REGULARIZATION STUDIES ON LDA FOR FACE RECOGNITION

Juwei Lu, K.N. Plataniotis, A.N. Venetsanopoulos

Bell Canada Multimedia Laboratory, The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, M5S 3G4, Canada

ABSTRACT

It is well-known that the applicability of Linear Discriminant Analysis (LDA) to high-dimensional pattern classification tasks such as face recognition (FR) often suffers from the so-called "*small sample size*" (SSS) problem arising from the small number of available training samples compared to the dimensionality of the sample space. In this paper, we propose a new LDA method that effectively addresses the SSS problem using a regularization technique. In addition, a scheme of expanding the representational capacity of face database is introduced to overcome the limitation that the LDA based algorithms require at least two samples per class available for learning. Extensive experimentation performed on the FERET database indicates that the proposed methodology outperforms traditional methods such as Eigenfaces and Direct LDA in a number of SSS setting scenarios.

1. INTRODUCTION

Face recognition (FR) systems, utilizing *linear discriminant analysis* (LDA) techniques have been shown to be very successful [4, 3, 6]. However, LDA-based methods often suffer from the so-called "*small sample size*" (SSS) problem, which exists in high-dimensional pattern recognition tasks where the number of training samples available for each subject is smaller than the dimensionality of the samples. For example, only $L \in [1,5]$ training samples per subject are available while the dimensionality is up to $J = 17154$ in the FR experiments reported here. As a result, the sample-based estimation for the between- and within-class scatter matrices is often extremely ill-posed. Traditional solutions to the SSS problem require the incorporation of a *principal component analysis* (PCA) step into the LDA framework. In this approach, PCA is used as a pre-processing step for dimensionality reduction and removal of the null spaces of the two scatter matrices. Then LDA is performed in the lower dimensional PCA subspace, as it was done for example in Fisherfaces [4]. However, it has been shown that the discarded null spaces may contain significant discriminatory information. To prevent this from happening, solutions without a separate PCA step, called *direct* LDA (D-LDA) methods have been presented recently in [3, 6].

The basic premise behind the D-LDA approaches is that the information residing in (or close to) the null space of the within-class scatter matrix is more significant for discriminant tasks than the information out of (or far from) the null space. Generally, the null space of a matrix is determined by its zero eigenvalues. However, due to insufficient training samples, it is very difficult to identify the true null eigenvalues. As a result, high variance is often introduced in the estimation for the zero or very small eigenvalues of the within-class scatter matrix, while the eigenvectors corresponding to these eigenvalues are considered the most significant feature bases in the D-LDA approaches.

To overcome the above problem, we propose a new LDA method for FR tasks in this paper. The LDA method developed here is based on a novel regularized Fisher's discriminant criterion, which is particularly robust against the SSS problem compared to the traditional one used in LDA. The purpose of regularization is to reduce the high variance related to the eigenvalue estimates of the within-class scatter matrix at the expense of potentially increased bias. It will be shown that, by adjusting the regularization parameter, we can obtain many LDA variants such as the D-LDA of [3] (hereafter YD-LDA) and the D-LDA of [6] (hereafter JD-LDA). The strength of regularization is dependent on the SSS situations. Extensive experiments indicate that there exists an optimal regularization solution for the proposed method, which outperforms some existing FR approaches including Eigenfaces [1], YD-LDA and JD-LDA. In addition, a scheme of expanding the representational capacity of face database is introduced to overcome the limitation that traditional LDA based algorithms cannot be applied to the extreme case where only one sample per class is available for learning. Furthermore, experimentation shows that the scheme also enhance the FR performance of the proposed LDA method.

This work is partially supported by a Bell University Labs grant. The authors would like to thank the FERET Technical Agent, the U.S. National Institute of Standards and Technology for providing the FERET database.

2. METHODS

2.1. A regularized Fisher's criterion

Given a training set, $\mathcal{Z} = \{\mathcal{Z}_i\}_{i=1}^{C}$, containing C classes with each class $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{C_i}$ consisting of a number of localized face images $\mathbf{z}_{ij}$, a total of $N = \sum_{i=1}^{C} C_i$ face images are available in the set. For computational convenience, each image is represented as a column vector of length $J(= I_w \times I_h)$ by lexicographic ordering of the pixel elements, *i.e.* $\mathbf{z}_{ij} \in \mathbb{R}^J$, where $(I_w \times I_h)$ is the image size, and $\mathbb{R}^J$ denotes the J-dimensional real space.

