Selecting Principal Components in a Two-Stage LDA Algorithm

Manli Zhu, Aleix M. Martínez
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH 43210
{zhum,aleix}@ece.osu.edu

Abstract

Linear Discriminant Analysis (LDA) is a well-known and important tool in pattern recognition with potential applications in many areas of research. The most famous and used formulation of LDA is that given by the Fisher-Rao criterion, where the problem reduces to a simple simultaneous diagonalization of two symmetric, positive-definite matrices, $\mathbf{A}$ *and* $\mathbf{B}$*; i.e.* $\mathbf{B}^{-1}\mathbf{AV} = \mathbf{V}\Lambda$*. Here,* $\mathbf{A}$ *defines the metric to be maximized, while* $\mathbf{B}$ *defines the metric to be minimized. However, when* $\mathbf{B}$ *has near-zero eigenvalues, the Fisher-Rao criterion gets dominated by these. While this works well when such small variances describe vectors where most of the discriminant information is, the results will be incorrect when these small variances are caused by noise. Knowing which of these near-zero values are to be used and which need to be eliminated is a challenging yet fundamental task in LDA. This paper presents a criterion for the selection of those vectors of* $\mathbf{B}$ *that are best for classification. The proposed solution is based on a simple factorization of* $\mathbf{B}^{-1}\mathbf{A}$ *that permits the re-ordering of the eigenvectors of* $\mathbf{B}$ *without the need to effect the end result. This allows us to readily eliminate the noisy vectors while keeping the most discriminant ones. A theoretical basis for these results is presented along with extensive experimental results to validate the claims.*

1. Instroduction

Over the years, Linear Discriminant Analysis (LDA) has been an instrumental tool in several areas of research. Growing in popularity, LDA has been tried on any imaginable problem. Succeeding in some cases, failing in many others. The fact that LDA has failed in several application has prompted researchers to define new (related) algorithms [3, 7]. However, a much overlooked problem of LDA lays on the instability of the definition of discriminability given by the Fisher-Rao criterion to be maximized [10].

The Fisher-Rao criterion is defined as $J(\mathbf{V}) = \frac{\mathbf{V}^T\mathbf{AV}}{\mathbf{V}^T\mathbf{BV}}$, where $\mathbf{V}$ are the discriminant feature vectors, and $\mathbf{A}$ and $\mathbf{B}$ are two symmetric, positive-definite matrices defining the metrics to be maximized and minimized, respectively [5, 11]. $\mathbf{A}$ is usually taken to be the between-class scatter matrix, $\mathbf{S}_B$. Typical metrics for $\mathbf{B}$ are the within-class scatter and the sample covariance matrices, $\mathbf{S}_W$ and $\mathbf{S}_m$ [6]. For simplicity, let us assume $\mathbf{A} = \mathbf{S}_B$ and $\mathbf{B} = \mathbf{S}_m$ (keeping in mind that our results extend to all other options). This means that we will be selecting our discriminant vectors from $\mathbf{V}$, where $\mathbf{S}_m^{-1}\mathbf{S}_B\mathbf{V} = \mathbf{V}\Lambda$ and $\Lambda = diag\,(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix of corresponding eigenvalues (discriminant values) with $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$.

A main reason for the instability of $J(\mathbf{V})$ comes from the fact that, in practical settings, some of the eigenvectors of $\mathbf{S}_m$ are generally associated to near-zero values. Close analysis of $J(\mathbf{V})$ reveals that this criterion is steered by such eigenvectors – because these will be the ones maximizing $J(\mathbf{V})$. Unfortunately, while some of these near-zero eigenvalues are due to their true underlying discriminant power, other near-zero eigenvalues are caused by small noise in the observations (i.e., noisy samples). *Discerning which of these near-zero eigenvalues are due to discriminant capabilities and which to noise is essential to the successful use of LDA.* And, this is indeed the goal of this paper.

