An Algorithm for Estimating Mixture Distribution of High Dimensional Vectors And Its Application to Character Recognition

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

Abstract

For statistical pattern recognition, in order to obtain high recognition accuracy, it is very important to estimate distribution precisely. In many cases, the distribution of feature vectors which are extracted from recognition objects is assumed to be normal, however it is more intricate and volatile in practice. It is thought to be more feasible to assume the distribution as mixed normal distribution. To estimate the mixed distribution precisely, a great number of training samples are required, especially in the case that the number of dimensions of feature vector is large. But unfortunately, compared with the number of dimensions, there are always not enough training samples. For this reason, the mixed normal distribution estimation is rarely used in recognition problems using high dimensional vectors, for example, character recognition. In this paper, by introducing Simplified Mahalanobis distance to the maximum likelihood estimates, the mixed normal distribution estimation algorithm for high dimensional vectors is proposed. As a practical application, the estimation algorithm is adopted to character recognition. A multi-template dictionary is constructed with consideration of the distribution of each category. The effectiveness of the proposed method is examined by experiments using Japanese characters.

1 Introduction

Significant achievements are made by statistical pattern recognition methods considering distributions of sample patterns in a feature space. For statistical pattern recognition, in order to obtain high recognition accuracy, it is very important to estimate distribution precisely. In many cases, the distribution of feature vectors which are extracted from recognition objects is assumed to be normal, however, they are much more intricate and volatile in practice. It is thought more feasible to assume the distribution as mixed normal distribution and then estimate the shape of it. To estimate the mixed distribution precisely, a large number of training samples is necessary, especially for treating high dimensional feature vectors. For some pattern recognition problems, such as Chinese and Japanese character recognition, very high dimensional feature vectors are used. But unfortunately, compared with the number of dimensions, there are always not enough training samples. For this reason, the mixed normal distribution estimation is rarely used in the recognition problems using high dimensional vectors, like character recognition.

Clustering problem is a common topic in pattern recognition field. In Chinese and Japanese character recognition research, various clustering methods are used to construct multi-template dictionaries for multi-font printed characters or handwritten characters [1], [2], [3]. Although these multi-template dictionaries have effects for recognizing, there are several problems that should be concerned. Considering the characteristic of distribution of category, it is thought unnecessary to increase all categories with a fixed number. Dividing a category without necessity will only increase the risk of mis-classification and enlarge the size of dictionary. Therefore, it is extremely important to select only categories that may be recognized incorrectly and limit the size of dictionary as small as possible. The other problem is as a category being partitioned into more classes, the number of training samples in one class becomes less and less. As a result, parameters are estimated with inadequate training samples, and they will be unreliable.

In this paper, the mixed normal distribution estimation algorithm for high dimensional vectors is proposed. This algorithm is based on the maximum likelihood estimates [4]. In the case that high dimensional vectors are used, the covariance matrix usually cannot be estimated accurately. To avoid the bad influence caused by this problem, a new probability density function using Simplified Mahalanobis distance [5], or SMD, is adopted instead of that of multivariate normal distribution. The SMD is a discriminant function originally proposed for character recognition.

As a practical application, the estimation algorithm is applied to character recognition. A new method of constructing a multi-template dictionary based on the estimated mixture distribution is proposed. First, by examining the relationship between two categories, the category with high misclassification possibility is selected and clustered. Then the distributions of all classes that belong to the partitioned category are estimated by the proposed estimation algorithm. Consequently, only the categories that are considered necessary to be partitioned into more classes are divided. Moreover, no matter how many classes a category is divided into, the parameters of each class are estimated with the whole training samples in that category. This evaluation process improves the reliability of parameters. The effectiveness of the proposed method is examined by experiments using Japanese characters.

The organization of the rest of the paper is as follows. In Section 2, the method of traditional maximum likelihood estimates of parameters of mixture normal distribution is described. Then a new algorithm for estimating mixture distribution is proposed. In Section 3, a method to construct a multi-template dictionary based on the estimated mixture distribution is proposed. Finally, in Section 4, experiments are carried out to confirm the effectiveness of the proposed method.

2 Mixture Distribution Estimation

2.1 Maximum Likelihood Estimates

First, the traditional maximum likelihood estimates of parameters of mixture normal distribution [4] is described in brief. Assume that the distribution of category k is defined as the following mixture normal distribution of $\mathcal{N}(k)$ classes.

