PARETO-OPTIMAL DISCRIMINANT ANALYSIS

Felix Juefei-Xu and Marios Savvides

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

ABSTRACT

In this work, we have proposed the Pareto-optimal discriminant analysis (PDA), an optimally designed linear subspace learning method that harnesses advantages across many well-known methods such as PCA, LDA, UDP and LPP. By optimizing over the joint objective function and carrying out an alternative coefficients updating scheme, we are able to obtain a linear subspace which is optimized to truly maximize the objective function in discriminant analysis. The proposed method also provides flexibility for formulating the linear transformation matrix in an overcomplete fashion, allowing for a sparse representation. We have shown, in the context of large scale unconstrained face recognition and illumination invariant face recognition, that our proposed PDA significantly outperforms other linear subspace methods.

***Index Terms*—** PCA, LDA, UDP, LPP, Discriminant Analysis, Sparse Representation

1. INTRODUCTION

Data with high dimensionality, such as facial images, has always been the Padora's box in the world of computer vision and pattern recognition. This is partially explained by the curse of dimensionality [1]. In order to overcome this, many techniques of dimensionality reduction have been developed to find a low dimensional representation $\mathbf{Y} \in \mathbb{R}^{l \times n}$ of the high dimensional data $\mathbf{X} \in \mathbb{R}^{d \times n}$. This can be easily achieved by a linear transformation matrix $\mathbf{W} \in \mathbb{R}^{d \times l}$, also known as linear subspace: $\mathbf{Y} = \mathbf{W}^\top \mathbf{X}$. In the past two decades, researchers have been devoted to finding *meaningful* subspace $\mathbf{W}$ such that the low dimensional representation can serve certain purposes.

Related Work: For instance, principal component analysis (PCA) [2] aims to find a subspace that well preserves the second-order statistics and captures maximal variability of the data, but does not care about the class separation for classification purpose. Discriminant analysis methods such as linear discriminant analysis (LDA) [3] and unsupervised discriminant projection (UDP) [4], on the other hand, seeks to obtain subspaces that are great for class separation. Manifold approaches such as Isomap [5], locally linear embedding (LLE) [6], and Laplacian eigenmap (LE) [7] conduct nonlinear dimensionality reduction, with the assumption that the high dimensional data lies on a low dimensional manifold embedded on an ambient space. Locality preserving projections (LPP) [8] is a direct linear approximation of LE and shares many of the data representation properties of nonlinear techniques such as Isomap, LLE, and LE. LPP finds linear projective subspaces that optimally preserve the neighborhood proximity structure of the data. In this work, we propose a Pareto-optimal[1] allocation of various linear subspace learning methods to optimally achieve class separation while retaining many nice properties brought by various leaning methods. We call this method Pareto-optimal discriminant analysis (PDA).

The main contributions are: (1) We propose the PDA that takes advantage of various subspace leaning methods by maximizing a joint objective function. (2) We employ an alternative coefficients updating scheme to obtain the optimal solution by manipulating the optimization function such that two sets of coefficients can be updated iteratively. (3) Flexibility for formulating the linear transformation matrix in an overcomplete fashion, allowing for a sparse representation.

2. FORMULATION OF PROPOSED METHOD

2.1. Motivation

We know that the objective functions from many linear subspace learning methods, such as PCA [2], LDA [3], UDP [4], and LPP [8], are all in the form of the generalized Rayleigh quotient as shown below:

$$J(\mathbf{w}) = \frac{\mathrm{tr}\left(\mathbf{w}^\top \mathbf{S}_i \mathbf{w}\right)}{\mathrm{tr}\left(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w}\right)} \tag{1}$$

where $(\mathbf{S}_i, \mathbf{Z}_i)$ are positive semi-definite matrices and can take value $(\mathbf{S}_T, \mathbf{I})$ from PCA, $(\mathbf{S}_B, \mathbf{S}_W)$ from LDA, $(\mathbf{S}_N, \mathbf{S}_L)$ from UDP, and $(\mathbf{S}_D, \mathbf{S}_{L'})$ from LPP. Solving each of the Rayleigh quotient problems would return a linear subspace spanned by the "optimal" projection direction vectors $\mathbf{w}$ that serve certain purposes in each of the subspace learning methods. Ideally, we would want to achieve a subspace that is spanned by projection vectors which are "optimal" in all purposes. We achieve this by formulating the Pareto-optimal discriminant analysis.

