Learning correspondences between visual features and functional features

Hitoshi MATSUBARA*, Katsuhiko SAKAUE and Kazuhiko YAMAMOTO
Electrotechnical Laboratory
Tsukuba, Ibaraki
JAPAN 305
e-mail: matsubar@etl.go.jp

Abstract

We have implemented a visual learning system MIRACLE-IV, which is capable of obtaining an internal structure of an object from a series of silhouette images with no initial explicit models about the object[1, 2, 3]. The images are derived from only one object, but the forms of the object are varied. The system is composed of two sub-systems: a model-acquisition part (the modeler) and an image-processing strategy part (the strategist). On the assumption that the object consists of hinges, slides and solids, the modeler learns the number of them in the object and the relationship between them. The strategist binds the functional features as hinges or slides with visual features in the actual image data. The image-processing sequence for the extraction of the visual feature is not given previously but is learned automatically through trial and error. In our research, mutual references between pattern information and symbol description play essential roles for learning. This paper describes how MIRACLE-IV learns correspondence between functional features and visual features.

1 Introduction

In this paper, we present an experimental visual learning system, which is called MIRACLE-IV (Multiple Image Recognition system Aiming for Concept LEarning – Intelligent Vision)[1, 2, 3]. MIRACLE-IV is capable of obtaining an internal structure of an object from a series of silhouette images with no initial explicit models about the object. The images are derived from only one object, but the forms of the object are varied. The system is composed of two sub-systems: a model-acquisition part (the modeler) and an image-processing strategy part (the strategist). On the assumption that the object consists of hinges, slides and solids, the modeler learns the number of them in the object and the relationship between them. The strategist binds the functional features as hinges or slides with visual features in the actual image data. In MIRACLE-IV, object models and image-processing strategies are not fixed but generated and renewed in a self-producing way.

The attempt to combine pattern recognition and symbol manipulation is an important theme of artificial intelligence [4, 5, 6]. Our research is placed as one of the attempts, but discriminated from other studies at the following point. It is assumed in other researches that pattern information is completely transformed into symbol description at the primary stage. In our research, however, an original image is referred at any time when additive pattern information is necessary for symbol manipulation. Namely, mutual references between pattern information and symbol description is the unique feature of MIRACLE-IV. This paper describes how MIRACLE-IV learns correspondence between visual features and functional features.

2 MIRACLE-IV Outlined

MIRACLE-IV is a system that can automatically learn the structural description of an object and the strategy of image processing from the presented series of silhouette images of various forms of the same object. The series of original images is shown in Figure 1. The system estimates an internal structure of the
Figure 1: A series of original images

Figure 2: Correspondence between the visual feature and the functional feature

The modeler knows beforehand that an object consists of hinges, slides and solids and that variations of the form are two-dimensional. It learns afterwards the number of hinges, slides and solids and their mutual relations. On the other hand, the strategist knows beforehand only physical constraints on the connection among individual processing modules. It accumulates afterwards strategic knowledge of image-processing sequences in the computer by organizing successful examples that have happened to be obtained through trial and error.

As mentioned above, knowledge of the modeler is explicitly separated from that of the strategist, and vice versa. The modeler knows nothing about image-processing modules, and the strategist knows nothing about internal structures. The learning process of MIRACLE-IV is done through the interaction between the modeler and the strategist. Such a two-subsystem-organization of MIRACLE-IV contributes to generality of the system because the strategic knowledge accumulated in the strategist is independent of a structure of a specific object. While the modeler manipulates the symbol description, the strategist processes the pattern information, and, if necessary, they exchange their information properly. In MIRACLE-IV, applying actual image data to concept learning has an essential meaning.

The mechanism of the modeler depends strongly on what kind of objects MIRACLE-IV accepts. The objects MIRACLE-IV accepts are simple tools and implements such as compasses, scissors and rulers. Practical tools are generally easy to make correspondence between their visual features and their functional features because there are no useless decorations on them.