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

where $\boldsymbol{\theta}_i^k$ is the ith parameter vector that includes $\boldsymbol{\mu}_i^k$ and Σ_i^k and b_i^k is the mixing parameter. $p_i^k(\boldsymbol{x}|\boldsymbol{\theta}_i^k)$ is the component density function that is normal $N(\boldsymbol{\mu}_i^k, \Sigma_i^k)$, which is described as

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

Here, $d_M(\boldsymbol{x}, \boldsymbol{\theta}_i^k)$ is the Mahalanobis distance, and the definition is as follows.

$$\begin{aligned} d_M(\boldsymbol{x}, \boldsymbol{\theta}_i^k) &= (\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t (\Sigma_i^k)^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i^k) & (3)\\ &= \sum_{t=1}^{n} \frac{1}{\lambda_{it}^k} ((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t \boldsymbol{\phi}_{it}^k)^2, & (4) \end{aligned}$$

where n is the number of dimensions of feature vector, λ_{it}^k is the tth eigenvalue of Σ_i^k sorted by descending order, and $\boldsymbol{\phi}_{it}^k$ is the eigenvector that corresponds to λ_{it}.

Denote the number of sample vectors of category k as M, and let $\boldsymbol{x}_j^k$ be the jth sample vector $(1 \leq j \leq M)$. In the traditional maximum likelihood estimates, parameters are estimated iteratively by the following equations.

$$\hat{b}_i^k = \frac{1}{M} \sum_{j=1}^{M} \tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k), \tag{5}$$

$$\hat{\boldsymbol{\mu}}_i^k = \frac{\sum_{j=1}^{M} \tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k)\boldsymbol{x}_j^k}{\sum_{j=1}^{M} \tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k)}, \tag{6}$$

$$\hat{\Sigma}_i^k = \frac{\sum_{j=1}^{M} \tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k)(\boldsymbol{x}_j^k - \hat{\boldsymbol{\mu}}_i^k)(\boldsymbol{x}_j^k - \hat{\boldsymbol{\mu}}_i^k)^t}{\sum_{j=1}^{M} \tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k)}, \tag{7}$$

where

$$\tilde{b}_i^k(\boldsymbol{x}_j^k, \hat{\boldsymbol{\theta}}_i^k) = \frac{p_i(\boldsymbol{x}_j^k|\hat{\boldsymbol{\theta}}_i^k)\hat{b}_i^k}{\sum_{t=1}^{\mathcal{N}(k)} p_t(\boldsymbol{x}_j^k|\hat{\boldsymbol{\theta}}_t^k)\hat{b}_t^k}. \tag{8}$$

2.2 Proposed Estimation Algorithm

In the case of using high dimensional feature vectors, comparing with the number of dimensions, the training samples are always not enough, hence covariance matrix usually cannot be estimated accurately [6], [7], [8]. Especially, the terms that correspond to the smaller eigenvalues in Eq. 4 will include much more error [7]. For this reason, the SMD is used instead of the Mahalanobis distance, and the term $(2\pi)^{n/2}|\Sigma_i^k|^{1/2}$ is considered as a constant. The 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,

$$\begin{aligned} & d_S(\boldsymbol{x}, \boldsymbol{\theta}_i^k) \\ &= \sum_{t=1}^{L} \frac{1}{\lambda_{it}^k} ((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t \boldsymbol{\phi}_{it}^k)^2 \\ &\quad + \sum_{t=L+1}^{n} \frac{1}{\lambda} ((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t \boldsymbol{\phi}_{it}^k)^2 & (9)\\ &= \sum_{t=1}^{L} \frac{1}{\lambda_{it}^k} ((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t \boldsymbol{\phi}_{it}^k)^2 \\ &\quad + \frac{1}{\lambda} \left\{ \|\boldsymbol{x} - \boldsymbol{\mu}_i^k\|^2 - \sum_{t=1}^{L} ((\boldsymbol{x} - \boldsymbol{\mu}_i^k)^t \boldsymbol{\phi}_{it}^k)^2 \right\}, \\ & & (10) \end{aligned}$$

where

$$\begin{aligned} \lambda &= \frac{1}{n-L} \sum_{t=L+1}^{n} \lambda_{it}^k & (11)\\ &= \frac{1}{n-L} \left\{ \mathrm{tr}\Sigma_i^k - \sum_{t=1}^{L} \lambda_{it}^k \right\}. & (12) \end{aligned}$$

Here, $\mathrm{tr}\Sigma_i^k$ denotes the trace of Σ_i^k.

