IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Designing Algorithm for Excellent Classification Rate for Face Recognition Applications Using Exponential Local Discriminant Embedding

[1] Poonam Tulsiram Bawankar; [2] Hemlata Dakhore

[1] M.E. (W.C.C.) IVth sem, CSE Deptt., G. H. Raisoni Institute of Engineering and Technology, RTMNU, Nagpur, Maharashtra, India-440028

[2]Assistant Professor, CSE department, G. H. Raisoni Institute of Engineering & Technology, RTMNU, Nagpur, Maharashtra, India-440028

Abstract - In this paper, we introduce a novel discriminant technique called "exponential LDE" (ELDE) for dimensionality reduction in Face recognition. The proposed ELDE can be seen as an extension of LDE framework in two directions [1]. First, the proposed framework overcomes the SSS problem without discarding the discriminant information that was contained in the null space of the locality preserving scatter matrices associated with LDE. Second, the proposed ELDE is equivalent to transforming original data into a new space by distance diffusion mapping (similar to kernel-based nonlinear mapping), and then, LDE is applied in such a new space. As a result of diffusion mapping, the margin between samples belonging to different classes is enlarged, which is helpful in improving classification accuracy. Identifying faces with facial expressions is also a challenging task, due to the deformation caused by the facial expressions [3]. To solve these issues, a preprocessing step was carried out after which Blur and Illumination-Robust Face recognition algorithm was performed. The test image and training images with facial expression are transformed to neutral face using Facial expression removal (FER) operation. Every training image is transformed based on the optimal Transformation Spread Function (TSF), and illumination coefficients. Local BinaryPattern (LBP) features extracted from test image and transformed training image is used for classification.

***Keywords** - Discriminant analysis, face recognition, featureextraction, graph-based embedding, local discriminant embedding (LDE), small-sample-size (SSS) problem*

1. Introduction

Face recognition is one of the renowned research areas in pattern recognition and computer vision considering its numerous practical uses in the area of biometrics, information security, access control and surveillance system. Recognizing a person or friend is an easy task for human beings. One can easily recognize a person from his biometric characteristics but in computer vision recognizing a person is one of the most challenging tasks. In most of the computer vision and pattern recognition problems,the large number of sensory inputs, such as images andvideos, is computationally challenging to analyze. In suchcases, it is desirable to reduce the dimensionality of the datawhile preserving the original information in the data distribution,allowing for more efficient learning and inference [1][2][5][6].

There are two main reasons for estimating a low-dimensional representation of high-dimensional data: reducing measurementcost of further data analysis and beating the curse of dimensionality. The dimensionality reduction can be achieved either by feature extraction or by feature selection. Feature extraction refers to methods that create a set of new features based on transformations and/or combinations of the original features, while feature selection methods select the most representative and relevant subset from the original feature set [7]. Feature extraction methods can be classified into two main classes: 1) linear methods and 2) nonlinear methods.

The nonlinear methods such as locally linear embedding (LLE) [8] and Laplacian eigenmaps (LE) [9] focus on preserving the local structures. Isomap [10] is a nonlinear projection method that globally preserves the data. It also attempts to preserve the geodesic distances between samples. The linear techniques have been increasingly important in pattern recognition [5], [11], [13] since they permit a relatively simple mapping of data onto a lower dimensional subspace, leading to simple and computationally efficient classification strategies. The classical linear embedding methods (e.g., principal component analysis (PCA), linear discriminant analysis (LDA), maximum margin criterion [14]), and locally LDA [15] are demonstrated to be computationally efficient and suitable for practical applications, such as pattern classification and visual recognition.PCA projects the samples along the directions of maximal variances and aims to preserve the Euclidean distances between the samples. Unlike PCA which is unsupervised, LDA [16] is a supervised technique. One limitation of PCA and LDA is that they only see the linear global Euclidean structure.

In [17], the authors assessed the performance of the quotient and difference criteria used in LDA. They also proposed a unified criterion that combines quotient-LDA and difference-LDA criteria. In [18], the authors propose to overcome all the limitations of LDA by formulating the linear feature extraction as a projection pursuit procedure,which selects from a large set of candidate projections the most discriminatory features by using AdaBoost with classifiability based criterion for multiclass case. The candidate projectionsare heuristically generated as the difference vectors between thenearest between-class boundary samples.nearest between-class boundary samples.In addition to the LDA technique and its variants [16], [19],[21], there have been recently a lot of interests in graph-based linear dimensionality reduction. Many dimensionality reductiontechniques can be derived from a graph whose nodes representthe data samples and whose edges quantify the similarityamong pairs of samples [18], [19]. Recent proposed methodsattempt to linearize some nonlinear embedding techniques. This linearization is obtained by forcing the mapping to be explicit, i.e., performing the mapping by a projection matrix.For example, locality preserving projections (LPP)and neighborhood preserving embedding [21] can be seen as linearized versions of LE and LLE, respectively. The mainadvantage of the linearized embedding techniques is that themapping is defined everywhere in the original space.

Someresearchers tried to remedy to the global nature of the linear methods (e.g., PCA, LDA, and LPP) by proposing localizedmodels [24]. In this paper, localized PCA, LDA, or LPP modelsare built using the neighbors of a query sample. The authorshave shown that the obtained localized linear models can outperformthe global models for face recognition and coarse headpose estimation problems. In [25], the authors have extendedthe LPP to the supervised case by adapting the entries of thesimilarity matrix according to the labels of the sample pair. In[26], the authors have proposed an enhanced supervised variantof LPP. The affinity matrix weights are modified in order totake into account label information as well as the similaritiesbetween pairs of samples. Moreover, the optimized criterionintegrates the uncorrelated and orthogonal constraints

Vageeswaran, Mitra, and Chellappa (2013), provoked by the problem of remote face recognition, the issue ofidentifying blurred and poorly illuminated faces is addressed.The set of all images obtained by blurring a given image is aconvex set given by the convex hull of shifted versions of theimage. Depending on this set-theoretic characterization, ablur-robust face recognition algorithm DRBF is suggested. Inthis algorithm we can easily incorporate existing knowledgeon the type of blur as constraints. Taking the low-dimensionallinear subspace model for illumination, the set of all imagesobtained from a given image is then shown by blurring andchanging its illumination conditions is a bi-convex set. Again,based on this set-theoretic characterization, a blur andillumination robust algorithm IRBF is suggested. Combining adiscriminative learning based approach like SVM would be avery promising direction for future work.

