A Method to Estimate the True Mahalanobis Distance from Eigenvectors of Sample Covariance Matrix

Masakazu Iwamura, Shinichiro Omachi, and Hirotomo Aso

Graduate School of Engineering, Tohoku University
Aoba 05, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan
{masa, machi, aso}@aso.ecei.tohoku.ac.jp

Abstract. In statistical pattern recognition, the parameters of distributions are usually estimated from training sample vectors. However, estimated parameters contain estimation errors, and the errors cause bad influence on recognition performance when the sample size is not sufficient. Some methods can obtain better estimates of the eigenvalues of the true covariance matrix and can avoid bad influences caused by estimation errors of eigenvalues. However, estimation errors of eigenvectors of covariance matrix have not been considered enough. In this paper, we consider estimation errors of eigenvectors and show the errors can be regarded as estimation errors of eigenvalues. Then, we present a method to estimate the true Mahalanobis distance from eigenvectors of the sample covariance matrix. Recognition experiments show that by applying the proposed method, the true Mahalanobis distance can be estimated even if the sample size is small, and better recognition accuracy is achieved. The proposed method is useful for the practical applications of pattern recognition since the proposed method is effective without any hyper-parameters.

1 Introduction

In statistical pattern recognition, the Bayesian decision theory gives a decision to minimize the misclassification probability as long as the true distributions are given. However, the true distributions are unknown in most practical situations. The forms of the distributions are often assumed to be normal and the parameters of the distributions are estimated from the training sample vectors. It is well known that the estimated parameters contain estimation errors and the errors cause bad influence on recognition performance when there are not enough training sample vectors.

To avoid bad influence caused by estimation errors of eigenvalues, there are some methods to obtain better estimates of the true eigenvalues. Sakai et al. [1, 2] proposed a method to rectify the sample eigenvalues (the eigenvalues of the sample covariance matrix), which is called RQDF. James and Stein indicated that the conventional sample covariance matrix is not admissible (which means there are some better estimators). They proposed an improved estimator of the sample covariance matrix (James-Stein estimator) [3] by modifying the sample eigenvalues.

However, estimation errors of eigenvectors of covariance matrix have not been considered enough and still an important problem. In this paper, we aim to achieve high-performance pattern recognition without many training samples and any hyper-parameters. We present a method to estimate the true Mahalanobis distance from the sample eigenvectors. First of all, we show the error of the Mahalanobis distance caused by estimation errors of eigenvectors can be regarded as the errors of eigenvalues. Then, we introduce a procedure for estimating the true Mahalanobis distance by deriving the probability density function of estimation errors of eigenvectors. The proposed method consists of two-stage modification of the sample eigenvalues. At the first stage, estimation errors of eigenvalues are corrected using an existing method. At the second stage, the corrected eigenvalues are modified to compensate estimation errors of eigenvectors. The effectiveness of the proposed method is confirmed by recognition experiments. This paper is based on the intuitive sketch [4] and formulated with statistical and computational approaches.

2 A Method to Estimate the True Mahalanobis Distance

2.1 The Eigenvalues to Compensate Estimation Errors of Eigenvectors

If all the true parameters of the distribution are known, the true Mahalanobis distance is obtained. Let $\boldsymbol{x}$ be an unknown input vector, $\boldsymbol{\mu}$ be the true mean vector, $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_d)$ and $\boldsymbol{\Phi} = \left(\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2 \cdots \boldsymbol{\phi}_d\right)$, where λ_i and $\boldsymbol{\phi}_i$ are the ith eigenvalue and eigenvector of the true covariance matrix. All the eigenvalues are assumed to be ordered in the descending order in this paper. The true Mahalanobis distance is given as

$$d(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Phi} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}). \tag{1}$$

In general, the true eigenvectors are unknown and only the sample eigenvectors $\left\{\hat{\boldsymbol{\phi}}_i\right\}$ are obtained. Let $\hat{\boldsymbol{\Phi}} = \left(\hat{\boldsymbol{\phi}}_1\ \hat{\boldsymbol{\phi}}_2 \cdots \hat{\boldsymbol{\phi}}_d\right)$. The Mahalanobis distance using $\hat{\boldsymbol{\Phi}}$ is

$$\hat{d}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \hat{\boldsymbol{\Phi}} \boldsymbol{\Lambda}^{-1} \hat{\boldsymbol{\Phi}}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}). \tag{2}$$

