Fast and Precise Discriminant Function Considering Correlations of Elements of Feature Vectors and Its Application to Character Recognition

Fang SUN[†], Shin'ichiro OMACHI[††], Nei KATO[†††], and Hirotomo ASO[††], *Members*

[†] Faculty of Science and Technology, Tohoku Bunka Gakuen University, Sendai-shi, 981-8550 Japan
[††] Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan
[†††] Graduate School of Information Sciences, Tohoku University, Sendai-shi, 980-8579 Japan

SUMMARY

During the late few years, research in recognition of handwritten Chinese and Japanese characters has matured significantly. However, in order to obtain high recognition rate, most character recognition systems have paid extremely expensive computational cost. For high performance character recognition systems, how to reduce the expensive computational cost is a very important problem now. Discriminant function is a very important factor for precise pattern recognition. The Mahalanobis distance is considered as an effective function. However, to calculate the Mahalanobis distance precisely, extremely large number of training samples are needed. In this paper, by investigating the relationship of elements of feature vector, a new discriminant function, called vector-divided Mahalanobis distance, is proposed. With the proposed method, high recognition performance can be obtained with less computational cost. Because the proposed method partitions high dimensional feature vector into several small number dimensional vectors, the ratio of the number of training samples to the number of dimensions becomes larger. This method is especially effect in the case of lack of training samples. The effectiveness of the proposed method is shown by the experimental results with the database ETL9B.

Keywords: character recognition, Mahalanobis distance, feature vector, ETL9B, vector-divided Mahalanobis distance.

1 Introduction

During the late few years, research in recognition of handwritten characters with database ETL9B[1], which is the largest public handwritten character database in Japan, has matured significantly[2, 3, 4]. More recently, Sun et al. have developed a handwritten character recognition system by using Modified Mahalanobis distance, and the average recognition rate is 98.24%[5]. Wakabayashi et al. obtain the high recognition rate of 99.05% using compression of higher dimensional features[3]. Suzuki et al. have proposed the image transformation method based on partial inclination detection (TPID)[4]. The recognition rate of 99.08% has been obtained. However, in order to obtain high recognition rates, most character recognition systems have paid expensive computational cost. For high performance character recognition systems, how to reduce the expensive computational cost becomes a very important problem now.

Effective discriminant function is a very important factor for precise pattern recognition system. There are a few well-known discriminant functions, such as the Euclidean distance, weighted Euclidean distance, city block distance, and the Mahalanobis distance. The Mahalanobis distance is derived by a probability density function of multivariate normal distribution, and so it is considered as an appropriate function if the distribution of samples is multivariate normal and sample patterns are enough. However, compared with the number of dimensions, the training samples are always not enough. For this reason, covariance matrix usually cannot be estimated accurately.

In the case of Chinese and Japanese character recognition, the number of dimensions of feature vectors are usually very large. The Improved Directional Element Feature[5] has 196 dimensions. The number dimensions of Weighted Direction Code Histogram[3] is 392, and that of Extended Peripheral Direction Contributivity[6, 7] is 1536. In the case of using high dimensional feature vectors, since there are not enough training samples usually, the estimation error will increase in eigenvalue expansion, especially in higher dimensions. Furthermore, there are other disadvantages of the Mahalanobis distance such as the computation time will reach $O(n^2)$ for n-dimensional feature vectors. The former problem is much more serious. From our simulation results, if the difference between computed Mahalanobis distance and true Mahalanobis distance is expected to be limited under 10%, the ratio of the number of samples to the number of dimensions must be greater than ten. That means if the number of dimensions is 1536, for one kind of character, over 15000 training samples are necessary. Since there are 2965 kinds of characters are included in the 1st JIS, forty-five million of training samples are needed. It is an impossible number to gather.

To resolve the above problems, some revisions of the Mahalanobis distance are proposed, such as Quasi-Mahalanobis distance (QMD)[10] and Modified Mahalanobis distance (MMD)[2]. Modified Quadratic Discriminant Function (MQDF) is also known that it can avoid the bad influence caused by a finite number of samples, and it can save the computation time by using a constant value instead of eigenvalues of higher dimensions[3].

