A Modification of Eigenvalues to Compensate Estimation Errors of Eigenvectors

Masakazu IWAMURA, Shin'ichiro OMACHI, and Hirotomo ASO

Graduate School of Engineering, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980–8579 Japan
E-mail: masa@aso.ecei.tohoku.ac.jp

Abstract

In statistical pattern recognition, parameters of distributions are usually estimated from training samples. It is well known that shortage of training samples causes estimation errors which reduce recognition accuracy. By studying estimation errors of eigenvalues, various methods of avoiding recognition accuracy reduction have been proposed. However, estimation errors of eigenvectors have not been considered enough. In this paper, we investigate estimation errors of eigenvectors to show these errors are another factor of recognition performance reduction. We propose a new method for modifying eigenvalues in order to reduce bad influence caused by estimation errors of eigenvectors. Effectiveness of the method is shown by experimental results.

1. Introduction

Statistical pattern recognition methods using the quadratic discriminant function and the Mahalanobis distance are optimal in some conditions. These methods essentially need to know the true distributions of the patterns, but it is usually impossible. Therefore, the parameters of distribution are estimated from training samples. It is well known that the parameters contain estimation errors when training samples are not enough.

To overcome the degradation in performance, regularized discriminant analysis [5] was proposed. As a method to approximate the true distribution, Sakai et al.[2] proposed a quadratic discriminant function with rectified eigenvalues which is called RQDF. They show a method to reduce estimation errors of eigenvalues by extending Fukunaga's analysis[1] which derives the deviation of estimation errors by perturbation theory.

Without estimating eigenvectors correctly, the effect of correcting eigenvalues is limited. Therefore, estimation errors of eigenvectors have to be corrected. However, the methods to correct estimation errors of eigenvectors have not been studied.

In this paper, first we investigate characteristics on estimation errors of eigenvectors experimentally. Examined feature vectors are extracted from real character images and artificial samples. As a result of the investigations, we show the possibility that estimation errors reduce recognition performance. In order to reduce bad influence caused by estimation errors of eigenvectors, we propose a new method to modify eigenvalues that compensates estimation errors of eigenvectors. Effectiveness of the method is shown by recognition experiments.