In attempting to solve this problem, researchers have previously looked at a related issue – that given by the singularity of $\mathbf{S}_m$. For example, when $\mathbf{S}_m$ cannot be inverted, LDA can be preceded by a Principle Component Analysis (PCA) step (i.e., $\mathbf{S}_m\mathbf{U} = \mathbf{U}\Lambda$) where the original dimensionality of the data is reduced to one of s-dimensions where $\mathbf{S}_m$ is no longer singular, $p \geq rank(\mathbf{S}_m) \geq s$ [1, 12, 4]. This method is usually referred to as PCA-LDA. Unfortunately, the performance of the PCA-LDA method varies widely when the set of PCs selected in the PCA step varies. The reason for this is simple: Since no prior information is available to determine which near-zero eigenvalues are caused by noise and which ones are highly discriminant, the PCA step is left to carry out a blind search and random

elimination of bases from $\mathbf{S}_m$ [9].

With this goal in mind, several strategies for the selection of the most adequate set of PCs of $\mathbf{S}_m$ have been proposed. A much used solution examines the plot of the eigenvalues (conveniently ordered from larger to smaller) and then searches for the "elbow" where the values fall sharply. Unfortunately, there exist too many applications where the eigenvalue plot drifts without any obvious cutting point [14]. Alternatively, the first k PCs that account for $d = \sum_{i=1}^{k} \lambda_{X_i} / \sum_{i=1}^{p} \lambda_{X_i}$ of the total variance can be used. Here, d must be specified by the user. In [7], it is suggested that a value of d between 70% and 90% preserves most of the information needed for representing most Gaussian-like distributions. But determining the most convenient value is problem-specific.

As mentioned earlier, an alternative is to modify the definition of the Fisher-Rao criterion. Most notably, [3, 2, 7] define the discriminatory power of a vector $\mathbf{u}_j$ as $J(\mathbf{u}_j) = \frac{\mathbf{u}_j^T \mathbf{S}_B \mathbf{u}_j}{\lambda_{X_j}}$, where $\mathbf{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ are the eigenvectors of $\mathbf{S}_m$, and $\{\lambda_{X_1}, \ldots, \lambda_{X_p}\}$ are the corresponding eigenvalues. However, it is unclear how many PCs would be needed to make this criterion useful in practice. Since the estimate of small eigenvalues are biased downwards, and the estimate of large eigenvalues are biased upwards, the reciprocal of eigenvalues can lead to an exaggerated influence of the low-variance PCs in the estimate of the discriminative power [7]. In addition, because $J(\mathbf{u}_j)$ is consistent with Fisher's criterion, it seems redundant to choose PCs according to $J(\mathbf{u}_j)$ before applying $J(\mathbf{V})$.

The primary problem with all the methods defined in the preceding paragraphs is that they bear no relation to the actual effectiveness of $\mathbf{S}_m^{-1}\mathbf{S}_B$. A more convenient solution would be given by a criterion that selects the eigenvectors $\mathbf{u}_i$ that are most consistent with those of $\mathbf{S}_B$. Note that noise will modify each matrix randomly and, hence, consistency (i.e. agreement) between $\mathbf{S}_B$ and $\mathbf{S}_m$ can (in general) only be due to a *real*, common discriminant source. Moreover, an analysis of the correlation between eigenvectors is invariant to their own (co-)variances. This means that our criterion will not be effected by either small or large eigenvalues.

To accomplish this, we will use a factorization of $\mathbf{S}_m^{-1}\mathbf{S}_B$ which permits the selection of those basis vectors of $\mathbf{S}_m$ that are best for maximizing $J(\mathbf{V})$. To show this, we will first prove that the results obtained with such a factorization are the same as those generated with $J(\mathbf{V})$. Noting that a reordering of the basis vectors of $\mathbf{S}_m$ does not change the end result, will facilitate the definition of our criterion. Our algorithm is thus a two-step process. First, we select those bases of $\mathbf{S}_m$ that are most correlated to the basis vectors of $\mathbf{S}_B$. Then, we use $J(\mathbf{V})$ to select the most discriminant vectors from $\mathbb{R}^p$, where p is the dimensionality of the original feature vectors.

Our theoretical analysis is detailed in Section 2. Experimental results are in Section 3. We conclude in Section 4.

