Two-Stage Computational Cost Reduction Algorithm Based on Mahalanobis Distance Approximations

Fang Sun[1], Shin'ichiro Omachi[2], Nei Kato[3], Hirotomo Aso[2], Shun'ichi Kono[1], and Tasuku Takagi[1]

[1] Faculty of Science and Technology, Tohoku Bunka Gakuen University
6-45-16, Kunimi, Aoba-ku, Sendai-shi, 981-8551 Japan
{fan,skono,ttakagi}@ait.tbgu.ac.jp
[2] Graduate School of Engineering, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan
{machi,aso}@ecei.tohoku.ac.jp
[3] Graduate School of Information Sciences, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan
kato@nemoto.ecei.tohoku.ac.jp

Abstract

For many pattern recognition methods, high recognition accuracy is obtained at very high expense of computational cost. In this paper, a new algorithm that reduces the computational cost for calculating discriminant function is proposed. This algorithm consists of two stages which are feature vector division and dimensional reduction. The processing of feature division is based on characteristic of covariance matrix. The dimensional reduction in the second stage is done by an approximation of the Mahalanobis distance. Compared with the well-known dimensional reduction method of K-L expansion, experimental results show the proposed algorithm not only reduces the computational cost but also improves the recognition accuracy.

1. Introduction

For many pattern recognition methods, usually high recognition accuracy is obtained at very high expense of computational cost. How to reduce the computational cost but keeping the high accuracy is a concerned problem. In the cases that high dimensional feature vectors are used, for example Chinese and Japanese character recognition, this problem is extremely serious.

Research in handwritten Chinese and Japanese character recognition has matured significantly [2, 3, 10]. Wakabayashi et al. obtain the high recognition rate by using weighted direction code histogram and compression of higher dimensional features [10]. Kato et al. have developed a handwritten character recognition system using improved directional element feature and asymmetric Mahalanobis distance [3]. These methods have obtained very high accuracy, however, the dimensionality of feature vectors are very large. The dimensionality of weighted direction code histogram is 392. The improved directional element feature has 196 dimensions. Some features have much more dimensions, for example, the extended peripheral direction contributivity [1, 7] has 1536 dimensions. In addition, because the number of kinds of Chinese and Japanese characters is extremely large, reducing the computational cost is an important problem. Of course, this problem is undoubtedly thought to be a common problem of most pattern recognition research, in this paper, as an example, Chinese and Japanese character recognition is considered.

Although computational cost reduction is the aim of this research, it does not mean that the recognition accuracy is not important. For precise pattern recognition system, effective discriminant function is a very important factor. Some research has shown the Mahalanobis distance gives good performance on pattern recognition. The Mahalanobis distance is derived by a probability density function of multivariate normal distribution. It is considered as an appropriate function if the distribution of samples is multivariate normal and there are enough sample patterns. However, compared with the dimensionality, the training samples are always not enough. For this reason, covariance matrix usually cannot be estimated accurately. Furthermore, there are other disadvantages of the Mahalanobis distance such as the computation time will reach $O(n^2)$ for n-dimensional feature vectors.

Considering the advantages and disadvantages of the Mahalanobis distance, based on two kinds of approximations of the Mahalanobis distance, in this paper, a new recognition algorithm that reduces the computational cost for calculating discriminant function is proposed. This algorithm consists of two stages which are feature vector division and dimensional reduction. The first stage of feature division is based on the characteristic of covariance matrix. The second stage is done by regarding the values of small eigenvalues as a constant.

There are some existing methods for reducing computational cost. One of the well-known dimensional reduction methods is the Karhunen-Loève (K-L) expansion [4]. Compared with the K-L expansion, experimental results show the proposed algorithm not only reduces the computational cost but also improves the recognition accuracy.

2. Two-stage computational cost reduction

2.1. Feature vector

The improved directional element feature [3] is used as the feature vector in this paper. The effectiveness of this feature has been shown with ETL9B [6], which is the largest handwritten character database in Japan. The improved directional element feature is calculated as follows.

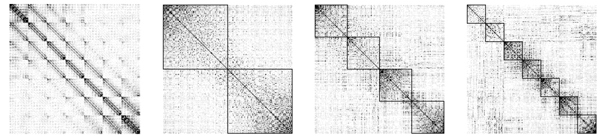
First an input image is normalized to 64×64 dots, and the contour of the image is extracted. Next, orientation, which is one of vertical, horizontal, and two oblique lines slanted at $\pm 45°$, is assigned for each pixel. Then the image is divided into 49 sub-areas of 16×16 dots where each sub-area overlaps eight dots with the adjacent sub-area. For 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. Feature vector division

The first stage of reduction is done by dividing feature vector. Here, the Mahalanobis distance is considered. Suppose the feature vector be n-dimensional vector. In the case of the improved directional element feature, $n = 196$. $\boldsymbol{\mu}$ and Σ denote mean vector and $n \times n$ covariance matrix. The squared Mahalanobis distance of vector $\boldsymbol{x}$ is defined as

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

Here, to avoid the bad influence caused by limited samples, regularization term is added to the sample covariance matrix [5]. In other words, $\Sigma = \hat{\Sigma} + \alpha I$ is used. Here, $\hat{\Sigma}$ is the sample covariance matrix, I is an identity matrix and α is a small positive constant.