In this paper, a new approach is considered. In order to approximate the Mahalanobis distance efficiently, a new discriminant function, called vector-divided Mahalanobis distance, is proposed. With the proposed method, high recognition performance can be obtained with less computational cost. Because the proposed method partitions 196-dimensional feature vector into several small number dimensional vectors, the ratio of the number of training samples to the number of dimensions becomes larger. Moreover, the computation time is reduced inverse proportion with division number. A method of dimension compression is proposed in [11]. However, the aim of the method is only to reduce computation time for recognition. The aim of our method is not only to reduce computational cost, but also to approximate the Mahalanobis distance without changing the original feature vectors. In order to achieve our aim, properties of covariance matrix are investigated. The validity of partitioning the covariance matrix is considered. Moreover, in order to suit the division of covariance matrix, component exchange algorithm is proposed. Experiments with the database ETL9B are done to show the effectiveness of the proposed method. In this paper, the Improved Directional Element Feature[5] with 196 dimensions is used to be the feature vector.

2 Feature Vector and Discriminant Function

2.1 Improved Directional Element Feature

The Improved Directional Element Feature[5] is used as the feature vector in this paper. It is calculated as follows. As shown in Fig. 1, first an input image is normalized to 64×64 dots, and the 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 ($= 4 \times 49$) dimensions.

2.2 Definition of Modified Mahalanobis Distance

Let n be the number of dimensions of the feature vector, in the case of the Improved Directional Element Feature, $n = 196$. $\boldsymbol{\mu}$ and Σ denote the mean vector and the n_by_n covariance matrix, respectively. The squared Mahalanobis distance of vector $\boldsymbol{x}$ is defined as

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

The squared Mahalanobis distance is abbreviated as Mahalanobis distance below.

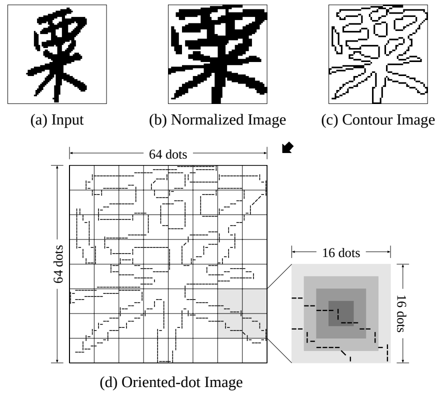

Figure 1: Improved directional element feature.

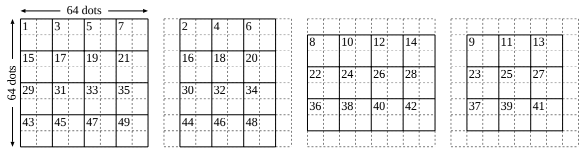

Figure 2: 49 sub-areas.

Let λ_k be the k-th eigenvalue of Σ sorted by descending order, and $\boldsymbol{\phi}_k$ be the eigenvector that corresponds to λ_k. If $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ is defined as $y_k = (\boldsymbol{x} - \boldsymbol{\mu})^t \cdot \boldsymbol{\phi}_k$, Eq.(1) can be rewritten as

$$d^2(\boldsymbol{x}) = \sum_{k=1}^{n} \frac{1}{\lambda_k} (\boldsymbol{x} - \boldsymbol{\mu}, \boldsymbol{\phi}_k)^2. \tag{2}$$

In the case of using the Mahalanobis distance, since there are not enough training samples usually, the covariance matrix cannot be calculated accurately. The estimation errors will increase in eigenvalue expansion, especially in the higher dimensions. In order to avoid this problem, the Modified Mahalanobis distance[5] is proposed as a discriminant function that can approximate the Mahalanobis distance by using small number m ($m < n$) of dimensions.

Here, Eq.(2) can be approximated as

$$d_M^2(\boldsymbol{x}) = \sum_{k=1}^{m} \frac{1}{\lambda_k + b} (\boldsymbol{x} - \boldsymbol{\mu}, \boldsymbol{\phi}_k)^2. \tag{3}$$

In Eq.(3), instead of λ_k in Eq.(2), $\lambda_k + b$ is employed, where b is a bias. $d_M^2(\boldsymbol{x})$ is called Modified Mahalanobis distance, or MMD. The bias is introduced to decrease the errors in eigenvalues caused by limited number of samples. According to the experimental results of the ETL9B, the MMD has shown quite good performance[2, 4]. However, because 150 of dimensions are necessary, the computational cost of the MMD is still very expensive.