2. A Two-step algorithm

2.1. Factorization of LDA

As is well-known, PCA provides the so-called spectral-decomposition of the (sample) covariance matrix of the data

$$\mathbf{S}_m = \lambda_{X_1}\mathbf{u}_1\mathbf{u}_1^T + \lambda_{X_2}\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_{X_P}\mathbf{u}_p\mathbf{u}_p^T,$$

where $\lambda_{X_1} \geq \lambda_{X_2} \geq \cdots \geq \lambda_{X_p}$ are the eigenvalues of $\mathbf{S}_m$ and $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ the corresponding eigenvectors [7]. Such representations have been extensively used to carry out all sorts of data analysis. For example, the first m (with $m \leq p$) PCs represent the m-dimensional subspace which carries most of the variance. And, when the last several PCs are associated to small changes, these can be used to detect near-constant linear relationships between variables [7].

Similar to the "spectral decomposition" of PCA, LDA can also be decomposed as a combination of eigenvalues and eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_m$ as [10]

$$\mathbf{S}_m^{-1}\mathbf{S}_B = \sum_{i=1}^{q}\sum_{j=1}^{p} \frac{\lambda_{B_i}}{\lambda_{X_j}} (\mathbf{u}_j^T \mathbf{w}_i)\mathbf{u}_j\mathbf{w}_i^T, \qquad (1)$$

where $\mathbf{S}_B = \sum_{i=1}^{q} \lambda_{B_i}\mathbf{w}_i^T\mathbf{w}_i$ and $q = rank(\mathbf{S}_B)$.

As in PCA, this can be used, for example, to eliminate some of the eigenvectors of $\mathbf{S}_m$. In general, we can select a subset of the p eigenvectors of $\mathbf{S}_m$ and use these to find the discriminant vectors of LDA. More formally, we define

$$\left(\mathbf{S}_m^{-1}S_B\right)_S = \sum_{j\in S}\sum_{i=1}^{q} \frac{\lambda_{B_i}}{\lambda_{X_j}} \mathbf{u}_j^T\mathbf{w}_i\mathbf{u}_j\mathbf{w}_i^T, \qquad (2)$$

where S is a subset of the set of s (with $s \leq p$) elements in $\{1, \cdots, p\}$. This selection is achieved using a criterion I. The r discriminant vectors of LDA are now given by the first r columns of $\mathbf{V}$, where

$$\left(\mathbf{S}_m^{-1}\mathbf{S}_B\right)_S \mathbf{V} = \mathbf{V}\Lambda. \qquad (3)$$

It is important to note that the method defined in Eqs. (2-3) is equivalent to the classical PCA-LDA algorithm iff the criterion for selecting the subset S is based on choosing the s eigenvectors $\mathbf{u}_i$ associated to the largest eigenvalues λ_{X_i}. This can be summarized in the following result (which is easy to prove using simple linear algebra manipulation).

Result 1 *The projection matrix* $\mathbf{V}$ *found by (3) and the projection matrix* $\mathbf{U}_s\mathbf{Y}$ *given by*

$$\mathbf{U}_s^T\mathbf{S}_B\mathbf{U}_s\mathbf{Y} = \mathbf{U}_s^T\mathbf{S}_m\mathbf{U}_s\mathbf{Y}\Lambda_Y,$$

are identical; where $\mathbf{U}_s = \{\mathbf{u}_j | j \in S\}$.

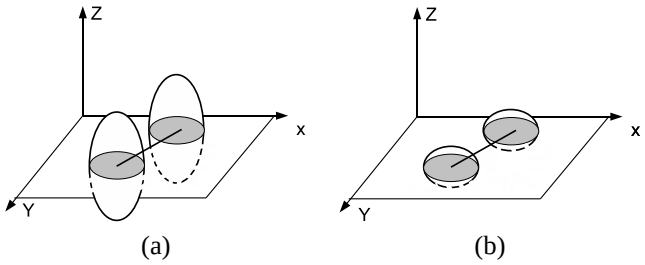


Figure 1. The discriminant information is not related to the eigenvalues of $\mathbf{S_m}$. In (a) the data has large variance about the $\mathbf{z}$ axis. In (b) the variance is small. In neither case $\mathbf{z}$ provides any discriminant information.