Patel, Wu, Biswas, Phillips, and Chellappa (2012)proposed a face recognition algorithm based on dictionarylearning methods that are robust to changes in lighting andpose is proposed. This entails using a relighting approachbased on a robust albedo estimation. Different experiments onpopular face recognition datasets have shown that the methodis efficient and can perform importantly better than manycompetitive face recognition algorithms. Learningdiscriminative dictionaries is that it can tremendously increasethe overall computational complexity which can make thereal-time processing very difficult. Discriminative methods aresensitive to noise. It is an interesting topic for future work todevelop and investigate the correctness of discriminativedictionary learning algorithm. This algorithm is potent to pose,expression and illumination variations.

The computation time is greatly reduced in our system byusing a standard neutral face image in the optical flow computation and warping process. The constrained opticalflow warping algorithm significantly improves the recognitionrate for face recognition from a single expressive face imagewhen only one training neutral image for each subject isavailable. Local discriminant embedding (LDE) is known as apowerful tool for discriminant analysis that is proposed toovercome some

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

limitations of the global LDA method. It extendsthe concept of LDA to perform local discrimination. Inthe case of a small training data set, however, LDE cannotdirectly be applied to high-dimensional data. This case is theso-called small-sample-size (SSS) problem which occurs whenthe number of samples is less than the feature dimension. TheSSS problem very often occurs when dealing with visual objectrecognition tasks includingthe face recognition problem. Theclassical solution to this problem was applying dimensionalityreduction on the raw data (e.g., using PCA).

1.1 System Design

A. Problem Definition

Local discriminant embedding (LDE) has been recently proposed to overcome some limitations of the global linear discriminant analysis method. In the case of a small training dataset, however, LDE cannot directly be applied to high –dimensional data. This case is the so-called small-sample-size (SSS) problem. The classical solution to this problem was applying dimensionality reduction on the raw data (e.g., using principal component analysis).

As well, the scheduled work systematically addresses face recognition under non-uniform motion blur and the merged effects of blur, illumination, and expression. For this, a preprocessing step for expression removal step was carried outafter which Blur and Illumination-Robust Face Recognition algorithm.

B. proposed Solution

In this paper, we introduce a novel discriminant techniquecalled "exponential LDE" (ELDE). The proposed ELDE can beseen as an extension of LDE framework in two directions. First,the proposed framework overcomes the SSS problem without discarding the discriminant information that was containedin the null space of the locality preserving scatter matricesassociated with LDE. Second, the proposed ELDE is equivalentto transforming original data into a new space by distance diffusion mapping (similar to kernel-based nonlinear mapping),and then, LDE is applied in such a new space. As a resultof diffusion mapping, the margin between samples belongingto different classes is enlarged, which is helpful in improvingclassification accuracy.

The use of matrix exponential for data embedding was usedin two recent works [23], [33]. In [23], the authors proposeexponential discriminant analysis (EDA) method that uses theexponential of the global within- and between-class scatter matrices.In [33], the authors propose exponential LPPs (ELPPs).The EDA method can solve the SSS problem, but it still inheritsthe global nature of LDA in the sense that it ignores the localstructures of data. Thus, the performance of EDA may not beoptimal. The proposed ELPP also solves the SSS problem,but it is an unsupervised technique that does not exploit thelabel information. The resulting learning based on ELPP cannotproperly model the underlying structure and characteristics ofdifferent classes. In contrast, our proposed ELDE is built upon a local discriminant technique and, therefore, able to overcomeall limitations of EDA, ELPP, and LDEThe use of kernels in the frameworks of machine learning [34] and linear embedding techniques was first introduced inthe 1990s. Kernel PCA (KPCA) was originally developed by Schölkopf in 1998 [35], while kernel Fisher discriminant (KFDanalysis (a kernelized version of the classical LDA) was firstproposed by Mika et al. in 1999 [36]. Subsequent research sawthe development of a series of KFD algorithms. Baudat and Anouar [37] extend the original KFD to deal with a multiclassclassification problem.

Lu et al. [17] generalized direct LDA(DLDA) [38] using the idea of kernels and presented kernel directdiscriminant analysis (KDDA). Their method was demonstrated effective for face recognition, but, as a nonlinear versionof DLDA, KDDA unavoidably suffers the weakness of DLDAin the sense that it overlooked the regular information providedby the nonnull space of the within-class scatter matrix [38]whenever the SSS problem occurs. Yang et al. [39] proposed a complete KFD (CKFD) method. The implemented methodwas intended to perform discriminant analysis in "double discriminantsubspaces": regular and irregular associated with thewithin-class scatter matrix after KPCA projection. Because of its ability to extract the most discriminatorynonlinear features, KFD and its variants have been found tobe very effective in many real-world applications. Althoughthe kernelized versions can give better results than the lineardiscriminant methods, the selection of the kernel type is stillan open problem. Furthermore, some KFD methods have manyparameters that should be tuned in advance (e.g., the CKFDmethod [39]).

In a face Recognition method, for a set of train images gcs and a test image, p, the identity of the test image is to be found. The test image may be blurred, illuminated, along with varied expressions. In testimage, p,

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

expressions are neutralized using Facial Expressionremoval (FER). The matrix Ac for each training (gallery) faceis generated. The test image, p can be stated as the convex consolidation of the columns of one of these matrices. Forrecognition task, the optimal TSF and illumination coefficients [Tc, αc, i] for each training image are computed [2]. Usingfacial expression removal, expressions are removed for transformed image.

2. System Architecture

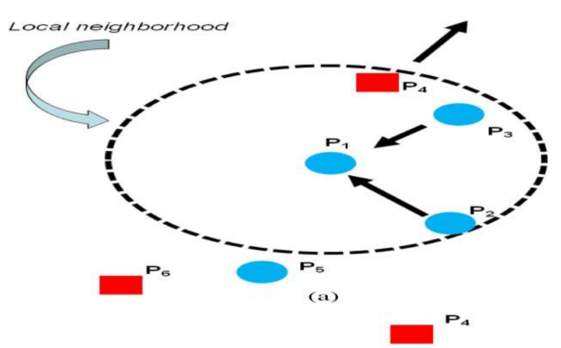
This paper is structured as follows. In Section I, we provide a review of the LDE method. In Section II, we describe ourproposed ELDE method. Section III provides a theoretical analysis of the proposed algorithm. we provide some concluding remarks in Section IV. Finally, The experimental results are presented in Section V.