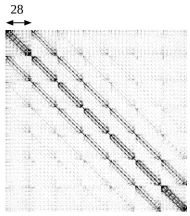

Figure 3: Mean covariance matrix of all the categories.

3 Division of Feature Vector

3.1 Property of Covariance Matrix

As described above, the Improved Directional Element Feature is produced by counting the number of each four directions in every 49 sub-areas. Therefore, it is thought that the same kind of directional elements of adjoining sub-areas have a high correlation, while elements of distant sub-areas have a low correlation. Here No.21 $\sim$ 200 sets of the ETL9B[1], are used to check the relationships among elements of the Improved Directional Element Feature. First, the covariance matrix of each kind of character included in the ETL9B (3036 kinds of characters in total) is calculated from the feature vectors of training samples. Then, the mean covariance matrix Σ_0 is computed. Fig.3 has shown the 196×196 covariance matrix Σ_0. The depth of black shows the absolute value of elements.

As shown in this figure, larger absolute values concentrate near the diagonal of the mean covariance matrix. Furthermore, the absolute value of covariance of each pair of elements at interval of 4 or 28 in a row or a column is comparatively large. That means these pairs of elements have strong correlations[2]. As shown in Fig.2, four kinds of directional elements are defined in each sub-area. The number of the same kind of directional element of two horizontally overlapped sub-areas has an interval of 4, and the number of the same kind of directional element of two vertically overlapped sub-areas has an interval of 28. Obviously, the overlapped sub-areas have strong correlations.

3.2 Vector-divided Mahalanobis Distance

Suppose covariance matrix is a block diagonal matrix that only consists of M numbers of $K \times K$ matrices ($n = M \times K$), while the other components are zero. In the other words, as shown in Eq.(4), suppose $\Sigma_1, \Sigma_2, ..., \Sigma_M$ are $K \times K$ square matrices and the other components are zero.

[1]These sets are the same sets as the sets used for the experiments in sections 4.1 $\sim$ 4.3.

[2]Because covariance matrices are calculated for the Mahalanobis distance, instead of using correlation matrices, the covariance matrices are used to check degrees of correlation.

$$\Sigma = \begin{bmatrix} \Sigma_1 & & & 0 \\ & \Sigma_2 & & \\ & & \ddots & \\ 0 & & & \Sigma_M \end{bmatrix}. \tag{4}$$

In this case, Eq.(1) can be written as

$$\begin{aligned} d^2(\boldsymbol{x}) &= (\boldsymbol{x}-\boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \\ &= \sum_{i=1}^{M} (\boldsymbol{x}_i-\boldsymbol{\mu}_i)^t \Sigma_i^{-1} (\boldsymbol{x}_i-\boldsymbol{\mu}_i) \\ &= \sum_{i=1}^{M} \sum_{k=1}^{K} \frac{1}{\lambda_{ik}} (\boldsymbol{x}_i-\boldsymbol{\mu}_i, \boldsymbol{\phi}_{ik})^2, \end{aligned} \tag{5}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{\mu}_i$ are K-dimensional vectors constructed by the $(K(i-1)+1)$th $\sim$ (Ki)th components of $\boldsymbol{x}$ and $\boldsymbol{\mu}$. λ_{ik} and $\boldsymbol{\phi}_{ik}$ are the kth eigenvalue and eigenvector of Σ_i. The computational cost is $O(K^2M) = O(nK)$. The cost is reduced to $1/M$ compared with Eq.(1) whose cost is $O(n^2)$. Here, as the same as the MMD, bias is added to the denominator of Eq.(5).

$$d^2(\boldsymbol{x}) = \sum_{i=1}^{M} \sum_{k=1}^{K} \frac{1}{\lambda_{ik}+b} (\boldsymbol{x}_i-\boldsymbol{\mu}_i, \boldsymbol{\phi}_{ik})^2. \tag{6}$$

The method of approximating the original covariance matrix by Eq.(4) is considered. First, instead of computing the $n \times n$ covariance matrix, feature vectors are divided into M numbers of K-dimensional vectors. Then M numbers of $K \times K$ matrices are computed. Finally, the approximation of the Mahalanobis distance is calculated using Eq.(6). This approximated Mahalanobis distance is called vector-divided Mahalanobis distance. By introducing this approximation, in the case of using the same number of training samples, the ratio of number of training samples to the number of dimensions becomes M times lager than using Eq.(1). It means more trustworthy covariance matrix can be expected.