Unfortunately, as mentioned above, the results obtained using this selection of eigenvectors (i.e., those associated to the largest eigenvalues) will not generally result in a reasonable solution. We therefore need to find a new, robust criterion for I. This we will do next.

2.2. Our criterion

A main advantage of Eq. (1) is that it defines a way to reconstruct $\mathbf{S}_m^{-1}\mathbf{S}_B$ using the eigenvalues and eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_m$ without any explicit computation of $\mathbf{S}_m^{-1}$. This will not only saves us from extensive computations, but will prove instrumental for the understanding of the role the eigenvectors of each matrix play.

In the decomposition of Eq. (1), $\mathbf{u}_j^T\mathbf{w}_i$ acts as a "switch," controlling which $\mathbf{u}_j$'s are correlated to the $\mathbf{w}_i$'s. To be able to play a role in the selection of discriminant vectors $\mathbf{V}$, $\mathbf{u}_j$ has to be correlated to the $span(\mathbf{w}_i)$, i.e., $\mathbf{u}_j^T\mathbf{w}_i \neq 0$ for at least one i in $\{1, \cdots, q\}$, regardless of the value of λ_{X_j}. This is illustrated in Fig. 1. In (a) the two distributions have a large variance about the $\mathbf{z}$ axis. In (b) this variance is near-zero. In both cases, however, the eigenvector align with $\mathbf{z}$ is not correlated to the $span(\mathbf{S}_B)$ which defines the $\mathbf{x}$-$\mathbf{y}$-plane. This means that the eigenvector aligned with $\mathbf{z}$ *can be eliminated from consecutive computations because this does not carry any discriminant information*. S will thus be constructed with those eigenvectors that define the $\mathbf{x}$-$\mathbf{y}$-plane.

The above result can be easily formulated in the form of a correlation-based criterion as

$$I_j = \sum_{i=1}^{q} \left(\mathbf{u}_j^T\mathbf{w}_i\right)^2, \quad (1 \leq j \leq p).$$

To further illustrate the use of this criterion, we will use the data of two publicly available databases – the AR-face [9] and ETH-80 sets [8]. The AR-face database consists of

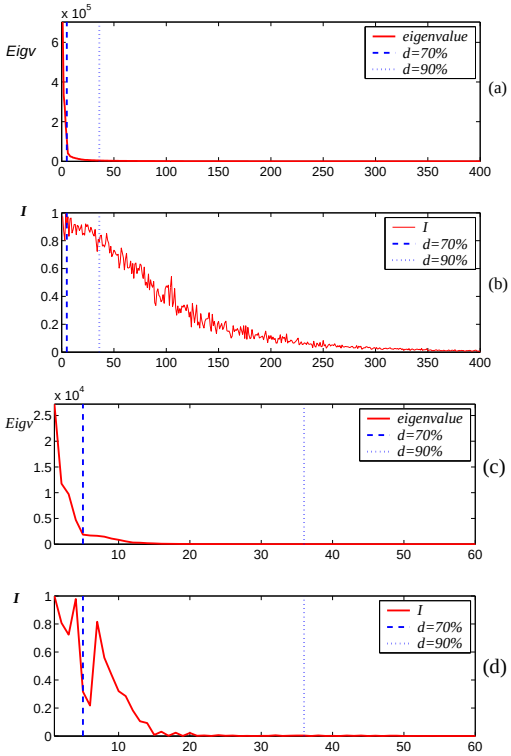


Figure 2. (a) Plotted here are the eigenvalues of $\mathbf{S}_m$ for the AR face database. (b) Shows the corresponding correlation values. (c-d) the same for the images of the ETH-80 set. Dashed and dotted lines represent the cutoff point with $d = 70\%$ and 90%, respectively.

frontal-view face images of 100 people. We use the first 13 images (taken during a first session, showing distinct facial composites, occlusions and illuminations) for training. The other 13 images are used for testing. Each image is cropped and vectorized to a standard size of 609 dimensions. The ETH-80 database contains images of eight object categories. Each category includes images of ten objects, each photographed from 41 orientations. In this case we cropped and vectorized each image to a standard size of 300 dimensions. To run our simulation, 1 of the 10 objects in each category is randomly chosen as testing vector. The remaining are used for training.

