Pattern Recognition 48 (2015) 732–745

Contents lists available at ScienceDirect

Pattern Recognition

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

Design and learn distinctive features from pore-scale facial keypoints

Dong Li, Kin-Man Lam*

Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

ARTICLE INFO

Article history:
Received 31 May 2014
Received in revised form
26 August 2014
Accepted 27 September 2014
Available online 16 October 2014

Keywords:
Pore-scale facial feature
Feature extraction
Face matching
Face reconstruction
Face recognition

ABSTRACT

Establishing correct correspondences between two faces with different viewpoints has played an important role in 3D face reconstruction and other computer-vision applications. Usually, face images are considered to lack sufficient distinctive features to establish a large number of correspondences on uncalibrated images. In this paper, we investigate pore-scale facial features, which are formed from pores, fine wrinkles, and hair. These features have many characteristics that make them suitable for matching facial images under different variations. Using both biological observation and computer-vision consideration, a new framework is devised for pore-scale facial-feature extraction and matching. The matching difficulty under various skin appearances of different subjects and imaging distortion is also analyzed. For further improving the matching performance and tackling distortions such as varying illuminations and unfocused blurring, a pore-to-pore correspondences dataset is established for training a more distinctive and compact descriptor. Experiments are conducted on a face database containing 105 subjects, and the results prove that the pore-scale features are highly distinctive; face images with a minimum resolution of 600 × 700 (0.4 mega) pixels contain sufficient details to perform a reliable matching in different poses. Generally, our algorithm can establish between 500 and 2000 correct correspondences on a pair of uncalibrated face images of the same person. Furthermore, the proposed methods can be applied to face recognition, 3D reconstruction, etc.

1. Introduction

Do two people have identical skin blocks on their faces? From a biological point of view, definitely not. It is a very challenging problem to match two facial-skin blocks from the computer-vision viewpoint, because of the requirements of finding accurate keypoint locations and extracting distinctive features in skin images. In general, face images are considered to lack sufficient distinctive features for tracking their geometry or for 2D direction feature matching on uncalibrated images. Hence, a lot of the literature requires the involvement of other man-made features, such - as structured lighting, special makeup, and markers; or it is based on stereo matching (1D direction feature matching) using calibrated images.

To the best of our knowledge, only a few studies have been reported in the literature that attempt to establish correspondences using uncalibrated face images. Lin et al. [1] employed the SURF features [2] on face images with viewpoints 45° apart (the face regions are of about 800 × 600 pixels), which typically obtained no more than 10 inliers (i.e. correctly matched keypoint pairs) out of a total of 30 matched candidates in 3 views (-45°, 0°, 45° from the frontal view). Thus, the classical structure-from-motion method with known camera intrinsic parameters, which extracts the correspondences of 3 views by employing RANSAC [3] on top of the PnP [4] algorithms, fails due to both the scarcity of absolute inliers and the small ratio of inliers to outliers detected.

Detecting and analyzing facial features is a fundamental task in computer vision, and it is also vital to face detection, pose estimation, landmark localization, and face recognition. Although great strides have been made in this area, it is still challenging to detect and obtain distinctive features from face images, especially on the pore scale. Therefore, rather than giving a historic review of the methods for facial-feature extraction and detection, we will categorize the methods into three different levels: primary facial features, marker-scale facial features, and pore-scale facial features.

Primary facial features include the eyes, eyebrows, nose, mouth, and the face boundary. There are many studies reported on face alignment [5,6], face recognition [7,8], and facial-organ detection [9,10]. Primary facial features are the easiest to detect, but they are difficult to locate precisely.

The definition of marker-scale facial features provided by [11] involves ten categories, such as freckle, mole, scar, and wrinkle. Park and Jain [12] proposed an automatic facial-mark (e.g. scars,

* Corresponding author.
E-mail addresses: dong.li@connect.polyu.hk (D. Li), enkmlam@polyu.edu.hk (K.-M. Lam).

http://dx.doi.org/10.1016/j.patcog.2014.09.026
0031-3203/© 2014 Elsevier Ltd. All rights reserved.

moles, and freckles) detection method. First, the active appearance model (AAM) is employed to detect and remove the eyes, eyebrows, nose, mouth, and face boundary. Then, the LoG blob detection is applied to detect the facial marks. With a database of 1225 images of 671 distinct subjects, experiment results showed that 90% of the subjects in the database have fewer than 15 marker-scale facial features, while the overall average is about 7. Thus, the number of marker-scale facial features is insufficient to establish the correspondences needed in real-world applications.

Pore-scale facial features include pores, fine wrinkles, and hair, which commonly appear in the whole face region. Due to small concavities where the incoming light is blocked, the pores are small, darker points, while the wrinkles are in the form of line structures. The hair also appears as small, darker points, and as lines. Lin and Tang [13] considered pore-scale facial features as repetitive texture units. A set of response vectors is extracted from skin regions using Gabor filters. Then, the skin texton distribution is used to represent skin textures. Cula et al. [14] developed two bidirectional texture models for use in skin texture recognition. Their research also considers the pore-scale facial features as textures, rather than identifying the pores in facial images.

To the best of our knowledge, there is no literature using the pore-scale facial features to perform keypoint detection and to establish precise correspondences on uncalibrated images. This is a challenging and difficult task because, intuitively, pore-scale facial features such as pores are similar to each other, so they are not distinctive. In this paper, we will solve the potential difficulties in extracting pore features from facial-skin images and analyze the relationship between the facial-skin appearance/image condition and the matching results. The proposed pore-scale feature-extraction framework is shown in Fig. 1; it mainly consists of quantity-driven detection, relative-position description, and candidate-constrained matching of pore-scale keypoints. Each of these steps is designed to identify pore-scale facial keypoints according to both biological observation and computer-vision consideration.

First of all, from a biological point of view, the quantity of pores on different faces should be similar even for people who have very different skin appearances; but the level of difficulty in detecting the pores varies. For pore-scale keypoint detection, a quantity-driven DoG detector is proposed in our algorithm. We use a Gaussian kernel to model the blob-shaped, pore-scale features so as to determine the number of DoG octaves for the detection. Unlike the general keypoint-detection methods, which use a constant threshold to detect keypoints, our method uses an adaptive threshold to extract a certain number of keypoints on a skin region. Furthermore, the DoGs sampling frequency is determined by the quantity of inliers in skin-matching experiments. In particular, the quantity of inliers and the repeatability of the keypoints are unified in our framework.

Based on the peak response of the modeled pore-scale feature on the DoG layers, the adaptive threshold is normalized as a new measure, namely the *Pore Index*, to characterize the skin appearance of different subjects. Thus, the Pore Index can also be used to analyze the difficulty level of matching faces under different image distortions and skin appearances. The pore-scale feature modeling and the Pore Index can provide us with a better understanding of the properties of pore-scale facial features.

We propose using relative-position information a description of the pore-scale facial features. For simplicity and convenience, the state-of-the-art feature-extraction method “Scale-Invariant Feature Transform (SIFT)” [15] is adapted to extract the relative-position information around a keypoint; this method is called Pore-SIFT (PSIFT). In order to include the information about the relative position of the pore-scale keypoints, a larger neighborhood size is needed than is the case for SIFT.

With this pore-scale feature-extraction framework, we can successfully establish a large number of correspondences on uncalibrated face images. To the best of our knowledge, no existing methods can establish a similar number of matched keypoints from two uncalibrated skin images of the same subject. In our experiments, we will study the influence of each stage of our proposed algorithm on the keypoint-matching performances. Based on the results, we can conclude that a fine-scale sampling rate, a fine pre-smoothing, an adaptive thresholding, and a relative-position descriptor are all important for achieving accurate results. The feature used for pore-scale keypoints should be robust to distortions such as noises, unfocused blurring, and reflectance. However, such distortions are hard to model or to be considered in the design of a feature-extraction framework. Hence, to further improve our algorithm, we establish a pore-to-pore correspondences dataset for learning a discriminant subspace in order to tackle the distortions and to achieve a more accurate pore-matching performance. In our algorithm, after projecting the PSIFT features onto a linear discriminant analysis (LDA) subspace, those intra-pore variations will be minimized, while the inter-pore distances will be maximized. Our PSIFT feature projected onto the LDA subspace is called LDAPSIFT, which can usually establish 30% more correspondences than PSIFT can. Both the LDAPSIFT and the PSIFT features are highly distinctive, and can be used to establish precise correspondences between uncalibrated face-image pairs.

With the prior knowledge of facial images, the use of a candidate-constrained matching scheme can significantly reduce the number of candidates considered in matching, so that the computational complexity of our algorithm can be reduced. Comprehensive experiments have also been conducted on all 105 subjects in a face database, in which the face images contain different types of distortions and variations. The experimental results will be analyzed and discussed comprehensively.

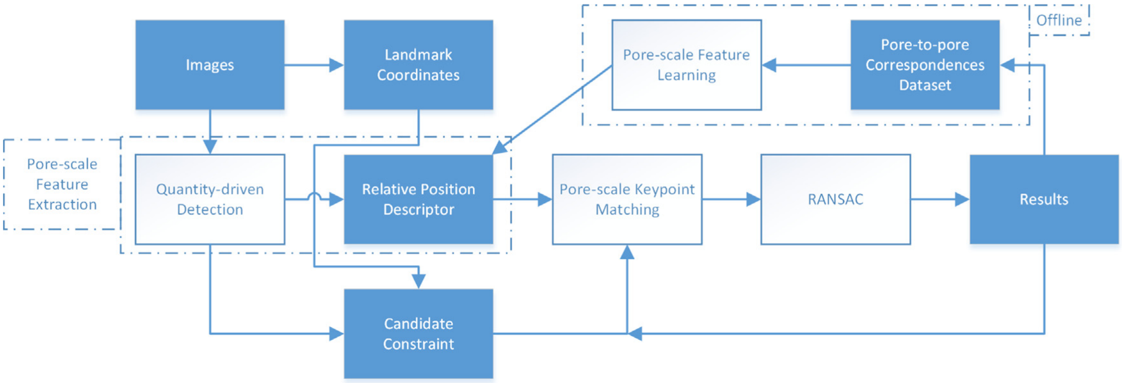
Offline
Pore-scale Feature Learning
Pore-to-pore Correspondences Dataset
Images
Landmark Coordinates
Pore-scale Feature Extraction
Quantity-driven Detection
Relative Position Descriptor
Pore-scale Keypoint Matching
RANSAC
Results
Candidate Constraint

Fig. 1. The structure of the proposed pore-scale facial-feature extraction and matching framework.

To the best of our knowledge, this is the first framework which is suitable for pore-scale facial-feature extraction and which can establish a large number of correspondences between uncalibrated face images. Our contributions are as follows:

- Our method identifies the pores in facial images, rather than considering pore-scale facial features as a kind of texture.
- We propose a new framework for pore-scale facial feature-extraction and description, and we describe how SIFT, a traditional local descriptor, can be adapted to form a pore-scale facial feature descriptor.
- Based on our framework, a pore-to-pore correspondences dataset containing 4240 classes of matched pore pairs is formed by the same pore keypoints from 4 face images of the same person with different poses. These 4 keypoints form a track, and can produce 6 keypoint matched pairs for each class. In other words, we have 6×4240 matched pore pairs in the dataset.
- Based on this pore-to-pore dataset, a learning-based pore-scale facial feature, namely LDAPSIFT, is proposed, which is more distinctive and compact than PSIFT.
- Our proposed methods can establish a large number of correspondences between uncalibrated face images of the same subject using the pore-scale facial features, which leads to many potential applications. Our work shows a way to merge general computer-vision approaches and face-based approaches.