(a) Original (b) $M = 2$ (c) $M = 4$ (d) $M = 7$

Figure 1. Original and exchanged covariance matrices.

Suppose the covariance matrix is a block diagonal matrix that consists of M numbers of $K \times K$ matrices ($n = M \times K$), while the other components are zero. In other words,

$$\Sigma = \begin{bmatrix} \Sigma_1 & & & 0 \\ & \Sigma_2 & & \\ & & \ddots & \\ 0 & & & \Sigma_M \end{bmatrix}, \qquad (2)$$

where $\Sigma_1, \Sigma_2, ..., \Sigma_M$ are $K \times K$ square matrices and the other components are zero. In this case, Eq.(1) can be written as

$$d^2(\boldsymbol{x}) = \sum_{i=1}^{M} (\boldsymbol{x}_i - \boldsymbol{\mu}_i)^t \Sigma_i^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_i), \qquad (3)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{\mu}_i$ are K-dimensional vectors constructed by $(K(i-1)+1)$th $\sim$ (Ki)th components of $\boldsymbol{x}$ and $\boldsymbol{\mu}$.

However, generally the covariance matrix is quite different from the block diagonal matrix denoted as Eq.(2). Obviously, if the values of non-diagonal components are not zero, they will cause negative effect. In order to make the covariance matrix closer to the block diagonal matrix, the components of the covariance matrix are exchanged by using a component exchange algorithm [9].

Fig.1 shows an example of applying the exchange algorithm. For the first 100 kinds of character images in the database ETL9B, the covariance matrix of each kind of character is calculated from the feature vectors. Then, the mean covariance matrix Σ_0 is computed. Fig.1(a) shows the 196×196 covariance matrix Σ_0. The depth of black shows the absolute value of element. Fig.1(b)$\sim$(d) show the results of exchanged covariance matrices in the cases of $M = 2, 4, 7$. Squares drawn by solid lines show the non-zero component areas. These results clearly show that the exchange algorithm changes the original covariance matrix closer to the block diagonal matrix of Eq.(2).

The order of elements of every reference vector is decided by the result of applying the component exchange algorithm. For an unknown input character, the order of elements of its feature vector is changed according to the order of each reference vector. 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 Eq.(3) is applied for classification.

By this stage's processing, recognition can be realized 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 dimensionality becomes larger. As a result, the estimated covariance matrix becomes much more reliable. Moreover, the computational time is reduced inverse proportion with the division number.

2.3. Dimensional reduction

The second stage of reduction is done by regarding the values of small eigenvalues as a constant. Eq.(3) can be written as follows.

$$d^2(\boldsymbol{x}) = \sum_{i=1}^{M} d_i^2(\boldsymbol{x}_i), \quad (4)$$

where

$$\begin{aligned} d_i^2(\boldsymbol{x}_i) &= (\boldsymbol{x}_i - \boldsymbol{\mu}_i)^t \Sigma_i^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_i) \\ &= \sum_{j=1}^{K} \frac{((\boldsymbol{x}_i - \boldsymbol{\mu}_i)^t \boldsymbol{\phi}_{ij})^2}{\lambda_{ij}}. \end{aligned} \quad (5)$$

λ_{ij} and $\boldsymbol{\phi}_{ij}$ are the jth eigenvalue and eigenvector of Σ_i. Replacing the value of $\lambda_{ij}(j > L)$ with the mean value of λ_{ij} $(j = L+1, ..., K)$, approximation of Eq.(5) is calculated [8].

In the case of using high dimensional feature vectors, comparing with the dimensionality, the training samples are always not enough, hence the covariance matrix usually cannot be estimated accurately. Especially, the terms that correspond to the smaller eigenvalues in Eq.(5) will include much more errors. For this reason, the mean value of λ_{ij} $(j = L+1, ..., n)$ is used instead of λ_{ij} $(t > L)$. Moreover, this processing can reduce the computational time to L/K of computing the original Mahalanobis distance.

3. Experiments

3.1. Performance of the proposed method

In order to confirm the effectiveness of the proposed method, experiments are carried out using the first 100 kinds of characters in the database ETL9B. There are 200 sets of image samples. The first 20 sets are used as evaluation samples, while the other 180 sets are used as training samples. First, experiments are carried out to exam the effect of the first stage processing. Three kinds of division numbers that are $M = 2, 4, 7$ are used for dividing the 196-dimensional feature vector. The second kind of experiments combines the two stages. Recognition accuracy and processing time are shown in Table 1. The case of $M = 1$ is the result of using the original Mahalanobis distance (Eq.(1)).