Of course, $d(\boldsymbol{x})$ and $\hat{d}(\boldsymbol{x})$ differ. Now, let $\hat{\boldsymbol{\Psi}} \equiv \boldsymbol{\Phi}^{\mathrm{T}} \hat{\boldsymbol{\Phi}}$ be *estimation error matrix* of eigenvectors. Since both $\boldsymbol{\Phi}$ and $\hat{\boldsymbol{\Phi}}$ are orthonormal matrices, $\hat{\boldsymbol{\Psi}}$ is also an orthonormal matrix. Substituting $\hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi} \hat{\boldsymbol{\Psi}}$ into Eq. (2), we obtain

$$\hat{d}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Phi} \left(\hat{\boldsymbol{\Psi}} \boldsymbol{\Lambda} \hat{\boldsymbol{\Psi}}^{\mathrm{T}}\right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}). \tag{3}$$

Comparing Eq. (3) and Eq. (1), $\left(\hat{\boldsymbol{\Psi}} \boldsymbol{\Lambda} \hat{\boldsymbol{\Psi}}^{\mathrm{T}}\right)^{-1}$ in Eq. (3) corresponds to $\boldsymbol{\Lambda}^{-1}$ (the true eigenvalues) in Eq. (1). If we can ignore the non-orthogonal elements of

$\left(\hat{\boldsymbol{\Psi}}\boldsymbol{\Lambda}\hat{\boldsymbol{\Psi}}^{\mathrm{T}}\right)^{-1}$, the error of the Mahalanobis distance caused by the estimation errors of eigenvectors will be regarded as the errors of eigenvalues. This means that even if eigenvectors have estimation errors, we can estimate the true Mahalanobis distance using certain eigenvalues. Now, let $\tilde{\boldsymbol{\Lambda}}$ be a diagonal matrix which satisfies $\boldsymbol{\Lambda}^{-1} \approx \left(\hat{\boldsymbol{\Psi}}\tilde{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}^{\mathrm{T}}\right)^{-1}$. Namely, $\tilde{\boldsymbol{\Lambda}}$ is defined as

$$\tilde{\boldsymbol{\Lambda}} = \mathrm{D}\left(\hat{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\Psi}}\right), \tag{4}$$

where D is a function which returns diagonal elements of the matrix. $\tilde{\boldsymbol{\Lambda}}$ is the eigenvalues which compensate estimation errors of eigenvectors. The justification of ignoring the non-diagonal elements of $\hat{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\Psi}}$ is confirmed by the experiment in Sect. 3.1.

$\tilde{\boldsymbol{\Lambda}}$ is defined by the true eigenvalues ($\boldsymbol{\Lambda}$) and estimation errors of eigenvectors ($\hat{\boldsymbol{\Psi}}$). $\hat{\boldsymbol{\Psi}}$ is defined by using the true eigenvectors ($\boldsymbol{\Phi}$). Since we assume that $\boldsymbol{\Phi}$ are unknown, we can not observe $\hat{\boldsymbol{\Psi}}$. However, we can observe the probability density function of $\hat{\boldsymbol{\Psi}}$ because the probability density function of $\hat{\boldsymbol{\Psi}}$ depends only on the dimensionality of feature vectors, sample size, the true eigenvalues and the sample eigenvalues, and does not depend on the true eigenvectors (See Appendix). Therefore, the expectation of $\hat{\boldsymbol{\Psi}}$ is observable even if the true eigenvectors are unknown. Let $\acute{\boldsymbol{\Psi}}$ be random estimation error matrix of eigenvectors. Eq. (4) is rewritten as

$$\acute{\tilde{\boldsymbol{\Lambda}}} = \mathrm{D}\left(\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Lambda}\acute{\boldsymbol{\Psi}}\right), \tag{5}$$

where $\acute{\tilde{\boldsymbol{\Lambda}}}$ is a diagonal matrix of the random variables representing the eigenvalues for the compensation.

