Pattern Recognition 44 (2011) 2565–2575

Contents lists available at ScienceDirect

Pattern Recognition

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

Face recognition based on the multi-scale local image structures

Cong Geng *, Xudong Jiang

School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Link, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 14 September 2010
Received in revised form
14 January 2011
Accepted 13 March 2011
Available online 24 March 2011

Keywords:
Face recognition
Feature extraction
Local image structure
Keypoint detection
Image matching
Template selection
Template synthesis

ABSTRACT

This paper proposes a framework of face recognition based on the multi-scale local structures of the face image. While some basic tools in this framework are inherited from the SIFT algorithm, this work investigates and contributes to all major steps in the feature extraction and image matching. New approaches to keypoint detection, partial descriptor and insignificant keypoint removal are proposed specifically for human face images, a type of non-rigid and smooth visual objects. A strategy of keypoint search for the nearest subject and a two-stage image matching scheme are developed for the face identification task. They circumvent the problem that local structures matched with those in probe disperse into many different gallery images. Although the proposed framework can work for single template per subject, a training procedure is developed for multiple samples per subject. It contains template selection, unstable keypoint removal and template synthesis to meet different requirements in face recognition applications. Each ingredient of the proposed framework is experimentally validated and compared with its counterpart in the SIFT scheme. Results show that the proposed framework outperforms SIFT and some holistic approaches to face recognition.

1. Introduction

The ability to recognize human faces is a demonstration of incredible human intelligence. Over the last three decades researchers from diverse areas have been making attempts to replicate this outstanding visual perception of human beings in machine recognition of faces [1]. However, there are still substantial challenging problems such as intraclass variations in three-dimensional pose, facial expression, make-up and lighting condition as well as occlusion and cluttered background. To deal with these difficulties, numerous algorithms have been proposed, which can be coarsely classified into two categories, holistic approaches and local feature/component based approaches.

Although human beings can easily recognize face images, it is unclear what features or image structures are used in the human intelligence for this recognition task. Therefore, holistic approaches do not explicitly utilize the image structure information in feature extraction. They take all pixels of a face image as initial features and extract a set of reliable and discriminative features based on machine learning from an available database. Since the principal component analysis (PCA) [2] and the linear discriminant analysis (LDA) [3] were introduced into face recognition, various holistic approaches have been extensively studied, such as Bayesian algorithm [4], the direct LDA [5], the null-space LDA [6], the dual-space LDA [7,8], the unified framework [9], the generalized LDA [10] and the locality preserving projections (LPP) [11]. Some of these approaches are summarized under a common framework graphically [12] and algebraically [13]. Recent developments of the holistic approaches include the marginal Fisher analysis (MFA) [12], eigenfeature regularization and extraction (ERE) [14], the sparse representation [15] and asymmetric PCA and LDA [16,17]. In general, 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 [18]. As a result, most approaches work on the normalized face images based on the manually identified landmarks. However, the recognition performance will deteriorate considerably if the manual process is replaced by an automatic landmark detection algorithm. Moreover, global features are sensitive to variations in facial expressions, poses and occlusions. Another intrinsic problem of all holistic approaches is their dependence to the training databases because knowledge about the face discrimination is generalized by machine learning from the face samples. A representative training database is necessary, which, however, is not available in many applications.

In contrast to holistic methods, local feature based approaches have the potential of more robust to variations in pose, scale, expression and occlusion [1,19]. Elastic bunch graph matching (EBGM) [20] and active appearance model (AAM) [21] fall into this category. However, the performances of both EBGM and AAM

* Corresponding author. Tel.: +65 98182956.
E-mail addresses: geng0007@ntu.edu.sg (C. Geng), exdjiang@ntu.edu.sg (X. Jiang).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.03.011

depend on a good selection of facial landmarks, which are often annotated manually. This makes the approaches semi-automatic and labor consuming.

One of the very fundamental problems arising when analyzing face images originates from the fact that face structures appear in different ways, depending upon the scales of observations. First, facial local structures are shown at different levels of scales, ranging from skin textures at fine scales, through eyes and mouth represented at median scales, to the shape of face contours at large scales. Second, the characteristic or the description of a local structure is strongly dependent on the scale at which the structure is modeled. Third, it is unknown in advance what the proper scales are to describe different local structures of unknown face images. To cope with these problems, an image representation that explicitly incorporates the notion of scale is a crucially important tool. In [22], multi-scale local features are extracted from a dense set of multi-scale image patches that deliver good generalization ability for face recognition. However, a multi-scale representation by itself contains no explicit information about what image structures are significant and what scales are appropriate to describe them. Thus, it is essential to complement a multi-scale process by explicit mechanisms for automatic scale selection.

The scale invariant feature transform (SIFT) [23,24] detects distinct local structures from images and selects appropriate scales to describe them automatically. It has shown good performance on object detection and some other machine vision applications [25–29]. Recently, some initial attempts apply the SIFT algorithm in the face recognition task. In [30] an overlapping sub-image matching strategy is used as the first attempt to explore the SIFT approach for face recognition. In [31], the SIFT descriptor is adopted to describe irregular local marks detected by a Hessian–Laplace detector. In [32–34], a graph is built on SIFT features. The recognition problem is modeled as a graph matching process. In [35], the authors propose a method based on SIFT and support vector machine. Fernandez and Vicente [36] combine Harris–Laplace and Difference-of-Gaussian detectors to detect both corner and blob structures in face images and use SIFT descriptors to represent them. In [37–39], salient regions are firstly identified from face images, and then SIFT features are extracted in each region. A modified keypoint descriptor and a redundant keypoint removal scheme are proposed for face recognition in [40,41]. Majumdar and Ward [42] rank the SIFT features according to their discriminative power and use the most discriminating ones for face recognition.

Despite all efforts above, there are still many outstanding issues and problems that need to be addressed and circumvented if we are to leverage the idea of SIFT and some of its good properties to solve the challenging face recognition problem. 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. This will cause heavy computational burden. In this paper, we first propose a training procedure to speed up the face recognition task if multiple training samples per subject are available. It contains template selection, unstable keypoint removal and template synthesis to meet different requirements of face recognition applications. Secondly, to enhance the identification performance, we analyze the merits and deficiencies of SIFT and propose new strategies for feature extraction and image matching, which leads to a new framework that overcomes limitations of SIFT in solving the face recognition problem. We propose a new approach to keypoint detection which can capture the information of many facial structures in the smooth area such as forehead, cheeks and chin. A partial descriptor is designed to represent the keypoints whose support areas exceed the face image. Our proposed detection approach and partial descriptor strategy produce a rich number of keypoints. As a significant keypoint should be distinct from others in terms of either its location or the image structures of its neighborhood, we further propose to remove keypoints based on their distinctiveness. A two-stage image matching scheme and a strategy of keypoint search for the nearest subject are developed to cater for the identification task. It circumvents the problem that the most similar local structures to the probe image disperse to many different gallery images. Finally, we perform the training procedure for multiple samples per subject based on our proposed feature extraction and matching framework to speed up the face recognition system with significantly better identification performance than that based on the original SIFT algorithm.

