Machine Vision and Applications (2013) 24:537–549
DOI 10.1007/s00138-012-0423-7

ORIGINAL PAPER

Fully automatic face recognition framework based on local and global features

Cong Geng · Xudong Jiang

Received: 30 May 2011 / Revised: 21 February 2012 / Accepted: 29 February 2012 / Published online: 22 March 2012

Abstract Face recognition algorithms can be divided into two categories: holistic and local feature-based approaches. Holistic methods are very popular in recent years due to their good performance and high efficiency. However, they depend on careful positioning of the face images into the same canonical pose, which is not an easy task. On the contrary, some local feature-based approaches can achieve good recognition performances without additional alignment. But their computational burden is much heavier than holistic approaches. To solve these problems in holistic and local feature-based approaches, we propose a fully automatic face recognition framework based on both the local and global features. In this work, we propose to align the input face images using multi-scale local features for the holistic approach, which serves as a filter to narrow down the database for further fine matching. The computationally heavy local feature-based approach is then applied on the narrowed database. This fully automatic framework not only speeds up the local feature-based approach, but also improves the recognition accuracy comparing with the holistic and local approaches as shown in the experiments.

Keywords Face alignment · Face recognition · Holistic approach · Multi-scale local feature

C. Geng (✉) · X. Jiang
School of Electrical and Electronic Engineering,
Nanyang Technological University,
Nanyang Link, Singapore 639798, Singapore
e-mail: geng0007@ntu.edu.sg

X. Jiang
e-mail: exdjiang@ntu.edu.sg

1 Introduction

In general, face recognition algorithms are divided into two categories based on the property of the features: holistic approaches and local feature-based approaches. Since the principal component analysis (PCA) [26] and the linear discriminant analysis (LDA) [2] were introduced into face recognition, various holistic approaches have been extensively studied [23], such as variants of LDA [6,22,47,50–52], marginal fisher analysis (MFA) [49], and eigenfeature regularization and extraction (ERE) [24], due to their good performance and low computational complexity. However, the holistic approaches require a preprocessing procedure to normalize the face image variations in pose and scale. This is not an easy task because it depends on the accurate detection of at least two landmarks from the face image [43]. Some algorithms for eye localization have been proposed based on the eyeball [7,11,18,25,30,32,39–42,48]. However, in many real applications the appearances of eyeball are not distinct or missing due to expressions, occlusions, illuminations or image noise. Hence, some algorithms localize multiple facial features like corners of eyes, nostrils, the tip of nose, corners of mouth, etc. Face alignment is performed based on these semantic features [3,9,12,31,53,54]. The same problem encountered in the detection of eyes remains. Moreover, in the training process, these semantic features are hand-annotated, which is very labor-consuming. In [21], an unsupervised approach is proposed for face alignment, which is not based on the localizations of semantic facial features (SF). As the performance of the face alignment algorithm influences the final recognition performance, many research papers on the holistic approaches report the recognition performance on the pre-normalized faces. The recognition performance will deteriorate considerably if the manual process is replaced by an automatic alignment algorithm.

In contrast to holistic methods, some local feature-based approaches [17,34,35] for face recognition are more robust to image variations in pose and scale. Furthermore, unlike the holistic approaches, the face normalization is an integrated part of the local approaches. Recently, some initial attempts apply the scale invariant feature transform (SIFT) [34,35] in the face recognition tasks [4,10,14–16,27,36,44]. Experimental results in [17] show that the performance of the local feature-based approach is significantly better than the popular holistic approaches. However, its computational time is much longer. For instance, to fulfill the face recognition task, one must search all the images in the database and compare each local feature in every image, which causes very heavy computational burden.

Although the computational complexity of holistic approaches is much lower than local feature-based methods, they need an accurate face alignment. On the contrary, some local feature-based approaches [17], which achieve better recognition performance than holistic approaches, are performed on unaligned face images. However, their computational burden is much heavier. To solve these two critical problems in holistic and local feature-based approaches, we propose a fully automatic face recognition framework (FAFF) based on both the local and global features. To speed up the local feature-based approach, we propose to use the holistic approach as a filter to retrieve some candidate face images from the whole gallery set. The selected face images have higher probabilities matching to the probe, than the remains in the gallery. They form a new gallery set with reduced size, on which we perform the local feature-based approach for face recognition. The reduction in the size of the gallery set will speed up the recognition process of local feature-based approaches. To solve the alignment problem in holistic approaches, we design a face alignment scheme based on multi-scale local features instead of relying on the semantic facial parts. In general, face alignment is performed based on the localizations of eyes, corners of mouth, nostrils and so on. However, in many real applications the appearances of facial parts are not distinct or missing due to expressions, occlusions, illuminations or image noise, which makes the alignment results unreliable. In face images, non-semantic facial features also hold distinct information, which can be utilized in the alignment process. Hence, we propose to align face images based on non-semantic multi-scale local features. The performance of our face alignment strategy is validated by face recognition tasks using local binary patterns (LBP) [1] and holistic approaches: LDA [2], UFS [47] and ERE [24]. Experimental results show that our alignment approach outperforms those based on localization of eyes [18,25,40–42], the localization of facial parts [12] and the congealing approach [21]. In our FAFF, we adopt the holistic approach ERE [24] to narrow down the database as it is one of the best performed holistic approaches [8]. For the local feature-based approach, we use the multi-scale local feature extraction and matching framework (LFEM) in [17], as it achieves better performance than many other face recognition approaches. We firstly align face images based on the multi-scale local features. Then ERE is performed on those well-aligned images to retrieve candidate faces from the whole gallery set. Finally, we perform LFEM on the narrowed gallery set. Our FAFF not only speeds up the local feature-based approach, but also improves the recognition performance. Our main contributions include:

1. We propose a face alignment approach based on multi-scale local features. Given an unaligned face image resulting from a face detector and a set of aligned face images in the data set, we build an automatic transformation mechanism, under which the unaligned face image can be precisely aligned for the recognition process.
2. A FAFF integrates the local feature-based approach LFEM and the holistic approach ERE in a cascaded way. We firstly use ERE as a filter to retrieve some candidate images which form a gallery set with reduced size, where we perform LFEM for face recognition. This framework not only speeds up the LFEM approach, but also achieves better recognition performance than LFEM, ERE and their parallel combination.

The rest of this paper is organized as follows. In Sect. 2, we give the outline of our FAFF, and briefly explain the rationale behind it. In Sect. 3, we describe the multi-scale LFEM [17]. In Sect. 4, face alignment based on multi-scale local features is introduced in detail. In Sect. 5, we fulfill the whole automatic face recognition framework by integrating ERE and LFEM in a cascaded way. In Sect. 6, experiments are conducted on AR [37], Georgia Tech (GT) [38] and ORL [46] databases and show the performances of our methods. This paper is finally concluded by discussion in Sect. 7.

2 Framework outline

Although some local feature-based approaches achieve better recognition performances than holistic approaches [17], their computational burden is much heavier. To speed up the recognition process of local feature-based approaches, in our FAFF, we firstly perform holistic approach to retrieve candidate images from the whole gallery set. The selected images have higher probabilities matching to the probe than the remains. Then we perform local feature-based approach for face recognition on the narrowed gallery set, composed of the retrieved candidate images. The reduction in the size of the gallery set will speed up the recognition process of

local feature-based approaches. Before we use the holistic approach as a filter, we must align face images into the same canonical pose, which is the premise of most holistic approaches. Usually, the alignment is done based on the localization of semantic facial parts like corners of eyes, eyeballs, corners of mouth, nostrils, etc. However, it is difficult to precisely detect these structures in many real applications. Besides semantic features, non-semantic facial features also hold distinct information. Hence, we propose to perform face alignment based on the non-semantic multi-scale local features.

