Fractional order singular value decomposition representation for face recognition

Jun Liu, Songcan Chen*, Xiaoyang Tan

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

Received 1 August 2006; received in revised form 28 January 2007; accepted 26 March 2007

Abstract

Face representation (FR) plays a typically important role in face recognition and methods such as principal component analysis (PCA) and linear discriminant analysis (LDA) have been received wide attention recently. However, despite of the achieved successes, these FR methods will inevitably lead to poor classification performance in case of great facial variations such as expression, lighting, occlusion and so on, due to the fact that the image gray value matrices on which they manipulate are very sensitive to these facial variations. In this paper, we take notice of the facts that every image matrix can always have the well-known singular value decomposition (SVD) and can be regarded as a composition of a set of base images generated by SVD, and we further point out that the leading base images (those corresponding to large singular values) on one hand are sensitive to the aforementioned facial variations and on the other hand dominate the composition of the face image. Then based on these observations, we subtly deflate the weights of the facial variation sensitive base images by a parameter $\alpha$ and propose a novel fractional order singular value decomposition representation (FSVDR) to alleviate facial variations for face recognition. Finally, our experimental results show that FSVDR can: (1) effectively alleviate facial variations; and (2) form an intermediate representation for many FR methods such as PCA and LDA to significantly improve their classification performance in case of great facial variations.

Keywords: Singular value decomposition (SVD); Fractional order singular value decomposition representation (FSVDR); Face representation (FR); Intermediate representation (IR); Face recognition

1. Introduction

Although we human beings can easily detect and identify faces in a scene, it is very challenging for an automated system to achieve such objectives. The challenges become more profound when large variations exist in the face images at hand, e.g., variations in illumination conditions, viewing directions or poses, facial expression, aging, and disguises such as facial hair, glasses, cosmetics and scarves. Despite of these challenges, face recognition has drawn wide attention from researchers in areas of machine learning, computer vision, pattern recognition, neural networks, and so on, thanks to the great need for face recognition in areas of access control, information security, law enforcement and surveillance, smart cards and so on [1–4].

Over the past decades, geometric feature-based methods [5–11] and appearance-based methods [12–15] are the two mainly employed face recognition methods. Geometric feature-based methods extract the relative position and other parameters of distinctive features such as eyes, mouth, nose, and chin as features, while appearance-based methods directly manipulate on the gray level values of the image pixels (e.g., a face image pattern is represented as an $r \times c$ matrix $A$, where $r$ and $c$ are, respectively, the numbers of rows and columns) and employ statistical tools to extract features for subsequent classification. Recently, it has been witnessed a strong trend away from geometry towards statistical and appearance-based models for face recognition [16], and this trend is supported by psychological, physiological, and biological studies dealing with vision in humans and animals [16–18]. A typical appearance-based face recognition scheme is given in Fig. 1, where the input is a face gray value matrix and the output is the given face’s class label. Generally speaking, this recognition scheme can be divided into two sequential stages: face
A natural choice for FR is the original gray value matrix (OGVM), whose classification scheme is depicted in route a in Fig. 1. Although OGVM is simple, it will encounter the following two problems: (1) the dimensionality of OGVM is much higher than the number of training samples (e.g., the dimensionality of a $100 \times 100$ image matrix is 10,000, while the number of training samples is often quite less) which leads to the so-called curse of dimensionality and consequently depresses the generalization ability of the correspondingly trained classifier; and (2) OGVM is very sensitive to facial variations such as expression, lighting, occlusion, etc., and thus will obtain poor classification performance in case of the aforementioned facial variations.

To obtain a good representation for face images, researchers have proposed many renowned dimensionality reduction (DR) methods that directly manipulate on OGVM (illustrated in route b in Fig. 1). The two most well-known DR methods are principal component analysis (PCA) [13,19] and linear discriminant analysis (LDA) [20,21].

Kirby and Sirovich [19] showed that any particular face can be (1) economically represented along the eigenpictures coordinate space, and (2) approximately reconstructed using just a small collection of eigenpictures and their corresponding projections (‘coefficients’). Turk and Pentland [13] applied PCA technique to face recognition, and proposed the well-known eigenfaces method. A recent major improvement on PCA is to directly manipulate on two-dimensional matrices (not one-dimensional vectors as in traditional PCA), e.g., two-dimensional PCA (2DPCA) [22], generalized low rank approximation of matrices [23], non-iterative generalized low rank approximation of matrices (NIGLRAM) [24] and so on. The advantages of manipulating on two-dimensional matrices rather than one-dimensional vectors are [22]: (1) it is simpler and straightforward to use for image feature extraction; (2) it is better in terms of classification performance; and (3) it is computationally more efficient. Based on the viewpoint of minimizing reconstruction error, the above PCA-based methods [13,22–24] are unsupervised methods that do not take the class labels into consideration.

Taking the class labels into consideration, LDA aims at projecting face samples to a subspace where the samples belonging to the same class are compact while those belonging to different classes are far away from each other. The major problem in applying LDA to face recognition is the so-called small sample size (SSS) problem (namely, the number of samples is far less than sample dimensionality), which leads to the singularities of the within-class and between-class scatter matrices. Recently, researchers have exerted great endeavor to deal with this problem. In Refs. [25,26], a PCA procedure was applied prior to the LDA procedure, which led to the well-known PCA+LDA or Fisherfaces method. In Refs. [27,28], samples were first projected to the null space of the within-class scatter matrix and then LDA was applied in this null space to yield the optimal (infinite) value of the Fisher’s linear discriminant criterion, which led to the so-called discriminant common vectors (DCV) method. In Refs. [29,30], LDA was applied in the range space of the between-class scatter matrix to deal with the SSS problem, which led to the LDA via QR decomposition (LDA/QR) method.

