

A MODEL FOR PERCEPTION OF STRUCTURAL IMAGE FEATURE

Hajime Enomoto

Naoki Yonezaki

Katsumi Nitta

Department of Computer Science
Tokyo Institute of Technology
Ookayama, Meguro-ku, Tokyo, Japan

A model to extract a structure feature of images belonging to one category is proposed, implemented and studied. This model consists of two parts i.e. an extraction of partial geometrical features and an extraction of structure feature by using them under the positional constraints. First, necessary conditions which geometrical features should satisfy is stated and a statistical method to select a key feature is described. Then the algorithm to extract structure feature as the maximal set of common subimages which satisfy the positional constraint is introduced. The structure feature is regarded as an intrinsic feature which is used when an animal perceives an image at a glance. The experiment was carried out about image of human faces and had a fairly good results.

1. INTRODUCTION

Perception process of images by computer is composed of extracting some features from images, describing them, and making the correspondence between objects and description. The process of correspondence needs much knowledge and takes much time[1]. The animal has an ability for efficient image perception which contains the fast algorithm to observe it initially and get the structure feature. It is obtained without knowledge and becomes the cue for understanding.

This paper proposes a model of perception along this fact and a learning schema to get the structure feature. The structure feature is a positional relation of local features which is essential to recognize objects. At first, necessary conditions for these features are discussed. Then, the problem of constructing the structure feature is treated. This method obtains the maximal matching of subimages which satisfy positional constraint. And finally, experiments for images of human faces show the efficiency of this schema.

2. FEATURE EXTRACTION

When an image is given, we can extract various features, such as edges, regions, and textures. If we have a model of objects, we do not have to get all these features. We need only the strong enough features which are appropriate as keys for recognition of images. But, in the learning schema, all features should be treated

equally without such a knowledge. The structure feature is constructed by the features that satisfy the following conditions.

- (a) They should not depend on the coordinate system and their position should not be affected by the operations that lose resolution.
- (b) They should be clustered into some categories and the relations between them should be considered more important than the absolute value.
- (c) They should include spatial informations. We divide an image into subimages, which are sets of adjacent pixels, and attach a feature vector $(X_1, X_2, \dots, X_N)$ to each subimage. Each component of this vector is the values on the feature operators.

These feature vectors are distributed spatially and if some relation between them is intrinsic for the image, we regard such relation as the structure feature[2].

3. SPATIAL MATCHING

3.1. Selection of Key Features and Labeling

Within all components of a feature vector, some components may be meaningless as the structure feature, and we must select key features. They are selected on a basis of statistical distribution of each feature of images which belong to the same kind.

(Selection Algorithm)

Step 1: Select one component X_i whose histograms show resemblance for all input images.

This resemblance is measured by the configuration of cores of clusters[3].

Step 2: If there are different numbers of clusters for X_i , some near clusters are merged.

Step 3: For each cluster of X_i , select the component X_j the distributions of which resemble each other for input images, and execute Step 2. This step is repeated until there does not exist the component that show the resemblance, and the sequence of components $(X_i, X_j, \dots)$ is obtained. Then, each subimage is labeled by an n-tuple $(Y_i, Y_j, \dots)$ where each component Y_k is the name of cluster of corresponding feature X_k to which the subimage belongs.

During these steps, the feature which shows peculiar property in a particular image is not selected, for the histograms of such features do not resemble those of other images, and it is considered as a meaningless feature for the structure feature which is common in images.

3.2 Positional Constraint

Here, we define the notion of "match" and "positional constraint".

If two subimages $S_1 \in I_1$, $S_2 \in I_2$, have the same labels, we say S_1 matches S_2 and it is denoted by $S_1 \approx S_2$.

If there are such subimages $S_1, S'_1 \in I_1$, $S_2, S'_2 \in I_2$ that satisfy the following conditions, we say they satisfy positional constraint and it is denoted by $(S_1, S'_1) \approx (S_2, S'_2)$. (In the following condition, $I(S), J(S)$ denote the row and the column index of S , respectively.)

