BOOSTING LINEAR DISCRIMINANT ANALYSIS 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

In this paper, we propose a new algorithm to boost performance of traditional Linear Discriminant Analysis (LDA)-based face recognition (FR) methods in complex FR tasks, where highly nonlinear face pattern distributions are often encountered. The algorithm embodies the principle of "divide and conquer", by which a complex problem is decomposed into a set of simpler ones, each of which can be conquered by a relatively easy solution. The AdaBoost technique is utilized within this framework to: 1) generalize a set of simple FR sub-problems and their corresponding LDA solutions; 2) combine results from the multiple, relatively weak, LDA solutions to form a very strong solution. Experimentation performed on the FERET database indicates that the proposed methodology is able to greatly enhance performance of the traditional LDA-based method with an averaged improvement of correct recognition rate (CRR) up to 9% reported.

1. INTRODUCTION

Face recognition (FR) systems, utilizing linear discriminant analysis (LDA) techniques have been shown to be very successful [1, 2]. However, the so-called "plug-in" covariance matrix estimates widely used in the LDA-based approaches often suffer from the so-called "*small sample size*" (SSS) problem often seen in high-dimensional pattern recognition tasks where the number of available training samples per subject (L) is smaller than the dimensionality of the samples (J). Recently, an effective SSS solution called Direct LDA (D-LDA), have been presented [1, 2]. Although may not be optimal in terms of CRR in some cases, the D-LDA of [2] (hereafter JD-LDA), enhanced by a simple regularization strategy, has been shown to be the more robust than the one of [1] against the SSS problem, performing well even when $L << J$, which is the case in many FR tasks.

Although successful in many cases, linear methods including the LDA-based ones often fail to deliver good performance when face patterns are subject to large variations in viewpoints, illumination or facial expression, which result in a highly nonlinear and complex distribution. The limited success of these methods should be attributed to their linear nature. There are two ways to handle the complex pattern distribution: 1) with nonlinear models, or 2) with a mixture of locally linear models (AMLLM). The main problem with most nonlinear methods such as those based on kernel machines is that the involved nonlinear parameters which significantly influence the performance of the FR systems, are very difficult to be optimized. In addition, these methods are computationally expensive compared to their linear counterparts, and tend to overfit quite often. On the other hand, AMLLM-based approaches embody the principle of "divide and conquer", by which a complex FR problem is decomposed into a set of simpler ones, in each of which a locally linear face distribution can be generalized and dealt with by a relatively easy linear solution. As such, the AMLLM-based methods are simpler, more cost effective and easier to implement compared to the nonlinear ones.

In this paper, we propose a new AMLLM-like method to boost the performance of the traditional LDA-based approaches in complex FR tasks. The main novelty existing in the method is the introduction of the machine-learning technique known as "boosting", which is able to boost an ensemble of weak learners slightly better than random guessing to a very accurate learner [3]. Boosting seems ideal to deal with two issues central to the AMLLM-like approaches: 1) the generalization of a set of simple linear solutions, each one aimed to a particular sub-problem; 2) the formation of a globally strong solution through the combination of the multiple local solutions. However, it is widely believed that boosting-like algorithms are not suited to a stable base learner such as LDA, because their effectiveness depends to a great extent on the base learner's "instability". To challenge the popular belief, we first propose a variable called "pairwise class discriminant distribution" (PCDD), which is used to build a strong connection between boosting and the LDA-based learner. Through PCDD, boosting can effectively manipulate the learner, so that it is focused on the hard to separate pairs of classes. Then, a cross-validation mechanism (CVM) is introduced to control the weakness (also function to a certain extent as instability) of the learner. With the integration of PCDD and CVM, both the ability that boosting controls the learner and the diversity of local LDA solutions produced are greatly enhanced. This, as will be seen in the experiments reported here, results in a significant boost to the FR performance.

2. METHODS

2.1. A JD-LDA Learner

Considering the robustness in the SSS conditions, the LDA approach chosen as the base learner is JD-LDA [2], which is briefly described here for completeness.

