Expression Subspace Projection for Face Recognition from Single Sample per Person

Hoda Mohammadzade, *Student Member, IEEE,* and Dimitrios Hatzinakos, *Senior Member, IEEE*

Abstract—Discriminant analysis methods are powerful tools in face recognition. However, these methods are not applicable under the single sample per person scenario because the within-subject variability cannot be estimated in this case. In the generic learning solution, this variability is estimated using images of a generic training set, for which more than one sample per person is available. However, because of rather poor estimation of the within-subject variability using a generic set, the performance of discriminant analysis methods is yet to be satisfactory. This is particularly the case when images are under drastic facial expression variation.

In this paper, we show that images with the same expression are located on a common subspace, which here we call it the *expression subspace*. We show that by projecting an image with an arbitrary expression into the expression subspaces, we can synthesize new expression images. By means of the synthesized images for subjects with one image sample, we can obtain more accurate estimation of the within-subject variability and achieve significant improvement in recognition. We performed comprehensive experiments on two large face databases: the Face Recognition Grand Challenge and the Cohn-Kanade AU-Coded Facial Expression database to support the proposed methodology.

Index Terms—Face recognition, facial expression, expression variation, expression transformation, expression subspace, LDA, generic training, single sample per person.

1 Introduction

Over the past two decades, significant advances have been achieved in face recognition techniques. However, there are still many challenges remaining. For example, face recognition in an uncontrolled environment still bears limitations due to illumination, pose and facial expression variation between gallery and probe images. Ref. [1] presents an excellent review of the advances and challenges in face recognition.

Unlike illumination and pose variation problems, the expression variation has not been given sufficient attention. Expression variation is a serious problem in many applications, such as surveillance and human-computer-interaction, where there is no control over the expression of the captured images.

The difficulty of recognizing a person's face, whose gallery and probe images differ in expression, originates from the fact that in the high dimensional face space, faces with the same expression but different identity might be closer to each other than faces with the same identity but different expressions. In other words, for example, the happy image of person A might be more similar to the happy image of person B rather than to the neutral image of person A, which can cause misidentification of person A.

Consequently, the face recognition methods that do not take into account the within-class variability of the face images, i.e., unsupervised methods, do not perform well under expression variation. On the other hand, supervised techniques such as discriminant analysis (DA) methods [2], [3], [4], [5], [6], [7], [8], [9] are very powerful for face recognition under expression variation [10], [11]. DA methods find the features of the face that are more robust to expression variation by learning the within-class variability of the subjects. However, this learning requires more than one sample per subject, and therefore, DA methods cannot be directly used under the single sample (SS) per subject scenario.

The SS problem exists in many face recognition applications such as law enforcement, surveillance and forensic applications. Moreover, adding more samples to the database is too costly, especially in large-scale applications.

Ref. [12] discusses the problem of recognizing from SS and reviews the existing approaches to solve this problem. These approaches include those that address the illumination, pose, and expression variation problem under the SS scenario. The most important methods addressing the expression variation problem under the SS scenario are as follows.

Martinez [13] proposed a local eigenspace approach, which is based on the fact that each expression affects some parts of the face more than others. He proposed to search for those areas that are less affected by a given expressions and then weight each

- *The authors are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada. Email: {hoda,dimitris}@comm.utoronto.ca*

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

local area accordingly.

Li et al. [14] used the idea of separating texture and shape information and projecting them into separate PCA spaces. They stated that the texture of an individual's face is relatively invariant, with slight variations due to illumination changes or blushing. On the other hand, certain shape features are also relatively invariant under some expressions, while some others undergo large variations. They proposed to construct separate eigenspaces for texture and the invariant shape features.

Lee and Kim [15] proposed a method to transform the probe image with an arbitrary expression to its corresponding neutral expression face. The expression transformation is based on the tensorface concept proposed by Vasilescu and Terzopoulos [16]. The transformed probe image is then matched with the gallery images which display neutral expression.

The performance of the above methods is not satisfactory and they require some or all of the following tasks: (1) the expression of the probe image is required to be determined, (2) the probe image is required to be warped to all the gallery images, (3) facial landmark points of the stored and probe images are required to be detected to fit 2D triangulated meshes to them, and (4) the algorithm must be trained on a training set in which every subject has an image of each expression. Performing these tasks is either error-prone and causes the propagation of the error to the recognition stage or they are not always feasible [15].

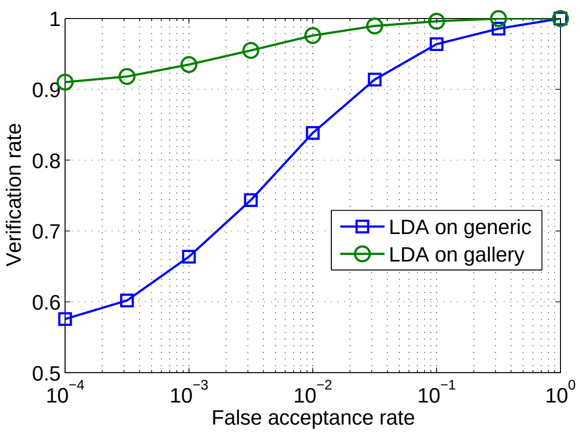


Fig. 1. Verification rate versus FAR for LDA trained on a generic set and LDA trained using two samples per gallery subject.

Another approach for recognition under the SS scenario is to use DA methods by training them using images of subjects different from the gallery subjects (i.e., pool of generic people) for which more than one sample is available [17]. In this case, the within-class variability is being estimated from images of generic people. Wang et al. [17] provided comprehensive examinations of the performance of state-of-the-art DA methods using generic training sets. They reported considerable improvement in the recognition performance compared to unsupervised methods. However, there is still a large gap between the performance of DA methods when they are trained using a generic set versus when they are trained using various images of each gallery subject. In other words, the expressionally robust features of the face found by a DA method are relatively *subject-specific*. There are two quantities that determine the discriminant projection bases: the between-class scatter matrix and the within-class scatter matrix. We found through experiments that, the estimation of the between-class scatter matrix using a generic training set is rather accurate, but what makes the performance of the generic-trained DA methods unsatisfactory is rather poor estimation of the within-class scatter matrix.

Fig. 1 shows the result of an experiment we performed to demonstrate the difference in the performance of LDA when it is trained using images of a generic set versus when it is trained using images of the gallery subjects. To perform this experiment, we formed a generic training set using the images of 100 subjects from the Face Recognition Grand Challenge (FRGC) database [18]. The images contained four expressions: neutral, happy, surprise and puffy cheeks. We used the neutral and happy images of another 150 subjects to form the gallery and probe sets. For each of these subject, we randomly selected one of its happy or neutral images as the gallery image. If the gallery image for that subject was happy, we chose its neutral images as the probe images and vice versa. The first time, we used only the generic set to train LDA ("LDA on generic" in Fig. 1). The second time, we augmented the generic training set by the gallery images and one surprise image for each gallery subject, and then trained LDA using this set ("LDA on gallery"). Note that, we added an image with an expression different from the expression of the gallery and probe images to the training set. One can clearly see the difference in the performance when LDA is trained using images of the gallery subjects.

The observation stated above motivated us to develop a method for synthesizing new expression images of the gallery subjects and add them to the generic set in order to obtain better estimation of the within-class variability of the galley subjects. In our proposed approach, the synthesis is performed by projecting the gallery images into specific subspaces, which here we call them expression subspaces. An expression subspace contains face images displaying that expression and is different from the subspace of another expression.

Moreover, we introduce a method for estimating the *synthesis error* in order to further improve the recognition performance. An important feature of the synthesis error is that it is orthogonal to the Euclidian distance between an input and its synthesized image.

This feature allows us to simply incorporate some statistics of the synthesis error in the calculation of the within-class scatter matrix, and thereby, further improve the recognition performance.

The proposed method does not perform any of the aforementioned tasks required by the existing methods and significantly outperforms them.

The rest of the paper is organized as follows. First, we introduce the expression subspaces and describe how we can synthesize new expression images using these subspaces. Next, we present the proposed face recognition method. Finally, we conclude the paper by presenting the experimental results.

2 Expression Subspaces

One of the main contributions of this work is to define expression subspaces and use them in synthesizing new expression images from only one image of a subject.

