



Multiple feature subspaces analysis for single sample per person face recognition

Yongjie Chu^{1,2} · Lindu Zhao¹ · Touqeer Ahmad²

Published online: 16 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Collecting samples is one of the main difficulties for face recognition, for example, in most of the real-world applications such as law enhancement, e-passport, and ID card identification, it is customary to collect a single sample per person (SSPP). Unfortunately, in such SSPP scenario, many presented face recognition methods suffer serious performance drop or fail due to their inability to learn the discriminative information of a person from a single sample. To address the SSPP problem, in this paper, we propose a multiple feature subspaces analysis (MFSA) approach, which takes advantage of facial symmetry. First, we divide each enrolled face into two halves about the bilateral symmetry axis and further partition every half into several local face patches. Second, we cluster all the patches into multiple groups according to their locations at the half face and formulate SSPP as a MFSA problem by learning a feature subspace for each group, so that the confusion between inter-class and intra-class variations of face patches is removed and more discriminative features can be extracted from each subspace. To recognize a target person, a k -NN classifier is employed in each subspace to predict the label of a face patch and majority voting strategy is used to identify the unlabeled subject. Compared with the state-of-the-art methods, MFSA is effortless and efficient in implementing, but achieves either better or competitive performance when recognizing face images taken in both constrained and unconstrained environment.

Keywords Single sample per person · Facial symmetry · Multiple feature subspaces analysis · Patch-based method

1 Introduction

Face recognition has proven to be a significant component in the field of biometrics recognition systems. Compared to other biometric techniques, e.g., fingerprints, iris scans, and speech recognition, two key advantages of face recognition

are its non-intrusiveness and mass identification ability. That is, it does not require the cooperation of the test subjects and can identify individuals among crowd. Due to these merits, face recognition techniques have been widely adapted in surveillance, access control, law enforcement etc., to name a few. Numerous methods have been proposed in the past two decades to robustly identify faces under controlled condition [1–5], and a great deal of progress has been made in recent years to rise to the challenge of large appearance variations of a face caused by illuminations [6–10], expressions [8,9,11–13], poses [6–9,12,13], and occlusions [8,13–16] in the unconstrained environment. These methods usually assume that multiple samples per person (MSPP) are available for training a face recognition system. However, in many real-world scenarios such as law enforcement, e-passport and ID card identification, only a single sample per person (SSPP) is enrolled or recorded. Though the unsupervised techniques like principal component analysis (PCA) [1], local binary pattern (LBP) [4], and Gabor features [5] can still be applied to deal with SSPP; such popular unsupervised methods suffer from serious performance drop as they are designed for

This work was partially supported by the National Natural Science Foundation of China (No. 71390333), the National Key Technology R&D Program of China during the 12th Five-Year Plan Period (No. 2013BAD19B05), the Special Fund for Basic Research in Central University, i.e., Fund for Graduate Research Project of Jiangsu Province (KYZZ15_0072), and the Study Abroad Program for Graduate Studies by China Scholarship Council.

✉ Lindu Zhao
ldzhao@seu.edu.cn
Yongjie Chu
yongjiechu@yeah.net

¹ Institute of Systems Engineering, Southeast University, Nanjing 211096, China

² Department of Computer Science and Engineering, University of Nevada, Reno, NV 89557, USA

Fig. 1 Partition of a face image



addressing MSPP face recognition. The conventional supervised methods like linear discriminant analysis (LDA) [17] and marginal fisher analysis (MFA) [18] fail to cope with SSPP problem due to unavailability of enough face images in the training set for good estimation of intrapersonal variations.

Lots of work have been done to deal with SSPP face recognition, which can be roughly categorized as global appearance methods such as projection-combined PCA ($(PC)^2A$) [19], Stringface [20], and uniform pursuit (UP) approach [21], and patch-based (local appearance) methods as Block LDA [22], discriminative multi-manifold analysis (DMMA) [23], sparse variation dictionary learning (SVDL) [24], etc. Because patch-based methods can easily avoid the effect of severely corrupted non-informative regions, they can generally achieve better performance than global appearance approaches and draw more attention from researchers. Most of the patch-based methods [22,25,26] are proposed to use all patches together to explore intra-class and inter-class information and then to construct a single subspace to extract features. Note that when a face image is partitioned into several patches, these patches represent different parts (semantics) of a face, such as eyebrows, eyes, cheeks, nose, and mouth. Though the patches from the same face share the same label or identity, local face patches at the same location (e.g., nose) of different persons are more similar than those at different locations (e.g., nose and mouth) of the same person. As a result, it is prone to confuse the inter-class and intra-class variations if only one subspace is constructed for feature extraction; this goes against extracting discriminative features for recognition.

Given a frontal face image, it is inherently symmetric about a bilateral symmetry axis, though not perfectly. Facial symmetry has previously been exploited to assist face detection [27] and pose estimation [28]. Recent research in the area of automatic face recognition has shown that there may be an advantage in utilizing facial symmetry to improve recognition accuracy. To handle the pose variations in face recognition, both [29] and [30] take advantage of facial symmetry. Further, [31] and [32] conduct extensive experiments on 2D and 3D databases to compare the face recognition

results using the average half face and full face; they draw the similar conclusion that the average half face tends to perform better than the full face. In [33], the facial symmetry is used to split a face sample into two to augment the number of samples in the training set, so that more variation information can be exploited. On the one hand, by utilizing the facial symmetry the dimensionality of data can be reduced, while the symmetry presented in the data can be preserved, whereas on the other hand, the training set can be augmented when there is limited number of training samples.

To benefit from advantages of patch-based methods and to overcome their disadvantages, in this paper, we propose a multiple feature subspaces analysis (MFSA) method for SSPP face recognition problem, where we try to exploit most of facial symmetry. In our method, specifically, each face image is divided into two halves about the bilateral symmetry axis, then the right half is mirrored to left across the symmetry axis, and then each half face is further partitioned into non-overlapping patches. Figure 1 displays an illustration of the partition of a frontal face image. Next, all these local face patches are clustered into different groups according to their locations (or semantics) at a half face level. The proposed MFSA is designed to learn a feature subspace for each group of patches, which enlarges the interpersonal margins and reduces the intrapersonal variations, so that the confusion between the inter-class and intra-class variations of face patches is removed and more discriminative features can be extracted in each subspace. In the recognition phase, each local patch of a probe face is projected into the corresponding feature subspace to perform feature extraction, and then its label is predicted by a k -nearest neighbor (k -NN) classifier; finally, we concatenate the predicted labels of all face patches and employ majority voting strategy to identify the unlabeled subject. Figure 2 illustrates the basic idea of the proposed approach.

Though the motivation of MFSA is to use the symmetry of frontal faces, we declare that it also works for non-frontal gallery faces, as we show in the experiments. The reasons is as follows. When the gallery faces are not frontal, MFSA still partitions these faces along the symmetry axis of images (not faces) and then learns feature subspaces for each group

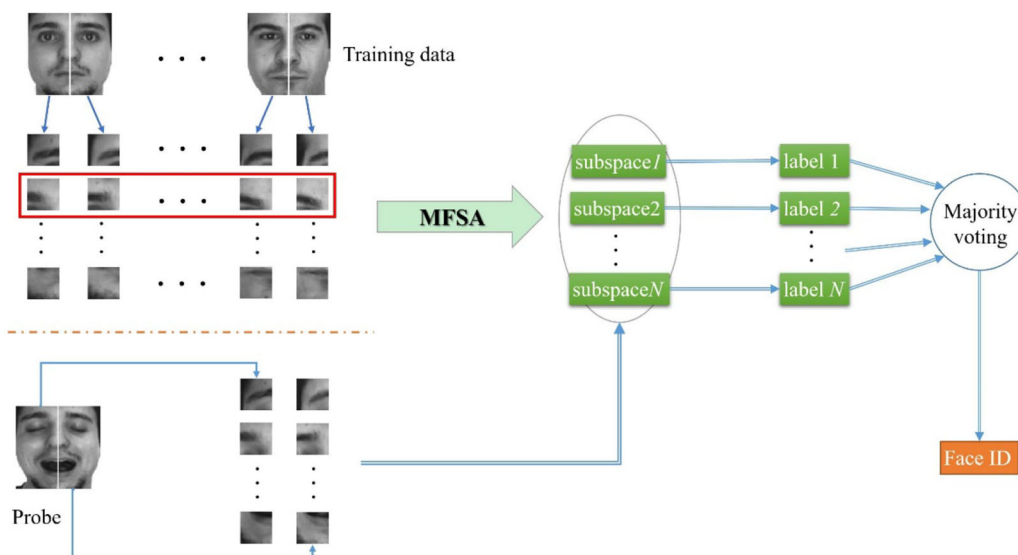


Fig. 2 Illustration of the proposed method

of patches. Note that in the testing phase, there are also many non-frontal faces in probe set. So there is some correspondence between gallery and probe faces about the non-frontal nature, and this makes MFSA work.

The main contributions of this paper can be summarized as follows: (i) facial symmetry and patch-based trick are unified to address SSPP problem, which not only enlarges the number of samples per subject, but also enables to explore sophisticated intra-class variations. To the best of our knowledge, it is the first attempt where facial symmetry and patch-based trick are integrated to deal with SSPP. (ii) A multiple feature subspaces analysis approach for SSPP face recognition is proposed. By using the facial symmetry and patch-based trick, the proposed MFSA transforms SSPP recognition into multiple discriminative subspaces learning problem, which avoids the disadvantage of the conventional patch-based methods that they tend to confuse the inter-class and intra-class variations. (iii) We conduct extensive experiments to evaluate the robustness of MFSA to pose, expression and occlusion variations. Experimental results show that in each case, MFSA is either competitive or superior to the state-of-the-art approaches for SSPP face recognition.

The rest of the paper is organized as follows. In Sect. 2, work related to SSPP face recognition is reviewed. In Sect. 3, we introduce the details of the proposed approach. Section 4 provides the experimental results and some discussions. Finally, we conclude the paper in Sect. 5.