Figure 1 visually gives the outline of our FAFF based on local and global features. It consists of offline training and online recognition. In the offline training process, we firstly extract local and global features from the face database. The images in the database are aligned manually, where we learn a common face template using multi-scale local features. In the online recognition process, we build an automatic transformation mechanism, under which the probe image is aligned with the common face template learned during the offline training. Then we perform the holistic approach on the well-aligned probe image and the face database to reduce the size of the gallery set. At last, the multi-scale local feature-based approach is applied on the narrowed gallery set and outputs the final image ID. In the following sections, we will introduce each ingredient of our FAFF in detail.

3 Local feature extraction and matching

3.1 Local feature extraction

3.1.1 Keypoint detection and scale selection

A keypoint is a pixel or the center of a local area that shows some characteristics different from its neighbors. Obviously, blob-like and corner-like structures are candidates of keypoints. A Laplace operator ∇^2 applied to the image $I(x, y)$ produces extrema at both blob-like and corner-like structures. Therefore, the spatial extrema of the Laplace image $\nabla^2 I(x, y)$ are keypoint candidates. However, local structures have different scales so that the extrema may appear in the images smoothed with different scales. To find the scale of a possible keypoint we need to detect the extrema in the scale space, too. As studied by Lindeberg [33], the normalized Laplacian of Gaussian (LoG), $\sigma^2\nabla^2 G$, provides scale invariance. Thus, automatic scale selection is done in the output images of the normalized LoG filter $O(x, y, \sigma^2) = \sigma^2\nabla^2 G(x, y, \sigma) * I(x, y)$, where $G(x, y, \sigma)$ is a Gaussian smooth filter with zero mean and diagonal covariance matrix $\sigma^2\mathbf{I}$.

To detect the blob-like and corner-like structures and represent them at the optimal scales, points of the normalized LoG images that are extrema in both spatial and scale spaces are selected. Unlike most other visual objects, which have

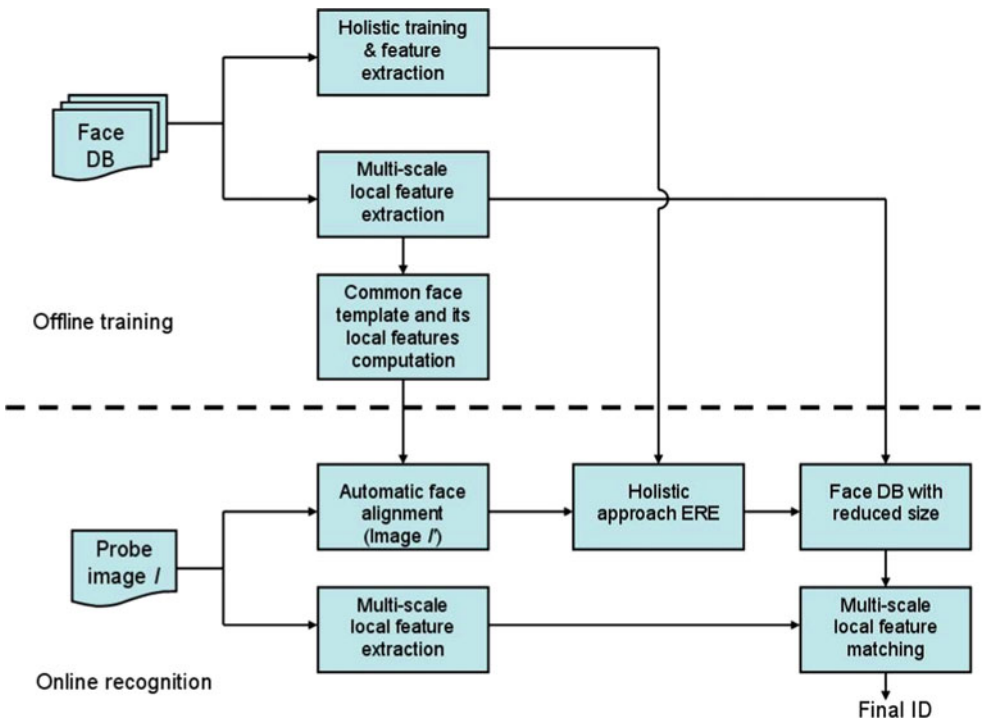


Fig. 1 Fully automatic face recognition framework based on both the local and global features

sharp corner structures with high contrast, human faces are round and smooth and hence have many indistinct structures that are yet very informative to differentiate persons. There are few obvious blobs and corners with high contrast, because the intensity changes in face images are gradual and slow in the most areas. On the other hand, the shape of the structures could be complex and some structures are close to each other or overlap. As a result, many local structures in the smooth area such as forehead, cheeks and chin cannot be detected due to the strict condition of the extrema in the 26 neighbors as proposed by Lowe [34,35]. To solve this problem, we proposed in [17] to compare a candidate point with its eight neighbors in the current scale and the corresponding one neighbor in the scale above and below. A keypoint will be selected if it is larger than all of these 10 neighbors or smaller than all of them. Figure 2 visually shows the proposed detection approach.

3.1.2 Descriptor representation

There are several descriptors proposed in the literature to represent the local image structures [5,20]. In this work, we adopt Lowe's descriptor, which is a set of histograms consisting of oriented gradients. In Lowe's SIFT framework, the support area proportional to the scale of the keypoint is divided into 4×4 blocks. An 8-bin oriented gradient histogram is computed in each block. Thus, a histogram vector $\mathbf{h}$ for each keypoint has $4 \times 4 \times 8 = 128$ dimensions. In the application of face recognition, we aim to distinguish different face images. Keypoints near the face edge carry important information about the shape of the facial contour are often of large scale and hence their support area will exceed the image area, if the image is cropped tightly to the face size. To make use of these important keypoints, we introduce a mask vector $\mathbf{m} = (m_1, m_2, \ldots, m_{128})^{\mathrm{T}}$ for each keypoint defined as

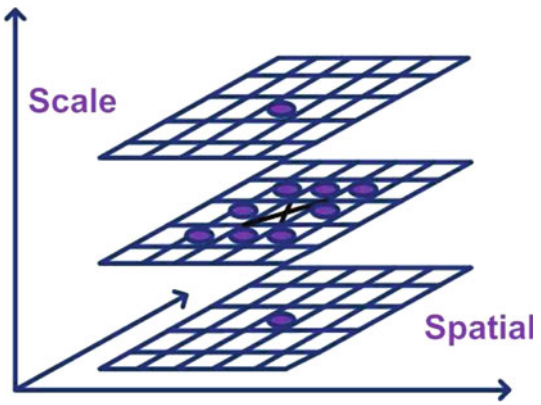


Fig. 2 Extrema of the normalized LoG images are detected by comparing a pixel (*marked with cross*) to its eight neighbors at the current scale and the corresponding pixels at the adjacent scales (*marked with circles*)

$$m_k = \begin{cases} 1, & \text{if the block is in the image;} \\ 0, & \text{if the block is outside the image.} \end{cases} \quad (1)$$

As face recognition is always following a face detection process, the rough position and size of a face in the image are known, thus, we can have the mask vector $\mathbf{m}$, which enables a partial descriptor. The feature vector $\mathbf{f}$ is computed by $\mathbf{f} = \mathbf{Mh}$, where $\mathbf{M}$ is a diagonal matrix whose diagonal elements are the elements of the mask vector $\mathbf{m}$. To invalid the contribution of the possible non-image region out of the descriptor matching, the similarity between two descriptors i and j is computed as

$$s_{ij} = \frac{\mathbf{f}_i^{\mathrm{T}} \mathbf{f}_j}{\sqrt{\mathbf{f}_i^{\mathrm{T}} \mathbf{M}_j \mathbf{f}_i \mathbf{f}_j^{\mathrm{T}} \mathbf{M}_i \mathbf{f}_j}} \quad (2)$$

It is easy to see that s_{ij} is a normalized similarity between two keypoints i and j in the common face region. It enables the participation of partial descriptors in the matching process.

