An Algorithm for Constructing a Multi-template Dictionary for Character Recognition Considering Distribution of Feature Vectors

Fang Sun, Shin'ichiro Omachi, and Hirotomo Aso
Department of Electrical Communications,
Graduate School of Engineering, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan
fan@ecei.tohoku.ac.jp

Abstract

In this paper, a new algorithm to construct a multi-template dictionary for character recognition is presented. It is a repetitive process of partitioning and estimating the distribution by observing both within-class distribution and between-class information. The effectiveness of the algorithm is shown by experimental results.

1. Introduction

A new algorithm to construct a multi-template dictionary for accurate and efficient recognition of handwritten Chinese and Japanese characters is proposed. The traditional multi-templates are constructed by partitioning the samples of a category into some fixed number of subclasses. The proposed algorithm will find an optimal number of templates for each category, since it is based on finding templates to be necessary to partition.

The Improved Directional Element Feature[7] with 196 dimensions is used to be the feature vector. As a classifier, the Euclidean distance is used.

2. Construction of a multi-template dictionary

Sample patterns, which are called *training patterns*, are prepared for each category. Assume that the distribution of training patterns of category k is regarded as mixture normal distribution of $\mathcal{N}(k)$ classes. At first, $\mathcal{N}(k) = 1$ for all k. By repeating the following two procedures, multi-template dictionary is constructed.

- Increase the number of classes of the category selected by partition conditions.
- Re-estimate the distribution of the selected category to get more reliable parameters.

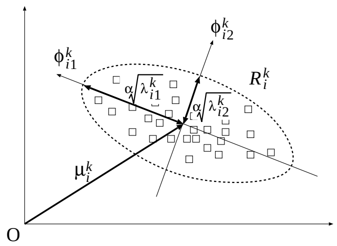

Figure 1. A feature region.

The partition conditions and estimation method are described in the following subsections.

2.1. Partition conditions

Feature regions of category k are the regions where vectors of category k distribute[1]. Let $\boldsymbol{\mu}_i^k$ be the ith mean vector of category k and Σ_i^k be the covariance matrix. Let λ_{it}^k and $\boldsymbol{\phi}_{it}^k$ be the eigenvalue and the unit eigenvector of tth principal component of Σ_i^k, respectively. The ith feature region of category k in the L-dimensional affine subspace[4] is defined as,

$$R_i^k = \left\{ \boldsymbol{u} \,\middle|\, \boldsymbol{u} = \boldsymbol{\mu}_i^k + \sum_{t=1}^{L} s_t \alpha \sqrt{\lambda_{it}^k} \boldsymbol{\phi}_{it}^k, \ \sum_{t=1}^{L} s_t^2 \leq 1 \right\}, \tag{1}$$

where $\alpha = 3.3$. An example is shown in Fig. 1. Because λ_{it}^k becomes small rapidly as the value t increases, small value L is used instead of the total number of dimensions n.

Let $\mathcal{N}(k)$ be the number of classes of category k. The region of category k should be partitioned finer if the following two partition conditions are satisfied.

$$\min_{\substack{l \neq k \\ 1 \leq j \leq \mathcal{N}(l)}} d_E(\boldsymbol{\mu}_i^k \pm \alpha \sqrt{\lambda_{it}^k} \boldsymbol{\phi}_{it}^k, \boldsymbol{\mu}_j^l) < \alpha \sqrt{\lambda_{it}^k}, \tag{2}$$

$$\mathcal{N}(k) < M. \tag{3}$$

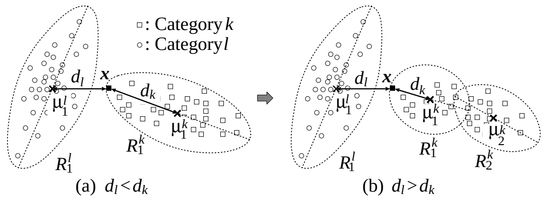

Figure 2. Condition for partition.

Here, $d_E(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the Euclidean distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. In Expression (2), $\boldsymbol{\mu}_i^k + \alpha\sqrt{\lambda_{it}^k}\boldsymbol{\phi}_{it}^k$ and $\boldsymbol{\mu}_i^k - \alpha\sqrt{\lambda_{it}^k}\boldsymbol{\phi}_{it}^k$ correspond to the ends of the region along with the tth axis of the ith feature region of category k. $\alpha\sqrt{\lambda_{it}^k}$ on the right side of Expression (2) is the distance between one of these points and $\boldsymbol{\mu}_i^k$. Expression (2) means that $\boldsymbol{\mu}_j^l$ is nearer to one of these ends that belongs to k than $\boldsymbol{\mu}_i^k$. In Expression (3), M is a constant to limit the number of classes of one category.