The conditional expectation of Eq. (5) given $\hat{\boldsymbol{\Lambda}}$ is calculated as

$$\begin{aligned}\tilde{\tilde{\boldsymbol{\Lambda}}} = \mathrm{E}\left[\acute{\tilde{\boldsymbol{\Lambda}}}\middle|\,\hat{\boldsymbol{\Lambda}}\right] &= \mathrm{E}\left[\mathrm{D}\left(\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Lambda}\acute{\boldsymbol{\Psi}}\right)\middle|\,\hat{\boldsymbol{\Lambda}}\right] \\ &= \mathrm{D}\left(\mathrm{E}\left[\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Lambda}\acute{\boldsymbol{\Psi}}\middle|\,\hat{\boldsymbol{\Lambda}}\right]\right),\end{aligned} \tag{6}$$

where $\tilde{\tilde{\boldsymbol{\Lambda}}} = \mathrm{diag}\left(\tilde{\tilde{\lambda}}_1, \tilde{\tilde{\lambda}}_2, \ldots, \tilde{\tilde{\lambda}}_d\right)$. The ith diagonal element of Eq. (6) is

$$\tilde{\tilde{\lambda}}_i = \mathrm{E}\left[\sum_{j=1}^{d}\left\{\acute{\psi}_{ji}\right\}^2\lambda_j\middle|\,\hat{\boldsymbol{\Lambda}}\right] \tag{7}$$

$$= \sum_{j=1}^{d}\mathrm{E}\left[\left\{\acute{\psi}_{ji}\right\}^2\middle|\,\hat{\boldsymbol{\Lambda}}\right]\lambda_j. \tag{8}$$

Letting

$$\left\{\tilde{\psi}_{ji}\right\}^2 = \mathrm{E}\left[\left\{\acute{\psi}_{ji}\right\}^2\middle|\,\hat{\boldsymbol{\Lambda}}\right], \tag{9}$$

we obtain

$$\tilde{\tilde{\lambda}}_i = \sum_{j=1}^{d} \left\{ \tilde{\psi}_{ji} \right\}^2 \lambda_j. \tag{10}$$

2.2 Calculation of Eq. (10)

We show a way to calculate Eq. (10). We will begin by generalizing the conditional expectation of Eq. (9). Let $f(\acute{\boldsymbol{\Psi}})$ be an arbitrary function of $\acute{\boldsymbol{\Psi}}$. The integral representation of the conditional expectation of $f(\acute{\boldsymbol{\Psi}})$ is given as

$$\mathrm{E}\left[f(\acute{\boldsymbol{\Psi}}) \middle| \hat{\boldsymbol{\Lambda}} \right] = \int_{\acute{\boldsymbol{\Psi}}} f(\acute{\boldsymbol{\Psi}}) \mathrm{P}(\acute{\boldsymbol{\Psi}} | \hat{\boldsymbol{\Lambda}}) d\acute{\boldsymbol{\Psi}}, \tag{11}$$

where $\mathrm{P}(\acute{\boldsymbol{\Psi}}|\hat{\boldsymbol{\Lambda}})$ is the probability density function of estimation errors of eigenvectors. Obtaining exact value of $\mathrm{P}(\acute{\boldsymbol{\Psi}}|\hat{\boldsymbol{\Lambda}})$ is difficult especially for large d. In this paper, Eq. (11) is estimated by Conditional Monte Carlo Method [5]. By assuming $f(\acute{\boldsymbol{\Psi}}) = \left\{ \acute{\psi}_{ji} \right\}^2$ in Eq. (11), $\left\{ \tilde{\psi}_{ji} \right\}^2$ of Eq. (9) is obtained. Therefore, we can calculate $\tilde{\tilde{\lambda}}_i$ in Eq. (10).

To carry out Conditional Monte Carlo Method, we deform the right side of Eq. (11). For the preparation of the deformation, let $\acute{\boldsymbol{\Sigma}}$ be a random symmetric matrix and $\acute{\boldsymbol{\Lambda}}$ be a random diagonal matrix that satisfies $\acute{\boldsymbol{\Sigma}} = \acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}$. Since the probability density function of estimation errors of eigenvectors ($\acute{\boldsymbol{\Psi}}$) is independent of the true eigenvectors ($\boldsymbol{\Phi}$), $\boldsymbol{\Phi} = \boldsymbol{I}$ is assumed without loss of generality, and $\acute{\boldsymbol{\Phi}} = \acute{\boldsymbol{\Psi}}$ immediately. Therefore, $\acute{\boldsymbol{\Sigma}} = \acute{\boldsymbol{\Phi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Phi}}^{\mathrm{T}}$. Hence the probability density of $\acute{\boldsymbol{\Sigma}}$ is given as the Wishart distribution (See Appendix). We have $\mathrm{P}(\acute{\boldsymbol{\Sigma}}) = \mathrm{P}(\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}) = \mathrm{P}(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}}) J(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}})$, where Jacobian $J(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}}) = \frac{d\acute{\boldsymbol{\Psi}} d\acute{\boldsymbol{\Lambda}}}{d\acute{\boldsymbol{\Sigma}}}$. We also have $\mathrm{P}(\acute{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}) = \mathrm{P}(\acute{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Lambda}}) J(\acute{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Lambda}})$ since $\hat{\boldsymbol{\Lambda}}$ is a realization of random variable $\acute{\boldsymbol{\Lambda}}$. Let $g(\acute{\boldsymbol{\Lambda}})$ be an arbitrary function and $G = \int_{\acute{\boldsymbol{\Lambda}}} g(\acute{\boldsymbol{\Lambda}}) d\acute{\boldsymbol{\Lambda}}$.