3.2 Local feature matching

To determine the identity of a probe face image based on a set of gallery images, local structures of the probe image represented by the keypoints and their descriptors are compared with those in the gallery. The gallery image whose local structures have the maximum similarity to the probe image establishes the identity of the probe.

3.2.1 Search the k-nearest neighbors of the nearest subject and affine transform estimation

In the identification tasks, there are many similar gallery images. As a result, the nearest keypoints to the probe often disperse to many candidates in the gallery, and hence the probability that the largest number of the nearest keypoints fall into the right candidate is low. This problem becomes severe if the gallery contains a large number of subjects. Moreover, multiple templates per subject in the gallery make it even worse. To circumvent the problem caused by multiple templates per subject, we propose to search the k-nearest neighbors of the nearest subject. The best candidate match of a keypoint in the probe image is found by identifying its first nearest neighbor in the keypoint set of all gallery images. The first nearest neighbor is defined as the gallery keypoint whose descriptor has the maximum similarity based on 2 to that of the probe keypoint. The subject ID of the first nearest neighbor is recorded. Then, we further identify the k-nearest neighbors that have the same subject ID as the first nearest neighbor so that the $(k+1)$th-nearest neighbor has a different subject ID. If two or more such nearest neighbors fall into a same gallery image, only the one with highest similarity is chosen from them. A candidate image is identified if at least three such k-nearest neighbors are found from it. We often

obtain multiple candidate images. The minimum similarity s_m of all the probe keypoints to their k-nearest neighbors is recorded for the second stage of the image matching.

Based on the correspondence between the keypoints in the probe image and those in a candidate gallery image found in the k-nearest neighbor search, we can compute the affine transform parameters between the two images. We follow Lowe's approach here [35]. Some k-nearest neighbors are rejected by this process due to their geometric inconsistence.

3.2.2 Further matching between probe and each candidate gallery image

Although the proposed method that searches the k-nearest neighbors of the nearest subject to a probe keypoint circumvents the problem of multiple templates per subject, the k-nearest neighbors often disperse to many different subjects if the gallery contains a large number of subjects. In general, the more subjects the gallery contains, the smaller the number of the k-nearest neighbors can be found in a candidate gallery image. This decreases the probability that the largest number of the nearest keypoints fall into the gallery image with the correct ID. This problem can be very severe if the gallery contains a large number of subjects. Thus, in an identification problem, the k-nearest neighbors of the probe keypoints found in a gallery image are often only a small portion of the keypoints that can be well matched with those in the probe image. Only considering the k-nearest neighbors in the whole database as the matched keypoints in a gallery image greatly weakens the discriminative power of the local structures of an image. Therefore, we propose to further search the keypoints in each single candidate gallery image that can well match with those in the probe.

We have obtained the six affine transform parameters m_1, m_2, m_3, m_4, t_x and t_y based on the matched keypoint pairs in the k-nearest neighbor search. We project the location of a probe keypoint $[x_p, y_p]$ to the gallery image $[x'_p, y'_p]$ by the affine transform as

$$\begin{bmatrix} x'_p \\ y'_p \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x_p \\ y_p \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{3}$$

The geometric distance d between the location $[x_g, y_g]$ of a gallery keypoint and that of a transformed probe keypoint is computed as $d = \sqrt{(x_g - x'_p)^2 + (y_g - y'_p)^2}$. If a gallery keypoint i is geometrically close to a transformed probe keypoint j, $d_{ij} < d_t$, and their descriptor is similar, $s_{ij} > s_t$, keypoint i is identified as a candidate matched with keypoint j. The thresholds d_t and s_t are chosen, respectively, to be the one-fourth of the translation bin width used in the Hough transform and the minimum similarity s_m of the probe keypoints to their k-nearest neighbors. If multiple gallery keypoints satisfy the above conditions, the one with the maximum descriptor similarity is chosen as the matched keypoint. If there is no gallery keypoint that satisfies the above conditions, the probe keypoint is not matched.

The thresholds d_t and s_t will affect the number of matched keypoints of two images. It is difficult to find the optimal thresholds for all applications. To reduce the sensitivity of the image matching to the thresholds d_t and s_t, instead of the number of matched keypoints, we proposed to use the accumulated similarities over all probe keypoints. The similarity of a probe keypoint j to a candidate gallery image is defined as

$$s_j = \begin{cases} \max(s_{ij})_{i \in \mathcal{I}_j}, & \text{if } \mathcal{I}_j \neq \emptyset; \\ 0, & \text{if } \mathcal{I}_j = \emptyset. \end{cases} \tag{4}$$

where $\mathcal{I}_j = \{i | d_{ij} < d_t \ \& \ s_{ij} > s_t\}$ and i is the index of the keypoint in the candidate gallery image. The similarity score of the probe image to the candidate gallery image S_{pg} is then the accumulated similarities of all probe keypoints: $S_{pg} = \sum_{j=1}^{q} s_j$, where q is the number of keypoints in the probe image. The identity of the probe image is established as that of the gallery image that has the highest similarity score S_{pg}.

4 Face alignment

The objective of our face alignment algorithm is not to localize facial feature points such as eye-brows, eyes, nose, mouth and contour of chin. The purpose of our alignment is to rectify face images into the same canonical pose for subsequent holistic recognition tasks. As mentioned in Sect. 1, face alignment algorithms based on localizations of facial parts are not reliable as the appearances of semantic facial features vary with expressions, illuminations, occlusions or image noise. Hence, we propose an approach for face alignment not just relying on the semantic features. In the field of face recognition, there is one interesting phenomenon: the variations between the images of the same identity due to expression, illumination and viewing direction are almost always larger than image variations due to change of face identity. This is because the appearance of faces is highly constrained, for example, any frontal view of a face has eyes on the sides, nose in the middle, mouth in the lower part. Moreover, the appearances of facial components of different identities may be similar. Our face alignment approach is inspired by this phenomenon.

4.1 Generate the common face template

Given a set of face images $\mathcal{O}$ in the training database, we align them in pose and scale with manually detected two eye coordinates. Our goal in this step is to learn a common face template based on these aligned face images $\mathcal{I}$. As mentioned

above, the similarities are high, among the facial componential appearances of different subjects. The mean face $\mathfrak{m}$ of $\mathcal{I}$ captures the common information of various identities and removes noises. The SIFT keypoints detected in $\mathfrak{m}$ tell us the locations of the common and stable features in $\mathcal{I}$. Figure 3a shows the mean face $\mathfrak{m}$ computed from $\mathcal{I}$ with SIFT keypoints. We further add extra keypoints to meet the symmetry property of face images, as shown in Fig. 3b. The keypoints in Fig. 3b are the anchor points in the common face template $\mathfrak{m}$.

