Published in IET Computer Vision
Received on 31st October 2008
Revised on 11th June 2009
doi: 10.1049/iet-cvi.2008.0065

ISSN 1751-9632

Preserving global and local information – a combined approach for recognising face images

K. Ruba Soundar[1] *K. Murugesan*[2]

[1]*Department of Computer Science and Engineering, PSR Engineering College, Sivakasi 626140, Tamil Nadu, India*
[2]*Principal, Bharathiyar Institute of Engineering for Women, Deviyakurichi – 636112, Tamil Nadu, India*
E-mail: rubasoundar@yahoo.com

Abstract: Face recognition can significantly impact authentication, monitoring and indexing applications. Much research on face recognition using global and local information has been done earlier. By using global feature preservation techniques like principal component analysis (PCA) and linear discriminant analysis (LDA), the authors can effectively preserve only the Euclidean structure of face space that suffers lack of local features, but which may play a major role in some applications. On the other hand, the local feature preservation technique namely locality preserving projections (LPP) preserves local information and obtains a face subspace that best detects the essential face manifold structure; however, it also suffers loss in global features which may also be important in some of the applications. A new combined approach for recognising faces that integrates the advantages of the global feature extraction technique LDA and the local feature extraction technique LPP has been introduced here. Xiaofei He *et al.* in their work used PCA to extract similarity features from a given set of images followed by LPP. But in the proposed method, the authors use LDA (instead of PCA) to extract discriminating features that yields improved facial image recognition results. This has been verified by making a fair comparison with the existing methods.

1 Introduction

Face recognition plays a vital role in applications such as identity authentication, access control and surveillance. Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [1–6]. Even though human can detect and identify faces in a scene with little or no effort, building an automated system that accomplishes such objectives is very challenging. The challenges are even more profound when one considers the large variations in the visual stimulus owing to illumination conditions, viewing directions or poses, facial expressions, aging and disguises such as facial hair, glasses or cosmetics. Face recognition research provides the cutting edge technologies for use in commercial, law enforcement and military applications. An automated vision system that performs the functions of face detection, verification and recognition will find countless unobtrusive applications, such as airport security and access control, building (embassy) surveillance and monitoring, human–computer intelligent interaction and perceptual interfaces, smart environments at home, office and cars that derive desirable facial features characterised by spatial frequency, spatial locality and orientation selectivity to cope with the variations due to illumination and facial expression changes [6–8].

A large number of dimensionality reduction techniques exist in the literature. In practical situations, when the image dimension is prohibitively large, one is often forced to use linear techniques. Consequently, projective maps have been the subjects of considerable investigation. Three popular forms of linear techniques are principal component analysis (PCA) [3], multidimensional scaling (MDS) and linear discriminant analysis (LDA) [3]. Each of these is an eigenvector method designed to model linear variabilities in high-dimensional data. The PCA method performs dimensionality reduction by projecting the original n-dimensional data onto the $k \ll n$-dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix. Thus, PCA builds a global linear model of the data. Classical MDS method finds an embedding that preserves pair-wise distances between data points, and it is equivalent to PCA when those distances are Euclidean. Both PCA and MDS are unsupervised learning algorithms, whereas LDA is a supervised learning algorithm. The LDA method searches for the projective axes on which the data points of different classes are far from each other (maximising between-class scatter), while constraining the data points of the same class to be as close to each other as possible (minimising within-class scatter).

In the work by Wang and Tang [9], a dual-space LDA approach for face recognition was proposed, which is based on a probabilistic visual model. The eigenvalue spectrum in the space of within-class scatter matrix is estimated, and discriminant analysis is simultaneously applied in the principal and subspaces of the within-class scatter matrix. In the work by Sergey Ioffe [10], a generative probability model is proposed, with which we can extract both the features and combine them for recognition. Also, in the work by Prince and Elder [11], they developed a probabilistic version of Fisher faces called probabilistic LDA. This method allows the development of non-linear extensions that are not obvious in the standard approach, but it suffers with its implementation complexity. Jain *et al.* [12] introduced the notion of hyper-features, whose properties of an image patch can be used to estimate the utility of the patch in subsequent matching tasks. These methods produce some higher results than the normal LDA method, but it did not consider the local manifold structure of the input images, which may play a major role in some applications. In this case, most of the signal lies in part of the subspace where the noise is also great.

A technique for direct visual matching of images for the purpose of face recognition and image retrieval, using a probabilistic measure of similarity based primarily on a Bayesian analysis of image differences, is introduced in [13]. This method replaces costly computation of non-linear (online) Bayesian similarity measures by inexpensive linear (offline) subspace projections and simple Euclidean norms, thus resulting in a significant computational speed-up for implementation with very large databases. Compared to the dimensionality reduction techniques, this method requires the probabilistic knowledge about the past information. In the work by Srisuk and Kurutach [14], the masked trace transform offers 'texture' information for face representation which is used to reduce the within-class variance as discussed. They also introduced a distance measure incorporating the weighted trace transform in order to select only the significant features from the trace transform.