Based on the preparation above, the right side of Eq. (11) can be deformed as

$$\begin{aligned}
&\int_{\acute{\boldsymbol{\Psi}}} f(\acute{\boldsymbol{\Psi}}) \mathrm{P}(\acute{\boldsymbol{\Psi}}|\hat{\boldsymbol{\Lambda}}) d\acute{\boldsymbol{\Psi}} \\
&= \int_{\acute{\boldsymbol{\Psi}}} f(\acute{\boldsymbol{\Psi}}) \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}}|\hat{\boldsymbol{\Lambda}})}{G} \left[\int_{\acute{\boldsymbol{\Lambda}}} g(\acute{\boldsymbol{\Lambda}}) d\acute{\boldsymbol{\Lambda}} \right] d\acute{\boldsymbol{\Psi}} \\
&= \int_{\acute{\boldsymbol{\Psi}} \times \acute{\boldsymbol{\Lambda}}} f(\acute{\boldsymbol{\Psi}}) \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Lambda}})}{\mathrm{P}(\hat{\boldsymbol{\Lambda}})} \frac{g(\acute{\boldsymbol{\Lambda}})}{G} d\acute{\boldsymbol{\Psi}} d\acute{\boldsymbol{\Lambda}} \\
&= \frac{1}{\mathrm{P}(\hat{\boldsymbol{\Lambda}})} \int_{\acute{\boldsymbol{\Sigma}}} f(\acute{\boldsymbol{\Psi}}) \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}})}{J(\acute{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Lambda}})} \frac{g(\acute{\boldsymbol{\Lambda}})}{G} J(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}}) d\acute{\boldsymbol{\Sigma}} \\
&= \frac{1}{\mathrm{P}(\hat{\boldsymbol{\Lambda}})} \int_{\acute{\boldsymbol{\Sigma}}} f(\acute{\boldsymbol{\Psi}}) w_0(\acute{\boldsymbol{\Sigma}}; \hat{\boldsymbol{\Lambda}}) \mathrm{P}(\acute{\boldsymbol{\Sigma}}) d\acute{\boldsymbol{\Sigma}},
\end{aligned} \tag{12}$$

where

$$w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}}) = \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}})}{J(\acute{\boldsymbol{\Psi}},\hat{\boldsymbol{\Lambda}})}\frac{J(\acute{\boldsymbol{\Psi}},\acute{\boldsymbol{\Lambda}})}{\mathrm{P}(\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}})}\frac{g(\acute{\boldsymbol{\Lambda}})}{G}. \tag{13}$$

Eq. (12) means that the expectation of $f(\acute{\boldsymbol{\Psi}})$ with probability density $\mathrm{P}(\acute{\boldsymbol{\Psi}}|\hat{\boldsymbol{\Lambda}})$ is the same as the expectation of $f(\acute{\boldsymbol{\Psi}})w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}})\frac{1}{\mathrm{P}(\hat{\boldsymbol{\Lambda}})}$ with probability density $\mathrm{P}(\acute{\boldsymbol{\Sigma}})$. Therefore, Eq. (11) can be calculated using the random vectors following normal distribution. By substituting Eq. (19) and Eq. (20) into Eq. (13), we have

$$w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}}) = \frac{|\hat{\boldsymbol{\Lambda}}|^{\frac{1}{2}(n-p-2)}\prod_{i<j}^{d}(\hat{\lambda}_i-\hat{\lambda}_j)}{|\acute{\boldsymbol{\Lambda}}|^{\frac{1}{2}(n-p-2)}\prod_{i<j}^{d}(\acute{\lambda}_i-\acute{\lambda}_j)}\frac{\exp\left(-\frac{n-1}{2}\operatorname{tr}\boldsymbol{\Lambda}^{-1}\acute{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\right)}{\exp\left(-\frac{n-1}{2}\operatorname{tr}\boldsymbol{\Lambda}^{-1}\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\right)}\frac{g(\acute{\boldsymbol{\Lambda}})}{G}. \tag{14}$$