The anchor points in $\mathfrak{m}$ tell us the possible locations of common features in $\mathcal{I}$. However, their descriptors provide little information, as $\mathfrak{m}$ is the mean face image. The support area of SIFT descriptors in $\mathfrak{m}$ is smoothed by the mean. We need to compute the descriptors directly from the individual images in the original training set $\mathcal{O}$. As there are pose variations, the locations of the detected keypoints in $\mathcal{O}$ cannot be used in the alignment process. We should project their locations into the coordinates of the well-aligned image set $\mathcal{I}$. Let $\mathcal{P}_q = \{p_i^q\}, q = 1, \ldots Q$, where Q is the number of images in the training set $\mathcal{O}$, represent the keypoint set detected in the qth image of $\mathcal{O}$, where p_i^q is the ith keypoint in $\mathcal{P}_q$. Suppose that the two eye coordinates of the qth face image in the set $\mathcal{O}$ are $[e_{1x}, e_{1y}; e_{2x}, e_{2y}]$, and the two eye coordinates in the corresponding well-aligned face image in the set $\mathcal{I}$ are $[a_{1x}, a_{1y}; a_{2x}, a_{2y}]$. Based on these two pairs of corresponding points, we can compute the similarity transformation parameters $[s, \theta, t_x, t_y]$ between image sets $\mathcal{O}$ and $\mathcal{I}$ as below:

$$\begin{bmatrix} s\cos\theta \\ s\sin\theta \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} e_{1x} & -e_{1y} & 1 & 0 \\ e_{1y} & e_{1x} & 0 & 1 \\ e_{2x} & -e_{2y} & 1 & 0 \\ e_{2y} & e_{2x} & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a_{1x} \\ a_{1y} \\ a_{2x} \\ a_{2y} \end{bmatrix} \tag{5}$$

Once we get the transformation parameters, we can project the location of the keypoint p_i^q, $[x_p, y_p]$, to the corresponding well-aligned coordinates $[x'_p, y'_p]$ by

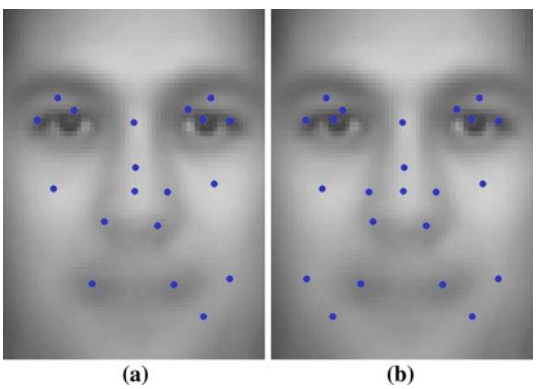


Fig. 3 Mean face $\mathfrak{m}$ with **a** SIFT keypoints; **b** symmetric keypoints $\mathcal{K}$

$$\begin{bmatrix} x'_p \\ y'_p \end{bmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{bmatrix} \begin{bmatrix} x_p \\ y_p \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{6}$$

We perform this projection on the locations of each keypoint in the set $\mathcal{P}$. Thus the keypoint descriptors of $\mathcal{P}$ capture various pose information, and their locations are well-aligned.

Let $\mathcal{K} = \{k_i\}$, $i = 1, \ldots t$, denote the anchor point set in $\mathfrak{m}$, where t is the number of anchor points in the common face template. Image set $\mathcal{O}$ is the original images with pose variations. The keypoint descriptors of $\mathcal{P}$ capture various pose information. To enhance the representative power of the template image $\mathfrak{m}$, we embed the descriptors of $\mathcal{P}$ into the anchor keypoint set $\mathcal{K}$. In a region R around the location of k_i, we search its neighbors in $\{\mathcal{P}_1, \ldots, \mathcal{P}_Q\}$. R is set to 1/6 times the image size in the experiments. If there are multiple keypoints in one face image falling into R, we select the one which is nearest to the location of k_i. In this way, around each anchor point k_i, we can find a series of keypoints $\mathcal{N}_i$ from different face images. Comparing with the scheme of one anchor point with one descriptor from one face image, variance of keypoints coming from the same semantic region of different faces enrich feature representation and are less subject to pose variations. Note that we only use the location information of anchor point k_i to locate candidate keypoints nearby in $\mathcal{P}$. We do not use the location or the descriptor of the anchor point k_i during the alignment process.

Now around each anchor point k_i, there is a series of keypoints $\mathcal{N}_i$. To make the number of keypoints in $\mathcal{N}_i$ less dependent to the number of images Q in the training database $\mathcal{O}$, we adopt hierarchical clustering [19] to group the descriptors of each keypoint set $\mathcal{N}_i$ into h clusters. The cluster center c_j^i, where $j = 1, \ldots, h$, is selected as the descriptor who has the largest accumulated cosine similarities among all the other descriptors in the same cluster. If the number of keypoints in $\mathcal{N}_i$ is smaller than h, we keep all the keypoints in $\mathcal{N}_i$. Hence, in the common face template $\mathfrak{m}$, the final number of keypoints is smaller than or equal to $t \times h$. And we denote these keypoints in the template image $\mathfrak{m}$ as final anchor point set $\mathcal{T}$.

4.2 Establish the feature correspondences

Now in the template image $\mathfrak{m}$, there are at most $t \times h$ anchor points extracted from various face images. Suppose that image I is the output of some face detector, which should be aligned into the same canonical pose as the template image $\mathfrak{m}$ for the subsequent holistic recognition process. SIFT keypoint set $\mathcal{B}$ is extracted from the image I. The best candidate match of a probe keypoint in $\mathcal{B}$ is found by identifying its nearest neighbor in the anchor point set $\mathcal{T}$. The nearest neighbor is defined as the anchor point whose descriptor has the maximum similarity to that of the probe keypoint.

Fig. 4 Sample images **a** before alignment; **b** after alignment

The nearest-neighbor search can only establish putative correspondences between keypoint sets $\mathcal{B}$ and $\mathcal{T}$. To eliminate spurious keypoint pairs, we further check their geometric consistencies. There are four parameters for each SIFT keypoint: 2D location, scale and orientation. We use Hough transform to cluster keypoint pairs with similar poses. The orientation bin size used in the Hough transform is 30°, the scale bin size is 2, and the location bin size is 0.25 times the image size. Each keypoint pair votes for the two closest bins in each dimension. Each bin with at least three keypoints is used to estimate the affine projection parameters that project the keypoint set $\mathcal{B}$ from the probe image I to the anchor point set $\mathcal{T}$. An affine transformation correctly accounts for 3D rotation of a planar surface under orthographic projection, but the approximation can be poor for 3D rotation of non-planar faces. A more general solution would be to solve for the fundamental matrix. However, a fundamental matrix solution requires at least seven keypoint pairs as compared to only three for the affine solution and in practice requires even more matches for good stability. Note that the affine transformation here is used to select keypoint pairs which are geometrically consistent. We can account for errors in the affine approximation by allowing large residual errors. When the number of keypoint pairs in the bin is larger than 3, least-squares solution is adopted to compute the affine transformation parameters. After the geometric verification, we can obtain keypoint pairs $\mathcal{B}_{\text{sub}}$ and $\mathcal{T}_{\text{sub}}$, which are subsets of $\mathcal{B}$ and $\mathcal{T}$, respectively. Note that some putative keypoint pairs are rejected by this process due to their geometric inconsistency. The anchor point set $\mathcal{T}_{\text{sub}}$ contains the final anchor points for the keypoint set $\mathcal{B}_{\text{sub}}$ to align to.

4.3 Face alignment by similarity transformation

The purpose of our face alignment is to rotate, resize and crop the output face images of face detectors automatically, which transforms them into canonical pose for the subsequent holistic recognition tasks. We do not want to change their structures. Hence, we adopt similarity transformation in the final alignment step. The similarity transformation gives the mapping of a model point $[x, y]$ to an image point $[u, v]$ in terms of an image scaling s, rotation θ, and translation $[t_x, t_y]$. We project the location of a probe keypoint $[x_p, y_p]$ in $\mathcal{B}_{\text{sub}}$ to its corresponding anchor keypoint $[x_p{}', y_p{}']$ in $\mathcal{T}_{\text{sub}}$ by the similarity transforma as

$$\begin{bmatrix} x_p & -y_p & 1 & 0 \\ y_p & x_p & 0 & 1 \\ & \cdots & & \\ & \cdots & & \end{bmatrix} \begin{bmatrix} s\cos\theta \\ s\sin\theta \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} x_p{}' \\ y_p{}' \\ \cdot \\ \cdot \end{bmatrix} \tag{7}$$

Each matched keypoint pair contributes two rows to the first and last matrices in 7. At least two matches are needed to provide a solution. We can write this linear system as

$$\mathbf{Ax} = \mathbf{b} \tag{8}$$

The least-squares solution for the parameters $\mathbf{x}$ can be determined by solving the corresponding normal equations

$$\mathbf{x} = [\mathbf{A}^{\mathbf{T}}\mathbf{A}]^{-\mathbf{1}}\mathbf{A}^{\mathbf{T}}\mathbf{b}, \tag{9}$$

which minimizes the sum of the squares of the distances from the projected model locations to the corresponding image locations.