Let $\mathbf{S}_b$ and $\mathbf{S}_w$ be the between- and within-class scatter matrices of the training set, respectively. The regularized Fisher's criterion, which is utilized in this work instead of the conventional one ($\Psi = \arg\max\limits_{\Psi} \frac{|\Psi^T \mathbf{S}_b \Psi|}{|\Psi^T \mathbf{S}_w \Psi|}$), can be expressed as follows:

$$\Psi = \arg\max_{\Psi} \frac{\left|\Psi^T \mathbf{S}_b \Psi\right|}{\left|\eta(\Psi^T \mathbf{S}_b \Psi) + (\Psi^T \mathbf{S}_w \Psi)\right|} \quad (1)$$

where $0 \leq \eta \leq 1$ is a regularization parameter. Although Eq.1 looks different from the conventional Fisher's criterion, it can be shown that they are exactly equivalent by the following theorem.

0-7803-8554-3/04/$20.00

Theorem 1 *Let $\mathbb{R}^J$ denote the J-dimensional real space, and suppose that $\forall \psi \in \mathbb{R}^J$, $u(\psi) \geq 0$, $v(\psi) \geq 0$, $u(\psi) + v(\psi) > 0$ and $0 \leq \eta \leq 1$. Let $q_1(\psi) = \frac{u(\psi)}{v(\psi)}$ and $q_2(\psi) = \frac{u(\psi)}{\eta \cdot u(\psi) + v(\psi)}$. Then, $q_1(\psi)$ has the maximum (including positive infinity) at point $\psi^* \in \mathbb{R}^J$ iff $q_2(\psi)$ has the maximum at point ψ^*.*

Proof: Since $u(\psi) \geq 0$, $v(\psi) \geq 0$ and $0 \leq \eta \leq 1$, we have $0 \leq q_1(\psi) \leq +\infty$ and $0 \leq q_2(\psi) \leq \frac{1}{\eta}$.

1. If $\eta = 0$, then $q_1(\psi) = q_2(\psi)$.
2. If $0 < \eta \leq 1$ and $v(\psi) = 0$, then $q_1(\psi) = +\infty$ and $q_2(\psi) = 1/\eta$.
3. If $0 < \eta \leq 1$ and $v(\psi) > 0$, then $q_2(\psi) = \frac{u(\psi)/v(\psi)}{1+\eta u(\psi)/v(\psi)} = \frac{q_1(\psi)}{1+\eta q_1(\psi)} = \frac{1}{\eta}\left(1 - \frac{1}{1+\eta q_1(\psi)}\right)$. It can be seen that in this case, $q_2(\psi)$ increases iff $q_1(\psi)$ increases.

Combining the above three cases, the theorem is proven.

The modified Fisher's criterion is a function of the parameter η, which controls the strength of regularization. Within the variation range of η, two extremes should be noted. In one extreme where $\eta = 0$, the modified Fisher's criterion is reduced to the conventional one with no regularization. In contrast with this, the strongest regularization is introduced in another extreme where $\eta = 1$. In this case, Eq.1 becomes $\Psi = \arg\max_{\Psi} \frac{|\Psi^T \mathbf{S}_b \Psi|}{|\Psi^T(\mathbf{S}_b + \mathbf{S}_w)\Psi)|}$, which as a variant of the original Fisher's criterion has been also widely used for example in [5, 6]. The advantages of introducing the regularization will be seen during the development of the new LDA algorithm proposed below.