2 Related work

Many exclusive approaches have been proposed to address the SSPP face recognition problem in the last few years.

These methods can be mainly classified into three categories: the unsupervised methods, supervised methods, and semi-supervised methods.

To boost the performance of traditional unsupervised methods, projection-combined PCA ((PC)²A) [19], enhanced projection-combined PCA (E(PC)²A) [34], and two-directional two-dimensional PCA ((2D)²PCA) [35] have successively been proposed. The idea behind these is to mine more global information. For instance, the first-order projection information in [19], the second-order projection information in [34], as well as the row and column information in [35] are collected by performing PCA on the limited gallery images. However, such improved methods extract only global information from the training images, which limits their effectiveness on SSPP. To further enhance the PCA-based unsupervised techniques, a uniform pursuit (UP) [21] approach and two-stage block-based whitened PCA (TS-BWPCA) [36] are designed. The UP [21] incorporates the neighborhood information into PCA to reduce the local confusion between the similar faces, while TS-BWPCA [36] is a coarse-to-fine scheme that extracts both global and local information of images by embedding local binary pattern (LBP) descriptor. To make more use of the local information in unsupervised scenario, [25] proposes to divide a face into smaller sub-images; PCA is then applied to each of these sub-images. The mentioned three approaches [21,25,36] can demonstrate good performance on some simple datasets, but fail to display promising results when significant variances are involved in the probe images, as they fail to exploit the intra-class information.

To utilize the supervised information and cover the intra-class variations in the training process, many supervised methods have been developed for SSPP problem, which can

be further categorized into three subclasses: virtual samples-based methods, generic set-based methods, and patch-based methods. These methods obtain multiple training samples per person so that within-class information can be extracted and discriminative features can be extracted. In the virtual samples-based methods, multiple virtual training images per subject are generated from the gallery images, which have the same size as that of gallery images. Generating virtual images can be accomplished by a small SVD perturbation [37], some transformations [38], or some decompositions [39]. However, such virtual images are generated highly based on the limited gallery images, resulting in substantial correlation and making the extracted features redundant. Generic set-based methods [24,40–47] introduce a separate image dataset for training, i.e., a generic set. The generic set which includes label information and possible variations of faces is used to estimate the interpersonal and intrapersonal variations of the gallery set. Occasionally, some learning tricks are further employed to enhance such approaches [40,41]. Most recently, sparse representation [8] has shown very effective face recognition performance and has been introduced to mine variations from generic set to address SSPP [24,43,45–47]. These sparse representation-based methods share a common idea, i.e., learning a variation dictionary from the generic set to approximate variations of probe faces. Since the generic and gallery sets do not necessarily share the similar inter-class and intra-class variations, the estimation of these variations using generic sets will be unfaithful, which results in degraded effectiveness of such methods. The last subclass is the patch-based methods [22,23,26,42,44,48–52]. As the name suggests, in patch-based methods, each image in the gallery set is divided into small blocks or patches; thus, there are multiple patches for each subject and intra-class variations can be measured. After obtaining the block images, [22] and [26] employ linear discriminant analysis (LDA) and self-organizing map (SOM), respectively, for feature extraction. Gao et al. [42] is a sparse representation-based method, which imposes some sparsity constraints when learning the reconstruction coefficients and the intra-class variance dictionaries. Lu et al. [23] proposes to formulate SSPP face recognition as a manifold–manifold matching problem and to extract features from multiple feature spaces to maximize the manifold margins of different persons. Yan et al. [49] improves [23] by employing multiple feature descriptors to learn manifolds. Zhang et al. [51] is also a manifold embedding method; it constructs two sparse graphs instead of using k -nearest neighbors to measure the similarity among samples during the manifold learning. Based on collaborative representation [53], a patch-based collaborative representation method is proposed in [50] by operating collaborative representation on patches and combining the recognition outputs of all patches. To make collaborative representation robust for SSPP problem, [52] proposes to further divide local patches

into overlapped blocks to capture the local structure relationship in the faces. A distinctive method is proposed in [48] by simulating the mechanism of fixation and saccades in human visual perception. This method uses dynamic image-to-class warping [54] technique for matching; it requires no training phase, but displays satisfactory performance for SSPP face recognition problem. The patch-based methods are more likely to be robust to local changes since the variation presented in a face is divided into small ones along with the partition, whereas such methods sometimes suffer from confusing the inter-class and intra-class information as we demonstrated previously.

Apart from the above stated methods, semi-supervised approaches are also applied to address the SSPP problem. For example, [55] proposes to utilize side information (weak label information) to calculate the within-class and between-class scatter matrices when there is no full-class label information. Yin et al. [56] presents a semi-supervised method named double linear regressions (DLR). DLR seeks the best discriminating subspace and preserves the sparse representation structure by first propagating the label information to the unlabeled data and then extracting features with the help of the propagated labeled dataset.

3 Proposed approach

3.1 Image partition and patches cluster

Let $M = [I_1, I_2, \dots, I_c]$ be the training set, I_i is the training image of the i th person with a size of $m \times n$, $1 \leq i \leq c$, c is the number of persons in the training set. Given a frontal face image, to utilize its symmetry, we first divide it into two halves about the bilateral symmetry axis and mirror the right half of a face to the left across the symmetry axis, and thus we get two highly similar half face images for the frontal training face image. Then, for each half face image, it is further partitioned into N non-overlapping local patches, and all patches are with the same size of $a \times b$, where $N = (m \times n)/(2(a \times b))$. We arrange these patches from one half face in the raster scan order (i.e., from left to right and top to bottom) and get a sequence containing N local image patches. Hence, two patch sequences are acquired for each face image. An illustration of face partitioning is shown in Fig. 1. For all the face images in the training set, we repeat the above process of face partition and get $2c$ patch sequences in total for the entire training set, i.e., two for each training sample. All the face patches inherit the label information of the original face image.

Given a patch sequence, the N local patches in it are sheared from N different locations of a half face. In different patch sequences, those local patches sheared from the same location generally have the same semantic, and thus

share more similarity. For example, the two patches shown in the rectangle drawn in Fig. 1 are intuitively more similar to each other than to any other patches, because they both are at the 6th location of a sequence and roughly represent the nose and upper lip. Based on this observation, we cluster local face patches sheared from the same location of all half face images into one group and obtain N such groups of local patches; each group consists of $2c$ patches with 2 for each subject. In Fig. 2, the red rectangle encloses one group of face patches.

3.2 Formulation of MFSA

3.2.1 Whitening processing

A group of local face patches includes $2c$ patches of c persons with 2 from each subject. As mentioned earlier, local patches in the same group have nearly the same semantic and represent almost the same part of different faces, so there lays significant similarity and correlation among them. It is well known that the whitening transformation can remove the correlations among the input data and reduce the redundancy [21]. Also, in the application of face recognition, whitening transformation is able to address the shortcoming of traditional PCA by lowering the weights of leading eigenvectors [57], which encode mostly illumination and expression rather than discriminative information. To de-correlate the training data in the same group, whitening is performed by scaling each principle direction of PCA to uniform the spread of the data.

Let $G_p = \{x_{p\kappa i}\}$ denote the p th group of local face patches, in which $p = 1, 2, \dots, N$, $\kappa = 1, 2$, $i = 1, 2, \dots, c$; $x_{p\kappa i}$ is the κ th patch of the i th person in the p th group, which is a d -dimensional column vector. The covariance matrix of face patches in group G_p can be computed by

$$S_p = \sum_{i=1}^c \sum_{\kappa=1}^2 (x_{p\kappa i} - \bar{x}_p)(x_{p\kappa i} - \bar{x}_p)^T, \tag{1}$$

where $\bar{x}_p = \frac{1}{2c} \sum_{i=1}^c \sum_{\kappa=1}^2 x_{p\kappa i}$ is the average face patch of G_p . The principle component directions are exactly the eigenvectors of the covariance matrix and can be acquired by matrix decomposition as

$$S_p = U \Lambda U^T, \tag{2}$$

where U is an orthogonal matrix and Λ is a diagonal matrix. The whitening transformation matrix is expressed as

$$\varphi_p = U \Lambda^{-1/2}. \tag{3}$$

With the help of whitening transformation, the whitened face patches are obtained by

$$X_{p\kappa i} = \varphi_p^T (x_{p\kappa i} - \bar{x}_p). \tag{4}$$

3.2.2 Discriminative subspaces learning

After acquiring the whitened data of N groups of face patches, our aim is to seek N discriminative subspaces $\{w_p | p = 1, 2, \dots, N\}$, $w_p \in R^{d_\ell \times d_p}$, where d_ℓ is the dimension of the whitened data vector, d_p is the dimension of the p th discriminative subspace, which can simultaneously maximize the inter-class separability and minimize the intra-class variance of all groups in the low-dimensional feature subspaces. To achieve this goal, we can empirically formulate the following optimization problem:

$$\begin{aligned} \max_{w_1, w_2, \dots, w_N} J_1(w_1, w_2, \dots, w_N) \\ = \frac{\sum_{p=1}^N \left(\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{s=1}^{k_1} A_{p\kappa i s} \left\| w_p^T X_{p\kappa i} - w_p^T Y_{p\kappa i s} \right\|^2 \right)}{\sum_{p=1}^N \left(\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{t=1}^{k_2} B_{p\kappa i t} \left\| w_p^T X_{p\kappa i} - w_p^T Z_{p\kappa i t} \right\|^2 \right)} \end{aligned} \tag{5}$$