There exist some of the spectral methods for dimensionality reduction purposes, namely distance preserving spectral methods and topology preserving spectral methods [15]. Actually, the first method involves pair-wise distances either directly or with some kind of weighting, which is proportional to the inverse of the Euclidean distances measured in the data space. In this the Euclidean distances to the K-nearest neighbours (KNN) alone are taken into account, while others are not considered and are replaced by those determined during the semi-definite programming step. On the other hand, in topology preserving spectral methods pair-wise distances are never used directly. Instead they are replaced with some kind of similarity measure, which most of the time is a decreasing function of the pair-wise distances.

There is another method for linear dimensionality reduction, called locality preserving projections (LPP) [16]. It builds a graph incorporating neighbourhood information of the data set. Using the notion of the Laplacian of the graph, we then compute a transformation matrix, which maps the data points to a subspace. This linear transformation optimally preserves local neighbourhood information in a certain sense. The representation map generated by the algorithm may be viewed as a linear discrete approximation to a continuous map that naturally arises from the geometry of the manifold [5]. The locality preserving character of the LPP algorithm makes it relatively insensitive to outliers and noise.

Earlier works based on PCA [17] or LDA [18] suffer from not preserving the local manifold of the face structure, whereas the research works on LPP [4] lack to preserve global features of face images. Some papers [1, 16] use the combination of both PCA and LPP that captures only the most expressive features, whereas our proposed work uses the combination LDA and the distance preserving spectral method LPP that captures the most discriminative features, which plays a major role in face recognition. Also those works that uses PCA captures the variation in the samples without considering the variance among the subjects. Hence in our proposed work, for the first time to our knowledge, we employ the combination of global feature extraction technique LDA and local feature extraction technique LPP to achieve a high-quality feature set called combined global and local preserving features (CGLPF) that captures the discriminate features among the samples, considering the different classes in the subjects which produces the considerable improved results in facial image representation and recognition. To reduce the effect of overlapping features, only a little amount of local features are eliminated by preserving all the global features in the first stage and the local features are extracted from the output of the first stage to produce good recognition result.

The rest of the paper is organised as follows. Section 2 describes eigenfaces obtained using PCA, Fisher faces obtained using LDA and Laplacian faces obtained using LPP. The newly proposed combined approach to extract CGLPF is given in Section 3. In Section 4, the facial images that are used and the results obtained using the CGLPF are presented. Also, a comparison of the CGLPF facial image recognition results with other methods like PCA, LDA and LPP on 400 images from Olivetti

doi: 10.1049/iet-cvi.2008.0065

Research Laboratory (ORL) image database, 400 images from University Manchester Institute of Science and Technology (UMIST) image database, 600 images formed by the combination of both of these databases and the 440 images from Indian face database [19] are presented. The paper is concluded with some closing remarks in Section 5.

2 PCA, LDA and LPP

In this section, the three pattern classification techniques for solving the face recognition problem, namely eigenface methods, Fisher face methods and Laplacian methods are shown. We approach this problem within the pattern classification paradigm, considering each of the pixel values in a sample image as a coordinate in a high-dimensional space (the image space).

2.1 Eigenfaces

As correlation methods are computationally expensive and require great amount of storage, it is natural to pursue dimensionality reduction schemes. A technique now commonly used for dimensionality reduction in computer vision – particularly in face recognition – is PCA [17, 20, 21]. PCA techniques, also known as Karhunen–Loeve methods, choose a dimensionality reducing linear projection that maximises the scatter of all projected samples. One approach to cope with the problem of excessive dimensionality of the image space is to reduce the dimensionality by combining features. Linear combinations are particularly attractive because they are simple to compute and analytically tractable. In effect, linear methods project the high-dimensional data onto a lower dimensional subspace.

More formally, let us consider a set of N sample images $\{x_1, x_2, \ldots, x_n\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of the c classes $\{X_1, X_2, \ldots, X_c\}$. Let us also consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where $m \ll n$. The new feature vectors $\boldsymbol{y}_k = R^m$ are defined by the following linear transformation

$$\boldsymbol{y}_k = \boldsymbol{W}^{\mathrm{T}} x_k \quad k = 1, 2, \ldots, N \tag{1}$$

