Integrating Discriminant and Descriptive Information for Dimension Reduction and Classification

Jie Yu, Qi Tian, *Senior Member*, *IEEE*, Ting Rui, Thomas S. Huang, *Fellow*, *IEEE*

Abstract *—In this paper, a novel hybrid dimension reduction technique for classification is proposed based on the hybrid analysis of principal component analysis (PCA) and linear discriminant analysis (LDA). LDA is known for capturing the most discriminant features of the data in the projected space while PCA is known for preserving the most descriptive ones after projection. Our hybrid technique integrates discriminant and descriptive information and finds a richer set of alternatives beyond LDA and PCA in a 2D parametric space, which fits a specific classification task and data distribution better. Theoretical study shows that our technique also alleviates the singularity problem of scatter matrix, which is caused by small training set, and increases the effective dimension of the projected subspace. In order to find the hybrid features adaptively and avoid exhaustive parameter searching, we further propose a boosted hybrid analysis method that incorporates a non-linear boosting process to enhance a set of hybrid classifiers and combine them into a more accurate one. Compared with the other techniques that aim at combining PCA and LDA, our approaches are novel because our method finds alternatives to LDA and PCA in a 2D parameter space and the boosting process provides enhancement and robust combination of the classifiers. Extensive experiments are conducted on benchmark and real image databases to compare our proposed methods to the state-of-the-art linear and non-linear discriminant analysis techniques. The results show the superior performance of our hybrid analysis methods.*

Index Terms—*Image classification, Information retrieval, Pattern recognition，Artificial intelligence*

Manuscript received in June 2006. This work was supported in part by the Army Research Office (ARO) grant under W911NF-05-1-0404, and by the Center of Infrastructure Assurance and Security (CIAS), the University of Texas at San Antonio.

Q. Tian is with the Department of Computer Science, University of Texas at San Antonio, San Antonio, TX 78249, USA (corresponding author; phone: (210) 458-5165; fax: (210) 458-4437; email: qitian@cs.utsa.edu).

J. Yu is with the Department of Computer Science, University of Texas at San Antonio, TX, USA (email: jyu@cs.utsa.edu).

T. Rui is with the Engineering Institute of Engineering Corps, PLA University of Science and Technology, Nanjing, China (email: rting@public1.ptt.js.cn).

T. Huang is with the Beckman Institute for Advance Science and Technology, University of Illinois at Urbana-Champaign, Illinois, USA (email: huang@ifp.uiuc.edu).

I. Introduction

CURSE of dimensionality is an impediment for most computer vision applications as the search space grows exponentially with the data dimension. Principal component analysis (PCA) [1, 2] and linear discriminant analysis (LDA) [3, 4, 5] are both well-known techniques for feature dimension reduction. LDA constructs the *most discriminant* features while PCA constructs the *most descriptive* features in the sense of packing most "energy".

LDA plays a key role in many research areas in science and engineering such as face recognition [6, 7], image retrieval [8, 9], and bioinformatics [10]. LDA is a simple algorithm that is used for both dimension reduction and classification. In either case, LDA aims at maximizing class separability in a low dimensional space by selecting the feature vectors ***w*** which maximize $\frac{|w^T S_B w|}{|w^T S_W w|}$, where S_B measures the variance between the class means, and S_w measures the variance of the samples in the same class.

PCA is a useful statistical technique that has found various applications in many fields [1, 2]. It is considered as one of the simplest and best-known *Data Analysis* techniques. Its goal is to replace the original (numerical) variables with new numerical variables called "Principal Components" that have the following properties: (1) They can be ranked by decreasing order of "importance" (this term can be given a precise meaning). The first few most "important" Principal Components account for most of the information in the data. In other words, one may then discard the original data set, and replace it with a new data set with the same observations, but fewer variables, without throwing away too much information. (2) These new variables are uncorrelated.