Were we to use the classical approach for selecting the eigenvectors of $\mathbf{S}_m$, we would need sort the eigenvectors $\mathbf{u}_i$ in descending order of eigenvalue. This is shown in Fig. 2(a) for those eigenvectors representing the AR-database, and in Fig. 2(c) for those drawn from the ETH-80 set. In each of this plots, we also show a dashed line to represent the cut where 70% of the total variance is kept and a dotted line to represent the 90% cut.

For comparison, Fig. 2(b,d) show the correlation value of the eigenvectors of $\mathbf{S}_m$ with those of $\mathbf{S}_B$. However, the order of the eigenvector is still the same as the one we had in Fig. 2(a,c). We did this to show that several of the eigenvalues associated to large eigenvalues are not very much correlated with $span(\mathbf{S}_B)$ whereas some eigenvectors associated to small eigenvalues are. Here, recall that those eigenvectors of $\mathbf{S}_m$ that are not correlated with $span(\mathbf{S}_B)$ cannot play a role when using $J(\mathbf{V})$. This clearly shows that when using the classical version of PCA-LDA we are indeed deleting eigenvectors of $\mathbf{S}_m$ that may contain highly discriminant information. In the AR set, even if we keep 90% of the variance!

Fortunately, we can now readily solve this problem with a simple re-ordering of the eigenvectors of $\mathbf{S}_m$. The new order will be given by the correlation value I_j (i.e. $I_1 \geq \cdots \geq I_p$). This is shown in Fig. 3.

To readily select those eigenvectors of $\mathbf{S}_m$ that are most correlated to the $span(\mathbf{S}_B)$, it is best to define a decreasingly monotonically function described by the discrete instances I_j. Since

$$\frac{1}{q}\sum_{j=1}^{p} I_j = \frac{1}{q}\sum_{j=1}^{p}\sum_{i=1}^{q}\left(\mathbf{u}_j^T\mathbf{w}_i\right)^2 = 1, \tag{4}$$

one can define the normalized result

$$f_j = \frac{1}{q}I_j = \frac{1}{q}\sum_{i=1}^{q}(\mathbf{u}_j^T\mathbf{w}_i)^2. \tag{5}$$

This function is monotonically decreasing and is guaranteed to start at $f_1 \approx \frac{1}{q}$ and end at $f_k = 0$ for some $k > 1$. Noting that when the number of classes is larger (i.e., q is large), the curve defined by f_j is less steep (because $span(\mathbf{w}_i)$ is a subspace of high dimensionality) and that when q is small, the curve goes to zero much faster (because $span(\mathbf{w}_i)$ represents a subspace of low dimensionality), allowing us to approximate the curve defined by $\{f_1, \ldots, f_p\}$ with an exponential function of the form $g_x = \lambda e^{-\lambda x}$, where $\lambda = f_1$. In Fig. 4 we show how g_x can approximate the shape of the original curve defined by the f_i describing the AR database. As seen in the figure, this approximation allow us to calculate the *confidence interval* h of g_x as

$$\int_0^a \lambda e^{-\lambda x}dx = h.$$

And, in turn, this is used to find the most adequate number s of eigenvectors of $\mathbf{S}_m$,

$$\begin{aligned} -e^{-\lambda x}|_0^a &= h \\ 1 - e^{-\lambda a} &= h \\ a &= -\ln(1-h)/\lambda, \end{aligned} \tag{6}$$

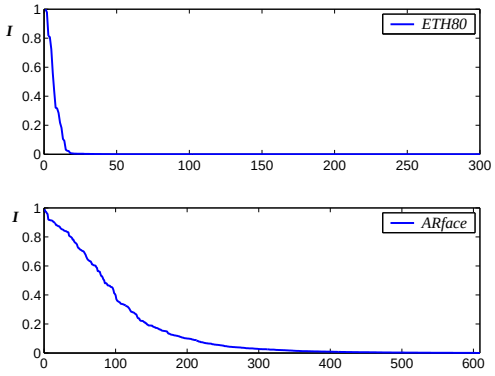


Figure 3. Shown here are the eigenvectors sorted according to the correlation value I_j.

