A Discriminant Function for Noisy Pattern Recognition

Shin'ichiro Omachi, Fang Sun, 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 a practical pattern recognition system, noisy pattern recognition is necessary and important. There are several basic ideas of recognizing noisy patterns, for example, constructing a dictionary with noisy patterns, applying different classifiers, or using filters to delete noise. In most conventional statistical pattern recognition methods, a feature vector is extracted from an object. The distribution of feature vectors is estimated for each category to select a candidate for an unknown input pattern. As it is known, the distribution of feature vectors will change if noise occurs. It is impossible to predict all kinds of noise that happen accidentally and irregularly. In this paper, an attempt to deal with noisy patterns is examined, and a new discriminant function is proposed. By introducing a *revision matrix* calculated from an unknown input pattern, the new function can be defined as a revised form of traditional function with the degree of detected noise of each individual pattern. The ability of the proposed function is evaluated by experiments on discrimination of two categories. Experimental results not only show the classification effectiveness of the proposed function, but also confirm that it is enormously important to detect the noise and to revise the discriminant function for noisy pattern recognition.

1 Introduction

Significant achievements are made by statistical pattern recognition methods considering distribution of sample patterns in feature space. In most conventional pattern recognition methods, the first step is to extract features from objects. These features are always expressed in form of feature vectors. Then, the distribution of feature vectors is estimated for each category using training data. Finally, the category with the maximum probability is selected to be the candidate of an unknown input pattern.

Because noisy pattern recognition is necessary and important for a practical pattern recognition system, it has lately attracted considerable attention. For example, in the field of character recognition, several methods for recognizing poor quality characters have been proposed. Hobby et al. [1] developed a method to enhance degraded document images by finding and averaging bitmaps of the same kind of symbols. It improves the display appearance and recognition accuracy. Chou et al. proposed a flexible matching method between template images and unknown character images [2]. A vector field, called character deformation field, is used for representing deformation. Rodríguez et al. exploited a two-stage classifier [3]. First, a multifont classifier is applied. Then, a specialized classifier rerecognizes the ambiguous patterns using the patterns whose certainty of correct classification is high.

With the spread of digital cameras, some studies on recognizing poor quality characters included in the images taken by digital cameras have been done [4], [5], [6]. Sawa et al. use Gaussian Laplacian filter to emphasize the image [4]. Then character segmentation and recognition are done with dynamic programming. The moving subtraction method is proposed by Kosai et al. [5]. It uses plural images by swinging a camera vertically and horizontally, and supplements the bad influence caused by the lowness of resolution. The method developed by Sawaki et al. prepares a multiple-dictionary to deal with the images under any conditions [6]. The environment condition of an image is estimated, and a relevant dictionary reflecting the condition is used for recognition.

Although the above approaches have obtained effects, it is thought there is a limit to apply multiple classifiers or dictionaries because noise is intricate and volatile. Since types of noise are always quite different from each other, no one can predict what kind of noise an unknown input pattern will carry. Because noise may change the appearance of an unknown input pattern, the feature vector extracted from a noisy pattern will be very different from clear ones. As an empirical result, in the case that the distribution is estimated with only noiseless samples, whereas the unknown input pattern is noisy, the recognition result is always unsatisfactory. On the other hand, even if the distribution is estimated with noisy samples, there is no guarantee that the type of noise of an unknown input pattern is included in training samples. Moreover, using a dictionary constructed by noisy patterns is also inappropriate to recognize clear patterns. Therefore, how to select training samples is not the most essential element of constructing a dictionary for recognizing both clear and noisy patterns.

The aim of our study is to develop a discriminant function that can cope with different qualities of patterns and different kinds of noise. For this purpose, a discriminant function is proposed that can dynamically quantify noise and formulate the relationship between noise and the change in distribution. Experiments are done to compare the proposed function with other attempts.

The organization of the rest of the paper is as follows. In Section 2, a new discriminant function that includes revision matrix is proposed. In Section 3, experimental methods are described and the experimental results are discussed. Finally, Section 4 shows that the proposed revision matrix can be introduced to various discriminant functions, and those functions can be applied to the case of treating high dimensional feature vectors.

2 A New Discriminant Function

In the course of observation, it is very difficult to avoid the addition of noise. In the case that noise is mingled in some elements of a feature vector, while the other elements are noiseless, the standard deviation of the noisy element will become larger according to the degree of noise. In this section, a discriminant function reflecting change in distribution according to the noise is defined.