2.2. A regularized LDA: R-LDA

Let us assume that $\mathcal{A}$ and $\mathcal{B}$ represent the null spaces of $\mathbf{S}_b$ and $\mathbf{S}_w$ respectively, while $\mathcal{A}' = \mathbb{R}^J - \mathcal{A}$ and $\mathcal{B}' = \mathbb{R}^J - \mathcal{B}$ denote the orthogonal complements of $\mathcal{A}$ and $\mathcal{B}$. The maximization of Eq.1 can be achieved by solving the eigenvalue problem of $(\eta\mathbf{S}_b + \mathbf{S}_w)^{-1}\mathbf{S}_b$ if sufficient training samples are available. Due to the SSS problem, often a degenerated $(\eta\mathbf{S}_b + \mathbf{S}_w)$ is obtained in FR tasks. Traditional methods, for example Fisherfaces [4], attempt to solve the problem by utilizing an intermediate PCA step to remove $\mathcal{A}$ and $\mathcal{B}$. However, it should be noted at this point that the maximum of the ratio in Eq.1 can be reached only when $\Psi^T\mathbf{S}_w\Psi = 0$ and $\Psi^T\mathbf{S}_b\Psi \neq 0$. This means that the discarded null space $\mathcal{B}$ may contain the most significant discriminatory information. On the other hand, there is no significant information, in terms of the maximization in Eq.1, to be lost if $\mathcal{A}$ is discarded. It is not difficult to see at this point that when $\Psi \in \mathcal{A}$, the ratio $\frac{|\Psi^T\mathbf{S}_b\Psi|}{|\Psi^T\mathbf{S}_w\Psi|}$ drops down to its minimal value, 0. Therefore, the intersection space $(\mathcal{A}' \cap \mathcal{B})$ is considered the optimal discriminant feature bases in the D-LDA approaches [3, 6].

In this work, we propose a *regularized* LDA (hereafter R-LDA) method, which following the D-LDA process of [3, 6], attempts to optimize the regularized Fisher's criterion of Eq.1. To this end, we first solve the complement space of $\mathbf{S}_b$, $\mathcal{A}'$. Let $\mathbf{U}_m = [u_1, \cdots, u_m]$ be the eigenvectors of $\mathbf{S}_b$ corresponding to its first $m(\leq C-1)$ largest nonzero eigenvalues Λ_b. The complement space $\mathcal{A}'$ is spanned by $\mathbf{U}_m$, which is furthermore scaled by $\mathbf{H} = \mathbf{U}_m\Lambda_b^{-1/2}$ so as to have $\mathbf{H}^T\mathbf{S}_b\mathbf{H} = \mathbf{I}$, where $\mathbf{I}$ is the $(m \times m)$ identity matrix. In the subspace spanned by $\mathbf{H}$, we then seek a set of feature bases, which minimize the denominator of Eq.1, $\eta\mathbf{I} + \mathbf{H}^T\mathbf{S}_w\mathbf{H}$. It is not difficult to see that the sought feature bases correspond to the $M(\leq m)$ eigenvectors of $\mathbf{H}^T\mathbf{S}_w\mathbf{H}$, $\mathbf{P}_M = [\mathbf{p}_1, \cdots, \mathbf{p}_M]$, with the smallest eigenvalues Λ_w. Combining these results, we can obtain the final solution, $\Psi = \mathbf{H}\mathbf{P}_M(\eta\mathbf{I} + \Lambda_w)^{-1/2}$, which is a set of optimal discriminant feature basis vectors. The detailed process to implement the R-LDA method is depicted in Fig.1.

It can be seen from Fig.1 that R-LDA reduces to YD-LDA and JD-LDA when $\eta = 0$ and $\eta = 1$, respectively. Varying the values of η within $[0, 1]$ leads to a set of intermediate D-LDA variants between YD-LDA and JD-LDA. Since the subspace spanned by Ψ may contain the intersection space $(\mathcal{A}' \cap \mathcal{B})$, it is possible that there exist zero or very small eigenvalues in Λ_w, which has been shown to be high variance for estimation in the SSS situations. As a result, any bias arising from the eigenvectors corresponding to these eigenvalues is dramatically exaggerated due to the normalization $(\mathbf{P}_M\Lambda_w^{-1/2})$. Against the effect, the introduction of the regularization helps to decrease the importance of these highly unstable eigenvectors, thereby reducing for some extent the variance.

Input: A training set $\mathcal{Z}$ with C classes: $\mathcal{Z} = \{\mathcal{Z}_i\}_{i=1}^C$, each class containing $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{C_i}$ face images, where $\mathbf{z}_{ij} \in \mathbb{R}^J$, and the regularization parameter η.

Output: An M-dimensional LDA subspace spanned by Ψ, an $J \times M$ matrix with $M \ll J$.

Algorithm:

Step 1. Express $\mathbf{S}_b = \Phi_b\Phi_b^T$, with $\Phi_b = [\Phi_{b,1}, \cdots, \Phi_{b,c}]$, $\Phi_{b,i} = (C_i/N)^{1/2}(\bar{\mathbf{z}}_i - \bar{\mathbf{z}})$, $\bar{\mathbf{z}}_i = 1/C_i\sum_{j=1}^{C_i}\mathbf{z}_{ij}$, and $\bar{\mathbf{z}} = 1/N\sum_{i=1}^{C}\sum_{j=1}^{C_i}\mathbf{z}_{ij}$.