An example is shown in Fig. 2. In Fig. 2(a), $d_k = \alpha\sqrt{\lambda_{11}^k}$ and $d_l = d_E(\boldsymbol{\mu}_1^k + \alpha\sqrt{\lambda_{11}^k}\boldsymbol{\phi}_{11}^k, \boldsymbol{\mu}_1^l)$. A vector $\boldsymbol{x}$ that exists at the end of the region of category k is recognized as in category l, because d_l is smaller than d_k. In this case, if category k is partitioned into two feature regions as shown in Fig. 2(b), d_k becomes smaller than d_l. That means, after increasing the number of classes of category k, the vector $\boldsymbol{x}$ can be correctly assigned to category k.

2.2. Estimate of distribution

Assume that the distribution of category k is described as the following mixture normal distribution of $\mathcal{N}(k)$ classes.

$$p^k(\boldsymbol{x}) = \sum_{i=1}^{\mathcal{N}(k)} b_i^k p_i^k(\boldsymbol{x}), \tag{4}$$

where b_i^k is the mixing parameter and $p_i^k(\boldsymbol{x})$ is the component density function that is normal $N(\boldsymbol{\mu}_i^k, \Sigma_i^k)$.

If axis t of region i of category k satisfies the partition conditions (2) and (3), the number of $\mathcal{N}(k)$ is increased. Then the parameters of each class of category k are re-estimated by the maximum likelihood estimate[2]. For initial centroids for maximum likelihood estimate, $\boldsymbol{\mu}_i^k \pm \varepsilon\boldsymbol{\phi}_{it}^k$ and $\boldsymbol{\mu}_j^k$ $(j \neq i)$ are used, where ε is a small constant value.

For the component density function,

$$p_i^k(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma_i^k|^{1/2}} \exp\left\{-\frac{1}{2}d_M(\boldsymbol{x}, \boldsymbol{\mu}_i^k)\right\}, \tag{5}$$

is commonly used. Here, $d_M(\boldsymbol{x}, \boldsymbol{\mu}_i^k)$ is the Mahalanobis distance between $\boldsymbol{x}$ and $\boldsymbol{\mu}_i^k$. However, because there are not enough training samples comparing with the number of dimensions, the covariance matrix cannot be estimated accurately. Therefore, the calculated Mahalanobis distance is unreliable. For this reason, *Simplified Mahalanobis distance*[6], or SMD, is used instead of the Mahalanobis distance. The term $(2\pi)^{n/2}|\Sigma_i^k|^{1/2}$ is considered as constant. SMD replaces λ_{it} $(t > L)$ of the Mahalanobis distance with the mean value of λ_{it} $(t = L+1, ..., n)$, and it is given as,

$$d_S(\boldsymbol{x}, \boldsymbol{\mu}_i^k) = \sum_{t=1}^{L} \frac{1}{\lambda_{it}^k}((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^T \boldsymbol{\phi}_{it}^k)^2 + \frac{1}{\lambda}\left\{||\boldsymbol{x} - \boldsymbol{\mu}_i^k||^2 - \sum_{t=1}^{L}((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^T \boldsymbol{\phi}_{it}^k)^2\right\}, \tag{6}$$

where

$$\lambda = \frac{1}{n-L}\sum_{t=L+1}^{n} \lambda_{it}^k. \tag{7}$$

SMD has the same statistical properties as the Mahalanobis distance, and the effectiveness has been shown[6]. Replacing the Mahalanobis distance by SMD,

$$\tilde{p}_i^k(\boldsymbol{x}) = C\exp\left\{-\frac{1}{2}d_S(\boldsymbol{x}, \boldsymbol{\mu}_i^k)\right\}, \tag{8}$$

is used as the estimated component density function. Here, C is a constant. $d_S(\boldsymbol{x}, \boldsymbol{\mu}_i^k)$ is the SMD between $\boldsymbol{x}$ and $\boldsymbol{\mu}_i^k$.

2.3. Algorithm for constructing a dictionary

The outline of the algorithm is as following. First, feature vectors are extracted from training patterns prepared for each category. Next, with these training patterns, the feature region of each category is estimated. Among the regions that satisfy (2) and (3), the one with largest λ_{it}^k $(1 \leq t \leq L)$ is selected, and the parameters of distribution of the selected category is re-estimated using the maximum likelihood estimate. This process is repeated until no category satisfies (2) and (3). Finally, each mean vector $\boldsymbol{\mu}_i^k$ is used to be the ith template of category k. The final set of templates is considered to be the dictionary. According to preliminary experimental results, $L = 5$ is chosen, because the value λ_{it}^k is found to be very small if t is larger than five.