As it is shown in Eq. 10, the SMD only uses the larger eigenvalues. That means the smaller eigenvalues that the error mostly included in will not bring any influence to the

SMD. The SMD also makes the maximum likelihood estimates of parameters possible even in the case that the number of samples is less than the number of dimensions. Moreover, the SMD has the same statistical properties as the Mahalanobis distance. Replacing the Mahalanobis distance by the SMD,

$$\tilde{p}_i^k(\boldsymbol{x}|\boldsymbol{\theta}_i^k) = C \exp\left\{-\frac{1}{2} d_S(\boldsymbol{x}, \boldsymbol{\theta}_i^k)\right\}, \qquad (13)$$

is used as the estimated component density function instead of Eq. 2. Here, C is a constant.

How to decide the initial parameter vector $\boldsymbol{\theta}_i^k$ is another problem for estimating the mixture distribution. It is very important to choose appropriate initial parameter vectors since the result of estimation depends on the vectors. In our approach, initial parameters are given by the result of clustering. First, *source mean vectors* $\boldsymbol{\mu}_i^k$ $(i = 1, ..., \mathcal{N}(k))$ are given. Then all the training vectors are clustered into $\mathcal{N}(k)$ classes by k-means algorithm [9]. Finally, mean vector and covariance matrix for each class are calculated and they are used as the initial parameters.

3 Application to Character Recognition

In this section, a new method to construct a multi-template dictionary for accurate and efficient recognition of handwritten characters is proposed. By considering the relation between distributions of different categories, the possible occurrence of incorrect recognition is assumed. Based on the proposed estimation algorithm, a clustering method to reduce mis-classification is developed. Here, the ETL9B [10] is used for the character recognition experiments. It is the largest public database in Japan and it includes both Chinese and Japanese handwritten characters. There are 200 samples of each character.

3.1 Feature Vector

The Improved Directional Element Feature [11] is used as the feature vector here. It is calculated as follows. As shown in Fig. 1, first an input image is normalized to 64×64 dots, and contour of the image is extracted. Next, orientation, which is one of vertical, horizontal, and two oblique lines slanted at $\pm 45°$, is assigned for each pixel. Then the image is divided into 49 sub-areas of 16×16 dots where each sub-area overlaps eight dots with the adjacent sub-area. (For example, meshed area of Fig. 1(d).) For each sub-area, a four-dimensional vector is defined to represent the quantities of the four orientations. Thus the total vector for one character has 196 ($= 49 \times 4$) dimensions.

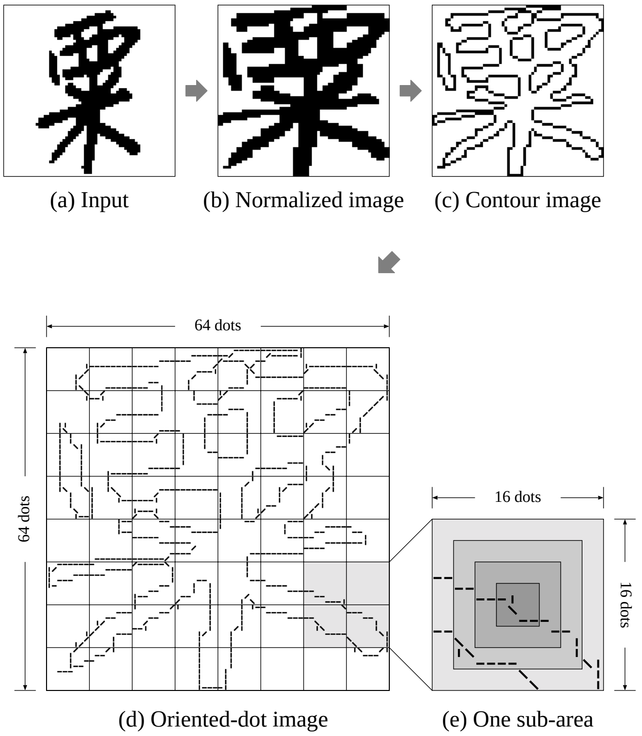

(a) Input
(b) Normalized image
(c) Contour image
(d) Oriented-dot image
(e) One sub-area

Figure 1: Improved Directional Element Feature.

