Neurocomputing 151 (2015) 471–480

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Nearest orthogonal matrix representation for face recognition

Jian Zhang [a,b], Jian Yang [a,*], Jianjun Qian [a], Jiawei Xu [c]

[a] Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China
[b] Department of Computer Engineering, Huaihai Institute of Technology, Lian yungang 222000, PR China
[c] School of Computer Science, University of Lincoln, UK

ARTICLE INFO

Article history:
Received 12 April 2014
Received in revised form
14 July 2014
Accepted 10 September 2014
Communicated by Ran He
Available online 28 September 2014

Keywords:
Face recognition
Image representation
Pattern classification
Singular vector decomposition (SVD)
Sketch image recognition

ABSTRACT

This paper presents a simple but effective method for face recognition, named nearest orthogonal matrix representation (NOMR). Specifically, the specific individual subspace of each image is estimated and represented uniquely by the sum of a set of basis matrices generated via singular value decomposition (SVD), i.e. the nearest orthogonal matrix (NOM) of original image. Then, the nearest neighbor criterion is introduced for recognition. Compared with the current specific individual subspace based methods (e.g. the sparse representation based classifier, the linear regression based classifier and so on), the proposed NOMR is more robust for alleviating the effect of illumination and heterogeneous (e.g. sketch face recognition), and more intuitive and powerful for handling the small sample size problem. To evaluate the performance of the proposed method, a series of experiments were performed on several face databases: Extended Yale B, CMU-PIE, FRGCv2, AR and CUHK Face Sketch database (CUFS). Experimental results demonstrate that the proposed method achieves encouraging performance compared with the state-of-the-art methods.

1. Introduction

In past few decades, face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. So far, the state-of-the-art recognition technologies can strike high accuracy under controlled environment, such as frontal faces with comfortable lighting conditions [1]. However, most existing face recognition technologies are still far from the perfection in uncontrolled cases of larger illumination, occlusion, disguise etc. [2]

Recently, reconstruction-representation-based methods, utilizing known samples as the basis to linearly rebuild the non-ideal query sample to eliminate the obstruction, has aroused widespread concerns in the field of robust face recognition. Derived from sparse representation and robust principle component analysis (RPCA) [3], an important relevant work named sparse representation based classifier (SRC) was firstly proposed in [4], where the rarely known samples are selected by L1-optimizer to reconstruct the query sample. With huge computational cost, SRC shows strong ability in dealing with sparse random pixel corruption and block occlusion. Relatively low-cost, Shi et al. advocated the L2-optimizer based regression method selecting all known samples to reconstruct the query sample in [5]. Subsequently, Zhang et al. analyzed the working principle of SRC and asserted the importance of collaborative representation strategy than L1-norm based sparsity constraint [6]. Thus, a collaborative representation based classifier (CRC) was proposed with ridge regression (L2-norm).

On the other hand, above representation residuals are usually measured by L1-norm or L2-norm corresponding to Gaussian or Laplacian distribution respectively. However, the distribution of representation residuals is really complicated to suppress the performance of above mentioned methods [7,8]. To this end, Yang et al. borrowed the idea of robust regression and proposed a regularized robust coding method [7,8]. He et al. further presented a correntropy based sparse representation (CESR) algorithm using the correntropy induced robust error metric [9,10].

Actually, the strategy of seeking the best linear representation in known samples, i.e. minimum linear reconstruction representation residuals, can be traced back to the nearest neighbor line (NFL) [11], which aims to extend the capacity of prototype features by computing a linear function to interpolate and extrapolate each sample pair in the same class. Chien and Wu further extended NFL and proposed the nearest feature plane (NFP) and the nearest feature space (NFS) methods for pattern classification [12]. As a special case of NFS, Naseem proposed linear regression based classifier (LRC) using the whole class samples to construct the linear prototype [13].

* Corresponding author.
E-mail addresses: zhangjianhhit@gmail.com (J. Zhang), csjyang@mail.njust.edu.cn (J. Yang), qjjtx@126.com (J. Qian), jxu@lincoln.ac.uk (J. Xu).

http://dx.doi.org/10.1016/j.neucom.2014.09.019
0925-2312/

Essentially, all successful application methods mentioned above are underpinned by the following assumption [4–16]:

Assumption 1. Specific individual subspace assumption: *samples from a specific object class are known to lie in a specific linear subspace, i.e. any test imagey from a specific class i has a specific individual subspace spanned by* $A_i = [a_{i1}, a_{i2}, ..., a_{iN}]$ *such that*

$$y = A_i w. \tag{1}$$

In SRC and CRC, the specific individual subspace of query sample is achieved by the linear regression (L1-optimizer or L2-optimizer). While the specific individual subspace of NFL, NFP, NFS and LRC is directly spanned by the known samples in one class. Despite different strategies, the specific individual subspace of all above methods is composed of the known samples. Thus, to guarantee sufficient representation of a query sample, relevant methods depend on the holding of following assumption heavily:

Assumption 2. Large sample size assumption: *there are sufficient known samples for each class, such that any query sample can be sufficiently represented using only the known samples in the same class.*

However, Assumption 2 does not really hold well. This leads to larger representation residuals and lower recognition rates of above mentioned methods. For example, none of the above mentioned methods can handle the case of the single image per person problem well. Towards this end, a simple but effective basis-acquisition technique was proposed, coined nearest orthogonal matrix representation (NOMR) explicitly utilizing Assumption 1 rather than implicitly using known samples. Specifically, the specific individual subspace of each image is firstly estimated and represented uniquely by the sum of a set of basis matrices generated via singular value decomposition (SVD), the nearest orthogonal matrix (NOM) of original image. Then, a simple nearest neighbor based criterion is introduced for recognition. This idea behind NOMR is that the visual face images can be divided into two parts by SVD: the special intrapersonal subspace basis corresponding to essence identity and the disguising space associated with various appearance changes such as illumination etc. Consistent with the Assumption 1, the same face images should have similar NOMs.

Compared with traditional specific individual subspace based methods, the proposed NOMR with precise, complete and expressive basis is more robust for alleviating the effect of illumination and heterogeneous (e.g. sketch face recognition), and more intuitive and powerful for handling the small sample size problem in face recognition.