In order to confirm the validity of the vector-divided Mahalanobis distance, simulation is done. Here, using the covariance matrix Σ_0 calculated in section 3.1, training samples are produced following the n-dimensional normal distribution $N(\mathbf{0}, \Sigma_0)$ that is supposed to be the true distribution. The details of simulation are as follows. First, by producing random numbers, N_t vectors which has the n-dimensional normal distribution $N(\mathbf{0}, \Sigma_0)$ are obtained to be training vectors. Each vector is separated into M numbers of K-dimensional vectors. Then, the covariance matrix $\hat{\Sigma}_i$ and the mean vector $\hat{\boldsymbol{\mu}}_i$ are calculated from the K-dimensional vectors. Eigenvalues and eigenvectors are computed using the $\hat{\Sigma}_i$ $(i = 1, ..., M)$. Furthermore, using Σ_0, other N_e vectors are randomly obtained to be evaluation vectors. The vector-divided Mahalanobis distance d^2 of each evaluation vector is computed. Suppose

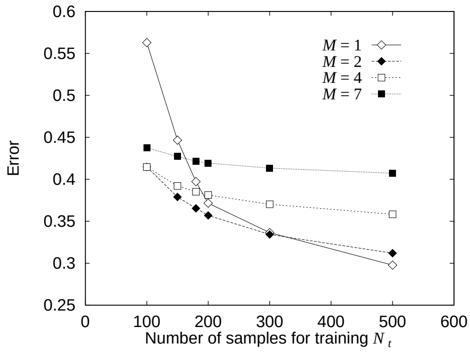

Figure 4: Result of simulation.

the distance d^2_{true} obtained from (1) with Σ_0 is the true value of the Mahalanobis distance. As it is known the expectation of the Mahalanobis distance will be different if numbers of training samples and dimensions have changed[13], the values d^2 and d^2_{true} of each N_e numbers of evaluation vectors are normalized to be the distribution of mean 0 and variance 1. The normalized values are denoted as $\tilde{d}^2$ and $\tilde{d}^2_{true}$. The error e of $\tilde{d}^2$ and $\tilde{d}^2_{true}$ is given as $e = |\tilde{d}^2 - \tilde{d}^2_{true}|$. The average of e among N_e vectors is calculated.

Fig.4 shows the errors of the two evaluation functions computed by various N_t, with $N_e = 10000$ and $M = 2, 4, 7$. In the same figure, the result calculated by the original function of the Mahalanobis distance but added bias b into the denominator, namely the result of the case of $M = 1$, is also displayed. In the case of $N_t < n$, because of the small sample size problem[14], the numbers of eigenvalues and eigenvectors are $N_t - 1$. Here, if $N_t < n$, $m = N_t - 1$, while if $N_t \geq n$, $m = n$. According to [5], $b = 5$. As shown in Fig.4, the errors of the original Mahalanobis distance are smaller if the numbers of training samples are large. However, if the number of training samples is small, especially if the number of training samples is smaller than the number of dimensions, the errors of the vector-divided Mahalanobis distance become smaller than that of the original Mahalanobis distance. If the number of training samples is $100 \sim 300$, the smallest error rate is obtained in the case of $M = 2$.

3.3 Exchange of Covariance Matrix Components

As shown in Fig.3, the covariance matrix calculated from the Improved Directional Element Feature is quite different from the block diagonal matrix denoted as Eq.(4). Obviously, if the values of non-diagonal components are not zero, they will cause negative effect. In order to solve this problem, components of covariance matrix are exchanged to approach the block diagonal matrix.