3.2 Feature Region

Here, the concept of *feature region* [12] is explained. Assume the sample vectors of category k are classified into several classes. The ith feature region of category k is a region where the sample vectors of class i of category k distribute, and it is defined as follows.

$$R_i^k = \left\{ \boldsymbol{x} \,\middle|\, d_M(\boldsymbol{x}, \boldsymbol{\theta}_i^k) \leq d_{\max} \right\}. \qquad (14)$$

Here, $d_{\max}$ is determined to make R_i^k include the vectors of category k as much as possible. If there are enough sample vectors compared with the number of dimensions, the values of the Mahalanobis distance will distribute according to biased F-distribution [13]. Then by deciding the percentage of the sample vectors that distribute in the region, the value of $d_{\max}$ can be determined theoretically. However, in character recognition process, it is difficult to prepare enough samples compared with the number of dimensions of feature vector. Here, $d_{\max}$ is determined to be the Mahalanobis distance between the mean vector and the farthest vector in category k. It is expressed as

$$d_{\max} = \max_j d_M(\boldsymbol{x}_j^k, \boldsymbol{\theta}_i^k). \qquad (15)$$

Here, $\boldsymbol{x}_j^k$ is the jth sample vector of class i of category k. An illustration of a feature region is shown in Fig. 2.

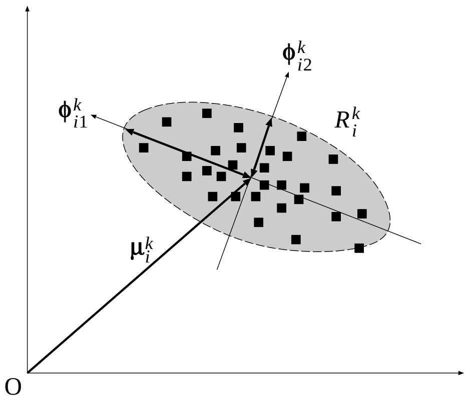


Figure 2: A feature region.

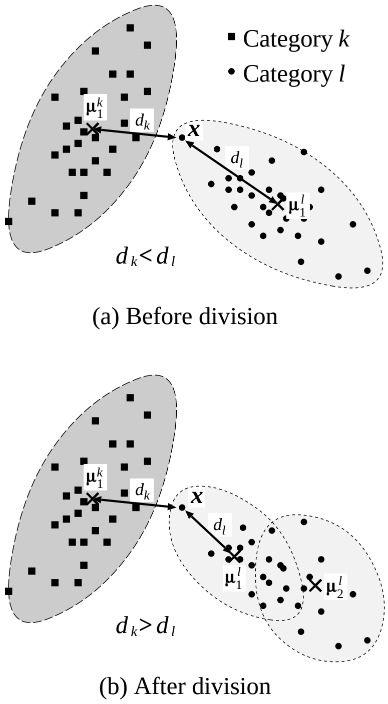
3.3 Relation Between Two Categories

When constructing a multi-template dictionary, dividing categories without necessity will only increase the risk of mis-classification and the computation cost. Therefore, it is extremely important to select and cluster the categories that may be recognized incorrectly.

Obviously, the categories regarded with mis-classification possibility are different according to the used classifier. In our previous work, the Euclidean distance is used to be the classifier [14]. The brief summary is as follows. The category considered necessary to be partitioned is selected. Fig. 3 shows an illustration. Denote the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{\mu}$ as $d_E(\boldsymbol{x}, \boldsymbol{\mu})$. In Fig. 3(a), a vector $\boldsymbol{x}$ that exists at the end of the region of category l is recognized as in category k, because $d_E(\boldsymbol{x}, \boldsymbol{\mu}_1^k)$ is smaller than $d_E(\boldsymbol{x}, \boldsymbol{\mu}_1^l)$. In this case, if category l is partitioned into two feature regions as shown in Fig. 3(b), $d_E(\boldsymbol{x}, \boldsymbol{\mu}_1^l)$ becomes smaller than $d_E(\boldsymbol{x}, \boldsymbol{\mu}_1^k)$. That means, after increasing the number of classes of category l, the vector $\boldsymbol{x}$ can be correctly recognized as category l.