3 The Modeler

We have selected a pair of compasses as an example object. By handling it with our hands, we can change its form in various ways. We let MIRACLE-IV see some silhouette images of the various forms in an order. The order, which is important to learning, will be discussed later in this section. We tell MIRACLE-IV that all the images are derived from only one object, "compasses", but do not tell what compasses are. So MIRACLE-IV does not have any explicit models of compasses initially. Under such circum-
stances, the modeler acquires the internal structure of the object through the interaction with the strategist. The acquired model consists of the knowledge about the number of the hinges in the object and the relation between the hinges and so on.

It is often said that learning something new requires a lot of previous knowledge about the thing. In machine learning, a dividing line between the already known and the newly learned should be set appropriately. It is very difficult for the computer to learn an internal structure of the object whose form changes freely. So we have to restrict the degree of freedom of changes. We assume that the change is caused by only hinges and slides. Namely, it is assumed that the modeler knows that the object consists of three constituents: hinge, slide and solid. We call this knowledge HSS assumption. On the HSS assumption, the modeler learns from a series of silhouette images how many hinges, slides or solids the object has and how they are connected.

Besides the HSS assumption, the modeler utilizes the following heuristics.

- The number of solids in the object is small.
- The frequency of hinges is higher than that of slides.
- The variations of the form are two-dimensional.

The order of input images must be determined carefully by man for incremental learning. A form which is slightly different than the form that has been input just before must be input one by one. To put it concretely, a form which is different in only one position of a hinge or a slide than the form input just before will be input. If we input a form which is different in more than two positions, the learning process of the modeler does not work well in general.

The modeler receives a shape description, such as a set of line segments, from the strategist. The modeler has to decide the correspondence between one set of line segments and the other. As form and direction of the object are varied under the HSS assumption, a flexible matching method is required for the modeler.

We adopt an extended relaxation method to make the constraint between adjacent line segments propagated. This is an extension of the relaxation method for the recognition of hand-printed KANJI(Chinese) characters[7]. In the correspondence process, the modeler infers locations of hinges or slides. The inference is based on the information about two kinds of relaxation parameters: angle and length. In essence, the modeler imagines the existence of a hinge by the difference of angles, and the existence of a slide by the difference of lengths.

We have not analysed the capability of the extended relaxation method quantitatively, but it works well in most cases under the above conditions.

4 The Strategist

The aim of the strategist is to bind the functional features of the object which are obtained by the modeler with the visual features in the raw image data. The automatic trial-and-error process of the strategist makes it possible to generate an appropriate image-processing sequence, which consists of various kinds of image processing modules (such as binarization module, boundary extraction module).

Some rule-based systems also have an automatic trial-and-error function to get to a goal. Their knowledge-bases to control the process should be constructed and tuned beforehand. In our strategist, however, only given are data-type constraint rules for the connection of processing modules. It can construct a strategic database through the accumulation and systematization of fortunate success examples in many trials.

A hinge which is a typical functional feature of compasses often has a circular shape. In the initial stage, however, the strategist knows nothing about visual features of a hinge. In this case, the modeler estimates existence possibility and predicted location of the hinge from the given silhouette images, and requests the strategist to detect the hinge by means of visual features. The strategist repeats the trial and error of image processing until the predicted location is detected as a visual feature; success will cause it to learn an image processing strategy corresponding to the hinge.

The correspondence between functional and visual features is independent of the kind of an object; we can see circular hinges in many tools or implements. If reliable strategies are sufficiently acquired, the strategist will be able to infer an internal struc-
Example of Learning Process

This chapter describes the incremental model-acquisition process of MIRACLE-IV by following an example sequence of images. In this case, we show MIRACLE-IV the silhouette images of the same "compasses" in four different forms one by one.