The numerator and denominator of Eq. (5) respectively formulate the total between-class and within-class scatters of all face patches. In Eq. (5), $Y_{p\kappa i s}$ represents the s th sample in the k_1 -nearest inter-class neighbors of $X_{p\kappa i}$; $Z_{p\kappa i t}$ denotes the t th sample in the k_2 -nearest intra-class neighbors of $X_{p\kappa i}$ in the group G_p ; $A_{p\kappa i s}$ and $B_{p\kappa i t}$ are two affinity matrices to characterize the similarity between $X_{p\kappa i}$ and $Y_{p\kappa i s}$ as well as that between $X_{p\kappa i}$ and $Z_{p\kappa i t}$, respectively. The affinity matrix A and B can be computed by some graph construction techniques, as [23,49] do, and we apply k -nearest neighbor method to calculate A and B as follows:

$$A_{p\kappa i s} = \begin{cases} \exp\left(-\|X_{p\kappa i} - Y_{p\kappa i s}\|^2 / \sigma^2\right), & \text{if } Y_{p\kappa i s} \in Ne_{\text{inter}}^{k_1}(X_{p\kappa i}) \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

$$B_{p\kappa i t} = \begin{cases} \exp\left(-\|X_{p\kappa i} - Z_{p\kappa i t}\|^2 / \sigma^2\right), & \text{if } Z_{p\kappa i t} \in Ne_{\text{intra}}^{k_2}(X_{p\kappa i}) \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where $Ne_{\text{inter}}^{k_1}(X_{p\kappa i})$ and $Ne_{\text{intra}}^{k_2}(X_{p\kappa i})$ represent the k_1 -nearest inter-class neighbors and k_2 -nearest intra-class neighbors of $X_{p\kappa i}$, respectively; k_1, k_2 , and σ are three empirically pre-specified parameters. Obviously, k_2 equals to 1 because there are only two samples from the same class in each patch group. According to the similarity measure, for two samples from different classes, if the distance between them in the original space is small, i.e., these two samples share more similarity than difference, $A_{p\kappa i s}$ will load a relatively large weight on the inter-class separability, so that the projections of these two samples in the feature subspace are more separable.

As the optimization in Eq. (5) involves N projection matrices that need to be simultaneously optimized, there is no closed-form solution to this optimization problem. Noting

that face patches in the same group have nearly the same semantic, for example, they all represent the eye, so there is high correlation among them; while those from different groups have the quite different semantics, for example, patches in one group represent the nose and those in the other group denote the mouth, and hence there is little correlation between different groups. Based on this observation, we make an assumption that the N groups of face patches are nearly independent to each other, such that each group can be treated separately. In view of this, we can degenerate the optimization problem defined in Eq. (5) to the following new formulation:

$$\begin{aligned} \max_{w_1, w_2, \dots, w_N} J_2(w_1, w_2, \dots, w_N) \\ = \sum_{p=1}^N \left(\frac{\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{s=1}^{k_1} A_{p\kappa is} \left\| w_p^T X_{p\kappa i} - w_p^T Y_{p\kappa is} \right\|^2}{\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{t=1}^{k_2} B_{p\kappa it} \left\| w_p^T X_{p\kappa i} - w_p^T Z_{p\kappa it} \right\|^2} \right). \end{aligned} \tag{8}$$

The optimization problem formulated in Eq. (8) can be solved by a parallel way. Given w_p , $p = 1, 2, \dots, N$, Eq. (8) can be rewritten as

$$\begin{aligned} \max_{w_p} J_2(w_p) &= \frac{\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{s=1}^{k_1} A_{p\kappa is} \left\| w_p^T X_{p\kappa i} - w_p^T Y_{p\kappa is} \right\|^2}{\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{t=1}^{k_2} B_{p\kappa it} \left\| w_p^T X_{p\kappa i} - w_p^T Z_{p\kappa it} \right\|^2} \\ &= \frac{\text{tr} \left(w_p^T \left(\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{s=1}^{k_1} A_{p\kappa is} (X_{p\kappa i} - Y_{p\kappa is})(X_{p\kappa i} - Y_{p\kappa is})^T \right) w_p \right)}{\text{tr} \left(w_p^T \left(\sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{t=1}^{k_2} B_{p\kappa it} (X_{p\kappa i} - Z_{p\kappa it})(X_{p\kappa i} - Z_{p\kappa it})^T \right) w_p \right)} \\ &= \frac{\text{tr} \left(w_p^T S_b^p w_p \right)}{\text{tr} \left(w_p^T S_w^p w_p \right)}, \end{aligned} \tag{9}$$

where

$$S_b^p = \sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{s=1}^{k_1} A_{p\kappa is} (X_{p\kappa i} - Y_{p\kappa is})(X_{p\kappa i} - Y_{p\kappa is})^T, \tag{10}$$

$$S_w^p = \sum_{i=1}^c \sum_{\kappa=1}^2 \sum_{t=1}^{k_2} B_{p\kappa it} (X_{p\kappa i} - Z_{p\kappa it})(X_{p\kappa i} - Z_{p\kappa it})^T \tag{11}$$

are the inter-class and intra-class scatter matrix for G_p , respectively. Consequently, the above subproblem is turned into solving an LDA subspace, and the projection matrix w_p can be acquired by solving the well-known eigen-decomposition problem:

$$(S_w^p + \gamma I_o)^{-1} S_b^p w_p = w_p \Psi, \tag{12}$$

where I_o is an identity matrix and γ is a regularization parameter, γI_o is added to S_w^p to avoid singularity problem, the value of γ is empirically set as 10^{-5} in this paper; Ψ is a diagonal matrix and its diagonal values are the first d_p largest eigenvalues of $(S_w^p)^{-1} S_b^p$, w_p is a matrix whose column vectors are the eigenvectors corresponding to the eigenvalues in Ψ . Reminding that the input data are firstly processed by the whitening transformation, we thus obtain the ultimate projection matrix of the feature subspace corresponding to the p th group of face patches, which can be formulated as:

$$W_p = \varphi_p w_p. \tag{13}$$

By solving the subproblem as (9) for each group in a parallel way, we get N feature subspaces for the N independent groups. The feature of a face patch can be trivially extracted by projecting the patch into its corresponding feature subspace. The N feature subspaces can be with various dimensions as they are independent. There is one point that needs to be noted. Since the dimension of face patches is quite small and further reduced by whitening processing, we use Eq. (11) to capture the intra-class variations in a very low-dimensional subspace. In such a low-dimensional subspace, Eq. (11) is able to well smooth the intra-class variations even if there are only two samples for each person in the training set. This is an important factor making the proposed approach effective for SSPP. *Algorithm 1* summarizes the detailed procedure of the proposed MFSA method.

Algorithm1: The proposed MFSA

Input: Group of face patches $\{G_p | p=1,2,\dots,N\}$, where $G_p = \{x_{p11}, x_{p21}, \dots, x_{p\kappa 1}, \dots, x_{p2c}\}$, $x_{p\kappa i} \in \mathbb{R}^d$, $i=1,2,\dots,c$, and $\kappa=1,2$; the value of the parameter k_1, k_2 and σ .

Output: Projection matrices of multiple feature subspaces: $\{W_p | W_p \in \mathbb{R}^{d \times d_p}, p=1,2,\dots,N\}$.

For $p=1,2,\dots,N$, repeat

Step1 (Whitening processing)

1.1 Computer the covariance matrix S_p of G_p according to Eq. (1), and do matrix decomposition on S_p by $S_p = U\Lambda U^T$.

1.2 Select the first d_ℓ largest eigenvalues from Λ to construct a diagonal matrix Λ_o , and select d_ℓ eigenvectors corresponding to the first d_ℓ largest eigenvalues from U to form a eigenvector matrix U_o .

1.3 Compute the transformation matrix by $\varphi_p = U_o \Lambda_o^{-1/2}$ and the whitened data by $X_{p\kappa i} = \varphi_p^T (x_{p\kappa i} - \bar{x}_p)$, in which $\bar{x}_p = \frac{1}{2c} \sum_{i=1}^c \sum_{\kappa=1}^2 x_{p\kappa i}$.

Step2 (Discriminative subspaces learning)

2.1 For the obtained whitened data, calculate two affinity matrices A and B as shown in Eq. (6) and Eq. (7), respectively.

2.2 Calculate inter-class scatter matrix S_b^p and intra-class scatter matrix S_w^p according to Eq. (10) and Eq. (11), respectively.

2.3 Solve the Eigen decomposition problem defined in Eq. (12).

2.4 Sort and select d_p eigenvectors $[V_{p1}, V_{p2}, \dots, V_{pd_p}]$ according to the first largest d_p eigenvalues.

2.5 Obtain the discriminative projection matrix $W_p = [V_{p1}, V_{p2}, \dots, V_{pd_p}]$.

Step3: (Output the feature subspaces)

Output the projection matrix of the p th feature subspace: $W_p = \varphi_p W_p$.

End

3.2.3 Recognition

In the recognition phase, given a probe sample T , we first partition it into $2N$ non-overlapping face patches and group them into N independent groups; each group consists of two patches (as described in Sect. 3.1 and demonstrated in Fig. 1). We name the face patches of a probe face as probe face patches. Let $G_{Tp} = \{\hat{x}_{p1}, \hat{x}_{p2}\}$ denote the probe face patches in the p th group, where $p = 1, 2, \dots, N$. Having obtained the learned projection matrices of feature subspaces $\{W_p | p=1, 2, \dots, N\}$, we project the probe patches in the p th group into the p th feature subspace to extract features. Different from other patch-based methods which concatenate the features extracted from different face regions to form a long vector and then classify it, we make classification of probe face patches in different subspaces independently. For a probe face patch \hat{x}_{pe} in G_{Tp} , $e = 1, 2$, one way to assign a label g to it is to apply the well-known Euclidean distance-based $1 - NN$ (nearest neighbors) classifier:

$$g = \arg \min_i dis(W_p^T \hat{x}_{pe}, W_p^T x_{pi\kappa}), \tag{14}$$

where $i = 1, 2, \dots, c, \kappa = 1, 2$, and $dis(W_p^T \hat{x}_{pe}, W_p^T x_{pi\kappa})$ is the Euclidean distance between the projections of the probe face patch and training face patches in the p th subspace.