[1] *Pareto optimality* is a concept in economics with applications in engineering. It is a state of economic allocation of resources in which it is impossible to make any one further better off without making at least one individual worse off. [9]

2.2. Pareto-optimal Discriminant Analysis

Let $\{\mathbf{\Psi}_i\}_{i=1}^N$ be independent events, each of which executes one of the subspace learning methods: $\Psi_i|\mathbf{w} = \mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})/\mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})$. So far we have 4 events (PCA, LDA, UDP, and LPP) denoted as $\Psi = (\Psi_1, \Psi_2, \Psi_3, \Psi_4)$. Pareto-optimal discriminant analysis (PDA) seeks an overall optimization framework from joint events. The optimal projection direction $\mathbf{w}$ over the joint events is achieved by:

$$\begin{aligned}\mathbf{w}^* &= \arg\max_{\mathbf{w}} (\Psi_1, \Psi_2, \Psi_3, \Psi_4|\mathbf{w}) \\ &= \arg\max_{\mathbf{w}} (\Psi_1|\mathbf{w})(\Psi_2|\mathbf{w})(\Psi_3|\mathbf{w})(\Psi_4|\mathbf{w}) \\ &= \arg\max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{S}_T \mathbf{w}}{\mathbf{w}^\top \mathbf{I} \mathbf{w}} \cdot \frac{\mathbf{w}^\top \mathbf{S}_B \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_W \mathbf{w}} \cdot \frac{\mathbf{w}^\top \mathbf{S}_N \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_L \mathbf{w}} \cdot \frac{\mathbf{w}^\top \mathbf{S}_D \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_{L'} \mathbf{w}} \\ &= \arg\max_{\mathbf{w}} J(\mathbf{w}) \end{aligned} \tag{2}$$

Now, let's consider a more general case where $(\mathbf{S}_i, \mathbf{Z}_i)$ can be taken from any discriminant analysis or linear subspace learning methods even not yet covered.

$$J(\mathbf{w}) = \prod_{i=1}^N \left(\frac{\mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})}{\mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})} \right)^{\alpha_i} \tag{3}$$

where α_i's are the mixing coefficients. Taking the logarithm on $J(\mathbf{w})$ will lead to:

$$\begin{aligned} L(\mathbf{w}) = \log J(\mathbf{w}) &= \sum_{i=1}^N \alpha_i \log \frac{\mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})}{\mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})} \\ &= \sum_{i=1}^N \alpha_i \log \mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w}) - \sum_{i=1}^N \alpha_i \log \mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w}) \end{aligned} \tag{4}$$

Taking derivative with respect to $\mathbf{w}$ and set to 0:

$$\frac{\partial L(\mathbf{w})}{\partial \mathbf{w}} = \sum_{i=1}^N \alpha_i \frac{2\mathbf{S}_i \mathbf{w}}{\mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})} - \sum_{i=1}^N \alpha_i \frac{2\mathbf{Z}_i \mathbf{w}}{\mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})} = 0$$

Let $u_i = \mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})$ and $v_i = \mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})$, we have:

$$\sum_{i=1}^N \alpha_i \frac{\mathbf{S}_i \mathbf{w}}{u_i} = \sum_{i=1}^N \alpha_i \frac{\mathbf{Z}_i \mathbf{w}}{v_i}$$

$$\sum_{i=1}^N \alpha_i \frac{u_i^{\alpha_i - 1} \mathbf{S}_i \mathbf{w}}{u_i^{\alpha_i}} = \sum_{i=1}^N \alpha_i \frac{v_i^{\alpha_i - 1} \mathbf{Z}_i \mathbf{w}}{v_i^{\alpha_i}}$$

$$\frac{\sum_i (u_i^{\alpha_i - 1} \alpha_i \mathbf{S}_i \prod_{j \neq i} u_j) \mathbf{w}}{\prod_{i=1}^N u_i^{\alpha_i}} = \frac{\sum_i (v_i^{\alpha_i - 1} \alpha_i \mathbf{Z}_i \prod_{j \neq i} v_j) \mathbf{w}}{\prod_{i=1}^N v_i^{\alpha_i}}$$