Step 2. Find the m eigenvectors of $\Phi_b^T\Phi_b$ with non-zero eigenvalues, and denote them as $\mathbf{E}_m = [\mathbf{e}_1, \cdots, \mathbf{e}_m]$.

Step 3. Calculate the first m most significant eigenvectors $(\mathbf{U}_m)$ of $\mathbf{S}_b$ and their corresponding eigenvalues (Λ_b) by $\mathbf{U}_m = \Phi_b\mathbf{E}_m$ and $\Lambda_b = \mathbf{U}_m^T\mathbf{S}_b\mathbf{U}_m$.

Step 4. Let $\mathbf{H} = \mathbf{U}_m\Lambda_b^{-1/2}$. Find eigenvectors of $\mathbf{H}^T\mathbf{S}_w\mathbf{H}$, $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_m]$ sorted in increasing eigenvalue order.

Step 5. Choose the first $M(\leq m)$ eigenvectors in $\mathbf{P}$. Let $\mathbf{P}_M$ and Λ_w be the chosen eigenvectors and their corresponding eigenvalues, respectively.

Step 6. Return $\Psi = \mathbf{H}\mathbf{P}_M(\eta\mathbf{I} + \Lambda_w)^{-1/2}$.

Fig. 1. The pseudo code implementation of the R-LDA method

2.3. Expansion of Representational Capacity of Face Database

The works described above are focused on the attempt to address the SSS problem from the viewpoint of improving the LDA algorithm. On the other hand, the problem can be approached by expanding the representational capacity of the available training database. For example, given a pair of prototype images belonging to a same class, Stan *et al.* [7] proposed a linear model, called the *nearest feature line* (NFL), to virtually generalize an infinite number of variants of the two prototypes under variations in illumination and expression. However, like LDA, the NFL method requires at least two training samples per subject to be available. To deal with the extreme case where only one training image per subject is available, J. Huang *et al.* [8] recently proposed a method, which constructs more samples by rotating and translating the prototype image. However, the method of [8] also introduces bias inevitably when face recognition is performed on a set of well-aligned face images for example along with the centers of the eyes as did in the experiments reported here.

To avoid the bias, an alternative approach to double the size of the training set is to introduce the mirrored versions of the training samples. Based on the symmetrical property of face object, intuitively it is reasonable to consider the mirrored view of a face image to be a real and bias-free sample of the face pattern. In addition to the training samples, the mirrored version of any test sample can be also utilized to enhance the FR performance. For example, we can verify the classification result of an given query with the result of its mirror. A recognition process is accepted only when the query and its mirror are given a same class label, otherwise the query is rejected to recognition. More sophisticated rules to combine the two results can be found in [9].

3. EXPERIMENTAL RESULTS

3.1. The FR Evaluation Design

A set of experiments are included in the paper to assess the performance of the proposed R-LDA method. To show the high complexity of the face patterns' distribution, a medium-size subset of the FERET database [2] is used in the experiments. The subset consists of 1147 gray-scale images of 120 people, each one having at least 6 samples. These images as depicted in Table 1 cover a wide range of variations in illumination, facial expression/details, pose angels and others. We follow the preprocessing sequence recommended in [2], which includes four steps: (1) eyes alignment, (2) removal of nonface portions, (3) histogram equalization, and (4) data normalization. Fig.2 depicts some sample images after the preprocessing sequence is applied. For computational convenience, each image is finally represented as a column vector of length $J = 17154$ prior to the recognition stage.

Table 1. No. of images divided into the standard FERET imagery categories, and the pose angle, α (degree), of each category.

Ct.	fa	fb	ba	bj	bk	ql	qr	rb	rc
No.	567	338	5	5	5	68	65	32	62
α	0	0	0	0	0	-22.5	+22.5	10	-10

Fig. 2. Some samples of eight people come from the normalized FERET evaluation database.

The number of available training samples per subject, L, has a significant influence on the strength of regularization. To study the sensitivity of the performance, in terms of *correct recognition rate* (CRR), to L, five tests were performed with various L values ranging from $L = 1$ to $L = 5$. For a particular L, the FERET subset is randomly partitioned into two datasets: a training set and a test set. The training set is composed of $(L \times 120)$ samples: L images per person were randomly chosen. The remaining $(1147 - L \times 120)$ images are used to form the test set. There is no overlapping between the two. To enhance the accuracy of the assessment, five runs of such a partition were executed, and all of the CRRs reported later have been averaged over the five runs.