Since calculating $\mathrm{P}(\hat{\boldsymbol{\Lambda}})$ is hard, though the formula of $\mathrm{P}(\hat{\boldsymbol{\Lambda}})$ is known, we show an another approach to obtain $\mathrm{P}(\hat{\boldsymbol{\Lambda}})$. When $f(\acute{\boldsymbol{\Psi}})=1$ is assumed in Eq. (12), $\frac{1}{\mathrm{P}(\hat{\boldsymbol{\Lambda}})}\int_{\acute{\boldsymbol{\Sigma}}} w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}})\mathrm{P}(\acute{\boldsymbol{\Sigma}})d\acute{\boldsymbol{\Sigma}}=1$. Then, we obtain a calculatable solution

$$\mathrm{P}(\hat{\boldsymbol{\Lambda}}) = \int_{\acute{\boldsymbol{\Sigma}}} w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}})\mathrm{P}(\acute{\boldsymbol{\Sigma}})d\acute{\boldsymbol{\Sigma}}. \tag{15}$$

Let us assume $f(\acute{\boldsymbol{\Psi}}) = \left\{\acute{\psi}_{ji}\right\}^2$. Since Eq. (12) is the right side of Eq. (9), by substituting Eq. (15) into Eq. (12), Eq. (9) is obtained as $\left\{\tilde{\psi}_{ji}\right\}^2 = \frac{\int_{\acute{\boldsymbol{\Sigma}}}\left\{\acute{\psi}_{ji}\right\}^2 w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}})\mathrm{P}(\acute{\boldsymbol{\Sigma}})d\acute{\boldsymbol{\Sigma}}}{\int_{\acute{\boldsymbol{\Sigma}}} w_0(\acute{\boldsymbol{\Sigma}};\hat{\boldsymbol{\Lambda}})\mathrm{P}(\acute{\boldsymbol{\Sigma}})d\acute{\boldsymbol{\Sigma}}}$. By replacing the integrals of both numerator and denominator with the averages of $\acute{\boldsymbol{\Sigma}}_k, k=1,\ldots,t$, and erasing $\frac{1}{t}$ of both numerator and denominator, we obtain

$$\left\{\tilde{\psi}_{ji}\right\}^2 = \frac{\sum_{k=1}^{t}\left\{\acute{\psi}_{k,ji}\right\}^2 w_0(\acute{\boldsymbol{\Psi}}_k\acute{\boldsymbol{\Lambda}}_k\acute{\boldsymbol{\Psi}}_k^{\mathrm{T}};\hat{\boldsymbol{\Lambda}})}{\sum_{k=1}^{t} w_0(\acute{\boldsymbol{\Psi}}_k\acute{\boldsymbol{\Lambda}}_k\acute{\boldsymbol{\Psi}}_k^{\mathrm{T}};\hat{\boldsymbol{\Lambda}})}, \tag{16}$$

where $\acute{\boldsymbol{\Psi}}_k$ is the eigenvectors of $\acute{\boldsymbol{\Sigma}}_k$ and $\acute{\psi}_{k,ji}$ is the ji element of $\acute{\boldsymbol{\Psi}}_k$.

From the discussion above, we have an algorithm to estimate $\tilde{\tilde{\lambda}}_i$ in Eq. (10). The algorithm is shown in Algorithm 1. n is the available sample size for training and $\breve{\boldsymbol{\Lambda}}$ is the matrix which represents the true eigenvalues or the corrected eigenvalues by an existing method for correcting estimation errors of eigenvalues. $g(\acute{\boldsymbol{\Lambda}}_k)=\frac{1}{t}$ in this paper.

2.3 Procedure for Estimating the True Mahalanobis Distance from the Sample Eigenvectors

When n sample vectors are available for training, we can estimate the true Mahalanobis distance by the following procedure.