Note that face patches at the same location of different faces are quite similar to each other and so are the extracted

feature from them. Moreover, as the input face patches are processed by whitening transformation, the dimension of the whitened data d_ℓ is much less than d , making the dimensions of the feature subspaces quite low. In such a situation, the $1 - NN$ classifier is difficult to find out the right label for a probe face patch in a low-dimensional feature subspace. To overcome this issue, a $k - NN$ classifier (the value of k is larger than 1) in this study is used in each subspace and the majority voting strategy is employed to predict the identity of a probe face. Specifically, with the $k - NN$ classifier, each probe face patch will be assigned k labels corresponding to the first nearest k neighbors in the relevant subspace, and the k labels form a label set, denoted by L_s . As a probe image is divided into $2N$ local patches, there will be $2N$ label sets $L_{s\tau}$, $\tau = 1, 2, \dots, 2N$, for one probe image. We concatenate all the $2N$ label sets to construct a long set $L = \{L_{s1}, L_{s2}, \dots, L_{s2N}\}$ and employ the majority voting strategy to output the most frequent label in L as the identity of the probe face. The recognition procedure is intuitively shown in Fig. 2.

3.2.4 Boosting MFSA by ccLDA

As MFSA treats each feature subspace independently and the final classification is based on the results of all subspaces, MFSA can be enhanced by improving the discriminative ability of each feature subspace. In this work, we replace LDA with class-cluster LDA (ccLDA) [58] to boost MFSA. ccLDA is proposed to overcome overfitting effect for small

training sample size problem. It is motivated by the fact that a cluster consisting of similar samples of different classes contains some information of intra-class variations and the mean vectors of different clusters contribute to distinguishing different classes. Therefore, ccLDA regularizes the between-class and within-class scatter matrices by between-cluster and within-cluster scatter matrices, respectively.

For a training dataset X , ccLDA first groups the training data into K non-overlapping clusters, i.e.,

$$X = X_1 \cup X_2 \cup \dots \cup X_K, \quad (15)$$

$$X_i \cap X_j = \emptyset \text{ for } \forall i \neq j \quad (16)$$

where X_i is the i th cluster. Based on the above clustering results and following the same way to calculate within-class scatter matrix $S_w^{(\text{class})}$ and between-class scatter matrix $S_b^{(\text{class})}$ of LDA, the within-cluster matrix $S_w^{(\text{cluster})}$ and between-cluster scatter matrix $S_b^{(\text{cluster})}$ can be calculated. Then, the total within-class scatter matrix S_w^{ccLDA} and total between-class scatter matrix S_b^{ccLDA} of ccLDA can be computed by

$$S_w^{ccLDA} = \alpha S_w^{(\text{class})} + (1 - \alpha) \frac{1}{T} \sum_{r=1}^T S_{wr}^{(\text{cluster})}, \quad (17)$$

$$S_b^{ccLDA} = \beta S_b^{(\text{class})} + (1 - \beta) \frac{1}{T} \sum_{r=1}^T S_{br}^{(\text{cluster})}, \quad (18)$$

in which T is the number of clustering results by running K -means algorithm T times, $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ are the regularization parameters that balance the weight between class-based scatter matrices and cluster-based scatter matrices. The objective function of ccLDA is formulated by

$$R_{opt} = \arg \max \frac{\text{Tr}(R^T S_b^{ccLDA} R)}{\text{Tr}(R^T S_w^{ccLDA} R)}. \quad (19)$$

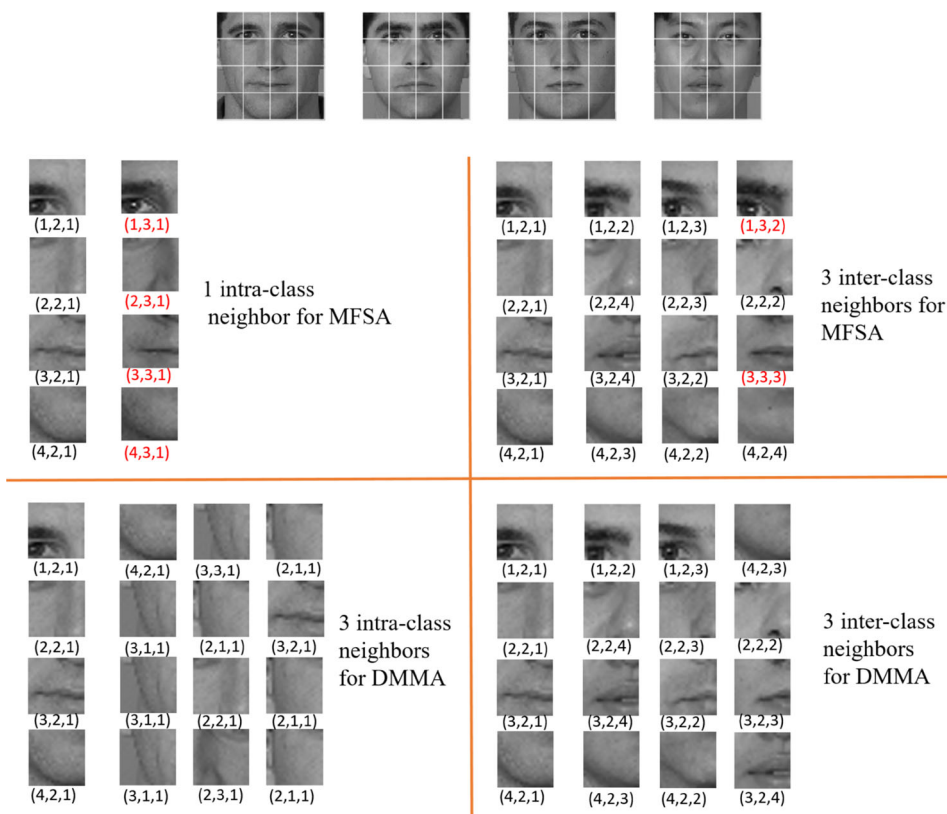
Similar to the solution of LDA, it can be solved by eigen-decomposition. The parameters involved in ccLDA, i.e., K , T , α , and β are usually determined empirically, and we set them to be 20, 40, 0.6, and 0.3, respectively, in our experiments. Note that there are only two samples per class in each group of face patches; thus, ccLDA is helpful to produce more discriminative feature subspaces, which further boost the recognition accuracy of MFSA.

3.3 Analysis of MFSA

The most similar work to our MFSA is discriminative multi-manifold analysis (DMMA) proposed in [23]. The two methods utilize local information of face patches to learn multiple feature spaces. Our MFSA is different from DMMA in at least three aspects.

- (1) MFSA employs the symmetry of faces to explore intra-class and inter-class variations according to face patch locations at a half face. We select four sample faces from FERET subset and partition them into 16 patches of the same size. Figure 3 displays the intra-class and inter-class neighbors of four face patches (in the second column) of the first subject; these intra-class and inter-class neighbors are used by two methods to explore the intra-class and inter-class variations, respectively. As can be seen from Fig. 3, all of the inter-class neighbors used by MFSA have the same semantics as the reference samples, and most of the inter-class neighbors used by DMMA have the same semantic as the reference samples, so there is not much difference between the inter-class neighbors used by these two methods (the right part of Fig. 3). However, there is a significant difference between the intra-class neighbors used by these two methods (the left part of Fig. 3). Specifically, because DMMA looks for intra-class neighbors of a reference patch from all other face patches of the same person and it does not utilize the symmetry of faces, the intra-class neighbors used by DMMA are quite different from the reference samples and involve more differences than similarities (lower left of Fig. 3); MFSA uses the symmetry of faces and mirrors one half face; the intra-class neighbor of a reference face patch is just the counterpart from the other half face (upper left of Fig. 3), so the reference patches and their corresponding neighbors have the same semantics and more similarities among them can be explored. Consequently, by using the symmetry of faces and learning multiple feature subspaces according to face patch locations, MFSA is able to estimate more accurate intra-class variations in each subspace and extract more discriminative features, though there is only one intra-class neighbor for each reference face patch. Please note that accurately estimating the underlying intra-class variations is the key to address SSPP face recognition. Therefore, MFSA has its superiority over DMMA in this aspect.
- (2) The way MFSA employed to recognize unlabeled faces is different from that of DMMA. Though both MFSA and DMMA aim to learn multiple feature subspaces, they are in different ways. MFSA learns a discriminative subspace for those face patches with the same semantics, while DMMA treats patches from one person as a manifold and learns a projection matrix for each subject. When recognizing an unlabeled face, MFSA independently predicts the labels of each face patch and then employs majority voting strategy to determinate the ultimate label of the probe face according to the united prediction labels of all patches. By using majority voting in our MFSA framework, those face patches which are difficult to be correctly classified are auto-

Fig. 3 Intra-class and inter-class neighbors used by MFSA and DMMA. The triple numbers below each face patch denote their original locations on the face, (i, j, k) denotes the face patch at the i th row and j th column of k th person. Patches with red numbers are mirrored across the symmetry axis (only for MFSA)



matically ignored. Thus, the effect of these challenging patches on the recognition can be reduced. Differently, DMMA employs the average construction error of all face patches to be the classification metric, and hence those challenging patches which score large construction errors will affect the ultimate score of a probe face. Therefore, MFSA concentrates those easily classified patches to output reliable results. Furthermore, owing to the fact that MFSA treats each feature subspace independently, techniques able to improve the discriminative ability of any feature subspaces of MFSA can be used to promote the performance of MFSA. For example, we adopt class-cluster LDA (ccLDA) [58] to boost the recognition accuracy of MFSA in the experiments.

- (3) MFSA obtains the multiple feature subspaces in a parallel way, while DMMA obtains its feature subspaces by an iterative algorithm, and hence MFSA is more readily and efficient to implement than DMMA.

4 Experimental analysis

We evaluate our proposed MFSA by conducting a number of experiments on three widely used databases, namely AR [59], FERET [60], and Labeled Faces in the Wild (LFW) [61] dataset. The following describes the details of the experiments and results.