2. Training for multiple image samples per subject

While multiple templates per subject in general increase the recognition rate as shown in the experiments later, they bring a great computational burden in the recognition process. It is possible to greatly reduce the computational complexity of the matching process by removing redundant information resides in the multiple training images. To meet different requirements for the computational complexity, we propose three schemes: template selection, unstable keypoint removal and template synthesis.

2.1. Template selection

From a set of training images of a subject, we want to select a subset of images serving as templates that best represent all training images of this subject in terms of differentiating it from others in the training database. Suppose in the training data set $\mathcal{D}$, each subject has a training set $\mathcal{S}$ with N images, $N>1$. To select q templates from them, $1\le q<N$, we pick a subset $\mathcal{P}_k$ that contains $N-q$ images of this subject out of the training data set $\mathcal{D}$. The remaining training images form a data set $\mathcal{G}_k$, $\mathcal{G}_k=\mathcal{D}-\mathcal{P}_k$. The candidate template set selected for this subject is then $\mathcal{T}_k=\mathcal{S}-\mathcal{P}_k$.

Feature extraction and matching procedures are applied to the probe set $\mathcal{P}_k$ and gallery set $\mathcal{G}_k$. The number of probe images that are correctly identified as the identity of $\mathcal{T}_k$ is recorded as n_k. Their similarity scores are accumulated and recorded as sc_k. In addition, the similarity scores of all the probe images to the most similar gallery images in $\mathcal{G}_k-\mathcal{T}_k$ are accumulated and recorded as sf_k. This process is applied to all subsets of the q-combination of the elements in set $\mathcal{S}$, denoted by $\mathcal{T}_k$, $k=1,2,\ldots,\binom{N}{q}$.

The subset $\mathcal{T}_k$ with the maximum value of n_k is selected as the templates of this subject. If there are multiple subsets $\mathcal{T}_k$ having the same maximum value of n_k, the subset $\mathcal{T}_k$ with the maximum value of sc_k among them is selected. If there are still multiple subsets having the same maximum values of n_k and sc_k, the subset $\mathcal{T}_k$ with the maximum value of sf_k among them is selected. Although the probability that multiple subsets $\mathcal{T}_k$ have the same maximum values of n_k and sc_k is zero for $N>2$, this event likely occurs for $N=2$. Larger value of sf_k indicates that the subset $\mathcal{T}_k$ is more dissimilar to the images of other subjects than the other subsets $\mathcal{T}_i$, $i\neq k$. The number of templates q can be determined by the event that the maximum value of n_k for $q+1$ is not smaller than that for q.

2.2. Unstable keypoint removal

If we have multiple training images per subject, we can check the repeatability of a keypoint in different images of the same subject. A keypoint with low repeatability is unstable and hence can be removed. Take a training image I_t from the training image

set $\mathcal{S}$ of a subject and call it probe image and call all its keypoints probe keypoints in this section. The descriptor of a probe keypoint is compared with all the keypoints of the other subjects in the training set $\mathcal{D}-\mathcal{S}$ and the similarity of its nearest neighbor is denoted by s^b. The minimum similarity of all the probe keypoints to their nearest neighbors is denoted by s^b_m. Then, the descriptor of this probe keypoint is compared with all the keypoints of the other images of the same subject. Its similarity to its nearest neighbor in the kth image I_k is denoted by s^w_k, $I_k \in \mathcal{S}, k \neq t$. The repeatability γ of this keypoint is initialized as zero and accumulated over all images I_k, $I_k \in \mathcal{S}, k \neq t$, by

$$\begin{cases} \gamma \Leftarrow \gamma+2 & \text{if } s^w_k > s^b \\ \gamma \Leftarrow \gamma+1 & \text{if } s^b_m < s^w_k \leq s^b \end{cases} \quad \forall k : I_k \in \mathcal{S},\ k \neq t. \tag{1}$$

A keypoint of the image I_t with a larger value of γ has a higher repeatability. The rationale behind the two conditions and two different values added to γ can be seen in our two-stage image matching process in Section 4. We can set a threshold T_1 to select keypoints with high repeatability. If the value of γ of a keypoint is smaller than T_1, it will be removed.

As many keypoints may have the same value of γ, we cannot well control the keypoint removal to some desirable number. To make the selection process more flexible, we further propose to distinguish keypoints with the same value of γ by the discriminative value δ defined as

$$\delta = \frac{\sum_{I_k \in \mathcal{S}}^{k \neq t} s^w_k}{(N-1)s^b}. \tag{2}$$

If the number of the keypoints satisfying $\gamma \geq T_1$ is larger than a desirable n but that satisfying $\gamma \geq T_1+1$ is smaller than n, we remove the keypoints satisfying $\gamma < T_1$ and the keypoints satisfying $(T_1 \leq \gamma < T_1+1)$ & $(\delta < T_2)$. In this way, we can keep a desirable number of keypoints by varying T_1 and T_2. This process is visually shown in Fig. 1.

2.3. Template synthesis

In general, multiple templates per subject lead to better recognition accuracy because they can represent different expressions, poses and illumination conditions of a subject. However, multiple templates will greatly increase the computational complexity of the recognition process. And in some applications with limited computational power, single template per subject might be required. Although the template selection algorithm proposed in Section 2.1 can select the most representative template, the representation power of a single image is limited. A solution is template synthesis. Fig. 2 visually shows the process of template synthesis.

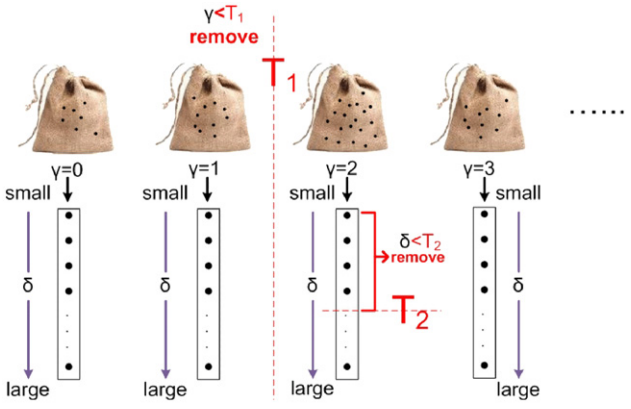

Fig. 1. An illustration of the process to remove keypoints with low repeatability.

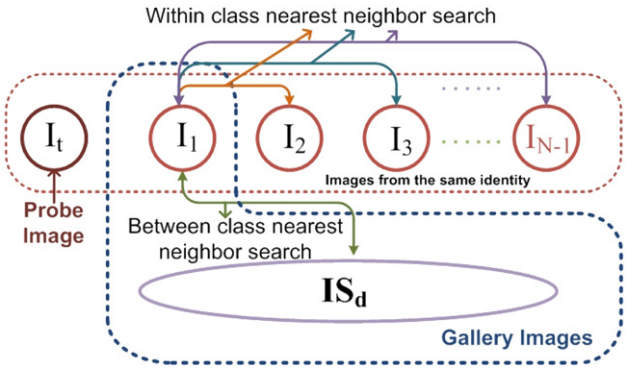

Fig. 2. An illustration of the process of template synthesis.