2. Estimation errors of eigenvectors

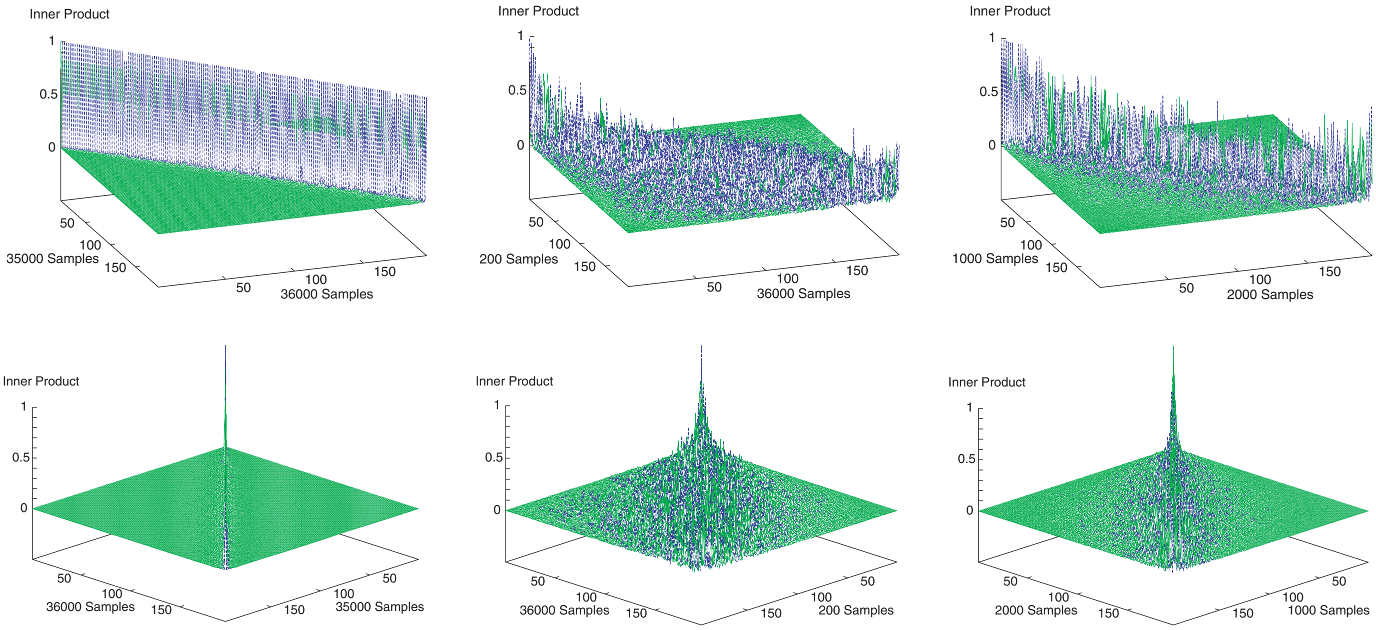
2.1. Evaluation method for errors

We investigate rotation angles of eigenvectors which show how estimation errors of eigenvectors are. Rotation angle of a vector can be measured by inner products of the vector and other vectors. A matrix $\boldsymbol{F}(N_1, N_2)$ is defined as

$$\boldsymbol{F}(N_1, N_2) = \begin{pmatrix} \hat{\boldsymbol{\phi}}_1^{N_1} \cdot \hat{\boldsymbol{\phi}}_1^{N_2} & \dots & \hat{\boldsymbol{\phi}}_1^{N_1} \cdot \hat{\boldsymbol{\phi}}_d^{N_2} \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\phi}}_d^{N_1} \cdot \hat{\boldsymbol{\phi}}_1^{N_2} & \dots & \hat{\boldsymbol{\phi}}_d^{N_1} \cdot \hat{\boldsymbol{\phi}}_d^{N_2} \end{pmatrix}, \qquad (1)$$

where d is the dimension of feature vector, N_1 and N_2 are the numbers of training samples, and $\hat{\boldsymbol{\phi}}_k^N$ is the kth eigenvector estimated from N training samples. Each diagonal component of the matrix represents the size of the error. Each off-diagonal component $\boldsymbol{F}_{ij}$ represents the length of component of vector $\hat{\boldsymbol{\phi}}_i^{N_1}$ to the direction of vector $\hat{\boldsymbol{\phi}}_j^{N_2}$, or that of vector $\hat{\boldsymbol{\phi}}_j^{N_2}$ to the direction of vector $\hat{\boldsymbol{\phi}}_i^{N_1}$.

In the following sections, the absolute value of each component of matrix $\boldsymbol{F}(N_1, N_2)$ is plotted in a three-dimensional view.



(a) $\boldsymbol{F}(36000, 35000)$ (b) $\boldsymbol{F}(36000, 200)$ (c) $\boldsymbol{F}(2000, 1000)$

Figure 1. Errors of real character images.

2.2. Errors of real character images

Using real character images, the relationship between the number of training samples and the estimation errors of eigenvectors is investigated.

Digit samples of NIST Special Database 19 are used. Each sample in the database is normalized nonlinearly[6] to 64 pixels square, and then 196-dimensional Directional Element Feature[4] is extracted.

The results of letter "0" are shown in Fig. 1. Each subfigure consists of two graphs from different viewpoints. From the figures, estimation errors of eigenvectors are ascertained and two kinds of tendencies can be observed: (i)the errors become greater as the difference of sample sizes $|N_1 - N_2|$ becomes larger as shown in Fig. 1(a) and Fig. 1(b), and (ii)the errors become greater as sample sizes N_1 and N_2 become smaller as shown in Fig. 1(a) and Fig. 1(c). These tendencies can be seen in the results of other letters.

2.3. Errors of artificial samples

Artificial samples are prepared as follows. First, we normalize real character images and extract feature vectors. Mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$ are calculated from 36,000 samples of letter "0". According to normal distribution $\boldsymbol{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, artificial samples are created by random numbers.