These DR methods have been proven to effectively lower the dimensionality of OGVM. Furthermore, in face recognition, PCA and LDA have become de-facto baseline approaches. However, despite of the achieved successes, these FR methods will inevitably lead to poor classification performance in case of great facial variations such as expression, lighting, occlusion and so on, due to the fact that the OGVM on which they manipulate is very sensitive to these facial variations. To mitigate this problem, in this paper, we propose a novel fractional order singular value decomposition representation (FSVDR), which acts as an intermediate representation (IR) between OGVM and DR (see route c in Fig. 1) for face recognition.

In literature, there have been a number of approaches that form an IR between OGVM and DR for face recognition. In Refs. [31,32], the authors first obtained the Gabor wavelet representation for face images and then utilized it as an IR for sub-
sequent DR utilizing the enhanced Fisher linear discriminant model [33] or independent component analysis [34]. The Gabor wavelet representation was reported to capture local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity. Besides Gabor wavelet, Er et al. [35] employed discrete cosine transforms (DCT) as an IR and then applied LDA subsequently for DR. It was reported that the low-frequency DCT coefficients accounted for a large area nonuniform illumination variations, and consequently the nonuniform illumination effect can be reduced by discarding several low-frequency DCT coefficients. Our FSVDR distinguishes from Gabor wavelet and DCT in that (1) FSVDR employs the well-known singular value decomposition (SVD), not Gabor wavelet or DCT and (2) FSVDR needs to tune a parameter \( \alpha \) to yield an IR and the choice of the parameter is both database and DR method dependent, while neither Gabor wavelet nor DCT has such parameter to be tuned. Besides, it is worthwhile to point out that, since FSVDR is still human face like (see Section 2.2), it can be applied prior to Gabor wavelet or DCT for forming an new IR for face recognition.

Furthermore, in face recognition, there have been studies of employing SVD to obtain representation for face images. Hong [36] proposed to apply SVD to each OGV to obtain singular values (SVs) to represent this face image, and then to perform classification based on these SVs. Cheng et al. [37] made use of the SVs as an IR, and then employed an optimal discriminant transformation to transform the SVs into a new space for subsequent classification. Although good classification performance was reported in Ref. [37], Tian et al. [38] pointed out that the SVs contained little useful information for face recognition and attributed the good performance reported in Ref. [37] to the small testing database. Comparing FSVDR with SVs [36,37], it is clear that the two methods are quite different. More specifically, the representation by SVs only employs the SVs, while our FSVDR utilizes not only the SVs, the left and right transformation matrices but also a parameter \( \alpha \) to yield the so-needed IR. Besides SVs, SVD was utilized to generate multiple virtual samples for face recognition with one sample per class in Ref. [39], which is a nonparametric method and is different from FSVDR.

Now, it is worthwhile to summarize our contributions in this paper as follows:

1. We take notice of the fact that every image matrix has the well-known SVD and point out both theoretically and experimentally that, for any given single face matrix \( A \), its leading base images (those corresponding to large SVs) are sensitive to facial variations such as illumination, occlusions, etc. Meanwhile, we show experimentally that these leading base images in fact dominate the composition of \( A \).

2. Based on the observations made in (1), our proposed FSVDR subtly deffates the weights of the facial variation sensitive base images by introducing a parameter \( \alpha \) and effectively alleviates the influence of the facial variations on face recognition. Furthermore, the relationship between FSVDR and OGV is theoretically studied.

3. We employ FSVDR as an IR for many well-known DR methods such as PCA, 2DPCA, NIGLRAM, PCA+LDA, DCV and LDA/QR. And the experimental results show that, by employing FSVDR as an IR, the classification performance of these methods can be significantly improved in case of great facial variations.

4. Our FSVDR offers a both database and DR method dependent IR, namely for different methods and different databases, the optimal value for parameter \( \alpha \) should be different. This is a very important characteristic of FSVDR, and is in accord with the main argument made in Refs. [12,40] that the DR methods can only obtain good performance under certain conditions. Further, based on our previous work in Ref. [28], we offer a heuristic criterion for choosing the parameter \( \alpha \) for the LDA-based methods.

In what follows, we present our proposed FSVDR in detail in Section 2, carry out extensive experiments to verify the effectiveness of the proposed FSVDR in Section 3, and draw an conclusion to this paper in Section 4.

2. Fractional order singular value decomposition representation

We will first analyze the SVD on each face image matrix pattern \( A \) in Section 2.1, and propose FSVDR in Section 2.2.

2.1. Singular value decomposition

Mathematically, every \( r \times c (r \geq c) \) without loss of generality) OGVM \( A \) can always have the SVD [41] as

\[
A = \tilde{U} \tilde{S} \tilde{V}^T,
\]

where \( \tilde{U} = [u_1, u_2, \ldots, u_r] \), \( \tilde{V} = [v_1, v_2, \ldots, v_c] \), \( \tilde{S} = (D \ 0)^T \), \( D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_c) \). 0 is a \( c \times (r-c) \) zero matrix and \( \lambda_i \)'s are the SVs in a nonincreasing order.

Further, assuming the rank of \( A \) to be \( k \), we have

\[
A = USV^T = \sum_{i=1}^{k} \lambda_i u_i v_i^T,
\]

where \( S = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) \), \( U = [u_1, u_2, \ldots, u_k] \) and \( V = [v_1, v_2, \ldots, v_k] \).

We have the following five properties for the SVD applied to \( A \).