As shown in Fig.5, if the ith $(1 \leq i \leq K)$ row of a covariance matrix is exchanged with

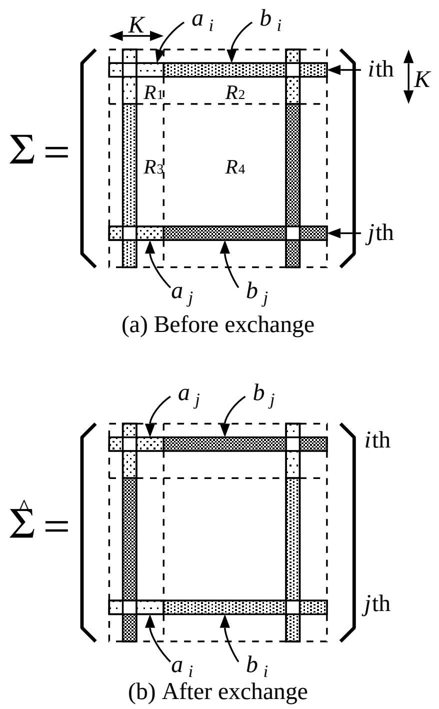

Figure 5: Exchange of elements of a covariance matrix.

the jth row ($K+1 \leq j$), the ith column and the jth column are exchanged simultaneously. As shown in Fig.5(a), four regions of the covariance matrix are defined as $R_1 \sim R_4$. The sum of the absolute values of $1 \sim K$ components that belong to the ith row is denoted as a_i. The sum of the absolute values of after the Kth components is denoted as b_i. In the same way, the sum of the absolute values of $1 \sim K$ components that belong to the jth row is denoted as a_j. The sum of the absolute values of after the Kth components is denoted as b_j. Here, the components of s_{ii}, s_{jj} and $s_{ij}(= s_{ji})$ are not included. Denote the sums of the absolute values of the ith row, the jth row, the ith column and the jth column in regions R_2 and R_3 are S_2 and S_3 before exchanging, S_2 and S_3 have the same value as

$$S_2 = S_3 = a_j + b_i + s_{ij}. \tag{7}$$

Moreover, denote the same kinds of sums after exchanging as $\hat{S}_2$ and $\hat{S}_3$. $\hat{S}_2$ and $\hat{S}_3$ also have the same value as

$$\hat{S}_2 = \hat{S}_3 = a_i + b_j + s_{ij}, \tag{8}$$

(See Fig.5(b)). In order to approach the block diagonal matrix shown in Eq.(4), the condition of $\hat{S}_2 < S_2$ must be satisfied. From the condition,

$$a_i - b_i \;\; < \;\; a_j - b_j, \tag{9}$$

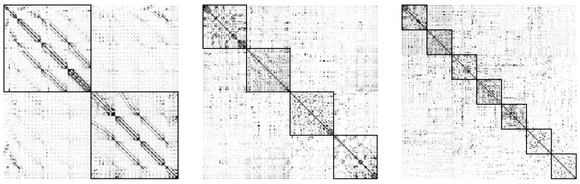

Figure 6: Exchanged covariance matrices.

is obtained. Moreover, based on property of covariance matrix, $s_{ii} + s_{jj} \geq 2s_{ij}$ holds. Therefore, in the case that

$$(a_i + s_{ii}) - (b_i + s_{ij}) < (a_j + s_{ji}) - (b_j + s_{jj}), \tag{10}$$

Eq.(9) holds.

The component exchange algorithm is the following. For each row of the covariance matrix, difference between the sum of the absolute values of $1 \sim K$ components and the sum of the absolute values after K components are calculated. K numbers of rows with larger values of difference are moved to $1 \sim K$th rows. Other rows with smaller values of difference are placed at after the Kth row. This operation is continued till convergence or certain number of times of repetition. As a result, the matrix with smaller values of components in region R_2 is obtained. Continuously, the region R_4 is considered to be a new matrix. The above operation is repeated. Obviously, if the division number is M, the above operation will be repeated $M - 1$ times.

By introducing the component exchange algorithm, it is possible to make the covariance matrix closer to the block diagonal matrix. Fig.6 shows the results of exchanged covariance matrices, where $M = 2, 4, 7$. Squares drawn by solid lines show the non-zero components areas. From these results, it is clear that the exchange algorithm changes the original covariance matrix closer to the block diagonal matrix as shown in Eq.(4).

For every kind of character, the order of elements of feature vector is changed according to the result of component exchange algorithm. n-dimensional feature vector is partitioned into M numbers of K-dimensional vectors. The elements of the ith K-dimensional vector are $K(i-1)+1 \sim Ki$ $(i = 1, ..., M)$. Then M numbers of $K \times K$ covariance matrices are computed. Eq.(6) is used to be the discriminant function.

