

TOWARD GENERATION OF 3-DIMENSIONAL MODELS OF OBJECTS USING 2-DIMENSIONAL FIGURES AND EXPLANATIONS IN LANGUAGE

Norihiro Abe, Fumihide Itho and Saburo Tsuji

Department of Control Engineering
Faculty of Engineering Science
Osaka University
Toyonaka Osaka Japan

ABSTRACT

We propose a system that can build a 3-dimensional model of an object with cylinder-like-body, given several 2-dimensional views and explanations on its substructures. The goal of our research is to build a qualitative model on relations between the object parts rather than a precise 3-D geometrical model. The system needs to match linguistic expressions to representation obtained from images. Also finding of correspondence between different views is important. Sometimes, it is difficult to get a complete model from a limited number of observations and explanations, the system can ask the user ambiguous points in terms of labels attached to the subparts. Several difficulties found in experimentation are discussed.

INTRODUCTION

There is the close relation between recognition of objects and learning of them. We human beings, build the object model by seeing them. There have been many researches that aim at autonomous building of 2 or 3-dimensional models [1-3], but most of them are either unnatural or difficult to use without much knowledge on their representation. In order to deal with a large set of objects, much effort of the user is needed. Oshima and Shirai [4] proposed a model learning system which uses multiple images observed from different positions. But image features are not always sufficient to get a reasonable model because the system is difficult to know which portions of an image are essential to the model creation.

Another way for a computer learn a model of an object is to understand the structure of the object given in natural language [5]. We have built an experimental system [6,7] which builds a model of a new object using an object model with similar features. This method was found useful to learning a 2-dimensional object without any rotational movements. The method was applied to a plot understanding system [8] using pictorial and linguistic knowledges. But we cannot apply it to imagery in which rotational movements often occur. For these reasons, a new method is proposed that uses verbal explanations along with several figures of an object. Explanation makes easier for the system to judge what portions of the given figures are important to a model building because named portions are often important parts. This also serves to resolve uncertainty by referring with a name to a known part of the object. However with only explanation, it is quite difficult to specify detailed relations and shape of the object. Figures can contribute to solve this problem since image analysis techniques have extracted them.

SYSTEM ORGANIZATION

Fig.1 shows the organization of this system. It consists mainly of three parts: language interpreter,

image-language communication part and model generator. Image-language communication part extracts significant features like a region or branch in order to find identities. This feature extraction is done by the same program as used in [6-8]. The result of the program are converted into frame structure as that of the linguistic interpreter is a same form. Model generator merges two frame representations into one frame which specifies the detailed model of an object. For the first image and explanation, one simple hypothetic model is created, and then it is modified by giving additional images and explanations. There are many difficult problems here. First, there may be some ambiguities in correspondence between features and their names. Second is difficulty in determining 3-dimensional structure. It is difficult to decide if a pattern in image is of 2-dimension (a kind of pattern on a surface) or 3-dimension (a protrusive structure). To know the precise shape of such a structure needs several observations from different view points. Then this system, mentioned as previously, makes a few hypotheses on a shape of an object body and its subparts; as a first approximation, it assumes that the object consists of a cylinder-like-body, and that any subfigures in the object are patterns on its surface except substructures clearly known as protrusive from the body. Of course these assumptions will often fail in building the true model because several patterns may be really protrusive structures. These wrong assumption cannot be revised until at least one another view is given.