**Property 1.** Denote \( A \) as

\[
A = [a_1, a_2, \ldots, a_r]^T,
\]

where \( a_i^T \) is a \( 1 \times c \) row vector that represents the \( i \)-th row of image matrix \( A \), and we have

\[
V^T C_{row} V = S^2,
\]

where

\[
C_{row} = A^T A = \sum_{i=1}^{r} a_i a_i^T.
\]

That is to say, \( v_j \) is the eigenvector of the covariance matrix \( C_{row} \) corresponding to eigenvalue \( \lambda_j^2 \), \( j = 1, 2, \ldots, k \).
Property 2. Denote $A$ as

$$A = [a_1, a_2, \ldots, a_c],$$

where $a_i$ is an $r \times 1$ column vector that represents the $i$th column of image matrix $A$, and we have

$$U^T C_{col} U = S^2,$$

where

$$C_{col} = AA^T = \sum_{i=1}^{c} a_i a_i^T,$$

namely, $u_j$ is the eigenvector of the covariance matrix $C_{col}$ corresponding to eigenvalue $\lambda_j^2$, $j = 1, 2, \ldots, k$.

Property 3. $u_i v_j^T$, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$ form a set of orthonormal base matrices for the space $R^{r \times c}$ and can be termed as base images.

Property 4. When projecting $A$ to the base images $u_i v_j^T$, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$, we have: (1) the coefficients are all nonnegative; (2) the coefficients are $\lambda_i$'s in the base images $u_i v_j^T$, $i = 1, 2, \ldots, k$; and (3) the coefficients are zeros in the base images complement to those in (2).

Property 5. When $A$ is regarded as the composition of $u v^T$ like bases, where $u$ and $v$ are $r \times 1$ and $c \times 1$ column vectors, respectively, then the least number of such bases is $k$ and $u_i v_j^T$, $i = 1, 2, \ldots, k$ provide a solution.

The proofs of these properties are simple and omitted here. Properties 1 and 2 reveal that (1) $u_i$'s corresponding to large $\lambda_i$'s capture the great horizontal variances among the column vectors of $A$, and (2) $v_j$'s corresponding to large $\lambda_i$'s capture the great vertical variances among the row vectors of $A$. Property 3 defines the concept of base images, and it is easy to conclude from properties 1–3 that the leading base images (those corresponding to leading SVs) capture the great variances in the face image $A$ itself. Properties 4 and 5 further offer some characteristics of these base images. Now, we experimentally demonstrate two important properties of the leading base images as follows.

2.1.1. Leading base images being sensitive to facial variations

We choose six images (under distinct facial variations) from the well-known AR face database [42] and show the original face images and the corresponding three leading base images in Fig. 2, from which we can clearly observe that the first base image $u_1 v_1^T$ well captures: (1) the light condition (left light on or right light on) of the face images and (2) the horizontal position of eyebrows, mouth (if not masked) and occlusions such as wearing glasses and wearing scarves. The reason behind these phenomena is as described in Properties 1 and 2, namely $u_1$ and $v_1$, respectively, capture the greatest variance of the column and row vectors of the face images and $u_1 v_1^T$ captures the greatest variance among $A$ itself. Theoretically, $u_2 v_1^T$ and $u_3 v_3^T$ also capture the great variances (although less than those by $u_1 v_1^T$) among the face image itself, but they are less obvious than $u_1 v_1^T$ in vision which may attribute to the fact that our eyes are good at detecting some larger variations (as in $u_1 v_1^T$) while not for some smaller variations (as in $u_2 v_2^T$ and $u_3 v_3^T$). Moreover, a thorough look at the second row of Fig. 2 show that $u_2 v_2^T$ does reveal the two reflected light spots on the black glasses. Finally, due to the fact that the leading base images capture the great variances within the face image itself, they are naturally more sensitive to facial variations such as expression, lighting, occlusion and so on, which can be typically read from $u_1 v_1^T$ in Fig. 2.

2.1.2. Leading base images dominating the composition of face image

The leading base images contribute a great deal to the composition of $A$, for which sake we define the cumulative energy contained in these leading images as

$$e_i = \frac{\sum_{j=1}^{i} ||\lambda_j u_j v_j^T||^2_F}{||A||^2_F},$$

where $||.||^2_F$ is the squared Frobenius norm. We plot the cumulative energy $e_i$ in Fig. 3, where the six face images are just the corresponding six face images shown in Fig. 2.
Fig. 3, one can easily read: (1) the first base images of the six face images here all possess an energy of over 85%; (2) the first few leading base images (e.g., 10) almost occupy all the energy (e.g., the cumulative energy $e_{10} > 99\%$) contained in $A$. As a result, the leading base images dominate the composition of face image.

From the discussion in Sections 2.1.1 and 2.1.2, we know that if we handle the facial variation sensitive leading base images nicely, the great facial variations within each face image can be effectively alleviated, which will be discussed in our proposed FSVDR in the next subsection.

2.2. Fractional order singular value decomposition representation

To alleviate the facial variations on face images, we propose a novel FSVDR, whose underlying ideas are that (1) the weights of the leading base images $u_i v_i^T$ should be deflated, since they are very sensitive to the great facial variations within the image matrix $A$ itself; (2) the weights of base images $u_i v_i^T$ corresponding to relatively small $\lambda_i$’s should be inflated, since they may be less sensitive to the facial variations within $A$, which can be read from Properties 1 and 2; and (3) the order of the weights of the base images $u_i v_i^T$ in formulating the new representation $B$ should be retained. More specifically, for each face image matrix $A$ which has the SVD in Eq. (2), its FSVDR $B$ is defined as

$$B = U S^\alpha V^T,$$

where $U$, $S$ and $V$ are the corresponding matrices in Eq. (2), and in order to achieve the above underlying ideas, $\alpha$ is a fractional parameter that satisfies:

$$0 \leq \alpha \leq 1.$$

It is easy to obtain the following properties with regard to FSVDR:

Property 6. The rank of FSVDR $B$ is $k$, i.e., identical to the rank of $A$.