In computer vision community, when comparing LDA with PCA, there is a tendency to prefer LDA to PCA, because, as intuition would suggest, the former deals directly with discrimination between classes, whereas the latter deals without paying particular attention to the underlying class structure. For example, when the data of each class can be represented by a single Gaussian distribution and share a common covariance matrix, LDA will outperform PCA. However, LDA as well as other discriminant analysis techniques are not guaranteed to work where the assumptions of the method do not hold. An interesting result is reported by Martinez and Kaka [11] that this is not always true in their study on face recognition. PCA might outperform LDA when the number of samples per class is small or when the training data non-uniformly sample the underlying distribution [12, 13].

Figure 1 shows two classical examples where PCA outperforms LDA in Fig. 1(a) and where LDA outperforms PCA in Fig. 1(b), respectively.

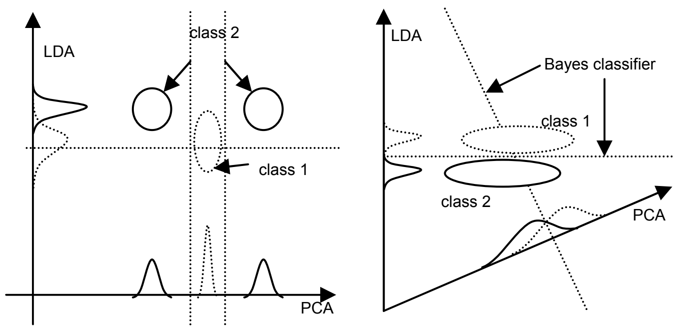


Figure 1. PCA outperforms LDA in (a), and LDA outperforms PCA in (b) ([13])

II. DISCRIMINANT AND DESCRIPTIVE FEATURE

A. *Problems*

The Small Sample Set problem and High Dimensionality problem are the two major challenges for many computer vision applications such as Content-based Image Retrieval (CBIR) [14]. These problems can be alleviated by discriminant analysis techniques, which are concerned with data in which each observation comes from one of several well-defined classes or populations. Usually, in discriminant analysis, assumptions are made about the structure of the populations and discriminant features are discovered for assigning future observation to one of the classes to minimize the probability of misclassification or meet some similar criteria.

In CBIR, the data structure of the image databases could be very complicated that each category should be modeled as one or multiple Gaussian mixture (e.g., red car and white car should be modeled as two Gaussians when color is the discriminating subspace and modeled as one Gaussian when shape is the discriminating subspace). If the true number of components *C* in mixture model is known *a priori*, then LDA is supposed to find an optimal linear transformation to separate them in a discriminating subspace. However, LDA faces several major problems such as effective dimension and regularization [15].

As we point out in Section I, PCA cannot take the semantic label information of the training samples to discover the discriminant features, which can facilitate classification. However, it outperforms LDA in some classification tasks. Mathematically PCA tries to maximize the covariance matrix of all samples in the projected space that is a more robust estimation than the within-class scatter matrix in LDA. Consequently, PCA is more robust to small sample set problem. Furthermore, PCA has effective dimension d_{PCA}, which is larger than that of LDA ($d_{LDA} = C-1$), which means more accurate data modeling can be achieved in higher dimensional subspace. Based on the above analysis, it is obvious that combining PCA and LDA may lead to discovering robust features for dimension reduction and classification.

B. *Related Work*

The strength of integrating discriminant and descriptive features has been studied in several research work and different related techniques that combine PCA and LDA have been proposed in the literature. Kriegman *et al.* proposed Fisherfaces to substitute PCA-based Eigenfaces [6]. Fidler and Leonardis further studied the above method on how to appropriately perform PCA to facilitate LDA [16]. Instead of applying PCA before LDA, research has been done in applying PCA and LDA simultaneously. Integrated PCA-FLD was proposed to combine PCA and LDA with one parameter [17,18]. Talukder and Casasent also proposed a linear combination of PCA and LDA (L.PCA-LDA) [19], in which they use a class-specific PCA instead of the normal PCA. Wang *et al.* proposed a Principal Discriminant Analysis (PDA) which finds a simple linear combination of the optimal projection found in PCA and LDA in processing chemical sensor arrays [20]. To alleviate the small sample set problem that LDA often fails to handle, null-space LDA [21], direct LDA [22] and dual-space LDA [23] have been proposed.