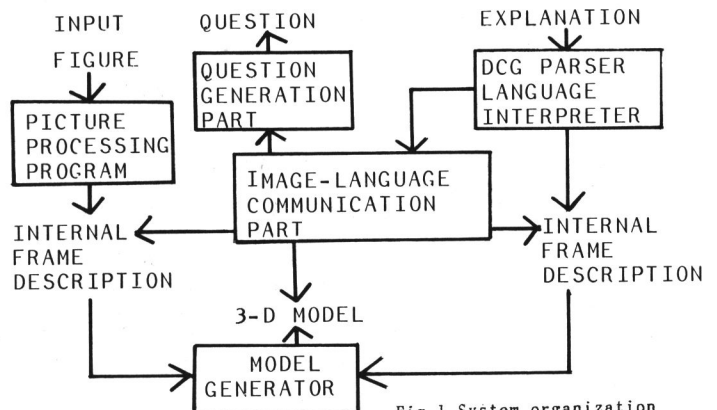


Fig.1 System organization

But this cannot be easily solved even if several other views are shown, as it is difficult to find true correspondences among parts in different views of the object. Conventional methods used in computer graphics have solved this problem by using precise 3-dimensional geometrical descriptions. However such an model building method is too rigorous to make the system unavailable. That is, the methods lack in flexibility. The description method used in this system just avails of relations among subparts with their names and various properties. The assumption that a main body is cylinder-like is just used only for inferring a subpart location when another observation of it is given from a different view point. Since this inference is based on a very coarse modeling, the predicted map of subparts locations may often be too vague to find unique correspondence among them. It is an advantage of the system that it has a possibility in deciding the correspondence by using various parts attributes together with their predicted relations. When some parts are left undecidable or unknown after all input figures have been shown, the system asks the user what they are. This man-machine communication method will be abbreviated in this paper.

An input sentence is converted into a tree structure. Next the word groups for shapes, colors and positional relations are extracted from the syntactic tree, and the system regards these noun groups as labels for some substructures of an object. The internal expression are generated based on the analyzed parsing tree.

LANGUAGE-IMAGE COMMUNICATION PART

The internal representation obtained by the image processing programs is also of a frame style. Since two frame representations, linguistic and pictorial one, describe the same object, they should be completely matched each other, but differences in several points cause matching to fail in finding correspondence of them. A unique correspondence is not always expected even if consistent pairings of slot values are got through so called graph matching method. There have been studied many matching methods for inductive learning[9]. Our matching method for mating two descriptions is different from them in the points that it at first checks just attributes to look for a clue to get good correspondences, then excludes contradictory combinations by the test of consistence in locational constraint.

KNOWLEDGE AND REPRESENTATION IN FIGURES

This system at first has no priori knowledge on a structure of an object to be modeled. Any generic model of objects is given, then features and structures of the object are not obtained until they are taught through learning. For example, see Fig.2 and 4. Given this figure, all regions such as B1, RE1, RA1, LP1, and all relations between them such that RP1 is included in RE1 and that RE1 is to the left of LP1 (from an observer) are extracted. Positional relations are described quite coarsely as shown in the figure; locations of subparts are described in terms of subregion denoted using U, M, D, L, C, R symbols. For example RE1 is located at LU of B1, and RP1 at RC of RE1. Here note that the descriptions of locations are different from that in explanations, where the relations are expressed relatively like "A is on the top of B". A translation from an observer centered framework into an object centered one is necessary to find correspondence between two representations.

MATHCING BETWEEN THE LINGUISTIC AND PICTORIAL FRAME

Let assume that six sentences in Fig.3 are given along with the figure. Each sentence is analyzed and converted into an internal form, which is then stored in the frame structures. Note that several parts not do not appear, which may be extracted from figures, or stated later.

Now let consider how we find correspondence between a linguistic and image expression. Given the sentences shown in Fig.3, the matcher can get complete pairs between labels and parts. If some of sentences are deleted from the explanation, ambiguities on correspondence between labels and parts are left. For example, if sentence (6) is deleted, then two possibilities are left concerning ARMs. When sentence (5) is abbreviated then there are left 6 ambiguities (note that sentence (5) is told). As we have many knowledges on shapes or properties of objects, we are apt to forget to state familiar portions to us, but it causes computer to fall in serious confusion. When no unique labelings are attained, the system retains all of them and tries to exclude contradictory correspondences by checking the rest of images and explanations. For example, the

ambiguity can be resolved on ARMs in the first case, if the correspondence between LA2 and LA1 in Fig.4 and 5 is found by the graph matcher and an explanation on arms is given concerning the figure 5. If nothing is elucidated to the figure, the ambiguity is left after the end of learning process unless it is resolved anywhere in learning. In such a case the system cannot help asking the user such ambiguous points.