Property 7. $u_i v_i^T$, $i = 1, 2, \ldots, k$ form a set of $u v^T$ like base images for the FSVDR $B$.

Property 8. When $0 < \alpha \leq 1$, the transformation of $A$ to $B$ is a bijection; when $\alpha = 0$, the transformation of $A$ to $B$ is not a bijection, but a surjection.

Property 9. Let $\beta = (\alpha - 1)/2$ and define $(V^T S V)^{\beta} = V^T S^\beta V$, then we have $B = A (A^T A)^{\beta}$, namely the FSVDR $B$ utilizes some cross product information contained in $A$, which has somewhat flavor of employing high-order information.

Property 6 shows that an intrinsic characteristic of $A$, the rank, is retained in the FSVDR $B$. From Properties 7 and 5, we know that the FSVDR $B$ in fact has the same $u v^T$ like base images as $A$, and considering the fact that these base images are the components to compose $A$ and $B$, we can say that the
information in $A$ is passed to $B$ nicely. Property 8 further shows the close relationship between the FSVDR $B$ and the original representation $A$. Property 9 says that the FSVDR $B$ utilizes some cross product information contained in OGVM $A$.

Now, we illustrate some FSVDR face images under different $z$ in Fig. 4, from which we can observe that:

1. The FSVDR $B$ is still like human face.
2. The FSVDR deflates the lighting condition in vision. Taking the two face images in the first row for example, when $z$ is set to 0.4 and 0.1, from the FSVDR $B$ alone, one can hardly tell whether the original face image matrix $A$ is of left light on or right light on.
3. The FSVDR $B$ reveals some facial details. In the original face images $A$ presented in the first row, neither the right eyeball of the left face image nor the left eyeball of the right face image is visible, however, when setting $z$ to 0.4 and 0.1 in FSVDR, the eyeballs become visible. Moreover, the jowls become more obvious in the FSVDR, as revealed in the face images of the first and the second rows.
4. The FSVDR $B$ cannot remove the occlusions such as glasses and scarves, due to the fact that it shares the same base images as the OGVM $A$. However, we can see from the second and third rows that the contrast between the face images and the occlusions decreases. As a result, in the FSVDR, the occluded images will be nearer to the faces without occlusions compared to those in the OGVM $A$.

The most important characteristic of the FSVDR lies in that it can form an IR for many DR methods such as PCA, LDA and so on (illustrated in route c of Fig. 1). Furthermore, when employing FSVDR as an IR for DR methods, the time complexities in training and testing are almost the same as the original DR methods. To elaborate this, we take DCV as an example in the following analysis. According to Refs. [27,28], the time complexity for training $N$ samples with dimensionality $d = rc$ is $O(N^2d)$, and the time complexity in testing any given unknown sample is $O(dc)$, where $C$ is the number of classes. For DCV based on FSVDR, on one hand, it consumes additional $O(Nd \max(r, c))$ in computing the FSVDR for $N$ samples, where $\max(r, c)$ is usually smaller than $N$, and thus the time complexity in training is still $O(N^2d)$, the same as original DCV; on the other hand, for any unknown sample, it takes additional $O(\max(r, c)d)$ in computing its FSVDR, and thus the time complexity in testing is also $O(dc)$ since $\max(r, c)$ is usually comparable to or less than $C$. And we will verify the effectiveness of FSVDR in the next section.

3. Experiments

In this section, we carry out extensive experiments to show that: (1) when directly applied to face recognition, FSVDR can yield significantly better classification performance than OGVM and the SVs (see Section 3.2); and (2) as an IR, the FSVDR can significantly improve the classification performance of quite a few DR methods such as PCA, PCA+LDA, LDA/QR, DCV, 2DPCA and NIGLRAM (see Section 3.3). Furthermore, we will experimentally visualize the samples by the FSVDR and OGVM to show the benefit brought by FSVDR in Section 3.4 and dwell on the problem of parameter choice in Section 3.5. Before reporting the experimental results, we first describe the database and the experimental setting in Section 3.1.

3.1. Database and experimental setting

3.1.1. Database description

We carry out experiments on three renowned face databases: AR [42], FERET [43] and YALE [25].

The AR database consists of over 4000 color images of 126 person’s faces (70 men and 56 women). Each person has 26 different images which were grabbed in two different sessions separated by two weeks, and 13 images in each session were recorded. The 13 images are, respectively, of neutral expression, smile, anger, scream, left light on, right light on, both light on, occlusion by glasses and left light on, occlusions by glasses and right light on, occlusions by glasses and both light on, occlusion by scarves and left light on, occlusions by scarves and right light on, occlusions by scarves and both light on. Fig. 5 illustrates the 26 image faces under different facial variations.
Fig. 5. An illustration of 26 images of one subject from AR face database.

The AR face database is one of the most well-known face recognition benchmarks. The Color FERET database contains a total of 11,338 facial images corresponding to 994 subjects, and the Gray FERET contains a total of 14,051 grayscale images corresponding to 1209 subjects. Here, we carry out experiments on the hardest subset of FERET Tests September 1996, whose testing samples have great facial variations in illumination. More specifically, we employ the gallery set that contains 1196 face images as training set and the fafc set that has 194 face images as testing set. The face images are preprocessed according to the CSU Face Identification Evaluation System [44] with a resolution of $75 \times 65$. The challenges of this FERET
subset are: (1) a large number of subjects (1196) in the training set, (2) one training sample per class and (3) great illumination variations in the testing set. Due to difficulty of this subset, we follow the CSU Face Identification Evaluation System [44] to report the Rank $k$ classification, where the testing sample is considered to be correctly classified so long as it belongs to the same class as one of its $k$ nearest neighbor (NN) samples in the training set.