I. REVIEW OF LDE

In order to make the paper self-contained and to present the main notations used in this paper, this section presents the LDE method [32]. It should be noticed that the marginal Fisher analysis method [19] and the LDE method are essentially the same. In [40], the authors proposed another variant of LDE by using a difference criterion instead of a quotient. Fu *et al*. [41] propose a variant of LDE in which the similarity between samples is depending on the relative angle instead of the Euclidean distance. In the sequel, capital bold letters denote matrices, and small bold letters denote vectors.

A. Intrinsic Graph and Penalty Graph

The objective of LDE is to estimate a linear mapping that simultaneously maximizes the local margin between heterogeneous samples and pushes the homogeneous samples closer to each other. The expected effect of LDE framework on data is shown in Fig. 1.



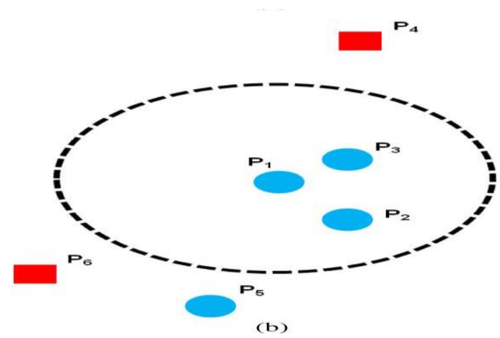


Fig. 1. (a) Center point has three neighbors. The points with the same color and shape belong to the same class. The within-class graph connects nearby points with the same label. The between-class graph connects nearby points with different labels. (b) After LDE, the local margins between different classes are maximized, and the distances between local homogeneous samples are minimized. (a) Original space. (b) Expected mapped space using LDE.

We assume that we have a set of *N*labeled samples *{***x***i}Ni*=1⊂R*D*. In order to discover both geometrical and discriminant structures of the data manifold, two graphs are built: the within-class graph *Gw*(intrinsic graph) and the between class graph *Gb* (penalty). Let *l*(**x***i*) be the class label of **x***i*. For each data sample **x***i*, two subsets *Nw*(**x***i*) and *Nb*(**x***i*) are computed. *Nw*(**x***i*) contains the neighbors sharing the same label with **x***i*, while *Nb*(**x***i*) contains the neighbors having different labels. One simple possible way to compute these two sets of neighbors associated with the local sample is the use of two nearest neighbor graphs: one nearest neighbor graph for homogeneous samples (parameterized by *K*1) and one nearest neighbor graph for heterogeneous samples (parameterized by *K*2). Note that *K*1 and *K*2 can be different and chosen with empirical values. Each of the graphs mentioned before, *Gw* and *Gb*, is characterized by its corresponding affinity (weight) matrices **W***w* and **W***b*, respectively. The entries of these symmetric matrices are defined by the following:

$$W_{w,ij} = \begin{cases} \text{sim}(\mathrm{x}_i, \mathrm{x}_j) & \text{if } \mathrm{x}_j \in N_w(\mathrm{x}_i) \text{ or } \mathrm{x}_i \in N_w(\mathrm{x}_j) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$W_{b,ij} = \begin{cases} 1 & \text{if } \mathrm{x}_j \in N_b(\mathrm{x}_i) \text{ or } \mathrm{x}_i \in N_b(\mathrm{x}_j) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Where, sim(**x***i*, **x***k*) is a real value that encodes the similarity between **x***i* and **x***k*. Without loss of generality, we assume that sim(**x***i*, **x***k*) belongs to the interval [0, 1]. Simple choices for this function are the kernel heat and the cosine.

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

B. Optimal Mapping

A linear embedding technique is described by a matrix transform that maps the original samples **x***i* into low-dimensional samples **A**T**x***i*. The number of columns of **A** defines the dimension of the new subspace. LDE method computes a linear transform **A** that simultaneously maximizes the local margins between heterogeneous samples and pushes the homogeneous samples closer to each other (after the transformation). Mathematically, This corresponds to :

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i,j} \left\| \mathbf{A}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j) \right\|^2 W_{w,ij} \qquad (3)$$

$$\max_{\mathbf{A}} \frac{1}{2} \sum_{i,j} \left\| \mathbf{A}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j) \right\|^2 W_{b,ij}. \qquad (4)$$

Using simple matrix algebra, the aforementioned criteria become respectively

$$J_{\text{homo}} = \frac{1}{2} \sum_{i,j} \left\| \mathbf{A}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j) \right\|^2 W_{w,ij} \qquad (5)$$

$$= \mathrm{tr}\left\{ \mathbf{A}^{\mathrm{T}} \mathbf{X} (\mathbf{D}_w - \mathbf{W}_w) \mathbf{X}^{\mathrm{T}} \mathbf{A} \right\} \qquad (6)$$

$$= \mathrm{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{X} \mathbf{L}_w \mathbf{X}^{\mathrm{T}} \mathbf{A}) \qquad (7)$$

$$J_{\text{hete}} = \frac{1}{2} \sum_{i,j} \left\| \mathbf{A}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j) \right\|^2 W_{b,ij} \qquad (8)$$

$$= \mathrm{tr}\left\{ \mathbf{A}^{\mathrm{T}} \mathbf{X} (\mathbf{D}_b - \mathbf{W}_b) \mathbf{X}^{\mathrm{T}} \mathbf{A} \right\} \qquad (9)$$

$$= \mathrm{tr}\left(\mathbf{A}^{\mathrm{T}} \mathbf{X} \mathbf{L}_b \mathbf{X}^{\mathrm{T}} \mathbf{A} \right) \qquad (10)$$

where tr(S) denotes the trace of the matrix S, X = (x1, x2, . . . , xN) is the data matrix, Dw denotes the diagonal weight matrix, whose entries are column (or row, since Ww is symmetric) sums of Ww, and Lw = Dw −Ww denotes the Laplacian matrix associated with the graph Gw. Given the two individual optimization objectives, (3) and (4), a unified criterion that should be maximized can be formulated as

$$J = \frac{J_{\text{hete}}}{J_{\text{homo}}} = \frac{\mathrm{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{X} \mathbf{L}_b \mathbf{X}^{\mathrm{T}} \mathbf{A})}{\mathrm{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{X} \mathbf{L}_w \mathbf{X}^{\mathrm{T}} \mathbf{A})} = \frac{\mathrm{tr}(\mathbf{A}^{\mathrm{T}} \widetilde{\mathbf{S}}_b \mathbf{A})}{\mathrm{tr}(\mathbf{A}^{\mathrm{T}} \widetilde{\mathbf{S}}_w \mathbf{A})} \qquad (11)$$