4.1 Datasets

We conduct experiments mainly on single sample per person recognition problem on the subsets of AR face database and FERET face database. The AR face database contains over 4000 face images of 126 subjects (70 men and 56 women), including frontal view faces with various facial expression, lighting conditions, and occlusions (sun glass and scarves). There are 26 images per person, taken in two different sessions (two weeks apart). Similar to the works in [8,23,49,62,63], two subsets of AR database are used in our work. One subset consists of 800 gray-level images of 100 persons (50 men and 50 women, 8 images per person), and images in this subset are with various expressions (e.g., neural, smile, anger, and scream). The other subset is used as probe set only, which includes 1200 gray-level images of 100 persons (50 men and 50 women, 12 images per person), and face images in this subset are occluded by sunglasses and scarves. In our experiments, the images in both subsets were taken from two different sessions. Face images in this database are aligned, cropped, and resized to 80×80 pixels. Experiments on the first and second subset are to show the impact of expression and occlusion variations on the recognition performance, respectively. Figure 4 shows 20 sample images of one subject from the subsets of AR database, in which the first and second rows are from Session 1 and Session 2, respectively; the first 4 columns with



Fig. 4 Sample face images from AR dataset



Fig. 5 Sample face images from FERET database

various expressions construct the first subset, and the remaining columns with occlusions form the second subset.

The FERET database consists of 13,539 facial images corresponding to 1565 subjects, who are diverse across ethnicity, gender, and age. To investigate the performance of the proposed MFSA method on pose variations in our experiments, similar to [23,49], we apply a subset of the FERET database, which consists of 1,000 gray-level images from 200 subjects, each has 5 images labeled as *ba*, *bd*, *be*, *bf*, and *bg*, i.e., with pose angle of 0, +25, +15, -15, and -25, respectively. Face images in this database are aligned, cropped, and resized to 80×80 pixels. Figure 5 shows 10 sample images of two subjects from the subset of FERET database.

The LFW contains images of 5749 individuals taken under an unconstrained setting. The complex surroundings of image capturing and inaccurate alignment of faces make the LFW data extremely challenging for face recognition in the SSPP setting. LFW-a is a subset of the LFW dataset, and the images in LFW-a have been aligned with a commercial software tool. Following the work of [42,50], we gather the subjects containing no less than ten samples and then get a dataset with 158 subjects from LFW-a database and further choose the first 10 images of 158 individuals to construct the face subset for evaluation. All images in this subset are resized to 32×32 pixels. Figure 6 displays some sample images of 3 subjects in LFW subset. It can be observed that images in this dataset are more complicated as they were captured in the unconstrained environment.

The basic parameters of MFSA are set as follows in our experiments. The values of parameters k_1 , k_2 , and δ in the formulation of MFSA are empirically tuned to be 1, 50, and 100, respectively. In the whitening process, we let PCA preserve

99.85% energy, which keeps about 25–40% of the maximum dimensions. The size of the face patches is varied for different face subsets, and the impact of the patch sizes is discussed in the following. As we will see, the dimension of input face patches is quite small compared with that of the whole face, and it is further reduced by whitening processing, and hence the dimensions of final feature subspaces are smaller. We can alternatively find the optimal dimension for each feature subspace to extract features, but for the simplicity of the experiments, we make all the feature subspaces have the same dimension in our experiments. Let dim be the dimension of the whitening subspace for one group, in our experiments, we experimentally set $dim-2$ to be the dimension of the feature subspaces to extract features. Finally, we set $k = 1 : 10$ of k -NN classifier and run MFSA on various values of k to report the best recognition accuracy. The values of k is also discussed in the following subsection. Moreover, to promote the performance of MFSA, we also adopt class-cluster LDA (ccLDA) [58] to extract features in each feature subspace and use “MFSA+” to denote ccLDA-based MFSA framework. Unless otherwise specified, we use pixel values as the input features in the following experiments.

4.2 Results and analysis

4.2.1 Face recognition with various expressions

We first evaluate the effectiveness of our method in the situation where various expressions are presented in the probe images. Similar to [23,49], a subset of AR database including face images with various expressions is used. For each subject, the neutral expression face from Session 1 is selected into the gallery set and other faces with different expressions from both sessions are used as probe sets. In these experiments, the size of each face patch is set to be 2×10 pixels. We compare our proposed MFSA with two categories of methods: non-generic learning methods and generic learning methods.

Non-generic learning methods do not require a generic set, so all 100 subjects are used in testing phase. Table 1 shows the comparison of the state-of-the-art nongeneric learning methods, the bold numbers denote the highest recognition

Fig. 6 Sample images in LFW database



Table 1 Performance (%) comparison with non-generic learning methods on the AR database (expression)

Method	Session 1				Session 2				
	Smi.	Ang.	Scr.	Avg.	Neu.	Smi.	Ang.	Scr.	Avg.
(PC) ² A [19]	97	87	62	82.0	77	74	67	40	64.5
E(PC) ² A [34]	97	87	63	82.3	77	75	68	41	65.3
2DPCA [3]	97	87	60	81.3	76	76	67	37	64.0
(2D) ² PCA [35]	98	89	60	82.3	71	76	66	41	63.5
SOM [26]	98	88	64	83.3	73	77	70	42	65.5
LPP [9]	94	87	36	72.3	86	74	78	20	64.5
SVD-LDA [37]	73	75	29	59.0	75	56	58	19	52.0
Block PCA [25]	97	87	60	81.3	77	76	67	38	64.5
Block LDA [22]	85	79	29	64.3	73	59	59	18	52.3
UP [21]	98	88	59	81.7	77	74	66	41	64.5
DMMA [23]	99	93	69	87.0	88	85	79	45	74.3
M ³ L [49]	99	94	72	88.3	90	89	79	49	76.8
LS-CRC [52]	97	92	67	85.3	89	82	77	45	73.3
SDMME [51]	100	99	69	89.3	94	84	82	49	77.3
MFSA	99	98	68	88.3	91	83	81	44	74.8
MFSA+	100	100	74	91.3	93	85	86	52	79.0

We run LS_CRC and SDMME algorithms by ourselves and other recognition rates are cited from [23] Bold denotes the highest recognition rates among compared method

rates among all methods. We can obviously observe that DMMA, M³L, LS-CRC, SDMME, MFSA, and MFSA+ perform much better than other methods. This is because all of the six methods consider the local structure information by using face patches. Compared with those well-performed approaches, MFSA achieves competitive recognition accuracy on Session 1, while achieves 2.5% lower average accuracy than the best one (SDMME) on Session 2. After employing cLDA to promote MFSA, MFSA+ gains the highest recognition rates nearly in all cases, and the averaged recognition rates on Session 1 and Session 2 are 2 and 1.7% higher than the second best one.

Generic learning methods select the first 80 subjects to construct the gallery and query set, while the other 20 subjects are used as the generic training set. So generic learning methods are evaluated with 80 subjects' face images. Table 2 exhibits the comparison of the recognition results with state-of-the-art non-generic learning methods. Though MFSA and MFSA+ are evaluated with 100 subjects' images, they achieve higher recognition accuracy than AGL, ESRC, and SVDL. Compared with CSR-MN, which adopts mixed

Table 2 Performance (%) comparison with generic learning methods on AR database (expression)

Method	Session 1	Session 2	Avg.
AGL [40]	77.9	55.8	66.9
ESRC [43]	85.4	70.4	77.9
SVDL [24]	86.3	74.2	80.3
CSR-MN [47]	91.7	80.1	85.9
MFSA	88.3	74.8	81.6
MFSA+	91.3	79.0	85.2

We run CSR-MN algorithm by ourselves and other recognition rates are cited from [44]

Bold denotes the highest recognition rates among compared method

norms to learn variation dictionary from the generic set and achieves best performance in this experiment, our MFSA+ can gain competitive recognition rates. According to Tables 1 and 2, the comparisons with these state-of-the-art methods not only demonstrate that MFSA is effective for coping with expression variations, but also verify that cLDA is very helpful for boosting the performance of MFSA.

Table 3 Recognition accuracy (%) of different methods on the FERET database (pose)

Method	<i>bd</i>	<i>be</i>	<i>bf</i>	<i>bg</i>	Avg.
(PC) ² A	15.5	36.0	66.5	24.0	35.5
E(PC) ² A	16.5	37.0	67.5	24.5	36.4
2DPCA	17.5	37.5	68.5	25.5	37.3
(2D) ² PCA	18.0	38.0	69.0	26.5	37.9
SOM	18.5	38.5	69.5	27.0	38.4
LPP	19.5	39.5	70.0	28.0	39.3
SVD-LDA	17.5	38.5	68.5	27.0	37.9
Block PCA	19.5	40.5	71.5	28.5	40.0
Block LDA	21.0	42.0	73.0	30.0	41.5
UP	21.0	41.0	73.0	32.0	41.8
DMMA	25.0	45.0	75.0	35.5	45.1
LS-CRC	32.5	45.0	77.5	35.0	47.5
SDMME	39.0	54.5	82.5	40.5	54.1
MFSA	59.5	75.5	85.0	50.5	67.6
MFSA+	61.0	76.5	85.0	52.0	68.6

We run LS-CRC and SDMME by ourselves and other recognition rates are cited from [23]

Bold denotes the highest recognition rates among compared method

4.2.2 Face recognition with various poses

In the previous subsection, we have validated effectiveness of MFSA when different expressions are presented in the probe faces while all the faces in gallery and probe sets are frontal. However, in many real applications, a face recognition system has to recognize a target person whose face is not frontal. To investigate the performance of the proposed MFSA in such scenarios, a subset of FERET database including face images with different poses is utilized. For each subject, the frontal face labeled as *ba* is selected as gallery image, and other faces with various poses are used as probe images. In these experiments, the size of each face patch is set as 8×2 pixels. On this dataset, we only compare our approaches with non-generic learning methods. Table 3 tabulates the recognition results of different methods.