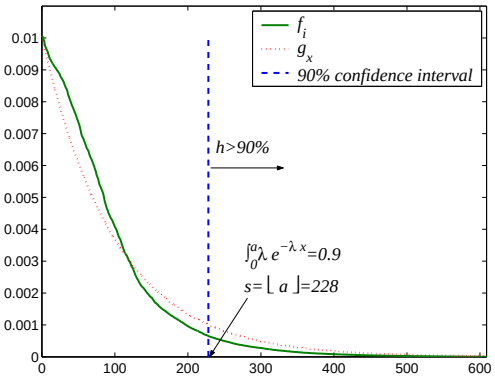


Figure 4. The confidence interval of the exponential function $g_x = \lambda e^{\lambda x}$ with probability 0.9, where $\lambda = f_1 = \frac{I_1}{q} \approx \frac{1}{99}$.

with $s = min(\lfloor a \rfloor, p)$. In Fig. 4 we show that, with a confidence interval of 90%, the optimal number of components s to use in the AR set is 228. After s has been determined the problem reduces to the use of Eq. (3).

3. Experimental results

We have tested our algorithm using three very distinct datasets. Two of these were the AR and ETH-80 databases defined earlier – for which the training and testing sets are also defined as above. The third was the Sitting Posture Distribution Maps (SPDM) of [13]. Here, the feature vectors were collected using a chair equipped with a pressure sensor sheet. Feature vectors correspond to $1,280$ pressure (force) values given at equally distributed points on the seating-pan of the chair. The goal is to classify each feature vector into one of the 10 possible sitting postures. This set includes five sample of each posture for a total of 50 people (i.e.,

5×50 samples per posture). From these 5 samples, three are randomly used for training and two for testing.

To illustrate our method, step by step, we first show the plot of the eigenvectors appropriately ordered according to the value of the criterion f_j for each of the database. This is shown in the first row of Fig. 5, where we have also plotted the approximation of these correlation values with the exponential function g_x. In this row of Fig. 5, we also see a dotted vertical line determining the set of eigenvectors within a 90% confidence interval.

In the second row of Fig. 5, we show the recognition rates (on a scale of 0 to 1) obtained for each of the possible values of eigenvectors of $\mathbf{S}_m$ used, i.e. s.[1] Here, we also use a dotted line to specify the recognition rate obtained with a confidence interval of 90%. As seen in the figure, our solution is very close to the maximum in all three cases. Most importantly, the exact same confidence interval can be efficiently used in all three datasets.

Fig. 5 also illustrates an important point. Note that, in all three cases, the shape of the curve describing the recognition rate decreases or remains stable after the inclusion of those eigenvectors that are within the 90% confidence interval. This demonstrates the theoretical point made earlier – those eigenvectors of $\mathbf{S}_m$ that are not correlated to $span(\mathbf{S}_B)$ are either associated to noise or do not contribute much to the discriminability.

For comparison purposes, we will now show results using the other commonly employed methods (that were presented earlier): *i)* where the number of eigenvectors is selected so that they account for d of the total variance, and *ii)* where we keep d of the discriminant power given by $J(\mathbf{u}_j)$. The plots of eigenvalue λ_{X_j} and discriminant power $J(\mathbf{u}_j)$ are shown in the first row of Figs. 6-7. In both cases we have used large dots to indicate the number of eigenvectors needed to account for d of the variance (Fig. 6) and d of the discriminant power given by $J(\mathbf{u}_j)$ (Fig. 7). Here, we used $d = 0.9, 0.91, \ldots, 0.99$ in Fig. 6-7 and the additional set $d = 0.5, 0.6, \ldots, 0.8$ in Fig. 7.