1. The first set of line segments (see Figure 3(1)) given by the strategist is kept as an initial structural description of the object. No structural information is gained by only one image, so the modeler does nothing particular. The modeler knows that hinges can be bent, but does not know that bent parts can be hinges. The model of "compasses" at the time consists of one solid.

$$\text{SOLID-A1} = (\text{a1 a2 a3 a4})$$

2. When the second set is given (see Figure 3(2-a)), the modeler makes correspondences between the first and the second. As a result, the inference is made that the bent part is a hinge. The modeler changes the structural description on this inference, tells the strategist the existence of "hinge" and its approximate position, and requests the strategist to find the other hinges by the visual feature. The strategist repeats the generation of image-processing sequences until the hinge position is detected as a visual feature. In this case, a circular shape is a cue to a hinge. Figure 3(2-b) illustrates the reply from the strategist. Each circle specifies the location where the strategist found a cue shape. Some of the locations, however, have never been bent so far; they are not confirmed as hinges but reserved as hinge candidates. The modeler does not know the visual feature (in this case "circle") of the hinge.

$$\text{SOLID-B1} = (\text{b1 b2})$$
$$\text{SOLID-B2} = (\text{b3 b4})$$
$$\text{HINGE-B1} = (\text{SOLID-B1, SOLID-B2})$$

3. Correspondences are made between the third set (see Figure 3(3)) and the structural description updated in 2. The hinge candidates contribute to the effective relaxation matching because the angles at the hinge candidates can be regarded as changeable. As a result, existence of another hinge is confirmed, therefore the structural description is renewed. The model at this time consists of three solids and two hinges.

$$\text{SOLID-C1} = (\text{c1 c2})$$
$$\text{SOLID-C2} = (\text{c3 c4})$$
$$\text{SOLID-C3} = (\text{c5 c6})$$
$$\text{HINGE-C1} = (\text{SOLID-C1, SOLID-C2})$$
$$\text{HINGE-C2} = (\text{SOLID-C2, SOLID-C3})$$

Figure 3: An example of learning process
4. The fourth set (See Figure 3(4)) is processed in the way similar to 3. The final structural description is that the object has three hinges and four solids, and that each hinge connects two solids. Unconfirmed hinge candidates are discarded at this time.

\[
\begin{align*}
\text{SOLID-D1} &= (d_1 \ d_2) \\
\text{SOLID-D2} &= (d_3 \ d_4) \\
\text{SOLID-D3} &= (d_5 \ d_6) \\
\text{SOLID-D4} &= (d_7 \ d_8) \\
\text{HINGE-D1} &= (\text{SOLID-D1}, \text{SOLID-D2}) \\
\text{HINGE-D2} &= (\text{SOLID-D2}, \text{SOLID-D3}) \\
\text{HINGE-D3} &= (\text{SOLID-D3}, \text{SOLID-D4})
\end{align*}
\]

The final structural description is a model of the object "compasses". In other words, MIRACLE-IV has learned a concept of the object "compasses". MIRACLE-IV has learned the correspondence between the visual feature ("circle") and the functional feature ("hinge").

The order of input images (1),(2),(3),(4) (4),(3),(2),(1) are adequate for the learning, but (1), (4), (2), (3), for example, is not adequate.

6 MIRACLE-IV as a learning system

From the standpoint of machine learning, the features of MIRACLE-IV are the following.

1. (Incremental) Inductive inference from examples.
2. Use of only positive examples.
3. Analogical reasoning.
4. Mutual reference between pattern information and symbol description.

7 Concluding Remarks

This paper describes how MIRACLE-IV learns correspondence between visual features and functional features. The learning process is illustrated by using "compasses" as an example.

Of course, MIRACLE-IV does not always learn correct models. It may learn incorrect models by its incorrect belief. However, if we consider that a man may learn an incorrect model from a series of silhouette images by his incorrect belief, MIRACLE-IV is a good model for human learning.

In this paper, functional features refer to hinges, slides and solids. The true function of "compasses" as a tool, however, is to draw circles. It is a challenging research theme to make a system that learns the function of "compasses" from their visual features.

References