In literatures, some methods with SVD for face recognition have been proposed but are quite different from ours. In [17], Hong et al. firstly proposed a singular value decomposition based face recognition method which uses the singular values of the original face image as the feature. In refs. [18–20], SVD based methods were further proposed for face recognition, which still used the singular values as the image representation features. In spite of perfect mathematical theory and good performance in small sample size databases, many experiments show that the above mentioned methods cannot get a satisfied recognition results on large databases with variations occlusion. In contrast to the methods with singular values, Tian et al. advocated that the left and right orthogonal matrix of SVD have more information for recognition [21]. They proposed a new feature extraction method which takes the projection coefficients with a proper selected orthogonal base of SVD (e.g. the average of one class images) as the feature for face recognition. This method probably leads to a better recognition performance with Bayesian classifier, but still falls into the category of using the singular value feature (i.e. the projection coefficients). In addition, the Non-negative Matrix factorization (NMF) is also established on the SVD. Lee et al. firstly use NMF to yield sparse representation of localized features to represent distributed parts over a face image in [22]. However, the NMF like Eigenface is established on the common subspace assumption rather than the specific individual subspace assumption.

The remainder of this paper is organized as follows. Section 2 presents the preliminary topics about the SVD. Section 3 develops the idea of the nearest orthogonal matrix representation for face recognition. Section 4 conducts extensive experiments to verify the validity of our approach. Section 5 offers our conclusions and future work.

2. Preliminaries of SVD

The singular value decomposition (SVD) is one of the matrix factorization methods, which can be described with the following Theorem in real field.

Theorem 1. ***SVD*** *[23]. If* $A \in R^{m \times n}$*, then there exist orthogonal matrix* $U = [u_1, \dots, u_m] \in R^{m \times m}$ *and* $V = [v_1, \dots, v_n]^T \in R^{n \times n}$ *such that*

$$A = U \Sigma V^T, \tag{2}$$

where $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_p, 0, \dots, 0)$ is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal and $p = \text{rank}(A)$.

The diagonal entries σ_i are denoted as the singular values of A, by convention arranged in non-increasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. The columns of U are termed left-singular vectors of A and the columns of V are called right-singular vectors of A.

In theory, the singular vector decomposition can be achieved by the eigen-decomposition. However, the SVD is really computed more efficiently. The specific algorithms can refer to the literature [24].

Naturally derived, the matrix $A \in R^{m \times n}$ can also be expressed as

$$A = \sum_{i=1}^{p} \sigma_i A_i = \sum_{i=1}^{p} \sigma_i u_i v_i^T \tag{3}$$

with the singular values and vectors. Here u_i and v_i are the ith columns of the corresponding singular vectors, σ_i are the ordered singular values, $p = \text{rank}(A)$. Generally, the Eq. (3) is coined as the separate model of SVD.

Removing the singular values from the Eq. (2), we will acquire

$$L := U \otimes V^T = \sum_{i=1}^{p} u_i v_i^T. \tag{4}$$

Eq. (4) is usually called the *symmetric orthogonalization* of the matric$A$ and the symbol $\otimes$ denotes the outer product of two matrices. It should be noted that L is unique since any sequence of sign choices for the columns of V determines a sequence of signs for the columns of U. In addition, L is also the *nearest orthogonal matrix* (defined as an orthogonal matrix Q which has the minimum$\|Q - A\|_F$) of A, which is underpinned by the following theorem:

Theorem 2. ***NOM****. Over all orthogonal matrices* Q, $||Q - A||_F$ *is minimized if and only if* $Q = L$*.*

Theorem 2 is the key of the orthogonal procrustes problem which was originally solved by Peter Schonemann in [25,26]. A detailed proof is given in [27]. The classical OPP asks how closely a matrix $A \in R^{m \times n}$ can be approximated by a second given matrix $B \in R^{p \times n}$ multiplied by a matrix $\Omega \in R^{m \times p}$ with orthogonal columns in the sense of Frobenius norm. This problem is equivalent to finding the nearest orthogonal matrix (NOM) to a given matrix $M = BA^T$. In addition, this also intuitively makes sense because an orthogonal matrix would have the decomposition UIV^T where $I = diag(\underbrace{1, \ldots, 1}_{p}, 0, \ldots 0) \in R^{m \times n}$.

3. Nearest orthogonal matrix representation for face recognition

3.1. SVD revisited for face representation

Algebra feature (e.g. the singular value feature) is one of the earliest image representations in face recognition society. In ref. [17], Hong et al. firstly proposed the concept of the algebra feature and provided the credible and solid theoretical proof of its advantages. However, the poor practical performance makes it widely questioned to gradually fade from the face recognition [21]. In this subsection, we will revisit SVD from the perspective of individual face subspace and reveal the role of SVD for face representation.

Suppose there is a grayscale face image $A \in R^{\boldsymbol{m} \times \boldsymbol{n}}$, we can easily get a separable form of this image by SVD with the formula (2). Fig. 1 shows an intuitive illustration of the separable form of a face image by SVD. Here,$A_i = u_i v_i^T$, $rank(A_i) = 1, i = 1, 2, \cdots, p$ and $p = rank(A)$.

As shown in Fig. 1, the face image A can be treated as a point in the subspace $S_A = span(A_1, A_2, \cdots, A_p)$. The σ_i is the coordinate on basis A_i. Now, our question turns to revealing the role of the basis and the coordinates (singular values) for face representation.

Here, we just use an example, which satisfies the Specific Individual Subspace Assumption, to reveal the role of SVD. In this example, the face image is selected from Extended Yale B database and divided into separable form by SVD (see Fig. 1).We treated the S_A as the identity individual face subspace of A. Then, we randomly change the singular values to generate some specific samples (points) in subspace S_A. The generating images are shown in Fig. 2. It can be seen that these images still preserve the identity information and the essential appearance details of the original image although the visual appearances has changed.

This observation not only supports the Assumption 1 but also reveals that the face images can be divided into two parts via SVD: the individual basis space corresponding to the essence identity information and the disguising space (with various appearance changes such as illumination etc.) associated with the singular values. In addition, this observation further proves that the singular values, i.e. $(\sigma_1, \ldots, \sigma_p)$, are not suitable for face recognition.

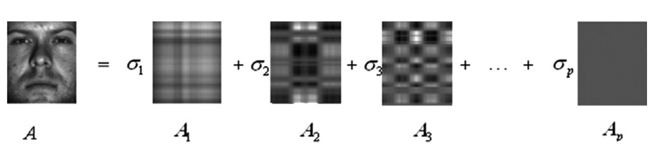

Fig. 1. An illustration of the separable form of one face image from Extended Yale B database by SVD. For the sake of clearly display, each A_i is normalized to [0,255].

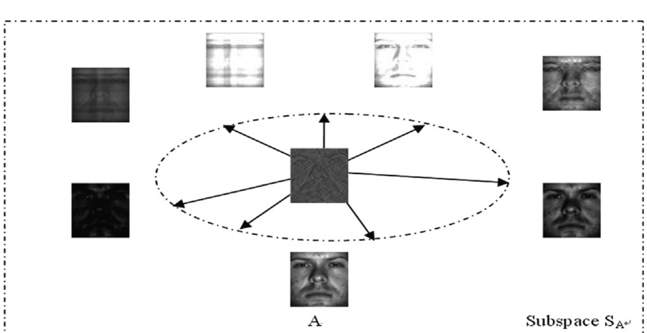

Fig. 2. Different visual appearances generated with different singular value of the same person in subspace S_A. A is the original image.