The YALE face database contains 165 gray level face images of 15 persons. There are 11 images per subject, and these 11 images are, respectively, under the following different facial expression or configuration: center-light, wearing glasses, happy, left-light, wearing no glasses, normal, right-light, sad, sleepy, surprised, and wink. In our experiment, the images are cropped to a size of $50 \times 50$, and the gray level values of all images are rescaled to $[0, 1]$. Fig. 6 shows the 11 images of one person from this database. On YALE face database, we perform two different experiments, YALE1 and YALE2, where the training and testing samples are given in Table 2.

### Table 2: Data partition on YALE face database

<table>
<thead>
<tr>
<th>Category</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>YALE1</td>
<td>c, f, h, i, j, k</td>
<td>a, b, d, e, g</td>
</tr>
<tr>
<td>YALE2</td>
<td>a, b, e, f, i, k</td>
<td>c, d, g, h, j</td>
</tr>
</tbody>
</table>

Fig. 6. An illustration of 11 images of one subject from YALE face database.

In our proposed FSVDR, one key problem is to choose the adequate value for $\alpha$. Instead of tackling this problem from the very beginning, we just let $\alpha$ change from 0 to 1 incremented by 0.05, and as a result, we will get a total of 21 classification results for a given method and a given experiment.

### 3.2. Comparison with OGVM and SVs for classification

In this subsection, we first compare the classification performance of FSVDR (illustrated in route d of Fig. 1, where the IR is FSVDR) and OGVM (illustrated in route a of Fig. 1) and report the experimental results in Figs. 7 and 8. Moreover, for illustration convenience, we list the best classification accuracies of FSVDR as a column in Table 3 with the name NN.

From Figs. 7 and 8 and Table 3 we can observe that the classification accuracies under FSVDR are significantly higher than those by OGVM. More specifically: (1) on AR1, the classification accuracy based on OGVM is 78.3%, and FSVDR achieves a best classification accuracy of 88.4%, an improvement of 10.1%, which attributes to that FSVDR can alleviate the variations caused by expression, lighting and time duration; (2) on AR2, OGVM can only achieve a classification accuracy of 56.7% which attributes to the occlusions by wearing glasses. However, the best classification accuracy for FSVDR is 89.7%, 33% higher than that of OGVM, which clearly shows that FSVDR can alleviate the occlusions caused by wearing glasses; (3) on AR3, OGVM can only achieve a classification accuracy of 11.0% which attributes to the occlusions by wearing scarves. For FSVDR, it can achieve a best classification accuracy of 82.3%, an improvement of 71.3%. And the reason behind the classification improvements is due to that FSVDR can effectively alleviate the occlusions caused by wearing scarves. Furthermore, comparing the classification performance (of either OGVM or FSVDR) on AR2 and AR3, we can see that in face recognition, the occlusion caused by wearing scarves is much
more harder than that by wearing glasses, which is in accord with the argument made in Refs. [45, 46]; (4) on AR4, FSVDR achieves a best classification accuracy of 81.0%, a significant improvement of 18.8% compared to OGVM’s 62.2%, which attributes to FSVDR’s alleviation of the facial variations in expression, lighting conditions, and occlusions on wearing glasses.
Table 3
The best classification accuracy (%) of the proposed FSVDR on the six independent experiments and by all the methods

<table>
<thead>
<tr>
<th>Data set</th>
<th>NN</th>
<th>DCV</th>
<th>LDA/QR</th>
<th>PCA+LDA</th>
<th>PCA1</th>
<th>PCA2</th>
<th>2DPCA</th>
<th>NIGLRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1 B</td>
<td>88.4</td>
<td>88.9</td>
<td>88.6</td>
<td>91.4</td>
<td>81.6</td>
<td>87.0</td>
<td>84.1</td>
<td>86.9</td>
</tr>
<tr>
<td>AR1 A</td>
<td>78.3</td>
<td>83.6</td>
<td>85.7</td>
<td>84.9</td>
<td>74.1</td>
<td>72.9</td>
<td>74.7</td>
<td>76.6</td>
</tr>
<tr>
<td>AR2 B</td>
<td>89.7</td>
<td>89.3</td>
<td>88.0</td>
<td>87.3</td>
<td>88.0</td>
<td>90.0</td>
<td>91.3</td>
<td>91.0</td>
</tr>
<tr>
<td>AR2 A</td>
<td>56.7</td>
<td>64.3</td>
<td>67.3</td>
<td>56.0</td>
<td>50.7</td>
<td>46.3</td>
<td>49.0</td>
<td>52.7</td>
</tr>
<tr>
<td>AR3 B</td>
<td>82.3</td>
<td>78.3</td>
<td>73.3</td>
<td>83.0</td>
<td>54.7</td>
<td>78.3</td>
<td>68.3</td>
<td>69.3</td>
</tr>
<tr>
<td>AR3 A</td>
<td>11.0</td>
<td>46.7</td>
<td>38.7</td>
<td>52.0</td>
<td>10.0</td>
<td>8.7</td>
<td>10.0</td>
<td>9.7</td>
</tr>
<tr>
<td>AR4 B</td>
<td>81.0</td>
<td>81.2</td>
<td>79.9</td>
<td>82.7</td>
<td>68.1</td>
<td>79.5</td>
<td>72.2</td>
<td>76.5</td>
</tr>
<tr>
<td>AR4 A</td>
<td>62.2</td>
<td>71.4</td>
<td>70.8</td>
<td>74.2</td>
<td>59.4</td>
<td>56.5</td>
<td>60.0</td>
<td>61.2</td>
</tr>
<tr>
<td>YALE1 B</td>
<td>80.0</td>
<td>82.7</td>
<td>84.0</td>
<td>85.3</td>
<td>61.3</td>
<td>69.3</td>
<td>76.0</td>
<td>81.3</td>
</tr>
<tr>
<td>YALE1 A</td>
<td>52.0</td>
<td>52.0</td>
<td>49.3</td>
<td>58.7</td>
<td>50.7</td>
<td>50.7</td>
<td>49.3</td>
<td>52.0</td>
</tr>
<tr>
<td>YALE2 B</td>
<td>85.3</td>
<td>94.7</td>
<td>86.7</td>
<td>92.0</td>
<td>73.3</td>
<td>77.3</td>
<td>85.3</td>
<td>85.3</td>
</tr>
<tr>
<td>YALE2 A</td>
<td>73.3</td>
<td>81.3</td>
<td>78.7</td>
<td>81.3</td>
<td>68.0</td>
<td>68.0</td>
<td>73.3</td>
<td>73.3</td>
</tr>
</tbody>
</table>