Given a training set containing C classes, $\mathcal{Z} = \{\mathcal{Z}_i\}_{i=1}^{C}$, with each class consisting of a number of face images: $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{C_i}$, a total of $N = \sum_{i=1}^{C} C_i$ face images are available in the set. Each image is represented as a column vector of length $J(= I_w \times I_h)$, *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. JD-LDA finds a set of optimal discriminant basis vectors, denoted as $\{\psi_m\}_{m=1}^{M}$ where $\psi_m \in \mathbb{R}^J$ and

The authors would like to thank the FERET Technical Agent, the U.S. National Institute of Standards and Technology (NIST) for providing the FERET database.

$M \ll J$, by optimizing a separability criterion, or equivalently solving the following eigenvalue problem,

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

where $\Psi = [\psi_1, \ldots, \psi_M]$, $\mathbf{S}_b$ and $\mathbf{S}_w$ are the between- and within-class scatter matrices of the training set respectively. For any input face images $\mathbf{z}$, its JD-LDA based representation $\mathbf{y}$ can be obtained by projecting $\mathbf{z}$ into a M-dimensional feature space spanned by Ψ, where the separability of different subjects is enhanced, thus $\mathbf{y} = \Psi^T \mathbf{z}$, where $\mathbf{y} \in \mathbb{R}^M$. Since JD-LDA is only a feature extractor, the subsequent classification in the JD-LDA learner is implemented using a classic nearest center classifier in the feature space based on a normalized Euclidean distance, which is given by

$$h(\mathbf{z}, i) = (d_{max} - d_{\mathbf{z},i})/(d_{max} - d_{min}) \quad (2)$$

where $d_{\mathbf{z},i} = \left\|\Psi^T(\mathbf{z} - \bar{\mathbf{z}}_i)\right\|$, $d_{max} = \max(\{d_{\mathbf{z},i}\}_{i=1}^{C})$, $d_{min} = \min(\{d_{\mathbf{z},i}\}_{i=1}^{C})$, and $\bar{\mathbf{z}}_i = \frac{1}{C_i}\sum_{j=1}^{C_i} \mathbf{z}_{ij}$ is the ith class center. The distance-based hypothesis $h(\mathbf{z}, i)$ has values in $[0, 1]$, and thus can function as required by the AdaBoost.M2 algorithm to indicate a "degree of plausibility" for labelling $\mathbf{z}$ as the class i.

2.2. Boosting the JD-LDA Learner (B-JD-LDA)

Since the boosting scheme proposed here (see Fig.1) is developed from AdaBoost.M2 [3], a sophisticated extension of the classic AdaBoost to the multi-class case, we first briefly review the AdaBoost.M2 algorithm, which attempts to overcome some limitations existing in those straightforward multi-class extensions such as AdaBoost.M1 [3] by introducing a sophisticated error measure called "pseudo-loss" (see step 5 in Fig.1) instead of the usual prediction error. The pseudo-loss is computed with respect to a distribution called "mislabel distribution", $\hat{D}$, defined over the set of all mislabels: B (see input in Fig.1). By manipulating the distribution, the boosting algorithm can focus the base learner not only on hard-to-classify samples, but more specifically, on the incorrect labels that are hardest to discriminate [3].

With these concepts and theories, we can start to design the algorithm to boost JD-LDA. First, we have to build the connection between the base learner and the boosting algorithm by introducing a new distribution called "pairwise class discriminant distribution" (PCDD), A_{pq}, which is defined on any pair of classes $\{(p, q) : p, q \in \{1, \cdots, C\}\}$, and computed at the t-th iteration by

$$A_t(p,q) = \begin{cases} \sum_{j=1}^{C_p} \hat{D}_t(\mathbf{z}_{pj}, q) + \sum_{j=1}^{C_q} \hat{D}_t(\mathbf{z}_{qj}, p), & \text{if } p \neq q \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Since $\hat{D}_t(\mathbf{z}_{ij}, y)$ indicates the difficult extent of distinguishing the incorrect label y on the sample $\mathbf{z}_{ij}$ based on the feedback from the hypothesis produced previously, intuitively $A_t(p, q)$ can be considered as a measure of how important it is to discriminate between classes p and q when design the current hypothesis h_t. A larger $A_t(p, q)$ value implies a worse separability between the two classes. It is therefore reasonable to manipulate the JD-LDA learner through $A_t(p, q)$, so that it is focused on the hard to separate pairs of classes. To this end, we define a variant of the between-class scatter matrix, which can be given as follows,