For the sake of simplicity, we consider neutral or happy expression in all the derivations throughout this paper, but they can be readily generalized to other expressions.

The vector representation of a $p \times q$-pixel face image can be obtained by row-wise or column-wise concatenation of the pixel values of the face image. Each such $p \times q$-dimensional vector represents a point in the $p \times q$-dimensional Euclidian space. We refer to these points as the face points.

The face points corresponding to images of the same expression lie on the same subspace, which we call the subspace of that expression. For instance, the neutral subspace contains neutral face images. An expression subspace can be approximated by applying PCA on a set of training face images displaying that expression. PCA finds the eigenvectors of the covariance matrix of the face vectors. Let $\{\mathbf{n}_1, ..., \mathbf{n}_M\}$ be the set of neutral face vectors. Their covariance matrix is obtained as

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^{M} (\mathbf{n}_i - \mathbf{m})(\mathbf{n}_i - \mathbf{m})^T \tag{1}$$

where $\mathbf{m}$ denotes the mean of the neutral face vectors and $\mathbf{C}$ denotes the covariance matrix. Because the dimension of the face vectors is very large, calculating the eigenvectors of their covariance matrix is computationally difficult. To solve this problem, a method is described in the well-known eigenface paper [19].

Let $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ be the set of the eigenvectors of $\mathbf{C}$ and $\{\lambda_1, ..., \lambda_k\}$ be their corresponding eigenvalues, which are sorted in descending order. We choose k such that 99% of the eigenvalue energy is used, i.e.,

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{M} \lambda_i} = 0.99 \tag{2}$$

The resulting neutral subspace is a subspace that passes through the neutral mean and is in the direction of the neutral eigenvectors. Every point in this subspace can be written as the summation of the neutral mean and a linear combination of the eigenvectors. Note that an expression subspace is therefore an *affine subspace*.

Fig. 2 shows the face points of pairs of happy and neutral images of 85 subjects from the Cohn-Kanade database [20] in a 3D space. The dimension of the points was reduced to three using PCA. The two planes represent the 2D neutral and happy subspaces. Note that as it is seen in Fig. 2, the happy and neutral subspaces are very different (they have different directions and zero crossings). We can further investigate this difference in the original high dimensional space by calculating the angles between their eigenvectors. Some statistical analysis of the angle between the eigenvectors of two expression subspaces is given in Appendix A.

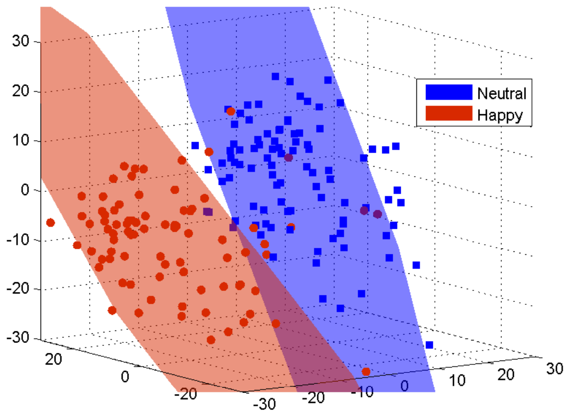


Fig. 2. Scatter plot of pairs of happy and neutral face points of 85 subjects from the Cohn-Kanade database. The two planes represent the neutral and happy subspaces.

An important property of a neutral subspace is that neutral face points are closer to it than happy face points. In order to show this property, we calculated the mean square error (MSE) between a large number of happy and neutral images and their projections into a neutral subspace using the FRGC database. The neutral subspace was constructed using the neutral images of 100 subjects. The MSE was then calculated for the neutral and happy images of 300 other subjects from this database. Fig. 3 shows the average of the MSE for different proportions of the eigenvalue energy of the neutral subspace. As it is known from the definition of PCA [19], the MSE increases by decreasing the eigenvalue energy of a PCA space. Fig. 3 shows that, the MSE of projection into the neutral subspace is larger for happy images than for neutral images, showing that the neutral images are closer to the neutral subspace than the happy images.

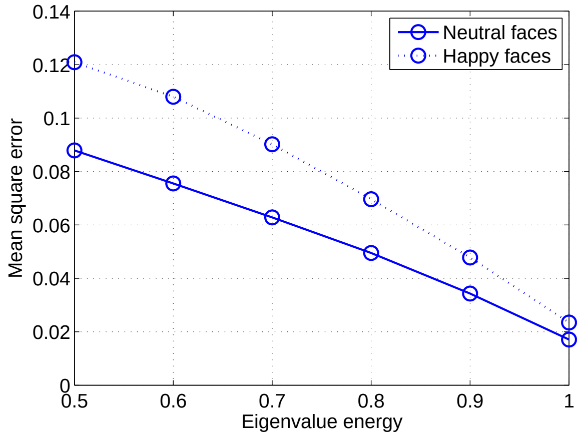


Fig. 3. Reconstruction error for neutral and happy faces projected to the neutral subspace versus different eigenvalue energies.

2.1 Synthesizing New Expression Images

In the following, for simplifying the explanations, we use the neutral subspace as an example of expression subspaces.

We know from geometry that, the projection of a point into a subspace is the nearest point of the subspace to that point. Based on this fact, *the projection of a face image with an arbitrary expression into the neutral subspace gives us an approximation of its neutral face image.*

Let $\mathbf{x}$ denote a face vector with an arbitrary expression. The approximation of the neutral face of the image $\mathbf{x}$ is obtained by reconstructing it from its projection coefficients into the neutral subspace as

$$\mathcal{N}(\mathbf{x}) = [\mathbf{u}_1 ... \mathbf{u}_k][\mathbf{u}_1 ... \mathbf{u}_k]^T(\mathbf{x} - \mathbf{m}) + \mathbf{m} \quad (3)$$

where the term $[\mathbf{u}_1 ... \mathbf{u}_k]^T(\mathbf{x} - \mathbf{m})$ gives the projection coefficients and $\mathcal{N}()$ denotes the neutral synthesis function. Fig. 4 illustrates the projection and reconstruction of the image $\mathbf{x}$ into a 1D neutral subspace (i.e., a line). The neutral subspace is in the direction of the neutral eigenvector $\mathbf{u}$ and passing through the neutral mean face $\mathbf{m}$. Also, $N(\mathbf{x})$ is the reconstruction from the projection of $\mathbf{x}$ into the neutral subspace, which is the approximation of the neutral face of $\mathbf{x}$.

Throughout the rest of this paper, we refer to "the reconstruction from the projection coefficients" of a vector simply by "the projection" of that vector.

2.2 Estimating the Synthesis Error

We call the distance between a true neutral image and its approximation (obtained by projecting its non-neutral image into the neutral subspace) the synthesis

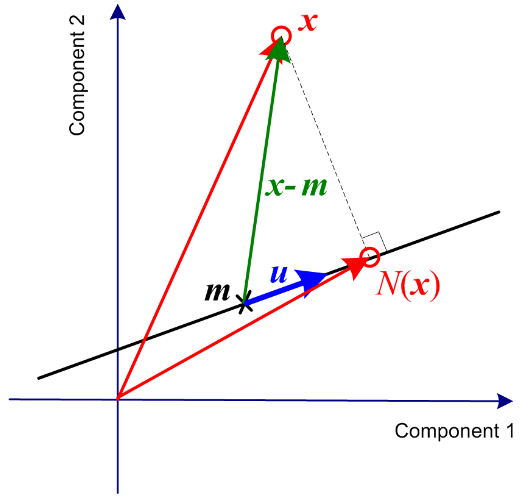


Fig. 4. Projection of an input face vector into a neutral subspace.

error, denoted as $\mathbf{e}_s$. We can estimate this error using a validation set as follows. Let $\mathbf{x}$ be the input image and assume that we want to synthesize its neutral image. Also, let $\{\mathbf{z}_i\}_{i=1}^{l}$ be the set of l nearest images to $\mathbf{x}$ from the validation set and $\{\mathbf{n}_i\}_{i=1}^{l}$ the set of their corresponding neutral images. The synthesis error for $\mathbf{x}$ is then obtained by averaging the synthesis error for $\{\mathbf{z}_i\}_{i=1}^{l}$ as

$$\mathbf{e}_s^x = \frac{1}{l}\sum_{i=1}^{l}(\mathbf{n}_i - \mathcal{N}(\mathbf{z}_i)) \quad (4)$$

The approximation of the neutral image of $\mathbf{x}$ is then obtained as

$$\mathcal{N}_e(\mathbf{x}) = \mathcal{N}(\mathbf{x}) + \mathbf{e}_s^x \quad (5)$$

where $\mathcal{N}_e(\mathbf{x})$ denotes the modified neutral synthesis function.