- (a) $S_1 \approx S_2$, $S'_1 \approx S'_2$.
- (b) If $I(S_1) < I(S'_1)$, $I(S_1) = I(S'_1)$, $I(S_1) > I(S'_1)$, then $I(S_2) < I(S'_2)$, $I(S_2) = I(S'_2)$, $I(S_2) > I(S'_2)$ hold, respectively.
- (c) if $J(S_1) < J(S'_1)$, $J(S_1) = J(S'_1)$, $J(S_1) > J(S'_1)$, then $J(S_2) < J(S'_2)$, $J(S_2) = J(S'_2)$, $J(S_2) > J(S'_2)$ hold, respectively.

When two images I_1, I_2 are given, a set of matched subimages of I_1 and I_2 is a set $M = \{S_1, S_2\}$ where $\forall (S_1, S_2) \in M [S'_1 \in I_1, S'_2 \in I_2, S_1 \approx S_2]$ and $\forall (S_1, S_2), (S'_1, S'_2) \in M [(S_1, S'_1) \approx (S_2, S'_2)]$. A set of matched subimages expresses a common structure of images. We regard the maximal set of matched subimages as the structure feature. It takes much time to obtain the maximal set, so we introduce a simple matching algorithm by which a good set can be obtained, though not a maximal one.

3.3 Simple Matching Algorithm

Let an image be a set of subimages $\{S_{ij} | 1 \leq i \leq M,$

$1 \leq j \leq N\}$, and let each subimage S_{ij} have a label L_{ij} . For each row i , the sequence of labels $L_{i1}, L_{i2}, \dots, L_{iN}$ is denoted by $L\langle i \rangle$. When two sequences $L\langle i \rangle$ and $L\langle j \rangle$ are given, the distance between them is defined as the reciprocal of the length of the longest common subsequence (LCS)[4], and denoted as $D(i, j)$. The matching algorithm is as follows:

Step 1: Let P be a set of all pairs of sequences $(L_1\langle i \rangle, L_2\langle j \rangle)$ where L_1, L_2 are from I_1, I_2 , respectively.

Step 2: Select a pair $(L_1\langle i_1 \rangle, L_2\langle j_1 \rangle)$ whose distance $D(i, j)$ is the smallest in P .

Step 3: From P , remove the pairs $(L_1\langle i \rangle, L_2\langle j \rangle)$ that do not satisfy $i > i_1, j > j_1$ or $i < i_1, j < j_1$.

Step 4: If P is empty or there is not a pair $(L_1\langle i \rangle, L_2\langle j \rangle)$ which satisfies $D(i, j) < 0$, go to step 5. Else, go to step 2.

Step 5: Change the labels of the subimages to "don't care symbol" that are not included in the selected rows, or that are not included in the LCS in step 2.

Step 6: Repeat from step 1 to step 5 for any pair of column vectors. In this case the "don't care symbol" is not included in LCS.

The subimages which are left unattached "don't care symbol" construct a set of matched subimages.

4. EXPERIMENT

We apply our algorithm to images of human face. Human faces have some subobjects such as hair, eyes, brows, and so on. However, we try to match the images without the knowledge about them. Input images are composed of 128×128 pixels and each pixel has 9 bit integer value as intensity. (Fig. 1)



Fig.1 Original images

These images are divided into subimages whose sizes are 4x4, and each subimage is attached a feature vector (X_1, X_2, X_3) where X_1, X_2, X_3 correspond to mean intensity, busyness and edge value, respectively[5][6]. Each component is categorized and consequently each subimage has a label composed of 3-tuple of cluster names. In Fig.2, concerning the first two images, the boundaries between subimages attached different labels are displayed.