Notice that the objective function $J(\mathbf{w})$ has appeared in the expression, where:

$$J(\mathbf{w}) = \prod_{i=1}^N \left(\frac{\mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})}{\mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})} \right)^{\alpha_i} = \frac{\prod_i^N u_i^{\alpha_i}}{\prod_i^N v_i^{\alpha_i}} \tag{5}$$

Let $\mathcal{S} = \sum_i (u_i^{\alpha_i - 1} \alpha_i \mathbf{S}_i \prod_{j \neq i} u_j)$, $\mathcal{Z} = \sum_i (v_i^{\alpha_i - 1} \alpha_i \mathbf{Z}_i \prod_{j \neq i} v_j)$ and plug in $J(\mathbf{w})$, then the derivation of PDA becomes: $\mathcal{S}\mathbf{w} = J(\mathbf{w})\mathcal{Z}\mathbf{w}$. Here, $\mathcal{S}$ and $\mathcal{Z}$ are nothing but the linear combinations of $\mathbf{S}_i$ and $\mathbf{Z}_i$. Since $\mathbf{S}_i$ and $\mathbf{Z}_i$ are positive semi-definite, so does $\mathcal{S}$ and $\mathcal{Z}$. We want to maximize $J(\mathbf{w})$, a common way to solve this is to treat $J(\mathbf{w})$ as the largest eigenvalue λ of matrix $\mathcal{Z}^{-1}\mathcal{S}$ (of course $\mathcal{Z}$ has to be invertible) as shown:

$$\mathcal{Z}^{-1}\mathcal{S}\mathbf{w} = \lambda \mathbf{w} \tag{6}$$

2.3. Alternative Updating of Coefficients

To solve the optimization in Equation (6), we employ an alternating rule for the updating of coefficients. Recall that:

$$\underbrace{\sum_i (u_i^{\alpha_i - 1} \alpha_i \mathbf{S}_i \prod_{j \neq i} u_j)}_{\mathcal{S}} \mathbf{w} = J(\mathbf{w}) \underbrace{\sum_i (v_i^{\alpha_i - 1} \alpha_i \mathbf{Z}_i \prod_{j \neq i} v_j)}_{\mathcal{Z}} \mathbf{w} \tag{7}$$

$$\mathcal{Z}^{-1}\mathcal{S}\mathbf{w} = J(\mathbf{w})\mathbf{w} \tag{8}$$

$$\mathcal{S} = \sum_i (u_i^{\alpha_i - 1} \alpha_i \mathbf{S}_i \prod_{j \neq i} u_j) \tag{9}$$

$$\mathcal{Z} = \sum_i (v_i^{\alpha_i - 1} \alpha_i \mathbf{Z}_i \prod_{j \neq i} v_j) \tag{10}$$

where $u_i = \mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})$ and $v_i = \mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})$, and the objective is:

$$J(\mathbf{w}) = \frac{\prod_i^N u_i^{\alpha_i}}{\prod_i^N v_i^{\alpha_i}} \tag{11}$$

Next, from Equation (7) and (11), an alternating coefficients updating scheme is implemented to obtain the final linear discriminant subspace $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K) \in \mathbb{R}^{d \times K}$ from the proposed Pareto-optimal discriminant analysis.

First, we initialize the mixing coefficients $\alpha_i^{(0)} = 1$. Since $\mathbf{S}_i$ and $\mathbf{Z}_i$ depends only on the data $\mathbf{X}$, they are used to initialize $\mathbf{w}^{(0)}$ according to Equation (7).

Second, we can obtain the value for $u_i^{(t)} = \mathrm{tr}(\mathbf{w}^\top \mathbf{S}_i \mathbf{w})$, $v_i^{(t)} = \mathrm{tr}(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})$, and also $J^{(t)}(\mathbf{w})$ from Equation (7).