B, FSVDR; A, OGVM.

and scarves; (5) on FERET, OGVM operates very poorly, with Rank 1, 11 and 21 classification accuracies only being 8.2%, 28.9% and 38.1%. This is due to a large number (1196) of distinct subjects in the training set and the great illumination variation in the testing samples. Favored by the proposed FSVDR, the best Rank 1 classification is 42.3%, which is higher than the Rank 21 classification of OGVM. Meanwhile, the best Rank 11 and 21 classification accuracies of FSVDR are significantly higher than those of OGVM; (6) on YALE1, OGVM achieves a classification accuracy of 52.0%, 28% lower than FSVDR’s best classification accuracy of 80.0%, which again verifies that FSVDR can alleviate such facial variations caused by lighting and wearing glasses; (7) on YALE2, FSVDR achieves a best accuracy of 85.3%, 12% higher than OGVM’s 73.3%, which attributes to FSVDR’s ability to alleviate the distinct expression variations and lighting changes.

Finally, the classification accuracies by directly applying SVs for classification on AR1, AR2, AR3, AR4, FERET, YALE1 and YALE2 are, respectively, 11.6%, 1.7%, 1.7%, 9.6%, 0.5%, 27.3% and 44.0%, which is in accord with the argument in Ref. [38] that the SVs contain little information for classification. Obviously, our FSVDR can achieve significantly better classification performance than FR by SVs.

3.3. An intermediate stage for several renowned DR methods

After verifying that FSVDR can yield significantly better classification performance compared to OGVM and SVs, we now move on to show that the FSVDR can improve the classification performance of two categories of renowned DR methods: LDA-based methods and PCA-based methods. More specifically, we will carry out experiments on the LDA-based methods such as PCA+LDA, DCV and LDA/QR in Section 3.3.1 and the PCA-based methods such as PCA, 2DPCA and NIGLRAM in Section 3.3.2.

3.3.1. LDA-based methods

We report the experimental results of employing FSVDR as an IR for LDA-based methods in Figs. 9 and 10. For illustration convenience, we list the best classification accuracies of FSVDR under optimal parameter z in Table 3, Figs. 13 and 14. Firstly, from Table 3, Figs. 13 and 14, we can observe that, when performing the LDA-based methods based on OGVM for face recognition, they can generally yield significantly better classification accuracies compared to those obtained on OGVM. For example, on AR3, the classification accuracy based on OGVM is 11.0%, but DCV, LDA/QR and PCA+LDA-based on OGVM, respectively, yield classification accuracies of 46.7%, 38.7% and 52.0%, or improvements of 35.7%, 27.7% and 41.0%. However, due to the great facial variations, the LDA-based methods based on OGVM cannot achieve a satisfactory classification performance on AR3. Secondly, from Figs. 9, 10, Table 3, Figs. 13 and 14, we can observe that when employing FSVDR as an IR, DCV, LDA/QR and PCA+LDA can achieve significantly higher classification accuracies than those based on OGVM. More specifically, (1) on AR1, AR4, and YALE2, the LDA-based methods utilizing FSVDR as an IR almost achieve 5.0% improvement in classification accuracies than the LDA-based methods based on OGVM; (2) on AR2 and YALE1, the LDA-based methods employing FSVDR as an IR witness over 20.0% improvement in classification accuracies compared to those based on OGVM; (3) on AR3, the classification accuracies of the LDA-based methods using FSVDR as IR are more than 30.0% higher than those based on OGVM.

As a result, the LDA-based methods can benefit from FSVDR, which acts as an IR to effectively alleviate facial variations such as expression, lighting, occlusions on wearing glasses and scarves and so on.

3.3.2. PCA-based methods

After verifying that the FSVDR can improve the classification performance of the LDA-based methods, we move on to show that the PCA-based methods can benefit from FSVDR, which acts as an IR. The experimental results under different z’s are reported in Figs. 11 and 12, and the best classification accuracies based on the FSVDR are reported in Table 3, Figs. 13 and 14. From these results, we can clearly see that the classification performance of the PCA-based methods employing FSVDR as an IR is significantly better than that of the PCA-based methods based on OGVM. More specif-
ically, on one hand, from the results reported on AR2, we can get that, the best classification accuracies based on FSVDR for PCA1, PCA2, 2DPCA and NIGLRAM are, respectively, 88.0%, 90.0%, 91.3% and 91.0%, which are 37.3%, 43.7%, 42.3% and 38.3% higher than those based on OGVM. On the other hand, the most significant improvement of classification performance can be typically drawn from the experiments on AR3, where PCA1, PCA2, 2DPCA and NIGLRAM based on FSVDR are, respectively, 44.7%, 60.6%, 58.3% and 59.6% higher than those based on OGVM.

As a result, from the results presented above, it is easy to conclude that when employing FSVDR as IR, the classification performance of the DR methods can be significantly improved.