2.1 Definition

Here, the Mahalanobis distance is considered. Let $\boldsymbol{\mu}$ and Σ be the mean vector and the $n \times n$ covariance matrix, respectively. The squared Mahalanobis distance from $\boldsymbol{\mu}$ to $\boldsymbol{x}$ is defined as

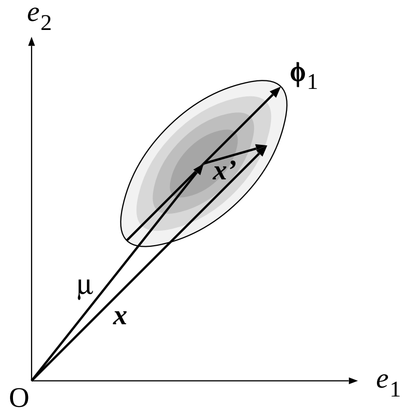
$$d^2 = (\boldsymbol{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}). \tag{1}$$

The squared Mahalanobis distance is abbreviated as the Mahalanobis distance below.

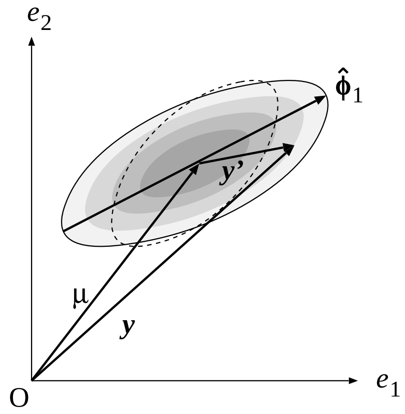
For simplicity, the case of two-dimensional normal distribution is discussed first. As shown in Fig. 1(a), e_1 and e_2 are the axes of the original coordinate. Let $\boldsymbol{\phi_1}$ be the eigenvector that corresponds to the first principal component. When b degrees of noise is added to the e_1-element while e_2-element is noiseless, we can observe that only the standard deviation of e_1-element becomes r_b times larger. Then the change in the distribution can be illustrated as Fig. 1(b).

In general, let $b(j)$ be the degree of noise added to the jth element of n-dimensional feature vector $\boldsymbol{x}$. Then we can observe that the mean vector is not changed and the standard deviation of jth element of $\boldsymbol{x}$ becomes $r_{b(j)}$ times larger where $r_{b(j)}$ is determined depending on the value $b(j)$. A diagonal matrix K is defined as

$$K = \begin{bmatrix} r_{b(1)} & & & 0 \\ & r_{b(2)} & & \\ & & \ddots & \\ 0 & & & r_{b(n)} \end{bmatrix}, \tag{2}$$



(a) Original distribution.



(b) Changed distribution.

Figure 1: Change in distribution.

which is called the revision matrix. For noiseless elements, say j, $r_{b(j)} = 1$.

Let $\boldsymbol{x}_i$ be a noiseless sample, and $\boldsymbol{y}_i$ be a sample with noise ($i = 1, 2, ..., N$). The mean vector $\boldsymbol{\mu}$ and unbiased covariance matrix Σ of noiseless samples are calculated as follows.

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i, \tag{3}$$

$$\Sigma = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^t. \tag{4}$$

The observation on the change of deviation by noise may be described as the following. Let

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{x}'_i, \tag{5}$$

and $\boldsymbol{y}_i$ be the corresponding noisy data such that

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{y}'_i. \tag{6}$$

If the jth element of $\boldsymbol{y}'_i$ is changed from $\boldsymbol{x}'_i$ as

$$y'_{ij} = r_{b(j)} x'_{ij}, \tag{7}$$

then the standard deviation $\hat{\sigma}_j$ of y'_{ij} is $r_{b(j)}$ times larger than that of x'_{ij}, since

$$\begin{aligned}
\hat{\sigma}_j &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y'_{ij})^2} \\
&= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_{b(j)}x'_{ij})^2} \\
&= r_{b(j)}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x'_{ij})^2}. \qquad (8)
\end{aligned}$$

From Eq. 7, $\boldsymbol{y}'_i$ can be written as

$$\boldsymbol{y}'_i = K\boldsymbol{x}'_i. \qquad (9)$$

The covariance matrix of noisy samples $\hat{\Sigma}$ is calculated using Eqs. 5, 6 and 9.