If the total scatter matrix $\boldsymbol{S}_{\mathrm{T}}$ is defined as

$$\boldsymbol{S}_{\mathrm{T}} = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^{\mathrm{T}} \tag{2}$$

where n is the number of sample images and $\mu \in R^m$ is the mean image of all samples, then after applying the linear transformation $\boldsymbol{W}^{\mathrm{T}}$, the scatter of the transformed feature vectors $\{y_1, y_2, \ldots, y_n\}$ is $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{T}}\boldsymbol{W}$. In PCA, the projection $\boldsymbol{W}_{\mathrm{opt}}$ is chosen to maximise the determinant of the total scatter matrix of the projected samples, that is

$$\begin{aligned} \boldsymbol{W}_{\mathrm{opt}} &= \arg\max_{w} |\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{T}}\boldsymbol{W}| \\ &= [w_1\, w_2\, i \ldots w_m] \end{aligned} \tag{3}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of n-dimensional eigenvectors of $\boldsymbol{S}_{\mathrm{T}}$ corresponding to the m largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as eigenpictures in [21] and eigenfaces in [17]. If classification is performed using a nearest neighbour classifier in the reduced feature space and m is chosen to be the number of images N in the training set, then the eigenface method is equivalent to the correlation method in the previous section.

A drawback of this approach is the scatter being maximised not only because of the between-class scatter that is useful for classification, but also due to the within-class scatter that, for classification purposes, has unwanted information. Much of the variation from one image to the next is due to illumination changes [22, 23]. Thus, if PCA is presented with the images of faces under varying illumination, the projection matrix $\boldsymbol{W}_{\mathrm{opt}}$ will contain principal components (i.e. eigenfaces) which retain, in the projected feature space, the variation due to lighting. Consequently, the points in the projected space will not be well clustered, and worse, if the classes may be smeared together.

It has been suggested that by discarding the three most significant principal components, the variation due to lighting is reduced. The hope is that if the first principal components capture the variation due to lighting, then better clustering of projected samples is achieved by ignoring them. Yet, it is unlikely that the first several principal components correspond solely to variation in lighting; as a consequence, information that is useful for discrimination may be lost.

2.2 Fisher faces

While PCA seeks directions that are efficient for representation, LDA seeks directions that are efficient for discrimination. Let us consider a set of N sample images $\{x_1, x_2, \ldots, x_n\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of the c classes, $\{X_1, X_2, \ldots, X_c\}$. Let us also consider a linear transformation that maps the original n-dimensional image space into an m-dimensional feature space, where $m \ll n$. This method selects the weight matrix $\boldsymbol{W}_{\mathrm{opt}}$ in such a way that the ratio of the between-class scatter and the within-class scatter is maximised as given below

$$\begin{aligned} \boldsymbol{W}_{\mathrm{opt}} &= \arg\max_{w} \left| \frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{B}}\boldsymbol{W}}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{W}}\boldsymbol{W}} \right| \\ &= [w_1\, w_2 \ldots w_m] \end{aligned} \tag{4}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalised eigenvectors of between-class scatter matrix $\boldsymbol{S}_B$ and within-class scatter matrix $\boldsymbol{S}_W$ corresponding to the m largest generalised eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$, that is

$$\boldsymbol{S}_B W_i = \lambda_i \boldsymbol{S}_W W_i, \quad i = 1, 2, \ldots, m \tag{5}$$

Note that there are at most $c - 1$ non-zero generalised eigenvalues, and so an upper bound on m is $c - 1$, where c is the number of classes [18].

2.3 Laplacian faces

The PCA and LDA aim to preserve the global structure, whereas the LPP [1, 16] seek to preserve the intrinsic geometry of the data and local structure. Let us consider a set of N sample images $\{x_1, x_2, \ldots, x_n\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of the c classes, $\{X_1, X_2, \ldots, X_c\}$. Consider the problem of representing all of the vectors in a set of n d-dimensional samples $x_1, x_2, \ldots, x_n$ with zero mean by a single vector $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ such that y_i represents x_i. Specifically, we find a linear mapping from the d-dimensional space to a line. We denote the transformation vector by $\boldsymbol{W}_{\mathrm{LPP}}$. That is, $\boldsymbol{W}_{\mathrm{LPP}}^{\mathrm{T}} x_i = y_i$. Actually, the magnitude of $\boldsymbol{W}_{\mathrm{LPP}}$ is of no real significance because it merely scales $\underline{y}_i$. The objective function of LPP is as follows

$$\min \sum_{ij} (y_i - y_j)^2 S_{ij} \tag{6}$$

where y_i is the one-dimensional representation of x_i and the matrix $\boldsymbol{S}$ is a similarity matrix.

The following are the steps to be carried out to obtain the Laplacian transformation matrix $\boldsymbol{W}_{\mathrm{LPP}}$.

1. *Constructing the nearest-neighbour graph:* Let G denotes a graph with k nodes. The ith node corresponds to the face image x_i. We put an edge between nodes i and j if x_i and x_j are 'close', that is, x_j is among k nearest neighbours of x_i or x_i is among k nearest neighbours of x_j. The constructed nearest neighbour graph is an approximation of the local manifold structure, which will be used by the distance preserving spectral method to add the local manifold structure information to the feature set.