As a classifier, the Mahalanobis distance or the SMD is much more effective than the Euclidean distance [15]. In order to obtain better recognition performance, the SMD is adopted to be the classifier in this study. The number of dimensions (L in Eq. 10) is 30. In the case of using the Mahalanobis distance, the situation of mis-classification as shown in Fig. 3(a) will not happen. Obviously, for the Mahalanobis distance, the situations of mis-classification are totally different from the Euclidean distance. It is useful to know the relation between two distributions of categories. As shown in Fig. 4, there are three considerable relations:

- Non-overlap: There is no overlapped area of two categories.
- Semi-overlap: There is an overlapped area of two categories, however samples of only one category distribute in the overlapped area.
- Overlap: There is an overlapped area of two categories, and samples of both categories distribute in the overlapped area.



(a) Before division

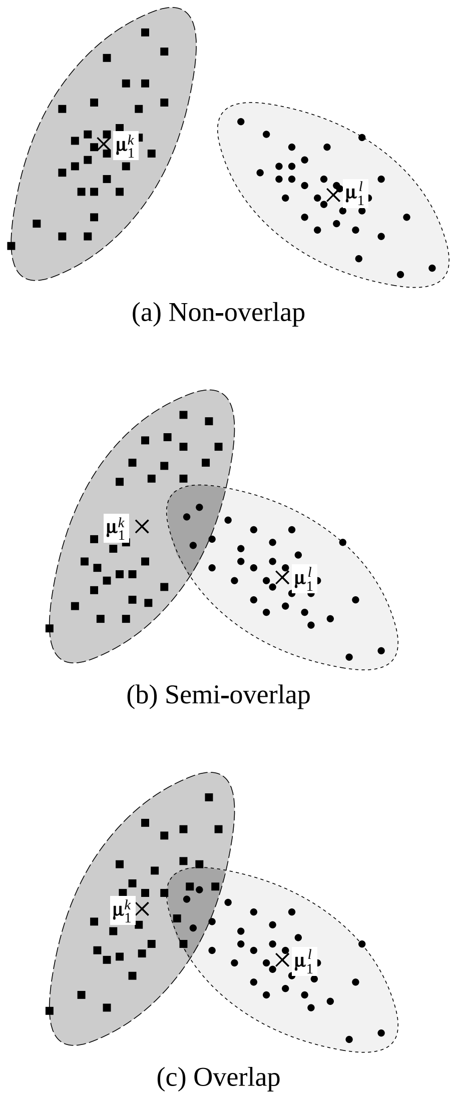
(b) After division

Figure 3: Method for avoiding mis-classification (Euclidean distance).

Among these three types of relations, in the case that distributions of two categories are "Semi-overlap" or "Overlap," a mis-classification may happen.

Among the pairs of two categories, the numbers of "Non-overlap," "Semi-overlap," and "Overlap" are investigated. The Japanese characters in the database ETL9B are used, and the number of kinds of characters is seventy-one. For each character, the first 180 samples are used for investigation. A part of the result is shown in Table 1, and this table is called *overlap-relation table*. In the table, "Sample" means the kind of character sample, and "Region" means the feature region. That is, (k, l)th number is the number of samples of category k within the feature region of category l. For example, fourteen samples of あ are in the feature region of お. These 14 images of あ are illustrated in Fig. 5. For comparison, some images of お are displayed in Fig. 6. Since the number of categories is 71, the total number of pairs of categories is 2485. The numbers of "Non-overlap," "Semi-overlap" and "Overlap" are 2404, 57 and 24, respectively.

Continuously, the selection standard of division is considered. An example of "Semi-overlap" relation is shown in Fig. 7(a). Some samples of category l are in the feature region of category k, but no sample of category k is in the feature region of category l. In this case, dividing category l does not change the situation. However, if category k is divided as shown in Fig.7(b), the overlapped area may dis-



(a) Non-overlap

(b) Semi-overlap

(c) Overlap

Figure 4: Three types of relations of two distributions.

Table 1: Results of investigation.