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^t \\
&= \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{y}'_i)(\boldsymbol{y}'_i)^t \\
&= \frac{1}{N-1}\sum_{i=1}^{N}(K\boldsymbol{x}'_i)(K\boldsymbol{x}'_i)^t \\
&= \frac{1}{N-1}\sum_{i=1}^{N}K(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^t K \\
&= K\Sigma K. \qquad (10)
\end{aligned}$$

In order to reflect the change in distribution, the following discriminant function that includes the revision matrix K is proposed.

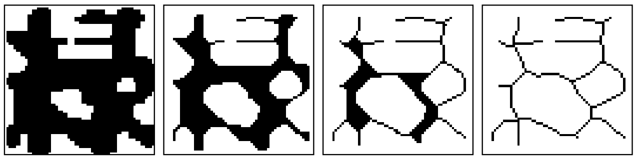
$$\begin{aligned}
\hat{d}^2 &= (\boldsymbol{x}-\boldsymbol{\mu})^t\hat{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
&= (\boldsymbol{x}-\boldsymbol{\mu})^t(K\Sigma K)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
&= (K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t\Sigma^{-1}(K^{-1}(\boldsymbol{x}-\boldsymbol{\mu})). \qquad (11)
\end{aligned}$$

Eq. 11 is called *Adaptive Mahalanobis distance*. The inverse matrix of K can easily be calculated since K is a diagonal matrix. Hence the computation time for recognition using Eq. 11 is not so large compared with recognition using Eq. 1.

2.2 *K* in Character Recognition

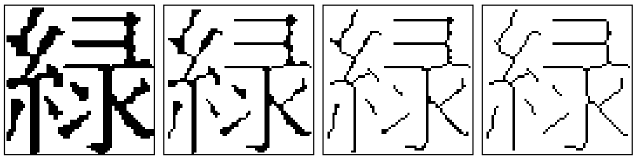
In this section, the method of how to decide the revision matrix K in a practical application is described. As the application, we consider the case of character recognition.

The authors have proposed an algorithm that detects blurred parts of character image and modify the discriminant function according to the detected blur [7]. Blurred part detection is done in the thinning process. Thinning is a repeating process of erasing a black pixel from boundaries of black pixels of character image. By scanning neighbor pixels around each black pixel, stroke width of character image is finally erased to one-pixel [8], [9]. Figs. 2(a) and 2(e) are examples of a normalized blurred image and a clear image, respectively. Thinning process of these images are shown by Figs. 2(b)∼2(d) and Figs. 2(f)∼2(h), respectively. In order to get completely thinned image of Fig. 2(d), ten times are needed to erase pixels on boundaries, while Fig. 2(h) requires only six times. If the repeating times in thinning process is limited to the number that a clear image is completely thinned, obviously it will not be enough for a blurred one and blurred area will remain. The pre-experimental result has shown usually six is the optimum number of times for thinning a clear image. An area can be regarded as a blurred one if any line width in the area is more than one-pixel after being thinned six times. Black region of Fig. 3 displays the detected blurred region of Fig. 2(c). After detecting the blurred part of input image, the *degree of blur* of each area is examined. Then, the standard deviation of the feature vectors of the areas with b degrees of blur and that of the areas with 0 degree is calculated. The ratio of these standard deviations is denoted as r_b. By the ratio r_b, the revision matrix K is constructed.



(a) Normalized (b) 3 times erased (c) 6 times erased (d) Skeletonized

(A) Blurred image.



(e) Normalized (f) 1 time erased (g) 2 times erased (h) Skeletonized

(B) Clear image.

Figure 2: Thinning process of blurred image and clear image.

Figure 3: Detected blurred area.

3 Experiments

We assume that revising a discriminant function according to detected noise is much more important than gathering noisy training samples for dictionary. In order to verify this assumption, four kinds of experiments are done to compare the effectiveness of two different dictionaries that are constructed with clear data and noisy data, and also, to confirm the ability of the proposed discriminant function of Eq. 11 and the Mahalanobis distance of Eq. 1.

Here, two appropriate categories, called Category A and Category B, which have n-dimensional normal distribution are given. Their distributions are assumed to be $N(\boldsymbol{\mu}_A, \Sigma_A)$ and $N(\boldsymbol{\mu}_B, \Sigma_B)$.

