are used in obtaining necessary information to identify the concept. By using two aspects of concept, image and language, relations between inputted objects are found and represented as a scene.

As shown in Fig.1a, a scene is inputted through video camera as visual stimulus. Applying a prescribed threshold to the image inputted, the



(a)



(b)

Fig.1 Schematic of the parallel processing system

image is transmitted into black and white image. The data is transferred to the other computer in order for two computers to be able to deal with the same data at a time. The contour of each of those objects is separated from its inside objects if any. Probable concepts are picked as a result of the image processing which will be described in the next section. Those concepts are sent to another computer for representing the meaning of the whole scene.

Fig.1b shows the schematic of the hierarchy of the protocols employed. The data processed is transferred between CPUs by using File Transfer Protocol(FTP).

3. LINEARIZATION AND CATEGORIZATION OF INPUTTED OBJECTS

Fig.2 shows the schematic of what kind of processing is made for an inputted object as a figure, in the level of image processing. The lines that touch the contour of object are regarded as part of the contour. So the objects inside the contour are only those which are independent of the contour. The contour and objects inside that contour are dealt with on separately on different computers. If the objects inside the contour also include other objects in their inside, such objects are memorized in data file temporally. After processing the former objects, the latter objects are dealt with. Thus a rough understanding of objects is performed and this process is to pick sets of probable concepts

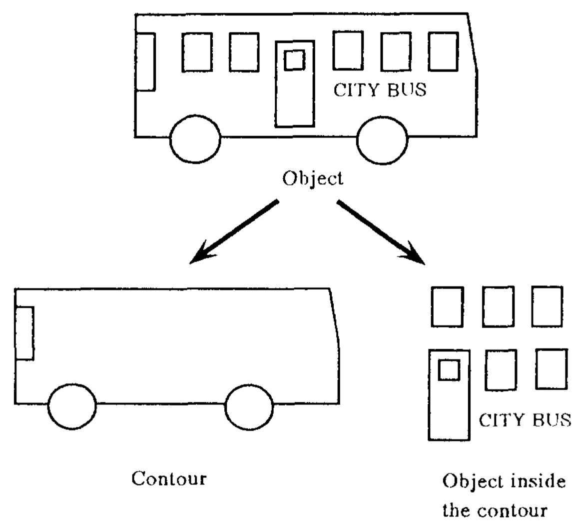
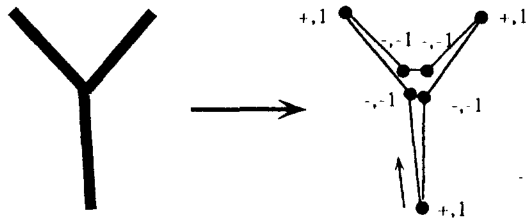


Fig.2 An example of separating inside objects from the contour



Convex-concave category +-+--+-
Acute-obtuse category 1-11-1-11-1

Fig.3 An example of categorization

for them. Objects' positions are also displayed on CRT for feature extraction.

3.1 FEATURE EXTRACTION FROM LINEARIZED FIGURES

An inputted object's shape is linearized employing the sectional linearization to extract its features[4]. After this process, an object is categorized by the ideas of convexity-concavity, acuteness-rectangle-obtuseness to represent the connection relation for every two successive line segments. Fig.3 shows an example of linearized figure with its category obtained. The symbols + and - denote convex and concave relations, respectively. In the acute-obtuse relation, 1, 0 and -1 are employed to indicate acuteness, rectangle and obtuseness, respectively.

3.2 FEATURE EXTRACTION FROM FIGURES INCLUDING ARCS AS PART

To extract features for an arc, the notion of difference in argument is employed here. Difference in argument is defined to be the angle

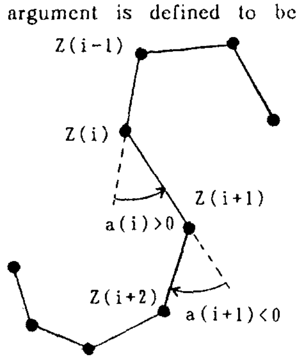


Fig.4 Definition of angle a(i)

between two successive line segments. Fig.4 shows difference in argument schematically. The angle between the line segments Z(i)-Z(i-1) and Z(i+1)-Z(i) is shown as a(i). This angle is restricted to the following range:

$$-\pi \leqq a(i) < \pi \qquad (\ i=0,1,2 \cdot\cdot\cdot\cdot,n-1\ )$$

Where n denotes the number of nodes; a(0) is the angle between the horizontal axis and the line segment Z(1)-Z(0). q(i) is defined as the total length of the line segments from Z(0) to Z(i). Therefore, the function $\theta(q)$ is defined as follows.