2. *Choosing the weights:* The weight matrix $\boldsymbol{S}$ of graph G models the face manifold structure by preserving local structure. If nodes i and j are connected, put

$$S_{ij} = e^{-\|x_i - x_j\|^2/t} \tag{7}$$

where t is a suitable constant. Otherwise, put $S_{ij} = 0$.

3. *Eigen map:* The transformation matrix $\boldsymbol{W}_{\mathrm{LPP}}$ that minimises the objective function is given by the minimum eigenvalue solution to the generalised eigenvalue problem. A detailed study about LPP and Laplace Beltrami operator is found in [1, 24]. The eigenvectors and eigenvalues for the generalised eigenvector problem are computed using (8).

$$\boldsymbol{XLX}^{\mathrm{T}} \boldsymbol{W}_{\mathrm{LPP}} = \lambda \boldsymbol{XDX}^{\mathrm{T}} \boldsymbol{W}_{\mathrm{LPP}} \tag{8}$$

where $\boldsymbol{D}$ is a diagonal matrix whose entries are column or row sums of $\boldsymbol{S}$, $D_{ii} = \Sigma_j S_{ji}$, $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{S}$ is the Laplacian matrix. The ith row of matrix $\boldsymbol{X}$ is x_j. Let $\boldsymbol{W}_{\mathrm{LPP}} = w_0, w_1, \ldots, w_{k-1}$ be the solutions of the above equation, ordered according to their eigenvalues, $0 \le \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{k-1}$. These eigenvalues are equal to or greater than zero because the matrices $\boldsymbol{XLX}^{\mathrm{T}}$ and $\boldsymbol{XDX}^{\mathrm{T}}$ are both symmetric and positive semi-definite. Note that the two matrices $\boldsymbol{XLX}^{\mathrm{T}}$ and $\boldsymbol{XDX}^{\mathrm{T}}$ are both symmetric and positive semi-definite since the Laplacian matrix $\boldsymbol{L}$ and the diagonal matrix $\boldsymbol{D}$ are both symmetric and positive semi-definite.

3 Combined global and local preserving features

The proposed combined approach that combines global feature preservation technique LDA and local feature preservation technique LPP to form the high-quality feature set CGLPF is described in this section. Actually, the CGLPF method is used to project face data to an LDA space for preserving the global information and then projecting to LPP space by using the distance preserving spectral methods, to add the local neighbourhood manifold information which may not be interested by LDA.

3.1 Preserving the global features

The mathematical operations involved in LDA, the global feature preservation technique, are analysed here. The fundamental operations are:

1. The data sets and the test sets are formulated from the patterns which are to be classified in the original space.

2. The mean of each data set μ_i and the mean of entire data set μ are computed.

$$\mu = \sum_i p_i \mu_i \tag{9}$$

where p_i is a priori probabilities of the classes.

3. Within-class scatter $\boldsymbol{S}_{\mathrm{w}}$ and the between-class scatter $\boldsymbol{S}_{\mathrm{b}}$ are computed using

$$\boldsymbol{S}_{\mathrm{w}} = \sum_j p_j{}^*(\mathrm{cov}_j) \tag{10}$$

$$\boldsymbol{S}_{\mathrm{b}} = \sum_j (x_j - \mu)(x_j - \mu) \tag{11}$$

doi: 10.1049/iet-cvi.2008.0065

where cov_j the expected covariance of each class is computed as

$$\text{cov}_j = \prod_i (x_j - \mu_i) \tag{12}$$

Note that $\boldsymbol{S}_\text{b}$ can be thought of as the covariance of data set whose members are the mean vectors of each class. As defined earlier, the optimising criterion in LDA is the ratio of between-class scatter to the within-class scatter. The solution obtained by maximising this criterion defines the axes of the transformed space.

The LDA can be a class-dependent or class-independent type. The class-dependent LDA requires L-class L-separate optimising criterion for each class denoted by C_1, C_2, $\ldots$, C_L and that are computed using

$$C_j = (\text{cov}_j)^{-1}\boldsymbol{S}_\text{b} \tag{13}$$

4. The transformation space for LDA, $\boldsymbol{W}_\text{LDA}$ is found as the eigenvector matrix of different criteria defined in (13).

In our experiments, we first apply the global feature extraction technique, LDA to preserve an optimum of 90% of global features and then applying the local feature extraction technique, LPP to preserve local features. Based on various experiments, we have selected the optimum value as 90%. Choosing a value less than 90% results in the removal of more local features with the discarded unimportant global features, whereas choosing a value more than 90% results in the constraint that makes the features more difficult to discriminate from one another.