Algorithm 1 Estimation of $\tilde{\tilde{\lambda}}_i$

1: Create nt sample vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{nt}$ that follows normal distribution $\mathrm{N}(\boldsymbol{0}, \breve{\boldsymbol{\Lambda}})$ by using random numbers.
2: **for** $k = 1$ to t **do**
3: Estimate the sample covariance matrix $\acute{\boldsymbol{\Sigma}}_k$ from $\boldsymbol{X}_{n(k-1)+1}, \ldots, \boldsymbol{X}_{nk}$.
4: Obtain the eigenvalues $\acute{\boldsymbol{\Lambda}}_k$ and the eigenvectors $\acute{\boldsymbol{\Psi}}_k$ of $\acute{\boldsymbol{\Sigma}}_k$.
5: **end for**
6: Calculate $\left\{\tilde{\Psi}_{ji}\right\}^2$ in Eq. (16).
7: Calculate $\tilde{\tilde{\lambda}}_i$ in Eq. (10)

1. The sample eigenvalues ($\hat{\boldsymbol{\Lambda}}$) and the sample eigenvectors ($\hat{\boldsymbol{\Phi}}$) are calculated from available n sample vectors.
2. The estimation errors of the sample eigenvalues are corrected by an existing method, e.g. Sakai's method [1, 2] or James-Stein estimator [3].
3. $\tilde{\tilde{\lambda}}_i$ is calculated by Algorithm 1.
4. Use $\tilde{\tilde{\lambda}}_i$ as eigenvalue with the sample eigenvectors for recognition.

3 Performance Evaluation of the Proposed Method

3.1 Estimated Mahalanobis Distance

The first experiment was performed to confirm that the proposed method has the ability to estimate the true Mahalanobis distance correctly from the sample eigenvectors. To show the ability, t_{ij}, e_{ij} and p_{ij} were calculated. t_{ij} is the true Mahalanobis distance between the jth input vector of class i and the true mean vector of the class to which the input vector belongs. e_{ij} was the calculated Mahalanobis distance from the true mean vectors, the true eigenvalues and the sample eigenvectors. p_{ij} was the one from the true mean vectors, the eigenvalues modified by the proposed method and the sample eigenvectors. Then, the average ratios of the Mahalanobis distances to the true ones $r_{\mathrm{e}} = \frac{1}{cs}\sum_{i=1}^{c}\sum_{j=1}^{s}\frac{e_{ij}}{t_{ij}}$ and $r_{\mathrm{p}} = \frac{1}{cs}\sum_{i=1}^{c}\sum_{j=1}^{s}\frac{p_{ij}}{t_{ij}}$ were compared, where c was the number of classes and s was the number of samples for testing. Here, $c = 10$ and $s = 1000$.

The experiments were performed on artificial feature vectors. The vectors of class i followed normal distribution $\mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ were calculated from feature vectors of actual character samples. The feature vectors of actual character samples were created as follows: The character images of digit samples in NIST Special Database 19 [6] were normalized nonlinearly [7] to fit in a 64×64 square, and 196-dimensional Directional Element Features [8] were extracted. The digit samples were sorted by class, and shuffled within the class in advance. $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ of class i were calculated from 36,000 feature vectors of the actual character samples of class i.

The parameter t of Algorithm 1 was 10,000. The average ratios of the Mahalanobis distances r_{e} and r_{p} are shown in Fig. 1. r_{e} are far larger than one

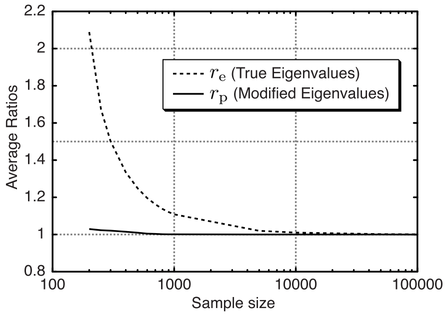

Fig. 1. The average ratios of the Mahalanobis distances.

for small sample sizes. However, r_p are almost one for all sample sizes. This means the true Mahalanobis distance is precisely estimated by the proposed method. Moreover, it shows the justification of the approximation of ignoring the non-diagonal elements of $\hat{\boldsymbol{\Psi}}^\mathrm{T}\boldsymbol{\Lambda}\hat{\boldsymbol{\Psi}}$ described in Sect. 2.

3.2 Recognition Accuracy

The second experiment was carried out to confirm the effectiveness of the Mahalanobis distance estimated by the proposed method as a classifier. Character recognition experiments were performed by using three kinds of dictionaries "Control," "True eigenvalue" and "Proposed method." The dictionaries had common sample mean vectors, common sample eigenvectors and different eigenvalues. "Control" had the sample eigenvalues, "True eigenvalue" had the true eigenvalues, and "Proposed method" had the eigenvalues modified by the proposed method. The recognition rates of the dictionaries were compared.