Where, the symmetric matrix Sb = XLbXT denotes the locality preserving between-class scatter matrix, and the symmetric matrix Sw = XLwXT denotes the locality preserving within class scatter matrix. The trace ratio optimization problem (11) can be replaced by the simpler yet inexact trace form

$$\max_{\mathbf{A}} \mathrm{tr}\left\{ (\mathbf{A}^{\mathrm{T}} \widetilde{\mathbf{S}}_w \mathbf{A})^{-1} (\mathbf{A}^{\mathrm{T}} \widetilde{\mathbf{S}}_b \mathbf{A}) \right\}. \qquad (12)$$

The aforementioned optimization problem has a closed-form solution due to the quadratic form. The columns of the sought matrix A are given by the generalized eigenvectors associated with the largest eigenvalues of the following:

$$\widetilde{\mathbf{S}}_b \mathbf{a} = \lambda \widetilde{\mathbf{S}}_w \mathbf{a}. \qquad (13)$$

In [40], the authors propose a difference criterion

$$J = \alpha J_{\text{hete}} - (1 - \alpha) J_{\text{homo}} \qquad (14)$$

where $0 < \alpha < 1$ is a balance parameter. The linear transform is found by maximizing the aforementioned criterion (14) under the constraint $\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{D}w\mathbf{X}^{\mathrm{T}}\mathbf{A} = \mathbf{I}$.

C. SSS Problem

In many real-world problems such as face recognition, both matrices **XL***b***X**T and **XL***w***X**T can be singular. This stems from the fact that, sometimes, the number of images in the training set *N* is much smaller than the number of pixels in each image *D*. This is known as the SSS problem. To overcome the complication of singular matrices, original data are first projected onto a PCA subspace or a random orthogonal space so that the resulting matrices **XL***b***X**T and **XL***w***X**T are nonsingular.

D. Blur, Illumination and Expression invariant face recognition

Fig 2. shows architecture for blur,illumination and expression invariant face recognition. Asimple pre-processing technique for removing expressionsfrom test image and transformed image were done to form areconstructed face image using wavelet transform. Wavelettransform are now used to handle such variations.

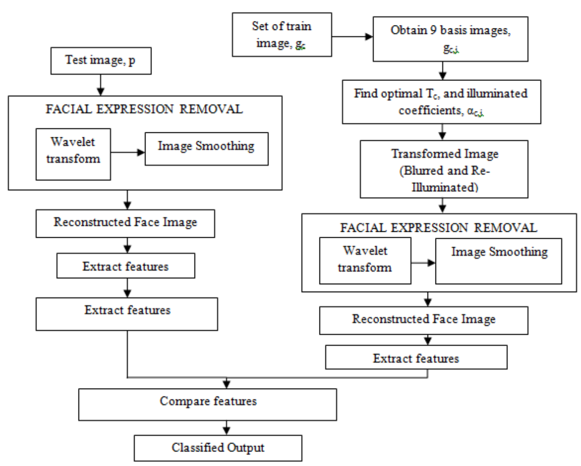


Fig 2: . Blur, Illumination and Expression invariant face recognition

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

The LBP features were used in DRBF. LBP features of bothBlurred test images and train images have been considered forthe recognition rate. The regions around the eyes are given thehighest weights [21]. In Figure 2, the white region has highest weights, the other colors have low weights and the darkerregions have very low weights [6]. The reconstruction error, rc becomes as follows when the weighting matrix has been used,

$$r_c = \min_T \| W (g - A_c T) \|_2^2 + \beta \| T \|_1, \text{ subject to } T \geq 0$$

The reconstruction error, rc has been responsive to smallpixel misalignments and hence, it is not preferable. Test image, p represents the input for which LBP features wereextracted.

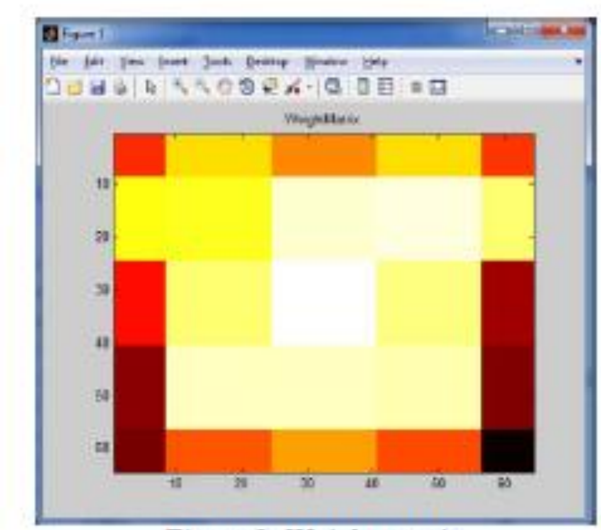


Figure 2. Weight matrix

For every train image, gc, Blur has been created through optimal Transformation Spread Function (TSF) Tc. Thisoptimal TSF was found by

$$T_c = \arg\min_T \| W (p - A_c T) \|_2^2 + \beta \| T \|_1, \text{ for } T \geq 0$$

Here,
Tc is optimal TSF
W is weighting matrix
p represents Blurred test image
Ac is matrix for each training face image
T is vector of weights

II. ELDE

As can be seen, solving the SSS problem associated withLDE relied on applying a PCA on the raw data. However, PCAeliminates the null space of the total covariance matrix of data[42]. With respect to LDE formulation, one can easily see thatthe use of PCA stage eliminates the null spaces associated withthe locality preserving scatter matrices **XL***b***X**T and **XL***w***X**T. Therefore, by using PCA as an initial stage in LDE, some discriminantinformation will not be handed over to the frameworkof LDE.