In fact, the specific individual subspace property of SVD from above example has already found its way into the face recognition and image processing [21,28,29]. In [28], Zhang et.al presented a method based on SVD perturbation to deal with the single image face recognition problem. This method generates several images of one person by changing its singular values and then employs the eigenface for face recognition. However, the high linear correlation of generated images failed to bring significant performance improvements. In [29], the specific individual subspace was successfully used to normalize illumination of images, which also provided a powerful proof to support our views.

3.2. Face representation and recognition with nearest orthogonal matrix

Inspired by the analysis in the previous section and Assumption 1, the natural idea is to directly advocate the basis generated via SVD to identity the original face image and then perform the recognition task. Obviously, the same face images should have the same or similar basis space. Specifically, our method mainly involves three steps. In the first step, the singular value decomposition is applied to a given face image $A \in R^{\boldsymbol{m} \times \boldsymbol{n}}$, so that the left singular vectors U and the right singular vectors Vcan be achieved as follows:

$$[U, \Sigma, V] = svd(A) \tag{5}$$

here, $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_p, 0, \ldots, 0)$ is a $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal and $p = \text{rank}(A)$.

In the second step, we first replace all non-zero singular value σ_i with 1 to get a new diagonal matrix $I = diag(\underbrace{1, \ldots, 1}_{p}, 0, \ldots 0) \in R^{m \times n}$. Then, a new representation of the original image A can be achieved with the following formula

$$A_{new} = UIV^T. \tag{6}$$

It is easy to verify that Eqs. (4) and (6) are equivalent, i.e. A_{new} is the nearest orthogonal matrix of A. Thus, we named the proposed face recognition method as nearest orthogonal matrix representation (NOMR). Fig. 3 shows the representation results of our method.

Finally, the nearest neighbor based criterion is introduced for recognition. Here, the distance between two arbitrary NOMs, B_{new} and C_{new}, is defined as

$$d(B_{new}, C_{new}) = ||B_{new} - C_{new}||_F. \tag{7}$$

where $||B_{new} - C_{new}||_F$ denotes the Frobenius norm (F-Norm) distance between two matrices.

Suppose that the known sample NOMs are $A_{\text{new}1}, A_{\text{new}2}, \ldots, A_{newN}$, where N is the total number of known samples, and that each of these samples is assigned to a given identity c_k. Given a query sample NOM B_{new}, if $d(B_{new}, A_{newl}) = \min_j ||B_{new} - A_{newj}||_F$ and $A_{newl} \in c_k$, then the decision is $B \in c_k$.

Remark 1. For any face image$A$, the reason for NOM application lies in its simplification of basis representation from $[A_1, A_2, \cdots, A_p]$ to $A_{new} = UIV^T$ as well as the equivalence in measuring two arbitrary samples, B and C with the simple and efficient F-Norm distance, i.e. $||B_{new} - C_{new}||_F = ||[B_1, B_2, \cdots, B_p] - [C_1, C_2, \cdots, C_p]||_F$. In addition, a specific individual subspace based representation method is highlighted in this paper rather than a more efficient subspace measurement such as L2-Hausdorff distance [30,31]. Although the NOM looks like a structural face image, it is a representation of the specific individual subspace of original image rather than the simple high-pass filtering.

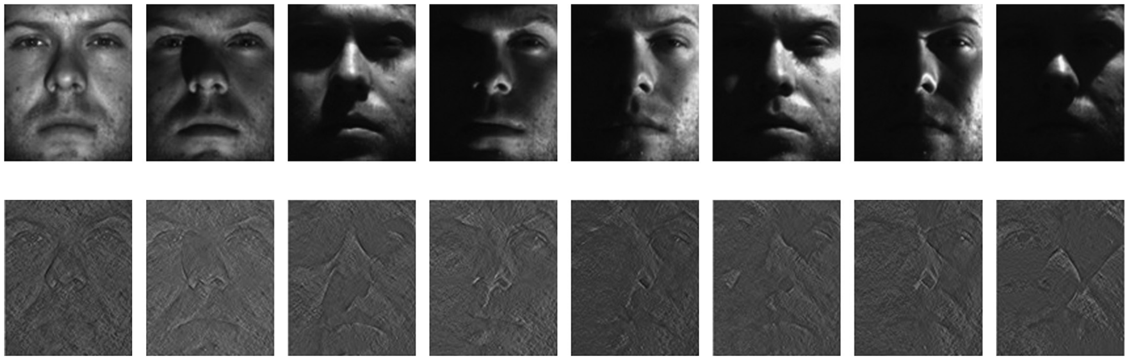

Fig. 3. Some face images of one person form Extended Yale B database and their nearest orthogonal matrix representations. Top row shows the original image and Bottom row shows the corresponding NOMR which is normalized to [0,255].

3.3. Further discussion on NOMR

As an image description or preprocessing method, the proposed NOMR really provides good features and innovative viewpoint for face image representation and recognition in theory.

Firstly, two reasons make the proposed method be more suitable for face recognition. One reason is that this method can express the intrapersonal variations efficiently (see Fig. 2). As we all know, the decision of some well-known methods such as SRC, LRC depends heavily on the reconstruction results. Thus, these methods tend to assume the over-complete requirements (i.e. with sufficient known samples) without exception. In practices, especially the small sample size problem, the over-complete assumption is too difficult to meet to span the real specific individual subspace. However, the linear algebra shows that the basis derived from the face image via SVD is complete for all images with the same size, which makes the proposed method express more variations of one person and achieve better performance than SRC, LRC etc. in the small sample size problems.

Another is its innovative occlusion elimination strategy. Different from traditional processing methods with every effort to restore the desirable images, the efficiency of NOMR is built on another processing strategy which only keeps the essential identify information underlying explanatory factors hidden in the observed milieu of low-level sensory data and removes all the appearance details. This processing strategy is based on a common sense that one observation combines useful information with interference information. Specifically, following this strategy in the face recognition problem, all observed images should be separated into two parts: one is the useful information sufficiently indicating the identity of the object; another is the interference information reflecting the variety of visual changes. However, how do we achieve this goal of stripping out the useful information from the observation? Fortunately, SVD could give a powerful support for achieving this strategy. The NOM denotes useful information indicating the identity and the residual image can be seen as the disguising information which confuses the recognition task.