Next, we want to update the mixing coefficients $\alpha_i^{(t+1)}$. Equation (11) can be written in matrix form as follows:

$$\begin{aligned} &\underbrace{\begin{pmatrix} \cdots \\ u_i^{\alpha_i^{(t)} - 1} \mathbf{S}_i \prod_{j \neq i} u_j \mathbf{w} \\ \cdots \end{pmatrix}^\top}_{\mathbf{D}_S} \boldsymbol{\alpha}^{(t+1)} \\ &= J(\mathbf{w}) \underbrace{\begin{pmatrix} \cdots \\ v_i^{\alpha_i^{(t)} - 1} \mathbf{Z}_i \prod_{j \neq i} v_j \mathbf{w} \\ \cdots \end{pmatrix}^\top}_{\mathbf{D}_Z} \boldsymbol{\alpha}^{(t+1)} \end{aligned} \tag{12}$$

$$\Rightarrow \underbrace{\mathbf{D}_S}_{d\times N} \cdot \underbrace{\boldsymbol{\alpha}}_{N\times 1} = J(\mathbf{w}) \cdot \underbrace{\mathbf{D}_Z}_{d\times N} \cdot \underbrace{\boldsymbol{\alpha}}_{N\times 1} \tag{13}$$

At this step, we update $\boldsymbol{\alpha}$ using Equation (13), which is an alternative way of forming Equation (7). Here, the $\boldsymbol{\alpha}$ are found by solving similar generalized eigenvalue problem as denoted in Equation (13). We only care about one single vector $\boldsymbol{\alpha}$ that corresponds to the largest $J(\mathbf{w})$, this step can either be solved using eigenvalue problem. Note that the α_i's on the exponential of u_i and v_i are taking values from previous iteration.

With the updated $\alpha_i^{(t+1)}$, eigen-decomposition is applied on Equation (11) to quickly find the new $\mathbf{w}_i^{(t+1)}$, and the new linear subspace $\mathbf{W}^{(t+1)}$ spanned by $\mathbf{w}_i^{(t+1)}$. We then evaluate each of the $\mathbf{w}_i^{(t+1)}$ using Equation (7) and store K vectors that correspond to the K highest $J(\mathbf{w}_i^{(t+1)})$. The new $J^{(t+1)}(\mathbf{w})$ is then obtained by: $J^{(t+1)}(\mathbf{w}) = \max_i J(\mathbf{w}_i^{(t+1)})$ and the corresponding $\mathbf{w}$ is denoted as $\mathbf{w}^{(t+1)}$ which is used to start the second iteration along with $J^{(t+1)}(\mathbf{w})$.

The updating process is stopped after certain number of iterations T. At each iteration, $\mathbf{W}$ is updated by keeping best vectors, leading to high value in $J^{(t+1)}(\mathbf{w})$, from all previous iterations. Eventually, we can obtain the optimal linear subspace $\mathbf{W}$ whose projection direction vectors $\mathbf{w}$ maximize the objective $J(\mathbf{w})$ in Equation (3) of the proposed Pareto-optimal discriminant analysis.

This alternative coefficients updating method is suited for this non-convex optimization problem where the solution might stuck in local minimum. Because $\mathbf{w}$ are $\boldsymbol{\alpha}$ are updated alternatively, whenever $\mathbf{w}$ is stuck, $\boldsymbol{\alpha}$ will bring $\mathbf{w}$ out of the valley. It is worth noted that due to the non-convex nature of the proposed updating scheme (involving solving eigen-decomposition), the algorithm does not converge. However, we only care about the updated $\mathbf{W}$ matrix after certain iterations T. In practice, the algorithm reaches fairly informative and discriminative $\mathbf{W}$ matrix after 50 iterations. This algorithm is listed in Algorithm 1.