The matrix exponential is widely used in applications suchas control theory and Markov chain analysis. In this section,the definition and properties of matrix exponential are brieflyintroduced. Given an $n \times n$ square matrix **S**, its exponential isdefined as follows:

$$\exp(\mathrm{S}) = \mathrm{I} + \mathrm{S} + \frac{\mathrm{S}^2}{2!} + \cdots + \frac{\mathrm{S}^m}{m!} + \cdots \qquad (15)$$

Where, **I** is the identity matrix with the size of $n \times n$. Theproperties of matrix exponential are listed as follows.

1) exp(**S**) is a finite matrix.
2) exp(**S**) is a full-rank matrix.
3) If matrix **S** commutes with **T**, i.e., **ST** = **TS**, then exp(**S** + **T**) = exp(**S**) exp(**T**).
4) For an arbitrary square matrix **S**, there exists the inverse of exp(**S**). This is given by

$$(\exp(\mathrm{S}))^{-1} = \exp(-\mathrm{S}).$$

5) If **T** is a nonsingular matrix, then exp(**T**−1**ST**) =**T**−1 exp(**S**)**T**.
6) If **v**1, **v**2, . . . ,**v***n*are eigenvectors of **S** that correspond to the eigenvalues *λ*1*, λ*2*, . . . , λn*, then **v**1, **v**2, . . . , **v***n* are also eigenvectors of exp(**S**) that correspond to the eigenvalues *eλ*1*, eλ*2*, . . . ,eλn*. It is also well known that the matrix is nonsingular.

A wide variety of methods for computing exp(**S**) were analyzed in [43]. The scaling and squaring method is one of the best methods for computing the matrix exponential. The exponential version of LDE is obtained by using the exponential of **S***b* and **S***w*. In other words, the eigenvaluesof these matrices are replaced with their exponential. Thisreplacement will have two beneficial effects: 1) solving the SSS problem and 2) introducing a distance diffusion mapping thatwill be explained in the next section. Thus, the new criterion to be maximized becomes

$$\max_{A} \mathrm{tr} \left\{ \left(A^T \exp(\tilde{S}_w) A \right)^{-1} \left(A^T \exp(\tilde{S}_b) A \right) \right\}. \qquad (16)$$

The columns of the sought matrix **A** are given by thegeneralized eigenvectors associated with the largest eigen values of the following equation:

$$\exp(\widetilde{\mathrm{S}}_b)\mathrm{a} = \lambda \exp(\widetilde{\mathrm{S}}_w)\mathrm{a}. \qquad (17)$$

It should be noticed that LDE is based on a two-order moment ($\mathbf{S}w$and $\mathbf{S}b$), whereas ELDE involves approximately a linear combination of different moments [exp ($\mathbf{S}w$) and exp($\mathbf{S}b$)],including a two-order moment [see (15)]. The main steps of ELDE algorithm are the following.

- Step 1) Normalize the feature vectors (vectorized images) $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N$.
- Step 2) Construct the similarity matrices $\mathbf{W}_w$ and $\mathbf{W}_b$ using (1) and (2).
- Step 3) Construct the locality preserving scatter matrices $\widetilde{\mathrm{S}}_w = \mathbf{X}\mathbf{L}_w\mathbf{X}^{\mathrm{T}}$ and $\widetilde{\mathrm{S}}_b = \mathbf{X}\mathbf{L}_b\mathbf{X}^{\mathrm{T}}$. Then, the optimal projection axes $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_k$ can be selected as the generalized eigenvectors of (17). These axes correspond to the first largest k eigenvector of $\exp(\widetilde{\mathrm{S}}_b)\mathrm{a} = \lambda \exp(\widetilde{\mathrm{S}}_w)\mathrm{a}$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$.
- Step 4) Project all samples into the obtained optimal discriminant vectors.

3. Theoretical Analysis of ELDE

A. Solving the SSS Problem

Whenever the SSS problem occurs, the classic solution consists in reducing the dimensionality of original data before applying the LDE framework. In other words, LDE sets the discriminant projection axes $\mathbf{a}i$ to the generalized eigenvectors of the following generalized eigenvalue decomposition problem:

$$\mathbf{X}_r\mathbf{L}_b\mathbf{X}_r^{\mathrm{T}}\mathbf{a} = \lambda\mathbf{X}_r\mathbf{L}_w\mathbf{X}_r^{\mathrm{T}}\mathbf{a}$$

Where, $\mathbf{X}r$ is the data matrix after dimensionality reductionusing PCA. Thus, using the data $\mathbf{X}r$ instead of the original data $\mathbf{X}$ in LDE formulation removes the null spaces of thelocality preserving scatter matrices $\mathbf{XL}b\mathbf{X}\mathrm{T}$ and $\mathbf{XL}w\mathbf{X}\mathrm{T}$. These null spaces may contain discriminant information. Asa consequence, some of significant discriminatory informationmay be lost due to this preprocessing PCA step.

B. Distance Diffusion Mapping

In pattern recognition field, the kernel trick becomes a well known tool that tackles the nonlinearity of data. The basic ideais to map input data space into a feature space having higherdimension through a kernel-based nonlinear mapping $\Phi(\mathbf{x})$. Forinstance, nonlinear support vector machines rely on the kerneltrick in order to classify data that are not linearly separable inthe input space [45].Therefore, ELDE might possess some similar properties ofthe kernel method. The only difference between ELDE and thekernel method is that ELDE involves the locality preservingscatter matrices, whereas the kernel method maps the featurevectors. However, the scatter matrices are derived from thesefeature vectors. Hence, ELDE might possess the property ofthe kernel method.Similar to the kernel method, for ELDE, there exists anonlinear mapping function Ψ such that the locality preservingscatter matrices are mapped into a new space, i.e.,

$$\begin{aligned} \Psi : \mathbb{R}^{D\times D} &\longrightarrow \mathbb{R}^{D\times D} \\ \widetilde{\mathbf{S}}_b &\longrightarrow \Psi(\widetilde{\mathbf{S}}_b) = \exp(\widetilde{\mathbf{S}}_b) = \exp(\mathbf{X}\mathbf{L}_b\mathbf{X}^{\mathrm{T}}) \\ \widetilde{\mathbf{S}}_w &\longrightarrow \Psi(\widetilde{\mathbf{S}}_w) = \exp(\widetilde{\mathbf{S}}_w) = \exp(\mathbf{X}\mathbf{L}_w\mathbf{X}^{\mathrm{T}}). \end{aligned}$$

Fig. 2 shows a geometrical interpretation of the two processes induced by the proposed ELDE method.