Once we obtain the transformation parameters $[s, \theta, t_x, t_y]$, we can transform the probe image I according to the 2-D spatial similarity transformation. Figure 4a shows some sample images before alignment. Figure 4b shows the corresponding images aligned by our approach.

5 FAFF based on local and global features

5.1 Eigenfeature regularization and extraction

Jiang [24] proposed a holistic approach named ERE, which facilitates a discriminative and a stable low-dimensional feature representation of the face image. Experiments comparing ERE with some other popular subspace methods on various databases show that ERE consistently outperforms others [8,24].

Given a set of properly normalized w-by-h face images, we can form a training set of column image vectors X_{ij}, where $X_{ij} \in \mathbb{R}^{n=wh}$. Let the training set contain p persons

and q_i sample images for person i. The number of total training sample is $l = \sum_{i=1}^{p} q_i$. For face recognition, each person is a class with prior probability of c_i. The within-class scatter matrix is defined by $\mathbf{S}^w = \sum_{i=1}^{p} \frac{c_i}{q_i} \sum_{j=1}^{q_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^{\mathrm{T}}$, where $\bar{X}_i = \frac{1}{q_i} \sum_{j=1}^{q_i} X_{ij}$. By solving the eigenvalue problem, we have

$$\Lambda = \Phi^{wT} \mathbf{S}^w \Phi^w, \tag{10}$$

where $\Phi^w = [\phi_1^w, \ldots, \phi_n^w]$ is the eigenvector matrix of $\mathbf{S}^w$, and Λ^w is the diagonal matrix of eigenvalues $\lambda_1^w, \ldots, \lambda_n^w$ corresponding to the eigenvectors.

The eigenspace $\mathbb{R}^n$ spanned by eigenvectors $\{\phi_k^w\}_{k=1}^n$ is decomposed into three subspaces: a reliable face variation dominating subspace (or simply face space) $\mathbf{F} = \{\phi_k^w\}_{k=1}^m$, an unstable noise variation dominating subspace (or simply noise space) $\mathbf{N} = \{\phi_k^w\}_{k=m+1}^r$ and a null space $\emptyset = \{\phi_k^w\}_{k=r+1}^n$, where r is the rank of $\mathbf{S}^w$. To determine the start point of the noise dominant region $m + 1$, we first find the median of non-zero eigenvalues $\lambda_{\mathrm{med}}^w = \mathrm{median}\{\forall \lambda_k^w | k \le r\}$. The distance between λ_{med}^w and the smallest non-zero eigenvalue is $d_{m,r} = \lambda_{\mathrm{med}}^w - \lambda_r^w$. The start point of the noise region $m + 1$ is estimated by

$$\lambda_{m+1}^w = \max\{\forall \lambda_k^w | \lambda_k^w < (\lambda_{\mathrm{med}}^w + \mu(\lambda_{\mathrm{med}}^w - \lambda_r^w))\}, \tag{11}$$

where μ is a constant. The optimal value of μ may be slightly larger or smaller than 1 for different applications.

The eigenspectrum is regularized by replacing the noise dominating λ_k^w in $\mathbf{N}$ with a model $\alpha/(k + \beta)$ and replacing the zero λ_k^w in the null space $\emptyset$ with the a constant. The regularized eigenspectrum $\tilde{\lambda}_k^w$ is given by

$$\tilde{\lambda}_k^w = \begin{cases} \lambda_k^w, & k < m; \\ \frac{\alpha}{k+\beta}, & m \le k \le r; \\ \frac{\alpha}{r+1+\beta}, & r < k \le n, \end{cases} \tag{12}$$

where $\alpha = \frac{\lambda_1^w \lambda_m^w (m-1)}{\lambda_1^w - \lambda_m^w}$, and $\beta = \frac{m\lambda_m^w - \lambda_1^w}{\lambda_1^w - \lambda_m^w}$. The training data are transformed to $\tilde{Y}_{ij} = \tilde{\mathbf{\Phi}}_n^{wT} X_{ij}$, where

$$\tilde{\mathbf{\Phi}}_n^w = \left[\phi_1^w / \sqrt{\tilde{\lambda}_1^w}, \ldots, \phi_n^w / \sqrt{\tilde{\lambda}_n^w} \right]. \tag{13}$$

After the feature regularization, a new total scatter matrix is formed by vectors $\tilde{Y}_{ij}$ of the training data as $\tilde{\mathbf{S}}^t = \sum_{i=1}^{p} \frac{c_i}{q_i} \sum_{j=1}^{q_i} (\tilde{Y}_{ij} - \bar{Y})(\tilde{Y}_{ij} - \bar{Y})^{\mathrm{T}}$, where $\bar{Y} = \sum_{i=1}^{p} \frac{c_i}{q_i} \sum_{j=1}^{q_i} \tilde{Y}_{ij}$.

The regularized features $\tilde{Y}_{ij}$ will be decorrelated for $\tilde{\mathbf{S}}^t$ by solving the eigenvalue problem similar to 10. The eigenvectors in the eigenvector matrix $\tilde{\mathbf{\Phi}}_n^t = [\tilde{\phi}_1^t, \ldots, \tilde{\phi}_n^t]$ are sorted in a descending order of the corresponding eigenvalues. The dimensionality reduction is performed by keeping the eigenvectors with the d largest eigenvalues $\tilde{\mathbf{\Phi}}_d^t = [\tilde{\phi}_k^t]_{k=1}^d = [\tilde{\phi}_1^t, \ldots, \tilde{\phi}_d^t]$, where d is the number of features usually selected by a specific application. The feature regularization and extraction matrix $\mathbf{U}$ is given by $\mathbf{U} = \tilde{\mathbf{\Phi}}_n^w \tilde{\mathbf{\Phi}}_d^t$, which transforms a face image vector $X, X \in \mathbb{R}^n$, into a feature vector $F, F \in \mathbb{R}^d$, by $F = \mathbf{U}^{\mathrm{T}} X$.

5.2 FAFF based on both the local and global features

Some local feature-based approaches achieve better recognition performance than holistic approaches [17], but their computational burden is much heavier. To speed up the local feature-based approach, we propose to use the holistic approach as a filter to retrieve candidate images from the whole gallery set. The retrieved images form a narrowed gallery set, on which we perform the multi-scale local feature-based approach for face recognition. The reduction in the size of the gallery set will significantly relieve its computational complexity.

Image I is the output of a face detector. Firstly, we align image I by the approach proposed in Sect. 4. Then the aligned image I' can be used as the input of the holistic approach ERE [24]. To calculate the similarities between the probe image I' and all the gallery images, each n-D face image vector is transformed into d-D feature vector F using the feature regularization and extraction matrix $\mathbf{U}$ obtained in the training stage. Then the cosine similarity measure between a probe feature vector F_p and a gallery feature vector F_g is applied,

$$S(F_p, F_g) = \frac{F_p^{\mathrm{T}} F_g}{\|F_p\|_2 \|F_g\|_2}, \tag{14}$$

where $\|\cdot\|_2$ is the norm 2 operator.

There are m similarity scores $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ between image I' and m gallery images. The scores $S_i, i = 1, \ldots, m$ are ordered as

$$S_{(1)} \ge S_{(2)} \ge \ldots \ge S_{(m)}, \tag{15}$$

and the gallery images with the top n highest similarity scores are selected as the input of the following multi-scale LFEM. The final image ID is the one with the highest similarity score of the n candidate gallery images computed by the LFEM approach. The rationale behind this cascaded process is as follows:

1. ERE is one of the best performed holistic approaches.
2. The probability that the true image ID is included in the top n gallery images is high.
3. The computational complexity of the ERE approach at the recognition stage is significantly lower than the LFEM approach.
4. The recognition performance of LFEM is better than the ERE approach.