4 Recognition Experiment

In order to confirm the effectiveness of the proposed method, recognition experiments are carried on with the ETL9B. The Improved Directional Element Feature is used as the

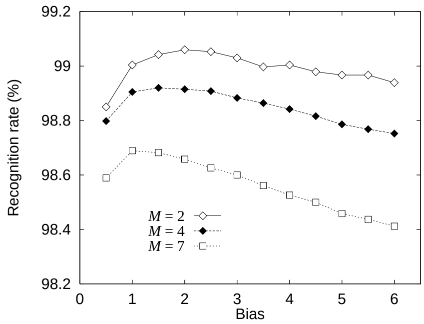

Figure 7: Experimental results (1).

feature vector. The proposed vector-divided Mahalanobis distance is adopted to be the discriminant function.

In sections 4.1 $\sim$ 4.3, the results of experiments that use the first 20 sets of the ETL9B as evaluation samples and the other 180 sets as training samples are described. Here, various values of bias b are tested. In section 4.4, a large scale experiment is done.

4.1 Number of Divisions

Experiments are carried on to exam the relation between the number of divisions and recognition rate. In order to reduce computational cost, recognition is divided into two steps that are rough classification and fine classification. Weighted Euclidean distance defined as Eq.(11) is used for rough classification. The fine classification objects are the top ten candidates of an unknown input character. The accumulated recognition rate of the top ten candidates of rough classification is 99.76%.

$$d_w^2(\boldsymbol{x}) = \sum_{j=1}^{n} \frac{1}{\sigma_j^2}(x_j - \mu_j)^2, \tag{11}$$

where x_j and μ_j are the jth element of $\boldsymbol{x}$ and $\boldsymbol{\mu}$. σ_j^2 is the variance of the jth element, and it is calculated from the training samples. σ_j^2 corresponds to the jth element of the covariance matrix Σ. Eq.(11) is the case of $M = 196$ of the proposed method.

For dividing the 196-dimensional feature vector, three kinds of division numbers that are $M = 2, 4, 7$ are used. Results are shown in Fig.7.

The horizontal axis of the figure shows the values of bias, while the vertical axis shows recognition rates. As shown in the figure, as the number of divisions becomes smaller, the recognition rate becomes higher. The reason of the obtained results is considered that because the increasing of the number of divisions, more components that are not near the diagonal are ignored. The highest recognition rate 99.06% is obtained in the case of $M = 2$ and $b = 2.0$.

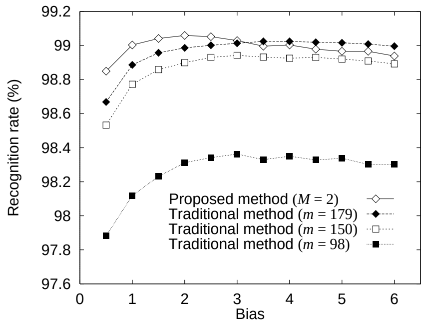

Figure 8: Experimental results (2).

4.2 Comparison with Conventional Method

In this section, in order to confirm the effect of the proposed method, results of conventional method that uses the MMD as the discriminant function are obtained to compare with the results of the proposed method. Here, three kinds of numbers of dimensions are tested. The first one is 98 which is the number of dimensions that the proposed method gets the best recognition rate. The second is 150 which is the same number of dimensions used in [2]. The last kind of number of dimensions is 179, which is the largest calculable number of eigenvectors in the case of using 180 training samples. For the conventional method, Eq.(11) is also adopted for the rough classification as described in section 4.1. Recognition results are shown in Fig.8. In the case of $m = 98$, compared with the result of the MMD, the recognition rate of the proposed method is much higher. From these results, it is thought popularly used methods those disregard higher dimensions have lost important information for classification. Although the proposed method ignores those components that are distant from the diagonal of covariance matrix, the bad influence is not as serious as those caused by disregarding higher dimensions. For the MMD, the best recognition rate of 99.03% is gained in the case of $m = 179$, $b = 4.0$. However, it does not exceed the best result of the proposed method.