As shown in Table 3, compared with recognition results on frontal faces, the recognition rates for all the methods drop largely, which implies that it is more challenging to recognize the non-frontal faces. Despite this, our MFSA and MFSA+ consistently outperform the 14 compared methods by a large margin on this tough subset. Compared to the best performance of other methods, which is generated by SDMME, MFSA is superior in accuracy by 20.5, 21, 2.5, 10% on the *bd*-*bg* subsets, respectively, and a gain of 13.5% on the average recognition rate. And as expected, MFSA+ further increases the recognition accuracy. Additionally, all techniques perform much better on *be* and *bf* than on *bd* and

bg; it is due to the fact that *be* and *bf* are with smaller pose angle (15°) than *bd* and *bg* (25°).

Though the proposed MFSA is designed by using the symmetry of frontal faces, it achieves the best performance on this subset in SSPP scenario, which demonstrates the effectiveness of the novel approach when dealing with pose variations. We conclude two reasons for it. First, MFSA is a patch-based method, and it runs on the partitioned face patches rather than the global images, which ease the significant variations caused by poses to some extent. This is the advantage over the holistic appearance-based methods. Second, despite a non-frontal face is not symmetric, the difference between the face patches at the same location of a half face is still smaller than that between patches at different locations. Thus, by clustering the face patches into various groups and extracting features from each group independently, MFSA eliminates the confusions between inter-class and intra-class variance when generating the projection matrices. This is the superiority of MFSA over other patch-based methods. As a result, the proposed MFSA is much more robust to the various poses presented in the global face images.

4.2.3 Face recognition with occlusions

Next, we investigate the robustness of the proposed approach to occlusions. In the following experiments, we choose a subset of AR database which consists of partially occluded faces. We use the same gallery set as in Sect. 4.2.1, the faces with sunglasses and scarves from both sessions as the probe sets. The size of each face patch is set as 4×4 pixels.

Different from the previous experiments, in this subsection, we employ *difference face patches* [54] instead of original face patches to extract features. After partitioning a half face and acquiring the patch sequence, a difference patch $\Delta f_j(x, y)$ is computed by subtracting $f_j(x, y)$ from its immediate neighboring patch $f_{j+1}(x, y)$ as:

$$\Delta f_j(x, y) = f_{j+1}(x, y) - f_j(x, y),$$

where $f_j(x, y)$ is the intensity of the pixel at coordinates (x, y) of the j th patch, $j \in \{1, 2, \dots, N\}$. Note that here the length of the difference patch sequence of a half face is $N - 1$. Figure 7 shows an intuitive illustration of the generation of difference patches. After obtaining the difference face patches of all half faces, we cluster them into $N - 1$ groups as what we have done before on the original face patches. As mentioned in [54], a difference patch can be viewed as the approximation of the first-order derivative of adjacent patches, which is helpful for enhancing the salient facial features representing textured regions such as eyes, nose, and mouth. As an occlusion part in the face is usually smooth, difference patches are not sensitive to it. This merit of difference patches will benefit feature extraction in our work.

Fig. 7 Generation of difference face patches

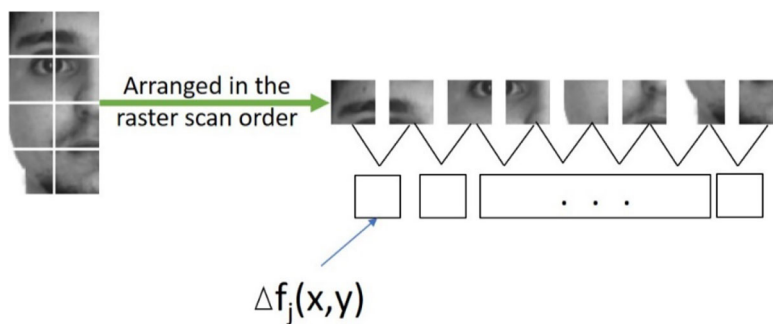


Table 4 Performance (%) comparison with non-generic learning methods on the AR database (occlusions)

Method	Session 1			Session 2		
	Sung.	Scarf	Avg.	Sung.	Scarf	Avg.
SRC [8]	40.8	25.3	33.1	22.9	13.3	18.1
RSC [13]	71.7	59.6	65.6	52.9	35.8	44.4
PCRC	88.8	90.4	89.6	69.6	75.4	72.5
LS-CRC	85.3	88.7	87.0	65.6	70.0	67.8
SDMME	87.7	89.3	88.5	70.0	72.4	71.2
MFSA	91.7	89.3	90.5	74.0	74.0	74.0
MFSA+	93.0	90.0	91.5	75.3	74.6	75.0

We run LS-CRC and SDMME by ourselves and other recognition rates are cited from [46]

Bold denotes the highest recognition rates among compared method

Table 5 Performance (%) comparison with generic learning methods on the AR database (occlusions)

Method	Session 1			Session 2		
	Sung.	Scarf	Avg.	Sung.	Scarf	Avg.
ESRC	86.2	83.8	85.0	65.4	57.5	61.5
SVDL	78.8	67.1	72.9	55.0	41.3	48.1
OPDL [45]	90.0	84.2	87.1	76.7	59.2	67.9
DMSC [46]	95.8	90.0	92.9	78.3	77.9	78.1
CSR-MN	93.8	91.5	92.7	77.6	75.9	76.8
MFSA	91.7	89.3	90.5	74.0	74.0	74.0
MFSA+	93.0	90.0	91.5	75.3	74.7	75.0

We run CSR-MN by ourselves and other recognition rates are cited from [46]

Bold denotes the highest recognition rates among compared method

Except for the difference patches, all other basic settings are the same as the previous experiments. Similar to Sect. 4.2.1, we compare our proposed MFSA with non-generic learning methods and generic learning methods. Table 4 displays the recognition results of various non-generic learning methods, while Table 5 shows the recognition accuracy comparison between MFSA and generic learning methods.

As observed from Table 4, the patch-based methods (the last five ones in Table 4) apparently gain higher recognition accuracy than global appearance-based methods (i.e., SRC and RSC), which implies that dividing faces into local patches is effective in handling the occlusions imposed on faces. Compared with all other non-generic learning methods, both MFSA and MFSA+ gain the best recognition performance in almost all cases, but only lag behind PCRC at 1.1 and 0.4% in Scarf case of Session 1. Specifically, the recognition accuracy of MFSA+ reaches more than 90% in each case of Session 1 and more than 74.5% in each case of Session 2; meanwhile, the average recognition rates in two sessions are the highest. Table 5 tabulates recognition results of our method and some generic learning methods. It can be seen from this table that the first four best methods are DMSC, CSR-MN, MFSA+, and MFSA. Neither MFSA nor MFSA+ achieves the highest recognition rates compared with these generic learning methods, but MFSA+ is quite competitive with the second best one CSR-MN. Note that our MFSA and MFSA are evaluated on all face images, while DMSC and CSR-MN use the 80 subjects for testing and remaining 20 subjects as external data for learning variations. We then try to evaluate MFSA+ by using only 80 subjects’ images just as DMSC and CSR-MN do. The recognition rates obtained by MFSA+ on the four cases are 95.8, 91.7, 79.6, and 77.7%, respectively. It is quite competitive with DMSC, while MFSA+ does not depend on any generic training data.

According to the above comparisons, we can conclude that our proposed method outperforms all compared non-generic learning methods and achieves either better or competitive recognition performance with most of generic learning methods. This demonstrates the robustness of MFSA to occlusion variations in single sample per person face recognition.

4.2.4 Face recognition on unconstrained faces

Previous experiments have shown the robustness of the proposed MFSA on expression, pose, and occlusion variations under constrained circumstances. Then, we further evaluate it on faces taken in unconstrained environment. Following

the work of [42,50], for those methods using generic set, 78 subjects are used as the generic dataset to learn the intra-class variance dictionaries and 80 subjects are used for evaluation. We then randomly choose one image as the gallery image for each subject and use nine images as the probe images for evaluation. The size of face patch is 4×4 pixels, and we employ *difference face patches* to extract features on this extensive database, as we have done in Sect. 4.2.3. As AGL, ESRC, and CSR-MN are generic set-based methods, we also evaluate the proposed MFSA using only 80 subjects' images. The performance of different methods based on 10 independent experiments on this dataset is listed in Table 6; we also conduct some extra experiments and list these recognition results in the same table.

Because faces in LFW dataset involve complicated variations in illumination, pose, expression, and occlusion, SSPP face recognition on this dataset is extremely challenging and exhibits poor performance. The results show that our approach achieves the third best performance only to ESRC and CSR-MN, that is to say, MFSA outperforms all compared non-generic learning methods and one generic learning method AGL. Compared with ESRC and CSR-MN, MFSA falls behind them obviously, and the differences are 6.33 and 8.88%, respectively. We want to declare that our approach is able to work without any generic datasets; however, ESRC and CSR-MN have to utilize knowledge learned from a generic set which has the same sample distribution as the probe set; more importantly, such a generic set is not always available in real-world applications.

Note that the gallery images used for evaluation are randomly selected, so they are usually non-frontal faces. We propose MFSA by taking advantage of facial symmetry, but it still works well when the face images in the gallery set are not frontal, which indicates that MFSA is practicable for both situations. Further, we choose one most frontal face image for each subject to construct a gallery set, and the rest 9 images are used for testing. In this situation, the recognition accuracy of MFSA is 18.57 and 24.58% when it is tested on 158 subjects' images and 80 subjects' images, which is about 4 and 6% higher than prior experimental results. MFSA+ still can obtain better performance than MFSA, though with small margin. Therefore, a good gallery image set is very helpful for boosting the performance of the proposed approach.

4.3 Discussion of parameters

To give more insight into the proposed approach, we here investigate the influence of two parameters on the recognition performance. For the sake of simplicity, we conduct the following experimental analysis only on AR and FERET subsets, but it is pertinent to explain the effect of parameters.