Sample	Region								
	あ	い	う	え	お	か	が	き	···
あ	—	0	0	0	14	0	0	0	···
い	0	—	0	0	0	0	0	0	···
う	0	0	—	0	0	0	0	0	···
え	0	0	0	—	0	0	0	0	···
お	0	0	0	0	—	0	0	0	···
か	0	0	0	0	0	—	44	0	···
が	0	0	0	0	0	0	—	0	···
き	0	0	0	0	0	0	0	—	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

appear and mis-classification may be avoided. For these reasons, the categories that include samples of other categories are decided to be the division targets.

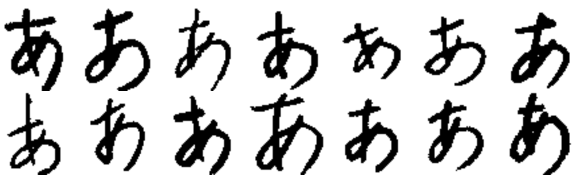


Figure 5: Character images of あ within the feature region of お.

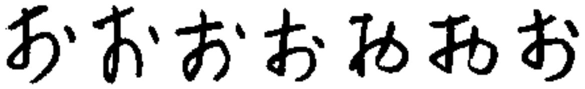
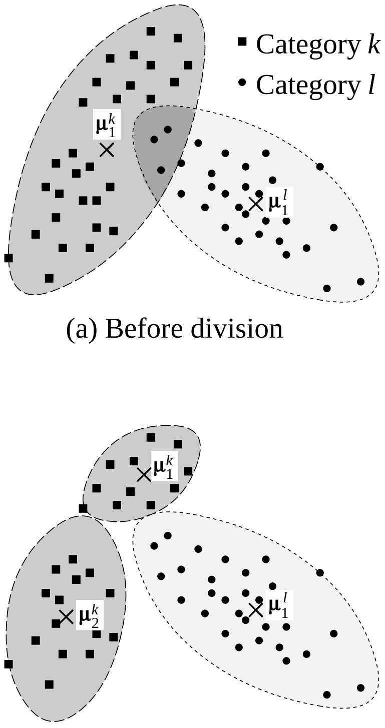


Figure 6: Character images of お.



(a) Before division

(b) After division

Figure 7: Method for avoiding mis-classification (Mahalanobis distance).

3.4 Construction of Dictionary

According to the above investigation, a method for constructing a multi-template dictionary is proposed. Training sample patterns are prepared for each category. Assume the distribution of training patterns of category k to be the mixture normal distribution of $\mathcal{N}(k)$ classes. At first, $\mathcal{N}(k) = 1$ for all k, and the maximum number of mixture for one category $\mathcal{N}_{\max}$ is defined. By repeating the following three proce-

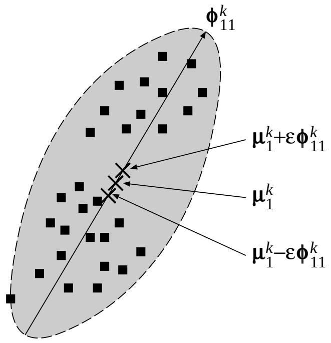

ϕ_{11}^k
$\mu_1^k+\varepsilon\phi_{11}^k$
μ_1^k
$\mu_1^k-\varepsilon\phi_{11}^k$

Figure 8: Source mean vectors.

dures, a multi-template dictionary is constructed.

1. Make an overlap-relation table like Table 1 using the training samples.
2. Select a category c such that $\mathcal{N}(c) < \mathcal{N}_{\max}$ whose feature region includes the largest number of samples of another category.
3. Increase the number of $\mathcal{N}(c)$ and estimate the parameters of mixture distribution of category c.

In our experiments, $\mathcal{N}_{\max} = 2$. For source mean vectors, $\boldsymbol{\mu}_1^k \pm \varepsilon\boldsymbol{\phi}_{11}^k$ are selected, where ε is a small constant value (see Fig. 8).

4 Experiments

In this section, experimental results of character recognition using the dictionary constructed by the proposed method are shown. In our study, objects of experiments are all kinds of handwritten Japanese characters in the database ETL9B, and the number of kinds of characters is seventy-one. For each character, the first 180 samples are used for training data and the rest of 20 samples are used for testing data. For comparison, the dictionary without devision (called single-template dictionary) and the dictionary constructed by dividing all the categories into two classes (called two-class dictionary) are also used for the experiments.