In order to give concrete examples about the aforementioned explanation, we used five face databases: Yale, Extended Yale, PF01, PIE, and FERET. Details about these databases are presented in the next section. For the first four databases, we fixed the neighborhood size for heterogeneous and homogeneous samples to seven, i.e., $K1 = K2 = 7$. For FERET database, we set these parameters to five. Table I shows the value of the distances *db* and *dw* for each face database. As can be seen for all databases, the between-class distance is greater than the within-class distance. We stress the fact that the Extended Yaleand PIE data sets contain face images taken under extreme illumination variation in addition to other types of variation.

In consequence, for these two data sets, the images of different persons (between-class samples) are close to each other or even mixed. Since the scores *db* and *dw* are sums of distances between pairs of images in the original image space, this can explain why the between-class distance *db* is not significantly larger than the within-class distance *dw* for the Extended Yale and PIE data sets (see Table I). Fig. 3 shows the largest 20 eigenvalues of the locality preserving scatter matrices $\mathbf{S}b$ and $\mathbf{S}w$ associated with Yale database.

IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

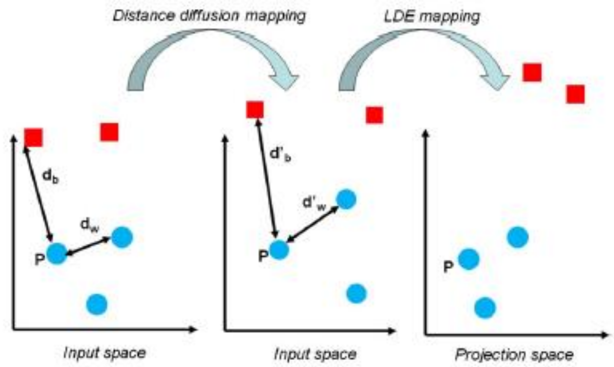


Fig. 2. Illustrative example showing the expected mapping performed by ELDE. First, distances between input samples (homogeneous and heterogeneous) are diffused in the input space. Second, the LDE algorithm is applied on these samples in order to obtain the projection.

TABLE I
FIRST ROW DEPICTS THE INTERCLASS SAMPLE DISTANCE d_b ASSOCIATED WITH FIVE FACE DATA SETS. THE SECOND ROW DEPICTS THE CORRESPONDING WITHIN-CLASS SAMPLE DISTANCE d_w

	Yale	Ext Yale	PF01	PIE	FERET
d_b	217988	466713	1104192	2356076	1916141
d_w	73889	459354	866265	2289714	1117366

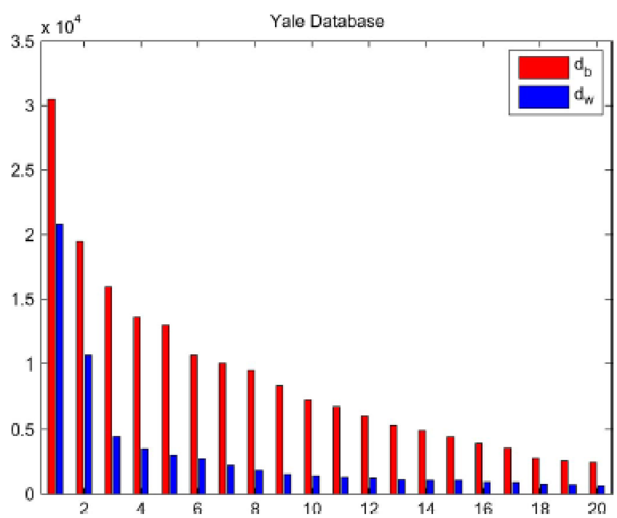


Fig. 3. First 20 eigenvalues of the locality preserving scatter matrices. The red bars correspond to the between-class locality preserving scatter matrix $\mathbf{X}\mathbf{L}_b\mathbf{X}^T$. The blue bars correspond to the within-class locality preserving scatter matrix $\mathbf{X}\mathbf{L}_w\mathbf{X}^T$.

To summarize, the use of the exponential of the localitypreserving scatter matrices has avoided the dimensionality reduction stage (PCA stage) and has implicitly introduced adistance diffusing mapping. This makes the proposed ELDEmore performant than LDE.

4. Blur and Illumination invariant face recognition

Test image, p and set of train images, gc were given as the input. Nine basis images are obtained for each training image. Using equation (4), by finding optimal TSF, Tc and illumination coefficients, αc, i, every image is transformed. LBP features are extracted for transformed image and test image. Both the LBP features are compared to find whether they almost match with each other.

$$[T_c, \alpha_{c,i}] = \operatorname{argmin}_{T,\alpha i} \| W (p - \textstyle\sum_{i=1}^{9} \alpha i\, A_{c,i} T) \|_2^2 + \beta \| T \|_1$$

subject to $T \geq 0$ (4)

Where,

$\alpha_{c,i}$, illumination coefficients
α_i, for i=1,2…9 are the linear coefficients
T_c, is optimal TSF (Transformation Spread Function)
W, weighting matrix
p, test image
A_c, matrix for each gallery face
T, vector of weights

Equation (4) is solved in two steps. The first step considers that there is no blur keeping hTm fixed and find the illumination coefficients, αc,i to form 9 basis images. These 9 basis images are relit using calculated illumination coefficients. From these images, matrix Ac is formed and Tc is solved. Using Tc, blur is created for all the basis images. This is the second step. This is done for nine iterations.

Algorithm 1: Blur, Illumination-Robust Face Recognition (BIRFR)
Input: Blurred, illuminated test image p, and a set of training images gc, c=1, 2… C.
Output: Identity of the test image.
1. For each gc,
2. Obtain nine basis images gc, i, i=1,2,….,9.
3. Find optimal TSF Tc and illumination coefficients αc, I in gc by solving equation (2).
4. Transform (blur and re-illuminate) gc using Tc and αc, i
5. Extract features of transformed gc.
6. Compare the features of p with those of the transformed gc.
7. Find the closest match of p.