Table 6 Performance (%) comparison among different methods on the LFW dataset

Method	(PC) ² A	E(PC) ² A	Block LDA	SVD-LDA	Fisher-faces	CRC	SRC	AGL
Accuracy \pm std.	8.36 \pm 0.81	8.40 \pm 1.41	4.47 \pm 0.87	6.99 \pm 1.15	12.69 \pm 1.34	14.57 \pm 1.78	7.40 \pm 1.35	13.33 \pm 2.26
Method	ESRC	CSR-MN	MFSA (158)	MFSA (80)	MFSA (158_fr)	MFSA (80_fr)	MFSA+ (158_fr)	MFSA+ (80_fr)
Accuracy \pm std.	24.89 \pm 1.87	27.44 \pm 1.06	14.88 \pm 0.90	18.56 \pm 0.16	18.57	24.58	19.36	26.23

MFSA (158) and MFSA (80), respectively, denote that the proposed method is evaluated on 158 subjects' images and 80 subjects' images; fr denotes when a frontal face of each person is used for training

Bold denotes the highest recognition rates among compared method

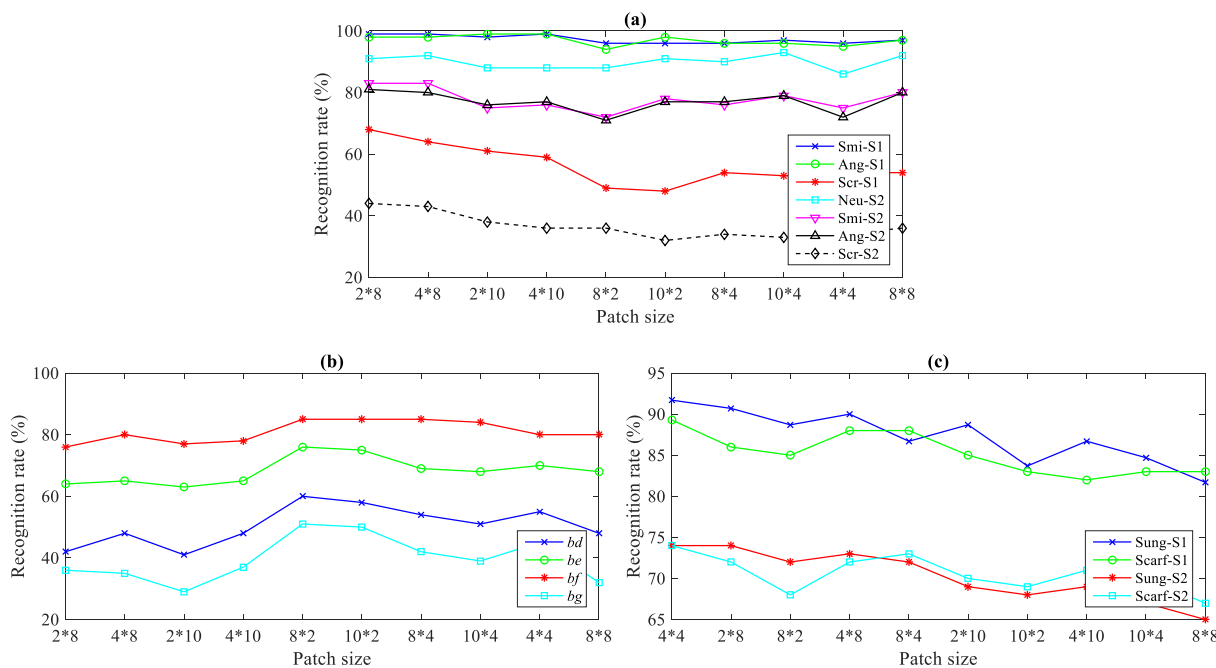


Fig. 8 Recognition rates (%) of the proposed method with respect to the patch size on **a** AR database (expression), **b** FERET database (pose) and **c** AR database (occlusion)

4.3.1 The effect of patch size

We let the values of parameters k_1 , k_2 , and δ still keep 1, 50, and 100, respectively, and then run MFSA on different values of k (from 1 to 10) to report the highest accuracy. The MFSA is tested on three face subsets with ten different patch sizes. The best recognition rates with respect to the patch size are shown in Fig. 8. We have the following observations from it:

- (1) Figure 8a shows the experimental results in the first AR subset. As can be seen, there is no sharp fluctuation in the rate curve of Smi-S1, Ang-S1, and Neu-S2, as these three probe sets show limited variations compared with the gallery set (Neu-S1). Nevertheless, obvious fluctuations are found on the other rate curves, especially in that of Scr-S1 and Scr-S2, which reveals that when the large expression variations are presented in the probe faces, the patch sizes do affect the recognition accuracy of the proposed MFSA. Furthermore, it is interesting to find that compared with the patch sizes of long length and short width, those patch sizes of narrow length and wide width perform better. The reason for this finding is that the variations of expressions mainly change the vertical structure of a face, so that more partitions are needed in this direction to extract more information. In this face subset, the patch size of 2×10 performs better than all other patch sizes.
- (2) Figure 8b exhibits the experimental results in FERET subset. The waved rate curves indicate that different

patch sizes affect the accuracy of MFSA in recognizing the non-frontal faces. The interesting finding is that patch sizes of long length and short width achieve higher rates than those of narrow length and wide width, which is just contrast with experiments on expressions. The reason is that the variations of poses mainly bring changes to the horizontal structure of a face; thus, more segmentations in horizontal direction are helpful to explore more information. Moreover, it is evident in the figure that the patch size of 8×2 achieves the highest recognition rates in all tests.

- (3) Figure 8c displays the experimental results in the second subset of AR, where the faces are occluded by sunglasses and scarves. We can also observe that the patch sizes have an impact on the recognition performance. Because the sizes and locations of the occlusions on the face are uncertain, according to this figure, smaller patch sizes are more likely to achieve higher recognition accuracy. We can see that the patch size of 4×4 displays the best performance in all cases in this subset.

4.3.2 The effect of k

In the testing phase, we use k -NN classifier and majority voting strategy to identify an unlabeled face. To examine the effect of the values of k on the recognition accuracy, some more experiments are conducted on the three subsets. We let the values of parameters k_1 , k_2 , and δ keep the same values as before and select the best patch size to do experiments.

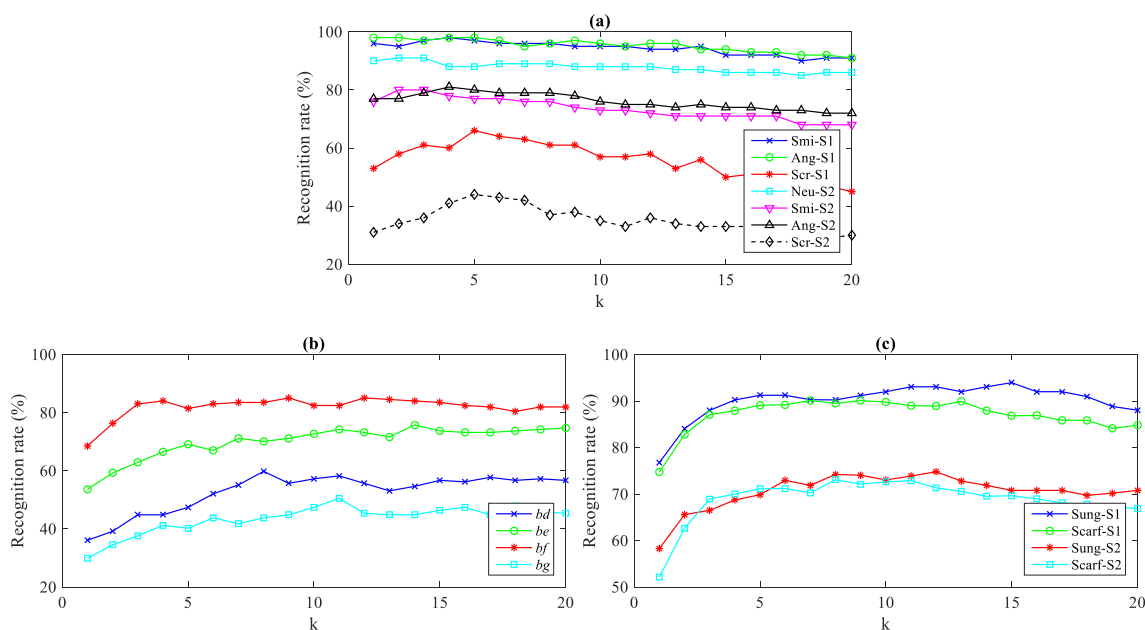


Fig. 9 Recognition rates (%) of the proposed method with respect to the parameter k on **a** AR database (expression), **b** FERET database (pose) and **c** AR database (occlusion)

We run MFSA on the value of k from 1 to 20. Figure 9 plots the recognition rates with respect to the parameter k .

On the whole, the recognition accuracy rises first and then drops along with the increasing value of k . It is validated in the figure that 1- NN classifier fails to help MFSA achieve promising results, as the recognition rates at $k = 1$ are worse than those at other values in all of the experiments; meanwhile, a larger value of k does not ensure higher recognition rate. In all the three subsets, the satisfying recognition rates can be achieved when the value of k is less than 10, which is a quite small number when comparing it to the number of face patches in each group. Therefore, a modestly small value of k is able to output good performance in the framework of MFSA, which enhances the efficiency of our approach.

5 Conclusions

We have proposed a novel multiple feature subspaces analysis to address the SSPP problem in face recognition, which is inspired by the symmetry of frontal faces. We partition each enrolled face image into non-overlapping face patches and cluster all patches into groups according to their locations at the half face, and then learn multiple discriminative feature subspaces for these groups to extract features. The new approach harvests the advantages of patch-based methods, while overcomes their disadvantages. Promising results on two widely used face databases demonstrate the robustness of the proposed method to variations in expression, pose, and occlusion. Furthermore, we evaluate the proposed

approach on an unconstrained face database, and the experimental results show its effectiveness in recognizing faces captured in the complicated environment.