$$\hat{\mathbf{S}}_{b,t} = \sum_{p=1}^{C} \phi_p \phi_p^T \quad (4)$$

where $\phi_p = (C_p/N)^{1/2} \sum_{q=1}^{C} A_t^{1/2}(p, q)(\bar{\mathbf{z}}_p - \bar{\mathbf{z}}_q)$. It should be noted at this point that the variant $\hat{\mathbf{S}}_{b,t}$ weighted by A_t embodies the principle behind the so-called "fractional-step" LDA depicted in [2, 4], that is, object classes that are difficult to be separated in the low-dimensional output spaces $(\Psi_1, \cdots, \Psi_{t-1})$ generalized in previous rounds, and thus can potentially result in misclassification, should be paid more attentions through more heavily weighting in the high-dimensional input space of the current (t-th) round, so that their separability is enhanced in the resulting feature space Ψ_t. Also, it is not difficult to see that the variant $\hat{\mathbf{S}}_{b,t}$ is equivalent to $\mathbf{S}_b$ when $A_t(p, q)$ is equal to a constant.

Similarly, the weighted within-class scatter matrix can be given as follows,

$$\hat{\mathbf{S}}_{w,t} = N \cdot \sum_{i=1}^{C} \sum_{j=1}^{C_i} D_t(\mathbf{z}_{ij})(\mathbf{z}_{ij} - \bar{\mathbf{z}}_i)(\mathbf{z}_{ij} - \bar{\mathbf{z}}_i)^T \quad (5)$$

where $D_t(\mathbf{z}_{ij}) = \sum_{y \neq y_{ij}} \hat{D}_t(\mathbf{z}_{ij}, y)$ is defined over $\mathcal{Z}$ as the sample distribution with similar meanings to the one defined in AdaBoost. As such, a larger value of $D_t(\mathbf{z}_{ij})$ implies a harder sample to those hypotheses generalized previously.

In addition to large margins, it has been found that the generalization error in boosting-like methods depends on the low or weak dependence among the hypotheses produced [5]. Obviously, hypotheses obtained through training with more overlapping samples will result in a stronger dependence among them. A way to avoid building similar hypotheses repeatedly is to artificially introduce some randomness in the construction of the training data. To this end, a modified PCDD is proposed below

$$\hat{A}_t(p,q) = \begin{cases} \sum_{j:g_t(\mathbf{z}_{pj})=q} D_t(\mathbf{z}_{pj}) + \sum_{j:g_t(\mathbf{z}_{qj})=p} D_t(\mathbf{z}_{qj}), & \text{if } p \neq q \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $g_t(\mathbf{z}) = \arg\max_{y \in \mathbf{Y}} h_t(\mathbf{z}, y)$. As an effect of $\hat{A}_t(p, q)$ instead of Eq.3, it can been seen that only those classes having the mislabelled samples by previous hypothesis h_{t-1} are contributed for the construction of the current hypothesis h_t (through $\hat{\mathbf{S}}_{b,t}$) in each iteration. Thus, by manipulating $\hat{A}_t(p, q)$, we can reduce the overlapping extent of training samples utilized for different hypotheses, and reach the goal of weakening the dependence among these hypotheses. Also, this has the effect of forcing every hypothesis produced to more specifically focus on the previously mislabelled samples, and helps to generalize a more diverse committee of hypotheses. On the other hand, the classification ability of the individual hypothesis h_t is to some extent weakened due to less training samples being used. This weakening may result in decrease in the margins of the training samples. However, it should be noted at this point that there appears to be a trade-off between weak dependence and large expected margins to achieve a low generalization error [5]. In many cases, the modification of Eq.6 may yield a better balance than $A_t(p, q)$, and thereby lead to a better classification performance.