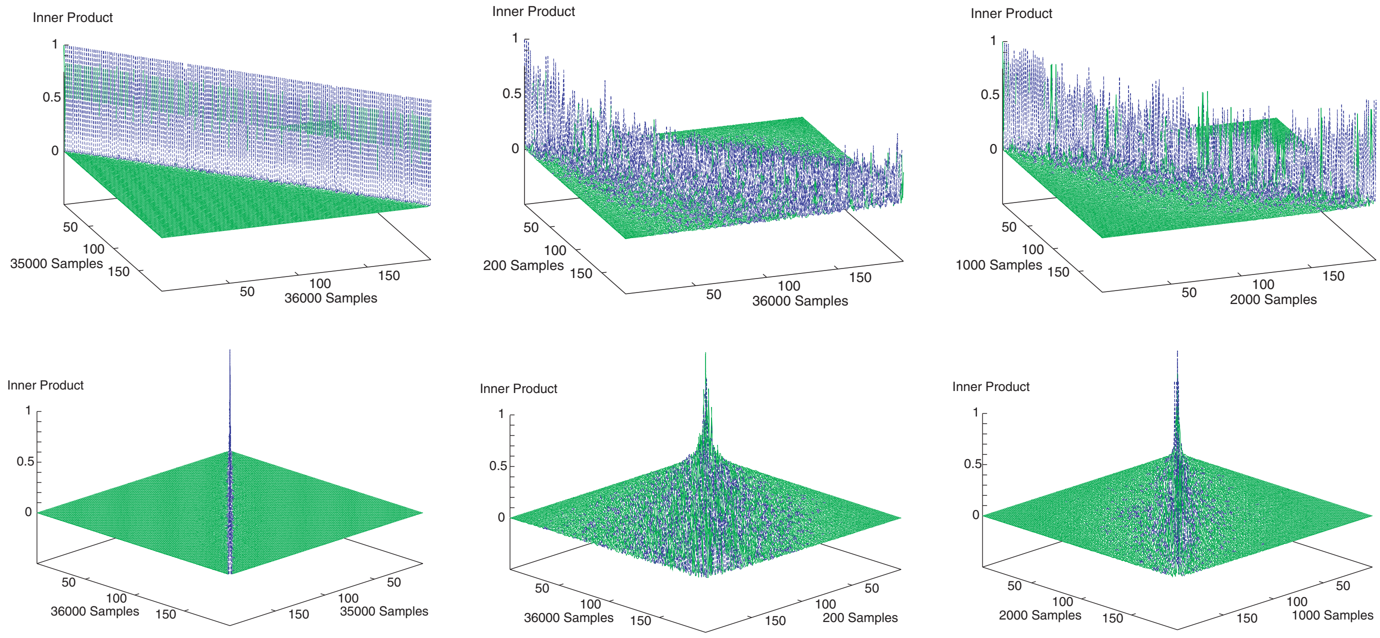
The results are shown in Fig. 2. The results are very similar to the results of real character images.

Artificial samples exactly are distributed according to normal distribution due to the process of construction. Estimation errors of artificial samples have same tendencies as that of real character images.

It is noted that the results of real character images and artificial samples indicate that estimation errors of eigenvectors seem not to depend on a sample set itself. Therefore the matrix $\boldsymbol{F}(N_1, N_2)$ calculated from artificial samples is valid for all the samples of the same size to use error correction.

3. Reducing bad influence of estimation errors of eigenvectors

We consider the influences of estimation errors of eigenvectors to recognition performance. Most statistical pattern recognition methods use a covariance matrix, which can decompose into the eigenvalues and the eigenvectors. It can be interpreted that eigenvectors are the directions to calculate the distance or the similarity, while eigenvalues are the weights corresponding to the directions. It seems that correcting only estimation errors of eigenvalues does not take



(a) $\boldsymbol{F}(36000, 35000)$ (b) $\boldsymbol{F}(36000, 200)$ (c) $\boldsymbol{F}(2000, 1000)$

Figure 2. Errors of artificial samples.

much effect on the performance.