First, the most representative template of a subject denoted by I_t is selected by the algorithm proposed in Section 2.1. Second, this template is used as a probe image and another image I_k from the same subject and all images of the other subjects serve as gallery images. If the probe image is correctly identified by the algorithm proposed in Section 4, the unmatched keypoints in I_k are candidates to be integrated into the template. Third, the repeatability and the discriminative value of each candidate is computed by the algorithm proposed in Section 2.2 if there are more than two training images of this subject. The stable candidate keypoints will be integrated into the template I_t. If there are only two training images of this subject, all unmatched keypoints in I_k will be integrated into the template I_t. Last, the geometrical affine transform between I_k and I_t established in the image matching process based on Eq. (3) is used to transfer the stable candidate keypoints in I_k to the template I_t. This process is repeated for all images I_k, $k \neq t$ of the same subject.

However, in the experiments, we find out that the performance of the original SIFT approach for face recognition is not effective. After the training procedure to speed up the face recognition system, the performance drops further, which is far from satisfactory. There are two main factors in the original SIFT algorithm which deteriorate the recognition performance. First, some keypoints representing distinct structures are missing, which causes that the number of keypoints detected is not sufficient to capture the distinct information of different identities. Second, in the matching procedure, after the Hough transform the number of keypoints left decreases drastically. Face identification based on the remaining few keypoints causes high misidentification probability. Hence, we propose a new framework for the feature extraction and matching to solve the problems mentioned above.

3. Local structure detection and representation

A local structure is represented by a keypoint that locates it and a descriptor that represents its intensity variations in a local support area. The scale of a local structure or of a keypoint determines the size of the support area. Thus, the keypoint detection and its scale determination are the most critical parts of finding the image local structures.

3.1. Keypoint detection and scale selection

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. To find the scale of a possible keypoint, Lowe [24]

proposed to use Difference-of-Gaussian (DoG) of nearby scales at $k\sigma$ and σ to approximate the normalized Laplacian-of-Gaussian.

To detect the blob-like and corner-like structures and represent them at the optimal scales, Lowe in his SIFT framework proposes to compare each sample point to its eight neighbors in the current DoG image (spatial space) and nine neighbors in the scale above and below (scale space). A SIFT keypoint is selected only if it is larger than all these 26 neighbor points or smaller than all of them. This keypoint detection method works well for rigid visual objects, which have sharp transitions between different sides of an object. In other words, there are distinct corner or blob structures with high contrast in such objects. However, human faces are non-rigid, round and smooth. 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 extreme value in the 26 neighbors. To show an example of how keypoints are missing in the detection process proposed by Lowe, we plot a one-dimensional signal I as in Fig. 3(a), which is the sum of two Gaussian structures at scales $\sigma_i^1 = 3.0$ and $\sigma_i^2 = 4.1$, respectively. At least two keypoints near the two scales should be detected in the vicinity of the two Gaussian peaks. Fig. 3(b) shows the DoG outputs at six successive scales in the SIFT framework that well cover the two scales of the signal I: $\sigma_1 = 1.6 \times 2^{1/3} \approx 2.02$, $\sigma_2 = 1.6 \times 2^{2/3} \approx 2.54$, $\sigma_3 = 1.6 \times 2^{3/3} = 3.20$, $\sigma_4 = 1.6 \times 2^{4/3} \approx 4.03$, $\sigma_5 = 1.6 \times 2^{5/3} \approx 5.08$ and $\sigma_6 = 1.6 \times 2^{6/3} = 6.40$. The DoG outputs in the vicinity of the two Gaussian peaks are marked and the details are shown in Fig. 3(c) and (d). From Fig. 3(c), we see that the sample point 1 (in red) at scale $\sigma_3 = 3.20$ is the local minimum at the current scale compared with the points 2 and 3. Although it is smaller than the three neighbors in blue (sample points $1'$, $2'$ and $3'$) in the scale below, it is not smaller than the neighbor point $2''$ in the scale above (in black). Hence, the sample point 1 would not be detected as a local minimum. Similarly, in Fig. 3(d) no keypoint will be detected by Lowe's method. In a word, the original detection approach cannot detect any keypoint at the tested six scales in the vicinity of the two Gaussian peaks correctly. The missing detection will be even worse for the complex two-dimensional structures. This will decrease the number of keypoints detected and also deteriorate the following matching performance.

To fully detect the keypoints with distinct structures, we propose 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 neighbors or smaller than all of them. In other words, we compare a candidate point only to its 10 neighbors rather than 26 neighbors. Fig. 4 visually shows the proposed approach. Obviously, keypoints detected by Lowe's approach form a subset of the keypoints detected by our approach. We can see from Fig. 3 that the proposed approach will successfully detect two keypoints (the sample point 1 in Fig. 3(c) and the sample point 1 in Fig. 3(d)) of signal I in the vicinity of the two Gaussian peaks.

Another detection example is shown in Fig. 5, where the blue points denote the ones detected by Lowe's approach and the red points denote some extra keypoints detected by the proposed approach. As we see, Lowe's approach detects few keypoints from the forehead, cheeks, chin and facial contour which still provide