With the introduction of $A_t(p, q)$, $\hat{A}_t(p, q)$, $\hat{\mathbf{S}}_{b,t}$ and $\hat{\mathbf{S}}_{w,t}$, we now can give a new boosting algorithm as depicted in Fig.1, where either $A_t(p, q)$ or $\hat{A}_t(p, q)$ can used to substituted for one another. It can been seen from Fig.1 that the JD-LDA learner at every iteration is tuned to conquer a particular sub-problem generalized by the feedback $\hat{D}_t$ in a manner of similar to automatic gain control, and they thereby can offer complementary information about

the patterns to be classified. The final solution is a mixture of T JD-LDA based FR systems by weighted linear combination.

2.3. A cross-validation mechanism to weaken the learner

It should be noted that JD-LDA itself has been a quite strong learner in terms of classification ability. As a consequence, $\epsilon_t = 0$ is often obtained so that the boosting process cannot go forward. To solve the problem, we have to artificially weaken the JD-LDA learner by introducing a sort of cross-validation mechanism, with which only a subset of the entire training set, $\mathcal{R}_t \subset \mathcal{Z}$, is utilized to train the individual JD-LDA learner. The subset $\mathcal{R}_t$ is formed in each iteration by choosing $r \leq L$ hardest examples per class based on values of $D_t(\mathbf{z}_{ij})$, thus $|\mathcal{R}_t| = C \cdot r$, where $|\mathcal{R}_t|$ denotes the size of $\mathcal{R}_t$. This strategy does not only weaken the JD-LDA learner, but also enhance the generalization ability of the overall algorithm due to the introduction of cross-validation. Moreover, since each time feeds the learner a different subset of the training examples, this essentially increases the diversity or weakens the dependence among the hypotheses produced.

3. EXPERIMENTAL RESULTS

3.1. The FR Evaluation Design

A set of experiments are included in this paper to assess the performance of the proposed boosting method (hereafter B-JD-LDA). To show the high complexity of the face patterns' distribution, a middle-size subset of the FERET database [6] is used in the experiments. The subset denoted as $\mathcal{G}$ consists of 606 gray-scale images of 49 people, each one having more than 10 samples. These images cover a wide range of variations in illumination, facial expression/details, acquisition time, races and others. We follow the preprocessing sequence recommended in [6]. Some examples obtained after preprocessing are depicted in Fig.2. For computational requirement, each image is finally represented as a column vector of length $J = 17154$.

Following standard FR practices, the database $\mathcal{G}$ is randomly partitioned into two subsets: the training set $\mathcal{Z}$ and test set $\mathcal{Q}$. The training set is composed of $|\mathcal{Z}| = L \cdot C$ images: L images per subject are randomly chosen. The remaining images are used to form the test set $\mathcal{Q} = \mathcal{G} - \mathcal{Z}$. To enhance the accuracy of the assessment, the correct recognition rates (CRRs) of all the methods evaluated here are averaged over five runs. Each run is executed on a random partition of the database $\mathcal{G}$. Also, it is empirically found that the selection between $A_t(p,q)$ and $\hat{A}_t(p,q)$ is data dependent. For the experiments reported here, B-JD-LDA with $\hat{A}_t(p,q)$ slightly outperforms the one with $A_t(p,q)$. Thus, for space limitations, only the results obtained by B-JD-LDA($\hat{A}_t(p,q)$) are reported here.

3.2. The FR Performance Comparison

Besides the proposed B-JD-LDA method and the stand-alone JD-LDA (without boosting, hereafter S-JD-LDA) method, the most well-known FR algorithm, the so-called Eigenfaces method [7], was also implemented to provide a performance baseline. For all the three methods, the CRR is a function of the number of the extracted feature vectors, M, and the number of available training samples per subject, L. Also, B-JD-LDA's performance is affected by r, the number of samples per subject that is used to train the based learner. Although a larger value of r will equivalently lead

---

Input: A set of training images $\mathcal{Z} = \{(\mathbf{z}_{ij}, y_{ij})_{j=1}^{C_i}\}_{i=1}^{C}$ with labels $y_{ij} = i \in \mathbf{Y} = \{1, \cdots, C\}$; the chosen weak learner is JD-LDA; and number of iterations T.

Let $B = \{(i,j,y) : i \in \mathbf{Y}, j \in \{1, \cdots, C_i\}, y \neq y_{ij}\}$.

Initialize $\hat{D}_1(\mathbf{z}_{ij}, y) = \frac{1}{|B|} = \frac{1}{N(C-1)}$, the mislabel distribution over B.