Fig. 5 shows some examples of the synthesized images. The first row contains the input images from the Cohn-Kanade database. These images belong to the few subjects who have given their permission for the use of their images in publications. The second row contains the synthesized images obtained by projecting the input image into the corresponding expression subspace, i.e., $\mathcal{H}(\mathbf{x})$, $\mathcal{N}(\mathbf{x})$, etc. The third row shows the synthesized images obtained by summing up the projected images and the estimated synthesis error for each input image, i.e., $\mathcal{H}_e(\mathbf{x})$, $\mathcal{N}_e(\mathbf{x})$, etc. The fourth row shows the ground-truth images. The number of the nearest neighbors l is 10 in this example. As seen in Fig. 5, the images in the third row, which are obtained using the modified synthesis function, better approximate the ground-truth images compared to the images in the second row.

Fig. 5. Examples of the synthesized images. The first (top) row: input images, the second row: projected images, the third row: projected images plus the synthesis error, and the fourth row: the ground-truth images.

3 Face Recognition Using Synthesized Images

In this section, we propose two face recognition methods. The first one is called the expression subspace projection (ESP) method, in which no estimation of the synthesis error is used. The second method is an extension of the first one, in which some statistics of the synthesis error is used to improve the estimation of the within-class scatter matrix. Although both methods are computationally simple and fast, the ESP method is much faster.

3.1 The ESP Method

Our proposed face recognition systems work in a generic learning framework. In these systems, generic training images are used to create the expression subspaces. New expression images are then synthesized from the gallery images as explained in Section 2.1. In the proposed ESP method, the generic training, gallery and synthesized images are then used to train LDA. LDA finds the discriminant projection bases by maximizing the Fisher-Rao's criterion [2] given by

$$\frac{|\mathbf{v}^T\mathbf{S}_b\mathbf{v}|}{|\mathbf{v}^T\mathbf{S}_w\mathbf{v}|} \tag{6}$$

where $\mathbf{v}$ is the projection basis, $\mathbf{S}_b = \sum_{i=1}^{c} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$ is the between-class scatter matrix, c is the number of classes, n_i is the number of the samples in class i, $\mathbf{m}_i$ is the mean of the samples of class i, and $\mathbf{m}$ is the global mean (mean of the samples of all classes). $\mathbf{S}_w = \sum_{i=1}^{c} \mathbf{S}_w^i$ is the within-class scatter matrix, where $\mathbf{S}_w^i = \sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \mathbf{m}_i)(\mathbf{x}_{ij} - \mathbf{m}_i)^T$ is the scatter matrix of class i and $\mathbf{x}_{ij}$ is the jth sample of that class.

When the generic set, gallery and synthesized images are used to train LDA, it can be shown from the definition of the within-class scatter matrix that, $\mathbf{S}_w$ of the gallery subjects is estimated from

$$\mathbf{S}_w^g \approx \mathbf{S}_w^t + \mathbf{S}_w^{g+s} \tag{7}$$

where $\mathbf{S}_w^g$ denotes the actual within-class scatter matrix of the gallery subjects (which is unknown), $\mathbf{S}_w^t$ denotes the within-class scatter matrix of the generic training images and $\mathbf{S}_w^{g+s}$ denotes the within-class scatter matrix of the gallery and synthesized images.

After obtaining the LDA projection bases, discriminant features of the gallery and probe images are extracted by projecting them into these bases. We observed that, the performance of an LDA-based face recognition system depends on the number of the LDA projection bases that is used. That is, using all of the generalized eigenvectors corresponding to all of the non-zero eigenvalues, as suggested by Belhumeur et al. [2], does not result in the best performance. We found through experiments that discarding the eigenvectors corresponding to the eigenvalues that are smaller than a threshold results in a better performance. The threshold that we used for the experiments in this paper is 0.01.

The cosine metric is used here for measuring the similarity between the feature vectors of the *galley and probe* images. We experimented with different commonly used similarity/distance measures, and concluded that cosine produces the best results. Note that the synthesized images are only used for training LDA and not for matching with the probe images. Fig. 6 shows the block diagram of the proposed ESP method.

3.2 The ESP-SEE Method

Instead of estimating the synthesis error for individual gallery images, we can use some statistics of the synthesis error to improve the estimation of the within-class scatter matrix as follows.

Theorem-The scatter matrix of class i can be written as

$$\mathbf{S}_w^i = \frac{1}{n_i}\sum_{j=1}^{\binom{n_i}{2}} \mathbf{d}_{ij}\mathbf{d}_{ij}^T \tag{8}$$

where $\mathbf{d}_{ij}$'s are pairwise distances of the samples of class i and $\binom{n_i}{2}$ denotes the number of two-combinations from a set of n_i elements. The proof of this Theorem is given in Appendix B. Fig. 7 illustrates the pairwise distances for a class with three samples.

Now, suppose we only have the happy image $\mathbf{h}$ of a gallery subject and we want to estimate the within-class scatter matrix for this subject using its synthesized neutral image. Fig. 8 illustrates the role of

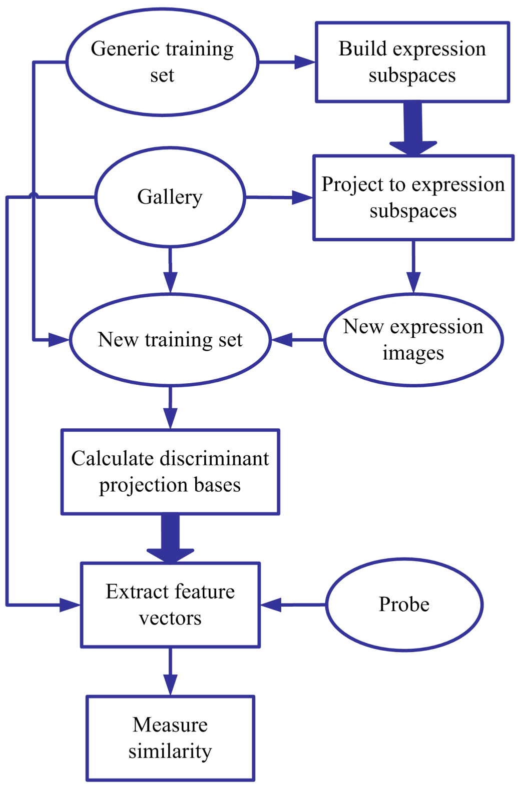


Fig. 6. Block diagram of the ESP method. Thick arrows convey projection bases.

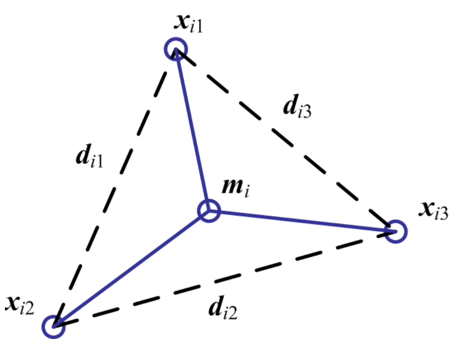


Fig. 7. Pairwise distances for a class with three samples.

a synthesized image in the estimation of the within-class scatter matrix. In this figure, the line represents the neutral subspace, the point $\mathbf{h}$ represents the gallery image, and the point $\mathbf{n}$ represents the true neutral image of that gallery subject which is unknown.(Although $\mathbf{n}$ may not lie exactly on the neutral subspace, we can approximate $\mathbf{n}$ with its projection into the neutral subspace given that the neutral subspace contains enough eigenvalue energy . That is, we can assume that $\mathbf{n}$ lies on the neutral subspace.) Also, the point $N(\mathbf{h})$ represents the synthesized neutral image, which is the approximation of $\mathbf{n}$. According to the theorem, the actual scatter matrix of class i is

$$\mathbf{S}_w^i = \frac{1}{2}\mathbf{d}\mathbf{d}^T \tag{9}$$

where as shown in Fig. 8, $\mathbf{d} = \mathbf{n} - \mathbf{h}$. In the ESP method, by using $\mathcal{N}(\mathbf{h})$ and $\mathbf{h}$ to train LDA, according to the theorem, the scatter matrix of class i is estimated from

$$\mathbf{S}_w^i \approx \frac{1}{2}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T \tag{10}$$

where $\tilde{\mathbf{d}} = \mathcal{N}(\mathbf{h}) - \mathbf{h}$, which we call the synthesized pairwise distance.