$$\theta(q)=a(0), \quad \text{if } q(0) \leqq q < q(1),$$
$$\theta(q)=\sum_{j=1}^{n-1} a(j), \quad \text{if } q(n-1) \leqq q < q(n).$$

An example is shown in Fig.5. $\theta(q)$ and q are measured at every node in order to obtain the graph of $\theta(q)$ with q at the bottom of Fig.5. The sign of a(i) at every node (i=0,1,..,5) corresponds to the convex-concave category of the figure in Fig.5. If the length of every line segment is identical for some successive line segments and the difference in argument a(i) for each of those is equal to each other, then $\theta(q)$ of the correponding nodes lie on a straight line. By using this graph, the feature extraction from figures will be described below.

An example of letter "2" is shown in Fig.6.

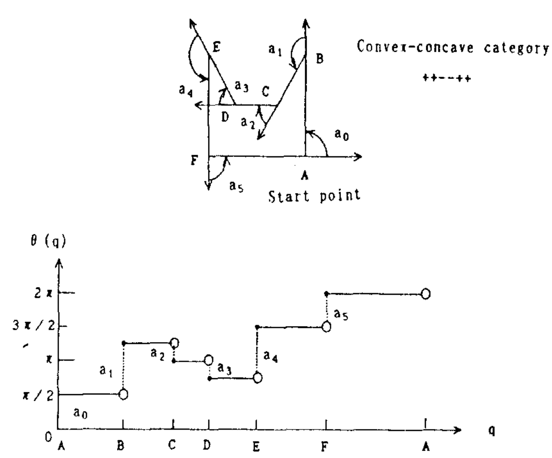


Fig.5 $\theta(q)$ of a figure

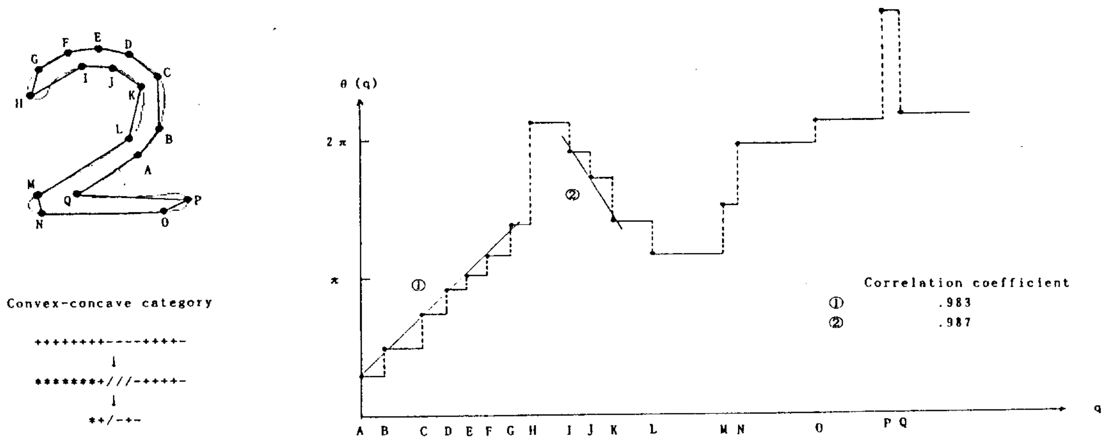


Fig.6 An example of letter "2"

After the letter is linearized, θ(q) and q are measured. The graph of θ(q) with q is also shown in the figure. Since inputted objects are drawn by hand, it is not always easy to distinguish a straight line from an arc. Then this system evaluates the correlation coefficient for the straight line obtained by using least squares. If this value is over 0.98 which is determined according to experimental results, this curved line is regarded as an arc. This method is applied for more than three successive convex nodes or for more than three successive concave nodes. In the case of "2" in Fig.6, the inclination of that straight line obtained by using least squares takes positive value between A and G. Between I and K, it is negative. The sign of that inclination expresses attributes. A positive one means convex-arc and a negative one means concave-arc. Whether a curved line is convex or concave is determined viewing from inside the object. By applying this function θ (q) to a curved line, figures with arcs are categorized. The symbols * and / denote convex-arc and concave-arc relations, respectively. Thus the category of the example "2" is represented as "*+/-+-."