Do for $t = 1, \cdots, T$:

1. Update the sample distribution: $D_t(\hat{D}_t)$, and the PCDD: A_t with Eq.3 or $\hat{A}_t$ with Eq.6.
2. If $t = 1$ then randomly choose r samples per class to form a learning set $\mathcal{R}_1 \subset \mathcal{Z}$.
   else choose r hardest samples per class based on D_t to form $\mathcal{R}_t \subset \mathcal{Z}$.
3. Train a JD-LDA learner with ($\mathcal{R}_t, D_t, A_t$ or $\hat{A}_t$).
4. Apply the trained JD-LDA($\mathcal{R}_t, D_t, A_t$ or $\hat{A}_t$) into $\mathcal{Z}$, and get back corresponding hypothesis with Eq.2, $h_t : \mathcal{Z} \times \mathbf{Y} \rightarrow [0,1]$.
5. Calculate the pseudo-loss of h_t:

$$\epsilon_t = \frac{1}{2} \sum_{(i,j,y) \in B} \hat{D}_t(\mathbf{z}_{ij}, y) \left(1 - h_t(\mathbf{z}_{ij}, y_{ij}) + h_t(\mathbf{z}_{ij}, y)\right).$$

6. Set $\beta_t = \epsilon_t / (1 - \epsilon_t)$.
   If $\beta_t = 0$, then set $T = t - 1$ and abort loop.
7. Update the mislabel distribution $\hat{D}_t$:

$$\hat{D}_{t+1}(\mathbf{z}_{ij}, y) = \hat{D}_t(\mathbf{z}_{ij}, y) \cdot \beta_t^{(1 + h_t(\mathbf{z}_{ij}, y_{ij}) - h_t(\mathbf{z}_{ij}, y))/2}.$$

8. Normalize $\hat{D}_{t+1}$ so that it is a distribution,

$$\hat{D}_{t+1}(\mathbf{z}_{ij}, y) \leftarrow \frac{\hat{D}_{t+1}(\mathbf{z}_{ij}, y)}{\sum_{(i,j,y) \in B} \hat{D}_{t+1}(\mathbf{z}_{ij}, y)}.$$

Output the final hypothesis,

$$h_f(\mathbf{z}) = \arg\max_{y \in \mathbf{Y}} \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t}\right) h_t(\mathbf{z}, y).$$

Fig. 1. The Algorithm of Boosting JD-LDA (B-JD-LDA).

Fig. 2. Some examples of six people after preprocessing.

to a stronger base learner, it is found in our experiments that B-JD-LDA would fail to perform well when too weak (*e.g.* $r = 2$) or too strong (*e.g.* $r = 4$) base learners are utilized as shown in Fig.3:A,B. The observations are consistent with the boosting theories discussed in [3]. Since space limitations prevent us from presenting all the results within the variation range of r, L and M, those depicted in Fig.3 and Table 1 are obtained only from several representative cases with $L = 5$, $r = 2, 3, 4$, and $M = 20$ being used.

T iterations of B-JD-LDA produced T JD-LDA hypotheses h_t, and each one was assigned a weight ($\log \frac{1}{\beta_t}$), through which

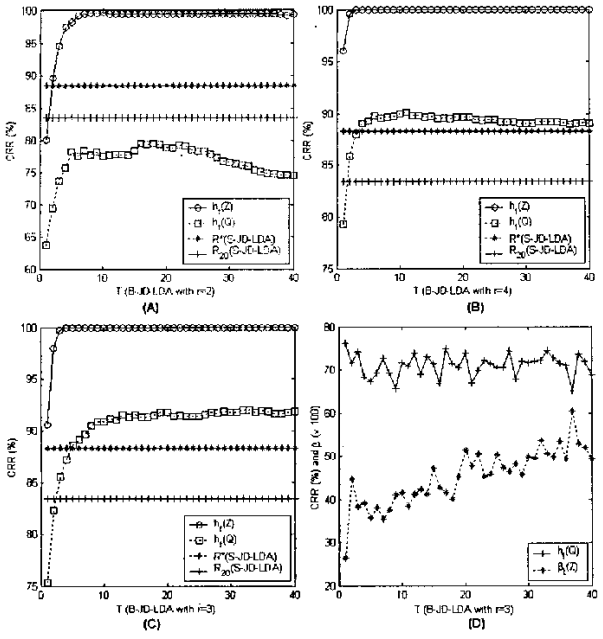

Fig. 3. CRR comparisons as a function of T. **A, B, C:** Training and test CRRs of B-JD-LDA with $r = 2, 3, 4$; **D:** Individual test CRRs of T generalized hypotheses, and their corresponding $\beta(\times 100)$.