### 3.4. Visualization of samples under the FSVDR

In this subsection, we will experimentally visualize the distribution of the training and testing samples in the reduced space to show the benefits brought by FSVDR. For this sake, we choose PCA+LDA as feature extraction method which can project the data from $C$ classes to a reduced $C-1$ dimensional space.

We carry out two independent experiments: (1) in Fig. 15(a) and (b), the 30 samples corresponding to three randomly chosen classes are from AR2, namely, samples (a–g) (see Fig. 5) are used for training and samples (h–j) are utilized for testing; and (2) in Fig. 15(c) and (d), the 30 samples corresponding to three randomly chosen classes are from AR3, namely, samples
(a–g) are used for training and samples (k–m) are utilized for testing.

Firstly, we look at the training samples presented in Fig. 15(a), from which we can see that, when employing OGVM (or setting \( z \) to 1 in the FSVDR), the samples from the same class are not compact, and the samples from different class are not far away. More specifically, on one hand, the four training samples from # 1 (denoted by square) are around \((-4.8, 4.4)\), but the other three training samples from # 1 are far away from \((-4.8, 4.4)\) and locate at \((-4.7, 10.4)\), \((-4.7, -10.4)\) and \((-4.8, -3.5)\), respectively. Tracing the experimental results, we know that the four training samples from # 1 that locate compactly around \((-4.8, 4.4)\) are samples (a–d), while the other three samples are samples (e–g), respectively. Similar observations can also be obtained from the training samples from # 2 and # 3. On the other hand, the Euclidean distance between \((-4.7,10.4)\) from # 1 to \((0.8, 10.5)\) from # 2 is about 5.5, quite less than 20.8, the Euclidean distance between \((-4.7,10.4)\) from # 1 and \((-4.7,-10.4)\) from the same class # 1. Obviously, the above given experimental results show that, when performing PCA+LDA on OGVM, we cannot come to the objective that the samples from the same class are compact and meanwhile the samples from different classes are far away in case of great facial variations.

Secondly, we look at the testing samples presented in Fig. 15(a), from which we can clearly see that the testing samples from the same class are not compact too and meanwhile the samples from different classes are not far away. Moreover,
the testing samples (of the three classes) corresponding to (h) are around a horizontal line $y = 2.7$, those corresponding to (i) around a horizontal line $y = 6.0$ and those corresponding to (j) around a horizontal line $y = -7.0$. The *Euclidean* distance between $(-2.5, 6.8)$ from #1 and $(-1.1, 6.8)$ from #2 is 1.4, quite less than 14, the *Euclidean* distance between $(-2.5, 6.8)$ from #1 and $(-2.9, -7.2)$ from the same class #1.

The results presented in Fig. 15(c) witness similar phenomenon as Fig. 15(a). The reason behind the phenomena in Fig. 15(a) and (c) is: when based on OGVM, the PCA+LDA is unable to compactly cluster the same class samples, which are under severe facial variations such as lighting, expression and occlusions.

Finally, we turn to the experimental results with FSVDR as an IR ($\alpha$ is set to 0.1) in Fig. 15(b) and (d), from which we can clearly see that: (1) the training samples belonging to the same class become very compact; (2) the training samples belonging to different classes are well separated from each other; (3) the testing samples from the same class are compact with each other; (4) generally speaking, these testing samples do not locate very near to the training with the same classes, which attributes to the fact that our proposed FSVDR does not remove the occlusions such as glasses and scarves, and on the contrary it just alleviates the influence of such occlusions; (5) despite that the testing samples do not locate very near to the training samples in the same classes, the testing samples can all be correctly classified, which attributes to the fact that the testing samples are farther away from the training samples of different classes compared to those of the same class.

Comparing the results presented in Fig. 15(a) and (c) with those in (b) and (d), we can clearly see that, FSVDR’s alleviation of the facial variations can help PCA+LDA achieve its objective (namely, samples from the same class are compact and samples from different classes are far away) and as a result can help improve the classification performance.
Fig. 14. A comparison of the best classification performance of the proposed FSVDR and OGVM utilizing different methods on YALE face database. Left bar: the best classification accuracy by FSVDR, right bar: the classification accuracy by OGVM.

3.5. The parameter $\alpha$

In FSVDR, $\alpha$ is a key parameter that should be tuned. From Figs. 7–12, we can observe that the classification accuracy curves are approximately unimodal and that there are many $\alpha$’s that can achieve superior performance to OGVM. Further, in our experiments, $\alpha = 0.1$ seems to be a good choice for NN and the PCA-based methods, and $\alpha = 0.4$ seems to a nice choice for the LDA-based methods.

Generally speaking, in designing automatic criterion for choosing adequate parameter $\alpha$, one should consider the following factors: (1) the smaller $\alpha$ is, the more the leading base
images (which are sensitive to facial variations) are deflated but meanwhile the discriminant information contained in the leading base images may be deflated too; (2) some face images have great facial variations (e.g., Fig. 5(h–m, u–z)) and are perhaps in favor of smaller $x$’s, while some face images have slight facial variations (e.g., Fig. 5(a, n)) and might be in favor of larger $x$’s; (3) the $x$ learned from the training set is a trade-off among all the training samples and thus is only applicable to the unknown sample from the similar distribution; and (4) each DR method has its specific application scope, which leads to the difficulty in designing a unique $x$ selection criterion for all the DR methods. As a result, the criterion for automatic choosing $x$ should be dependent on the training samples, the given testing sample and the specific DR method.