3.2. The FR Performance Comparison

Besides YD-LDA, JD-LDA and R-LDA, the so-called Eigenfaces method [1], was implemented to provide a performance baseline. The CRRs obtained by R-LDA as a function of (M, η) are depicted in Fig.3, where η started from 10^{-4} instead of zero in case $\mathbf{H}^T\mathbf{S}_w\mathbf{H}$ is singular. Also, a quantitative comparison of the best CRRs obtained by the four methods with corresponding parameter values, is summarized in Table 2.

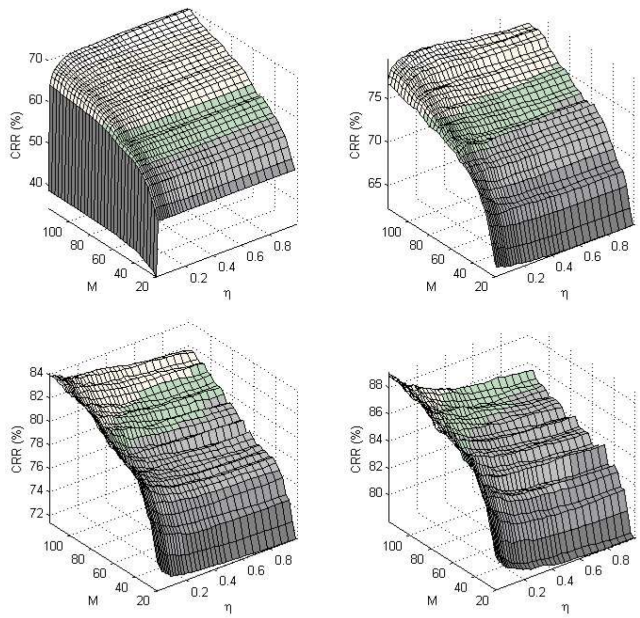

Fig. 3. CRRs obtained by R-LDA as a function of (M, η). **Top**: $L = 2, 3$; **Bottom**: $L = 4, 5$.

The parameter η controls the strength of regularization, which balances the tradeoff between variance and bias in the estimation for the small eigenvalues of the within-class scatter matrix. Varying the values of η within $[0, 1]$ leads to a set of intermediate LDA variants between YD-LDA and JD-LDA. In theory, YD-LDA with no bias introduced should be the best performer among these variants if sufficient training samples are available. It can be observed at this point from Fig.3 and Table 2 that the CRR peaks gradually moved from the right side toward the left side ($\eta = 0$) that is the case of YD-LDA as L increases. Small values of λ have been good enough for the regularization requirement in many cases ($L \geq 4$) as shown in Fig.3. However, it also can be seen from Fig.3 and Table 2 that YD-LDA performed poorly when $L = 2, 3$. This should be attributed to the high variance in the estimate of $\mathbf{S}_w$ due to insufficient training samples. In these cases, even $\mathbf{H}^T\mathbf{S}_w\mathbf{H}$ is singular or close to singular, and the resulting effect is to dramatically exaggerate the importance associated with the eigenvectors corresponding to the smallest eigenvalues. Against the effect, the introduction of regularization helps to decrease the larger eigenvalues and increase the smaller ones, thereby counteracting for some extent the bias. This is also why JD-LDA outperforms YD-LDA when L is small. Although R-LDA is the top performer amongst all the methods compared in Table 3, the determination of its optimal parameter values is computationally demanding as it is based on exhaustive searches. A fast and cost effective R-LDA parameter optimization method will be the focus of future research.

Table 2. Comparison of best found CRRs (%) and their corresponding parameter values without using mirrored samples.

$L=$	1	2	3	4	5
PCA	47.50	59.58	66.71	67.92	68.85
(M^*)	119	159	217	289	327
YD-LDA	–	38.85	76.85	84.02	89.07
(M^*)	–	116	117	114	114
JD-LDA	–	70.32	78.14	81.95	85.78
(M^*)	–	116	114	115	112
R-LDA	–	70.36	79.70	84.14	89.07
(M^*)	–	116	117	116	114
(η^*)	–	0.867	0.22	0.016	1e-4

Table 3. Comparison of best CRRs (%) and their corresponding parameter values obtained by R-LDA using different mirror schemes. In R-LDA$^{\mathrm{m3}}$, Rej. denotes the reject rate (%).