Algorithm 1: Alternative updating of coefficients

input : $\{\mathbf{S}_i\}_{i=1}^N, \{\mathbf{Z}_i\}_{i=1}^N, K, T$
output: $\mathbf{W} = \{\mathbf{w}_i\}_{i=1}^K$

```
Initialize α_i^(0) to be all 1 ;
Compute u_i^(0) and v_i^(0) using α_i^(0) ;
Compute w^(0) and J(w)^(0) using Equation (7) ;
for t = 0 : T do alternative updating of α and w
    Optimize Equation (13) to obtain α_i^(t+1) ;
    Optimize Equation (11) to obtain top K optimal w_i^(t+1)
    corresponding to K highest J(w_i^(t+1)) using Equation (7)
    ;
    Obtain new J^(t+1)(w) = max_i J(w_i^(t+1)) ;
    Update W by keeping best w's corresponding to highest
    J^(t+1)(w) ;
    Update u_i^(t+1) and v_i^(t+1) ;
```

2.4. Sparse Coding

The choice of K in updating the $\mathbf{W} \in \mathbb{R}^{d\times K}$ matrix is critical in Pareto-optimal discriminant analysis. If $K < d$, then $\mathbf{W}$ becomes a standard linear transformation matrix that maps the high dimension data sample $\mathbf{x} \in \mathbb{R}^d$ to a lower dimension representation $\mathbf{y} \in \mathbb{R}^K$ via $\mathbf{y} = \mathbf{W}^\top \mathbf{x}$, just like in the traditional PCA, LDA, LPP, and UDP method. However, if $K > d$, matrix $\mathbf{W}$ becomes overcomplete, and the data sample $\mathbf{x}$ of dimension d can be represented by the overcomplete matrix $\mathbf{W}$ and an even higher dimension representation $\mathbf{y}$ as follows: $\mathbf{x} = \mathbf{W}\mathbf{y}$. Since this is an underdetermined system, there are infinite number of solutions for the choice of the coefficient vector $\mathbf{y}$. However, if we enforce sparsity on $\mathbf{y}$, we can greedily obtain a unique sparse coefficient vector $\mathbf{y}$ of dimension K. This can be done by any pursuit algorithm such as the orthogonal matching pursuit (OMP) [10], which solves the following optimization problem: $\text{minimize}_y \|\mathbf{x} - \mathbf{W}\mathbf{y}\|_2^2 \ \text{subject to} \ \|\mathbf{y}\|_0 \leq \kappa$, where the sparsity level is captured by parameter κ. In our experiments, we find that an overcomplete formulation for matrix $\mathbf{W}$ yields better recognition performance. In the following experiments, we choose $K = 4d$ as a result of cross-validation on a small validation set.

3. DATABASES

The first database we use is a large-scale NIST's **FRGC ver 2.0 Database** [11] which has three components: the generic *training* set, the *target* set and the *probe* set. FRGC Experiment 4 is the hardest experiment in the FRGC protocol where face images captured in *controlled* indoor setting are matched against *uncontrolled* outdoor conditions, where harsh lighting conditions considerably alter the appearance of the face image. The second database we use is the **YaleB and Extended YaleB Database**. This database contains the most harsh illumination conditions and is well known for illumination invariant face recognition. We want to test the performance of the proposed PDA compared to other subspace learning methods under the scenario of recognizing faces under various lighting changes. The YaleB [12] contains 5,760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). The extended YaleB [13] contains 16,128 images of 28 human subjects under 9 poses and 64 illumination conditions. We combine both databases, with $10 + 28 = 38$ unique subjects in total. We only choose the frontal image with all the illumination variations. So there are 64 images for each subjects. So the entire database contains $38 \times 64 = 2,432$ images.

These two databases are among the most widely used databases in face recognition tasks [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] because they cover both unconstrained face recognition and illumination invariant face recognition.

Table 1. Performance on FRGC Experiment 4 Evaluation.

	VR at 0.001 FAR	*EER*	*Rank1 ID Rate*
PCA	0.164	0.209	0.733
LDA	0.425	0.115	0.794
UDP	0.285	0.143	0.829
LPP	0.102	0.366	0.627
PDA	**0.676**	**0.068**	**0.918**

Table 2. Performance on YaleB Evaluation.

	VR at 0.001 FAR	*EER*
PCA	0.107	0.379
LDA	0.831	0.042
UDP	0.172	0.375
LPP	0.196	0.379
PDA	**0.957**	**0.011**

4. EXPERIMENTS

Experimental Setup: The normalized cosine distance (NCD) is adopted to compute similarity matrix between target and probe images. The performance is analyzed using verification rate (VR) at $\mathbf{0.001}$ (0.1%) false accept rate (FAR), equal error rate (EER), rank-1 identification rate, and the receiver operating characteristic (ROC) curves.