The experiments were performed on the feature vectors of actual character samples and the artificial feature vectors described in Sect. 3.1. Since the true eigenvalues are not available for feature vectors of actual character samples, the eigenvalues corrected by Sakai's method [1, 2] were used in place of the true eigenvalues. The parameter t of Algorithm 1 was 10,000. 1,000 samples were used for testing. The recognition rates of the experiment on the use of feature vectors of actual character samples and artificial feature vectors are shown in Fig.2(a) and Fig.2(b) respectively. Both of the figures show that the recognition rate of "Proposed method" is higher than that of "True eigenvalue." Therefore, the effectiveness of the proposed method is confirmed. Although the feature vectors of actual character samples does not follow normal distribution, the proposed method is effective. When the number of training samples is small, the difference between the recognition rates of "True eigenvalue" and "Proposed method" is large. The difference decreases as the sample size increases. This seems to depend on the amount of estimation errors of eigenvectors. The figures also show that the recognition rate of "True eigenvalue" is higher than that of "Control."

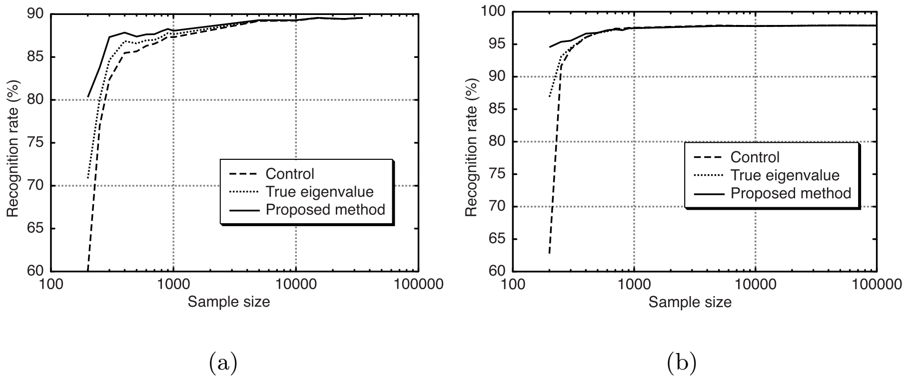

Fig. 2. The recognition rates. (a)On the use of the feature vectors of actual character samples. (b)On the use of the artificial feature vectors.

4 Conclusion

In this paper, we aimed to achieve high-performance pattern recognition without many training samples. We presented a method to estimate the true Mahalanobis distance from the sample eigenvectors. Recognition experiments show that the proposed method has the ability to estimate the true Mahalanobis distance even if the sample size is small. The eigenvalues modified by the proposed method achieve better recognition rate than the true eigenvalues. The proposed method is useful for the practical applications of pattern recognition since the proposed method is effective without any hyper-parameters, especially when the sample size is small.

References

1. Sakai, M., Yoneda, M., Hase, H.: A new robust quadratic discriminant function. In: Proc. ICPR. (1998) 99–102
2. Sakai, M., Yoneda, M., Hase, H., Maruyama, H., Naoe, M.: A quadratic discriminant function based on bias rectification of eigenvalues. Trans. IEICE **J82-D-II** (1999) 631–640
3. James, W., Stein, C.: Estimation with quadratic loss. In: Proc. 4th Berkeley Symp. on Math. Statist. and Prob. (1961) 361–379
4. Iwamura, M., Omachi, S., Aso, H.: A modification of eigenvalues to compensate estimation errors of eigenvectors. In: Proc. ICPR. Volume 2., Barcelona, Spain (2000) 378–381
5. Hammersley, J.M., Handscomb, D.C.: 6. In: Monte Carlo Methods. Methuen, London (1964)
6. Grother, P.J.: NIST special database 19 handprinted forms and characters database. Technical report, National Institute of Standards and Technology (1995)

7. Yamada, H., Yamamoto, Saito, T.: A nonlinear normalization method for handprinted kanji character recognition — line density equalization —. Pattern Recognition **23** (1990) 1023–1029
8. Sun, N., Uchiyama, Y., Ichimura, H., Aso, H., Kimura, M.: Intelligent recognition of characters using associative matching technique. In: Proc. Pacific Rim Int'l Conf. Artificial Intelligence (PRICAI'90). (1990) 546–551
9. Muirhead, R.J.: Aspects of Multivariate Statistical Theory. John Wiley & Sons, Inc., New York (1982)