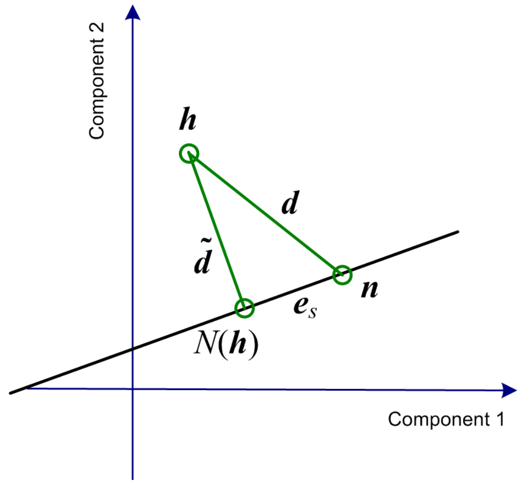


Fig. 8. Geometric view of a synthesized image.

The synthesis error is also depicted in Fig. 8. It is shown in Appendix C that, *the synthesis error is orthogonal to the synthesized pairwise distance*,

$$\tilde{\mathbf{d}} \perp \mathbf{e}_s \tag{11}$$

This orthogonality allows us to simply use some statistics of the synthesis error to improve the estimation of the within-class scatter matrix. As seen in Fig. 8, the pairwise distance $\mathbf{d}$ can be written as the summation of the synthesized pairwise distance and the synthesis error,

$$\mathbf{d} = \tilde{\mathbf{d}} + \mathbf{e}_s \tag{12}$$

Then, using (11) and (12), it can be easily shown that

$$\mathbf{d}\mathbf{d}^T = \tilde{\mathbf{d}}\tilde{\mathbf{d}}^T + \mathbf{e}_s\mathbf{e}_s^T \tag{13}$$

where the second term on the right represents the second moment of the synthesis error.

In the next section, we describe an approach for estimating the second moment of the synthesis error using a validation set. For now, let $\mathbf{M}_e = \mathbf{e}_s\mathbf{e}_s^T$ be the estimation of the second moment of the synthesis error. By substituting (13) in (8), it can be easily shown that, the within-class scatter matrix of the gallery subjects can be estimated as

$$\mathbf{S}_w^g \approx \mathbf{S}_w^{g+s} + (\sum_{i=1}^{c} \frac{n_i - 1}{2})\mathbf{M}_e \quad (14)$$

where c denotes the number of the gallery subjects and n_i denotes the number of the images of class i consisting of the gallery and synthesized images (i.e., $n_i - 1$ is the number of synthesized images for class i).

Finally, by adding the generic training images to the set of gallery and synthesized images, the estimation of the within-class scatter of the gallery subjects is obtained from

$$\mathbf{S}_w^g \approx \mathbf{S}_w^t + \mathbf{S}_w^{g+s} + (\sum_{i=1}^{c} \frac{n_i - 1}{2})\mathbf{M}_e \quad (15)$$

Once the within-class scatter matrix is calculated, the calculation of the between-class scatter matrix and the remaining steps are the same as those in the ESP method. Fig. 9 shows the block diagram of the ESP-SEE method.

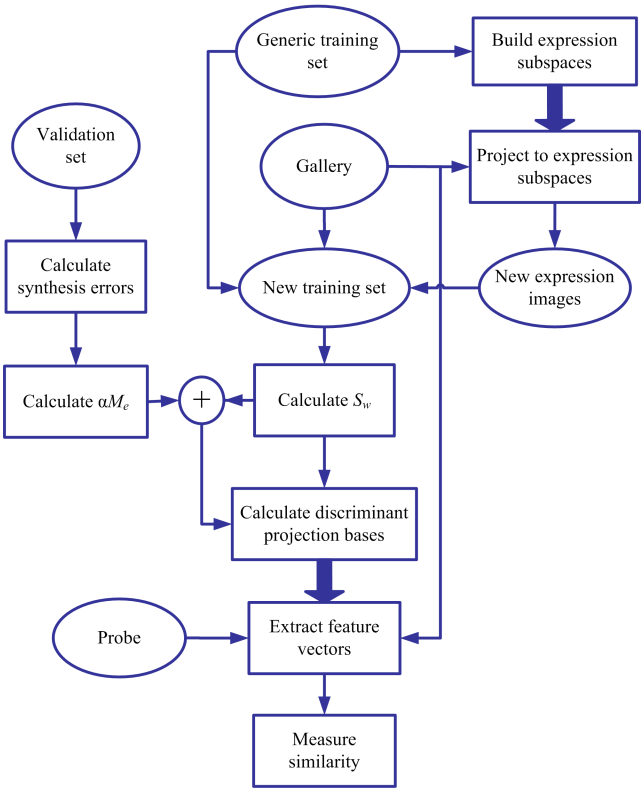


Fig. 9. Block diagram of the ESP-SEE method. Thick arrows convey projection bases.

3.2.1 *Estimating the Second Moment of the Synthesis Error*

We estimate the second moment of the synthesis error using a validation set as follows. To create a validation set, we divide the subjects of the generic training set into two sets: the training set, which is used to create the expression subspaces, and the validation set.

Let $\mathbf{n}_i$ and $\mathbf{h}_i$ be the neutral and happy images of class i in the validation set. Also, let $\mathcal{N}(\mathbf{h}_i)$ be the synthesized neutral image of $\mathbf{h}_i$. Then, the synthesis error is obtained as

$$\mathbf{e}_s^i = \mathbf{n}_i - \mathcal{N}(\mathbf{h}_i) \quad (16)$$

The estimation of the second moment of the synthesis error can then be obtained by averaging as

$$\mathbf{M}_e = \frac{1}{\sum_{i=1}^{V} w_i} \sum_{i=1}^{V} \sum_{j=1}^{w_i} \mathbf{e}_s^{ij}(\mathbf{e}_s^{ij})^T \quad (17)$$

where V is the number of validation subjects, $\mathbf{e}_s^{ij}$'s are the synthesis errors for class i, and w_i is the number of the synthesis errors obtained for class i.

4 Experimental Results

In this section, we demonstrate the performance of our face recognition methods and compare them with the generic learning method proposed by Wang et al. [17], which to the best of our knowledge produces the best results compared to other published methods. This method, in which LDA is trained using only the images of the generic people, is called here as "generic LDA".

We also compare the performance of our proposed methods with the adaptive generic learning method [21], which has been recently proposed to address the problem of applying DA methods under SS scenario. In this method, the within-class variability and the class mean of the subjects with SS are inferred by a predicting model learned from the generic training set. The overall within-class and between-class scatter matrices for the gallery subjects are then estimated using the predicted within-class variability and class means.

We denote our proposed framework for training LDA as "GGS LDA", as in that, LDA is trained using the images of a generic training set, the gallery and the synthesized images from the gallery. This framework can be used in conjunction with other expression synthesis/transformation methods.

To the best of our knowledge, the only expression synthesis/transformation method existing in the literature, is the transformation method of Lee et al. [15], in which the tensorface concept [16] is used to transform the expression of an image. In the face recognition method proposed by Lee et al. [15], the gallery images are assumed to be neutral and the probe images are assumed to be non-neutral. The expression of the probe images is then transformed to neutral to match them with the gallery images. Lee et al. used various neutral images per subject to train LDA for recognition. Therefore, in fact, they tried to solve the problem of recognizing from single expression, but not recognizing from SS. We found through experiments that the performance of their method, when it is used for recognition from SS, is very poor.

We implemented the transformation method of Lee et al. and used it in our GGS-LDA framework (denoted here as "Lee+GGS LDA"), in order to compare the performance of our proposed synthesis method with their transformation method.