In the following, instead of aiming at finding an automatic criterion that is applicable to all methods and databases, we try to look for an automatic criterion for the LDA-based methods. Furthermore, in order to ensure that the learned $x$ on the training set is applicable to the testing set, we only consider the case that the testing set has similar distribution as the training set (in fact, this is also almost a basic and common assumption of statistical learning theory, and one can easily get that AR1 and AR4 satisfy this condition, while others not).

Firstly, we introduce a mean square variance (MSV) criterion defined in Ref. [28]:

$$\text{MSV} = \frac{1}{C} \sum_{i=1}^{C} SV_i,$$  \hspace{1cm} (12)

where $SV_i$ is the standard variance of the $i$th class defined as

$$SV_i = \frac{1}{d} \sum_{k=1}^{d} \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (x_{ijk} - m_{ik})^2,$$ \hspace{1cm} (13)

where $x_{ijk}$ and $m_{ik}$, respectively, denote the $k$th element of the $d$-dimensional samples $x_{ij}$ and class mean $m_{ij}$, $C$ is the number of classes, and $N_i$ is the number of training samples contained in the $i$th class. In Ref. [28], we argued that when MSV is relatively small, LDA-based methods such as DCV will operate well, and on the contrary, when the MSV value is relatively high, LDA-based methods such as DCV will operate poorly. A justification for this criterion is given as follows: (1) the smaller $\text{MSV}$ is, the compact the same class samples are, and on the contrary, the bigger $\text{MSV}$ is, the looser the same class samples are; (2) when the same class samples are very loose, these samples will lead to biased estimation of the class mean, within-class and between-class scatter matrices, while on the contrary, when the same class samples are compact, the estimation of the class mean, within-class and between-class variance matrices may be much more reliable; (3) when the same class samples are compact, it is more likely that these samples can nicely depict the Gaussian distribution from which they are generated; and (4) considering the fact that LDA is a special case of the Bayesian decision theory [20] under the assumption that the $C$ classes samples are, respectively, from $C$ Gaussian distributions with equal covariance, then it is essential for the same class samples to be compact, namely $\text{MSV}$ to be small in the LDA-based methods.

Based on the above argument, we give a heuristic criterion to automatically choose an adequate $x$ for the LDA-based methods

$$x_{opt} = \arg \min_x \text{MSV}(x),$$ \hspace{1cm} (14)

where $\text{MSV}(x)$ is the MSV of the training samples $x_{ij} = \text{vec}(B_{ij}^j)$, $j = 1, 2, \ldots, N_i$, $i = 1, 2, \ldots, C$, $B_{ij}^j$ is the FSVDR of the face image samples $A_{ij}$ under parameter $x$ and $A_{ij}$ denotes the $j$th face image from the $i$th class.

We report the experimental results on AR1 and AR4 in Fig. 16, from which we can see that, on AR1 and AR4, when $x$ is set to 0.4, $\text{MSV}(0.4)$ achieves the optimal value. Turning to the results presented in Fig. 9, we can clearly see that, when $x$ is set to 0.4, the LDA methods (such as DCV, LDA/QR and PCA+LDA) based on FSVDR can yield near optimal classification performance that is significantly higher than those based on the OGVM. To further verify that the learned $x$ from a distribution is applicable to the testing samples from the similar distribution, we carry out the following five experiments by utilizing the same testing samples as AR1 and the first $p$ face images of each person for training, where $p$ changes from $3$ to $7$ incremented by $1$. We employ the learned $x = 0.4$ on AR1 for FSVDR in the five experiments and present the results in Fig. 17, from which we can see that: (1) the learned $x = 0.4$ on AR1 is effective on the five experiments and the LDA methods based on FSVDR can obtain significantly superior perfor-
Fig. 17. Performance under different number of training samples per class with $\alpha = 0.4$ on AR1.
tomatically set $\alpha$ adequate values for PCA-based methods such as PCA, 2DPCA and NIGL-RAM; (2) modify the MSV criterion to take the testing sample into consideration so that it can be applicable to the case that the testing sample differs from the training samples to a large extent (e.g., AR2, AR3, FERET, YALE1 and YALE2); (3) carry out research to set sample dependent parameter $\alpha$, since different face samples are affected by facial variations differently (note that, in our FSVDR, we set a universal value for all the face samples); (4) set different $\alpha$ values for different singular values in order to better suppress noise and meanwhile retain discriminant information; (5) compare FSVDR with intermediate representations such as Gabor wavelet and DCT for face recognition; and (6) combine FSVDR and Gabor wavelet or DCT to form a new intermediate representation for face recognition.

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments that enhance this paper, Martinez for providing the AR database, the FERET program for the FERET database, and Belhumeur et al. for the YALE database. This work is supported by the National Natural Science Foundation of China under Grant No. 60473035, the Natural Science Foundation of Jiangsu Province under Grant No. BK2006187, and the Innovation Foundation of NUAA under Grand No. Y0603-042.

References

[38] Y. Tian, T. Tan, Y. Wang, Y. Fang, Do singular values contain adequate information for face recognition, Pattern Recognition 36 (3) (2003) 649–655.


About the Author—SONGCAN CHEN received his B.S. degree in mathematics from Hangzhou University (now merged into Zhejiang University) in 1983. In December 1985, he completed his M.S. degree in computer applications at Shanghai Jiaotong University and then worked at the Nanjing University of Aeronautics and Astronautics (NUAA) in January 1986 as an assistant lecturer. There he received a Ph.D. degree, in 1997, in communication and information systems. Since 1998, as a full-time professor, he has been with the computer science and engineering department at NUAA. His research interests include pattern recognition, machine learning and neural computing. In these fields, he has authored or coauthored over 70 scientific journal papers.

About the Author—JUN LIU received his B.S. degree from Nantong Institute of Technology in 2002. From 2002, he is a Ph.D student in the Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include pattern recognition and computer vision, and he has authored or coauthored over 10 papers.