If we calculate the matrix $\boldsymbol{F}(N_1, N_2)$ exactly, we can correct eigenvectors. However, calculating $\boldsymbol{F}(N_1, N_2)$ is too difficult because we can obtain the matrix only stochastically and degree of freedom about eigenvectors is larger than that of eigenvalues. Therefore, we will consider a modification of eigenvalues to compensate estimation errors of eigenvectors.

3.1. A method of modifying eigenvalues

If N samples per category can be used for training, we propose the following procedure to reduce bad influence caused by estimation errors of eigenvectors.

1. The parameters of the distribution are estimated from these N samples.
2. Eigenvalues of the sample covariance matrix are corrected by some known method for correcting estimation errors of eigenvalues. The corrected kth eigenvalue is denoted as λ'_k.
3. Each corrected eigenvalue λ'_k is modified by the next formula.

$$\tilde{\lambda}_k = \sum_{i=1}^{d} \left\{ \boldsymbol{F}(\infty, N)_{ik} \right\}^2 \lambda'_i. \tag{2}$$

$\tilde{\lambda}_k$ is the kth eigenvalue which is modified by this method and used for recognition. The matrix $\boldsymbol{F}(\infty, N)$ is estimated stochastically.

Eq. (2) is derived as follows. Let $\boldsymbol{\Sigma}$ be the true covariance matrix and $\hat{\boldsymbol{\Sigma}}$ be the sample covariance matrix estimated from N samples. The matrices can be decomposed into the eigenvalues and the eigenvectors as $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T$ and $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Phi}}^T$, where $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Phi}}$ are denoted as

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_d\right), \tag{3}$$
$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_d \end{pmatrix}, \tag{4}$$
$$\hat{\boldsymbol{\Lambda}} = \mathrm{diag}\big(\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_d\big), \tag{5}$$
$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} \hat{\boldsymbol{\phi}}_1 & \hat{\boldsymbol{\phi}}_2 & \cdots & \hat{\boldsymbol{\phi}}_d \end{pmatrix}. \tag{6}$$

$\boldsymbol{\phi}_k$ is the corresponding eigenvector to eigenvalue λ_k, and $\hat{\boldsymbol{\phi}}_k$ corresponds to $\hat{\lambda}_k$. It is assumed that the true eigenvalues $\boldsymbol{\Lambda}$ is estimated as $\boldsymbol{\Lambda}' = \mathrm{diag}\left(\lambda'_1, \lambda'_2, \cdots, \lambda'_d\right)$. If we could know the true eigenvectors $\boldsymbol{\Phi}$, the covariance matrix can be estimated as $\boldsymbol{\Sigma}_{\text{est}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{\Phi}^T$. However, we have $\hat{\boldsymbol{\Phi}}$ instead of $\boldsymbol{\Phi}$. Therefore, we desire to find modified eigenvalues $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}\big(\tilde{\lambda}_1, \tilde{\lambda}_2, \cdots, \tilde{\lambda}_d\big)$ to satisfy $\boldsymbol{\Sigma}_{\text{est}} \simeq \hat{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Phi}}^T$. Then, following equation is obtained.

$$\hat{\mathbf{\Phi}}\tilde{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}^T \simeq \mathbf{\Phi}\mathbf{\Lambda}'\mathbf{\Phi}^T$$
$$\tilde{\mathbf{\Lambda}} \simeq \hat{\mathbf{\Phi}}^T\mathbf{\Phi}\mathbf{\Lambda}'\mathbf{\Phi}^T\hat{\mathbf{\Phi}} = \boldsymbol{F}(\infty, N)^T\mathbf{\Lambda}'\boldsymbol{F}(\infty, N) \qquad (7)$$

Eq. (2) is the kth diagonal component of Eq. (7).

4. Recognition experiments

We carry out experiments to show the ability of the proposed method. Digit samples of NIST Special Database 19 are used for the experiments. $1{,}000$ samples are used for testing. The other samples are used for training within $35{,}000$ samples. The quadratic discriminant function[3] is used. Matrix $\boldsymbol{F}(\infty, N)$ is calculated from artificial samples.