We further show an example that the NOMR is valid for face recognition with illumination changes. As shown in Fig. 4, images A and B are of the same identity with different illumination conditions, while C and A are of the same illumination condition with different identities. $\overline{A}$, $\overline{B}$ and $\overline{C}$ are the NOMR of A, B and C respectively. If we use the nearest neighbor rule with F-Norm distance to seek for the correct class of B, the distance are $||A-B||_F = 58.0430$ and $||C-B||_F = 57.2180$ respectively. It is obvious that $||A-B||_F > ||C-B||_F$. In term of the nearest neighbor decision rule, the original images produce the wrong classification result. However, the correct classification result can be got in the NOMR space with $||\overline{A}-\overline{B}||_F = 1.9948 < ||\overline{C}-\overline{B}||_F = 1.9961$.

What's more, this processing philosophy gives a new insight into handling well the different image modalities recognition problems such as the sketch face image recognition [32,33]. The major challenge of this problem is to match images in different modalities. For example, sketches are a concise representation of human faces, often containing shape exaggeration and having different textures than photos. Different type images just have the different visual appearance, but we believe that the same object should have lied in the same basis space, i.e. the same NOMR. In Section 4, we perform a sketch face image recognition experiment on CUHK and the experimental results show an excellent performance of the proposed NOMR.

Secondly, the NOMR is a non-parameter two-dimensional (2D) holistic description method and can be easily achieved by SVD. Without parameter tuning and learning process, NOMR can save time-costing and, more importantly, avoids the embarrassment in many parametric representation methods. It is noted that the nearest orthogonal matrix of A are uniquely determined, i.e. the NOM is unique, although the SVD of A are not unique (there are $2^{\max(m,n)}$ possible SVD's for a given matrix A). These features make our method simpler, faster, more stable and possess more representation meaning.

Finally, as shown in Fig. 3, the NOM of the original image is still a face image. This feature means that many feature descriptors, e.g. Gabor [27], LBP [34] etc. [35–37] can be further applied after NOMR. In fact, our proposed representation method can be seen as a data preprocessing process, thus the state-of-the-art feature extraction methods or classifiers can be used to further enhance the performance of the classification system.

Remark 2. It should be noted that the NOM representing the specific individual subspace is one kind of useful information hidden in the observation for recognition task. This idea neither excludes the residence of some information in favor of recognition problem from the disguising information (refers to the residual image by removing the NOM), nor opposes the existence of other more helpful representation form for recognition task.

Indeed, the NOMR depends on the image alignment heavily. However, the image alignment is beyond the scope of this article. Thus, without loss of generality, the proposed method just considers the condition that the images have been aligned.

4. Experiments

Five publicly available databases including the Extended Yale B database [38], the CMU-PIE database [39], the FRGCv2 database [40], the AR database [41] and the CUHK sketch face database [32,33] are used in our experiments.

The proposed NOMR is tested and compared with state-of-the-art specific individual subspace based methods: NFL [11], LRC [13], SRC [4], CRC [6], and CESR [10]. In above methods, the L_1-optimizer will be solved by using the matlab function "BPDN_homotopy_function" from the L_1_homotopy package (http://users.ece.gatech.edu/%7Esasif/homotopy/); The L_2-optimizer will be solved by using the "Tikhonov regularization" to avoid the singular condition. CRC and LRC are tuned to achieve their best performance by choosing the optimal regression parameters, and the parameter settings of other methods follow the authors' suggestions. It should be mentioned that all experiments are done on the original face images, without any image preprocessing and feature extraction step.

In order to highlight the representation power of NOMR, we also compared it with some selected and related representation methods: PCA [42], LDA [43], NMF [22], Singular Values (SV) [17] and Improved Singular Values (ISV) [21]. Here, the nearest neighbor rule (NN) with F-Norm distance is used for classification. In addition, the method of the original face image representation and NN is taken as the Baseline.

What's more, the data were randomly permuted 50 times, thus all measures are reported as the average.

4.1. Experiment using the extended Yale B database

The extended Yale B database contains about 2432 frontal face images of 38 individuals under 9 posed and 64 illumination conditions [38]. The 64 images of a person in a particular pose are acquired at camera frame rate of 30 frames/s, so there is only small change in head pose and facial expression. All frontal-face images marked with P00 are used, and each image is resized to 80×80 pixels in our experiments.

In the first experiment, we randomly selected k (1, 4, 8, 16 and 32) images of each subject to form the known set and the remaining images are used for testing. Note that the Fisherface cannot work well with few samples per class. Therefore, we only provide the recognition rates when there are k (4, 8, 16 and 32) known samples per class. For Eigenface and Fisherface, we further extract 200 dimensions and 37 dimensions from the down-sample

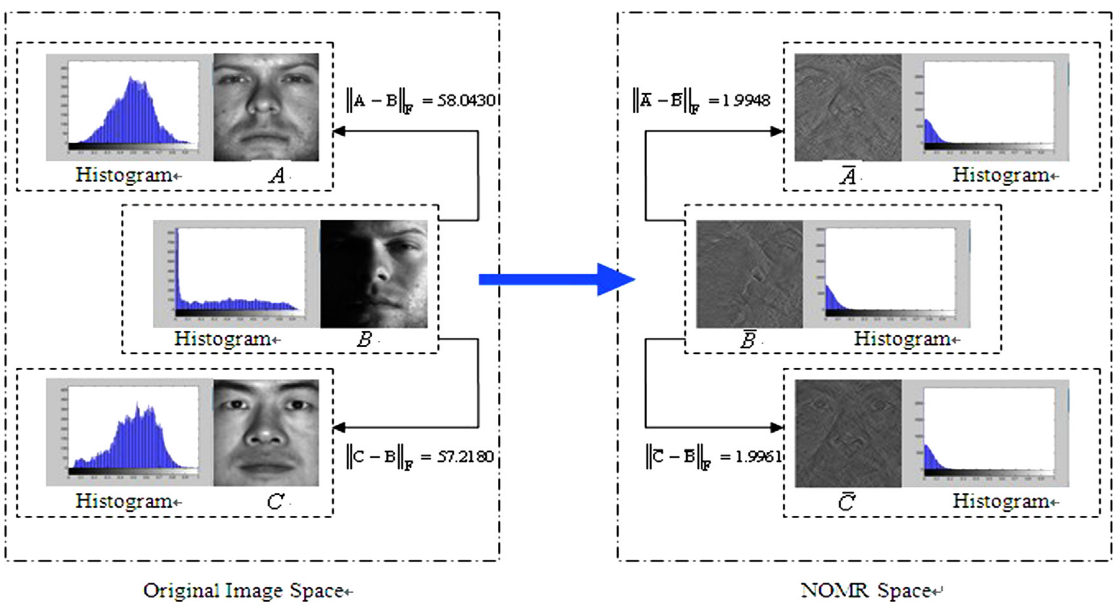

Fig. 4. An example of Original Image fails while NOMR succeeds.