As we will elaborate in Section III, compared to these related work, our proposed methods are novel because: 1) we use two parameters to control the balance between PCA and LDA. Thus, our methods can search a 2D parameter space and could find combination that better fits the classification task and data. Moreover, 2) other methods use simple parameter selection using cross-validation which could suffer from a biased training data set, e.g. not representative or imbalanced data set. In our method we use AdaBoost to provide robust combination and it can enhance the classifier iteratively. We conducted experiments to compare our methods to these related ones. The experiment results in Section IV will show the superior performance of our techniques.

III. THE HYBRID FEATURE DIMENSION REDUCTION

A. *Hybrid Discriminant Analysis*

Based on the analysis of the problems and properties of PCA and LDA, we are motivated to propose a framework called Hybrid Discriminant Analysis (HDA) that can capture both discriminant and descriptive features and unify both techniques. We design a novel optimal function as

$$W_{opt} = \arg\max_{W} \frac{|W^T[(1-\lambda)\cdot S_B + \lambda \cdot S_\Sigma]W|}{|W^T[(1-\eta)\cdot S_W + \eta \cdot I]W|} \quad (1)$$

where λ, η are two parameters, S_Σ is the covariance matrix of all the training samples, and *I* is an identity matrix. The range of the parametric pair (λ,η) is from $(0,0)$ to $(1,1)$.

With different (λ,η) values, the equation (1) provides a rich set of alternatives to PCA and LDA in a 2D parametric space as shown in Figure 2. $(\lambda=0,\eta=0)$ reduces to the full LDA; $(\lambda=1,\eta=1)$ recovers the full PCA; $(\lambda=0,\eta=1)$ gives a subspace that is mainly defined by maximizing the scatters among all the classes with minimal effort on clustering each class; $(\lambda=1,\eta=0)$ gives a subspace that mainly preserves the most energy while minimizing the scatter matrices of within-classes; $(\lambda=\frac{1}{2},\eta=\frac{1}{2})$ gives a subspace that is discriminative while preserving as much energy as possible, a trade-off between LDA and PCA.

Clearly LDA and PCA are the special cases in the hybrid discriminant analysis. Since (λ,η) values can be any real number from $(0,0)$ to $(1,1)$, more alternatives of the feature dimension reduction schemes between and beyond PCA and LDA, which haven't been studied before, can be easily obtained by setting parameter pair (λ,η).

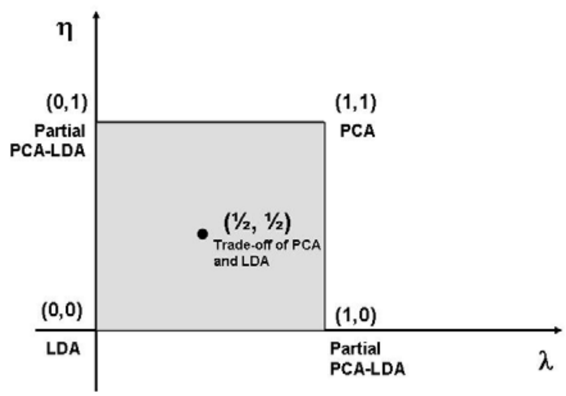


Figure 2. The hybrid discriminant analysis in the parametric space (λ,η).