The second row in Figs. 7-6 shows the recognition rates obtained when s of the eigenvectors are kept. Note that these (recognition rate) curves are different in each image and different to those shown in the second row of Fig. 5, because the eigenvectors are sorted according to distinct criteria.

We note that there is no consistent value of d that can obtain good results in all three datasets. This is opposed to our method, where the same value can be consistently used in all three databases.

[1] Note that when s is equal to the maximum number of eigenvectors, the solution shown in the figure is that of Fisher-Rao's LDA.

4. Conclusions

This paper presents a two-stage LDA method where the PCA dimensions (e.g., eigenvectors of $\mathbf{S}_m$) that are less correlated to the $span(\mathbf{S}_B)$ are eliminated and discriminant vectors are computed using the decomposition of $\mathbf{S}_m^{-1}\mathbf{S}_B$ define in Eq. (3). Theoretical justification as well as experimental results have been presented. Comparison to most used techniques have shown the superiority of the proposed approach.

5. Acknowledgments

This work is supported in part by the US National Institutes of Health under grant R01 DC 005241.

References

[1] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *ECCV (1)*, pages 45–58, 1996.

[2] W.-C. Chang. On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics*, 32(3):267–275, 1983.

[3] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, London, 1982.

[4] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *JOSA-A*, 14:1724–1733, 1997.

[5] R. A. Fisher. The statistical utilization of multiple measurements. *Annals of Eugenics*, 8:376–386, 1938.

[6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[7] I. T. Jolliffe. *Principle Component Analysis*. Springer, 2002.

[8] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR03*, pages II:409–415, 2003.

[9] A. M. Martínez and A. C. Kak. Pca versus lda. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

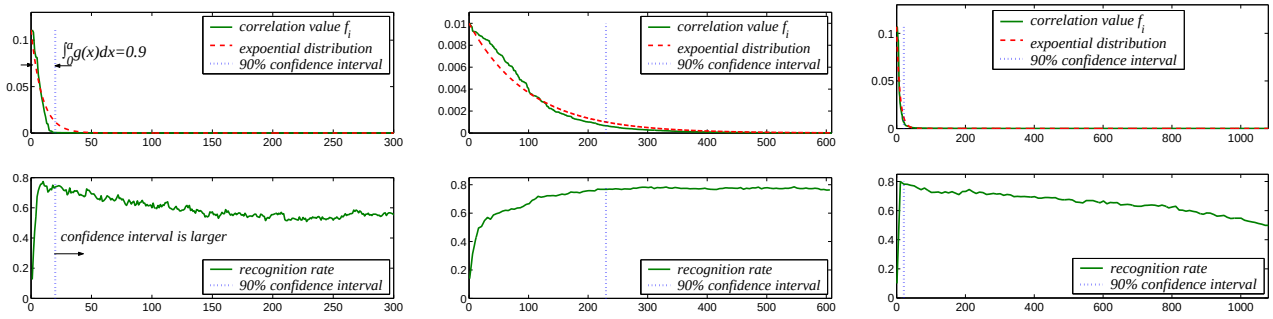
[10] A. M. Martínez and M. Zhu. Where are linear feature extraction methods applicable? *IEEE Trans on Pattern Analysis and Machine Intelligence*, 27(12):1934–1944, 2005.

[11] C. R. Rao. The utilization of multiple measurements in problems of biological classification. *J. Royal Statistical Soc., B*, 10:159–203, 1948.