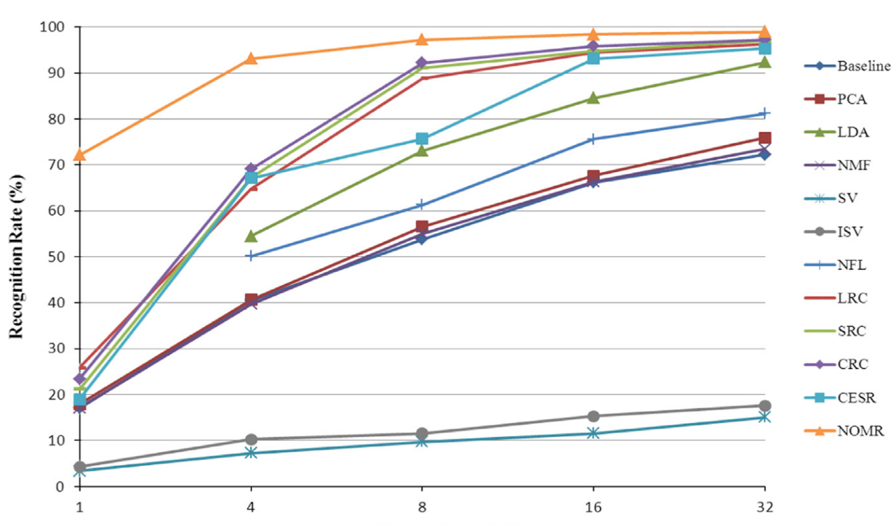

Fig. 5. The average recognition rates (%) on Extended Yale B database.

images respectively. Fig. 5 shows the average recognition rates of Baseline, PCA, LDA, NMF, SV, ISV, NFL, LRC, SRC, CRC, CESR and NOMR versus known sample size.

From Fig. 5, we can see that the proposed NOMR consistently outperforms other methods irrespective of the variation of known sample set size. Especially with the small sample size, such as 4 known samples per class, the NOMR achieves the best recognition rate of 93.13%, which is over 20% higher than the nearest competitor CRC. For the single sample problem, the NOMR significantly outperforms other methods.

In the second experiment, we test the performance of NOMR under various lighting conditions on the Extended Yale B database. The database includes five subsets; subset 1 consisting of 266 images (7 images per subject) under nominal lighting conditions was used as the known samples set, while all others were used for testing. Subsets 2 and 3, each consisting of 12 images per subject, characterize slight-to-moderate luminance variations, while subset 4 (14 images per subject) and subset 5 (19 images per subject) depict severe lighting variations (see Fig. 6). Here, each image is resized to a spatial resolution of 96×84. The results are shown in Fig. 7.

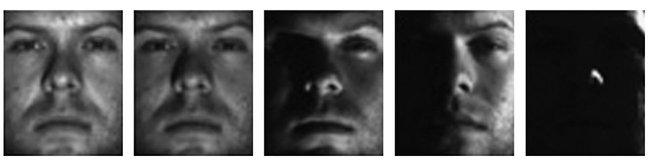

Fig. 6. From left to right, each image illustrates sample from subsets 1, 2, 3, 4, and 5, respectively.

From Fig. 7, we can see that the proposed NOMR shows excellent performance for either moderate lighting variations or severe lighting variations, yielding the highest recognition accuracy for almost all subsets than other methods. Especially, for both subsets 4 and 5 with extreme lighting conditions, the proposed NOMR achieves the best results of 87.1% and 83.4% among all methods. Some robust methods like SRC and CESR do not seem very robust to extreme illumination changes. Both LRC and CRC seem less sensitive to illumination changes than other methods excluding NOMR.

4.2. Experiment using CMU PIE database

The CMU PIE face database contains 68 subjects with over 40,000 face images [39]. Images of each person were taken across 13 different poses, under 43 different illumination conditions, and with 4 different expressions.

Here we use a subset containing images of pose C27 (a nearly frontal pose) of 68 persons, each with 21 different direction illuminated images (see Fig. 8). All images are manually aligned, cropped and resized to be 64×64 pixels [44]. In our experiment, some images of each subject are randomly selected for training,

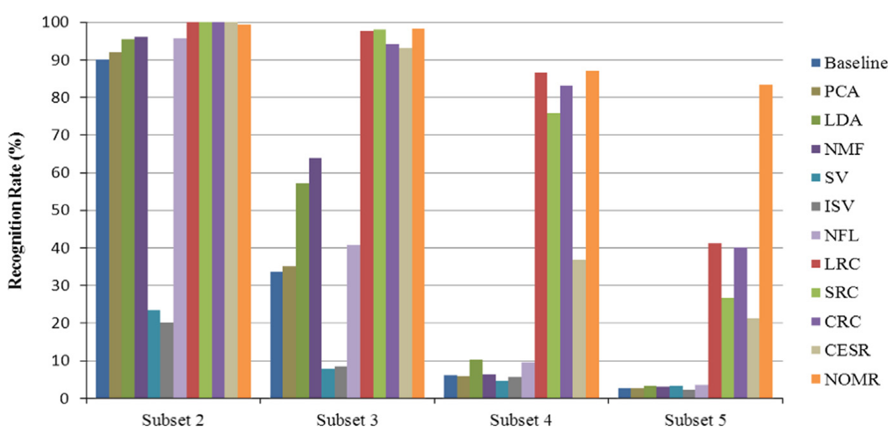

Fig. 7. The recognition rates (%) on Extended Yale B database under various lighting conditions.

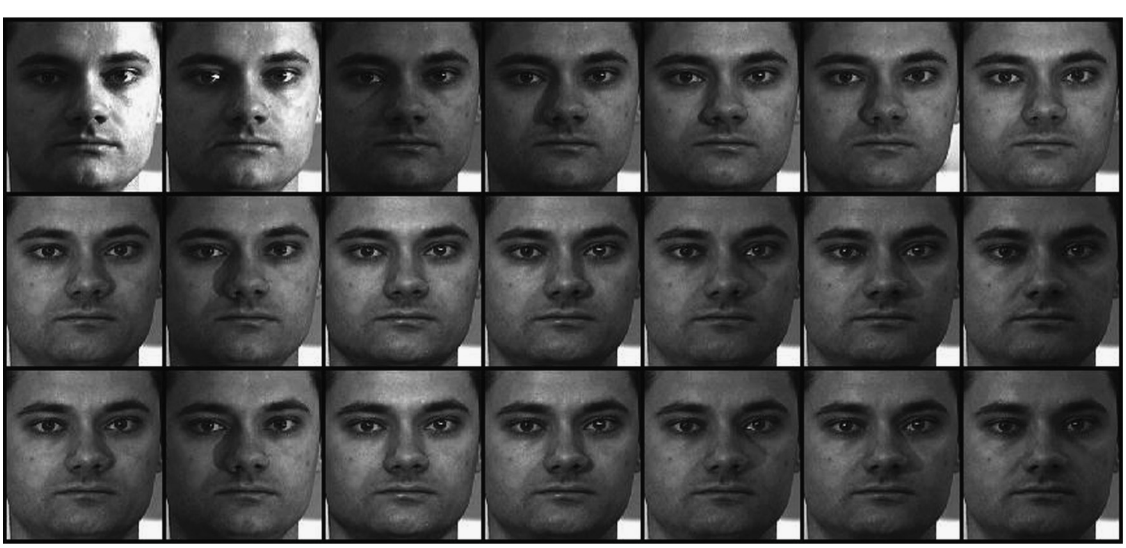

Fig. 8. Illustration of one person from the subset pose C27 of CMU PIE.