For FRGC Database: We strictly follow NIST's FRGC Experiment 4 protocol which involves 1-to-1 matching of $8,014$ uncontrolled probe images to $16,028$ controlled target images ($\sim$ 128 million pair-wise face match comparisons). The similarity matrix is of size $8,014 \times 16,028$.

For YaleB Database: We randomly select 32 images per subject for training purpose, and the remaining 32 images for testing in a 1-to-1 matching fashion. Therefore, in the testing stage, there are totally $1,216$ images from 38 classes. The similarity matrix is of size $1,216 \times 1,216$.

Experimental Results: *For FRGC Database:* Table 1 shows the VR at 0.1% FAR, EER, and Rank-1 ID rate. Fig. 1 shows the ROC curves. From the results, the proposed PDA method significantly outperforms PCA, LDA, UDP, and LPP. The main reason for this is: PDA is optimally designed to harness the good properties across all the aforementioned subspace learning methods. The proposed PDA achieves a high 67.6% VR at 0.1% FAR and 91.8% rank-1 identification rate.

For YaleB Database: Table 2 shows the VR at 0.1% FAR and EER for YaleB evaluation and Fig. 1 shows the ROC curves. From the results, the proposed PDA method significantly outperforms PCA, LDA, UDP, and LPP. When lighting conditions changes, LDA learns a subspace that is excellent for performing illumination invariant face recognition. Our PDA has achieved even higher results (95.7% VR at 0.1% FAR), a significant improvement over other subspace learning methods.

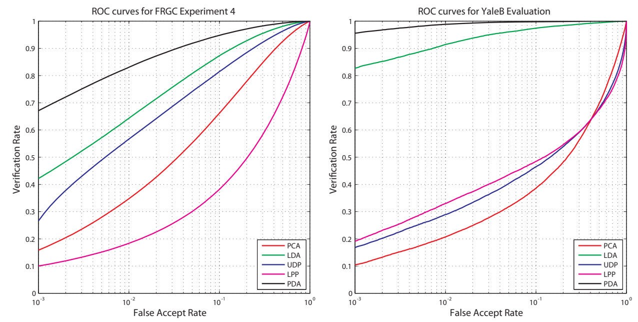

Fig. 1. ROC curves for FRGC Experiment 4 evaluation (left) and YaleB evaluation (right). Please zoom in for details.

5. CONCLUSION

In this work, we have proposed the Pareto-optimal discriminant analysis (PDA), an optimally designed linear subspace learning method that harnesses advantages across many well-known methods such as PCA, LDA, UDP and LPP. By optimizing over the joint objective function and carrying out an alternative coefficients updating scheme, we are able to obtain a linear subspace which is optimized to truly maximize the objective function in discriminant analysis. The proposed method also provides flexibility for formulating the linear transformation matrix in an overcomplete fashion, allowing for a sparse representation. We have shown, in the context of large scale unconstrained face recognition and illumination invariant face recognition, that our proposed PDA significantly outperforms other linear subspace learning methods.

6. REFERENCES

[1] A. K. Jain, R. P. W.Duin, and Jianchang Mao, "Statistical pattern recognition: a review," *TPAMI*, vol. 22, no. 1, pp. 4–37, 2000.

[2] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *CVPR*, Jun 1991, pp. 586–591.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *TPAMI*, vol. 19, no. 7, pp. 711–720, July.

[4] Jian Yang, D. Zhang, Zhong Jin, and Jing-Yu Yang, "Unsupervised discriminant projection analysis for feature extraction," in *ICPR*, 2006, vol. 1, pp. 904–907.

[5] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 22, pp. 2319–2323, Dec 2000.

[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 22, pp. 2323–2326, Dec 2000.

[7] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *NIPS*, Jun 2001.

[8] Xiaofei He and P. Niyogi, "Locality preserving projections," in *NIPS*, 2003.

[9] Wikipedia, "Pareto Optimality," `http://en.wikipedia.org/wiki/Pareto_efficiency`.

[10] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Application to Wavelet Decomposition," in *Asilomar Conf. on Signals, Systems and Comput.*, 1993.