3.1 Testing Data

By generating random numbers, n-dimensional N_E vectors whose distribution is $N(\boldsymbol{\mu}_A, \Sigma_A)$ are obtained, and this set of vectors is called E_A. Another set E_B is produced in the same way that includes n-dimensional N_E vectors whose distribution is $N(\boldsymbol{\mu}_B, \Sigma_B)$. Here, $n = 196$ and $N_E = 10000$. The value of n is selected from the number of dimensions of a feature vector, called the Improved Directional Element Feature [10], that is proposed for Chinese character recognition. Using sample character images, $\boldsymbol{\mu}_A$, Σ_A, $\boldsymbol{\mu}_B$ and Σ_B are calculated. Moreover, a set of vectors E'_A is obtained by adding noise to each vector in E_A as the following way. Let the mean vector of set E_A be

$$\boldsymbol{\mu}_A = (\mu_{A1}, \mu_{A2}, ..., \mu_{An}),$$

and the diagonal components of Σ_A be σ^2_{A1}, σ^2_{A2},..., σ^2_{An}. A vector $\boldsymbol{v}$ in E_A is denoted as

$$\boldsymbol{v} = (v_1, v_2, ..., v_n) \in E_A.$$

Select k elements randomly from n elements of $\boldsymbol{v}$. Denote these selected elements as $v_{a_1}, v_{a_2}, ..., v_{a_k}$. The element $v_{a_i} (1 \leq i \leq k)$ is replaced by a value whose distribution is one-dimensional normal $N(\mu_{Aa_i}, (r\sigma_{Aa_i})^2)$ and obtained by random numbers. Here, r represents the *ratio* of standard deviations mentioned in Section 2.1 as r_b. In the same way as making E'_A, noise is added to each vector in E_B, and a new set E'_B is generated.

In the following, how distribution changes if different levels of noise is added to certain elements of vector $\boldsymbol{v}$ is investigated. By using Fisher's linear discriminant [11], elements of E_A and E_B are projected onto the line that discriminates these categories optimally. The histogram on this line is expressed in Fig. 4(a). This figure shows that E_A and E_B can be discriminated by linear discriminant function. Moreover, error rate using Eq. 1 is estimated by resubstitution method [12], which gives the lower bound of the true error rate. Mean vectors and covariance matrices are calculated from the sets E_A and E_B, then samples in the same sets E_A and E_B are classified by Eq. 1. The rates of misclassification is 0%.

After producing E'_A and E'_B under the condition that $k = 4$ and $r = 10.0$, elements of E'_A and E'_B are projected onto the line that discriminates these categories optimally. The histogram on this line is illustrated in Fig. 4(b). Furthermore, Fig. 4(c) shows the terrible case of noise when $k = 12$ and $r = 20.0$. In these two cases, E'_A and E'_B cannot be discriminated linearly. Error rates of (b) and (c) are estimated using Eq. 1 by resubstitution method, which are 7.51% and 21.4%. These results show that E'_A and E'_B cannot be discriminated by the Mahalanobis distance either.

3.2 Training Data

Training data is similarly generated as the testing data. By producing random numbers, N_T vectors with distribution $N(\boldsymbol{\mu}_A, \Sigma_A)$ and N_T vectors with distribution $N(\boldsymbol{\mu}_B, \Sigma_B)$ are obtained. These sets are different from E_A and E_B, and these are denoted as T_A and T_B. Here, $N_T = 10000$. T'_A and T'_B are obtained by adding noise to each vector in T_A and T_B in the same way as E'_A and E'_B, respectively. The pair of noiseless sets (T_A, T_B) and the pair of noisy sets (T'_A, T'_B) are adopted as training data. Here, a set of mean vector and covariance matrix calculated from training data is called dictionary. The dictionaries of the noiseless sets T_A and T_B are ($\tilde{\boldsymbol{\mu}}_{T_A}$, $\tilde{\Sigma}_{T_A}$) and ($\tilde{\boldsymbol{\mu}}_{T_B}$, $\tilde{\Sigma}_{T_B}$). The dictionaries of noisy sets T'_A and T'_B are ($\tilde{\boldsymbol{\mu}}_{T'_A}$, $\tilde{\Sigma}_{T'_A}$) and ($\tilde{\boldsymbol{\mu}}_{T'_B}$, $\tilde{\Sigma}_{T'_B}$).