The difference of the proposed hybrid analysis from the existing variants for linear discriminant analysis alleviates the two problems discussed in Section II.A: (i) If we examine the denominator of equation (1), it is a full rank matrix and consequently HDA is more robust to the small sample set problem. The difference from simple regularization is that we also consider preserving the descriptive features in the nominator of equation (1). (ii) Due to the full rank of the nominator, HDA has *effective dimension* up to the dimensionality of data (D), while for LDA it is at most $C-1$ ($C-1 << D$ usually). This gives the hybrid approach significantly higher capacity for informative density modeling, for which FDA (Fisher Discriminant Analysis, C=2) has virtually none. The above two differences are responsible for the robust performance of the hybrid discriminant analysis. It is worth mentioning that this framework can easily adopt variants of PCA and LDA to generate a new set of projections that discovers different features. For example, if we substitute S_B with the data covariance and S_W with the noise covariance (which can be estimated analogous to equation (1), but over examples sampled from the assumed noise distribution), we obtain *oriented* PCA (OPCA) [2], which aims at finding a direction that describes most variance in the data while avoiding known noise as much as possible.

B. Boosted Hybrid Discriminant Analysis

Given the data distribution and classification task, the optimal projection of HDA that offers the best classification performance could lie outside of the line of PCA and LDA in the parametric space of (λ,η), as we indicated in Section III.A. We have to search the whole parametric space to find the best parameter setting. This will result in extra computational complexity. It is also true that the best pair (λ^*,η^*) we found for one particular dataset is often different from that of another dataset and therefore this cannot lead to a generalization as illustrated in Figure 3. Note that our HDA is a dimension reduction technique and it is always followed by a classifier, e.g., nearest neighbor, SVM [24] or Bayesian, in a classification task. For simplicity, we call the HDA projection and the trained classifier in its projected space as HDA classifier in the rest of the paper.

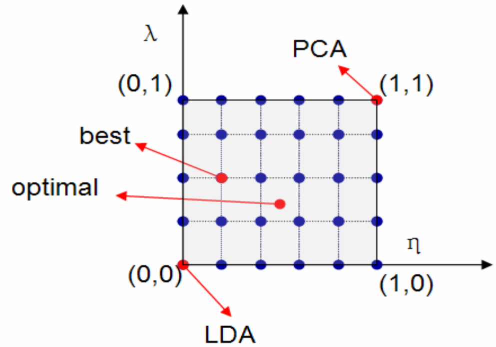


Figure 3. Best HDA classifier found in the parameter space by subsampling may not be the one with optimal setting.

AdaBoost is proposed to combine weak classifiers into a powerful one with bounded error rate [25]. It has been successfully applied in many research projects, such as Viola and Jones's fast face detection system [26], while its application in a dimension reduction scheme has not been found in literature. We adopt the idea of AdaBoost into a Boosted HDA framework. Unlike most of the existing approaches that boost individual features to form a composite classifier, our scheme boosts both the individual features and a set of weak classifiers. Our algorithm is shown below:

Algorithm **AdaBoost with HDA as weak learner**

Given: Training Sample set X and corresponding label Y
K HDA classifiers with different (λ,η)

Initialization: weight $w_{k,t=1}(x) = 1/|X|$

AdaBoosting

For $t = 1,\ldots,T$

For each classifier $k = 1,\ldots,K$ do

- Find the optimal projection based on weighted mean for all the samples, positive samples and negative samples and weighted scatter matrices in the following way. Note that $\sum_{x\in X} w_{k,t}(x) = 1$.
  - (i) Update weighted mean μ_{all}, μ_p, and μ_n in similar way as follows
    $$\mu_{all} = \sum w_{k,t}(x)\cdot x / \sum w_{k,t}(x)$$
  - (ii) Update within-class and between-class scatter matrices and co-variance matrix
- Train the base classifier on the data projected by the optimal projection W found in Eq. (1)
- Get the confidence-rated prediction on each sample $h_{k,t}(\mathbf{x}) \in (-1,1)$
- Compute the weight of each classifier based on its classification error rate $\varepsilon_{k,t}$
  $$\alpha_{k,t} = \frac{1}{2}\ln\left(\frac{1-\varepsilon_{k,t}}{\varepsilon_{k,t}}\right)$$
- Update the weight of each sample
  $$w_{k,t+1}(\mathbf{x}) = w_{k,t}(\mathbf{x})\exp(-\alpha_{k,t}\cdot h_{k,t}(\mathbf{x})\cdot y)$$