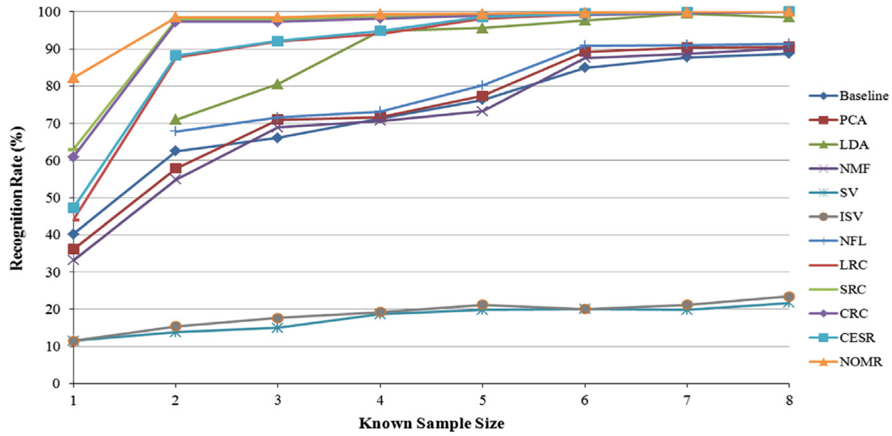

Fig. 9. The average recognition rates (%) on CMU PIE database.

Fig. 10. Illustration of one person from the subset of FRGCv2 database.

and the remaining for test. Note that the Fisherface cannot work well with few samples per class. Therefore, we only provide the recognition rates when there are k (2, 3, 4, 5, 6, 7 and 8) known samples per class. For Eigenface and Fisherface, we further extract 200 dimensions and 67 dimensions from the down-sample images respectively. The average recognition rates versus the known sample size are shown in Fig. 9.

We can see that the proposed NOMR always achieves the best results, but the traditional specific individual subspace based methods like SRC, LRC, CRC and CESR also achieve competitive results in this test. Note that the illumination conditions of images in the CMU-PIE database are much better than that in the Extended Yale B database in foregoing experiment. It seems that the traditional specific individual subspace based methods are insensitive to relatively slight illumination changes. In addition, when the known sample size is 1, i.e. the single sample problem, the proposed NOMR can still get the amazing 82.31% recognition rate, about 20% better than the second-highest SRC.

4.3. Experiment using the FRGCv2 database

The FRGC version 2.0 is a large public scale face image database, including controlled and uncontrolled images [40]. This database contains 12,776 training images (6360 controlled images and 6416 uncontrolled ones) from 222 individuals16,028 controlled target images and 8014 uncontrolled query images from 466 persons for the FRGCv2.0 Experiment 4. The controlled images have good image quality, while the uncontrolled images display poor image quality such as large illumination variations, low resolution of the face region, and possible blurring. Here, we use a subset (222 subjects having 36 samples in the training set) of the Experiment 4 to test the proposed NOMR. The face region of each image is first cropped from the original high-resolution still images and resized to a spatial resolution of 32×32. Fig. 10 shows some cropped face images.

In this experiment, we randomly select 18 images of each person for training and the rest images for testing. The comparison of NOMR and its competing methods is shown in Table 1. From Table 1, we can see that NOMR still achieves the competitive

Table 1
The average recognition rates (%) of NOMR and its competing methods on the FRGCv2 database.

Methods	Recognition rate (%)
Baseline	72.42 ± 1.48
PCA	71.86 ± 3.35
LDA	80.16 ± 2.16
NMF	72.68 ± 3.28
SV	11.46 ± 1.12
ISV	12.26 ± 1.23
NFL	73.26 ± 2.87
LRC	78.26 ± 2.22
SRC	89.26 ± 2.44
CRC	90.18 ± 1.88
CESR	83.96 ± 4.87
NOMR	89.62 ± 1.22

results with the recognition rate $89.62 \pm 1.22\%$. SRC gets the similar result of $89.26 \pm 2.44\%$ with ours. CRC gets the best result of 90.18 ± 1.88, just slightly higher than the proposed NOMR.

4.4. Experiment using AR database

Real face disguise recognition is a challenging problem in FR society. In this experiment, we will test NOMR performance in real face disguise with the AR database [41]. The AR face contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of most persons are taken in two sessions (separated by two weeks). Each section contains 13 color images and 126 individuals participated in both sessions. Here, the subset consists of 3120 images from 120 individuals (26 samples per class, 65 men and 55 women) is used for testing. In this experiment, we evaluate the robustness of NOMR in dealing with real disguise. The images were resized to 30×30. Different from [4–6], we test these method with more challenging condition: for each subject, 7 images were randomly selected from 14 images with only illumination change and expressions as known samples, while others are divided into two separate subsets for testing (with sunglasses and scarf, 6 sample per subject per Session, see Fig. 11)

From Fig. 12, we observe that NOMR achieves the best recognition rate for the test with scarf. For the test with sunglasses, where the occlusion level is relatively low, the traditional specific individual subspace based methods like SRC, LRC, CRC, CESR, NFL and our proposed NOMR can achieve good results. There is no significant performance difference between NOMR and other methods. However, NOMR significantly outperforms others when the test images with scarves.

4.5. Experiment using CUHK sketch face database:

Face photo-sketch image recognition is to match a face sketch drawn by an artist to one of many face photos in the database. In law enforcement, it is desired to automatically search photos from police mug-shot databases using a sketch drawing when the photo of a suspect is not available.

We test the performance of NOMR on the CUHK face sketch database [32,33] compared with Baseline, PCA, NMF, SV, ISV, LRC, SRC, CRC, CESR and the Sketch Transform Method (STM) proposed in [33]. The database CUHK includes 188 people faces from the Chinese University of Hong Kong (CUHK) student database. There are 376 faces in total. For each face, there is a sketch drawn by an artist based on a photo taken in a frontal pose, under normal lighting condition, and with a neutral expression (see Fig. 13).