3.3 Methods

The noisy testing data sets E'_A and E'_B are discriminated by the traditional and Adaptive Mahalanobis distance. Two kinds of functions are used as the discriminant functions for each of the dictionaries of noiseless and noisy training data. Therefore, four kinds of experiments are done in total. In these experiments, various values of k and r are examined. The revision matrix K can be defined by the parameter r, which actually is the level of added noise. The revision matrix K is defined for each category A and B, and its effectiveness of discriminating noisy patterns is determined.

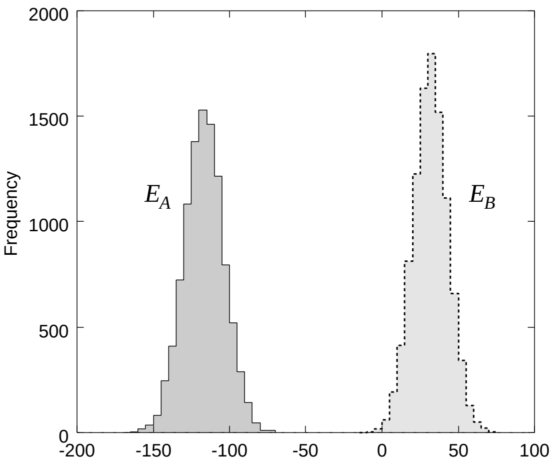
Table 1 shows the combination of mean vectors, covariance matrices, and discriminant functions for each experimental method. Qualitative meaning of each method is described as follows.

C-TM: Clear data – Traditional Mahalanobis.

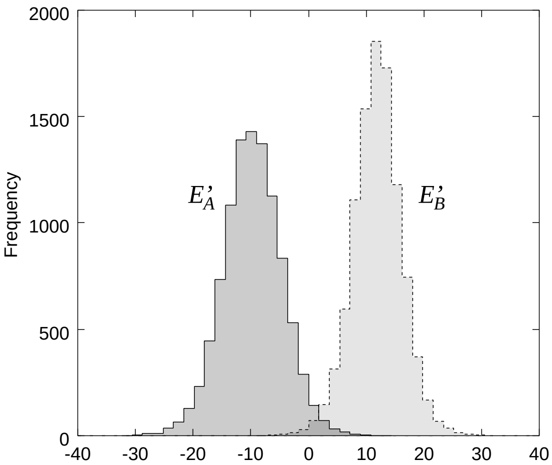
Dictionaries are made with noiseless data, and the traditional Mahalanobis distance (Eq. 1) is used for discrimination.

N-TM: Noisy data – Traditional Mahalanobis.

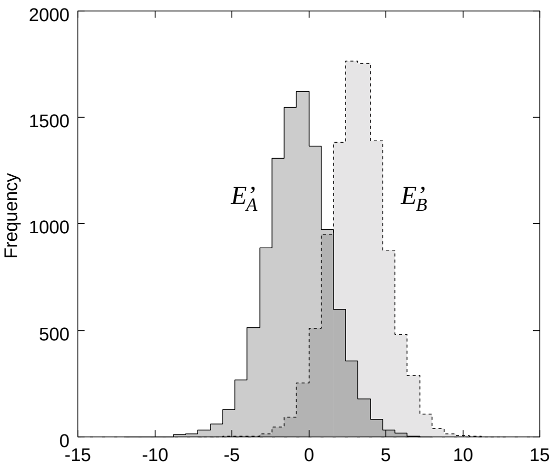
Dictionaries are made with noisy data, and the traditional Mahalanobis distance (Eq. 1) is used for discrimination.



(a) E_A and E_B.



(b) E'_A and E'_B ($k = 4$, $r = 10.0$).



(c) E'_A and E'_B ($k = 12$, $r = 20.0$).

Figure 4: Distribution of testing data with different degree of noise.

Table 1: Four kinds of experiments.

Method	Function	Dictionary of Category A	Dictionary of Category B
C-TM	Eq. 1	$(\tilde{\boldsymbol{\mu}}_{T_A}, \tilde{\Sigma}_{T_A})$	$(\tilde{\boldsymbol{\mu}}_{T_B}, \tilde{\Sigma}_{T_B})$
N-TM	Eq. 1	$(\tilde{\boldsymbol{\mu}}_{T'_A}, \tilde{\Sigma}_{T'_A})$	$(\tilde{\boldsymbol{\mu}}_{T'_B}, \tilde{\Sigma}_{T'_B})$
C-AM	Eq. 11	$(\tilde{\boldsymbol{\mu}}_{T_A}, \tilde{\Sigma}_{T_A})$	$(\tilde{\boldsymbol{\mu}}_{T_B}, \tilde{\Sigma}_{T_B})$
N-AM	Eq. 11	$(\tilde{\boldsymbol{\mu}}_{T'_A}, \tilde{\Sigma}_{T'_A})$	$(\tilde{\boldsymbol{\mu}}_{T'_B}, \tilde{\Sigma}_{T'_B})$

C-AM: Clear data – Adaptive Mahalanobis.