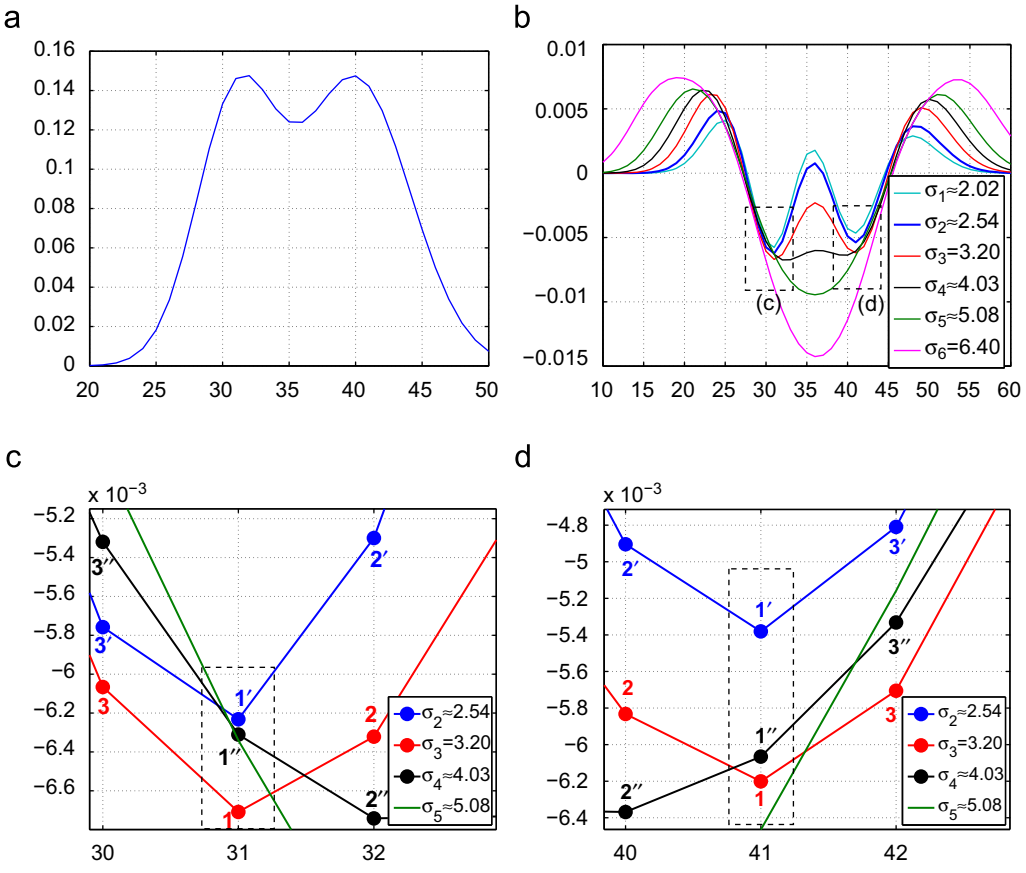

Fig. 3. Problem of the SIFT keypoint detection. (a) Input signal I. (b) DoG outputs at six different scales. (c,d) Local minima of DoG outputs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

useful information in telling different people although the change of these areas are subtle, while our approach can successfully detect these keypoints.

3.2. Descriptor representation

There are several descriptors proposed in the literature to represent the local image structures. Some efforts [43,44] were made to produce discriminative and low-dimensional descriptors learned from large databases. 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, keypoints near the face edge carry important information about the shape of the facial contour. However, these keypoints that represent the shapes of the facial contour are often of large scale and hence their support areas will exceed the image area, if the image is cropped tightly to the face size. To make use of these important keypoints, we introduce the partial descriptor. Readers can refer to our previous work [40,41] for details.

3.3. Insignificant keypoint removal

The above process will detect a rich number of keypoints. To reduce the matching time, Lowe [24] in his framework proposed to remove keypoints with low contrast and on the edges. This is an effective way to remove unstable keypoints detected from rigid visual objects where there are sharp transitions between different sides of an object. However, faces are non-rigid, round and smooth objects. There are few straight edges in face images. The intensity changes in face images are gradual and slow and hence the blob and corner structures are not significantly different from their neighboring pixels. Therefore, the keypoint removal scheme in the SIFT framework is inappropriate for face images. For instance, the blue points in Fig. 5(a) show the initially detected keypoints by Lowe's approach. Fig. 5(b) shows the remaining blue keypoints after applying a threshold on the minimum contrast of candidate keypoints. Fig. 5(c) shows the final remaining blue keypoints after further removing keypoints with high edge responses. The keypoints marked with ellipse and rectangle are removed, which do represent distinctive structures such as wrinkles and mouth corners.

Therefore, for face images, we propose to remove keypoints based on their distinctiveness from others. A significant keypoint should be distinct from others in terms of either its location or the image structures of its neighborhood. Thus, an initially detected keypoint is removed if and only if the spatial Euclidian distances from it to any other keypoint is smaller than a threshold t_e and the similarity between their descriptors is higher than a threshold t_c. This process removes insignificant keypoints and hence retains distinctive ones. Finally, as we will see in the next section, the final image matching is based on the similarities of keypoint descriptors and their relative geometrical locations.

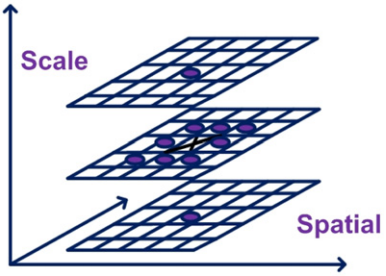

Fig. 4. Extrema of the DoG images are detected by comparing a pixel (marked with cross) to its eight neighbors at the current scale and the corresponding pixels at the adjacent scales (marked with circle).

4. Face identification

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 image. The image matching algorithm in Lowe's framework [24] contains two stages. First, the nearest neighbor in terms of the descriptor of every keypoint in the probe image is searched from all the gallery images. Gallery images that have at least three nearest neighbors are picked out as candidate images. Second, all nearest neighbors in a candidate image undergo a geometrical verification based on a set of affine transform parameters estimated by the clustered nearest neighbors in the Hough transform. The image matching score is then computed based on the number of nearest neighbors that coincide with the affine transform. This image matching algorithm works well for visual object detection where gallery contains few and quite dissimilar objects to be detected from a probe image that may contain several objects against a cluttered background. However, it is problematic if this algorithm is applied to an identification problem where the gallery contains a lot of similar objects.

Fig. 6 shows a wrong case when applying the matching algorithm of the SIFT framework [24] to the face identification. Fig. 6(a) shows the candidate images that contain at least three nearest neighbors to the probe keypoints. Geometric verification

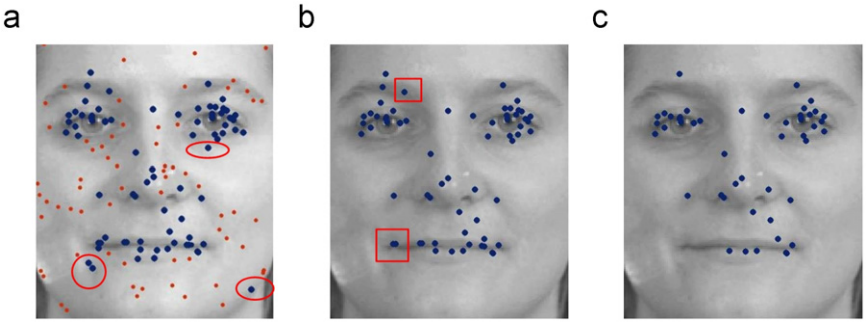

Fig. 5. (a) Initial keypoints detected by the original method (in blue) and extra keypoints detected by the proposed approach (in red). (b) Remaining blue keypoints after keypoint removal by low contrast. (c) Remaining blue keypoints after further keypoint removal based on the ratio of principal curvatures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

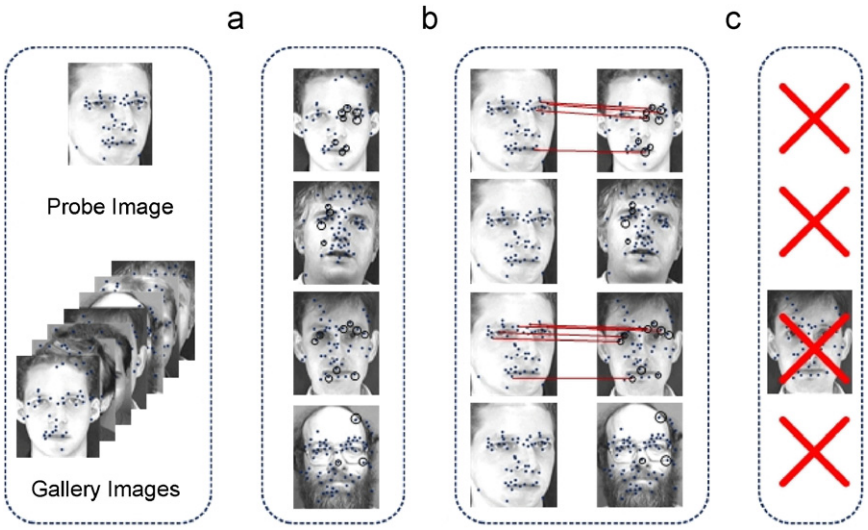

Fig. 6. The process of keypoints matching strategy proposed by Lowe: the number of keypoints matched in the right image is 4 and the number of keypoints matched in the wrong image is 5. (a) Nearest neighbor search. (b) Geometric verification. (c) Decision.