The rest of this paper is organized as follows: Section 2 describes a model for the pore-scale facial feature, and introduces the quantity-driven pore-scale facial-feature detection scheme and the *Pore Index*. Section 3 presents the pore-scale facial-feature descriptor based on the relative positions of keypoints, and a pore-to-pore dataset is constructed so that discriminant projections are learned which minimize the ratio of the intra-pore variations to the inter-pore differences. Section 4 introduces our proposed candidate-constrained matching scheme. Section 5 shows the experimental results for using the pore-scale facial features for matching. The statistics of the Pore Indices of 420 images are used to analyze the difficulty of face matching with different facial-skin appearances. Some potential applications are also discussed. Finally, Section 6 summarizes this paper and discusses some directions for our future work.

2. Quantity-driven detection and pore index

Pore-scale facial features – such as pores, fine wrinkles and hair – are darker than their surroundings in a skin region, and those features which are blob-shaped or endpoint/corner-shaped provide stable locations for matching purposes. Therefore, we apply the DoG detector for keypoint detection on multi-scales, which is given as follows:

$$\begin{aligned} D(x,y,\sigma) &= L(x,y,k\sigma) - L(x,y,\sigma) \\ &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y), \end{aligned} \tag{1}$$

where the scale space of an image $L(x,y,\sigma)$ is the convolution of the image $I(x,y)$ and the Gaussian kernel

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-(x^2+y^2)/2\sigma^2\right). \tag{2}$$

We construct the DoG in octaves, which have the σ doubled in the scale space. Each octave has N_s DoG layers, so the factor k in (1) is defined as

$$k = 2^{1/N_s}. \tag{3}$$

Unlike the general DoG detector, which localizes both the scale-space maxima and minima of the DoG, we detect only the maxima of the DoG to locate the darker keypoints in face regions. An example of the pore-scale facial-feature responses with DoG is shown in Fig. 2(c).

2.1. Pore-scale facial-feature modeling

A blob-shaped pore-scale keypoint is a small, darker point due to its small concavity, where incident light is likely blocked. There is no sharp or clear boundary around a pore keypoint. For a better understanding of such pore-scale facial features and the sampling frequency selected in the detection, we model the blob-shaped skin pores using a Gaussian function, as follows:

$$pore(x,y,\sigma) = 1 - 2\pi\sigma^2 G(x,y,\sigma), \tag{4}$$

where σ is the scale of the pore model. This models a skin pore with the gray-level intensities normalized to $[0,1]$, as illustrated in Fig. 3. Then, the DoG response of a pore, denoted as, D_{pore}, can be computed as follows:

$$\begin{aligned} &D_{pore}(x,y,\sigma_1,\sigma_2) \\ &\quad = [G(x,y,k\sigma_1) - G(x,y,\sigma_1)] * pore(x,y,\sigma_2) \end{aligned}$$

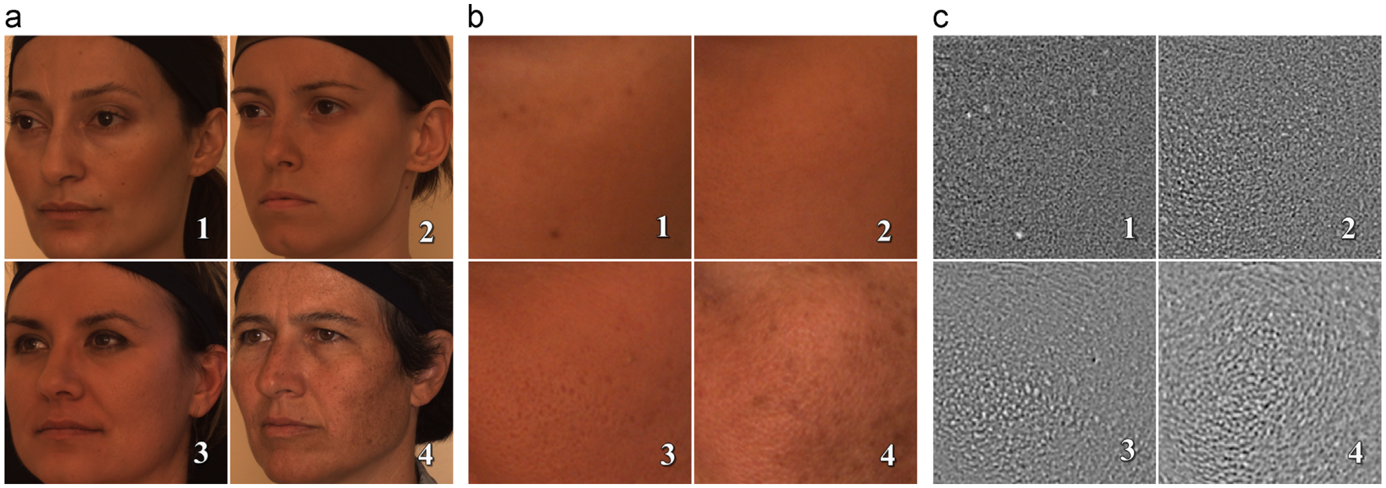

Fig. 2. (a) Four face images with different skin conditions from the Bosphorus face database, (b) zoomed-in local skin-texture images, and (c) the DoG of the zoomed-in local skin-texture images. The four images in (a) are named Subjects 1, 2, 3 and 4 in this paper.

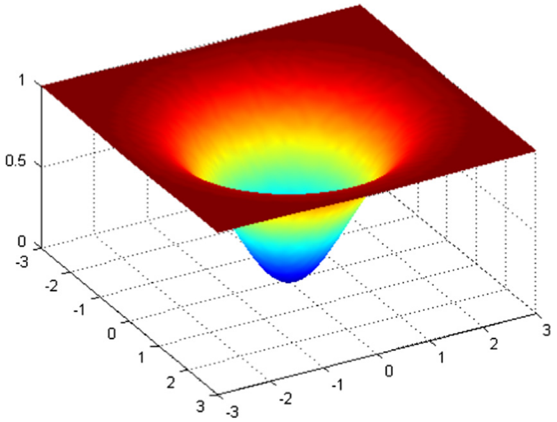

Fig. 3. A skin pore is modeled using a 2D Gaussian function with the gray-level intensities normalized to [0, 1] and $\sigma = 1$, where the coordinates (0, 0) represent the pore location.

$$= \int_{-\infty}^{+\infty} [G(u,v,k\sigma_1) - G(u,v,\sigma_1)] \cdot pore(x-u, y-v, \sigma_2)\, du\, dv, \tag{5}$$

where σ_1 is the scale of the pore model and σ_2 is the scale of the DoG filter. The magnitude of the DoG should be maximized so as to be selected as the maximum when compared to its 26 neighbors in a $3 \times 3 \times 3$ region. Hence, the response function at the center ($x=0$, $y=0$) of the pore is determined as follows:

$$\begin{aligned} D_{pore}(x=0, y=0, \sigma_1, \sigma_2) &= \int_{-\infty}^{+\infty} [G(u,v,k\sigma_1) - G(u,v,\sigma_1)] \cdot pore(-u, -v, \sigma_2)\, du\, dv \\ &= \sigma_2^2/(\sigma_1^2 + \sigma_2^2) - \sigma_2^2/(k^2\sigma_1^2 + \sigma_2^2). \end{aligned} \tag{6}$$

The maximum of D_{pore} is determined by taking the derivative of this function with respect to σ_1 and setting it at zero, giving

$$\hat{\sigma}_1 = k^{-1/2}\sigma_2. \tag{7}$$

Thus, the scale of the o-th octave's second-final layer (the (N_s+1)-st layer in a total of N_s+2 layers) is $\sigma_{1,o,N_s+1} = 2^o\sigma_0$, where σ_0 is the initial scale of the original image. The σ_0 is usually set at a value larger than 0.8, which is the minimum needed to prevent significant aliasing [16]. Consequently, when three DoG octaves (i.e. $o=3$) are constructed, the largest scale of the detected Gaussian pore function $\sigma_2 = k^{1/2}\hat{\sigma}_{1,o=3,N_s+1} = k^{1/2}2^3\sigma_0 > 6.4(where\ k>1, \sigma_0>0.8)$, which is large enough for the detection of pore-scale facial keypoints. Therefore, the number of DoG octaves N_o is set at 3 for all the experiments in this paper.

2.2. Quantity of keypoints

From the biological point of view, different people should have a similar quantity of pores in their facial skin, while from the computer-vision viewpoint, the quantity of keypoints detected directly affects the number of inliers available in the matching process. On the other hand, if many of the keypoints are noises or are unstable, it will be hard to find the inliers using RANSAC. Hence, via experiments, we have found that an appropriate number of keypoints that can densely cover the whole skin region is about 5000. However, this number is affected by the dense DoG responses at hairy (e.g. bearded) areas, which need to be re-weighted or discarded. In order to evaluate skin conditions precisely, we simply crop a hairless cheek region, which is about 7% (in size) of a whole face region, in our experiments. Fig. 4 illustrates the position and the size of the region to be extracted from a face image. Facial-landmark detection [17] or face-parsing [18] methods can be used for automatic, pre-defined skin-region cropping. With this cropped region, we set the desired number of keypoints, N_k, to within the range [450, 500].

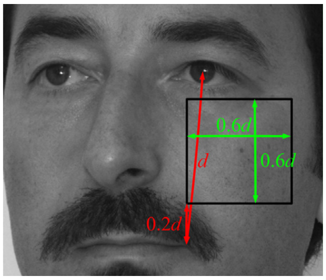

Fig. 4. Illustration of a cropped skin region in a face image for keypoint detection. (d is the distance between the right eye center and the right mouth corner.)

2.3. Pore Index and adaptive threshold

Substitute (7) into (6), the peak value P of the DoG response of a pore D_{pore} is given as follows:

$$P = D_{pore}(\hat{\sigma}_2) = (k-1)/(k+1). \tag{8}$$

This equation displays two very useful properties: (a) The maximized response is independent of the scale of the pores, so it is also invariant to image resolutions. (b) The peak value is relevant to the sampling frequencies in scale, due to the fact that the factor k is determined by the number of layers per octave N_s, shown in (3). Hence, the peak value is used to normalize the response with different sampling frequencies in scale in our sampling-frequency evaluation (i.e. the experiments in Section 2.4).