[11] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, "Overview of the face recognition grand challenge," in *CVPR*, jun 2005, vol. 1, pp. 947–954.

[12] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *TPAMI*, vol. 23, no. 6, pp. 643–660, jun 2001.

[13] Kuang-Chih Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *TPAMI*, vol. 27, no. 5, pp. 684–698, may 2005.

[14] F. Juefei-Xu and M. Savvides, "Single Face Image Super-Resolution via Solo Dictionary Learning," in *ICIP*, Sept 2015.

[15] F. Juefei-Xu and M. Savvides, "Encoding and Decoding Local Binary Patterns for Harsh Face Illumination Normalization," in *ICIP*, Sept 2015.

[16] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K. Pal, C. Bhagavatula, and M. Savvides, "Investigating the Feasibility of Image-Based Nose Biometrics," in *ICIP*, Sept 2015.

[17] F. Juefei-Xu, D. K. Pal, and M. Savvides, "NIR-VIS Heterogeneous Face Recognition via Cross-Spectral Joint Dictionary Learning and Reconstruction," in *CVPRW*, June 2015.

[18] F. Juefei-Xu, D. K. Pal, K. Singh, and M. Savvides, "A Preliminary Investigation on the Sensitivity of COTS Face Recognition Systems to Forensic Analyst-style Face Processing for Occlusions," in *CVPRW*, June 2015.

[19] K. Seshadri, F. Juefei-Xu, D. K. Pal, and M. Savvides, "Driver Cell Phone Usage Detection on Strategic Highway Research Program (SHRP2) Face View Videos," in *CVPRW*, June 2015.

[20] F. Juefei-Xu and M. Savvides, "Image Matching Using Subspace-Based Discrete Transform Encoded Local Binary Patterns," Sept. 2013, US Patent US 2014/0212044 A1.

[21] F. Juefei-Xu and M. Savvides, "Facial Ethnic Appearance Synthesis," in *ECCVW*, 2015, pp. 825–840.

[22] F. Juefei-Xu and M. Savvides, "Weight-Optimal Local Binary Patterns," in *ECCVW*, 2015, pp. 148–159.

[23] F. Juefei-Xu, Dipan K. Pal, and M. Savvides, "Hallucinating the Full Face from the Periocular Region via Dimensionally Weighted K-SVD," in *CVPRW*, June 2014.

[24] F. Juefei-Xu and M. Savvides, "Subspace Based Discrete Transform Encoded Local Binary Patterns Representations for Robust Periocular Matching on NIST's Face Recognition Grand Challenge," *TIP*, vol. 23, no. 8, pp. 3490–3505, aug 2014.

[25] F. Juefei-Xu and M. Savvides, "An Image Statistics Approach towards Efficient and Robust Refinement for Landmarks on Facial Boundary," in *BTAS*, Sept 2013, pp. 1–8.

[26] F. Juefei-Xu and M. Savvides, "An Augmented Linear Discriminant Analysis Approach for Identifying Identical Twins with the Aid of Facial Asymmetry Features," in *CVPRW*, June 2013, pp. 56–63.

[27] F. Juefei-Xu and M. Savvides, "Unconstrained Periocular Biometric Acquisition and Recognition Using COTS PTZ Camera for Uncooperative and Non-cooperative Subjects," in *WACV*, Jan 2012, pp. 201–208.

[28] F. Juefei-Xu, Khoa Luu, M. Savvides, T.D. Bui, and C.Y. Suen, "Investigating Age Invariant Face Recognition Based on Periocular Biometrics," in *IJCB*, Oct 2011, pp. 1–7.

[29] F. Juefei-Xu and M. Savvides, "Can Your Eyebrows Tell Me Who You Are?," in *ICSPCS*, Dec 2011, pp. 1–8.

[30] F. Juefei-Xu, M. Cha, M. Savvides, S. Bedros, and J. Trojanova, "Robust Periocular Biometric Recognition Using Multi-level Fusion of Various Local Feature Extraction Techniques," in *DSP*, 2011.

[31] F. Juefei-Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust Local Binary Pattern Feature Sets for Periocular Biometric Identification," in *BTAS*, sep 2010, pp. 1–8.