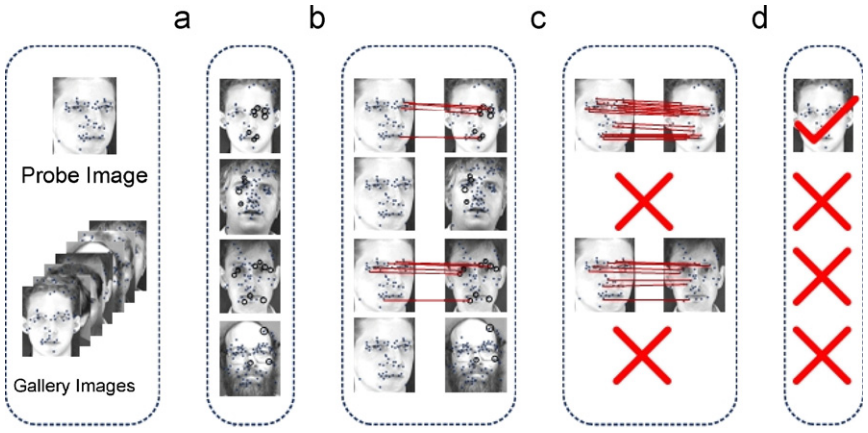

Fig. 7. The process of keypoints matching strategy proposed by us: the number of keypoints matched in the right image is 16 and the number of keypoints matched in the wrong image is 7. (a) Nearest neighbor search. (b) Geometric verification. (c) Individual image matching. (d) Decision.

is then performed between the probe keypoints and their nearest neighbors in each candidate image as shown in Fig. 6(b). Only those matches (linked by straight lines) with coherent relative locations are kept. The final decision is based on the number of the matched keypoints, which is apparently wrong since the third candidate image has more matched keypoints than the correct (the first) candidate.

The problem occurs at the very first stage. From Fig. 6, we can see that only the keypoints that fulfill the nearest neighbor condition are considered for further processing. 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 alleviate this problem, we propose a new matching framework for face recognition as shown in Fig. 7, which has the following new features:

1. To find a set of candidate gallery images, we search the *k-nearest* neighbors of the *nearest subject*, where k is determined by the second nearest subject.
2. A second matching process is introduced into the framework that *re-matches* the probe keypoints and keypoints in *each individual candidate gallery image* to augment the matched keypoint set.
3. Compute the similarity scores between the probe image and the candidate gallery images based on the *accumulated keypoint similarities*.

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

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 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. Note that we do not adopt the match rejection mechanism in the SIFT framework that is based on the similarity ratio between the first and the second nearest neighbors because it is not appropriate for the identification task. Fig. 7(a) shows the k-nearest neighbors (marked by circle) found in the gallery.

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 [24]. Fig. 7(b) shows the keypoint pairs used to calculate the affine transformation parameters. Note that some k-nearest neighbors are rejected by this process due to their geometric inconsistency.

4.2. 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

$$d = \sqrt{(x_g - x'_p)^2 + (y_g - y'_p)^2}. \tag{4}$$

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 satisfying the above conditions, the probe keypoint is not matched. The matched keypoint pairs are linked by straight lines in Fig. 7(c). It shows a significant increase of the matched keypoints of the correct gallery image from Fig. 6(b) or Fig. 7(b) while the number of the matched keypoints of the incorrect gallery image remains almost the same. We see that the proposed matching algorithm greatly increases the discriminative power of the image local structures.

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

$$s_j = \begin{cases} \max\limits_{i \in \mathcal{I}_j}(s_{ij}) & \text{if } \mathcal{I}_j \neq \emptyset, \\ 0 & \text{if } \mathcal{I}_j = \emptyset, \end{cases} \tag{5}$$

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:

$$S_{pg} = \sum_{j=1}^{q} s_j, \tag{6}$$

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}.

5. Experimental results

We shall first validate each ingredient of the proposed feature extraction and matching framework for face recognition by comparing it with the counterpart of the SIFT framework and some holistic methods, PCA [2], LDA [3] and ERE [14], one of the state-of-the-art holistic approaches to the face recognition, on FERET database 1, FERET database 2 and ORL database in Section 5.1. Then, in Section 5.2 we evaluate the efficacy of the unstable keypoints removal approaches proposed in Sections 3.3 and 2.2. Finally, performances of the training procedure for multiple samples per subject proposed in Section 2 based on our feature extraction and matching framework are compared with that based on the original SIFT algorithm on ORL, Georgia Tech (GT) and AR databases in Sections 5.3 and 5.4.

While images are preprocessed and normalized for the holistic approaches following the CSU face identification evaluation system with manually detected two eye coordinates, this preprocessing and normalization procedure is not applied for the multi-scale local structure based approaches. Fig. 8 shows some image samples used by holistic methods and samples used by the proposed and the SIFT frameworks.

5.1. Validation of feature detection and image matching

5.1.1. Results on FERET database 1

There are 2388 images comprising 1194 persons (two images Fa/Fb per person) selected from the FERET database [45]. Images are cropped into the size of 65×75 pixels. In the first experiment, images of 250 people are randomly selected for training, and the remaining images of 944 people are used for testing (FERET 1a). As

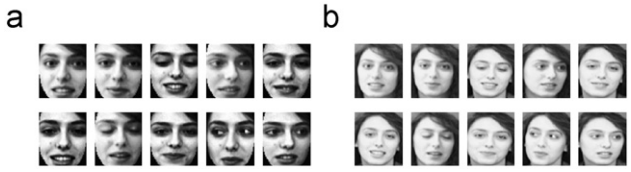

Fig. 8. Images of a sample subject used in (a) holistic approaches and (b) multi-scale local structure based approaches (the proposed and SIFT frameworks).

in the SIFT framework and our proposed approach, there are no training process, so we only use the images of 944 people for the experiments. In the second experiment, more training samples (497 people) are randomly selected, and the remaining images of 697 people are used for testing (FERET 1b). Table 1 gives the rank one recognition rates of the holistic approaches PCA [2], LDA [3] and ERE [14], together with the SIFT framework in [24], the proposed feature detection with the SIFT matching approach (PFD-SIFTM) and the proposed feature detection and matching framework (PFDM).