4.1 Datasets

We performed our experiments on the Cohn-Kanade AU-Coded Facial Expression database [20] and the Face Recognition Grand Challenge (FRGC) v2.0 [18] database, which are, to the best of our knowledge, the most comprehensive facial expression databases with a large number of subjects and various expressions.

The FRGC database contains 4007 images from 464 subjects displaying various expressions. The most displayed expressions are neutral, happy, surprise and puffy cheeks.

The Cohn-Kanade database contains video sequences from 97 people, each performing a series of one to six facial expressions. The facial expressions include happy, surprise, fear, disgust, anger and sadness. Each subject might have one to six video sequences and each video sequence starts with a neutral expression and ends to a target non-neutral expression. The last frame in each video sequence displays the target expression with its utmost intensity.

In order to use the Cohn-Kanade database, we created an image dataset from the video sequences in this database. This dataset consists of two to seven images from each subject with different expressions. For each subject, we selected the last frame of each of its video sequences as the non-neutral images and the first frame of one of its videos sequences as the neutral image. The resulting dataset consists of 497 images from 96 subjects with as many as seven expressions per subject.

In order to align face images, we detected the eye region of the images using the PCA technique described in [22]. In order to implement this method, first we manually detect the eye coordinates in a few images to be used for training. We then rotate, scale and shift the training images so that their eye centers locate at the same pixel coordinates across all of the images. The resulting eye regions are then used to construct a PCA space representing the eye region (so-called the eigen-eye space). Now, in order to detect the eye region in an image, we search for the region that has the smallest distance from the eigen-eye space, i.e., the smallest reconstruction error. The candidate eye region is also required to be rotated and scaled until its reconstruction error is minimized. After the eye region is detected, the same rotation and scaling is applied to the entire image and the image is shifted so that its eye region locates at the same pixel coordinates across all of the images.

After aligning face images, we mask and normalize them to have zero mean and unit variance. The normalization is required to reduce the illumination variation of the images.

4.2 Experimental setup

For each dataset, we randomly selected one third of the subjects as the test set, one third as the validation set and one third as the generic training set. Note that, there was no overlap between the subjects used for these three sets. We also experimented with different numbers of subjects (out of the selected one third) for the generic training set to investigate the effect of the size of this set on each algorithm.

We formed the gallery and probe sets using the images of the test subjects as described in the next paragraph. We then performed face recognition using each algorithm by calculating the similarity scores for every pair of gallery and probe images. We repeated the random division of the dataset and performing face recognition for 60 times. Finally, for each algorithm, we calculated the verification rates at different thresholds using the similarity scores over all the trials.

In order to provide a comprehensive analysis, we performed the face recognition test on each pair of expressions separately. For example, once we used only neutral and happy expressions for the test set. In this case, for each testing subject, we randomly selected one of its happy or neutral images for the gallery set. If the gallery image for that subject was a happy one, we chose its neutral image for the probe set and vice versa. Note that, only one image per subject is used to form the gallery set. For the sake of brevity, we only report experiments where either the gallery or probe image for each subject is neutral. However, in applying the proposed method, both of the gallery and probe images can be non-neutral. Note that, this means that unlike many face recognition systems, enrollment is not required to be performed under neutral expression.

4.2.1 *Calculating the Statistical Significance of the Results*

In order to determine whether the difference between the performance rates of two methods is significant, we perform *hypothesis testing* [23]. Let x_1 and x_2, where $x_1 > x_2$, be the number of true positives for the methods A and B, respectively and n be the total number of positive samples (i.e., $\frac{x_1}{n}$ and $\frac{x_2}{n}$ are the verification rates.). Also, let p_1 and p_2 be the actual probabilities of correctly verifying a positive sample using the methods A and B, respectively. We wish to test the assumption that $p_1 - p_2 = 0$ (the *null hypothesis*) against the assumption that $p_1 > p_2$ (the *alternative hypothesis*). That is, we wish to test whether the observed numbers of true positive samples x_1 and x_2 for the given number of samples n support the rejection of the null hypothesis. If these numbers support the rejection of the null hypothesis, we conclude that the difference between the observed verification rates is significant.

The number of true positives x for a given method can be viewed as a random variable with binomial distribution, where the probability of success p is the probability of correctly verifying a positive sample and the number of trials is equal to the total number of positive samples n. We know that, for a sufficiently large sample size n, the binomial distribution converges to the normal distribution $N(np, np(1-p))$. Let x_1 and x_2 be two random variables with binomial distributions with probabilities of success p_1 and p_2 and the numbers of trials n_1 and n_2, respectively. It can be shown that, under the assumption that $p_1 = p_2$,

$$q = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{(\frac{n_1+n_2}{n_1 n_2})(\frac{x_1+x_2}{n_1+n_2})(1 - \frac{x_1+x_2}{n_1+n_2})}} \tag{18}$$

has approximately a standard normal distribution (i.e., $N(0,1)$). We use q as the *test statistic* for our hypothesis testing. In our case, $n_1 = n_2$, and therefore, q simplifies as

$$q = \frac{x_1 - x_2}{\sqrt{(\frac{(x_1+x_2)(2n-x_1-x_2)}{2n})}} \tag{19}$$

In the hypothesis testing, we find a region on the real line where under the null hypothesis, the density of the test statistic is negligible. This region is called the *critical region* of the test. If the observed test statistic falls in the critical region, we reject the null hypothesis. The critical region is obtained according to the chosen level of significance. For 95% level of significance, the critical region for our hypothesis test is $q > z_{0.95}$, where $z_{0.95}$ is the standard normal percentile and is equal to 1.64.

4.3 Results

Fig. 10 shows the ROC curves for the experiment on the Cohn-Kanade database using both of our methods: ESP and ESP-SEE, and the generic LDA ("Gen LDA"). The number of the subjects in the generic training set was 30 for this experiment. We were unable to apply the transformation method of Lee et al. [15] for the Cohn-Kanade database because this method requires every training subject to have an image of each expression. In both of the Cohn-Kanade and FRGC databases, each subject may not have images of every expression. We need a large number of generic training subjects in order to have enough subjects with every expression. As seen in Fig. 10, our proposed methods achieve significant improvement over the generic LDA.

Fig. 11 shows the ROC curves for the experiment on the FRGC database using the two proposed methods and the generic LDA. The number of the subjects in the generic training set was 30 for this experiment. Again it can be seen that, the proposed methods achieve significant improvement over the generic LDA.

Fig. 12 shows the ROC curves for the experiment on the FRGC database using 100 subjects for the generic training set. In this case, we were able to apply the transformation method of Lee et al. As mentioned before, for applying this method, we used our proposed GGS-LDA framework. As it can be seen in Fig. 12, the Lee+GGS-LDA method outperforms the generic LDA, showing the success of our GGS-LDA framework in conjunction with other transformation/synthesis methods. However, our proposed synthesis methods outperform the transformation method of Lee et al. Other advantages of our proposed synthesis method over the transformation method of Lee et al. are: (1) the simplicity of the proposed synthesis methods, i.e., the expression of an image is transformed simply by projecting it into another subspace, (2) it is not required to determine the expression of the input image, and (3) the algorithm is not required to be trained on a training set in which every subject has an image of each expression.

For implementing the proposed ESP-SEE method, we estimated the second moment of the synthesis error using the validation set. We also experimented with estimating the synthesis error for each individual image using its nearest neighbors from the validation set and then training LDA using the modified synthesized images (as described in Section 2.2). The performance of the ESP-SEE method using these two implementation approaches is the same, but the former requires much less computations.

To summarize the results, the verification rate at 0.1% false acceptance rate (FAR) for each expression on the Cohn-Kanade and FRGC database are shown in Tables 1 and 2, respectively. Also, the results using some other sizes of the generic training set are shown in these tables.