4.1 Comparison of the Dictionaries

The dictionary is constructed by the algorithm described in Section 3.4. The overlap-relation table of this dictionary is shown in Table 2. All the pairs of eight characters in the table become “Non-overlap.”

The numbers of three relations for three kinds of dictionaries are summarized in Table 3. The first three rows indicate the numbers of “Non-overlap,” “Semi-overlap” and “Overlap”. Comparing results of the proposed method with the single-template, the numbers of “Semi-overlap” and “Overlap” decrease tremendously. These results have shown that the possibility of mis-classification is reduced by dividing categories. The fourth row in the table displays the total number of classes. By using the proposed method, 47 out of 71 categories are divided, and the total number of classes is 118. In other words, 24 categories are decided unnecessary to be divided. With smaller size of dictionary, the proposed method has obtained the same results as the two-class method. It proves the proposed method is able to choose the categories that need to be divided.

Table 2: Number of overlapped samples after constructing dictionary.

Sample	Region								
	あ	い	う	え	お	か	が	き	···
あ	—	0	0	0	0	0	0	0	···
い	0	—	0	0	0	0	0	0	···
う	0	0	—	0	0	0	0	0	···
え	0	0	0	—	0	0	0	0	···
お	0	0	0	0	—	0	0	0	···
か	0	0	0	0	0	—	0	0	···
が	0	0	0	0	0	0	—	0	···
き	0	0	0	0	0	0	0	—	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Table 3: Results of division.

	Single-template	Two-class	Proposed method
Non-overlap	2404	2476	2476
Semi-overlap	57	6	6
Overlap	24	3	3
# of classes	71	142	118

4.2 Experimental Results

Experiments of recognizing testing data are carried out. The results (error rates) of the three methods are shown in Table 4. The error rate of the proposed method is the smallest among these three methods. These experimental results have shown the effectiveness of the proposed method. The fact that the proposed method is better than two-class shows that dividing a category without necessity not only enlarges the size of dictionary but also increases the risk of mis-classification.

Fig.9 shows the mistaken characters of one set of seventy-one characters by three methods. The wrong selected can-

Table 4: Experimental results.

	Single-template	Two-class	Proposed method
Error rate	5.9%	6.5%	5.1%

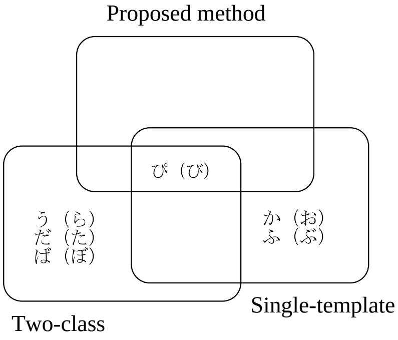


Figure 9: Mistaken characters.

didates are also shown in the parentheses. Even though the results of single-template are correct, う, だ and ば are mistaken by the two-class method. The reason of failure is that the two-class method clusters all categories by a fixed number without considering any between-class information of distributions of samples, even in the cases of う, だ and ば that need not to be partitioned. In contrast, the proposed method clusters categories determined necessary to be partitioned, and estimates the region of each category precisely. As shown in Fig.9, except ぴ, which is mis-classified by all methods, there is no more wrong candidate selected by the proposed method.

The difference between the proposed method and most conventional multi-template methods is the proposed method uses all training samples to estimate parameters of a class. In order to compare with method that calculates the parameters of a class only with the training samples belong to the class, another experiment is carried out. As the clustering method, LBG algorithm[16] is used. The error rate of this experiment is 5.7%. This result has shown that estimating mixture distribution by the proposed method is more effective than conventional clustering method.

5 Conclusions

In order to estimate the distribution of feature vectors in a feature space, it is more feasible to assume the distribution as mixed normal distribution. However, the mixed normal distribution estimation is rarely used in high dimensional recognition problems, because usually there are not enough training patterns.

In this paper, a new algorithm to estimate the parameters of the mixed normal distribution is proposed. This algorithm is based on the maximum likelihood estimates. By introducing the Simplified Mahalanobis distance to the estimation process, the maximum likelihood estimates of parameters can be done in the case of inadequate training samples. The proposed algorithm is effective in treating the high dimensional vectors.