Dictionaries are made with noiseless data, and proposed discriminant function (Eq. 11) is used for discrimination.

N-AM: Noisy data – Adaptive Mahalanobis.

Dictionaries are made with noisy data, and proposed discriminant function (Eq. 11) is used for discrimination.

3.4 Experimental Results

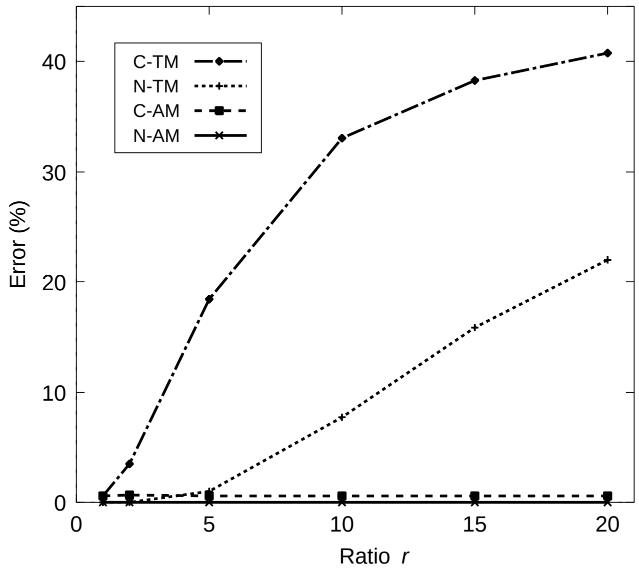
Fig. 5 shows the error rates of classification. Fig. 5(a) represents the results that the number of noisy elements is fixed to four and various values of r are adopted. Fig. 5(b) illustrates the results that the ratio r is fixed to 10.0 and various values of k are used.

The figures have shown that the error rates of C-TM and N-TM increase with k or r becoming larger. The error rate of C-TM is relatively high even for small k and r. Comparing with the result of C-TM, the error rate of N-TM is much smaller, that means it is rather effective to make dictionaries with noisy patterns for recognizing image with noise.

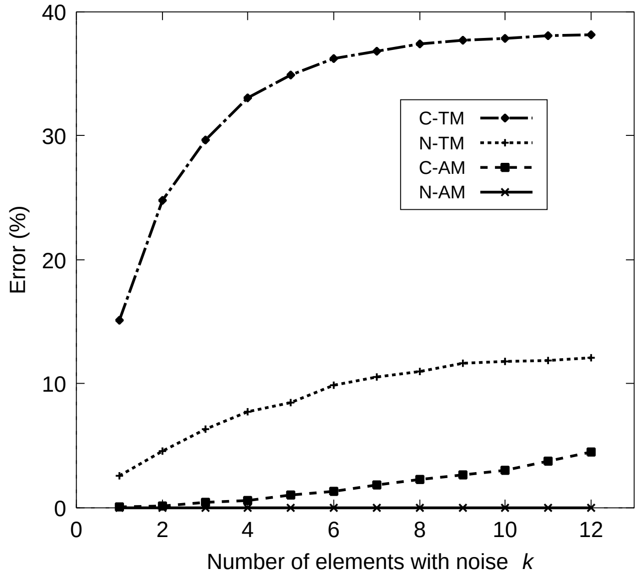
The error rate of C-AM increases with k becoming larger, while it is not sensitive to r. In other words, error rate increases as the number of noisy elements becomes larger but it is not sensitive to degree of noise. If r is small, the result of N-TM is better than C-AM. However, if $r \geq 5$, the error rate of C-AM is smaller than that of N-TM.

Moreover, the error rate of N-AM is 0% under any parameters. As described in Section 3.1, error rate of E'_A and E'_B using the traditional Mahalanobis distance under the condition that $k = 12$ and $r = 20.0$ is more than 21.4%. (Distribution of this case can be seen in Fig. 4(c).) However, the error rate of N-AM here is 0% even if $k = 12$ and $r = 20.0$.