The verification rates using the adaptive generic learning method (denoted as "Adapted LDA") are also shown in Tables 1 and 2. While the adapted LDA method outperforms the generic LDA on the FERET and a passport database as reported in [21], the generic LDA outperforms the adapted LDA method here on the FRGC and Cohn-Kanade database. One possible explanation for this observation is that, the adapted LDA method might be more suitable for situations with less expression variations as it is the case for the images in the FERET and the passport

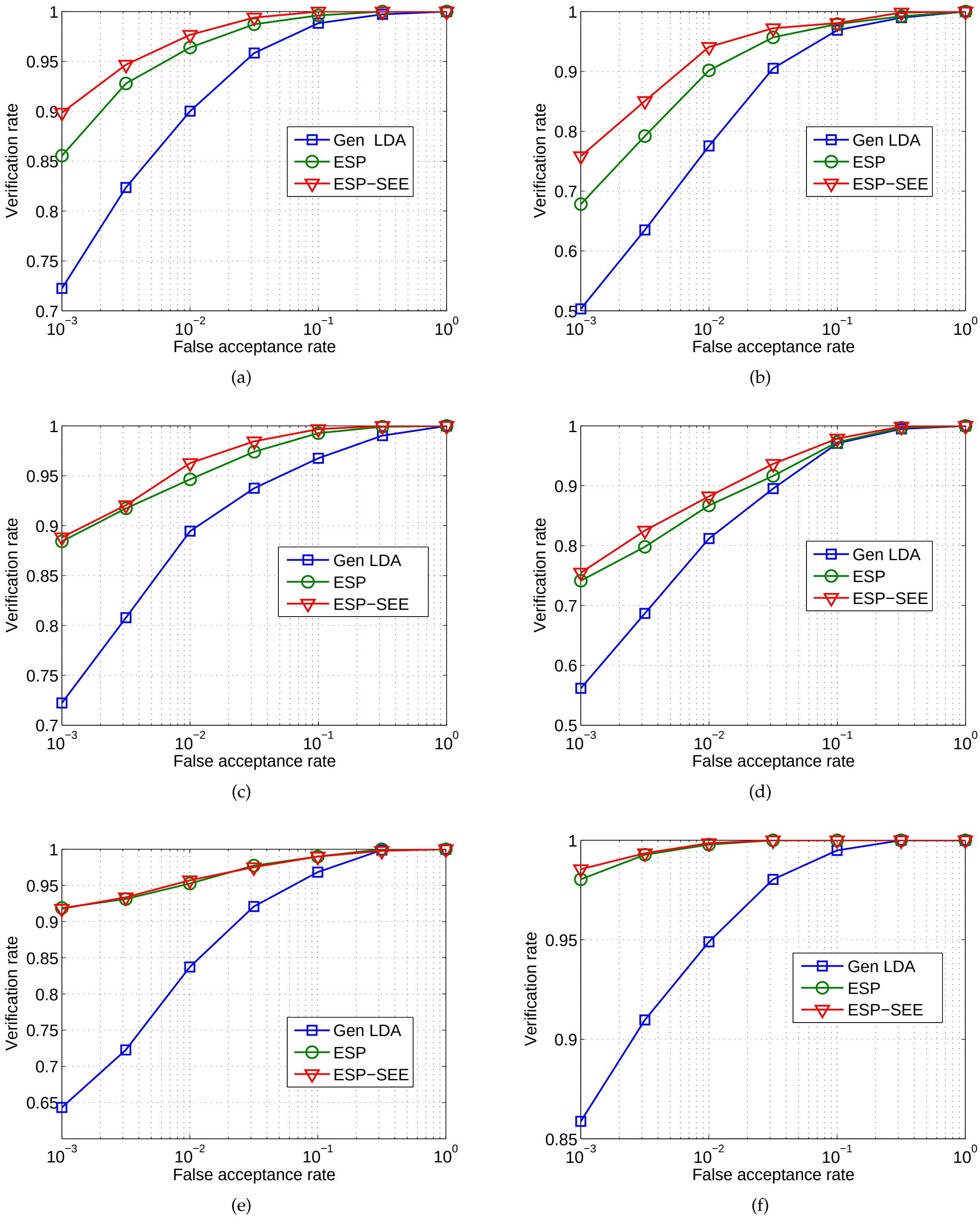
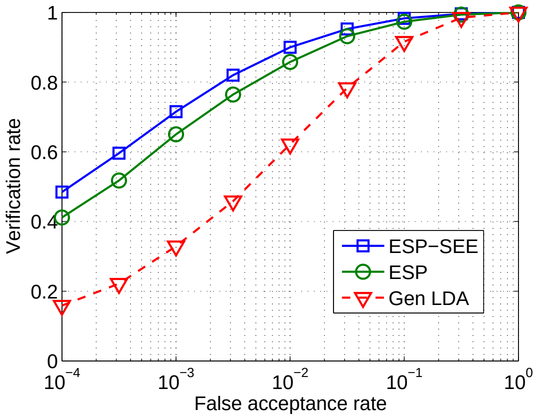
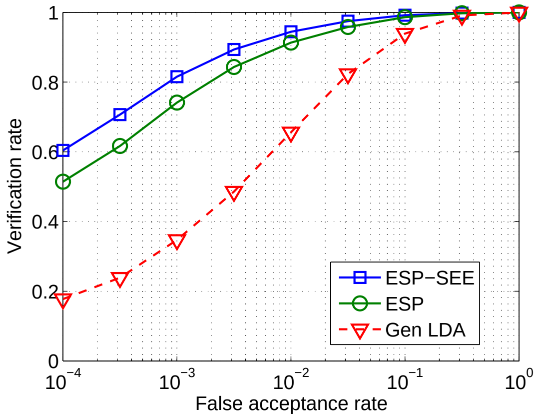


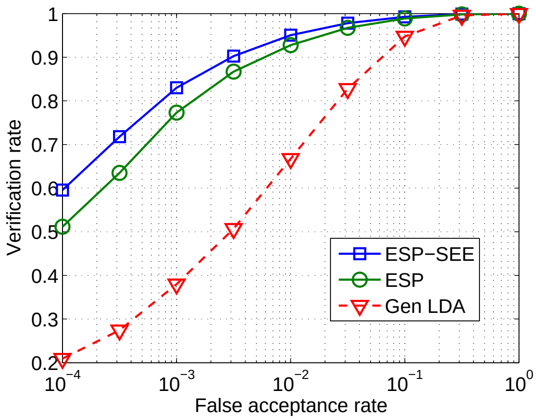
Fig. 10. ROC curves for the two proposed methods and generic LDA method on the Cohn-Kanade database. The number of subjects in the generic training set is 30. Expressions used for the test set are: (a) neutral and happy, (b) neutral and surprise, (c) neutral and fear, (d) neutral and disgust, (e) neutral and anger, and (f) neutral and sadness.



(a)



(b)



(c)

Fig. 11. ROC curves for the two proposed methods and the generic LDA on the FRGC database. The number of subjects in the generic training set is 30. Expressions used for the test set are: (a) neutral and happy, (b) neutral and surprise, and (c) neutral and puffy cheeks.

database used for the experiments in [21].

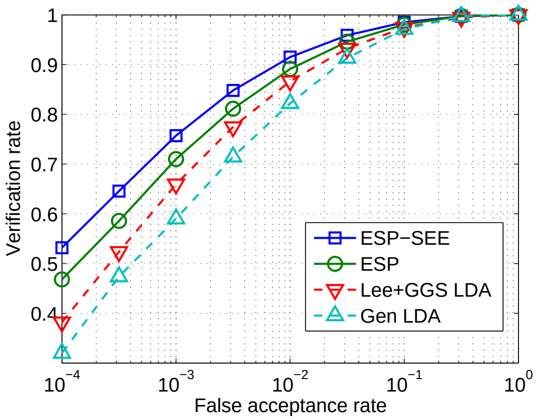
To determine the statistical significance of the results, the test statistic for our proposed methods versus the generic LDA and Lee+GGS-LDA methods are shown in Tables 3 and 4. The decision to reject or accept the null hypothesis is based on the 95% level of significance, i.e., if $q > z_{0.95} = 1.64$, we reject the null hypothesis and conclude that the improvement in the performance is significant, otherwise, we conclude that the observed results do not support the rejection of the null hypothesis (i.e., accept).

To calculate the test statistic (see (19)) for each experiment, we set the total number of positive samples n equal to the total number of subjects in the database that display the corresponding expression (this number is shown inside the brackets in Tables 3 and 4). It should be noted that, although in each trial of our resampling approach, one third of the subjects are used for the test set, by repeating the trials for numerously enough number of times, we can be confident that all of the subjects in the database have been used for the test set. Moreover, for calculating the verification rates, we concatenate the similarity scores from all of the trials. Therefore, the number of independent positive samples for each experiment is equal to the total number of subjects in the database that display the corresponding expression.