A Probability Density Function of Estimation Errors of Eigenvectors

Let the d-dimensional column vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be random sample vectors from normal distribution $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. The distribution of the random matrix $\boldsymbol{W} = \sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}}$ is the Wishart distribution $\boldsymbol{W}_d(n, \boldsymbol{\Sigma})$. The probability density function of $\boldsymbol{W}$ is

$$\mathrm{P}(\boldsymbol{W}|\boldsymbol{\Sigma}) = \frac{v(d,n)|\boldsymbol{W}|^{\frac{1}{2}(n-d-1)}}{|\boldsymbol{\Sigma}|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{W}\right), \tag{17}$$

where $v(d,n) = \frac{1}{2^{\frac{1}{2}nd} \pi^{\frac{1}{4}d(d-1)} \prod_{i=1}^{d} \boldsymbol{\Gamma}\left[\frac{1}{2}(n+1-i)\right]}$.

Let $\acute{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{t=1}^{n} (\boldsymbol{X}_t - \acute{\boldsymbol{\mu}}) (\boldsymbol{X}_t - \acute{\boldsymbol{\mu}})^{\mathrm{T}}$ be the sample covariance matrix and $\acute{\boldsymbol{\mu}} = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{X}_t$ be the sample mean vector. The distribution of $\acute{\boldsymbol{\Sigma}}$ is given as $\boldsymbol{W}_d(n-1, \frac{1}{n-1}\boldsymbol{\Sigma})$ and the probability density function is as follows [9]:

$$\mathrm{P}(\acute{\boldsymbol{\Sigma}}|\boldsymbol{\Sigma}) = (n-1)^{\frac{1}{2}(n-1)d} v(d, n-1) \frac{|\acute{\boldsymbol{\Sigma}}|^{\frac{1}{2}(n-d-2)} \exp\left(-\frac{n-1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \acute{\boldsymbol{\Sigma}}\right)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}(n-1)}}. \tag{18}$$

$\boldsymbol{\Sigma}$ and $\acute{\boldsymbol{\Sigma}}$ are decomposed into $\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathrm{T}}$ and $\acute{\boldsymbol{\Phi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Phi}}^{\mathrm{T}} = \boldsymbol{\Phi}\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}$. Eq. (18) is deformed as

$$\begin{aligned} \mathrm{P}(\boldsymbol{\Phi}\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}|\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathrm{T}}) &= (n-1)^{\frac{1}{2}(n-1)d} v(d, n-1) \\ &\quad \cdot \frac{|\acute{\boldsymbol{\Lambda}}|^{\frac{1}{2}(n-d-2)} \exp\left(-\frac{n-1}{2} \operatorname{tr} \boldsymbol{\Lambda}^{-1} \acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}\right)}{|\boldsymbol{\Lambda}|^{\frac{1}{2}(n-1)}}. \end{aligned} \tag{19}$$

Since the right side of Eq. (19) is independent of $\boldsymbol{\Phi}$, the left side of Eq. (19) is simplified as $\mathrm{P}(\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}}|\boldsymbol{\Lambda})$, and denoted as $\mathrm{P}(\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}})$ by omitting the condition.

Finally, the probability density function of estimation errors of eigenvectors are given as $\mathrm{P}(\acute{\boldsymbol{\Psi}}|\acute{\boldsymbol{\Lambda}}) = \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}},\acute{\boldsymbol{\Lambda}})}{\mathrm{P}(\acute{\boldsymbol{\Lambda}})} = \frac{\mathrm{P}(\acute{\boldsymbol{\Psi}}\acute{\boldsymbol{\Lambda}}\acute{\boldsymbol{\Psi}}^{\mathrm{T}})}{\mathrm{P}(\acute{\boldsymbol{\Lambda}})J(\acute{\boldsymbol{\Psi}},\acute{\boldsymbol{\Lambda}})}$, where Jacobian $J(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}})$ is as follows [9]:

$$J(\acute{\boldsymbol{\Psi}}, \acute{\boldsymbol{\Lambda}}) = \frac{\prod_{i=1}^{d} \boldsymbol{\Gamma}\left[\frac{1}{2}(d+1-i)\right]}{2^d \pi^{\frac{1}{4}d(d+1)}} \frac{1}{\prod_{i<j}^{d} (\acute{\lambda}_i - \acute{\lambda}_j)}. \tag{20}$$