To determine the threshold which can result in $N_k \in [450, 500]$ from the cropped skin region, the binary search method is performed on a threshold list. The threshold list is set within the range $[0, 0.2 \times P]$, where P is the peak value of the DoG response of a modeled pore. In other words, an adaptive threshold τ is searched, which can be considered as the product of a coefficient R_{pore} and the peak value P. This coefficient R_{pore} is called the *Pore Index*, and R_{pore} is the ratio of the adaptive-peak threshold τ used in the PSIFT detector to the modeling-peak value P of the DoG images, defined as follows:

$$R_{pore} = \tau/P. \tag{9}$$

Therefore, the Pore Index R_{pore} represents the roughness/contrast of the skin. In Section 5.4, we will show how the kinds of distortions affect the Pore Index. In Section 5.6, a method for analyzing the difficulty of matching based on the Pore Index will be described.

2.4. Quantity-driven parameter selection

2.4.1. Matching performance measurement

Most of the pore-scale facial features are tiny and of low contrast. Thus, the selected parameters should be robust to noise or blurring for real, practical applications. However, the noise and blur kernels are hard to model. Hence, rather than using uniform-noise-added images as in [15], a skin dataset cropped from face images in the Bosphorus database [19] was used in our experiments. The database contains face images at different poses, captured using unsynchronized and uncalibrated cameras. The four subjects shown in Fig. 2(a), together with 16 other subjects selected randomly, are used to generate a skin dataset. In this experiment, skin regions in different poses are grouped to form four different datasets (10° and 20°, 10° and 30°, 20° and 30°, and 20° and 45°). Fig. 4 illustrates the cropping scheme: all of the facial-skin images are cropped from the same region of the respective face images.

Before measuring the matching performance, several notions should first be introduced. To match a keypoint in one face image to that in another face image, the Euclidean distances between the keypoint descriptor and that of another keypoint on the other face are computed. The best-matched keypoint is determined by the nearest-neighbor rule, i.e. the keypoint in the other face image with the minimum Euclidean distance. The distance ratio is defined as the ratio between the Euclidean distances of the best-matched keypoint and the second-best keypoint. A matched keypoint is accepted if the distance ratio is smaller than a threshold, which is determined empirically by experiments. Hence, the number of matched keypoints is far fewer than the number of keypoints in each image. The matched pairs of keypoints are then verified using RANSAC to fit the epipolar constraint in order to capture the inliers.

In [15], repeatability is used to measure the matching performance of the descriptors with different parameters, which is defined as the number of inliers divided by the smaller number of keypoints from the two images under consideration. Because the number of pore-scale keypoints detected in a face image is predefined by an adaptive threshold, the number of inliers is therefore proportional to the repeatability in our proposed framework. Furthermore, whether or not the inliers can be successfully identified by RANSAC is closely dependent on the inlier rate. The inlier rate is defined as the ratio of the number of inliers identified to the total number of matches.

2.4.2. Sampling frequency in scale

In this section, the sampling frequency in scale is experimentally determined. The quantity of inliers is used to evaluate the matching performance at different sampling frequencies. The inlier rate is also used to show the robustness of the different sampling frequencies. Each keypoint in one face image is matched to the keypoints in another face image. The matched keypoint is accepted if the distance ratio is smaller than 0.8, which is determined empirically by experiments. The matched pairs of keypoints are then verified using RANSAC, where the distance threshold used is set at 0.001, considering the fact that the images are unsynchronized and the facial appearances are non-rigid.

Fig. 5(a) and (b) shows the experimental results for determining the optimal number of layers (scales) N_s in each DoG octave. The results were obtained using 3–8 scales per octave; 8 layers per octave is the maximum that can prevent significant aliasing. All the results are the average of the four datasets. Fig. 5 (a) shows that the average number of inliers of the 20 subjects is significantly improved when more scales are used, although all of the subjects cannot achieve the highest inlier rate. The average inlier rates in Fig. 5(b) are always more than 85%; this reflects the robustness of PSIFT. To ensure that a sufficient number of inliers can be densely located in the whole facial-skin region, the quantity of inlier candidates is more important than the inlier rate in our algorithm. Therefore, 8 scales are sampled in each octave in our

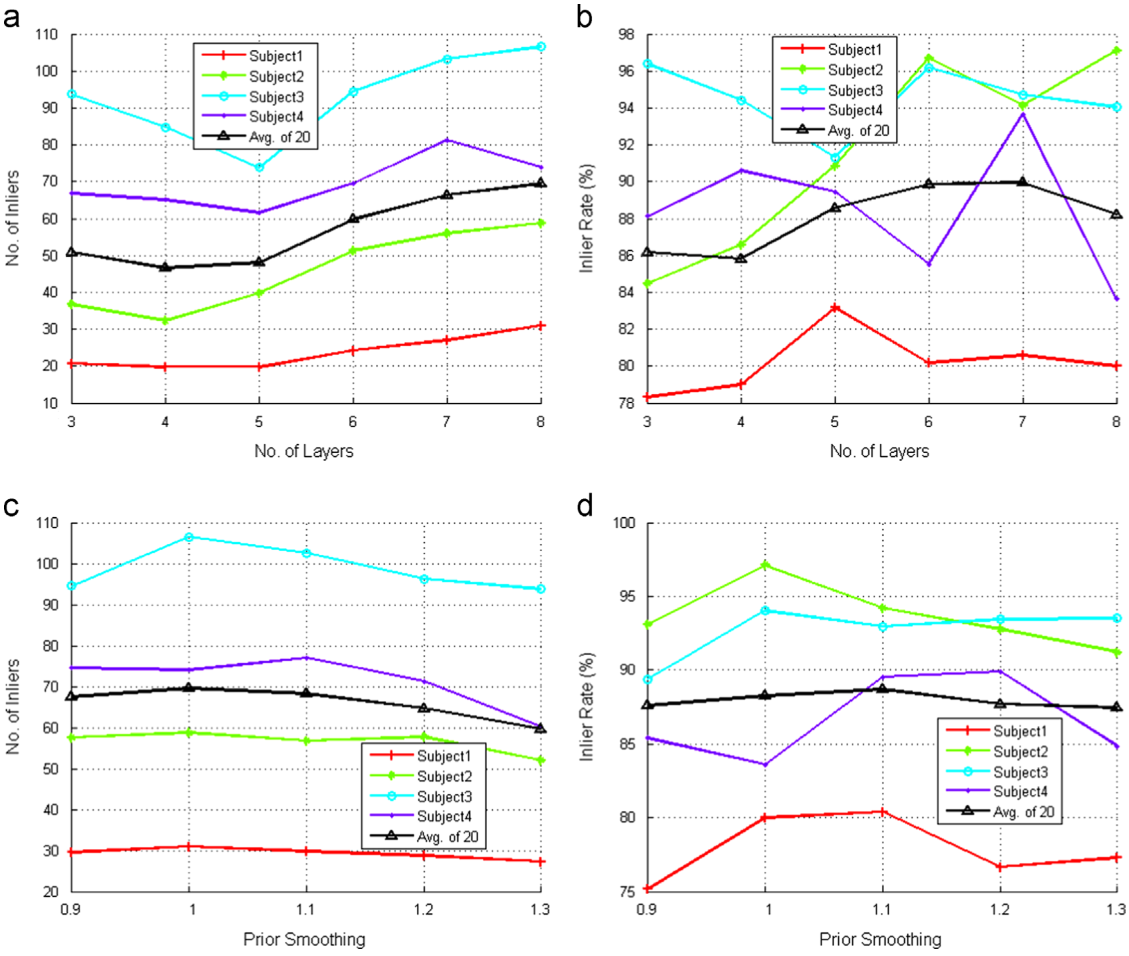

Fig. 5. (a) The numbers of inliers with different numbers of scales or layers sampled per octave, (b) the inlier rate with different numbers of scales sampled per octave, (c) the numbers of inliers detected with different values of the prior smoothness σ_0, and (d) the inlier rates with different values of the prior smoothness σ_0.

algorithm, and this setting is used for all the experiments in this paper.

2.4.3. Sampling frequency in the spatial domain

Fig. 5(c) and (d) shows the matching performance when the prior smoothing σ_0 varies. The results show that the largest number of inliers is obtained when the prior smoothing σ_0 is set at 1. Furthermore, the average inlier rates of the 20 subjects are almost a constant. Although the highest inlier rate is not obtained when $\sigma_0 = 1$, the quantity of inliers is more important for the face matching application. Furthermore, the outlier candidates can be effectively excluded from the matching process in the next step. Therefore, we choose to set the prior smoothing σ_0 at 1 for all the experiments in this paper.

2.4.4. Eliminating the edge responses

A poorly defined peak in the DoG scale space will have a large principal curvature across the edge, but a small one in the perpendicular direction. The principal curvature is represented by the eigenvalues of a 2×2 Hessian matrix H, r is the ratio between the larger eigenvalue and the smaller one, and $(r+1)^2/r$ is at a minimum when the two eigenvalues are equal, and it increases with r. In our experiments, the threshold r_{th} of r is set at 3.

3. Relative-position descriptor and discriminant learning

In this section, we will describe the local PSIFT descriptor, which is adapted from SIFT to extract the relative-position information about neighboring pores. Using the PSIFT descriptors, keypoints from two facial-skin regions can be matched. In order to further improve the matching accuracy, a pore-to-pore correspondences dataset is constructed and used to learn a discriminant subspace for pore-feature matching. The performance of these two features will be evaluated in Section 5.

3.1. Relative-position descriptor

Fig. 2 shows some sample results based on a DoG layer in an octave. The lighter points on the DoG, as shown in Fig. 2(c), represent the responses of the feature points. These points are indeed very similar to each other when they are observed individually: most of them are blob-shaped, and the surrounding region of each keypoint has almost the same color. However, unlike man-made textures, the relative positions of the pores are unique. Hence, the descriptor should extract not only the information around the keypoints themselves, but also the information in a neighborhood wide enough to include the neighboring pore-scale features. Thus, both the number of subregions and the support size of these subregions used in the SIFT descriptor are enlarged, as shown in Table 1. In addition, the keypoints are not assigned a main orientation, because most of the keypoints do not have a coherent orientation.

3.2. Pore-to-pore correspondences dataset

Using the proposed pore-scale feature-extraction framework, we have built a pore-to-pore correspondences dataset so that the learning for pore-pair matching can be conducted. This is the first dataset of its kind. In the following, we will describe how this dataset is generated.

The Bosphorus face database [19], which includes 105 distinct subjects, is used to develop the pore-to-pore correspondences dataset. Those face images with 10°, 20°, 30° and 45° poses are used. A total of 420 skin-region images are obtained by cropping from the cheek region of the face images. Fig. 4 shows the cheek region cropped from a face image, and Fig. 2 illustrates the 4 face samples whose corresponding cropped skin regions are used to establish the dataset.

With the cropped skin regions, we detect their pore-scale keypoints and generate the corresponding PSIFT descriptors based on our proposed pore-scale extraction framework. Then, for each subject, its pore keypoints at one pose are matched to the corresponding pore keypoints at an adjacent pose; this establishes three sets of matched keypoint pairs of 10° and 20°, 20° and 30°, 30° and 45°. Two keypoints are matched if they have the smallest Euclidean distance between their PSIFT descriptors, and if their distance ratio is smaller than 0.9. Then, the RANSAC algorithm [3] is applied to the matched candidates to identify those inliers which satisfy the epipolar constraint. During each RANSAC iteration, a candidate fundamental matrix is calculated using the eight-point algorithm [20], followed by non-linear refinement. After finding a set of matches between each image pair, we organize the matched keypoints to form *tracks*. A track is a set of matched keypoints across the face images of a subject at different poses. If a track contains more than one keypoint in the same image, it is considered to be inconsistent and is removed. We choose only those consistent tracks containing 4 keypoints corresponding to the 10°, 20°, 30° and 45° pose. Finally, 4240 tracks are established, which are then used to learn a discriminant subspace for pore-scale keypoint matching.