5. For the LFEM approach, comparing image I' with n gallery images is much faster than comparing it with all the m gallery images ($n < m$).

6 Experimental results

In this section, we will test our proposed approaches on AR [37], GT [38] and ORL [46] databases. Firstly, we will validate our face alignment strategy through LBP [1] and three holistic face recognition approaches: LDA [2], UFS [47] and ERE [24]. Secondly, we will report the results of our fully automatic face recognition framework based on both the local and global features (FAFF).

6.1 Validation of our face alignment strategy

Eight face alignment approaches are compared: adaboost eye detector [18,25], eye localization by pixel differences (PD) [41], eye localization by rank order filter (ROF) [40], eye localization by cascaded asymmetric principal and discriminative analysis (C-APCDA) [22,42], localization of semantic facial features [12], congealing [21], our proposed approach (Prop.) and manual alignment (MA).

6.1.1 Results on AR database

The color images in AR database are converted to gray-scale and cropped into the size of 60×85 pixels. We conduct two sets of experiments on this database.

In the first set, 75 persons with 14 non-occluded images per person are selected, which makes the database containing 1,050 images. At the alignment stage, for the C-APCDA eye detector [42], the first seven images per subject serve as training images and are aligned manually by two eye coordinates. The remaining seven images per subject serve as the output of the face detector, which should be aligned. For our proposed alignment approach, the first seven images per subject are aligned manually to generate the face template. At the recognition stage, images are divided into 5×5 windows for the LBP [1] approach. For the three holistic approaches LDA [2], UFS [47] and ERE [24], the first 7 images of all subjects are used in the training and gallery sets, which are normalized manually. The remaining seven images of all subjects serve as probe, which are aligned by different approaches. The best recognition performances of the holistic approaches LDA [2], UFS [47] and ERE [24] over all possible numbers of features are recorded.

To verify the performance of our alignment approach on occluded face images, in the second set of experiments, we select 55 persons with 12 occluded images per person (some sample images are shown in Fig. 5). The first 6 images of all 55 subjects are used in the training and gallery sets and the remaining 6 images of all subjects serve as probe. Obviously, alignment approaches that are only based on eye localizations fail in this case because of the occlusion. Therefore, we compare the performances of three alignment approaches: localization of semantic facial features [12], our proposed approach and the manual alignment.

Fig. 5 Sample images of occluded face

Table 1 Recognition rate on AR database

	LBP (%)	LDA (%)	UFS (%)	ERE (%)
SET 1				
MA	98.10	93.90	95.05	95.05
Adaboost [18,25]	95.81	89.71	91.05	90.67
PD [41]	87.24	80.38	82.86	82.29
ROF [40]	73.52	69.71	70.48	71.24
C-APCDA [42]	94.67	89.52	90.10	89.14
SF [12]	91.05	84.76	87.05	86.48
Prop.	95.43	92.95	92.95	94.10
SET 2				
MA	86.67	72.12	72.12	80.30
SF [12]	50.91	23.64	27.58	32.12
Prop.	76.06	58.79	60.00	62.73

Table 1 shows the rank one recognition rates on AR database of four recognition algorithms based on seven different alignment approaches. From this table, we can see that in the first set of experiments, the recognition performance achieved by our alignment approach, though slightly worse than manual alignment, outperforms all other automatic alignment approaches. Because in our alignment approach we do not just depend on the positions of semantic facial parts, our approach can get more reliable alignment results. Figure 6a gives the cumulative matching curves obtained from seven different alignment approaches based on the ERE approach. The cumulative recognition performance obtained by our alignment approach is significantly better than other automatic alignment approaches. The second set of experiments is conducted on the occluded face images. The results in Table 1 show that our proposed alignment approach significantly outperforms the alignment approach based on localizations of semantic facial features [12]. The recognition

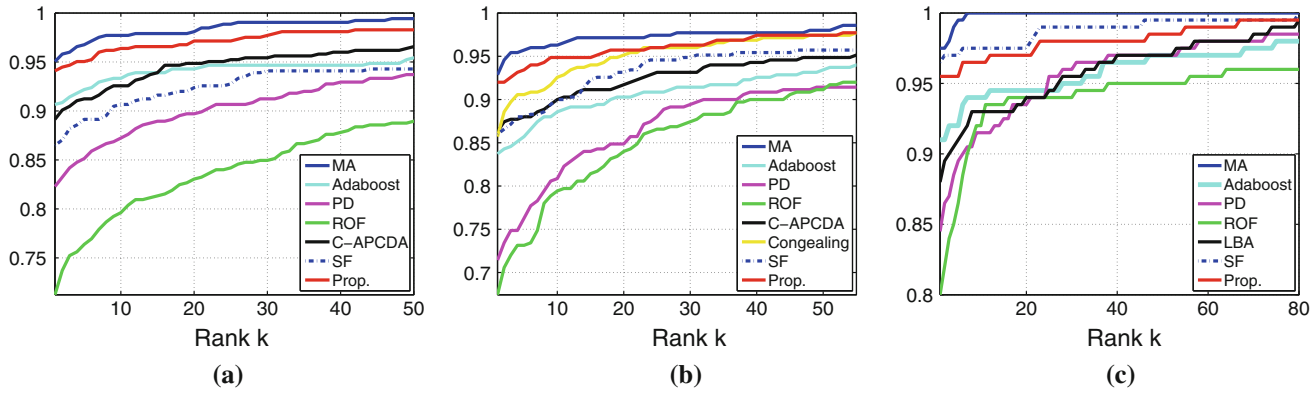


Fig. 6 Cumulative matching curves of different alignment approaches obtained by ERE on **a** AR database; **b** GT database; **c** ORL database

rates are quite low because all face images in the training/gallery and probe sets are occluded.

6.1.2 Results on GT database

The GT database consists 750 color images of 50 subjects (15 images per subject). These images have large variations in both pose and expression and some illumination changes. Images are converted to gray scale and cropped into the size of 60 × 80. At the alignment stage, for the C-APCDA approach [42], the first eight images per subject serve as training images and are aligned manually based on the two eye positions. The remaining seven images per subject serve as the output of the face detector, which should be aligned. For our proposed approach, the first eight images per subject are aligned manually to generate the template. At the recognition stage, images are divided into 5 × 5 windows for the LBP approach [1]. For the three holistic approaches LDA [2], UFS [47] and ERE [24], the first eight images of all subjects are used in the training and gallery sets, which are normalized manually. The remaining seven images of all subjects serve as probe images, which are aligned by different approaches. The best recognition performances of the holistic approaches LDA [2], UFS [47] and ERE [24] over all possible numbers of features are recorded.

Table 2 shows the rank one recognition rates on GT database of four recognition algorithms based on eight different alignment approaches. Comparing different automatic face alignment approaches, the performance of our face alignment approach is significantly better than eye-based approaches: Adaboost [18,25], PD [41], ROF[40] and C-APCDA [42], semantic feature-based approach [12] and the unsupervised approach: Congealing [21]. Figure 6b gives the cumulative matching curves obtained from eight different alignment approaches based on the ERE approach. The cumulative recognition performance obtained by our alignment approach is significantly better than other automatic alignment approaches.