[12] D. L. Swets and J. J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

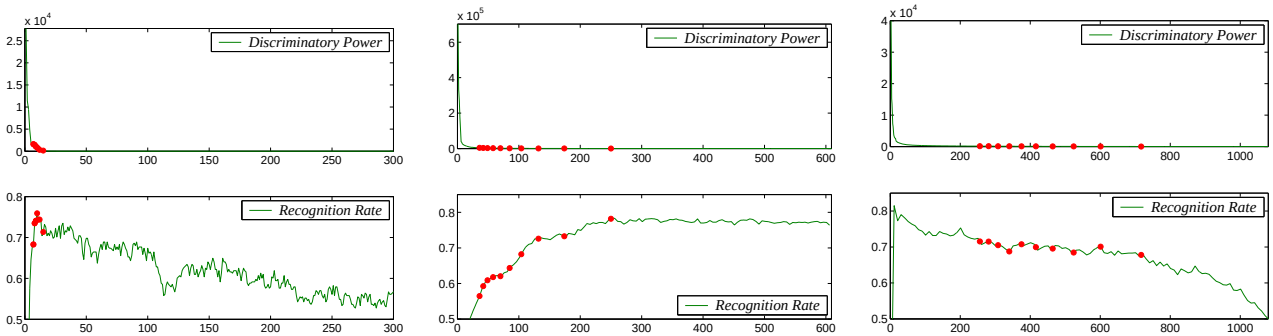
[13] H. Z. Tan, L. A. Slivovsky, and A. Pertland. A sensing chair using pressure distribution sensors. *IEEE Trans. on Mechatronics*, (3):261–268, 2001.

[14] A. Webb. *Statistical Pattern Recognition*. Newnes, 1999.



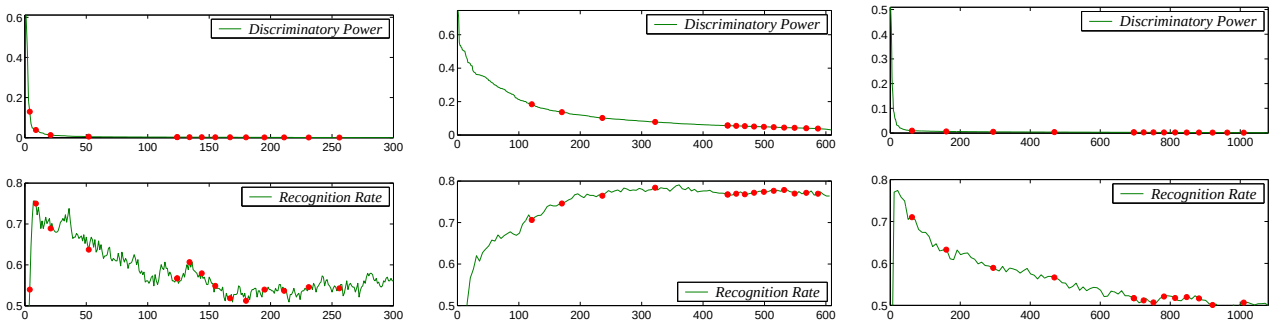
(a) (b) (c)

Figure 5. Selecting PCs according to correlation values f_i. In the first row, we sort eigenvectors in a decreasing order of f_i for three data-sets: (a) ETH-80. (b) ARface. (c) SPDM. The second row shows the corresponding recognition rates. The vertical dotted line represents the confidence interval with probability 90%.



(a) (b) (c)

Figure 6. Selecting eigenvectors according to the eigenvalue. In the first row, we sort the eigenvectors in a decreasing order of eigenvalue. In the second row, red dots show the recognition rate using first s PCs determined by $d = 0.9, 0.91, 0.92, \cdots, 0.99$, where $d = \frac{\sum_{i=1}^{s} \lambda_{X_i}}{\sum_{i=1}^{p} \lambda_{X_i}}$. (a) ETH-80. (b) ARface. (c) SPDM.



(a) (b) (c)

Figure 7. Selecting eigenvectors according to the discriminant power $J(\mathbf{u}_j)$. In the first row, we sort the eigenvectors in a decreasing order of $J(\mathbf{u}_j)$. The second row shows the recognition rates using first s PCs determined by $d = 0.5, 0.6, 0.7, 0.8, 0.9, 0.91, 0.92, \cdots, 0.99$, where $d = \frac{\sum_{j=1}^{s} J(\mathbf{u}_j)}{\sum_{j=1}^{p} J(\mathbf{u}_j)}$. (a) ETH-80. (b) ARface. (c) SPDM.