3.3. Discriminant learning

The extracted features should also be robust to distortions such as noises, unfocussed blurring and reflectance. However, such distortions are hard to model and are a challenge in the design of the feature-extraction framework. To tackle such distortions, a supervised learning procedure based on LDA is proposed.

The set of matched pore-keypoint tracks is used for learning, whereby 4 pore keypoints in a track form a class of matched pore keypoints. LDA is employed, which maximizes the discrimination between different classes, and minimizes the variance within the same class. In other words, this method learns a set of projection vectors which maximizes the ratio of the between-class scatter to the within-class scatter. The between-class scatter matrix S_B is defined as $S_B = \sum_{i=1}^{c} N_i(\mu_i-\mu)(\mu_i-\mu)^T$ and the within-class scatter matrix S_W is defined as $S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i}(x_k-\mu_i)(x_k-\mu_i)^T$, where μ_i is the mean descriptor of the class/track X_i, N_i is the number of examples in the class/track X_i, and c (equal to 4240 in our

Table 1
Parameters of the PSIFT and SIFT descriptors.

Parameters	PSIFT	SIFT
No. of subregions	8 × 8	4 × 4
Support size of total subregions	48 × scale of keypoints	12 × scale of keypoints
Support size of each subregion	6 × scale of keypoints	3 × scale of keypoints
No. of orientation bins	8	8
Dimension of the feature	512	128

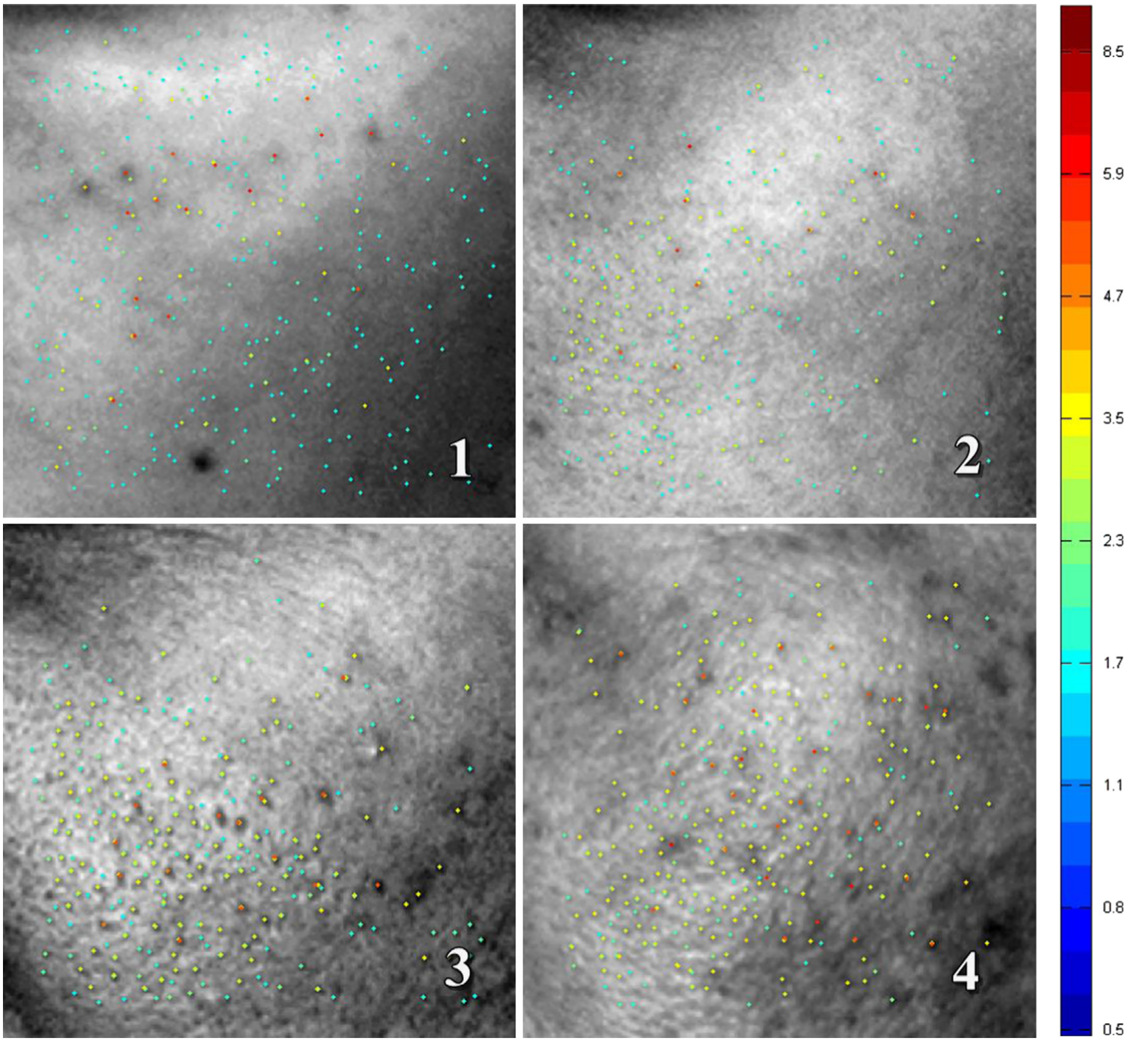

Fig. 6. Keypoint visualization based on Subjects 1–4's skin images corresponding to Fig. 2 (different colors for the keypoints indicate their scale, as shown in the color bar). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 2
Pore Index statistics (Average/Normalized Standard Deviation) of Subjects 1–4's skin images with different poses.

Subject	Avg/NStd
1	0.0147/0.0527
2	0.0176/0.0255
3	0.0239/0.0237
4	0.0308/0.0394

pore-to-pore dataset) is the number of classes/tracks. We seek the optimal projection W_{opt}, which maximizes the ratio of the determinant of the between-class scatter matrix of the projected examples to the determinant of the within-class scatter matrix of the projected examples, *i.e.*

$$W_{opt} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} \tag{10}$$

and the projection W_{opt} is the eigenvector with the largest eigenvalue of the following generalized eigen system: $S_B W_{opt} = \lambda S_W W_{opt}$. After projecting a PSIFT descriptor onto this LDA subspace, we normalize the projected descriptor to unit magnitude to form the LDAPSIFT descriptor.

4. Candidate-constrained matching

In order to achieve a more efficient and accurate matching, a candidate-constrained matching scheme based on two stages of matching will be presented.

In real applications, it is often the case that some features from two face images do not have any correct matches. This is because the two facial regions considered in the matching do not overlap. Furthermore, noise and blurring caused by a camera out of focus or in motion can have a significant influence on the matching result: some keypoints may not be detected in the second image, or the local-feature description may vary according to facial expressions. Hence, a *candidate-constrained matching scheme* based on two stages of matching is proposed to narrow the matching candidates and to achieve accurate matching, based on both the inter- and the intra-scale facial information.

First, the inter-scale facial information includes the relative locations of the pore-scale facial features and the primary facial features, such as the eyes and mouth, and the rough epipolar constraint, which is estimated by RANSAC in the first matching. Using this information, we can dramatically narrow the searching range in the face image to be matched, which can help reduce the number of outliers. The searching range can be considered only within a certain vertical range. In our experiments, the searching area in the vertical direction is limited to 10% of the image height. Thus, the searching area is narrowed to 20% of the whole face in the first matching. In the second matching, the estimated epipolar constraint and its relative location to the primary facial features are used, and the searching area can then be narrowed to about 5%.

The intra-scale facial information about facial features include the scales and the local descriptions of the keypoints, which are used in the matching process to further narrow the search of matched candidates. Face images captured from different views can be scaled so as to have a similar resolution. Usually, the higher-resolution face image can be down-scaled and aligned to the lower-resolution one. Hence, the scale of the detected keypoints in the two images should be similar. The scale of a keypoint here is the scale of the DoG layer where the magnitude is maximum. We define *scale ratio* as the ratio of the scales of the two keypoints from the two face images to be matched. Hence, we can further narrow the matching candidates according to the scale ratio. Two keypoints may be matched if their scale ratio is within the range [0.5, 2].

Finally, for those keypoints that satisfy the scale ratio constraint, their Euclidean distances are computed. A pair of keypoints with the closest distance is matched if their distance ratio is smaller than a certain threshold. The estimation of the positions of the keypoints to be matched based on facial features is summarized in Algorithm 1, while the candidate-constrained matching scheme is summarized in Algorithm 2. Based on the matching scheme, between 1000 and 3000 matches between two face images can be established. Then, RANSAC is employed to find the inliers and estimate the fundamental matrix robustly.

Algorithm 1. Estimation of a candidate's region.

```
Given two images $\mathbf{I}_1$ and $\mathbf{I}_2$, with a keypoint at $(x_2, y_2)$ in $\mathbf{I}_2$,
  a possible region of the corresponding keypoint $(\hat{x}_1, \hat{y}_1)$ in
  $\mathbf{I}_1$ is to be estimated.
For image $\mathbf{I}_2$, $(x_2^{LE}, y_2^{LE})$, $(x_2^{RE}, y_2^{RE})$, $(x_2^{LM}, y_2^{LM})$ and $(x_2^{RM}, y_2^{RM})$ are
  the coordinates of the left eye, right eye, left mouth corner
  and right mouth corner, respectively.
  The fundamental matrix is defined as $F$.
Connect the points $(x_2^{LE}, y_2^{LE})$ $(x_2^{LM}, y_2^{LM})$ and the points
  $(x_2^{RE}, y_2^{RE})$ $(x_2^{RM}, y_2^{RM})$, to form two lines, which divide the
  second image $\mathbf{I}_2$ into the left, the center, and the right
  subregions, denoted as $\mathbf{I}_2^L$, $\mathbf{I}_2^C$, and $\mathbf{I}_2^R$, respectively.
Compute the similarity transformation$T^L$, based on the
  coordinates of the left eye and the left mouth corner in $\mathbf{I}_2$
  and $\mathbf{I}_1$.
Compute the homography transformation $T^C$, based on the
  correspondences of the left eye, right eye, left mouth
  corner, and right eye corner in $\mathbf{I}_2$ and $\mathbf{I}_1$.
Compute the similarity transformation $T^R$, based on the
  coordinates of the right eye and the right mouth corner in
  $\mathbf{I}_2$ and $\mathbf{I}_1$.
**Input**: $(x_2, y_2)$, $\mathbf{I}_2^L$, $\mathbf{I}_2^C$, $\mathbf{I}_2^R$, $T^L$, $T^C$, $T^R$ and $F$.
Compute the epipolar line $l_1$ based on $(x_2, y_2)$ and $F$, and
  define the direction of $l_1$ as $\theta_{l1}$.
**if** $(x_2, y_2)$ in $\mathbf{I}_2^L$ **then**
  Transform $(x_2, y_2)$ to $(x'_1, y'_1)$ using $T^L$.
  Project $(x'_1, y'_1)$ to $l_1$, and compute $(\hat{x}_1, \hat{y}_1)$.
**else if** $(x_2, y_2)$ in $\mathbf{I}_2^C$ **then**
  Transform $(x_2, y_2)$ to $(x'_1, y'_1)$ using $T^C$.
  Project $(x'_1, y'_1)$ to $l_1$, and compute $(\hat{x}_1, \hat{y}_1)$.
**else if** $(x_2, y_2)$ in $\mathbf{I}_2^R$ **then**
  Transform $(x_2, y_2)$ to $(x'_1, y'_1)$ using $T^R$.
  Project $(x'_1, y'_1)$ to $l_1$, and compute $(\hat{x}_1, \hat{y}_1)$.
**end if**
**Output**: the ellipsoid region $\mathbf{R}(x, y)$ satisfies
  $(x \times \cos(\theta_{l1}) + y \times \sin(\theta_{l1}) - \hat{x}_1)^2 / a^2$
  $+ (-x \times \sin(\theta_{l1}) + y \times \cos(\theta_{l1}) - \hat{y}_1)^2 / b^2 \le 1,$     (11)
  where $a = 0.32 \times H_1$ and $b = 0.04 \times H_1$. Thus, the area of
  this ellipsoid region is about $\pi \times 0.32 \times 0.04 \times$
  $W_1/H_1 \approx 5\%$ of the total area of the first face image.
```