Table 2 Recognition rate on GT database

	LBP (%)	LDA (%)	UFS (%)	ERE (%)
MA	90.29	92.00	91.43	92.86
Adaboost [18,25]	85.43	82.29	81.71	83.71
PD [41]	72.00	64.29	69.14	71.43
ROF [40]	84.29	60.29	64.00	67.43
C-APCDA [42]	85.43	83.43	83.71	86.29
Congealing [21]	83.43	84.29	82.29	85.71
SF [12]	85.43	83.43	84.57	86.00
Prop.	84.29	92.57	90.57	92.00

6.1.3 Results on ORL database

Images of ORL database are cropped into the size of 50 × 57. The ORL database contains 400 images of 40 people (10 images per person). At the alignment stage, for the learning-based approach [42], the first five images per subject serve as training images and are aligned manually by the two eye coordinates. The remaining five images per subject serve as the output of the face detector, which should be aligned. For our proposed alignment approach, the first five images per subject are aligned manually to generate the template. At the recognition stage, for the LBP approach [1], images are divided into 3 × 3 windows. For the three holistic approaches LDA [2], UFS [47] and ERE [24], the first five images of all subjects are used in the training and gallery sets, which are normalized manually. The remaining five images of all subjects serve as probe images, which are aligned by different approaches. The best recognition performances of the holistic approaches LDA [2], UFS [47] and ERE [24] over all possible numbers of features are recorded.

Table 3 shows the rank one recognition rates on ORL database of four recognition algorithms based on seven different alignment approaches. Comparing different automatic alignment approaches, our proposed approach achieves

Table 3 Recognition rate on ORL database

	LBP (%)	LDA (%)	UFS (%)	ERE (%)
MA	95.0	92.5	83.5	97.0
Adaboost [18,25]	87.5	88.5	77.0	91.0
PD [41]	85.5	80.5	69.5	84.5
ROF [40]	86.5	74.5	65.5	80.0
C-APCDA [42]	85.5	83.0	62.5	88.0
SF [12]	93.5	93.0	81.5	96.5
Prop.	94.0	93.0	83.0	95.5

comparable performance as the semantic feature-based approach SF [12], but consistently significantly outperforms others over different face recognition algorithms. Figure 6c gives the cumulative matching curves obtained from seven different alignment approaches based on the ERE approach. The cumulative recognition performance obtained by our alignment approach is ,though slightly worse than the SF [12] approach, significantly better than other automatic alignment approaches.

6.2 Validation of our FAFF

In this section, we will validate our FAFF based on the holistic approach ERE [24] and the multi-scale LFEM [17], on three databases AR, GT and ORL. Our FAFF utilizes the classification results of both the ERE and the LFEM approaches. From some points of view, it is a kind of score fusion [45] for multiple algorithms. To validate the efficacy of our proposed framework, we compare it with ERE, LFEM and two score fusion approaches: sum rule with equal weights (SR_{ew}) [45] and sum rule with weights learned by LIBLINEAR (SR_{lw}) [13]. The experimental settings in this section are similar to those in Sect. 6.1. The only difference is that in this section, the images used in the holistic approaches are aligned automatically by our proposed face alignment strategy. Table 4 gives the percentage of the gallery images retrieved by the holistic approach ERE at two different levels r_1 and r_2. Our FAFF approach achieves the same recognition performance as LFEM by comparing one probe with r_1 images in the gallery. And by comparing with r_2 images, the recognition performance of our framework reaches the best. Table 5 gives

Table 4 Percentage of the gallery images selected by ERE where FAFF achieves the same recognition performance as LFEM on the original database at r_1 and the best performance at r_2

AR database		GT database		ORL database	
r_1	r_2	r_1	r_2	r_1	r_2
40 %	40 %	15 %	80 %	20 %	40 %

Table 5 Recognition rate on AR, GT and ORL databases

	AR (%)	GT (%)	ORL (%)
ERE	94.10	92.00	95.5
LFEM	99.05	95.43	98.0
SR_{ew}	98.48	96.86	98.5
SR_{lw}	99.05	96.29	98.0
FAFF	99.05	97.43	99.0
LBP	95.43	84.29	94.0
LDA	92.57	92.57	93.0
UFS	92.95	90.57	83.0
AFS	97.14	94.00	–

the rank one recognition rate of the holistic approach ERE [24], local feature based approach LFEM, score fusion of ERE and LFEM by sum rule with equal weights (SR_{ew}) [45] and learned weights (SR_{lw}) [13], and the best recognition performance of FAFF by selecting r_2 images of all the gallery. To compare the performances of FAFF and other face recognition algorithms, we also report the recognition accuracies of LBP, LDA, UFS and the attribute face service (AFS) [28,29]. Because currently the AFS cannot handle grayscale images, we only give its performance on the AR and GT databases.

From Table 5, we see that on AR database, the best recognition rate is achieved by LFEM, SR_{lw} and FAFF. However, the proposed FAFF is about twice as faster as LFEM and SR_{lw} because the number of gallery images is reduced to 40% by ERE (from Table 4) and the time consumption of ERE is negligible comparing to LFEM. On GT database, the FAFF approach outperforms all other approaches in Table 5 and is yet less time consuming than LFEM, SR_{ew} and SR_{lw} because the number of gallery images is reduced to 80% by ERE. SR_{ew} and SR_{lw} outperform LFEM at a price of greater computational complexity as they parallel combine LFEM with ERE. In contrast, the proposed FAFF not only speeds up the LFEM, but also enhances its accuracy more than the two parallel combination approaches SR_{ew} and SR_{lw}. According to Table 4, the proposed FAFF can achieve the same recognition rate as LFEM by comparing the probe with only 15% gallery images. Similarly, on ORL database, by comparing the probe with 20% gallery images, the proposed FAFF reaches the same performance as LFEM. Increasing the number of images to 40% of the whole gallery set, the recognition rate of the proposed FAFF is the highest among all approaches in Table 5.

7 Conclusion

Face recognition algorithms can be divided into two categories based on the types of features: holistic and local

feature-based approaches. Holistic approaches become popular due to their efficacy and efficiency. However, they depend on careful positioning of the face images into the same canonical pose. This is not an easy task because it depends on the accurate detection of at least two landmarks from the face image. Besides eye detection, some face alignment algorithms rely on other semantic facial parts. However, in real applications the appearances of these semantic features may not be distinct or missing due to expressions, occlusions, illuminations or noise. Local feature-based approaches to face recognition [17,34,35] are more robust to image variations in pose and scale than holistic ones. The multi-scale LFEM proposed in [17] even achieves better recognition accuracy than many popular holistic methods. However, its computational burden is much heavier.

In this paper, we propose a FAFF based on both the local and global features. To relieve the computational burden of local feature-based approaches, we firstly apply the holistic method to retrieve candidate images from the whole gallery set. The selected gallery images have higher probabilities matching to the probe than the remains, which form a new gallery set with reduced size. Then the local feature-based approach is performed on the narrowed gallery set. The reduction in the size of the gallery set relieves the computational burden of local feature-based approaches. Furthermore, the recognition accuracy is better than the holistic approach, local feature-based approach and their parallel combination. As the local features can be used to align images automatically, we propose an alignment strategy based on non-semantic multi-scale local features. Given a set of aligned images in the training data set, a common face template with common keypoints is trained. Putative correspondences are established between the keypoint sets from the unaligned probe face image and the learnt face template. Geometric verifications are performed to eliminate spurious matches with inconsistent poses.