An arc can be distinguished from any polygon by using the following facts. In case of a regular triangle, or a square or, in general, a regular polygon, the correlation coefficient for each of them is exactly 1. In case of an arc, however, it takes the value from 0.98 to less than 1. There is a small possibility that the above value takes 1. But even in such a case, it is easy to differentiate a regular triangle or a square from an arc. The reason is as follows. That is, for an arc, the connection angle at every node is obtuse, while for the cases of regular triangle and square, it is acute or rectangle, respectively. Next, in the case in which the object is identified to be a general polygon, the correlation coefficient takes on the value in the same range as the case of an arc. Hence the figure may be an arc instead. For such objects, increasing the closeness of approximation, linearization operation is iterately applied. If the number of nodes increases with the number of iterated trials of linearization, then it is inferred to be an arc.

4. PERIPHERAL PROCEDURES

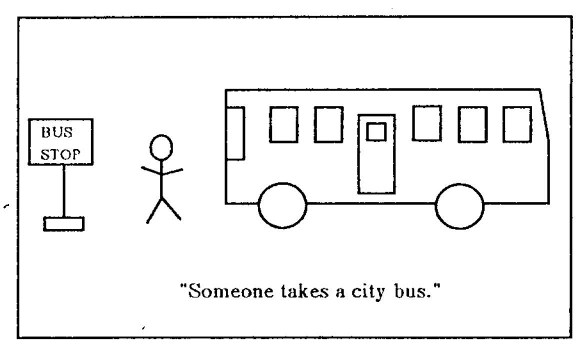


Fig.7 An example of scene

There are some other procedures necessary in interpretation of a scene. One is verbal analysis including parsing. In the present system, an experienced sentence is analyzed and stored in the memory, forming some verbal structures by each grammatical case for predicate representation. The part of speech called "joshi," in Japanese is crucial in parsing. For "joshi," there are the following ones: "wo," "de," "ga," "ni," "no," "ha," "to" and "he." The form which words are stored in has to be so built as to be able to compare between sentences of different structure and also to reproduce its original form as a natural sentence. An experienced sentence is processed by the above methods and is memorized as the expression form [5] for a scene. This form is shown below.

verb(n(1),n(2),n(3),n(4),n(5),n(6),n(7),n(8))

n(1)=noun+"joshi:wo"	n(2)=noun+"joshi:de"
n(3)=noun+"joshi:ga"	n(4)=noun+"joshi:ni"
n(5)=noun+"joshi:no"	n(6)=noun+"joshi:ha"
n(7)=noun+"joshi:to"	n(8)=noun+"joshi:he"

"ni" includes "niha." "de" includes "deha" and "nioite."

If there is any absent "bunsetu" in this form, "*" is placed instead at its position. An example is shown in Fig.7. In Fig.7, it seems that a bus is approaching to a bus stop. The expression form and the scene file for Fig.7 are shown below.

The expression form :

take(*,*, a person, a city bus,*,*,*,*)

Here is employed another procedure surrounding the process of the interpretation of a scene; the scene file. The scene file is formed by various relations which maintains between objects in a scene in the imagerial level.The expression form and the scene file help each other to derive the meaning of the scene.

The scene file :

a bus stop($xmax_1$,$ymax_1$),($xmin_1$,$ymin_1$)
attributes : "BUS STOP"
a human($xmax_2$,$ymax_2$),($xmin_2$,$ymin_2$)
a city bus($xmax_3$,$ymax_3$),($xmin_3$,$ymin_3$)
attributes : window, door

Sentences are experienced in advance to acquire concept's structures. Concept's structures described in this paper incorporates a human-like way in which attention is paid to a particular portion of the whole. This is called localization [6]. Concept's structures are a set of local hierarchy. Attributes are extracted from inputted sentences by using verbal analysis mentioned above.