Algorithm 2. Candidate-constrained matching.

```
Assume that there are $N_{k1}$ and $N_{k2}$ keypoints detected in $\mathbf{I}_1$
  and $\mathbf{I}_2$, respectively. The coordinates and the scale of the $i$-
  th keypoint in $\mathbf{I}_1$ are denoted as $(x_1^i, y_1^i)$ and $\sigma_1^i$,
  respectively. Similarly, the coordinates and the scale of the
  $j$-th keypoint in $\mathbf{I}_2$ are denoted as $(x_2^j, y_2^j)$ and $\sigma_2^j$,
  respectively. Without loss of generality, assume that
  $N_{k1} < N_{k2}$. The height of the image $\mathbf{I}_1$ is $H_1$.
Matching the keypoints is established from $\mathbf{I}_2$ to $\mathbf{I}_1$. After
  the first stage of matching, the fundamental matrix is $F_1$.
**for** Stage $(t=1,2)$ **do**
  **for** the $j$-th keypoint in $\mathbf{I}_2$ **do**
    Initialize a candidate list $\mathbf{p}$, which includes all the
    keypoints in $\mathbf{I}_1$ ($size(\mathbf{p}) = N_{k1}$).
    **for** the $i$-th keypoint in $\mathbf{I}_1$ **do**
      **if** $t == 1$ **and** $|y_2^j - y_1^i| < 0.1 \times H$ **and**
      $0.5 \le |\sigma_2^j / \sigma_1^i| \le 2$ **then**
        $\mathbf{p}$ is not updated.
      **else if** $t == 2$ **and** $(x_1^i, y_1^i) \in \mathbf{R}(x_2^j, y_2^j)$ (described in
      Algorithm 1) **and** $0.5 \le |\sigma_2^j / \sigma_1^i| \le 2$ **then**
        $\mathbf{p}$ is not updated.
      **else**
        Remove the $i$-th keypoint from the candidate list
        $\mathbf{p}$ of $\mathbf{I}_1$.
      **end if**
    **end for**
    **if** the *distance ratio* based on the reduced candidate
    list $\mathbf{p}$ is smaller than $\delta$, which is a constant between
    0.8 and 0.9. **then**
      A match is established.
    **end if**
  **end for**
  Using RANSAC, compute the fundamental matrix $F$ and
  determine the inliers.
**end for**
```

5. Experiments

In this section, we will evaluate the performances of our proposed pore-scale facial features in terms of accuracy for pore identification and skin/face matching. Two evaluation criteria are used based on different testing datasets: the receiver operating characteristics (ROC) curves and equal error rate (EER) using the pore-to-pore correspondences dataset, and the number of inliers and the repeatability from RANSAC using an uncalibrated dataset. Both of these two dataset are generated from the Bosphorus database [19]. The face images used in the experiments are the original size in Bosphorus database (about 1400 × 1200 pixels) unless otherwise specified.

Three types of distortions – namely low resolution, noise, and blurring – are simulated to evaluate the robustness of the pore-scale facial features. The variations of the Pore Indices for facial-skin images with these distortions are discussed. We have also collected the statistics of the Pore Indices for 420 face images. Finally, based on these results, a method that can estimate the difficulty level of face matching based on the Pore Index is presented.

Our methods will be compared with the state-of-the-art matching approach based on SIFT [15]. In the experiments, the number of inliers detected is used to evaluate the performance of the matching task. All the parameters of SIFT are set at the recommended values in [15] except the peak threshold for the DoG detection. In [15], the peak threshold of the SIFT detector is set at 0.03 for the best performance, which results in extracting as few as 10 keypoints in face images. Therefore, the peak threshold of the SIFT detector in our experiments is reduced and set at 0.005, which leads to detecting a similar number of keypoints as the PSIFT detector. In our LDAPSIFT learning procedure, 2120 tracks are used for training. The remaining 2120 tracks are used for testing.

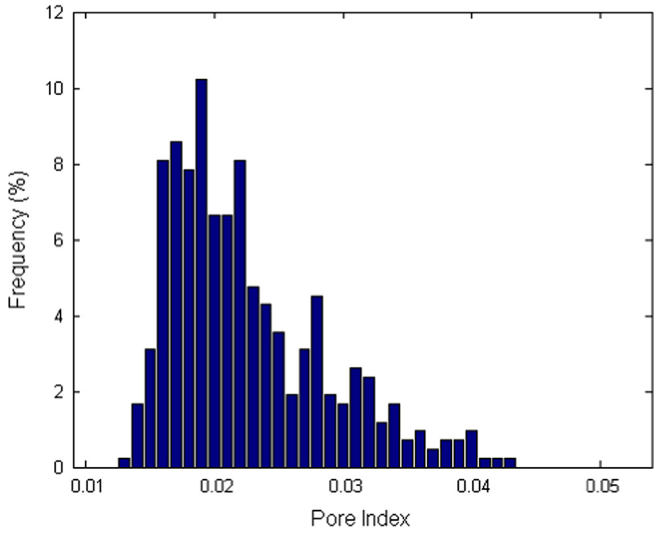

Fig. 7. The frequency of subjects with different Pore Indices (the bin size is 0.001 and the number of bins is 41).

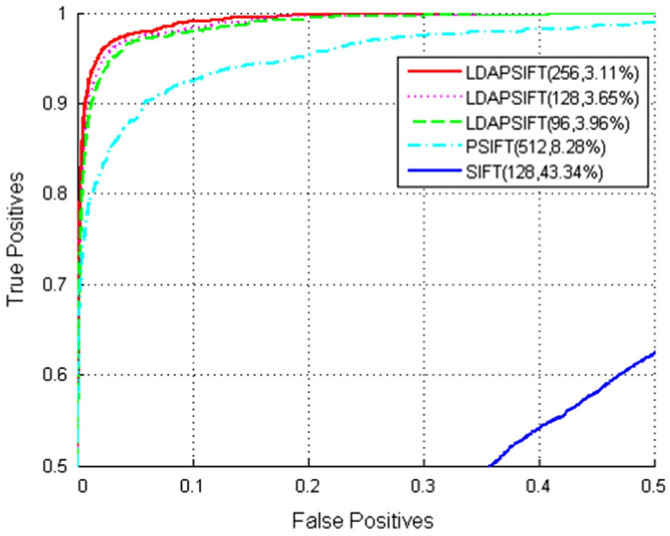

Fig. 8. ROC curves of the different descriptors. In the parenthesis, the first number is the dimension of the descriptor and the second one is the EER.

5.1. Pore-scale facial-feature statistics and visualization

Fig. 2 shows four face samples with the corresponding zoomed-in, local skin-texture images and the corresponding DoG layers. These faces have very different skin conditions. Subject 1's skin is very smooth and fine, while Subjects 2 and 3 have ordinary skin conditions. The pores on Subject 2 are smaller than those on Subject 3. In addition to the pore-scale facial features, Subject 4 has a greater number of marker-scale facial features. Nevertheless, the total numbers of PSIFT keypoints detected are similar because we have used our quantity-driven detection scheme. We define the *normalized standard deviation* (*Nstd*) of the Pore Indices as the standard deviation of the Pore Indices divided by the corresponding average of the Pore Indices, which is used to measure the difference between the Pore Indices in various face images of the same subject. Fig. 6 illustrates the keypoint-detection results, where different colors for the keypoints represent the scales detected on the corresponding DoG layers. The average and the normalized standard deviation of the Pore Indices in 4 images of each of the subjects are summarized in Table 2. We find that the Pore Indices can effectively represent the roughness/contrast of the skin images. Later, the average and the normalized standard deviation of the Pore Indices are used to analyze the difficulty level of face matching; further details will be described in Section 5.6.

The statistics of the Pore Indices for 420 images from the Bosphorus database [19] are collected based on the cropped skin regions, as illustrated in Fig. 4. The frequencies of the different Pore Indices for the face images are shown in Fig. 7. The distribution of the Pore Indices reflects the distribution of people's skin appearances.

5.2. Descriptor learning based on the pore-to-pore dataset

Based on the PSIFT framework, 4240 pore tracks were produced from 100 subjects; the details have been described in Section 3.2. In our LDAPSIFT learning procedure, we randomly select 2120 tracks from 55 subjects for LDAPSIFT training. Then, the remaining 2120 tracks from the other 45 subjects are used for testing the performances of the different descriptors. Every track is composed of 4 pore keypoints from facial images of the same subject at 10°, 20°, 30°, 45° difference from the corresponding frontal-view image. In our experiments, those pores at the 10° are chosen to form the gallery set, while those pores at 45° form the testing set. The ROC curves and the EERs of the different descriptors are presented in Fig. 8.

SIFT can achieve an EER of only 43.34%, while LDAPSIFT and PSIFT produce a significant improvement over SIFT: less than 4% and 8.28%, respectively. The results for LDAPSIFT and PSIFT show that the pore-scale facial features are distinctive and can make much more accurate identification among a huge number of pores. The ROC curves of LDAPSIFT with different descriptor dimensions are also given. LDAPSIFT with a dimension of 128 is used in the

Table 3
Skin matching results.