These results confirm that it is not enough to make a dictionary with noisy data for recognizing extremely noisy patterns. The more important matter is to detect the noise and to revise the discriminant function. These experimental results clearly prove that the proposed method supplements the existing statistical methods and has high ability of classification.



(a) The number of noisy elements is fixed to $k = 4$.



(b) The ratio is fixed to $r = 10.0$.

Figure 5: Experimental results of traditional and Adaptive Mahalanobis distance.

4 Application to the Other Discriminant Functions

Although the Mahalanobis distance is a very valid criterion, in the case of treating high dimensional vectors, e.g., recognizing character patterns, it is known that the Mahalanobis distance has disadvantages such as the large computation time and bad influence caused by limited samples [13], [14], [15]. To resolve these problems, various modifications of the Mahalanobis distance are proposed. In this section, it is shown that the revision matrix K can be introduced to those functions. Typical three are Quasi-Mahalanobis distance [16] (QMD), Modified Mahalanobis distance [17] (MMD), and Simplified Mahalanobis distance [18] (SMD). To explain these functions, Eq. 1 is rewritten as

$$d^2 = \sum_{i=1}^{n} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 \tag{12}$$

$$= \sum_{i=1}^{m} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 + \sum_{i=m+1}^{n} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2. \tag{13}$$

Here, n is the number of dimensions of feature vector, and λ_i is the ith eigenvalue of Σ sorted by descending order, and $\boldsymbol{\phi}_i$ is the eigenvector that corresponds to λ_i.

The QMD replaces λ_i of the second term in Eq. 13 with λ_{m+1}. In the MMD, the second term in Eq. 13 is neglected, and instead of λ_i in the first term, $\lambda_i + b$ is used, where b is a bias ($b = 5$). The SMD replaces λ_i of the second term in Eq. 13 with the mean value of λ_i $(i = m+1, ..., n)$. Denote the QMD, the MMD and the SMD as d_Q, d_M and d_S, these are written as,

$$d_Q^2 = \sum_{i=1}^{m} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 + \frac{1}{\lambda_{m+1}} \sum_{i=m+1}^{n} ((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2, \tag{14}$$

$$d_M^2 = \sum_{i=1}^{m} \frac{1}{\lambda_i + b}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2, \tag{15}$$

$$d_S^2 = \sum_{i=1}^{m} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 + \sum_{i=m+1}^{n} \frac{1}{\lambda}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 \tag{16}$$

$$= \sum_{i=1}^{m} \frac{1}{\lambda_i}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 + \frac{1}{\lambda}\left\{ ||\boldsymbol{x}-\boldsymbol{\mu}||^2 - \sum_{i=1}^{m}((\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\phi}_i)^2 \right\}, \tag{17}$$

where

$$\lambda = \frac{1}{n-m} \sum_{i=m+1}^{n} \lambda_i \tag{18}$$

$$= \frac{1}{n-m} \left\{ \mathrm{tr}\Sigma - \sum_{i=1}^{m} \lambda_i \right\}. \tag{19}$$

Here, $\mathrm{tr}\Sigma$ denotes the trace of Σ. Effectiveness of these functions for character recognition has been shown [17], [18] with the large handwritten Japanese and Chinese character database ETL9B [19]. So, the revision matrix K should be introduced for such functions rather than the original Mahalanobis distance for practical applications.

Denote the QMD, the MMD and the SMD with K as $\hat{d}_Q$, $\hat{d}_M$ and $\hat{d}_S$, respectively. These are written as follows.

$$\begin{aligned}\hat{d}_Q^2 &= \sum_{i=1}^{m} \frac{1}{\lambda_i}((K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t \boldsymbol{\phi}_i)^2 \\ &\quad + \frac{1}{\lambda_{m+1}} \sum_{i=m+1}^{n} ((K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t \boldsymbol{\phi}_i)^2, \qquad (20)\end{aligned}$$

$$\hat{d}_M^2 = \sum_{i=1}^{m} \frac{1}{\lambda_i + b}((K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t \boldsymbol{\phi}_i)^2, \qquad (21)$$

$$\begin{aligned}\hat{d}_S^2 &= \sum_{i=1}^{m} \frac{1}{\lambda_i}((K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t \boldsymbol{\phi}_i)^2 \\ &\quad + \frac{n-m}{\mathrm{tr}\Sigma - \sum_{i=1}^{m} \lambda_i} \Big\{ ||K^{-1}(\boldsymbol{x}-\boldsymbol{\mu})|| \\ &\quad - \sum_{i=1}^{m} ((K^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))^t \boldsymbol{\phi}_i)^2 \Big\}. \qquad (22)\end{aligned}$$

The computation time of each of the above functions is not so large compared with the function without K.