From this table, we can see that the recognition rates of holistic approaches PCA, LDA and ERE increase with the increase of the number of training images (from 500 to 994). Comparing the performance of SIFT and PFD-SIFTM, we can see that the keypoints detected by our approach can capture more distinct information than Lowe's keypoints. And the proposed feature detection and matching framework (PFDM) can further improve the recognition performance. Fig. 9 shows two cumulative matching curves [46] of PCA, LDA, ERE, SIFT, PFD-SIFTM and PFDM on FERET database 1a and FERET database 1b. From this two figures, we can see that our proposed feature detection and matching framework achieves significantly better performance than the original SIFT approach over all ranks.

5.1.2. Results on FERET database 2

This database is constructed, same to one data set used in [14], by choosing 256 subjects with at least four images per subject. And we use the same number of images (four) per subject for all subjects. The first 512 images of the first 128 subjects are used for training, and the remaining 512 images serve as testing images. In the SIFT framework and our proposed approach, we only use the last 512 images for testing as there are no training process. The size of the image is 130×150 pixels, same as that in [14]. The ith ($i = 1,2,3,4$) images of all the testing subjects are chosen to form a gallery set, and the remaining three images per subject serve as the probe images to be identified from the gallery set. Table 2 shows the average recognition rates (ARR) and corresponding standard deviations (Std) over the four probe sets, each of which has a distinct gallery set.

The recognition rates are lower than those obtained in Section 5.1.1 due to the increase of variations of probe images. However, our proposed approach still outperforms the holistic approaches PCA, LDA and ERE. And the higher recognition rates achieved by PFD-SIFTM and PFDM compared with SIFT, show that our keypoint detection and matching strategies are more appropriate for face recognition task than Lowe's SIFT framework.

5.1.3. Results on ORL database

The ORL database contains 400 images from 40 subjects taken at different times, varying the lighting, facial expressions and facial details. Images of the ORL database are resized into 50×57 pixels. Each subject has 10 images with index i from 1 to 10. The images from each subject with the same index i are picked out to form the gallery set and the remaining 360 images serve as the probe set. The rank one recognition rate is computed by the number of correctly identified probe images over 360. Ten runs of experiments are performed where each run has distinct gallery images. The average recognition rate (*ARR*) and its standard deviation (Std) over the 10 runs are recorded as indications of the recognition performance.

Table 3 gives the recognition rates of the SIFT framework in [24], the proposed feature detection with the SIFT matching approach (PFD-SIFTM) and the proposed feature detection and matching framework (PFDM). No keypoint is removed in this experiment.

The recognition rate is low because the large variations of the face pose and expression in this database cannot be well represented by a single template per subject. Table 3 demonstrates that the proposed feature detection and matching approaches significantly outperform the counterparts in the SIFT framework. The large difference between the average recognition rates relative to their standard deviations shows the statistical significance of the experimental results.

Table 1
Rank one recognition rate on FERET database 1.

Database	PCA (%)	LDA (%)	ERE (%)	SIFT (%)	PFD-SIFTM (%)	PFDM (%)
FERET 1a	83.16	89.72	94.81	93.33	97.67	97.88
FERET 1b	85.8	96.41	97.13	94.41	98.42	98.71

Table 2
Average recognition rate and its standard deviation with single template per subject on FERET database 2.

	PCA (%)	LDA (%)	ERE (%)	SIFT (%)	PFD-SIFTM (%)	PFDM (%)
ARR	74.41	80.79	83.07	81.58	90.17	92.06
Std	2.03	1.54	1.43	0.89	0.75	1.37

Table 3
Average recognition rate and its standard deviation with single template per subject on ORL database.

	SIFT (%)	PFD-SIFTM (%)	PFDM (%)
ARR	57.84	77.22	81.56
Std	2.52	1.64	1.85

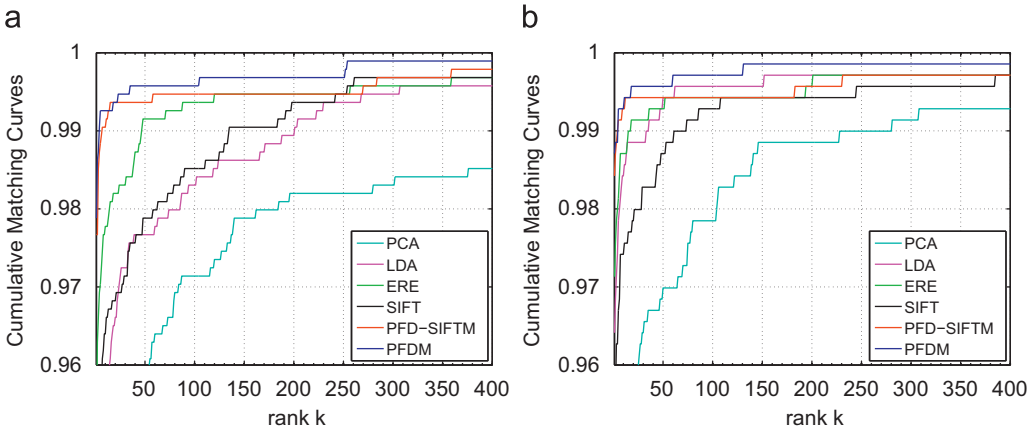

Fig. 9. (a) Cumulative matching curves on FERET 1a. (b) Cumulative matching curves on FERET 1b. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Validation of keypoint removal

To validate the method of the insignificant keypoint removal proposed in Section 3.3, it is compared with the keypoint removal approach in the SIFT framework that is based on the low contrast and high edge response. Experiment setting is the same as that in Section 5.1.3. Fig. 10 shows the rank one recognition rates against the average number of remaining keypoints after these two keypoint removal approaches, respectively. Note that the proposed keypoint detection and image matching scheme is applied in these experiments. Fig. 10 demonstrates that the proposed keypoint removal method consistently outperforms that of the SIFT framework at different amount of the removed keypoints. However, this experiment shows that the gain in recognition accuracy by the keypoint removal is insignificant. Removing the keypoints more than 20% of the initially detected reduces the recognition accuracy. Therefore, the keypoint removal should not target at improving the recognition accuracy but at reducing the computational complexity of the recognition process.

In Section 2.2, we propose an approach to remove unstable keypoints if multiple training images per subject are available. To test the effect of this approach, we choose the first five samples per person of the ORL database as the training set, and the last five samples per person as the probe set. One template per subject is selected from the training images using the template selection method proposed in Section 2.1 and the keypoints of the selected template are removed based on the method proposed in Section 2.2 with the help of the other training images. Table 4 gives the average number of the remaining keypoints (*ANo.*) per template and the recognition rate (*RR*) using different threshold T, $T=T_1+T_2$. It shows again that the gain in recognition accuracy by the keypoint removal is insignificant. Removing the keypoints by this method more than 30% reduces the recognition accuracy. Therefore, the keypoint removal by this method should also not target at improving the recognition accuracy but at reducing the computational complexity of the recognition process.