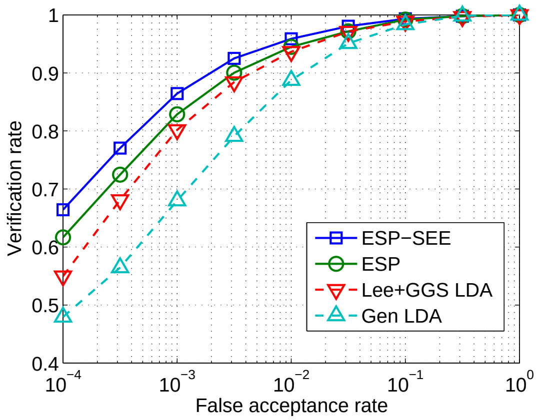
As shown in Tables 3 and 4, the performance improvement achieved by the proposed methods over the generic LDA and Lee+GGS-LDA methods is significant in most of the experiments. The only exception is the experiment using the puffy-cheeks expression on the FRGC database for our proposed ESP-SEE method versus Lee+GGS-LDA method, which is due to the insufficient number of subjects in the database which display this expression. However, the performance improvement in this case is significant at 90% level of significance ($q > z_{0.90} = 1.28$).

In addition, in Tables 1 and 2, we have shown the verification results using PCA. The low verification rates for PCA shows the difficulty of recognition under the SS scenario using unsupervised methods.

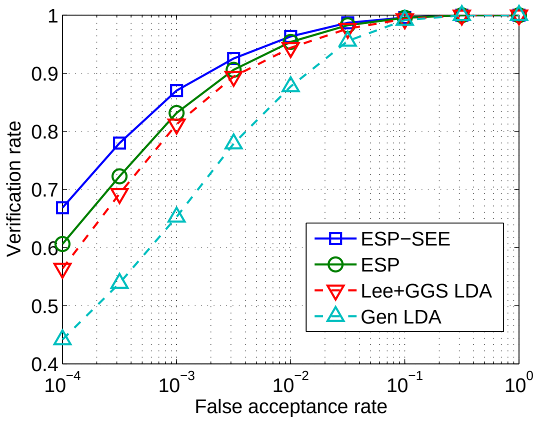
As shown in Table 2, recognizing happy images is more difficult than the other expressions in the FRGC



(a)



(b)



(c)

Fig. 12. ROC curves for the two proposed methods, generic LDA, and Lee+GGS LDA on the FRGC database. The number of subjects in the generic training set is 100. Expressions used for the test set are: (a) neutral and happy, (b) neutral and surprise, and (c) neutral and puffy cheeks.

TABLE 1
Verification rates at 0.1% FAR for the Cohn-Kanade database.

Expression	Training Subjects	PCA	Gen LDA	Adapted LDA	Lee+GGS LDA	ESP	ESP-SEE
Happy	20	0.04	0.57	0.48	NA	0.85	0.90
	30	0.06	0.72	0.61	NA	0.85	0.90
Surprise	20	0.00	0.34	0.28	NA	0.55	0.64
	30	0.01	0.50	0.46	NA	0.68	0.76
Fear	20	0.21	0.58	0.45	NA	0.85	0.87
	30	0.26	0.72	0.56	NA	0.88	0.89
Disgust	20	0.08	0.46	0.29	NA	0.73	0.77
	30	0.11	0.56	0.40	NA	0.74	0.77
Anger	20	0.35	0.60	0.43	NA	0.90	0.90
	30	0.43	0.64	0.54	NA	0.92	0.92
Sadness	20	0.47	0.78	0.68	NA	0.98	0.98
	30	0.47	0.86	0.75	NA	0.98	0.98
Average	20	0.19	0.55	0.43	NA	0.81	0.84
	30	0.22	0.67	0.55	NA	0.84	0.87

TABLE 2
Verification rates at 0.1% FAR for the FRGC ver2.0 database.

Expression	Training Subjects	PCA	Gen LDA	Adapted LDA	Lee+GGS LDA	ESP	ESP-SEE
Happy	30	0.08	0.33	0.31	NA	0.65	0.72
	50	0.08	0.46	0.41	NA	0.70	0.76
	100	0.10	0.59	0.53	0.66	0.71	0.76
	150	0.10	0.61	0.58	0.66	0.71	0.76
Surprise	30	0.11	0.35	0.33	NA	0.74	0.82
	50	0.13	0.50	0.46	NA	0.78	0.84
	100	0.16	0.68	0.63	0.80	0.83	0.86
	150	0.16	0.74	0.71	0.82	0.84	0.88
Puffy Cheeks	30	0.17	0.38	0.35	NA	0.77	0.83
	50	0.18	0.52	0.45	NA	0.80	0.86
	100	0.21	0.65	0.60	0.81	0.83	0.87
	150	0.22	0.70	0.63	0.81	0.83	0.87
Average	30	0.12	0.35	0.33	NA	0.72	0.79
	50	0.13	0.49	0.44	NA	0.76	0.82
	100	0.15	0.64	0.59	0.76	0.79	0.83
	150	0.16	0.68	0.64	0.76	0.79	0.84

TABLE 3
Statistical significance for the verification results on the Cohn-Kanade database.

Expression (No. of Subjects)	Training Subjects	ESP vs. Gen LDA	ESP-SEE vs. Gen LDA
Happy (85)	20	4.00 (Reject)	4.87 (Reject)
	30	2.05 (Reject)	3.00 (Reject)
Surprise (88)	20	2.80 (Reject)	3.98 (Reject)
	30	2.42 (Reject)	3.57 (Reject)
Fear (61)	20	3.29 (Reject)	3.57 (Reject)
	30	2.22 (Reject)	2.36 (Reject)
Disgust (47)	20	2.66 (Reject)	3.09 (Reject)
	30	1.82 (Reject)	1.92 (Reject)
Anger (38)	20	3.01 (Reject)	3.01 (Reject)
	30	2.92 (Reject)	2.92 (Reject)
Sadness (79)	20	3.86 (Reject)	3.86 (Reject)
	30	2.74 (Reject)	2.74 (Reject)

TABLE 4
Statistical significance for the verification results on the FRGC ver2.0 database.

Expression (No. of Subjects)	Training Subjects	ESP vs. Gen LDA	ESP-SEE vs. Gen LDA	ESP-SEE vs. Lee+GGS LDA
Happy (246)	30	7.09 (Reject)	8.66 (Reject)	NA
	50	5.39 (Reject)	6.82 (Reject)	NA
	100	2.79 (Reject)	4.02 (Reject)	2.44 (Reject)
	150	2.34 (Reject)	3.58 (Reject)	2.44 (Reject)
Surprise (215)	30	8.11 (Reject)	9.88 (Reject)	NA
	50	6.05 (Reject)	7.50 (Reject)	NA
	100	3.60 (Reject)	4.43 (Reject)	1.66 (Reject)
	150	2.54 (Reject)	3.69 (Reject)	1.73 (Reject)
Puffy Cheeks (122)	30	6.15 (Reject)	7.18 (Reject)	NA
	50	4.61 (Reject)	5.73 (Reject)	NA
	100	3.19 (Reject)	4.01 (Reject)	1.29 (Accept)
	150	2.38 (Reject)	3.21 (Reject)	1.28 (Accept)

database. In the Cohn-Kanade database, the most difficult expression for recognition is "disgust" and the easiest one is "sadness". Clearly, the level of the deformation of the face is highest under the disgust expression and lowest under the sadness expression.

The intensity of the surprise expression is much higher in the Cohn-Kanade database compared to the FRGC database, which resulted in higher verification rates for the latter. Another difference between the two databases is that, for the Cohn-Kanade database, all the images of each subject were taken under the same illumination condition. This makes the recognition task easier compared to when there is illumination variation between images of each subject, which is the case in the FRGC database. It might be for this reason that the verification rates for the happy expression are higher for the Cohn-Kanade database compared to that for the FRGC database.

It is clear from the verification rates shown in Tables 1 and 2 that, the proposed methods are very useful especially when the size of the generic training set is small. When the number of the subjects in the generic training set is small, the estimation of the within-class scatter matrix using only the generic training set is considerably poor and the role of the synthesized images is more pronounced. On the other hand, as the number of the subjects in the generic training set increases, this set contains more variability of subjects, and therefore, the within-class variability of the gallery subjects can be better estimated using only the generic training set.