DISCUSSION AND CONCLUSION

In Fig.7, an application of the methods proposed here in the present paper is schematically shown. The details in relation to Fig.7 can be found in Reference[5]. As is already described, an object's contour is separated from its inside objects, if any. Their features are extracted by the linearization. From their features and the letters of "BUS STOP," the objects are identified

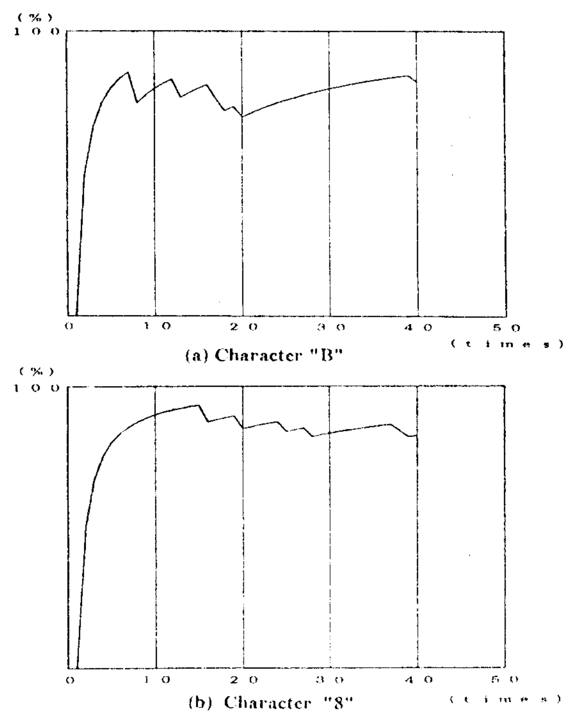


Fig.8 The learning processes for "B" and "8," as confusing characters. The decrease in cognition rate in the course of learning is due to an unlearned category. Only the categories in the simplified level is applied ic recognition.

to be a bus stop, a man and a bus, respectively. A context provides additional information in understanding the object which is identified to be a bus. The context of the scene can be derived from the letters "BUS STOP." From these letters, the bus is understood as a city bus, and not any other kind of bus. As is already mentioned, in order for the system to do this, characters must be recognized as well as figures in a scene in which both figures and letters are depicted. So the system tries to identify objects in the scene, assuming both possibilities by utilizing the function of parallel processing available on the LAN. The present system processes characters in the same manner as figures within a unique hierarchical structure of categorizing experienced objects. This fact is worth noting. In recognizing the word "BUS STOP," "B" is the most complicated letter. Any object in a scene is identified by searching a category to which its features belong[2]. Objects such as "B" and "8" are difficult to distinguish from each other without using the arc detection.

As already mentioned, the present system has the mechanism of learning through experiencing. The leaning process for recognizing "B" and "8" are shown, in the form of progress in recognition rate with number of learning times, in Figs.8a and 8b. The successful recognitions are around 85% in both cases. This is due to the fact that any special processing is considered to distinguish from figures in recognizing characters, in the process shown in Figs.8a and 8b.

Several advantages of the present system are as follows: Considering an object's contour and its inside objects is effective in identifying it. By using the notions of expression form and scene file, the present system derives the meaning of the objects as a whole. Application of personal computer network to a scene understanding system is effective in the parallel processing from imagerial and verbal viewpoints.

REFERENCES

[1] Y.Sakai, "Inference and Estimation Using Experience-Based Knowledge," '92KACC, Seoul, Korea, Oct.19-21,1992, pp.636-641.

[2] Y.Sakai, and M.Kitazawa, "Personal Computer Networking for Concept Understanding," ibid., 1992, pp.624-629.

[3] Y.Sakai, and K.Nomura, "The Verbal Aspect of Concept Understanding," ibid., 1992, pp.630-635.

[4] Y.Sakai, and T.Murahashi, "Experience Based Realization of Conceptual Functions for Man-Machine Interfaces," Proc. of IECON'91, Kobe, Japan, Oct.28-Nov.1, 1991, pp.1605-1610.

[5] M.Kitazawa, and Y.Sakai, "Intelligent Scene Recognition by Personal Computer Networking," to be published in Proc. of 2nd IEEE International Workshop on Robot and Human Communication, Tokyo, Japan, Nov.3-5, 1993.

[6] Y.Sakai, and H.Noyama, "Imagery for Man-Machine Interfaces Applying the Notions of Localization and Experience Sequence," Proc. of Int. Conf. on Manufacturing Systems and Environment, Tokyo, Japan, May 28-June 1, 1990, pp.325-330.