3.2 Adding local features

The LPP technique to preserve local features is applied and the transformation matrix $\boldsymbol{W}_\text{LPP}$ is obtained as described in Section 2.3. We have used the distance preserving spectral method instead of the topology preserving spectral methods because, in all distance preserving methods, the eigensolver is applied to a dense matrix, whose entries are Euclidean distances and only the top eigenvectors are used to form the solution of a maximisation problem that consists of maximising the variance in the embedding space, which is directly related to the associated eigenvalues. On the other hand, in the case of topology preserving methods, the eigensolver is applied to a sparse matrix and the bottom eigenvectors are used to form the solution of a minimisation problem, which generally corresponds to a local reconstruction error or distortion measure [15].

By considering the transformation space $\boldsymbol{W}_\text{LDA}$ and $\boldsymbol{W}_\text{LPP}$, the embedding is done as follows

$$x \rightarrow y = \boldsymbol{W}^\text{T} x, \quad \boldsymbol{W} = \boldsymbol{W}_\text{LDA}\boldsymbol{W}_\text{LPP}, \\ \boldsymbol{W}_\text{LPP} = [w_0, w_1, \ldots, w_{k-1}] \tag{14}$$

where y is a k-dimensional vector and $\boldsymbol{W}_\text{LDA}$, $\boldsymbol{W}_\text{LPP}$ and $\boldsymbol{W}$ are the transformation matrices of LDA, LPP and CGLPF methods.

In our proposed work, for the first time to our knowledge, we employ the combination of LDA and LPP that captures the discriminate features among the samples considering the different classes in the subjects which produces considerable improved results. Also, the linear mapping obtained using our CGLPF best preserves the global discriminating features and the local manifold's estimated intrinsic geometry in a linear sense.

4 Experimental results and discussion

In this section, the images that are used in this work and the results of facial image recognition obtained with the newly proposed CGLPF feature set are presented. For face recognition the features obtained by the PCA, LDA, LPP and CGLPF described in previous sections are used. For classification experiments, the facial images from the facial image databases ORL, UMIST, combination of both ORL and UMIST and Indian face database are used. The ORL database contains a total of 400 images containing 40 subjects each with 10 images in different poses. Similarly, the UMIST database contains a total of 400 images having 20 subjects, each with 20 different posed images. We also form a combination of ORL and UMIST database images to obtain 600 images, wherein 400 images (40 subjects $\times$ 10 images) are from ORL database and 200 images (20 subjects $\times$ 10 images) are from UMIST database that differ in poses, expressions and lighting conditions. For verifying the reliability of our algorithm in more real-time images with more expression changes, we applied our algorithm in the Indian face database which contains 440 images (40 subjects $\times$ 11 images). Figs. 1 and 2 show sample images used in our experiments collected from ORL and the UMIST face databases, respectively.

The images of ORL, UMIST and combined databases are already aligned and no additional alignments are done by us. However, the images in the Indian face database have background information with non-facial components and in different orientations such as front, looking left, looking right, looking up, looking up towards left, looking up towards right and looking down. Also, these images show various emotions like neutral, smile, laughter and sad/disgust. These are used in our experiments without making any alignment. For normalisation purpose, we make all the images into equal size of 50×50 pixels by doing the bilinear image resizing.

The database is separated into two sets as follows: (i) the testing set is formed by taking one image for each subject, so totally 40, 20, 60 and 40 testing images from ORL, UMIST, ORL + UMIST and Indian face databases,

Figure 1 *Sample set of images collected from ORL database*

Figure 2 *Sample set of images collected from UMIST database*

respectively, and (ii) the training set is formed by taking all the images excluding the image used for testing, thus the training set contains 360, 380, 540 and 400 images from ORL, UMIST, ORL + UMIST and Indian face databases, respectively. Since one cannot always expect that all testing images will form a part of the training image set, we exclude the testing image from that of training image set for all subjects in our analysis.

In the experimental phase, considering the ORL database with 40 subjects each having 10 images in different poses, we take the first image of the first subject as the testing image.

doi: 10.1049/iet-cvi.2008.0065

The top matching nine images are found from a set of 360 images excluding the first image of each of the 40 subjects. If the top matching images lie in the same row (subject), then it is treated as a correct recognition. The percentage of classification is defined as the ratio of the number of correct recognition to the total number of top matching images. Likewise, we take the first image of other 39 subjects for correct recognition process and the percentage of classification is calculated. From these 40 values of percentage of classification, the average value is obtained and is shown in the first row of Table 1 for different methods such as PCA, LDA, LPP and CGLPF. The second row entries in Table 1 are obtained by the same procedure explained above, but by taking second images of all 40 subjects. This is followed for all ten images of all 40 subjects and the average values of classification percentage are tabulated in Table 1. Similar analyses are carried out using UMIST, combination of ORL and UMIST and Indian face databases and are given in Tables 2–4, respectively. In addition to the average percentage of classification values, the statistical parameters namely standard deviation and standard error are also calculated and tabulated in Tables 1–4 for showing the statistical significance of our proposed method. We consider the Euclidian distance between the training and testing images as a measure of similarity.