To see how the estimation of the within-class scatter matrix using the synthesized images converges to the actual within-class scatter matrix of the gallery images, we calculated the difference between these two matrices using various numbers of training subjects. Fig. 13 shows this difference for the FRGC database when happy and neutral expressions are used for the test set. The plot for other pairs of expressions is similar. To calculate the actual within-class scatter, we used both the gallery and probe image for each subject. To measure the difference between the synthesized and actual within-class scatter matrix for each subject, we summed up the square of the element-wise differences between the two matrices and divided by the total number of the elements. We then averaged the result over all of the testing subjects. As it can be seen in Fig. 13, the matrix difference decreases by increasing the number of the training subjects, which shows the convergence of the synthesized within-class scatter matrix to the actual one.

For performing the above experiments, the expression subspaces were constructed by using 99% of the eigenvalue energy. To investigate the effect of the proportion of the eigenvalue energy on the recognition performance, we experimented with the other values of this proportion. Fig. 14 shows the results for the ESP-SEE method on the FRGC database when the happy and neutral expressions are used for the test set. As it is seen, by increasing the eigenvalue energy better recognition performance has been achieved, showing that better approximation of the unavailable expression images can be obtained.

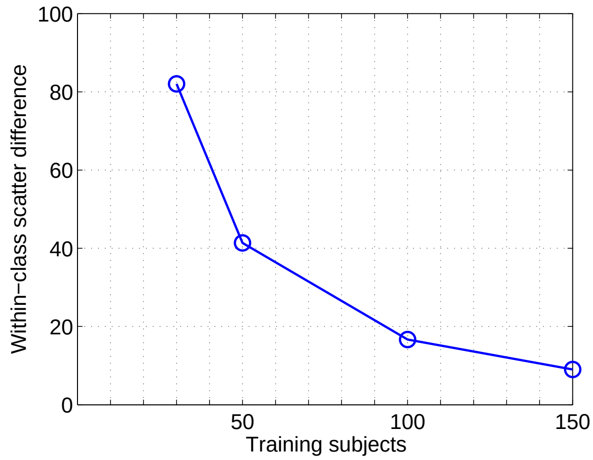


Fig. 13. Difference between the synthesized and actual within-class scatter matrix versus various number of training subjects.

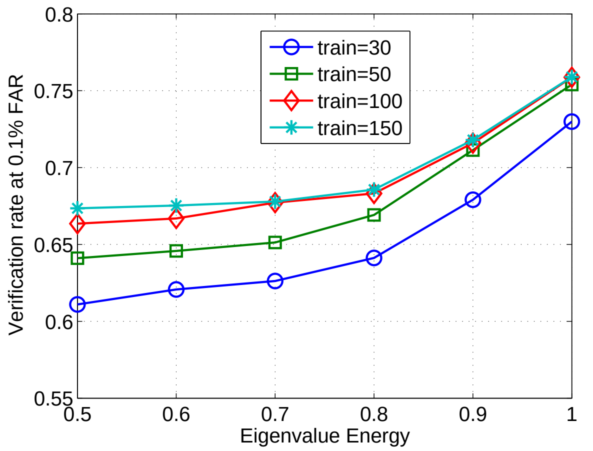


Fig. 14. Verification rate at 0.1% FAR versus the eigenvalue energy of the expression subspace using the ESP-SEE method on the FRGC database.

5 Conclusion

In this paper, we introduced expression subspaces for approximating new expression images from one image of a subject. We then proposed two methods to improve recognition of expression-variant faces from SS. The ESP method uses the synthesized images along with the gallery images and a generic set of images to train a DA algorithm. To improve the performance of ESP, the ESP-SEE method was proposed which uses a validation set to estimate the second moment of the synthesis error. This estimation is used to improve the estimation of the within-class scatter matrix obtained from synthesized images. Using the FRGC and the Cohn-Kanade databases, we demonstrated that the proposed methods significantly improve the recognition rate compared to the existing methods and do not need the error-prone tasks required by them.

An important advantage of the proposed method is its simplicity; the expression of an image is transformed simply by projecting it into another subspace. Another advantage of the proposed methods is that both the gallery and probe images can have non-neutral expressions.

As a straightforward DA method, we used LDA in this work. We expect that the use of more advanced DA methods can provide more improvement using the proposed approach.

The proposed solution can address other face recognition problems such as recognizing occluded faces, faces wearing glasses and age-variant faces. It can be also used to address general pattern recognition problems, e.g., for the problem of recognizing objects that are misclassified under similar circumstances.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, pp. 399–459, 2003.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

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

[4] M. Loog and R.P.W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The chernoff criterion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 732–739, 2004.

[5] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *Annals of Statistics*, vol. 23, pp. 73–102, 1995.

[6] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385–2404, 2000.

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

[8] M. Zhu and A.M. Martinez, "Subclass discriminant analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274–1286, 2006.

[9] Y. Liu, K. Schmidt, J. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *Computer Vision and Image Understanding*, vol. 91, pp. 138–159, 2003.

[10] H. Mohammadzade and D. Hatzinakos, "An expression transformation for improving the recognition of expression-variant faces from one sample image per person," *Proc. IEEE Int. Conf. Biometrics: Theory Applications and Systems*, pp. 1–6, 2010.

[11] H. Mohammadzade and D. Hatzinakos, "Correspondence normal difference: An aligned representation of 3D faces to apply discriminant analysis methods," *Proc. Int. Conf. Digital Signal Processing*, pp. 1–8, 2011.

[12] X. Tan, S. Chen, Z. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition*, vol. 39, pp. 1725–1745, 2006.

[13] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 748–763, 2002.

[14] X. Li, G. Mori, and H. Zhang, "Expression-invariant face recognition with expression classification," *Proc. Third Canadian Conf. Computer and Robot Vision*, p. 77, 2006.

[15] H.S. Lee and D. Kim, "Expression-invariant face recognition by facial expression transformations," *Pattern Recognition Letters*, vol. 29, pp. 1797–1805, 2008.

[16] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," *Proc. European Conf. Computer Vision*, pp. 447–460, 2002.

[17] J. Wang, K.N. Plataniotis, J. Lu, , and A.N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognition*, vol. 39, pp. 1746–1762, 2006.

[18] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 947–954, 2005.

[19] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 37, no. 1, pp. 2–86, 1991.

[20] T. Kanade, J.F. Cohn, and Y.L. Tian, "Comprehensive database for facial expression analysis," *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, 1999.

[21] Y. Su, S. Shan, X. Chen, and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2699–2706, 2010.

[22] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 19, pp. 696–710, 1994.

[23] A. Papoulis and S.U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 4th edition, 2002.

Hoda Mohammazade received the BSc degree (with honours) from Amirkabir University of Technology (Tehran Polytechnic), Iran, in 2004 and the MSc degree from the University of Calgary, Canada, in 2007, both in Electrical Engineering. She is currently pursuing her PhD degree in Electrical Engineering at the University of Toronto, Canada. Her research interests include signal and image processing, pattern recognition and biometric systems.

Dimitrios Hatzinakos received the Diploma degree from the University of Thessaloniki, Greece, in 1983, the M.A.Sc degree from the University of Ottawa, Canada, in 1986 and the Ph.D. degree from Northeastern University, Boston, MA, in 1990, all in Electrical Engineering. In September 1990 he joined the Department of Electrical and Computer Engineering, University of Toronto, where now he holds the rank of Professor with tenure. Since November 2004, he is the holder of the Bell Canada Chair in Mutimedia, at the University of Toronto. Also, he is the co-founder and since 2009 the Director and the chair of the management committee of the Identity, Privacy and Security Institute (IPSI) at the University of Toronto. His research interests and expertise are in the areas of Multimedia Signal Processing, Multimedia Security, Multimedia Communications and Biometric Systems. Since 2008, he is an Associate Editor for the IEEE Transactions on Mobile Computing. He is a senior member of the IEEE, a Fellow of the Engineering Institute of Canada (EIC), and member of EURASIP, the Professional Engineers of Ontario (PEO), and the Technical Chamber of Greece.