Algorithm 2: Blur, Illumination and Expression-Robust
Face Recognition (BIEFR)
Input: Blurred, illuminated and expression-variated test image, p, and a set of training images gc, c=1, 2… C.
Output: Identity of the test image.

1. Obtain neutral face from p and gc for C-face classes using FER.
2. For each gc,
3. Obtain nine basis images gc, i, i=1,2,….,9.
4. For each gc,
5. Find optimal TSF Tc and illumination coefficients αc, I by solving equation (3.4).
6. Transform (blur and re-illuminate) gc using computed Tc and αc, i.
7. Extract LBP features of transformed gc.
8. Compare features of neutralized p and transformed gc.
9. Find closest match of p.

5. Performance Evaluation

In this section, we evaluate the proposed method on both synthetic and real data. For the synthetic data, we visualize the projection axes and the embedded data. For real data, we evaluate recognition accuracy for classification tasks using five public face databases.

A. Synthetic Data

We use the following three synthetic data sets.

Set 1) Two-dimensional data set (Fig. 4). This two-class set contains 600 points. Each class (300 points) is formed by a mixture of three Gaussians.

Set 2) Two-dimensional data set (Fig. 5). This three-class set contains 600 points. Each class (200 points) is formed by a mixture of two Gaussians.

Set 3) Three-dimensional data set [Fig. 6(a)]. This two-class set contains 600 points. The first class is generated from a single Gaussian with zero mean and $0.5\mathbf{I}_3$ covariance, where $\mathbf{I}_3$ is the 3×3 identity matrix. The second class is generated from a mixture of three Gaussians. The first one has 100 points with [1, 4, 0] mean and $0.5\mathbf{I}_3$ covariance, the second one has 100 points with $[2\sqrt{3}, -2, 0]$ mean and $0.5\mathbf{I}_3$ covariance, and the last one has 100 points with $[-2\sqrt{3}, 2, 0]$ mean and $0.5\mathbf{I}_3$ covariance.

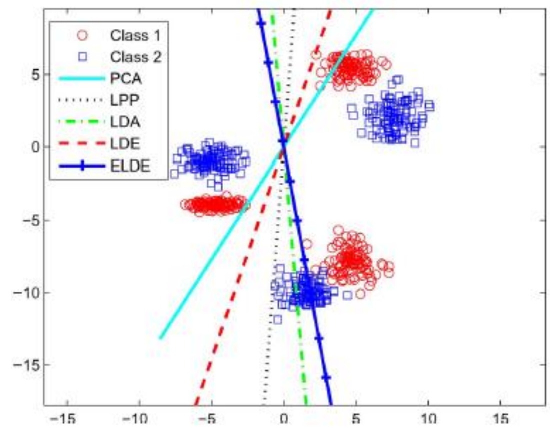


Fig. 4. Example of the multimodal data set representing two classes of 2-D samples. Each class is distributed as three separated Gaussians having different parameters. The projection direction of the proposed ELDE, together with that of four linear methods, is plotted. We can see that every method has provided a different direction according to the objective function used.

Fig. 4 shows the direction of the first projection axis (firsteigenvector) obtained by PCA, LDA, LPP, LDE, and ELDE(proposed method) when applied on Set 1. As can be seen,every embedding method has provided a different projectiondirection according to the objective function used. Fig. 5 showsthe direction of the first projection axis (first eigenvector)obtained by the five embedding methods when applied on Set 2.We can observe that the LDA method is confused in this case,while LDE and ELDE methods can still find the projectiondirection with the strongest discriminative power.

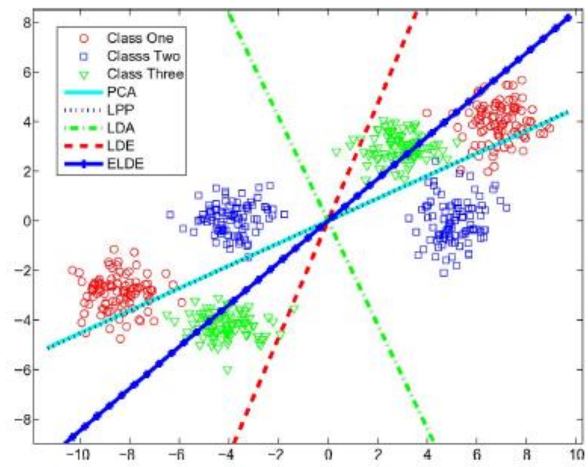


Fig. 5. Example of the multimodal data set representing three classes of 2-D samples. Each class is distributed as two separated Gaussians having different parameters. The projection direction of the proposed ELDE, together with that of four linear methods, is plotted. We can see that LDA is confused in this case, while LDE and ELDE methods can still find the projection direction with the strongest discriminative power.

Fig. 6 shows the embedding of Set 3. Fig. 6(b)–(d) shows theembedded data of Set 3 in a 2-D subspace (first two eigenvectors)using LDA, LDE, and ELDE, respectively. Fig. 6(e) showsthe embedded data using ELDE in a 3-D subspace. As can beseen, both LDE and ELDE succeeded to provide projected datahaving good discrimination. We can also appreciate that ELDEhas provided clusters that are more compact than the clustersobtained by LDE.

In brief, based on the aforementioned three examples, we canobserve the following: 1) PCA only aims to find the directionin which the data structure is maximally preserved, and theremay not exist any discrimination on such direction; 2) LDAmay confuse when the data distribution is more complicated;3) LPP can give the same direction as PCA (Fig. 5); and4) LDE and ELDE can find the discriminative directions basedon local analysis, and it does not make any assumptions on thedistributions of the data points.

6. Conclusion and Future Work

A robust face recognition system for unconstrained environment was developed using ELDE algorithm. In this algorithm, LBPfeatures were extracted for the blurred, illuminated, expressionvariated probe image. Every image in the gallery set wastransformed using optimal TSF and their LBP features wereextracted. A simple pre-processing step, FER was carried outand the reconstructed face images have been used for furtherprocessing. LBP features of transformed image and blurredprobe image were compared to find the best match.It was observed that for ELDE algorithm, when Cropped Yale was used, the recognition rate obtained was 81.986% and for Yale face dataset, 87.88%. It was observed that for ELDE used, the recognition rate noticed was 82% and for Yale facedataset, 67.996%. The system works effortlessly and is robustto conditions like blur, illumination and expressions. The results were improved when expression was removed.