In this experiment, the CUHK is divided into two subsets: Subset 1 includes 88 photo-sketch pairs; Subset 2 includes others. The STM takes Subset 1 as training data and Subset 2 for testing with its best $d2$ distance strategy (described in [33]). For other methods, including ours, we directly use the photos in Subset 2 as the known data and the corresponding sketches for the testing. We compute the cumulative match score (CMS) and recognition rates with the feature space dimensions 50×40 obtained by the down-sample method. The cumulative match score (CMS) measures the percentage of "the correct answer is in the top n matches," where n is called the rank. The cumulative match scores and the recognition rates are shown in Figs. 14 and 15.

We can see that the proposed NOMR achieves the impressing 100% accuracy when the rank is larger than 3. The SRC, CRC and LRC just obtain the 95% accuracy at most with the rank larger than 7. The Baseline and PCA have the 52% and 67% accuracy at most. As for the sketch transform method (STM), the results show that it gives the 96% accuracy with the tenth rank match. The NMF, SV,

Fig. 11. The top row illustrates samples with Sunglasses; the bottom row illustrates samples with Scarf.

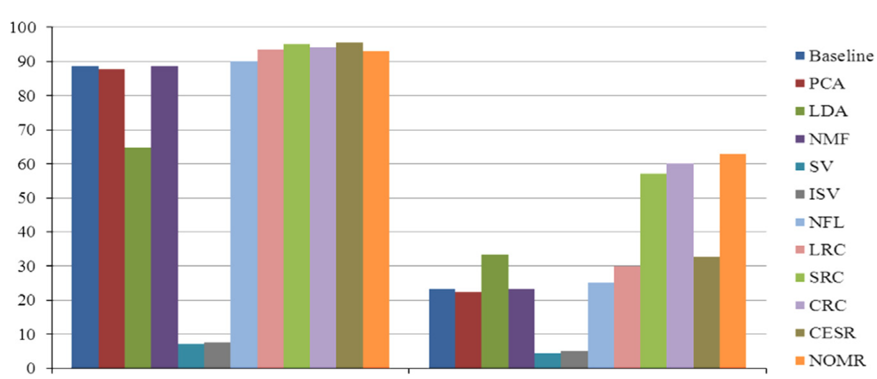

Fig. 12. The average recognition rate (%) on AR database with real face disguise.

Fig. 13. Illustration of some photo and sketch face images from CUHK.

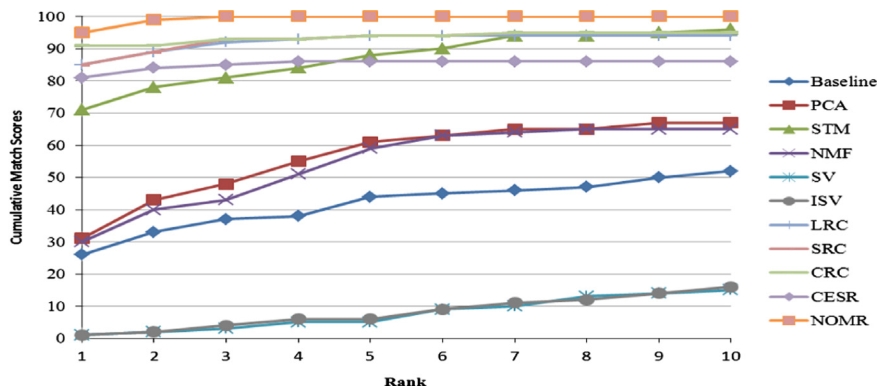

Fig. 14. The Cumulative Match scores with CUHK sketch face database.

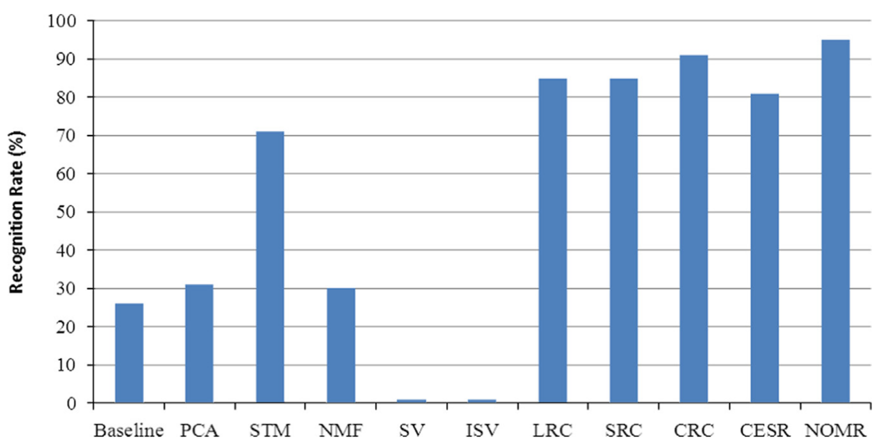

Fig. 15. The recognition rate on CUHK sketch face database.

ISV and CESR obtain the 65%, 15%, 16% and 86% accuracy respectively when the rank meets ten. It is surprising that our proposed NOMR achieves the best 95% recognition rate versus 85% of SRC, 85% of LRC, 91% of CRC, 26% of Baseline, 31% of PCA, 30% of NMF, 81% of CESR, 1% of SV, 1% of ISV and 71% of STM. It seems that the specific individual subspace based methods are insensitive to the sketch heterogeneous problem.

5. Conclusion and future work

This paper presents a novel face recognition method based on the specific individual subspace assumption and uses the nearest orthogonal matrix to represent the original image. The nearest orthogonal matrix representation (NOMR) not only holds the identity features of the original image, but also validly reduces the obstruction of illumination, occlusion and disguise. The NOMR is both simple and practical since it does not involve the parameters and learning process. Experimental results demonstrate that the proposed NOMR can achieve encouraging recognition rate as opposed to the state-of-the-art face recognition methods (SRC, LRC, CRC, CESR and so on). Especially, for the small sample size problem and sketch face recognition, the NOMR significantly outperforms other methods.

Our future work mainly includes applying NOMR to local image representation, object detection and other recognition tasks. In addition, we will explore the performance of the NOMR with other classification algorithms in large scale databases.

Acknowledgments

The authors would like to thank the editor and the anonymous reviewers for their critical and constructive comments and suggestions. This work was partially supported by the National Science Fund for Distinguished Young Scholars under Grant nos. 61125305, 61472187, 61233011, 61373063 and 31270745, the Key Project of Chinese Ministry of Education under Grant no. 313030, the 973 Program no. 2014CB349303, the Priority Academic

Program Development of Jiangsu Higher Education Institutions, Fundamental Research Funds for the Central Universities No. 30920140121005, and Program for Changjiang Scholars and Innovative Research Team in University No. IRT13072.