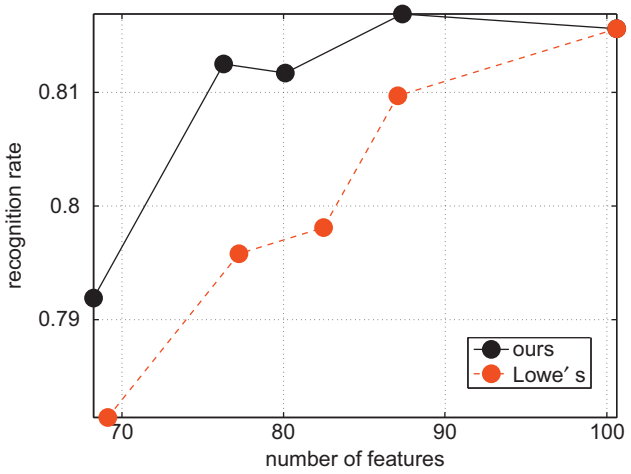

Fig. 10. Rank one recognition rate of the two keypoint removal approaches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Average number of the remaining keypoints per template and recognition rate using different threshold T.

T	0	1	2	2.5	3	3.5	4	5
ANo.	101	93	86	81	75	69	63	36
RR%	85	85.5	86	86	85.5	84.5	83	80

5.3. Validation of single template selection and synthesis

To test the effect of the proposed template selection approach against the random template and the effect of the template synthesis, we choose the first five samples per person of the ORL database as the training images, and the last five samples per person as the probe images. We first pick out one template per subject randomly from the training set to form a gallery set and use it and the probe set to test the recognition accuracy. This random template picking-up is repeated 20 times and the average recognition rate and its standard deviation are computed. Then, one template per subject is selected from the training images using the template selection method proposed in Section 2.1. Recognition rate on the gallery formed by the selected templates and the probe set is tested. Finally, some stable keypoints of the remaining training images are integrated into the selected template based on the template synthesis approach proposed in Section 2.3. The synthesized template is basically a single template, which has 20% more keypoints on average than that before synthesis. Table 5 shows the recognition rates (ARR or RR). It demonstrates the contribution of the template selection and synthesis to the recognition accuracy. However, the single template selected and synthesized cannot remarkably increase the recognition rate because the representation power of a single template for lighting, pose and expression variations is limited.

5.4. Recognition performances of multiple templates per subject

Three databases, ORL, GT and AR are applied to test the face recognition performance with multiple templates per subject. For ORL database, the first five samples per subject form the training and gallery sets, and the remaining five samples serve as the probe images. The Georgia Tech (GT) Face Database contains 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 resized into 60×80 pixels. Similarly, the GT database is partitioned into the training or gallery set consisting of the first eight samples per subject and the probe set consisting of the remaining seven samples. The color images in AR database are converted to gray-scale and cropped into the size of 120×170 pixels. In the experiments, 75 persons with 14 non-occluded images per person are selected, which makes the database containing 1050 images. The first seven images per subject serve as training and gallery images and the remaining seven images as probe images. The best recognition performances of the holistic approaches PCA [2], LDA [3] and ERE [14] are recorded. The SIFT framework with default parameter in [24] and the proposed feature detection and matching framework (PFDM) are applied on the three databases with and without the proposed template selection. Table 6 shows the rank one recognition rates on ORL, GT and AR databases. The second row under the name of the database shows the number of templates per subject selected by the approach proposed in Section 2.1.

On the ORL database, the multiple templates per subject greatly enhance the recognition accuracy for both the SIFT and the

Table 5
Recognition rate of single template on ORL database.

Template	Random	Random	Selected	Synthesized
Algorithm	SIFT	PFDM	PFDM	PFDM
ARR/RR	59.9%	80.6%	85%	89%
Std	4.60%	2.48%	N.A.	N.A.

Table 6
Recognition rate on ORL, GT and AR databases.

ORL database						
5	5	5	5	5	4	4
PCA	LDA	ERE	SIFT	PFDM	SIFT	PFDM
85.5%	92.5%	97%	90%	99.5%	89%	98.5%
GT database						
8	8	8	8	8	7	7
PCA	LDA	ERE	SIFT	PFDM	SIFT	PFDM
80.57%	90.71%	92.86%	80.57%	95.71%	78.6%	94.57%
AR database						
7	7	7	7	7	6	6
PCA	LDA	ERE	SIFT	PFDM	SIFT	PFDM
93.52%	94.1%	95.43%	97.14%	99.81%	96.77%	99.43%

proposed approach. While the recognition rate of SIFT increases from 59.9% to 90%, the proposed approach achieves 99.5% recognition rate with five templates per subject, which is better than the holistic approach ERE. The template selection of only four out of the five slightly decreases the recognition rate. Note that the results of template selection based on our proposed framework significantly outperform that based on the original SIFT algorithm. On the GT database, the proposed approach achieves a higher recognition rate than SIFT and the holistic approaches on eight templates per subject. On the AR database, the proposed approach achieves a remarkably high recognition rate of 99.81% comparing to 95.43% of ERE and 97.14% of SIFT. Reducing the number of the templates per subject from seven to six by the proposed template selection algorithm worsens the recognition performance of the proposed approach a little. Table 6 shows that the deployment of multiple templates of a subject is an effective way to circumvent the problems caused by various intraclass variations of pose, expression and illumination. It seems that more templates of a subject than necessary cause no harm to the recognition accuracy. This might be attributed to the proposed scheme that searches the k-nearest neighbors of the nearest subject.

6. Conclusion

Face recognition based on the multi-scale local features has the potential to be more robust to variations in pose, scale, expression and occlusion than the holistic approaches. However, local feature based methods are far more complex in both feature extraction and matching/classification compared with the holistic approaches that are mainly based on the machine learning. The difficult face recognition task cannot be well accomplished based on just one bright idea. Rather, a sophisticated system with well-designed individual components is required to cope with this challenging problem. This paper presents a face recognition framework based on the multi-scale local image features with scale selection. While some basic tools such as DoG filter, HoG descriptor and Hough transform are inherited from the SIFT framework, this work investigates and contributes to all major steps in the feature extraction and matching for the challenging face recognition task.

The SIFT framework is designed and works well for general object recognition. However, it is not optimal for the face recognition because face is a non-rigid, round and smooth object. Its intensity has slow and gradual change but the local structure may be complex due to the spatial overlap of different local structures. The proposed keypoint detection, partial descriptor and insignificant point removal extract and retain more useful local structure features for the face recognition compared with the counterparts in the SIFT framework.

The nearest neighbor search of the local structures suffers the problem that the nearest neighbors of different probe keypoints disperse into different templates of the correct subject. The proposed search of the k-nearest neighbors of the nearest subject solves this problem. Due to the inevitable intraclass variation, we cannot expect that a probe local structure is always nearest to the corresponding one of the same subject among all structures of all subjects in the gallery. The proposed second matching stage matches the local structures of the probe image individually with each candidate gallery image. This solves the problem caused by the large amount of subjects in the gallery.