Parameters of distributions are calculated from training samples. The relationship between the number of training samples and recognition rates is observed. Recognition rates of three cases are compared. The first one is that sample mean vectors, sample eigenvectors and eigenvalues corrected by Sakai's method are used for recognition. This case means that only estimation errors of eigenvalues are corrected, and it is labeled "Sakai's Method." The second one is that sample mean vectors, sample eigenvectors and modified eigenvalues by applying the proposed method are used. This case means that estimation errors of eigenvalues and eigenvectors are corrected, and it is labeled "Proposed Method." The difference of these recognition rates shows the effectiveness of the proposed method. As the third case, the result of normal quadratic discriminant function with statistical parameters from samples are shown for reference. It is labeled "Quadratic Discriminant Function."

The results of the experiments are shown in Fig. 3. The results shows following facts:

- Comparison of "Quadratic Discriminant Function" and "Sakai's Method" shows that correction of estimation errors of eigenvalues is effective.

- Comparison of "Sakai's Method" and "Proposed Method" shows that the proposed method is effective. Particularly, the proposed method is more effective when sample size is small.

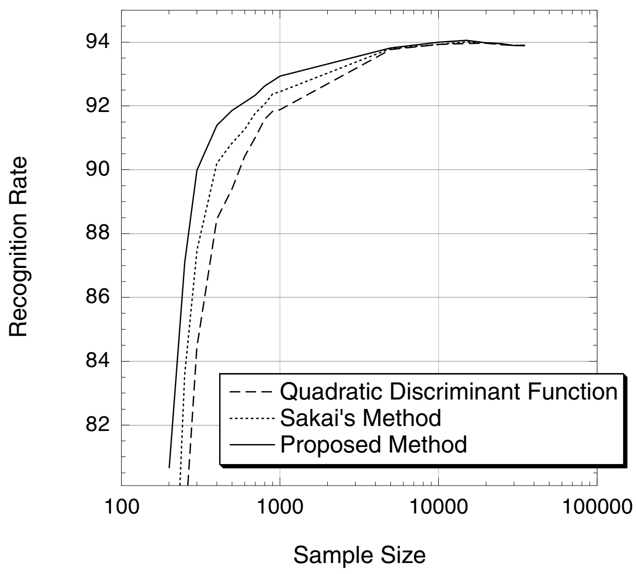


Figure 3. Experimental results.

5. Conclusions

In this paper, we investigated the estimation errors of eigenvectors and showed the possibility that estimation errors of eigenvectors reduce recognition performance. And, we proposed a method to reduce bad influence caused by estimation errors of eigenvectors, and ascertained the effectiveness of the method by the recognition experiments.

Estimation errors of eigenvalues have been considered and researched so far. However, estimation errors of eigenvectors have not been considered. Our investigations indicate the importance of considering estimation errors of eigenvectors. The proposed method modifies eigenvalues to compensate estimation errors of eigenvectors. The recognition experiments show the properness of the method. The method is useful for the recognition applications which are hard to get many samples because the effectiveness of the method is noteworthy when sample size is small.

Obtaining matrix $\boldsymbol{F}(\infty, N)$ theoretically is the task to be solved.

References

[1] K. Fukunaga. *Introduction to statistical pattern recognition*, pages 425–435. Academic Press, 2nd edition, 1990.

[2] M. Sakai, M. Yoneda, and H. Hase. A new robust quadratic discriminant function. In *Proc. ICPR*, pages 99–102, 1998.

[3] J. Schürmann. *Pattern classification*. John Wiley & Sons, Inc., 1996.

[4] N. Sun, Y. Uchiyama, H. Ichimura, H. Aso, and M. Kimura. Intelligent recognition of characters using associative matching technique. In *Proc. Pacific Rim Int'l Conf. Artificial Intelligence (PRICAI'90)*, pages 546–551, November 1990.

[5] A. Webb. *Statistical Pattern Recognition*, pages 34–36. Oxford University Press Inc., 1999.

[6] H. Yamada, Yamamoto, and T. Saito. A nonlinear normalization method for handprinted kanji character recognition — line density equalization —. *Pattern Recognition*, 23:1023–1029, 1990.