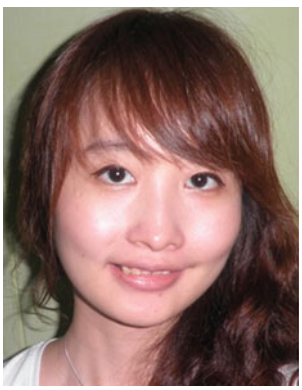
Experimental results demonstrate that the proposed FAFF not only speeds up the local feature-based approach for face recognition, but also improves the recognition accuracy over the holistic approach, local feature-based approach and their parallel combination.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 2037–2041 (2006)
2. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
3. Berg, T., Berg, A., Maire, M., White, R., Teh, Y., Learned-Miller, E., Forsyth, D.: Names and faces in the news. In: CVPR (2004)
4. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of sift features for face authentication. In: Workshop on Computer Vision and Pattern Recog, pp. 35–40 (2006)
5. Brown, M., Hua, G., Winder, S.: Discriminative learning of local image descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **33**(1), 43–57 (2011)
6. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **27**(1), 4–13 (2005)
7. Chen, Y., Kubo, K.: A robust eye detection and tracking technique using gabor filters. In: IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, vol. 1, pp. 109–112 (2007)
8. Choi, J., Neve, W., Plataniotis, K., Ro, Y.: Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks. IEEE Trans. Multimed. **13**(1), 14–28 (2011)
9. Cootes, T., Edwards, G., Taylor, C.: Active apprearnce models. In: ECCV (1998)
10. Cruz, C., Sucar, L., Morales, E.: Real-time face recognition for human-robot interaction. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 1–6 (2008)
11. D'Orazio, T., Leo, M., Cicirelli, G., Distante, A.: An algorithm for real time eye detection in face images. ICPR **3**, 278–281 (2004)
12. Everingham, M., Sivic, J., Zisserman, A.: "Hello! My name is... Buffy"—automatic naming of characters in TV video. In: BMVC, pp. 889–908 (2006)
13. Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C.: Liblinear: a library for large linear classification. JMLR **9**, 1871–1874 (2008)
14. Fernandez, C., Vicente, M.: Face recognition using multiple interest point detectors and sift descriptors. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 1–7 (2008)
15. Geng, C., Jiang, X.: Face recognition using sift features. In: ICIP, pp. 3313–3316 (2009)
16. Geng, C., Jiang, X.: Sift features for face recognition. In: IEEE International Conference on Computer Science and Information Technology, pp. 598–602 (2009)
17. Geng, C., Jiang, X.: Face recognition based on the multi-scale local image structures. Pattern Recognit. **44**(10–11), 2565–2575 (2011)
18. Hadid, A., Heikkila, J., Silver, O., Pietikäinen, M.: Face and eye detection for person suthentication in mobile phones. In: IEEE International Conference on Distributed Smart Camera, pp. 101–108 (2007)
19. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning. Springer, Berlin (2009)
20. Hua, G., Brown, M., Winder, S.: Discriminant embedding for local image descriptors. In: ICCV, pp. 1–8 (2007)
21. Huang, G., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: ICCV, pp. 1–8 (2007)
22. Jiang, X.D.: Asymmetric principal component and discriminant analyses for pattern classification. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 931–937 (2009)
23. Jiang, X.D.: Linear subspace learning-based dimensionality reduction. Signal Process. Mag. **28**(2), 16–26 (2011)
24. Jiang, X.D., Mandal, B., Kot, A.: Eigenfeature regularization and extraction in face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 383–394 (2008)
25. Jung, S., Chung, Y., Yoo, J., Moon, K.: Real-time face verification for mobile platforms. Advances in visual computing pp. 823–832 (2008)
26. Kirby, M., Sirovich, L.: Application of karhunen–loeve procedure for the characterization of human faces. IEEE Trans. Pattern Anal. Mach. Intell. **12**(1), 103–108 (1990)
27. Kisku, D., Tistarelli, M., Sing, J., Gupta, P.: Face recognition by fusion of local and global matching scores using ds theory: an

evaluation with uni-classifier and multi-classifier paradigm. In: IEEE Computer Vision and Pattern Recognition, pp. 60–65 (2009)
28. Kumar, N.: Attribute face service. http://afs.automaticfacesystems.com/
29. Kumar, N., Berg, A., Belhumeur, P., Nayar, S.: Describable visual attributes for face verification and image search. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10), 1962–1977 (2011)
30. Li, G., Cai, X., Li, X., Liu, Y.: An efficient face normalization algorithm based on eye detection. In: IEEE International Conference on Intelligent Robots and Systems, pp. 3843–3848 (2006)
31. Li, S., Yan, S., Zhang, H., Cheng, Q.: Multi-view face alignment using direct appearnce models. In: IEEE International Conference on Automatic Face and Gesture Recognition (2002)
32. Lin, D., Yang, C.: Real-time eye detection using face circle fitting and dark-pixel filtering. In: ICME, pp. 1167–1170 (2004)
33. Lindeberg, T.: Feature detection with automatic scale selection. IJCV **30**, 79–116 (1998)
34. Lowe, D.: Object recognition from local scale-invariant features. ICCV **2**, 1150–1157 (1999)
35. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60**, 91–110 (2004)
36. Luo, J., Ma, Y., Takikawa, E., Lao, S., Kawade, M., Lu, B.: Person-specific sift features for face recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 593–596 (2007)
37. Martinez, A., Benavente, R.: Cvc technical report #24. Technical report (1998)
38. Nefian, A.: Embedded bayesian networks for face recognition. In: ICME, pp. 133–136 (2002)
39. Park, C., Kwak, J., Park, H., Moon, Y.: An effective method for eye detection based on texture information. In: IEEE International Conference on on Information Technology Convergence, pp. 586–589 (2007)
40. Ren, J., Jiang, X.: Eye detection based on rank order filter. In: IEEE International Conference on Information, Communications and Signal Processing, pp. 1–4 (2009)
41. Ren, J., Jiang, X.: Fast eye localization based on pixel differences. In: ICIP, pp. 2733–2736 (2009)
42. Ren, J., Jiang, X., Yuan, J.: A fast and accurate cascade subspace face/eye detector on mobile devices. In: International Conference on Computer Vision (ICCV), Workshop on Mobile Vision, pp. 84–91 (2011)
43. Riopka, T., Boult, T.: The eyes have it. In: ACM SIGMM Workshop on Biometrics Methods and Applications, pp. 9–16 (2003)
44. Rosenberger, C., Brun, L.: Similarity-based matching for face authentication. In: ICPR, pp. 1–4 (2008)
45. Ross, A., Jain, A.: Information fusion in biometrics. Pattern Recogn. Lett. **24**(13), 2115–2125 (2003)
46. Samaria, F., Harter, A.: Parameterisation of a stochastic model for human face identification. In: 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142 (1994)
47. Wang, X., Tang, X.: A unified framework for subspace face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1222–1228 (2004)
48. Wu, J., Trivedi, M.: A binary tree for probability learning in eye detection. In: Workshop on Computer Vision and Pattern Recognition, pp. 170–177 (2005)
49. Yan, S., Xu, D., Zhang, B., Yang, Q., Zhang, H., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
50. Yang, J., Frangi, A., Yang, J., Zhang, D., Jin, Z.: KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 230–244 (2005)
51. Ye, J., Janardan, R., Park, C., Park, H.: An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 982–994 (2004)
52. Zheng, W., Tang, X.: Fast algorithm for updating the discriminant vectors of dual-space lda. IEEE Trans. Info. Forensics Secur. **4**(3), 418–427 (2009)
53. Zhou, Y., Gu, L., Zhang, H.: Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In: CVPR (2003)
54. Zhou, Y., Zhang, W., Tang, X., Shum, H.: A bayesian mixture model for multi-view face alignment. In: CVPR (2005)

Author Biographies



Cong Geng received the B.Eng. degree in electrical and electronic engineering from the Wuhan University, China, in 2006. Since 2007, she has been a PhD candidate in the Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her research interests include biometrics, pattern recognition, computer vision, and machine learning. She is working on feature extraction and matching using feature based methods for face recognition.

Xudong Jiang received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China (UESTC) in 1983 and 1986, respectively, and the PhD degree from Helmut Schmidt University Hamburg, Germany, in 1997, all in electrical engineering. From 1986 to 1993, he was a lecturer at UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1993 to 1997, he was with Helmut Schmidt University Hamburg, as a scientific assistant. From 1998 to 2004, he was with the Institute for Infocomm Research, A*STAR, Singapore, as a lead scientist and the head of the Biometrics Laboratory where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition (FVC'00). Since 2003, he has been a faculty member in Nanyang Technological University, Singapore. Currently, he is a tenured associate professor. Dr. Jiang has published over 100 papers and holds 7 patents. His research interest includes signal/image processing, pattern recognition, computer vision, machine learning and biometrics.