Method	Avg. no. of inliers	Repeatability (%)	No. of image pairs on which more than 20 inliers
LDAPSIFT	89.33	18.01	102
PSIFT	73.86	14.89	96
SIFT detector+PSIFT	25.94	5.95	44
PSIFT detector+SIFT	8.65	1.74	11
SIFT	3.66	0.79	5

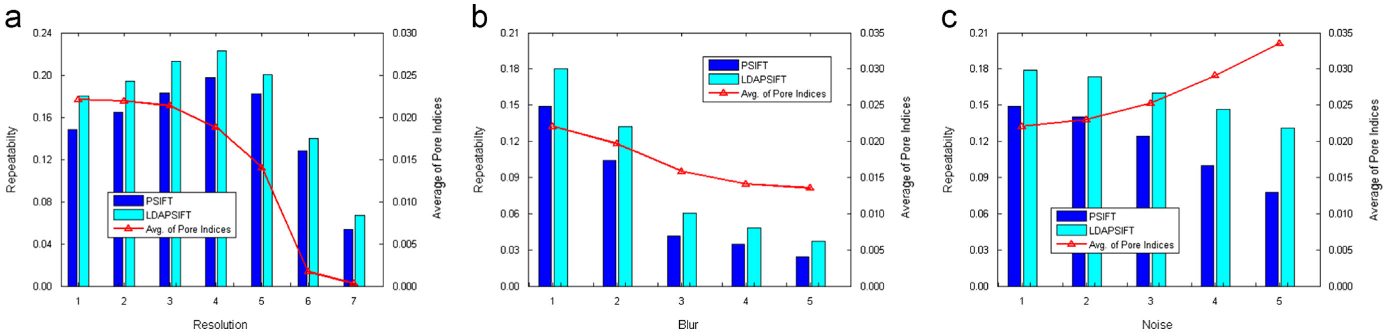

Fig. 9. Respective numbers of inliers using PSIFT and LDAPSIFT under three types of distortion: (a) low resolution, (b) blurring, and (c) uniform noise.

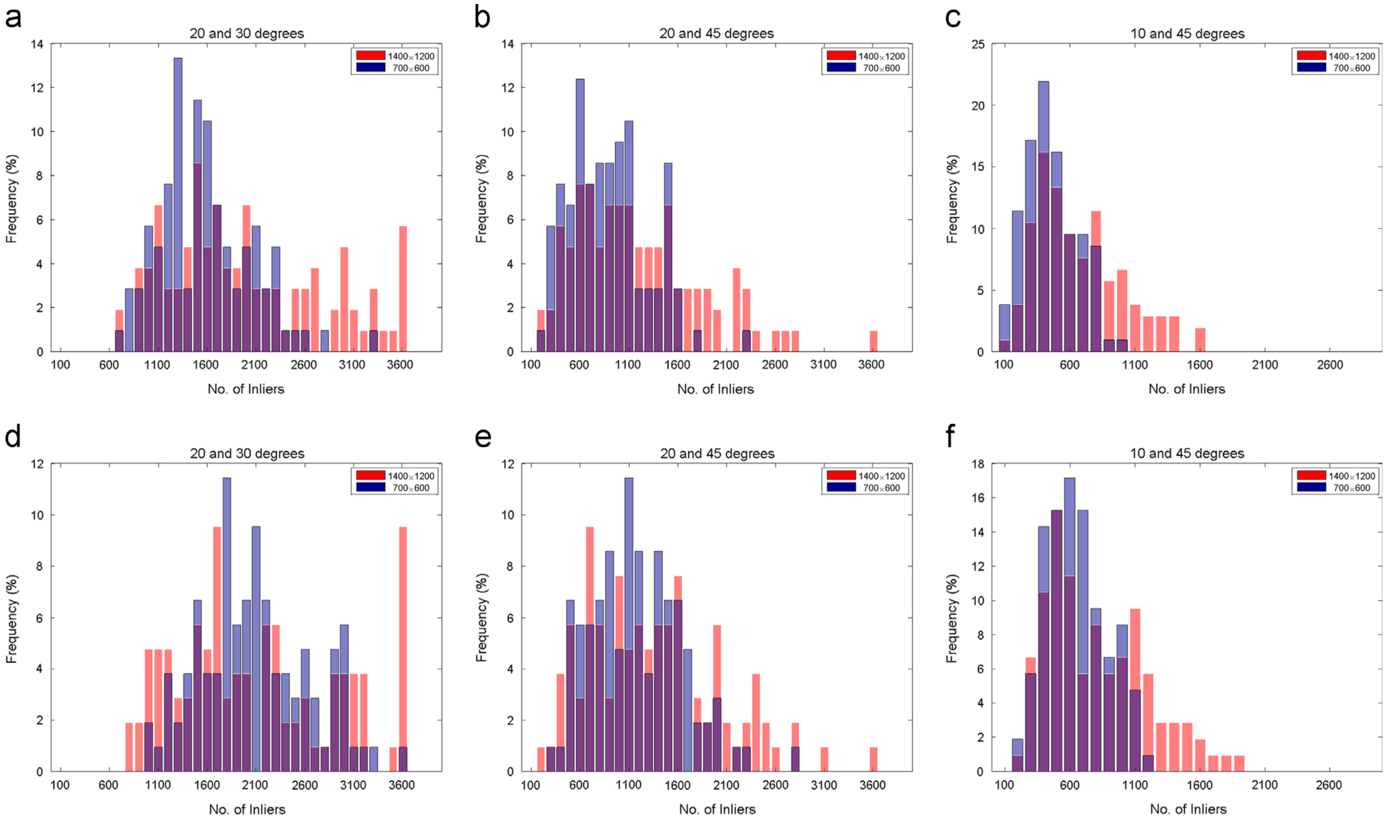

Fig. 10. Distributions of the number of inliers for all 105 subjects in the Bosphorus database using PSIFT and LDAPSIFT for the face images at two different resolutions (700 × 600 and 1400 × 1200) and under three different pose variations: PSIFT (a) 10° (images at 20° and 30°), (b) 25° (images at 20° and 45°), and (c) 35° (images at 10° and 45°); LDAPSIFT (d) 10° (images at 20° and 30°), (e) 25° (images at 20° and 45°), and (f) 35° (images at 10° and 45°).

rest of the experiments in this paper; this is far fewer than that of PSIFT (whose required dimension is 512).

5.3. Skin matching based on the Bosphorus dataset

In this section, we evaluate the performance of each stage of our algorithm in terms of skin matching. We use 105 skin-region pairs cropped from 210 face images, which were captured at 10° and 45° to the right of the frontal view in the Bosphorus database [19], as shown in Figs. 2 and 5. Considering the fact that the dataset is uncalibrated and unsynchronized, the distance threshold used in RANSAC is set at 0.0005 to fit the epipolar constraint.

To investigate the performance of each stage of our algorithm, we combine the different detectors and descriptors in two different ways, namely SIFT detector+PSIFT descriptor and PSIFT detector+SIFT descriptor. Having changed the default peak threshold from 0.03 in [15] to 0.005 in our experiment, the SIFT detector can detect a similar average number of keypoints over 210 skin-region images to the PSIFT detector. Table 3 illustrates the number of inliers after RANSAC, the repeatabilities, and the number of image pairs which have more than 20 inliers, for each of the methods. Only 5 and 44 out of 105 image pairs have more than 20 inliers using the SIFT descriptor with the SIFT detector and with the PSIFT detector, respectively, i.e. most of the cases fail to match the pore keypoints. The SIFT detector with the PSIFT descriptor fails most of the cases, too. To achieve a good performance, the PSIFT detector with the PSIFT descriptor or with the LDAPSIFT descriptor should be considered; it can match, on average, 73.86 and 89.33 inliers, respectively, from the image pairs.

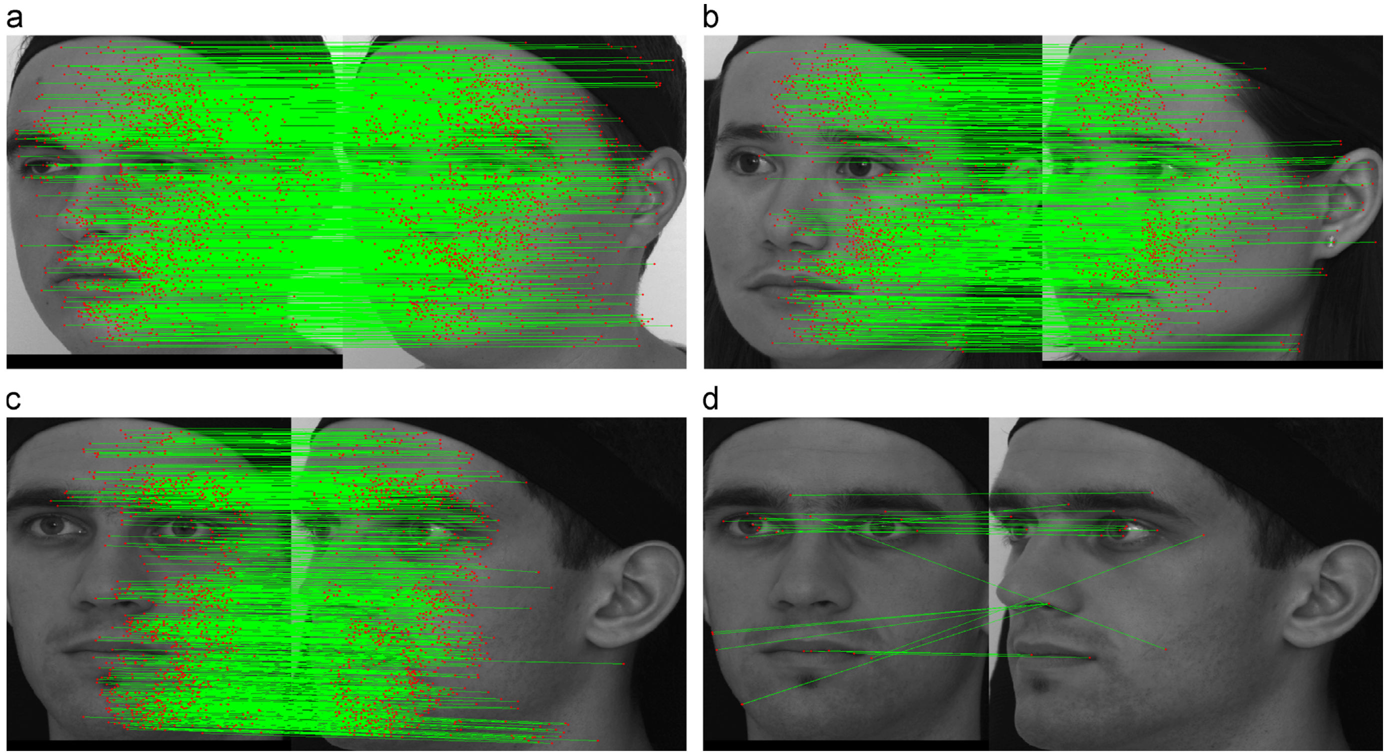

Fig. 11. PSIFT matching results for those pairs of faces with the median number of inliers of the 105 subjects for three pose variations: (a) 10° pose difference, 1522 inliers detected; (b) 25° pose difference, 858 inliers detected; (c) 35° pose difference, 441 inliers detected; and (d) SIFT matching results for the image pair in (c): 28 matches are established, but no inliers can be found via RANSAC.