As a practical application, a new method to construct a multi-template dictionary for character recognition is proposed. The proposed method is a repetitive process of clustering and estimating mixture distribution by observing both within-class distribution and between-class information. With this method, only the numbers of classes of those categories determined to need much more classes are increased. Moreover, no matter how many classes a category is divided into, the parameters of each class are estimated with the whole number of training samples in that category. This evaluation operation improves the reliability of parameters. The effectiveness of the method has been affirmed by experiments using handwritten characters. The results have shown that low error rate is achieved with the small size dictionary constructed by the proposed method. This means, the optimal number of classes for each character are prepared in the dictionary by the proposed method. The proposed mixture distribution estimation algorithm makes a new approach to pattern recognition using high dimensional vectors, such as character recognition.

This method can be easily applied for recognition of digits, alphabets, and all kinds of Chinese and Japanese characters. To show the effectiveness for those characters is the future problem. The proposed estimation algorithm can be adopted to other problems using high dimensional vectors, and it is also the future work.

Acknowledgments

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Encouragement of Young Scientists, 10780216, 1998.

References

[1] T. Hai, T. Morishita, Y. Kabuyama, Y. Izaki, and E. Yamamoto, “A Study of Multiple Template Matching for Handwritten Kanji Character Recognition,” *IEICE Technical Report*, PRL81-42, 1981.

[2] M. Shiono, “Fundamental Experiment on Handwritten KANJI Recognition by Multidictionary Template Matching Method,” *Transactions of Information Processing Society Japan*, vol.27, no.9, pp.853–859, Sept. 1986.

[3] S. Katoh and H. Takahashi, “A Multi-Font Kanji Recognition Method Using a Layered Template Dictionary,” *Trans. IEICE*, vol.J74-D-II, no.1, pp.8–18, Jan. 1991.

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

[5] F. Sun, S. Omachi, and H. Aso, “Precise Selection of Candidates for Handwritten Character Recognition Using Feature Regions,” *IEICE Trans. Inf. & Syst.*, vol.E79-D, no.5, pp.510–515, May 1996.

[6] S.J. Raudys, A.K. Jain, “Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.13, no.3, pp.252–264, March 1991.

[7] T. Takeshita, F. Kimura, and Y. Miyake, “On the Estimation Error of Mahalanobis Distance,” *Trans. IEICE*, vol.J70-D, no.3, pp.567–573, March 1987.

[8] M. Sakai, M. Yoneda, and H. Hase, “A New Robust Quadratic Discriminant Function,” *Proc. 14th Int’l Conf. Pattern Recognition (ICPR’98)*, pp.99–102, Aug. 1998.

[9] J.A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, Inc., 1975.

[10] T. Saito, H. Yamada, and K. Yamamoto, “On the Data Base ETL9 of Handprinted Characters in JIS Chinese Characters and Its Analysis,” *Trans. IEICE*, vol.J68-D, no.4, pp.757–764, April 1985.

[11] N. Sun, M. Abe, and Y. Nemoto, “A Handwritten Character Recognition System by Using Improved Directional Element Feature and Subspace Method,” *Trans. IEICE*, vol.J78-D-II, no.6, pp.922–930, June 1995.

[12] H. Aso, K. Echigo, and M. Kimura, “Structural Properties of Character Feature Space and Evaluation of Effectiveness of Features,” *Trans. IEICE*, vol.J76-D-II, no.11, pp.2285–2294, Nov. 1993.

[13] I.T. Young, “Further Consideration of Sample and Feature Size,” *IEEE Trans. Inf. Theory*, vol.IT-24, no.6, pp.773–775, June 1978.

[14] F. Sun, S. Omachi, and H. Aso, “An Algorithm for Constructing a Multi-template Dictionary for Character Recognition Considering Distribution of Feature Vectors,” *Proc. 14th Int’l Conf. Pattern Recognition (ICPR’98)*, pp.1114–1116, Aug. 1998.

[15] N. Kato, M. Abe, and Y. Nemoto, “A Handwritten Character Recognition System by Using Modified Mahalanobis Distance,” *Trans. IEICE*, vol.J79-D, no.1, pp.45–52, Jan. 1996.

[16] Y. Linde, A. Buzo, and R. M. Gray, “An Algorithm for Vector Quantizer Design,” *IEEE Trans. on Comm.*, vol.COM-28, no.1, pp.84–95, Jan. 1980.