The output of the Tables 1–4 shows that the CGLPF performs better than the other approaches like PCA, LDA and LPP for various databases. It is because, in CGLPF

Table 1 Comparison of average percentage of classification of the proposed CGLPF method with the existing methods for ORL database

Testing image number	PCA	LDA	LPP	CGLPF
1	69.4444	87.6667	72.5	89.7222
2	66.9444	89.1111	71.3889	91.9444
3	67.2222	86.4444	72.5	91.1111
4	65.2778	86.3333	69.7222	89.7222
5	66.9444	88.1667	71.1111	91.3889
6	63.6111	88.8333	66.3889	87.7778
7	68.6111	87.6667	72.5	89.1667
8	67.5	89.3333	70.5556	89.7222
9	65.2778	83.3889	68.8889	86.9444
10	63.3333	82.3889	66.1111	85.5556
average	66.41665	86.93333	70.16667	89.30555
standard deviation	2.009273	2.369323	2.390253	2.029645
standard error	0.635388	0.749246	0.755864	0.64183

Table 2 Comparison of average percentage of classification of the proposed CGLPF method with the existing methods for UMIST database

Testing image number	PCA	LDA	LPP	CGLPF
1	34.4737	52.3684	33.1579	55.2778
2	36.3158	55.2632	36.0526	58.8889
3	39.4737	63.6842	40.2632	62.5
4	41.0526	65.2632	38.6842	67.2222
5	42.8947	65.7895	40.5263	64.7222
6	44.7368	67.8947	41.8421	66.6667
7	43.6842	64.4737	42.3684	65.8333
8	45	60.7895	43.1579	60.8333
9	45	59.2105	41.8421	60.5556
10	42.3684	54.2105	40	57.7778
11	41.3158	55	39.7368	57.2222
12	40.5263	58.1579	38.9474	59.4444
13	44.2105	63.6842	42.8947	63.8889
14	47.1053	65.7895	45	67.2222
15	46.8421	67.1053	42.8947	69.1667
16	47.1053	66.5789	42.6316	69.7222
17	45	65	41.8421	68.6111
18	42.8947	66.3158	40.2632	69.1667
19	41.3158	67.8947	37.8947	68.3333
20	36.8421	59.7368	34.7368	62.5
average	42.40789	62.21053	40.23684	63.77778
standard deviation	3.56993	4.98094	3.00156	4.53615
standard error	0.79826	1.11377	0.67117	1.01431

both the global and local features are preserved, whereas in the other cases either global feature or local feature is preserved and the other one is ignored since it affects the results. In our experiments, the global features are preserved in the Fisher feature spare and the local features are preserved using the KNN graph-based distance preserving spectral method LPP. Instead of simple KNN-based approach, the topology preserving spectral method such as locally linear embedding (LLE) can also be employed for adding the local features in the LDA space. We have conducted the experiments using the LLE also, but the results obtained show very poor percentage of classification. It is because the main objective function of the LLE is to

Table 3 Comparison of average percentage of classification of the proposed CGLPF method with the existing methods for combination of ORL and UMIST databases

Testing image number	PCA	LDA	LPP	CGLPF
1	65.7407	81.6667	69.2593	84.0741
2	65.5556	85.3704	71.4815	86.2963
3	68.5185	85.7407	73.1481	88.7037
4	70.5556	86.1111	72.4074	88.3333
5	70.9259	86.2963	74.8148	88.7037
6	70.1852	84.6296	71.6667	87.963
7	72.5926	85	73.1481	87.2222
8	69.8148	85.1852	72.4074	87.4074
9	66.1111	84.4444	69.8148	86.6667
10	61.8519	78.8889	67.7778	79.0741
average	68.18519	84.33333	71.59259	86.44445
standard deviation	3.2766	2.31164	2.10092	2.93686
standard error	1.03615	0.731003	0.66437	0.92872

minimise distances between neighbouring points, whereas the LPP aims to maximise the distances between non-neighbouring points. This minimising objective function can lead to degenerate solutions, such as an embedding having identical coordinates for all data points.

Fig. 3 shows the comparison chart of the overall percentage of classification obtained with the proposed CGLPF method and the other three existing methods, excluding the images used for testing in the training image set. However, as in most of the cases, if the testing images also form a part of the training image set, then the results obtained are shown in Fig. 4. The error bars are also shown in Figs. 3 and 4 for showing the statistical significance of our proposed method.