End for each classifier
End for t
The final prediction $H(\boldsymbol{x}) = sign(\sum_{k,t} \alpha_{k,t} \cdot h_{k,t}(\boldsymbol{x}))$

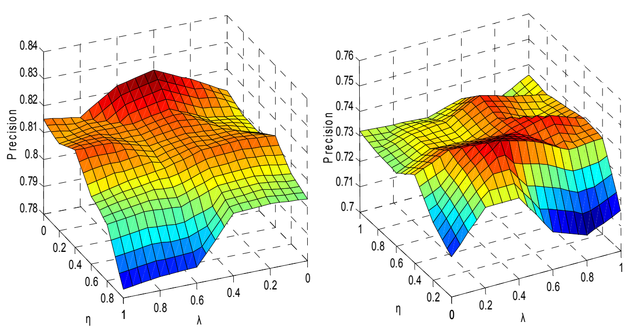
IV. EXPERIMENTS AND ANALYSIS

Extensive experiments have been performed to evaluate the hybrid discriminant analysis and its boosted variant and compare to state-of-the-art dimension reduction techniques on various benchmark image databases for image classification. In the experiments, the original data are first projected to the lower subspace and Bayesian classifier based on Gaussian models is applied on the projected data for classification. The distribution of the labeled samples, *i.e.*, training data set, is learned and the unlabeled test samples will be classified with the learned distribution. Empirical observations suggest that the transformed image data often approximates a Gaussian in the lower subspace, and so in our implementation, we use low-order Gaussian mixture to model the transformed data. Since it is assumed that there is no priori information available for the class number, we consider it as a two-class classification problem.

A. Comparison of PCA, LDA and HDA

In the first experiment, we compare the performance of feature dimension reduction techniques using PCA, LDA and HDA. UCI benchmark data sets are used in the experiments. We used two different data sets: Heart and Breast-Cancer. The data dimensions of these data sets are 13 and 9, respectively. The sizes of the training sets are 170 and 200, and the sizes of the testing data sets are 100 and 77 for these two datasets, respectively.

Figure 4 shows the classification precision using PCA $(\lambda = 1, \eta = 1)$, LDA $(\lambda = 0, \eta = 0)$ v.s. the other different parameter settings (λ, η) of HDA. The projected subspace dimension is one. For Heart data, the precision is 79.3% for PCA, 81.2% for LDA, and 83.9% for the HDA for the best pair $(\lambda = 0.4, \eta = 0)$. For Breast-Cancer data, the precision is 73.6% for PCA, 71.5% for LDA, and 75.6% for the HDA for the best pair $(\lambda = 0.4, \eta = 0.2)$.



(a) Heart Dataset (b) Breast-Cancer Dataset

Figure 4. Performance of hybrid discriminant analysis on (a) Heart Dataset, and (b) Breast-Cancer Dataset with PCA at $(\lambda = 1, \eta = 1)$ and LDA at $(\lambda = 0, \eta = 0)$

From the results in Figure 4, we can find that LDA performs better than PCA on the Heart dataset while on the Breast-Cancer dataset PCA outperforms LDA. It verifies the previous research conclusion we mentioned in Section I that there is no guarantee that LDA always perform better than PCA. Consequently, we can conclude that descriptive information may help classification as discriminant one does and integrating both of them yields more accurate classifier. It is also clear that the performances on different data sets have significantly different patterns over the parameter space. It is hard to find some specific setting that always gives optimal results. Under that circumstance, our boosting scheme is desirable in that it enhances and combines a set of classifiers into a more accurate classifier without parameter searching.