References

[1] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E..Tabassi, M. Bone, Face Recognition Vendor Test 2002: Evaluation Report, FRVT 2002, March 2003.

[2] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 34 (4) (2003) 399–485.

[3] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.H. Romeny, J.B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Comput. Vis. Graph. Image Process. 39 (1987) 355–368.

[4] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[5] Q.-F. Shi, A. Eriksson, Anton van den Hengel, C.-H. Shen, Is face recognition really a compressive sensing problem?, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[6] L. Zhang, M. Yang, X.-C. Feng, Sparse representation or collaborative representation: which helps face recognition, in: IEEE International Conference on Computer Vision (ICCV), 2011.

[7] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: CVPR, 2011.

[8] M. Yang, L. Zhang, J. Yang, D. Zhang, Regularized robust coding for face recognition, IEEE Trans. Image Process. 22 (5) (2013).

[9] R. He, W.S. Zheng, B.G. Hu, X.W. Kong, A regularized correntropy framework for robust pattern recognition, Neural Comput. 23 (2011) 2074–2100.

[10] R. He, W.S. Zheng, B.G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. PAMI 33 (8) (2011) 1561–1576.

[11] S.Z. Li, J. Lu, Face recognition using nearest feature line method, IEEE Trans. Neural Netw. 10 (2) (1999) 439–443.

[12] J.T. Chien, C.C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Trans. Pattern Anal. Mach. Intell 24 (12) (2002) 1644–1649.

[13] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (no. 11) (2010) 2106–2112.

[14] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[15] R. Barsi, D. Jacobs, Lambertian reflection and linear subspaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 218–233.

[16] G. Shakhnarovich, B. Moghaddam, Face recognition in subspaces, in: Stan Z. Li, Anil K. Jain (Eds.), Handbook of Face Recognition, Springer-Verlag, 2004 (35 p.).

[17] Z.-Q. Hong, Algebraic feature extraction of image for recognition, Pattern Recognit. 24 (3) (1991) 211–219.

[18] Y.-Q. Cheng, et al., Human face recognition method based on the statistical model of small sample size, SPIE Proc. Intell. Robots Comput. Vis. 1607 (1991) 85–95.

[19] L. Hong, A.K. Jain, Integrating faces and fingerprint for personal identification, IEEE Trans. Pattern Anal. Mach. Intell. 20 (12) (1997) 1295–1307.

[20] Y.-H. Wang, T.-N. Tan, Y. Zhu, Face identification based on singular values decomposition and data fusion, Chin. J. Comput. 23 (6) (2000) 013.

[21] Y. Tian, T.-N. Tan, Y.-H. Wang, Y.-C. Fang, Do singular values contain adequate information for face recognition? Pattern Recognit. 36 (3) (2003) 649–655.

[22] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (21) (1999) 788–791.

[23] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334.

[24] D. James, K. William, Accurate singular values of bidiagonal matrices, Soc. Ind. Appl. Math. J. Sci. Stat. Comput. 11 (5) (1990) 873–912.

[25] P.-H. Schönemann, A generalized solution of the orthogonal procrustes problem, Psychometrika 32 (1) (1996) 1–10 (March).

[26] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK (1994) 414–415.

[27] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[28] D.Q. Zhang, S.C. Chen, Z.H. Zhou, A new face recognition method based on SVD perturbation for single example image per person, Appl. Math. Comput. 163 (2005) 895–907.

[29] H. Demirel, G. Anbarjafari, M.N.S. Jahromi, Image equalization based on singular value decomposition," in: Proceedings of 23rd ISCIS, 2008, pp. 1–5.

[30] L. Wang, X. Wang, J. Feng, Subspace distance analysis with application to adaptive Bayesian algorithm for face recognition, Pattern Recognit. 39 (3) (2006) 456–464.

[31] L. Wang, X. Wang, J. Feng, Intrapersonal subspace analysis with application to adaptive Bayesian algorithm for face recognition, Pattern Recognit. 38 (4) (2005) 617–621.

[32] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 31 (2009).

[33] X.O. Tang, X.Y. WANG, Face sketch recognition, IEEE Trans. Circ. Syst. Video Technol. 14 (1) (2004).

[34] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.

[35] J.J. Qian, J. Yang, G.W. Gao, Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction, Pattern Recognit. 46 (10) (2013) 2724–2739.

[36] D. Zhang, Z.H. Zhou, (2D) 2PCA: two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (1) (2005) 224–231.

[37] S.R. Zhou, J.P. Yin, J.M. Zhang, Local binary pattern (LBP) and local phase quantization (LBQ) based on Gabor filter for face representation, Neurocomputing 116 (2013) 260–264.

[38] K.C. Lee, J. Ho, D. Driegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. PAMI 27 (5) (2005) 684–698.

[39] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-PIE,, Image Vis. Comput. 28 (5) (2010) 807–813.

[40] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 947–954.

[41] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998.

[42] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.

[43] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, J. Optical Society. America 14 (1997) 1724–1733.

[44] D. Cai, X. He J. Han, Spectral Regression for Efficient Regularized Subspace Learning, IEEE 11th International Conference on Computer Vision (ICCV 2007).

Jian Zhang received his B.S. degree and M.S. degree in 2001 and 2006, respectively. Now, he is working for his Ph.D. degree in the Nanjing University of Science and Technology (NUST), on the subject of Pattern Recognition and Intelligence System. His current interests include pattern recognition, computer vision and machine learning.

Jian Yang received the BS degree in mathematics from the Xuzhou Normal University in 1995. He received the M.S. degree in applied mathematics from the Changsha Railway University in 1998 and the Ph.D. degree from the Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 80 scientific papers in pattern recognition and computer vision. His journal papers have been cited more than 1600 times in the ISI Web of Science, and 2800 times in the Web of Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and IEEE Trans. Neural Networks, respectively.

Jianjun Qian received the B.S. and M.S. degrees in 2007 and 2010, respectively. Currently, he is pursuing the Ph. D. degree at the School of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), Nanjing, China. His research interests include pattern recognition, computer vision and image feature extraction in particular.

Jiawei Xu received the B.S. and M.S. degrees in computer engineering from Shanghai University of Engineering Science and Technology, Shanghai, China, 2007 and Hallym University, Korea, 2010, respectively. Now he is a PhD student in the School of Computer Science, University of Lincoln, UK. His research interests include computer vision, human attention models, and visual cortex modeling. He was a pattern classification engineer in JTV Co. Ltd, Beijing during the year of 2011.