Since the images in the ORL database are aligned to the middle of the face image without any background, the variation in the images are only due to the internal expression changes. This is because of the LPP, which give better results than PCA for the ORL database. For the UMIST database, the images are captured under various angles, which is a global variation and hence the PCA works better than the LPP. For the case of combined database, only partial database from the UMIST is used, hence the results of ORL database dominated the net result. In all the above three cases, the LDA performs well than PCA and LPP, because of its discriminating nature to

Table 4 Comparison of average percentage of classification of the proposed CGLPF method with the existing methods for Indian face database

Testing image number	PCA	LDA	LPP	CGLPF
1	47.25	66.25	72.6667	77.6667
2	48.75	71.75	75.8333	83.3333
3	45.25	66.25	71.8333	75.8333
4	46.75	69.25	77.6667	76.5
5	45.25	66.25	71.8333	75.8333
6	44.75	67.75	72.6667	79.1333
7	45.25	69.25	75.8333	81.0333
8	47.25	68.25	71.8333	77.6667
9	46.75	66.25	72.6667	79.1333
10	45.25	67.75	71.8333	72.6667
11	46.75	69.25	75.8333	75.8333
average	46.29545	68.02273	73.68181	77.69393
standard deviation	1.29099	1.81123	2.14563	2.98841
standard error	0.40825	0.57276	0.67851	0.94502

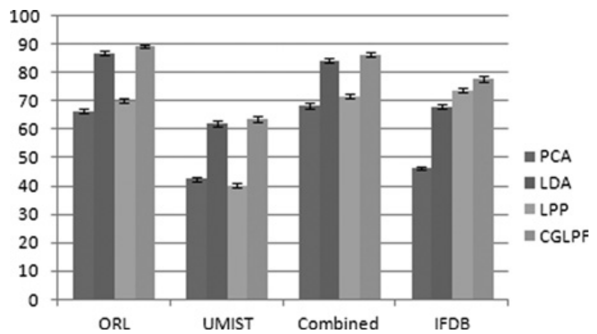

Figure 3 *Overall comparison of average percentage of classification of proposed CGLPF method with existing methods excluding testing images in the training image set*

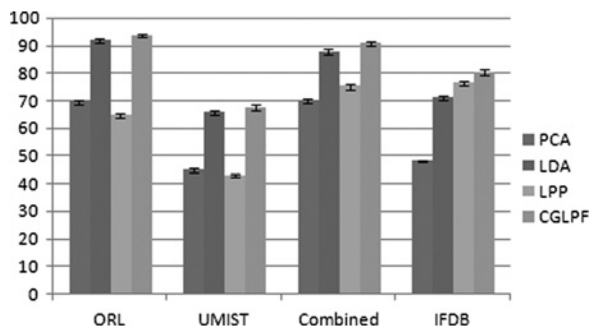

Figure 4 *Overall comparison of average percentage of classification of proposed CGLPF method with existing methods excluding testing images in the training image set*

doi: 10.1049/iet-cvi.2008.0065

Table 5 Average percentage of classification of the proposed CGLPF method using ORL, UMIST, combined and Indian face databases without repetition, repeating for 2, 5 and 10 times

Database	Without repetition	Two times repetition	Five times repetition	Ten times repetition
ORL	89.30555	89.47223	89.27778	89.33332
UMIST	63.77778	63.88889	63.84445	63.91667
combined	86.44445	86.43103	86.44505	86.44447
IFDB	77.69393	77.71273	77.70327	77.69483

recognise the images. This is the reason why we employ CGLPF method, wherein the global feature extraction LDA and local feature extraction LPP are integrated to yield a good percentage of classification rather than combining PCA and LPP as in [1]. For the Indian face database, the LPP outperforms PCA and LDA. This is because the images in the database are with constant background and hence it is necessary to distinguish the images only by local variations. Since the large background area and the non-facial part of the images dominate the discriminating local features, the LDA produces only lower results. But our proposed method CGLPF that combines the LDA and LPP uses the advantages of LDA for the first three cases and for the last case it uses the advantages of LPP. The results shown in Fig. 3, which is compared with Fig. 4, seem to have improved slightly; this is because, when the testing images are also part of the training image set, the same trained images will take part surely in the closest matching images. More experiments have been carried out to illustrate that the proposed method is not affected due to overfitting, and the results are indicated in Table 5, wherein the same actual database is again repeated for two, five and ten times to yield a redundant enlarged database and the average percentage of classification are calculated. A close consistency in results proves that the proposed method is not affected due to overfitting.

It is the nature that the time complexity is increasing when using the combined schemes compared to using the techniques individually. But in our proposed method, the training is done offline and the testing is done in the real time or online. In the online phase, it is only going to project the testing image into the CGLPF feature set which is having only lower dimensions compared to the cases where the techniques are used individually. Hence when we employ our method in real-time applications, there is no delay in the online, and the offline delay does not cause any considerations in the real-time processing.