B. Comparison of HDA to the State-of-the-art Variants of Discriminant Analysis

In this experiment, we compare our hybrid analysis with the state-of-the-art variants of discriminant analysis algorithms such as discriminant-EM algorithm (DEM) [9], kernel DEM (KDEM) [27], BDA [28], kernel BDA (KBDA) [28], and their regularized algorithms [29]. The data sets used in the experiments are the MIT facial image dataset (2358 images) and non-face images (2958 images) from Corel database [30]. All the face and non-face images are scaled down to 16×16 gray-scale images and normalized feature vector of dimension 256 is used to represent the image. We set the size of the training set as 100, 200, 400, and 800, respectively. Compared with the feature vector dimension of 256, the training sample size is set from relatively small to relatively large.

Several conclusions can be drawn from results in Table 1: (i) Regularization is very important for sample-based estimations such as DEM and BDA. Even simple regularization can significantly improve the classification performance, *e.g.*, for DEM and BDA in average by 15%~40%. We have tested different regularization methods. The classification performance can be further improved in average by 1~5% using a parameterized way [29]. As the size of the training samples increases and becomes larger than the number of feature dimension, the error rate drops since the possibility of $|W^T S_w W| = 0$ in LDA becomes smaller that means the regularization becomes less important. (ii) When compared with the regularized DEM and BDA, the PCA-LDA pair performs much better. (iii) When compared with the regularized KDEM and KBDA, The PCA-LDA pair performs better than KBDA, and comparable to KDEM. It should be noted only linear transformation is used in the hybrid analysis, but it is more efficient than nonlinear algorithms such as KDEM and KBDA. The performances of other hybrid pairs are very close to the KDEM and KBDA. All these show the robust performance of the hybrid analysis.

Table 1. Comparison of DEM, BDA, KDEM, KBDA, and the hybrid pairs

Error Rate (%)	Size of Training Set			
	100	200	400	800
DEM w/o regulation	48.75	49.43	16.7	11.2
DEM w/ simple regulation	10.5	19.3	15.0	9.0
BDA w/o regulation	49.2	49.9	50.0	20.85
BDA w/ simple regulation	34.7	25.4	18.5	19.3
KDEM w/ simple regularization	6.93	1.43	0.7	0.5
KBDA w/ simple regularization	3.04	2.89	2.58	1.44

PCA-LDA (λ^*,η^*)	1.73 (0.4, 0.2)	2.2 (0.2, 0)	1.5 (0.2, 0)	0.73 (0.2, 0.2)

C. *Comparison of Boosted HDA with the State-of-the-art PCA-LDA Related Techniques*

To evaluate how well our boosted Hybrid Discriminant Analysis can discover the discriminant and descriptive features of face images, we test it on three benchmark face image databases, Harvard [6], ATT [31] and UMIST [32], with change of illumination, expression and head pose, respectively. We randomly chose one person's face images as positive and the rest face images of others are considered as negative. On Harvard data set, we used 220 images as training set and 440 as testing set. 134 images were used for training and 267 images were used for testing on ATT data set. 188 images and 376 images were used in the experiment on UMIST data set for training and testing respectively. The experiments are repeated 100 times to get average performance. For comparison purpose, five state-of-the-art PCA and LDA related techniques mentioned in Section II.B are also tested on the same databases.

Table 2. Comparison of HDA, Boosted HDA and the state-of-the-art PCA-LDA techniques

Error Rate (%)	Harvard Dataset			ATT Dataset	UMIST Dataset
	Subset 1	Subset 2	Subset 3		
Eigenface	1.2	5.4	25.3	28.1	38.3
Fisherface [6]	0.7	1.4	3.7	19.5	31.2
N.PCA-LDA [16]	0.8	1.1	2.8	17.6	32.9
L.PCA-LDA [19]	0.6	1.3	2.2	16.3	23.6
PDA [20]	0.9	1.2	3.4	17.9	29
HDA	0.4	0.7	2.3	11.3	27.9
Boosted HDA	**0.3**	**0.5**	**1.9**	**7.3**	**18.5**

The results are listed Table 2 with smallest error rate in bold. Our HDA performs better than other PCA-LDA related techniques on 4 of 5 tests and it is comparable to L. PCA-LDA on UMIST data set. Eigenface performs the worst among these techniques because it is an unsupervised method and no discriminant information is discovered for classification. The boosted version provides best classification in all 5 tests with the error rate decreasing at least 21%. It shows that our methods are more robust to the changes of illumination, expression and pose than other techniques because descriptive feature is captured along with discriminant one.