The results of the first stage have shown that as the number of divisions becomes larger, the processing time becomes less rapidly. However, recognition accuracy also decreases a little. The reason is that with the number of divisions becoming larger, more components that are not near the diagonal of covariance matrix are ignored and useful information is lost. On the other hand, as shown with the results of two stages, the processing time is extremely short. The highest recognition accuracy 99.30% is obtained with $M = 2$. This is even better than the result of the original Mahalanobis distance. The reasons are considered as follows. Since the proposed method partitions the original feature vector into several small number of dimensional vectors, compared with the dimensionality of divided vector, the number of training samples relatively becomes larger. Accordingly, the covariance matrices calculated by the proposed method in the first stage are more precise than the matrix computed by the original method. Furthermore, in the second stage, only comparatively precise eigenvalues are adopted, the bad influence caused by errors included in smaller eigenvalues are avoided.

However, the approximation of the Mahalanobis distance in the first stage is only valid if many pairs of components of feature vector have little correlations. The proposed method needs to be tested whether it functions properly with other kinds of features.

3.2. Comparison with conventional method

There are some conventional methods for reducing computational cost. One of the well-known dimensional reduction methods is the K-L expansion. In order to confirm the effect of the proposed method, experiments for comparing with the K-L expansion are carried out. For the K-L expansion, 180 training samples are used to calculate the covariance matrix of each kind of character. Then, the mean covariance matrix is computed. The eigenvectors of this covariance matrix that corresponds to P large eigenvalues are used as the axes of new subspace.

Fig.2 shows the results of recognition accuracy and computation time per character using various P. The result of the proposed method that achieves the highest recognition accuracy ($M = 2$) is also displayed. The recognition accuracy of the K-L expansion is very low in the case of using the same processing time of the proposed method. For the K-L expansion, the highest recognition accuracy of 99.25% is obtained in the case of $P = 165$. Although the computational time is eight times longer than the proposed algorithm, it does not exceed the best result of the proposed method.

Table 1. Results.

Number of divisions M		1	2	4	7
First stage	Recognition accuracy (%)	99.15	99.10	98.95	98.95
	Time (sec/character)	0.169	0.083	0.040	0.027
Two stages	Recognition accuracy (%)	99.15	99.30	99.05	98.95
	Time (sec/character)	0.021	0.021	0.020	0.022

4. Conclusions

For many pattern recognition methods, high recognition accuracy is obtained at very high expense of computational cost. How to reduce the computational cost but keeping the high accuracy is a concerned problem. In the cases that high dimensional feature vectors are used, for example Chinese and Japanese character recognition, this problem is extremely serious.

In this paper, a new recognition algorithm that reduces the computational cost for calculating discriminant function was proposed. This algorithm consists of two stages which are feature vector division and dimensional reduction. The processing of feature division is based on the characteristic of covariance matrix. The dimensional reduction in the second stage is done by an approximation of the Mahalanobis distance.

The effectiveness of the proposed algorithm was confirmed by the feature vectors extracted from character images. The experimental results have clarified that the proposed algorithm not only reduces the computational cost but also improves the recognition accuracy. Moreover, the comparison experiments have shown the results of proposed method exceed the results of the K-L expansion.

Computational cost reduction is undoubtedly thought to be a common problem of most pattern recognition research. The focus of this paper is Chinese and Japanese character recognition. The algorithm proposed here can be used in any cases of limited training samples. To investigate the effectiveness of this algorithm in other pattern recognition problems using other kinds of feature vectors is a future work.

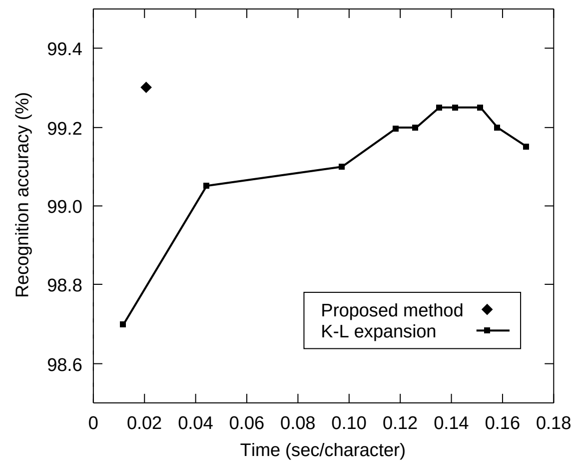


Figure 2. Comparison with K-L expansion.

References

[1] N. Hagita, S. Naito, and I. Masuda. Handprinted Chinese characters recognition by peripheral direction contributivity feature. *Trans. IEICE*, J66-D(10):1185–1192, October 1983.

[2] T. W. Hildebrandt and W. Liu. Optical recognition of handwritten Chinese characters: advances since 1980. *Pattern Recognition*, 26(2):205–225, 1993.

[3] N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto. A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(3):258–262, March 1999.

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

[5] S. Raudys. Evolution and generalization of a single neurone: I. single-layer perceptron as seven statistical classifiers. *Neural Networks*, 11(2):283–296, February 1998.

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

[7] H. Sakano, H. Kida, and N. Mukawa. An analysis of a character classifier using genetic algorithm. *Trans. IEICE*, J80-D-II(7):1687–1694, July 1997.

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

[9] F. Sun, S. Omachi, N. Kato, and H. Aso. Fast and precise discriminant function considering correlations of elements of feature vectors and its application to character recognition. *Trans. IEICE*, J81-D-II(9):2027–2034, September 1998.

[10] 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*, J79-D-II(5):765–774, May 1996.