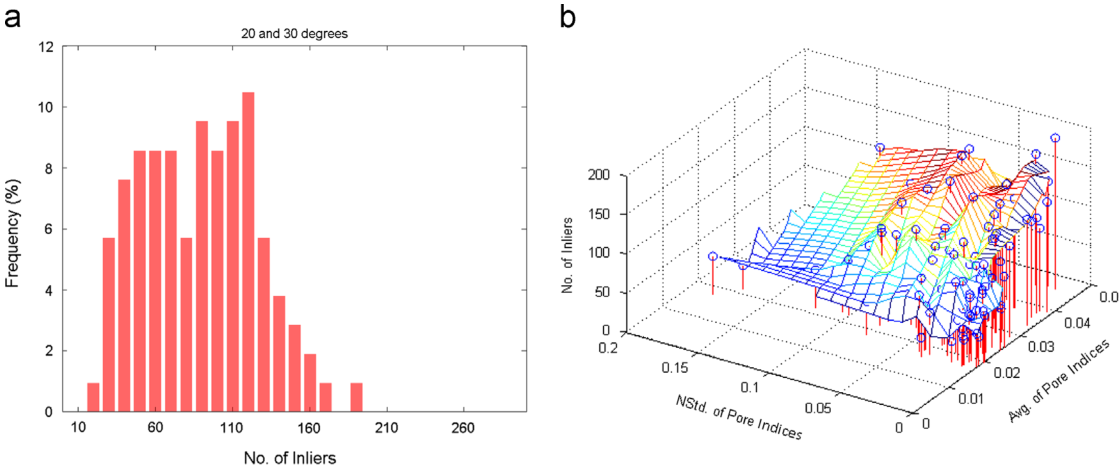

Fig. 12. (a) Histogram of the number of inliers, and (b) 3D distribution of skin-matching results, based on the PSIFT skin matching of all 105 subjects in the Bosphorus database, with poses at 20° and 30°.

5.4. *Robustness evaluation with simulated distortions*

In this section, three types of distortions – namely low resolution, blurring, and uniform noise – are considered to evaluate the robustness of PSIFT and LDAPSIFT for the skin-matching task. The gallery and the testing set used in Section 5.3 are used as the original image pairs.

Bicubic interpolation is used to generate lower-resolution images with down-sampling factors of 0.875, 0.75, 0.625, 0.5, 0.375, and 0.25. For the blurring distortion, only the testing set is blurred and the scales of the Gaussian kernels are set at 0.8, 1.6, 2.4, and 3.2. The noisy images are generated by adding 0.5%, 1%, 1.5%, and 2% of uniform noise to the images, i.e. a random number from the uniform interval $[-0.01, 0.01] \times noi$ (where the factor $noi = 0.5, 1, 1.5, 2$) is added to a face image whose pixel values are in the range [0, 1]. The repeatability of the keypoints is used to measure the robustness of PSIFT and LDAPSIFT.

Fig. 9(a)–(c) illustrates the average repeatability under these three types of distortions. The average value of the Pore Indices is also given as a measure of image quality under the different distortions. PSIFT and LDAPSIFT are both scale-invariant, so they are robust to resolution variations in certain regions, as shown in Fig. 9(a). We find that the repeatability initially increases with a lower down-sampling rate. But when the down-sampling rate is lower than 0.5, the repeatability starts to decrease. This is because the down-sampling process can eliminate some noise in the images, which increases the repeatability. However, when the down-sampling rate is further lowered, the details in skin regions are also eliminated. This causes a decline in the repeatability. The repeatability declines dramatically when the down-sampling

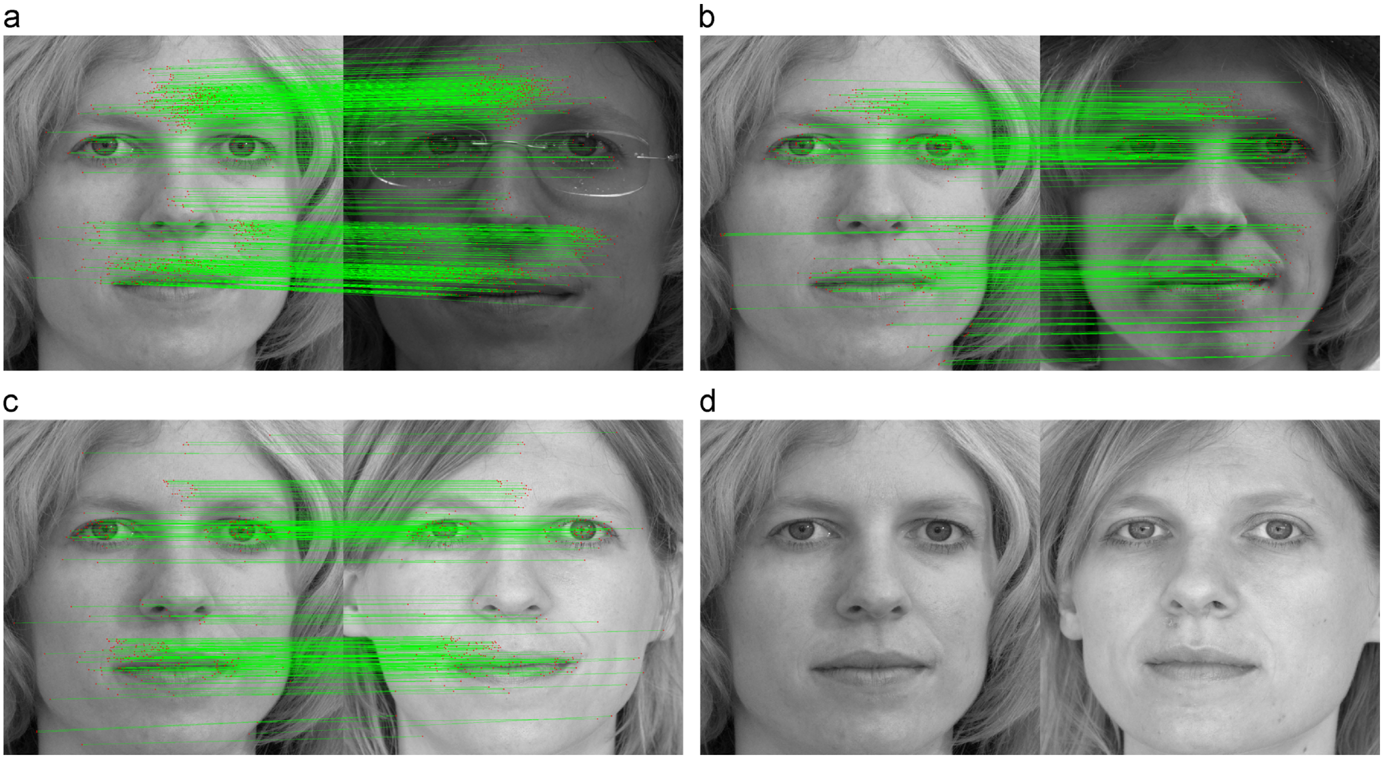

Fig. 13. Preliminary results for differentiating between identical twins.

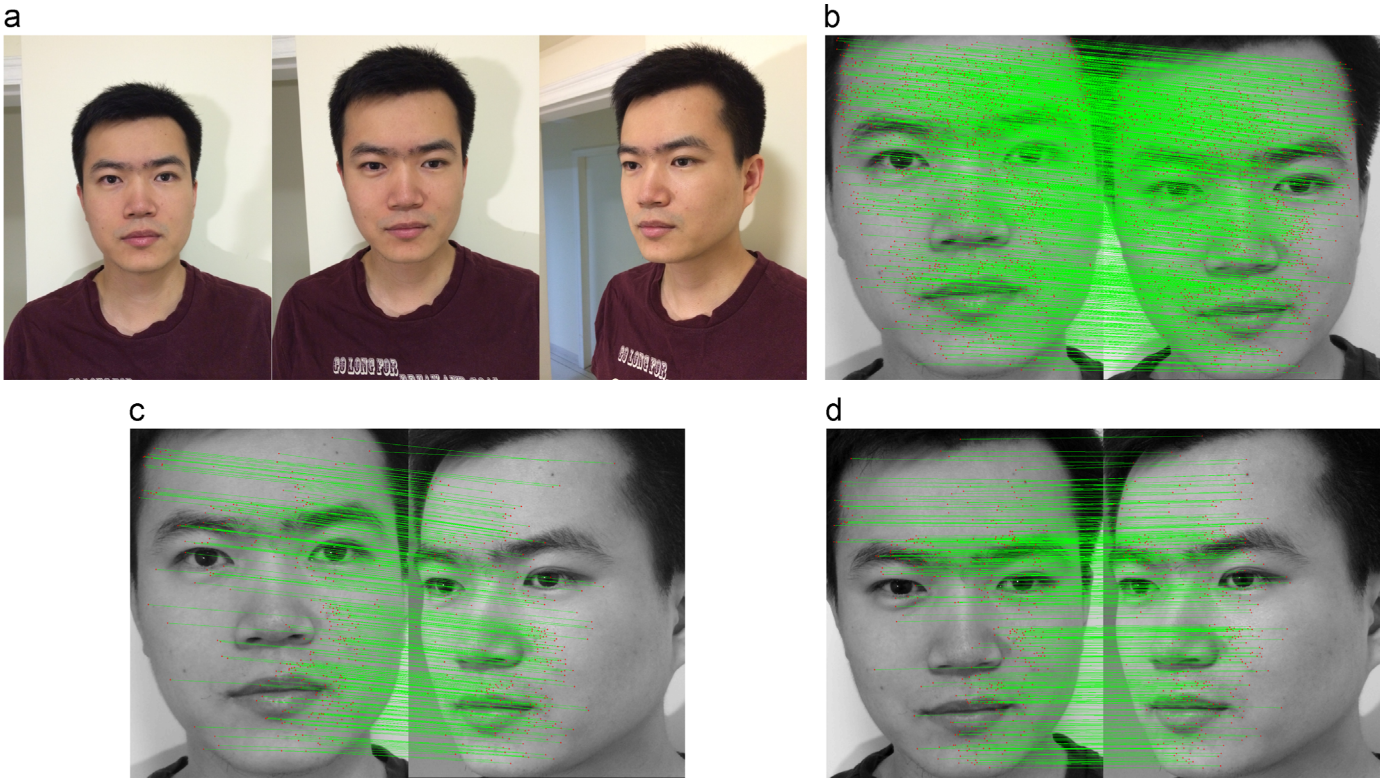

Fig. 14. (a) Three images captured by an iPhone 5S, and (b)–(d) the corresponding matching results.

factor is set lower than 0.375 (the 6th bar). Hence, the lowest resolution of a face image that can generate a sufficient number of pore-scale facial features is about 600 × 500 pixels, which is lower than the capabilities of most digital cameras nowadays. When the down-sampling rate is larger than 0.75 (the 3rd bar), the average Pore Indices is approximately a constant. When the face images are further down-sampled, some of the peaks on the DoG start to be distorted. Hence, the Pore Index decreases with a higher down-sampling rate. Blurring also decreases the high-frequency information in face images, so both the Pore Indices and the repeatability decline with blurring, as shown in Fig. 9(b). As the uniform noise introduces high-frequency content to faces, the Pore Index increases with the magnitude of the noise, as shown in Fig. 9 (c). Most of the pore-scale facial features are tiny and of low contrast. Thus, high-frequency information is easily distorted. Consequently, the repeatability decreases when the strength of such distortion increases.