5 Conclusions

A new approach that combines the global and local information preserving features has been implemented and tested for standard facial image databases like ORL and UMIST. The feature set created is an extension to the Laplacian faces used in Xiaofei He *et al.*, where they use the PCA only for reducing the dimension of the input image space, and we use LDA for preserving the discriminating features in the global structure. The CGLPF feature set created using the combined approach retains the global information and local information, which makes the recognition insensitivity to absolute image intensity and insensitivity to contrast and local facial expressions.

In several applications, the feature insensitivity to different image transformations is not only a desired property but also a necessary one. In this work, it is observed that the CGLPF shows a good percentage of recognition and it is superior to the conventional methods when the images are subjected to various expressions and pose changes. Therefore the CGLPF feature set obtained through the combined approach seems to be an attractive choice for many facial-related image applications.

6 References

[1] HE X., YAN S., HU Y., NIYOGI P., ZHANG H.-J.: 'Face recognition using Laplacian faces', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (3), pp. 328–340

[2] BATUR A.U., HAYES M.H.: 'Linear subspace for illumination robust face recognition'. Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2001

[3] BELHUMEUR P.N., HESPANHA J.P., KRIEGMAN D.J.: 'Eigenfaces vs. Fisher faces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720

[4] BELKIN M., NIYOGI P.: 'Laplacian eigenmaps and spectral techniques for embedding and clustering'. Proc. Conf. on Advances in Neural Information Processing System, 2001

[5] BELKIN M., NIYOGI P.: 'Using manifold structure for partially labeled classification'. Proc. Conf. on Advances in Neural Information Processing System, 2002

[6] BRAND M.: 'Charting a manifold'. Proc. Conf. on Advances in Neural Information Processing Systems, 2002

[7] CHUNG F.R.K.: 'Spectral graph theory'. Proc. Regional Conf. Series in Mathematics no. 92, 1997

[8] CHANG Y., HU C., TURK M.: 'Manifold of facial expression'. Proc. IEEE Int. Workshop on Pattern Analysis, 2003

[9] WANG X., TANG T.: 'Dual-space linear discriminant analysis for face recognition'. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, July 2004, vol. 2, (27), pp. 564–569

[10] IOFFE S.: 'Probabilistic linear discriminant analysis'. Proc. European Conf. on Computer Vision, 2006, vol. 4, pp. 531–542

[11] PRINCE S.J.D., ELDER J.H.: 'Probabilistic linear discriminant analysis for inferences about identity'. Proc. IEEE Int. Conf. on Computer Vision, 2007

[12] JAIN V., FERENCZ A., LEARNED-MILLER E.: 'Discriminative training of hyper-feature models for object identification', 2008

[13] MOGHADDAM B., JEBARA T., PENTLAND A.: 'Bayesian face recognition', *Pattern Recognit.*, 2000, **33**, (11), pp. 1771–1782

[14] SRISUK S., KURUTACH W.: 'Face recognition using a new texture representation of face images'. Proc. Electrical Engineering Conf., Cha-am, Thailand, 6–7 November 2003, pp. 1097–1102

[15] JOHN A., VERLEYSEN M.: 'Nonlinear dimensionality reduction' (Springer, 2007), pp. 233–238

[16] HE X., NIYOGI P.: 'Locality preserving projections'. Proc. Conf. on Advances in Neural Information Processing Systems, 2003

[17] TURK M., PENTLAND A.: 'Eigenfaces for recognition', *J. Cognit. Neurosci.*, 1991, **3**, (1), pp. 71–86

[18] ZHAO W., CHELLAPPA R., PHILLIPS P.J.: 'Subspace linear discriminant analysis for face recognition'. Technical report CAR-TR-914, Center for Automation Research, University of Maryland, 1999

[19] JAIN V., MUKHERJEE A.: 'The Indian face database', 2002. Available at: http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/

[20] LEVIN A., SHASHUA A.: 'Principal component analysis over continuous subspaces and intersection of half-spaces'. Proc. European Conf. on Computer Vision, 2002

[21] SIROVITCH L., KIRBY M.: 'Low-dimensional procedure for the characterization of human faces', *J. Opt. Soc. Am. A*, 1987, **2**, pp. 519–524

[22] HALLINAN P.: 'A low-dimensional representation of human faces for arbitrary lighting conditions'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1994, pp. 995–999

[23] MOSES Y., ADINI Y., ULLMAN S.: 'Face recognition: the problem of compensating for changes in illumination direction'. European Conf. on Computer Vision, 1994, pp. 286–296

[24] DIAZ-GARCIA J.A.: 'Derivation of the Laplace-Beltrami operator for the zonal polynomials of positive definite Hermitian matrix argument', *Appl. Math. Sci.*, 2007, **1**, (4), pp. 191–200

 IET Comput. Vis., 2010, Vol. 4, Iss. 3, pp. 173–182
 doi: 10.1049/iet-cvi.2008.0065