GRAPH MATCHING BETWEEN CONSECUTIVE FIGURES

Let consider a case two figures are given as shown in Fig.4 and 5. Besides labeling process, now assuming that each part has a label given in Fig.4, the system must find correspondence between them in order to make clear uncertain points concerning identification of relationship among subparts. Matching the description of Fig.4 against that of Fig.5 requires a partial matching method between them, because a change in a view point causes visible substructures to be occluded and new substructures not observed so far to appear. This partial match is at first tried by finding subparts with similar attributes in terms of color, shape and their substructures. General constraints on the graph match are listed in Fig.6. For an example, following correspondences are found at first stage ($\leftrightarrow$ means a strong correspondence, and $\rightleftharpoons$ means a weak one).

RE1 $\leftrightarrow$ LE2	LE1 $\leftrightarrow$ LE2
H1 $\leftrightarrow$ H2 (a)	H1 $\leftrightarrow$ H2 (b)
M1 $\leftrightarrow$ M2	M1 $\leftrightarrow$ M2

(1) DBAQ HAS LEFT-EYE, RIGHT-EYE, NOSE, AND MOUTH IN HIS BODY.

```
*EXIST(DBAQ LEFT-EYE)
*EXIST(DBAQ RIGHT-EYE)
*EXIST(DBAQ MOUTH)
*EXIST(DBAQ NOSE)
*EXIST(DBAQ BODY)
*PART-OF(LEFT-EYE BODY)
*PART-OF(RIGHT-EYE BODY)
*PART-OF(MOUTH BODY)
*PART-OF(NOSE BODY)
```

Fig.3 Explanations and their internal representations

(2) DBAQ HAS LEFT-ARM AND RIGHT-ARM BESIDE HIS BODY.

```
*EXIST(DBAQ LEFT-ARM)
*EXIST(DBAQ RIGHT-ARM)
*ATTACH(LEFT-ARM BODY)
*ATTACH(RIGHT-ARM BODY)
```

(3) DBAQ HAS HAIRS ON THE TOP OF HIS BODY.

```
*EXIST(DBAQ HAIRS)
*ATTACH(HAIRS BODY)
*RELATE(HAIRS ON BODY)
```

(4) DBAQ'S LEFT-EYE IS TO THE RIGHT OF HIS NOSE.

```
*RELATE(LEFT-EYE TO-RIGHT NOSE)
```

(5) DBAQ'S RIGHT-EYE IS TO THE LEFT OF HIS NOSE.

```
*RELATE(RIGHT-EYE TO-LEFT NOSE)
```

(6) DBAQ'S LEFT-ARM IS ON THE RIGHT OF HIS BODY.

```
*RELATE(LEFT-ARM ON-RIGHT BODY)
```

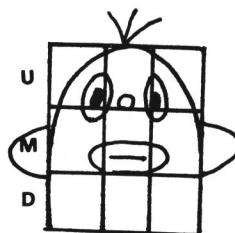


Fig.2 L C R

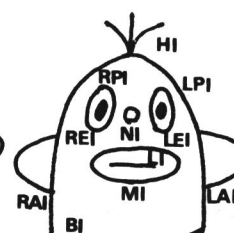


Fig.4 A front view

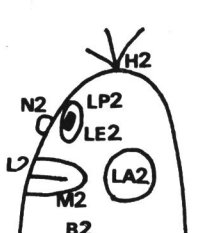


Fig.5 A side view

Where, there are correspondences with regard to subparts L1 L2, LP2 RP1, RP2, but these are used to support the above correspondence (a), (b). Concerning LA1, RA1, N1 and N2, LA2, as they are just one close region with the same color and no substructures, they cannot serve to restrict possibility of correspondence. Next using the above fact as a clue, relations toward the horizontal and vertical direction are checked to see if the object is turned upside down or not. If so, the relation toward right and left direction must be reversed when a matching is attempted for the rest of structures of the object. Note that the assumption that a body is cylinder-like is used here. In either case for (a) or (b), the object is not turned upside down, then it can be decided that horizontal relations among subparts must be preserved between these two figures. By the way when the first figure was given, whereas LA1, RA1, H1 were clearly recognized as projecting structures, other parts are hypothesized as patterns on B1. On the other hand in the second figure, just N2 and H2 are identified as protruding ones. A transformation of projecting structures on a surface between one and another view point must be taken into consideration so as to decide correspondence among the rest of parts of the object. In Fig.4, locations of H1, LA1 and RA1 are not clear, because this figure cannot specify the precise location of them. Then their location are assumed to be on the side or top of the body, which means the first 3D-relational model of the object that has been generated by the first figure and explanation. For this description, the description gained from Fig.5 is checked to see which parts are matched each other.