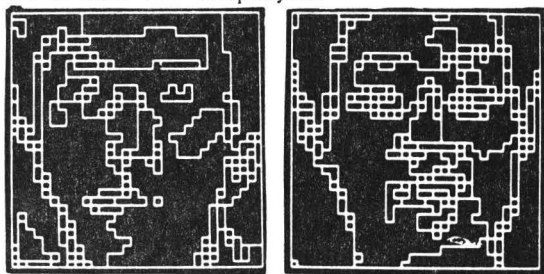


Fig.2 Labels attached to subimages

By comparing the histograms of these features, clusters of X_1 are merged, and new labels are attached to each subimage. (Fig. 3)

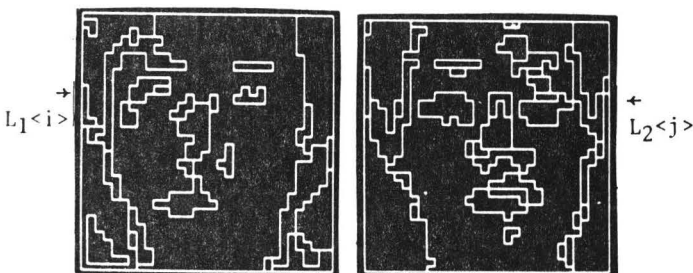


Fig.3 Merged labels

LCS of sequences of these labels is calculated by the simple matching algorithm. For example, let $L_1<i>, L_2<j>$ be given as following sequences. (They are sequences of labels which are indexed by arrows in Fig.3, and each letter expresses a label attached to a subimage.)

$L_1<i> = \text{baadbdbdbbbdbbddd bbbbaaaa}$

$L_2<j> = \text{baaaadbdbdddbbbdddbbbccdbddade}$

LCS of these sequences is calculated in Step 2 and following result is obtained. (Labels that are not included in LCS are changed into "don't care symbol" $*$.) The distance between them is 1/16.

$L_1<i> = \text{baadbdb**b****dbb*b*dd**bb***a***}$

$L_2<j> = \text{baa**dbbb*d*bbb*dd**b****b**a***}$

The result of matching first and second original images are shown in Fig. 4. In this image, only the regions in which more than one-third of subimages are left unattached "don't care symbol" are displayed. It takes about 3 minutes to match two images.

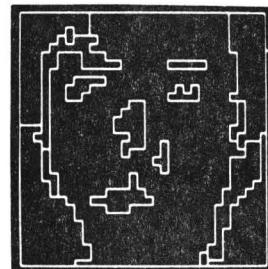


Fig.4 Matched image of two images

To match many images, this algorithm is applied one by one. The result of matching four images of Fig.1 is shown in Fig. 5. Some principal sub-objects are remained because the relative positions of them are common for all images and each of them has a same local feature.

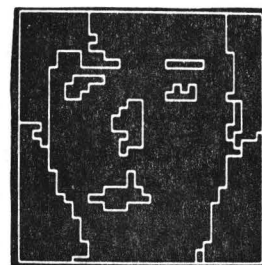


Fig.5 Matched image of four images

5. CONCLUSION

We have presented an algorithm to get a structure feature of images. This feature corresponds to the structure we get from images at a glance, and it is used as the cue to understand them. We used the relation between sub-images and from statistical property, we got the structure feature. The advantage of this method is that we can keep from noises, because noises which happen irregularly are removed while matching process. The results of experiments show the maximal set of subimages has enough information as the structure feature.

REFERENCES

- [1] Winston, P.H. "The Psychology of Computer Vision." McGraw-Hill, 1975.
- [2] Enomoto, H. et al. "Statistical Approach for Construction and Estimation of Syntactic Pattern Recognition Mechanism." in *Proc. IJCPR-78*, Kyoto, Japan, November, 1978, pp.359-363.
- [3] Fukada, Y. "Spatial Clustering Procedures for Region Analysis." In *Proc. IJCPR-78*, Kyoto, Japan, November, 1978, pp.329-331.
- [4] Hirschberg, D. "Algorithm for the Longest Common Subsequence Problem." *JACM* 24:4 (1977) pp.664-675.
- [5] Ohlander, R.B. "Analysis of Natural Scenes." PhD thesis, Computer Science Dept. Carnegie Mellon University, April, 1975.
- [6] Rosenfeld, A. "Digital Image Processing." Academic Press, 1976.