they were linearly combined to form a mixture h_f. Although each hypothesis has its own focus, a hypothesis producing higher CRR should be given a larger weight overall. The β_t and the CRRs of h_t applied to the test set $\mathcal{Q}$ individually are shown in Fig.3:D, from which it can been seen that, although the test CRR of individual $h_t(\mathcal{Q})$ is only around 70%, their mixture h_f can boost the CRR up to more than 92% as shown in Fig.3:C, where it can be also observed that B-JD-LDA continued to improve the test CRRs ($h_f(\mathcal{Q})$) when an appropriate value of $r = 3$ was used in the learner, even long after the training CRRs ($h_f(\mathcal{Z})$) had reached 100%, clearly showing the beautiful property of the boosting algorithm as a large margin classifier against the overfitting.

Table 1. Comparisons of the CRRs (%) in five runs.

runs	Eigenfaces		S-JD-LDA		B-JD-LDA
	R^*/M^*	R_{20}	R^*/M^*	R_{20}	R_{20}/T^*
1st	78.4/63	73.1	88.9/30	83.7	91.4/15
2nd	77.8/147	71.5	88.9/30	86.7	93.6/39
3rd	73.1/121	67.9	88.6/32	84.8	93.6/18
4th	78.9/91	71.7	87.3/46	81.7	91.7/36
5th	72.3/131	65.9	87.8/36	80.3	93.1/27
Ave.	76.1/111	70.0	88.3/35	83.4	92.7/27

In addition, a quantitative comparison regarding the CRRs on the test set $\mathcal{Q}$ among the three methods is summarized in Table 1, where R_{20} denotes the CRR with $M = 20$, R^* denotes the CRR with the best found M^*, T^* denotes the iteration number used to find the reported R_{20}, and $r = 3$ was used in B-JD-LDA. Due to the computational demand, the optimal M for B-JD-LDA was not sought in the experiment. However, it can be clearly seen from Table 1 that B-JD-LDA even with the sub-optimal CRRs, R_{20}, has greatly outperformed Eigenfaces and S-JD-LDA with their best found results, R^*. The averaged improvement of B-JD-LDA (R_{20}) against S-JD-LDA (R^*) is around 4.5%, while the improvement is up to 9.3% given the same value of $M(= 20)$. Also, it should be noted at this point that only $T^* = 27$ iterations, in average, are required to find an excellent result using the B-JD-LDA framework. Such a computational cost is affordable for most personal computers.

4. CONCLUSION

A novel method for face recognition has been introduced in this paper. The proposed method overcomes the limitations of traditional LDA techniques by utilizing a boosting algorithm to form a mixture of LDA models, which can be used to address the nonlinearity commonly encountered in complex FR tasks. With the introduction of the PCDD, a strong connection between the boosting algorithm and the LDA-based learners is built. By manipulating the PCDD, a set of LDA sub-models can be produced in a manner of automatic gain control. Unlike most traditional mixture models that are based on cluster analysis, these sub-models are generalized in the context of classification error minimization. The effectiveness of the proposed method including boosting power and robustness against overfitting has been demonstrated through experimentation using the FERET database. It is anticipated that in addition to JD-LDA, the performance of other LDA variants may be greatly enhanced through the B-JD-LDA framework.

5. REFERENCES

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

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

[3] Yoav Freund and Robert E. Schapire., "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55(1), pp. 119–139, 1997.

[4] Rohit Lotlikar and Ravi Kothari, "Fractional-step dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 623–627, 2000.

[5] Alejandro Murua, "Upper bounds for error rates of linear combinations of classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 591–602, May 2002.

[6] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick 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.

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