$L=$	1	2	3	4	5
R-LDA$^{\mathrm{m1}}$	56.69	70.80	82.49	86.93	91.12
(M^*)	118	118	117	114	108
(η^*)	0.092	0.30	0.016	1e-4	1e-4
R-LDA$^{\mathrm{m2}}$	56.79	71.16	82.36	86.93	90.90
(M^*)	119	118	119	112	112
(η^*)	0.7	0.3	0.016	1e-4	1e-4
R-LDA$^{\mathrm{m3}}$	60.19	74.16	84.09	88.18	91.91
Rej.	7.95	6.13	2.90	2.07	1.50
(M^*)	119	119	119	112	111
(η^*)	1	0.4	0.016	1e-4	1e-4

The LDA based algorithms require at least two training samples for each class. However, with the introduction of the mirrored training samples, it becomes possible to break the limitation. In the results depicted in Table 3, where R-LDA$^{\mathrm{m1}}$, R-LDA$^{\mathrm{m2}}$ and R-LDA$^{\mathrm{m3}}$ denote the versions of R-LDA that is trained with a combined set consisting of the training samples and and their mirrors. The difference between the three versions is that R-LDA$^{\mathrm{m1}}$ uses only the test set, R-LDA$^{\mathrm{m2}}$ uses only the mirrored test set, while R-LDA$^{\mathrm{m3}}$ combines the results of the test set and its mirror in a way introduced in Section 2.3. Not surprisingly, it can be seen from Table 3 that R-LDA$^{\mathrm{m1}}$ and R-LDA$^{\mathrm{m2}}$ have a very close performance. This means that to recognize a test sample, we can use either the sample or its mirror. Compared to R-LDA in Table 2, the performance improvement of R-LDA$^{\mathrm{m1}}$ and R-LDA$^{\mathrm{m2}}$ is up to 2.02% in average over $L = 2 \sim 5$. Also, the reject rates obtained by R-LDA$^{\mathrm{m3}}$ indicate that the recognition is incorrect in most cases when the sample and its mirror are given different labels. Therefore, compared to R-LDA$^{\mathrm{m1}}$ and R-LDA$^{\mathrm{m2}}$, another CRR improvement up to 2.1% in average over $L = 1 \sim 5$ is obtained by R-LDA$^{\mathrm{m3}}$. These results obtained by the three R-LDA versions demonstrate that the mirrors of face images provide not only more samples, but also complemental information, which is useful to enhance the FR performance.

4. CONCLUSION

A new LDA method for face recognition has been introduced in this paper. The proposed method is based on a novel regularized Fisher's discriminant criterion, which is particularly robust against the SSS problem compared to the traditional one used in LDA. It has been shown that a series of traditional LDA variants including the recently introduced YD-LDA and JD-LDA can be derived from the proposed R-LDA framework by adjusting the regularization parameter. Also, a scheme to double the size of face databases is introduced, so that R-LDA can be carried out in the extreme case where only one training sample available for each subject. The effectiveness of the proposed method has been demonstrated through experimentation using the FERET database.

R-LDA can be seen as a general pattern recognition method capable of addressing the SSS problem. We expect that in addition to FR, R-LDA will provide excellent performance in applications, such as image/video indexing, retrieval, and classification.

5. REFERENCES

[1] Matthew A. Turk and Alex P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[3] Yu Hua and Yang Jie, "A direct LDA algorithm for high-dimensional data - with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, October 2001.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[5] K. Liu, Y.Q. Cheng, J.Y. Yang, and X. Liu, "An efficient algorithm for foley-sammon optimal set of discriminant vectors by algebraic method," *Int. J. Pattern Recog. Artif. Intell.*, vol. 6, pp. 817–829, 1992.

[6] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using LDA based algorithms," *IEEE Trans. on Neural Networks*, vol. 14, no. 1, pp. 195–200, January 2003.

[7] Stan Z. Li and Juwei Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, vol. 10, pp. 439–443, March 1999.

[8] J. Huang, P.C. Yuen, W.S. Chen, and J.H. Lai, "Component-based LDA method for face recognition with one training sample," in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Nice, France, October 2003.

[9] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, March 1998.