5 Conclusions

For most statistical methods of pattern recognition, to give the exact expression of distribution of feature vectors is the first step of accurate recognition. However, in the case that noise is included in a pattern, the feature vector will be quite different from the same kind of clear pattern. Since noise occurs irregularly and accidentally, it is thought difficult to estimate a dictionary that can cope with all kinds of noise by using a lot of kinds of noisy training patterns.

In this paper, considering the characteristic of noise, a new discriminant function for recognizing noisy pattern is presented. The revision matrix reflecting the change in distribution of category caused by the detected noise is constructed. Very satisfactory performance of the modified discriminant function has been shown by experiments with artificial data. These results not only indicate the effectiveness of the proposed function, but also confirm that it is important to detect the noise and to revise the discriminant function for noisy pattern recognition.

Moreover, it is shown that the revision matrix can be easily applied to various discriminant functions. In some cases, especially when the dimension of the feature vector is high, the modified version of the Mahalanobis distance is more effective for pattern recognition than the original one. These discriminant functions can be used for practical noisy pattern recognitions such as character recognition, speech recognition, face recognition, and so on.

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] J.D. Hobby and T.K. Ho, “Enhancing Degraded Document Images via Bitmap Clustering and Averaging,” *Proc. Fourth Int’l Conf. Document Analysis and Recognition (ICDAR’97)*, pp.394–400, Aug. 1997.

[2] T.R. Chou and F. Chang, “Optical Chinese Character Recognition for Low-Quality Document Images,” *Proc. Fourth Int’l Conf. Document Analysis and Recognition (ICDAR’97)*, pp.608–611, Aug. 1997.

[3] C. Rodríguez, J. Muguerza, M. Navarro, A. Zárate, J.I. Martín, and J.M. Pérez, “A Two-Stage Classifier for Broken and Blurred Digits in Forms,” *Proc. 14th Int’l Conf. Pattern Recognition (ICPR’98)*, pp.1101–1105, Aug. 1998.

[4] K. Sawa, S. Tsuruoka, T. Wakabayashi, F. Kimura, and Y. Miyake, “Low Quality String Recognition for Factory Automation,” *Proc. Fourth Int’l Conf. Document Analysis and Recognition (ICDAR’97)*, pp.475–478, Aug. 1997.

[5] J. Kosai, M. Okamoto, K. Kato, and K. Yamamoto, “A Proposal of Character Recognition Method on Low Resolution,” *Proc. Meeting on Image Recognition and Understanding (MIRU’98)*, vol.2, pp.257–262, July 1998.

[6] M. Sawaki, H. Murase, and N. Hagita, “Character Recognition in Bookshelf Images by Automatic Template Selection,” *Proc. 14th Int’l Conf. Pattern Recognition (ICPR’98)*, pp.1117–1120, Aug. 1998.

[7] S. Omachi, F. Sun, and H. Aso, “Precise Recognition of Blurred Chinese Characters by Considering Change in Distribution,” *Proc. of The 10th Scandinavian Conference on Image Analysis (SCIA’97)*, pp501–506, June 1997.

[8] H. Aso, “Thinning Algorithm Suitable for Parallel Processing,” *Trans. IEICE*, vol.J76-D-II, no.9, pp.2148–2150, Sept. 1993.

[9] C.J. Hilditch, “Linear Skeleton from Square Cupboards,” In: *Machine Intelligence 6*, B. Meltzer & D. Michie, Eds., Edinburgh University Press, pp.403–420, 1969.

[10] 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.

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

[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition, Second Edition*, Academic Press, Inc., 1990.

[13] 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.

[14] 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.

[15] 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.

[16] M. Kurita, S. Tsuruoka, S. Yokoi, and Y. Miyake, "Handprinted "Kanji" and "Hiragana" character recognition using weighting direction index histograms and quasi-Mahalanobis distance," *IEICE Technical Report*, PRL82-79, pp.105–112, Jan. 1983.

[17] 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.

[18] 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.

[19] 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.