Moreover, as shown in Fig.7 and Fig.8, if the number of dimensions of divided vector becomes smaller, the value of bias corresponding to the peak of recognition rate becomes smaller, too. Originally, the bias is introduced to reduce the negative influence caused by errors included in eigenvalues. If the value of bias becomes smaller, it is thought that the eigenvalues are comparatively accurate. Because the proposed method partitions the 196-dimensional feature vector into several small number of dimensional vectors, compared with the number of dimensions of divided vector, the number of training samples relatively becomes larger. For this reason, the covariance matrix calculated by the proposed method is much more reliable. However, if the number of division is too large, more components that are not near the diagonal of covariance matrix are ignored. As a result, useful information

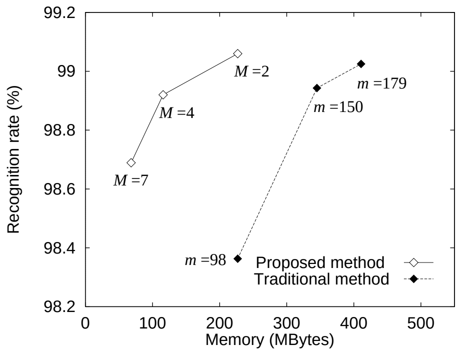

Figure 9: Recognition rate and memory use.

is lost. In our experiments conditions, the optimum number of divisions is two.

4.3 Computational Cost

As described in section 3.1, the proposed method has reduced the computation time and memory to $1/M$ of the Mahalanobis distance. The computational cost and recognition rate of the proposed method and the conventional method are shown in Fig.9. The numbers of divisions of the proposed method and the numbers of dimensions of the conventional method are also shown in the figure. The values of bias are selected to obtain the best recognition rate. As shown in the figure, in order to get the same level of recognition ability, the memory cost of the proposed method is much less than the conventional method.

For practical recognition system, available memory is limited. By using the proposed method, it is possible to construct a recognition system with little cost of memory.

4.4 Experiment with all sets of ETL9B

In order to verify the performance of the proposed method, experiments are carried out with all sets of the ETL9B. In these experiments, rough classification is not performed. Every twenty sets out of the 200 sets of the ETL9B is considered as a group, which makes ten groups in total, named Group A through J. In rotation, nine groups are used as the training data, and the excepted one group is employed as test data. Based on the results obtained in sections 4.1 and 4.2, parameters that give the best performance are selected. For the conventional method MMD, the number of dimensions is 179, and bias b is 4.0. For the proposed method, the number of divisions is 2, and the bias b is 2.0. The proposed algorithm is implemented in C language on HP C160. The recognition rates are shown in Table 1. Two kinds of average processing time are displayed in Table 2. From these tables, it is shown that the proposed method gives a very satisfactory performance with very low computational cost.

Table 1: Recognition rates.

Group	Conventional	Proposed
A	99.10%	99.10%
B	98.63%	98.67%
C	98.96%	98.95%
D	98.63%	98.73%
E	98.72%	98.72%
F	98.51%	98.57%
G	98.46%	98.47%
H	98.60%	98.67%
I	98.53%	98.49%
J	98.50%	98.53%
Average	98.66%	98.69%

Table 2: Time for recognition.

Method	Average processing time	
	Pre-processing and feature extraction	Classification
Conventional	0.006 sec/character	3.72 sec/character
Proposed	0.006 sec/character	2.04 sec/character

As shown in Table 1, the recognition rates of Group A of both the conventional method and the proposed method are higher than the result of Group A of rough classification (See sections 4.1 and 4.2). Although the accumulated recognition rate of top ten candidates of rough classification is extremely high as 99.76%, some correct answers missed by the rough classification are selected by the proposed method. It proves that the proposed method has a very high recognition ability.

5 Conclusions

In this paper, by investigating correlation of elements of feature vector, the vector-divided Mahalanobis distance is proposed. A new character recognition method is developed based on the vector-divided Mahalanobis distance. First, the covariance matrix is calculated from the feature vectors used for character recognition. By examining the characteristic of the relationship of elements of feature vector, the original Mahalanobis distance is considered that it can be computed approximately from divided vectors. The effect of this idea is evidenced by simulation. Moreover, in order to suit the division of covariance matrix, component exchange algorithm is proposed. The effectiveness of the proposed method is shown by the experimental results with the database ETL9B. The results have proved that the vector-divided Mahalanobis distance gives very good performance especially in the case of small number of training samples, and its computational cost is reduced drastically.