The performance of MFSA framework can be further enhanced by considering the following factors. First, MFSA utilizes the symmetry of faces, but the symmetry axis is not fixed for various faces, especially for the non-frontal faces. Thus, finding the correct symmetry axis of faces automatically will be beneficial for MFSA. Second, different face patches may have various discriminative abilities. So it will be a good alternative to treat face patches differently and endow them with various weights. Third, in the procedure of computing the multiple discriminative subspaces, we have assumed face patch groups independent. As a result, the global information, i.e., the relationship among those face patches from the same face, is ignored. To integrate the holistic information of face images in the framework will also do good to the performance of MFSA. In the future, we will attempt to improve MFSA from these interesting research directions.

References

1. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings CVPR'91., IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE (1991)
2. Heisele, B., Ho, P., Poggio, T.: Face recognition with support vector machines: global versus component-based approach. In Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. Proceedings. IEEE (2001)

3. Yang, J., et al.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(1), 131–137 (2004)
4. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
5. Liu, C., Wechsler, H.: A Gabor feature classifier for face recognition. In: Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. Proceedings. IEEE (2001)
6. Wagner, A., et al.: Toward a practical face recognition system: robust alignment and illumination by sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(2), 372–386 (2012)
7. De Marsico, M., et al.: Robust face recognition for uncontrolled pose and illumination changes. *IEEE Trans Syst Man Cybern Syst* **43**(1), 149–163 (2013)
8. Wright, J., et al.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
9. He, X., et al.: Face recognition using Laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005)
10. Cheng, Y., et al.: Illumination-insensitive features for face recognition. *Vis. Comput.* **33**(11), 1483–1493 (2017)
11. Kakadiaris, I.A., et al.: Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 640–649 (2007)
12. Cox, D., Pinto, N.: Beyond simple features: a large-scale feature search approach to unconstrained face recognition. In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011). IEEE (2011)
13. Yang, M., et al.: Robust sparse coding for face recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2011)
14. Zhou, Z., et al.: Face recognition with contiguous occlusion using Markov random fields. In: ICCV. (2009)
15. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In: *Computer Vision—ECCV 2010*. pp. 448–461. Springer (2010)
16. Zhao, L., et al.: Fisher discrimination-based $l_{2,1}$ -norm sparse representation for face recognition. *Vis. Comput.* **32**(9), 1–14 (2016)
17. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
18. Yan, S., et al.: A parameter-free framework for general supervised subspace learning. *IEEE Trans. Inf. Forensics Secur.* **2**(1), 69–76 (2007)
19. Wu, J., Zhou, Z.-H.: Face recognition with one training image per person. *Pattern Recognit. Lett.* **23**(14), 1711–1719 (2002)
20. Chen, W., Gao, Y.: Recognizing partially occluded faces from a single sample per class using string-based matching. In: *Computer Vision—ECCV 2010*. pp. 496–509. Springer (2010)
21. Deng, W., et al.: Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach. *Pattern Recognit.* **43**(5), 1748–1762 (2010)
22. Chen, S., Liu, J., Zhou, Z.-H.: Making FLDA applicable to face recognition with one sample per person. *Pattern Recognit.* **37**(7), 1553–1555 (2004)
23. Lu, J., Tan, Y.-P., Wang, G.: Discriminative multimifold analysis for face recognition from a single training sample per person. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 39–51 (2013)
24. Yang, M., Van Gool, L., Zhang, L.: Sparse variation dictionary learning for face recognition with a single training sample per person. In: Proceedings of the IEEE International Conference on Computer Vision. (2013)
25. Gottumukkal, R., Asari, V.K.: An improved face recognition technique based on modular PCA approach. *Pattern Recognit. Lett.* **25**(4), 429–436 (2004)
26. Tan, X., et al.: Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. *IEEE Trans. Neural Netw.* **16**(4), 875–886 (2005)
27. Saha, S., Bandyopadhyay, S.: A symmetry based face detection technique. In: Proceedings of the IEEE WIE National Symposium on Emerging Technologies. (2007)
28. Pathangay, V., Das, S., Greiner, T.: Symmetry-based face pose estimation from a single uncalibrated view. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG'08. IEEE (2008)
29. González-Jiménez, D., Alba-Castro, J.L.: Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry. *IEEE Trans. Inf. Forensics Secur.* **2**(3), 413–429 (2007)
30. Passalis, G., et al.: Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(10), 1938–1951 (2011)
31. Harguess, J., Aggarwal, J.: A case for the average-half-face in 2D and 3D for face recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE (2009)
32. Harguess, J., Aggarwal, J.: Is there a connection between face symmetry and face recognition? In: 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE (2011)
33. Li, S., et al.: The casia nir-vis 2.0 face database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2013)
34. Chen, S., Zhang, D., Zhou, Z.-H.: Enhanced (PC) 2 A for face recognition with one training image per person. *Pattern Recognit. Lett.* **25**(10), 1173–1181 (2004)
35. Zhang, D., Zhou, Z.-H.: (2D) 2PCA: two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing* **69**(1), 224–231 (2005)
36. Biao, W., Wenming, Y., Weifeng, L.: Two-stage block-based whitened principal component analysis with application to single sample face recognition. *IEICE Trans. Inf. Syst.* **95**(3), 853–860 (2012)
37. Zhang, D., Chen, S., Zhou, Z.-H.: A new face recognition method based on SVD perturbation for single example image per person. *Appl. Math. Comput.* **163**(2), 895–907 (2005)
38. Shan, S., et al.: Extended Fisherface for face recognition from a single example image per person. In: IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002. IEEE (2002)
39. Gao, Q.-X., Zhang, L., Zhang, D.: Face recognition using FLDA with single training image per person. *Appl. Math. Comput.* **205**(2), 726–734 (2008)
40. Su, Y., et al.: Adaptive generic learning for face recognition from a single sample per person. In: *Computer Vision and Pattern Recognition*. (2010)
41. Si, S., Tao, D., Geng, B.: Bregman divergence-based regularization for transfer subspace learning. *IEEE Trans. Knowl. Data Eng.* **22**(7), 929–942 (2010)
42. Gao, S., et al.: Neither global nor local: regularized patch-based representation for single sample per person face recognition. *Int. J. Comput. Vis.* **111**(3), 365–383 (2015)
43. Deng, W., Hu, J., Guo, J.: Extended SRC: undersampled face recognition via intraclass variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(9), 1864–1870 (2012)
44. Zhu, P., et al.: Local generic representation for face recognition with single sample per person. In: *Asian Conference on Computer Vision*. Springer (2014)
45. Wei, C.P., Wang, Y.C.F.: Undersampled face recognition with one-pass dictionary learning. In: IEEE International Conference on Multimedia and Expo. (2015)

46. Yu, Y.F., et al.: Discriminative multi-scale sparse coding for single-sample face recognition with occlusion. *Pattern Recognit.* **66**, 302–312 (2017)
47. Li, Z.M., Huang, Z.H., Shang, K.: A customized sparse representation model with mixed norm for undersampled face recognition. *IEEE Trans. Inf. Forensics Secur.* **11**(10), 1–1 (2016)
48. Wei, X., Li, C.-T.: Fixation and saccade based face recognition from single image per person with various occlusions and expressions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* (2013)
49. Yan, H., et al.: Multi-feature multi-manifold learning for single-sample face recognition. *Neurocomputing* **143**, 134–143 (2014)
50. Zhu, P., et al.: Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In: *European Conference on Computer Vision.* (2012)
51. Zhang, P., et al.: Sparse discriminative multi-manifold embedding for one-sample face identification. *Pattern Recognit.* **52**(C): 249–259 (2016)
52. Liu, F., et al.: Local structure based multi-phase collaborative representation for face recognition with single sample per person. *Inf. Sci.* **s 346–347**: 198–215 (2016)
53. Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: which helps face recognition? In: *International Conference on Computer Vision.* (2011)
54. Wei, X., et al.: Dynamic image-to-class warping for occluded face recognition. *IEEE Trans. Inf. Forensics Secur.* **9**(12), 2035–2050 (2014)
55. Kan, M., et al.: Side-Information based linear discriminant analysis for face recognition. In: *BMVC.* (2011)
56. Yin, F., et al.: Double linear regressions for single labeled image per person face recognition. *Pattern Recognit.* **47**(4), 1547–1558 (2014)
57. Deng, W., Hu, J., Guo, J.: Gabor-Eigen-Whiten-Cosine: a robust scheme for face recognition. In: *Analysis and Modeling of Faces and Gestures.* (2005)
58. Pang, Y., Wang, S., Yuan, Y.: Learning regularized LDA by clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **25**(12), 2191–2201 (2014)
59. Martinez, A.M.: The AR face database. In: *CVC Technical Report.* **24**, (1998)
60. Phillips, P.J., et al.: The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
61. Huang, G.B., et al.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition.* (2008)
62. Weng, R., et al.: Robust Feature Set Matching for Partial Face Recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 601–608 (2013)
63. Wei, X., Li, C.-T., Hu, Y.: Face recognition with occlusion using dynamic image-to-class warping (DICW). In: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG).* IEEE (2013)



Yongjie Chu received the Bachelor degree in business management and Master degree in operations research from Qufu Normal University, Rizhao, China, in 2006 and 2010, respectively. He is currently pursuing the Ph.D. degree in systems engineering at Southeast University, China. His research interest includes machine learning, pattern recognition, and face recognition.



Lindu Zhao received his Bachelor degree in industrial psychology from Hangzhou University, China, in 1988, Master degree in computer application technology from Jiangsu University, China, in 1993, and Ph.D. degree in systems engineering from Southeast University, in 1997. From 1997–1999, he was a post-doctor in electronics and communications at Southeast University. Since 1999, he has been Assistant professor with Institute of Systems Engineering, Southeast University. His research interests include analysis and decision of complex systems, information fusion, machine learning and emergence management, etc.



Touqeer Ahmad received his B.S. (computer engineering), M.Sc. (computer science) and M.S. (computer science and engineering) degrees, respectively, from NUCES-FAST, Lahore, Pakistan, University of York, UK and University of Nevada, Reno, USA in 2008, 2011 and 2014. Currently, he is a Ph.D. candidate and graduate research assistant in the Department of Computer Science and Engineering at University of Nevada, Reno. His research interests are applied machine learning, computer vision and deep learning.