References

[1] Fadi Dornaika and Alireza Bosaghzadeh," Exponential Local Discriminant Embedding and Its Application to Face Recognition", IEEE TRANSACTIONS ON CYBERNETICS, VOL. 43, NO. 3, JUNE 2013.

[2] L. Maaten, E. Postma, and J. Herik, "Dimensionality reduction: A comparative review," TiCC, Tilburg Univ., Tilburg, The Netherlands, TiCC TR 2009-005, 2009

[3] Anubha Pearline," Hemalatha. M ,"Face Recognition Under Varying Blur, Illumination and Expression in an Unconstrained Environment", Vol. 14 CIC 2016 Special Issue International Journal of Computer Science and Information Security (IJCSIS) https://sites.google.com/site/ijcsis/ ISSN 1947-5500

[4] Abhijith Punnappurath and Ambasamudram Narayanan Rajagopalan, "Face Recognition Across Non-Uniform Motion Blur, Illumination, and Pose", IEEE Transactions On Image Processing, Vol. 13, No. 7, pp. 2067- 2082, Jul. 2015.

[5] L. Saul, K. Weinberger, F. Sha, J. Ham, and D. Lee, "Spectral methods for dimensionality reduction," in *Semisupervised Learning*. Cambridge, MA: MIT Press, 2006.

[6] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extension: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[7] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[8] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[9] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[10] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[12] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.

[13] A. M. Martinez and M. Zhu, "Where are linear feature extraction methods applicable?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1934–1944, Dec. 2005

[14] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 157–165, Jan. 2006.

[15] T. Kim and J. Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 318– 327, Mar. 2005.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1990.

[17] Y. Tao and J. Yang, "Quotient vs. difference: Comparison between the two discriminant criteria," *Neurocomputing*, vol. 18, no. 10–12, pp. 1808– 1817, Jun. 2010.

[18] Y. Su, S. Shan, X. Chen, and W. Gao, "Classifiability-based discriminatory projection pursuit," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2050–2061, Dec. 2011.

[16] D.-Q. Dai and P. C. Yuen, "Face recognition by regularized discriminant analysis," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 4, pp. 1080–1085, Aug. 2007.

[17] J. Lu, K. Plataniotis, and A. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 117–126, Jan. 2003.

[18] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1027–1061, May 2007.

[19] S. Yan, D. Xu, B. Zhang, and H.-J. Zhang, "Graph embedding: A general framework for dimensionality reduction," in *Proc. Int. Conf. Comput. Vis Pattern Recognit.*, 2005, pp. 830–837.

[20] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Conf.Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2003.

21] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[22] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognit.*, vol. 43, no. 12, pp. 4165–4176, Dec. 2010.

[23] T. Zhang, B. Fang, Y. Tang, Z. Shang, and B. Xu, "Generalized discriminant analysis: A matrix exponential approach," *IEEE Trans. Syst., Man,Cybern. B, Cybern.*, vol. 40, no. 1, pp. 186–197, Feb. 2010. [24] Y. Fu, Z. Li, J. Yuan, Y. Wu, and T. S. Huang, "Locality versus globality: Query-driven localized linear models for facial image computing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 12, pp. 1741–1752, Dec. 2008.

[25] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006.

[26] W. Wong and H. Zhao, "Supervised optimal locality preserving projection," *Pattern Recognit.*, vol. 45, no. 1, pp. 186–197, Jan. 2012.

[27] F. Wang, X. Wang, D. Zhang, C. Zhang, and T. Li, "Marginface: A novel face recognition method by average neighborhood margin maximization," *Pattern Recognit.*, vol. 42, no. 11, pp. 2863–2875, Nov. 2009.

[28] B. Alipanahi, M. Biggs, and A. Ghodsi, "Distance metric learning vs. Fisher discriminant analysis," in *Proc. AAAI Conf. Artif. Intell.*, 2008, pp. 598–603.

[29] A. Globerson and S. Roweis, "Metric learning by collapsing classes," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2006, pp. 451–458.

[30] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Dec. 2009.

[31] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," in *Advances in Neural Information Processing Systems*, vol. 15. Cambridge, MA: MIT Press, 2003, pp. 505–512.

[32] H. Chen, H. Chang, and T. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 846–853.

[33] S.Wang, H. Chen, X. Peng, and C. Zhou, "Exponential locality preserving projections for small sample size problem," *Neurocomputing*, vol. 74, no. 17, pp. 3654–3662, Oct. 2011.

[34] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.

[35] B. Schölkopf, A. Smola, and K.Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[36] S. Mika, G. Rätsch, B. Schölkopf, A. Smola, J. Weston, and K. Müller, "Invariant feature extraction and classification in kernel spaces," in *Proc.Adv. Neural Inf. Process. Syst.*, 1999, pp. 526–532.

[37] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, Oct. 2000.

[38] L. Chen, H. Liao, J. Lin, M. Kao, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, Oct. 2000.

[39] J. Yang, A. F. Frangi, D. Z. J.-Y. Yang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230–244, Feb. 2005.

[40] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. Int. Joint Conf. Artif. Intell.*, 2007, pp. 708–713.

[41] Y. Fu, M. Liu, and T. S. Huang, "Conformal embedding analysis with local graph modeling on the unit hypersphere," in *Proc. IEEE Int. Conf.Comput. Vis. Pattern Recognit.*, 2007, pp. 1–6.

[42] H. Yu and H. Yang, "A direct LDA algorithm for high-dimensional data: With application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, Oct. 2001.

[43] C. Moler and C. V. Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM Rev.*, vol. 45, no. 1, pp. 3–49, Mar. 2003.

[44] N. Higham, "The scaling and squaring method for the matrix exponential revisited," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 4, pp. 1179–1196, 2005.

[45] J. Shawe-Taylor and N. Cristianini, *Support Vector Machines and OtherKernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ.Press, 2000.

[46] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.