In this paper, the effectiveness of the proposed method is confirmed by the feature vectors extracted from characters. The proposed method can be used in any cases of limited training samples. To investigate the effectiveness of this method in the case of using other feature vectors is a future work. Also it is important to estimate the proposed method with much more variety of handwritten documents.

Acknowledgment. The authors wish to thank the members of the Electrotechnical Laboratory who provided the database ETL9B. This work was supported in part by the Ministry of Education, Science, Sports and Culture Grant-in-Aid for Scientific Research (B), No.08458072.

References

[1] 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.

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

[3] T. Wakabayashi, Y. Deng, S. Tsuruoka, F. Kimura, and Y. Miyake, "Accuracy Improvement by Nonlinear Normalization and Feature Compression in Handwritten Chinese Character Recognition," *Trans. IEICE*, vol.J79-D-II, no.5, pp.765–774, May 1996.

[4] M. Suzuki, N. Kato, H. Aso, and Y. Nemoto, "A Handprinted Character Recognition System Using Image Transformation Based on Partial Inclination Detection," *IEICE Trans. Inf. & Syst.*, vol.E79-D, no.5, pp.504–509, May 1996.

[5] 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.

[6] N. Hagita, S. Naito, and I. Masuda, "Handprinted Chinese Characters Recognition by Peripheral Direction Contributivity Feature," *Trans. IEICE*, vol.J66-D, no.10, pp.1185–1192, October 1983.

[7] H. Sakano, H. Kida, and N. Mukawa, "An Analysis of a Character Classifier Using Genetic Algorithm" *Trans. IEICE*, vol.J80-D-II, no.7, pp.1687–1694, July 1997.

[8] 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.

[9] 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.

[10] 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, 1983.

[11] N. Hagita and I. Masuda, "Dimensional Reduction of Direction Contributivity Feature Vector for Recognition of Handprinted Kanji Characters," *IEICE Technical Report*, PRL85-36, 1985.

[12] J. Tsukumo, "Improved Algorithm for Direction Pattern Matching and its Application for Handprinted Kanji Character Classification," *IEICE Technical Report*, PRU90-20, 1990.

[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] K. Fukunaga, "Introduction to Statistical Pattern Recognition," 2nd edition, Academic Press, pp.39–40, 1990.

Authors

Fang Sun received the B.E., Master of Information Sciences and Dr.Eng. degrees from Tohoku University, Japan, in 1994, 1996 and 1999, respectively. She is now a research associate at Faculty of Science and Technology, Tohoku Bunka Gakuen University, Japan. Her research interest is character recognition. Dr. Sun is a member of the Institute of Electronics, Information and Communication Engineers.

Shin'ichiro Omachi received his B.E., M.E. and Doctor of Engineering degrees in Information Engineering from Tohoku University, Japan, in 1988, 1990 and 1993, respectively. He has worked as a research associate at the Education Center for Information Processing at Tohoku University from 1993 to 1996. He is now an associate professor at Graduate School of Engineering, Tohoku University. His research interests include pattern recognition and parallel processing of images. Dr. Omachi is a member of the IEEE, the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, and the Japanese Society of Artificial Intelligence.

Nei Kato received his M.S. and the Dr.Engineering. degrees in Information Engineering from Tohoku University in 1988 and 1991, respectively. He joined the Computer Center at Tohoku University in 1991 and now he is an associate professor at the Graduate School of Information Sciences at Tohoku University. He has been engaging in research of pattern recognition, neural networks, and computer networking. Dr. Kato is a member of IEEE, the Institute of Electronics, Information and Communication Engineers and the Information Processing Society of Japan.

Hirotomo Aso received his B.E., M.E., and Doctor of Engineering degrees in Electrical Engineering from Tohoku University, Japan, in 1968, 1970, and 1974, respectively. He was with the Department of Information Engineering, Tohoku University in 1973, and later joined the Faculty of Engineering, Nagoya University from 1979 to 1986. He is now a professor at Tohoku University. He is presently engaged in research of character pattern recognition, cellular automata, concurrent program schema, and parallel processing. He has received the prize of research encouragement of IEICEJ in 1978 and the prize of achievements (Gyoseki-sho) of IEICEJ in 1992. Dr. Aso is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, the Japanese Society of Artificial Intelligence, Japanese Cognitive Science Society, IEEE, EATCS, and ACM.