D. *Efficiency of Boosted HDA*

As we can imagine, for simply searching the parametric space, the larger the searched space, the better is the performance of the best single classifier. However, the exhaustive search means more computational costs. Table 3 shows the boosted PCA-LDA classifier and the best single classifier of PCA-LDA analysis in the different search space. Due to the space limit, only the performance on Heart dataset is shown. Similar results are obtained on other datasets. The range of (λ,η) is between $(0,0)$ and $(1,1)$. The searching step size of λ and η is 0.25, 0.2, 0.167, and 0.1 resulting in the searching space size 16, 25, 36, and 100, respectively.

Table 3. Comparison of the Boosted HDA and best single HDA classifier on Heart dataset

Error Rate (%)		Best single (λ*, η*)	Boosted HDA		
			$T=1$	$T=3$	$T=5$
# of classifiers	16	17.85 (0.33, 0)	17.03	16.45	16.45
	25	17.2 (0.5, 0.25)	16.95	16.6	16.5
	36	17.6 (0.4, 0)	16.9	16.55	16.0
	100	16.95 (0.5, 0.1)	17.05	16.37	15.9

Although Boosted HDA did not provide big performance boost in this experiment, such a minor performance enhancement may be important for performance-sensitive tasks such as biometric identification. Most importantly, from Table 3, we find the boosted PCA-LDA classifier is not sensitive to the size of the search space, *e.g.*, the boosted PCA-LDA classifier from a weak set of 16 single classifiers achieves the better performance (16.45%) than the best single classifier (16.95%) of search space size 100 after three iterations. Instead of searching a large parametric space to find the better single classifier, the boosted PCA-LDA classifier provides a more efficient way to combine a small set of weak classifiers into a more powerful one.

V. Conclusion and Discussion

In this paper, we propose a novel hybrid feature dimension reduction scheme for classification. It integrates discriminant and descriptive information of the data and provides a richer set of alternatives to traditional techniques. LDA and PCA are unified in this framework as two special cases. Optimal projection that fits the data distribution and classification task can be found in a 2-D parameter space. In order to find both discriminant and descriptive features adaptively, boosting process is adopted, which also avoids parameter searching.

The novel HDA not only compensates for regularization that is afflicted by all sample-based estimation methods, but also increases the effective dimension of the projected subspace which enables complex data modeling in the projected space with high dimensionality. The boosted HDA provides three desirable properties. First, it can provide a unified solution to find optimal features for specific application and database. Second, it avoids parameter searching which is easy to bias to the training data set. Finally, the weighted training scheme in boosting adds indirect non-linearity and adaptivity to the linear methods and thus enhances the classification by iterations. The extensive experiments on benchmark data sets, facial image data sets and real image databases show the superior performance of our proposed methods.

Many interesting issues are worth investigating in the future. 1) Although our methods work well in the experiments, further theoretical study on the limitation of the integration and boosting scheme is expected. 2) We are interested in combining the Generalized PCA (GPCA) analysis [33] with LDA in a unified framework. 3) Some work is ongoing on applying the hybrid analysis in biometrics (face, fingerprint, iris classification) and non-computer vision area such as gene expression-based microarray classification in bioinformatics.

REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*. 2nd edition, New-York: Springer-Verlag, 2002.

[2] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks*, New York: Wiley, 1996.

[3] R. A. Fisher, "The use of multiple measurement in taxonomic problems," *Annals of Eugenics*, vol. 7, pp.179-188, 1936.