As mentioned previously there have been two possible correspondences, but preservation of horizontal relations excludes the case (a) because $RE1 \leftrightarrow LE2$, $H1 \leftrightarrow H2$, $M1 \leftrightarrow M2$ contradicts the assumption that the body is of cylinder-like structure. On the other hand the case (a) leads without contradiction to a conclusion that LA2 can be LA1 and N2 be N1.

Then the 3-dimensional structure of N1 is rewritten from pattern to projection and correctness of other structures is augmented as a consequence of this matching.

IMPLEMENTATION AND DISCUSSION

The examples shown above are real results of the implemented system. Although this system is not completely implemented, we have found many difficulties during the experimentation. In this research the image processing programs extract subregions independently from the rest of this system, then meaningless correspondences are generated at the first stage. The top down extraction mechanism concerning image processing is needed so as to reduce meaningless combinations between figures and parts names.

Since the examples used in this study are rather simple, then a reasonable conclusion is obtained. Especially, the examples lead to a model of the object without asking the user ambiguous points. We have also experimented examples with several ambiguities on both labeling process and matching of figures. It is however too difficult to elucidate the case including both of the ambiguities in a limited space. We have not treated a peculiar case where a region or line are observed as a line or point, respectively. And also we have not considered precise estimation of a shape transformation except for the relation between ellipses and circles.

Now, we have just experimented a few objects with a cylinder-like-body, consequently we must extend this method so as to apply it to more objects not modeled by the first approximation that a body is of the cylinder-like. A knowledge organization mechanism must be constructed in order to build concept of the learned objects and to use experience in learning of other objects. The implemented system has no such a function.

SAME-SHAPE; must not match a region against a branch.
 SAME-COL; parts with different colors should not be paired.
 SAME-INCL; for subparts A of B and C of D, when A does not matched to C even if B and D can be paired, they are not regarded as a pair.
 SAME-REL; if there is an extreme difference of location between the predicted one and that of the model, the prediction is not right.

Fig.6 Constraints on grap matching

REFERENCES

1. Marr,D.; Analysis of Occluding Contour, Proc. Roy. Soc.,B., Vol.197,No.1129,pp.441-475 (1981)
2. Brooks,R.,A.; Symbolic Reasoning Among 3-D Models and 2-D Images, Artificial Intelligence, pp.285-348 (1981)
3. Popplestone,R.J. Et al.; Forming modelos of plane-and-cylinder faceted bodies from light stripes, 4th-IJCAI, pp.664-668 (1975)
4. Oshima,M. And Shirai,Y.; Object Recognition using Three-dimensional Information. 7th-IJCAI, PP.601-606 (1981)
5. Asin,G.,J.; Hierarchical Representation of Three-Dimensional Objects Using Verbal Models, IEEE PAMI-3 No.2, pp.197-204 (1981)
6. Abe,N., Itoh,F. And Tsuji,S.; Toward learning of Object Models Using Analogical Object and Verbal Instruction., 2nd-AAAI, pp.362-366 (1982)
7. Abe,N. And Tsuji,S.; A Learning of Object structures by Verbalism., COLING-82, pp.1-6 (1982)
8. Abe,N., Soga, I and Tsuji,S.; A Plot Understanding System on Reference to both Image and Language., 7th-IJCAI, PP.77-84 (1981)
9. Dietterich,T.,G. And Michalski,R.,S.; Inductive Learning of Structural Descriptions, Artificial Intelligence, Vol.16, pp.257-294 (1982)
10. Winston,P.,H.; Learning New Principles from Precedents and Exercises, AIM 632 of MIT, pp.1-59