Fig. 3 shows an example. If the first principal component of region R_1^k is selected, initial centroids are set as $\boldsymbol{\mu}_1^k \pm \varepsilon\boldsymbol{\phi}_{11}^k$ as shown in Fig. 3(a). Then the distributions are estimated as two-class normal. The result is shown in Fig. 3(b). Continuously, if the first principal component of R_2^k, and if the second principal component of new R_2^k are selected in turn, final feature regions are displayed in Fig. 3(d).

3. Experiments

In this section, experimental results of character recognition using the dictionary constructed by the proposed al-

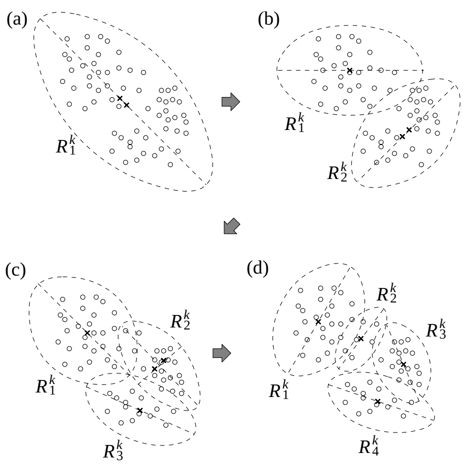

Figure 3. Initial centroids.

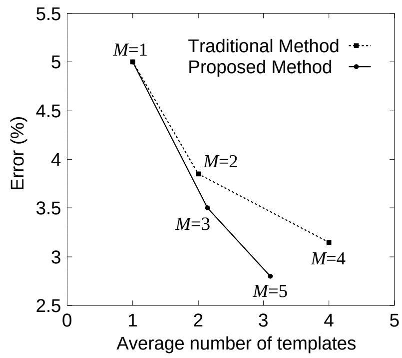

Figure 4. Experimental results.

gorithm are shown. In our study, objects of experiments are handwritten Chinese and Japanese characters. The first 100 kinds of Chinese and Japanese characters in the database ETL9B[5] are used for experiments. For each category, 200 samples are prepared. The first 20 samples are used as test patterns and the other 180 samples are used as training patterns. A dictionary is constructed by the algorithm described in Section 2. M in Expression (3) varies several values.

For comparison, experiments using another dictionary with fixed number (M) of classes for each category are carried out. This method is called traditional method. In this case, LBG algorithm[3] is used to partition the samples in each category.

Experimental results are shown in Fig. 4. Horizontal axis shows the number of templates per category, and vertical axis shows error rate. The value of M is also indicated in the figure. These results show that low error rate is achieved with a small size dictionary constructed by the proposed algorithm. For example, in the case of the proposed method with the average number of templates about three ($M = 5$), the error rate is smaller than the traditional method when the number of templates is four.

4. Conclusions

A new algorithm to construct a multi-template dictionary for accurate and efficient recognition of handwritten Chinese and Japanese characters is proposed. The proposed algorithm is a repetitive process of partitioning and estimating the distribution by observing both within-class distribution and between-class information. With this algorithm, only the number of templates of those categories determined to need much more templates is increased. As a result, the optimal number of templates for each category are prepared in the dictionary. For estimating distributions of samples with the maximum likelihood estimate, SMD is used in the component density function instead of the Mahalanobis distance. The experimental results have shown that by considering distributions of samples, the criterion becomes much more reliable even in the case of definite training samples.

The effectiveness of the algorithm has been affirmed by experiments using Chinese and Japanese handwritten character database. The results have shown that low error rate is achieved with the small size dictionary constructed by the proposed algorithm. This means, suitable number of templates for each category can be estimated by the proposed algorithm.

This algorithm can be easily applied for recognition of digits, alphabets, and all kinds of Chinese and Japanese characters. To show the effectiveness for such characters is the future problem.

References

[1] H. Aso, K. Echigo, and M. Kimura. Structural properties of character feature space and evaluation of effectiveness of features. *Trans. IEICE*, J76-D-II(11):2285–2294, November 1993.

[2] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons Inc., New York, 1973.

[3] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Trans. on Comm.*, COM-28(1):84–95, January 1980.

[4] E. Oja. *Subspace Methods of Pattern Recognition*. Research Studies Press, England, 1983.

[5] T. Saito, H. Yamada, and K. Yamamoto. On the data base ETL9 of handprinted characters in JIS Chinese characters and its analysis. *Trans. IEICE*, J68-D(4):757–764, April 1985.

[6] F. Sun, S. Omachi, and H. Aso. Precise selection of candidates for handwritten character recognition using feature regions. *IEICE Trans. Inf. & Syst.*, E79-D(5):510–515, May 1996.

[7] N. Sun, M. Abe, and Y. Nemoto. A handwritten character recognition system by using improved directional element feature and subspace method. *Trans. IEICE*, J78-D-II(6):922–930, June 1995.