The deployment of multiple templates per subject is a solution to the large variations of the face pose and expression. This, however, imposes great computational burden on the recognition process, and hence we cannot arbitrarily increase the number of templates of a subject in the gallery. In addition, a face database collected in practice often contains some very similar or even identical images, which need be trimmed off. Different training schemes including template selection, unstable keypoint removal and template synthesis are proposed to meet different requirements in the face recognition applications.

References

[1] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (2003) 399–458.
[2] M. Kirby, L. Sirovich, Application of Karhunen–Loeve procedure for the characterization of human faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 103–108.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 711–720.
[4] B. Moghaddam, T. Jebara, A. Pentland, Bayesian face recognition, Pattern Recognition 33 (2000) 1771–1782.
[5] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, Pattern Recognition Letters 26 (2005) 181–191.
[6] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana, Discriminative common vectors for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 4–13.
[7] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, Z. Jin, Kpca plus lda: a complete kernel Fisher discriminant framework for feature extraction and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 230–244.
[8] W. Zheng, X. Tang, Fast algorithm for updating the discriminant vectors of dual-space lda, IEEE Transactions on Information Forensics and Security 4 (2009) 418–427.
[9] X. Wang, X. Tang, A unified framework for subspace face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 1222–1228.
[10] J. Ye, R. Janardan, C. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 982–994.
[11] X. He, S. Yan, Y. Hu, P. Niyogi, H.J. Zhang, Face recognition using laplacian faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 328–340.
[12] S. Yan, D. Xu, B. Zhang, Q. Yang, H. Zhang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007) 40–51.
[13] S. Ji, J. Ye, Generalized linear discriminant analysis: a unified framework and efficient model selection, IEEE Transactions on Neural Networks 19 (2008) 1768–1782.
[14] X.D. Jiang, B. Mandal, A. Kot, Eigenfeature regularization and extraction in face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2008) 383–394.
[15] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 210–227.
[16] X.D. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 931–937.
[17] X.D. Jiang, Linear subspace learning based dimensionality reduction, IEEE Signal Processing Magazine 28 (2) (2011) 16–26.
[18] T. Riopka, T. Boult, The eyes have it, in: Proceedings of the 2003 ACM SIGMM Workshop on Biometrics Methods and Applications, pp. 9–16.
[19] X. Tan, S. Chen, Z. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recognition 39 (2006) 1725–1745.

[20] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 775–779.
[21] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 681–685.
[22] J. Wright, G. Hua, Implicit elastic matching with random projections for pose-variant face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[23] D.G. Lowe, Object recognition from local scale-invariant features, in: IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157.
[24] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.
[25] D.G. Lowe, Local feature view clustering for 3d object recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 682–688.
[26] M. Brown, D.G. Lowe, Invariant features from interest point groups, in: British Machine Vision Conference, pp. 656–665.
[27] M. Brown, D.G. Lowe, Recognising panoramas, in: IEEE International Conference on Computer Vision, vol. 2, 2003, pp. 1218–1225.
[28] K. Yan, R. Sukthankar, Pca-sift: a more distinctive representation for local image descriptors, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 506–513.
[29] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 1615–1630.
[30] M. Bicego, A. Lagorio, E. Grosso, M. Tistarelli, On the use of sift features for face authentication, in: Workshop on Computer Vision and Pattern Recognition, 2006, pp. 35–40.
[31] D.H. Lin, X.O. Tang, Recognize high resolution faces: from macrocosm to microcosm, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 1355–1362.
[32] D.R. Kisku, A. Rattani, E. Grosso, M. Tistarelli, Face identification by sift-based complete graph topology, in: IEEE Workshop on Automatic Identification Advanced Technologies, 2007, pp. 63–68.
[33] C. Rosenberger, L. Brun, Similarity-based matching for face authentication, in: IEEE International Conference on Pattern Recognition, 2008, pp. 1–4.
[34] D.R. Kisku, A. Rattani, M. Tistarelli, P. Gupta, Graph application on face for personal authentication and recognition, in: International Conference on Control, Automation, Robotics and Vision, 2008, pp. 1150–1155.
[35] L.C. Zhang, J. Chen, Y. Lu, P. Wang, Face recognition using scale invariant feature transform and support vector machine, in: International Conference for Young Computer Scientists, 2008, pp. 1766–1770.
[36] C. Fernandez, M.A. Vicente, Face recognition using multiple interest point detectors and sift descriptors, in: IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–7.
[37] C. Cruz, L.E. Sucar, E.F. Morales, Real-time face recognition for human–robot interaction, in: IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–6.
[38] Y.B. Han, J.Q. Yin, J.P. Li, Human face feature extraction and recognition base on sift, in: International Symposium on Computer Science and Computational Technology, vol. 1, 2008, pp. 719–722.
[39] D.R. Kisku, M. Tistarelli, J.K. Sing, P. Gupta, Face recognition by fusion of local and global matching scores using ds theory: an evaluation with uni-classifier and multi-classifier paradigm, in: IEEE Workshop on Computer Vision and Pattern Recognition, 2009, pp. 60–65.
[40] C. Geng, X.D. Jiang, Face recognition using sift features, in: IEEE International Conference on Image Processing, 2009, pp. 3313–3316.
[41] C. Geng, X.D. Jiang, Sift features for face recognition, in: IEEE International Conference on Computer Science and Information Technology, 2009, pp. 598–602.
[42] A. Majumdar, R.K. Ward, Discriminative sift features for face recognition, in: Canadian Conference on Electrical and Computer Engineering, 2009, pp. 27–30.
[43] G. Hua, M. Brown, S. Winder, Discriminant embedding for local image descriptors, in: IEEE International Conference on Computer Vision, 2007.
[44] M. Brown, G. Hua, S. Winder, Discriminative learning of local image descriptors, IEEE Transaction on Pattern Analysis and Machine Intelligence 33 (2011) 43–57.
[45] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The feret evaluation methodology for face recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 1090–1104.
[46] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha, A.W. Senior, The relation between the roc curve and the cmc, in: IEEE Workshop on Automatic Identification Advanced Technologies, 2005, pp. 15–20.

Cong Geng received the B.Sci. degree in electrical and electronic engineering from the Wuhan University, China, in 2006. Since 2007, she has been a Ph.D. candidate in the Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her research interests include pattern recognition, computer vision, image processing and biometrics. She is working on feature extraction and matching using local 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 and electronic 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 2002, he was with Nanyang Technological University (NTU), Singapore, as a senior research fellow, where he developed a system that achieved the most efficient and the second most accurate fingerprint verification at the International Fingerprint Verification Competition (FVC'00). From 2002 to 2004, he was a lead scientist and the head of the Biometrics Laboratory at the Institute for Infocomm Research, Singapore. He joined NTU as a faculty member in 2004. Currently, he is a tenured associate professor and serves as the director of the Center for Information Security, the School of Electrical and Electronic Engineering, NUT, Singapore. His research interest includes pattern recognition, image processing, computer vision and biometrics. He is a senior member of the IEEE.