[4] R.A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, vol. 8, pp.376-386, 1938.

[5] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., 2001.

[6] P. N. Belhummeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 711-720, 1997.

[7] K. Etemad and R. Chellapa, "Discriminant analysis for recognition of human face images," *Journal of Optical Society of American,* 14(8): 1724-1733, 1997.

[8] D. Swets and J. Weng, "Hierarchical discriminant analysis for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(5), 396-401, 1999.

[9] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant EM algorithm with application to image retrieval," *Proc. of IEEE Conf. Computer Vision and Pattern Recognition,* South Carolina, June 13-15, 2000.

[10] W J. Ewans and G. R. Grant, *Statistical methods in bioinformatics*, Springer-Verlag, 2001.

[11] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 228-233, February 2001.

[12] R. Beveridge, K. She, B. Draper, and G. Givens, "A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, Hawaii, vol. 1, pp. 535-542, 2001.

[13] M. Zhu, A. M. Martinez, and H. Tan, "Template-based recognition of sitting postures," *Proc. IEEE Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, Madison, WI, 2003.

[14] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans PAMI*, pp. 1349-1380, December 2000.

[15] J. Friedman, "Regularized discriminant analysis," *Journal of American Statistical Association*, vol. 84, no. 405, pp. 165-175, 1989.

[16] S. Fidler and A. Leonardis, "Robust LDA classification by subsampling," *IEEE Conf. CVPR Workshop*, June 2003.

[17] Cheng, X., Chen, Y.R., Tao, Y., Wang, C.Y., Kim, M.S., Lefcourt, "A novel integrated PCA and FLD method on hyperspectral imaging feature extraction method for cucumber chilling damage inspection," *Transactions of ASAE*, 47:1313-1320, 2004

[18] Xuemei Cheng, Yang Tao, Yud-Ren Chen, Xin Chen, "Integrated PCA-FLD method for hyperspectral imagery feature extraction and band selection," *IEEE International Symposium on Biomedical Imaging*, April 6-9, 2006

[19] A. Talukder and D. Casasent, "A general methodology for simultaneous representation and discrimination of multiple object classes," *Optical Engineering, special issue on "Advances in Recognition techniques"*, March 1998.

[20] M. Wang, A. Perera-Lluna and R. Gutierrez-Osuna, "Principal discriminant analysis for small-sample-size problems: application to chemical sensing," *Proc. of IEEE SENSORS*, Vienna, October 2004.

[21] J. Ye and T. Xiong, "Null space versus orthogonal linear discriminant analysis," *ACM Intl. Conf. on Machine Learning*, 2006

[22] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with applications to face recognition," *Pattern Recognition*, 34, 2067—2070, 2001

[23] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2004

[24] V. Vapnik, *The nature of statistical learning theory*, second edition, Springer-Verlag, 2000.

[25] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence,* 14(5):771-780, Sep., 1999.

[26] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2001

[27] Q. Tian, Y. Wu, J. Yu, and T.S. Huang, "Self-supervised learning based on discriminative nonlinear features for image classification," *Pattern Recognition, Special Issue on Image Understanding for Digital Photographs*, Vol. 38, No. 6, pp. 903-917, 2005.

[28] X. Zhou and T.S. Huang, "Small sample learning during multimedia retrieval using biasMap," *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, Hawaii, December 2001.

[29] Q. Tian, J. Yu, T. Rui, and T.S., Huang, "Parameterized discriminant analysis for image classification," *Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME 2004)*, June 27-30, Taipei, Taiwan, 2004.

[30] Corel Corporation. http://www.corel.com/

[31] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," *IEEE Workshop on Applications of Computer Vision,* Sarasota FL, December 1994.

[32] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. PAMI*, Vol. 20, 1998.

[33] K. Huang, Y. Ma, and R. Vidal, "Minimum effective dimension for mixtures of subspaces: a robust GPCA algorithm and its applications," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, 2004.