5.5. *Face matching based on the Bosphorus dataset*

In this section, experiments were conducted using the Bosphorus database [19]. The face images in this database were captured unsynchronized and from different views. The subjects were filmed at different angles by rotating the chair they were sitting in to align with stripes placed on the floor indicating the corresponding angles. All 105 subjects in the Bosphorus database were used to evaluate the performance of our method under different skin conditions. Four samples are illustrated in Fig. 2. The distance threshold used in RANSAC is set at 0.0001, considering the fact that the images are unsynchronized and that the facial appearances are non-rigid. Fig. 10 shows the frequency of subjects with respect to the different numbers of inliers detected for three sets of image pairs with different combinations of poses (20° and 30°, 20° and 45°, and 10° and 45° poses) based on PSIFT and LDAPSIFT. The original-size images contain distortions such as reflectance and out-of-focus blurring. LDAPSIFT and PSIFT can achieve a slight improvement when the images are down-sampled. Compared to PSIFT, LDAPSIFT is more robust to different kinds of distortion and can establish denser matches. Fig. 11 shows three samples of the matching results based on PSIFT, with the median number of inliers detected with respect to the three pose combinations, and with an image resolution of 700×600.

5.6. *Skin-matching difficulty analysis*

In this section, we will revisit the experimental results for PSIFT and LDAPSIFT, and analyze the face-matching difficulty based on the statistics of the Pore Indices of the matched image pairs. Fig. 10 (a) shows the performance of PSIFT and LDAPSIFT in terms of the frequency of subjects with respect to the different numbers of inliers detected when the two faces under 20° and 30° poses, and when the resolutions of the face images are 700×600 and 1400×1200, respectively. Correspondingly, if the skin region is cropped from the face images at their original resolution, i.e. 1400×1200, the histogram of the different numbers of inliers for skin matching is shown in Fig. 12(a). The experimental results are also shown in the 3D plot in Fig. 12(b), with the number of inliers, the normalized standard deviation, and the average Pore Indices being the three axes of the 3D plot. Lower-resolution distortion, or blurring, can reduce the high-frequency information in the original face images, which leads to a lower Pore Index, as shown in Fig. 9 (a) and (b). In contrast, the noise distortion will introduce more high-frequency variations, which results in a higher Pore Index, as shown in Fig. 9(c). Thus, if the Pore Indices of two images of the same person are very different, at least one of those two images is distorted. For such a pair of face images, it is relatively hard to establish a large number of correct matches. In Fig. 12(b), a mesh is generated based on the points in the 3D plot. We find that the image pairs with a large average value and a small normalized standard deviation of the Pore Indices can establish more correspondences than the others can.

5.7. *Applications of our approach*

In this section, we will further discuss the potential applications of our proposed approach, and show some preliminary results for the applications. Our work shows a potential way to merge general computer-vision approaches and face-based approaches. For example, the classical shape-from-motion method with known camera intrinsic parameters, which extracts correspondences of 3 views on top of our method, can be used in 3D face reconstruction.

We have conducted a simple classification experiment where we attempt to discriminate among different subjects based on their cropped cheek-region skin images. We randomly select 20 subjects from the 105 subjects in the Bosphorus database [19]. The skin image captured from a frontal-view image is designated as the gallery set for each of the subjects, while the corresponding image regions of the other 4 views are designated as the testing set. The number of inliers, after RANSAC, is used as a similarity measure. Both PSIFT and LDAPSIFT can achieve a 100% recognition rate based on using only a single image per class as the gallery. A similar experiment setting was reported in [14], which used 20 subjects and 4 skin regions per subject. For each skin region, 26 images were designated as the gallery, and 6 images were used for testing. The result was obtained by taking the majority vote of the 4 regions, and a recognition rate of only 73% was achieved. The significant improvement using our method is due to the fact that PSIFT and LDAPSIFT can distinguish every pore-scale facial keypoint in a skin block, rather than treating skin images as texture, as in [14].

Another potential application, i.e. differentiating between identical twins, is shown in Fig. 13. The images are from the ND-Twins-2009-2010 database. Fig. 13(a) and (b) shows two image pairs of the same subject with different illuminations. We find that our method can handle images captured in outdoor scenes. This is due to the fact that only the contrast of the pore keypoints, rather than their relative position, is affected by illumination. Fig. 13(c) shows an image pair of the same subject taken one year apart. Fig. 13 (d) shows a pair of images of twins, i.e. the query is an imposter. No correspondences are established by RANSAC in this pair of images.

To show the robustness of our method and its potential in mobile applications, three face-matching results are shown in Fig. 14 based on images captured by an iPhone 5S. Note that the environmental luminance was relatively low, the aperture was set at F2.2, the ISO was set at 320, and the shutter speed was set from 1/15 s to 1/17 s by the iOS automatically. Thus, the images suffer from noises due to the high ISO, and blurring due to the slow shutter speed. However, our method can still successfully establish a huge number of correspondences between the images.

6. Conclusion and discussion

In this paper, we have proposed a new framework to extract pore-scale facial features from facial skin. Our method identifies every pore in the facial images, rather than considering pore-scale facial features as a kind of texture. We have modeled the blob-shaped pore-scale features using a Gaussian kernel, and analyzed the relationship between the strength of the response and the number of DoG layers used. A new measure, namely the Pore Index, is proposed to analyze the relationship between facial-skin image conditions and the difficulty level of face matching, which also reflects the adaptive threshold for keypoint detection. The PSIFT descriptor is designed to extract the relative-position information about the neighborhood of the keypoints. A pore-to-pore correspondences dataset, including 4×4240 PSIFT features and 6×4240 matched pairs, is established. Based on this pore-to-pore dataset, a learning-based pore-scale facial feature LDAPSIFT–which is more distinctive and compact than the existing methods–is proposed for tackling different kinds of distortions. We have shown via experiments that pore-scale facial features are sufficiently distinctive to track a face's geometry. Furthermore, LDAP-SIFT can efficiently reduce the computation time of feature matching to 8% of PSIFT. For the feature-matching stage, PSIFT needs 1.45 s, while LDAPSIFT needs only 0.12 s on an Intel i7 3.4 GHz CPU with 8 threads and 8 GB Ram PC under the MATLAB R2012b programming environment. The runtime can be further reduced by parallel computing techniques; this makes it affordable for large-scale recognition/identification applications.

In our future work, we will further investigate the properties of the pore-scale facial features, and further improve the matching performance when the face images have a larger baseline. This work allows for accurate 3D face reconstruction based on a number of 2D face images, and the pore-scale facial features can be used as a new biometric feature for person identification.

Conflict of interest

None declared.

References

[1] Y. Lin, G. Medioni, J. Choi, Accurate 3d face reconstruction from weakly calibrated wide baseline images with profile contours, in: CVPR. 2010, pp. 1490–1497.

[2] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (surf), Comput. Vis. Image Underst. 110 (3) (2008) 346–359.

[3] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.

[4] V. Lepetit, F. Moreno-Noguer, P. Fua, Epnp: an accurate o(n) solution to the pnp problem, Int. J. Comput. Vis. 81 (2) (2009) 155–166.

[5] I. Matthews, S. Baker, Active appearance models revisited, Int. J. Comput. Vis. 60 (2) (2004) 135–164.

[6] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, Robust and efficient parametric face alignment, in: ICCV, IEEE, 2011, pp. 1847–1854.

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

[8] L. Wiskott, J.M. Fellous, N. Kuiger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.

[9] D.W. Hansen, Q. Ji, In the eye of the beholder: a survey of models for eyes and gaze, IEEE Trans. Pattern Anal. Mach. Intell. 32 (3) (2010) 478–500.

[10] P. Wang, Q. Ji, Multi-view face and eye detection using discriminant features, Comput. Vis. Image Underst. 105 (2) (2007) 99–111.

[11] N. Spaun, Facial comparisons by subject matter experts: their role in biometrics and their training, in: Advances in Biometrics, Lecture Notes in Computer Science, vol. 5558, Springer, Berlin, Heidelberg, 2009, pp. 161–168.

[12] U. Park, A.K. Jain, Face matching and retrieval using soft biometrics, IEEE Trans. Inf. Forensics Secur. 5 (3) (2010) 406–415.

[13] D. Lin, X. Tang, Recognize high resolution faces: from macrocosm to microcosm, in: CVPR, vol. 2, 2006, pp. 1355–1362.

[14] O.G. Cula, K.J. Dana, F.P. Murphy, B.K. Rao, Skin texture modeling, Int. J. Comput. Vis. 62 (1–2) (2005) 97–119.

[15] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[16] J. Morel, G. Yu, Is sift scale invariant? Inverse Probl. Imaging 5 (1) (2011) 115–136.

[17] X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in: CVPR, 2012, pp. 2879–2886.

[18] B. Smith, L. Zhang, J. Brandt, Z. Lin, J. Yang, Exemplar-based face parsing, in: CVPR, 2013, pp. 3484–3491.

[19] A. Savran, N. Alyz, H. Dibekliolu, O. Eliktutan, B. Gkberk, B. Sankur, et al., Bosphorus database for 3d face analysis, in: Biometrics and Identity Management, Lecture Notes in Computer Science, vol. 5372, Springer, Berlin, Heidelberg, 2008, pp. 47–56.

[20] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Second ed., Cambridge University Press, Cambridge, UK, 2004, ISBN: 0521540518.

Dong Li received the bachelor's degree in Computer Science, in 2006, and the master's degree in Computer Science, in 2009, both from Tianjin University, Tianjin, and the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong SAR, China, in 2014.

Kin-Man Lam received the Associateship in Electronic Engineering with distinction from The Hong Kong Polytechnic University (formerly called Hong Kong Polytechnic) in 1986, the M.Sc. degree in communication engineering from the Department of Electrical Engineering, Imperial College of Science, Technology and Medicine, London, U.K., in 1987, and the Ph.D. degree from the Department of Electrical Engineering, University of Sydney, Sydney, Australia, in August 1996.

From 1990 to 1993, Prof. Lam was a lecturer at the Department of Electronic Engineering of The Hong Kong Polytechnic University. He joined the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University again as an Assistant Professor in October 1996. He became an Associate Professor in 1999, and is now a Professor. Prof. Lam was actively involved in professional activities. He has been a member of the organizing committee or program committee of many international conferences. In particular, he was the Secretary of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), the Technical Chair of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP 2004), a Technical Co-Chair of the 2005 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2005), a secretary of the 2010 International Conference on Image Processing (ICIP 2010), a Technical Co-Chair of 2010 Pacific-Rim Conference on Multimedia (PCM 2010), and a General Co-Chair of the 2012 IEEE International Conference on Signal Processing, Communications, & Computing (ICSPCC 2012), which was held in Hong Kong in August 2012. Prof. Lam was the Chairman of the IEEE Hong Kong Chapter of Signal Processing between 2006 and 2008.

Currently, he is the VP-Member Relations and Development of the Asia-Pacific Signal and Information Processing Association (APSIPA) and the Director-Student Services of the IEEE Signal Processing Society. Prof. Lam serves as an Associate Editor of IEEE Trans. on Image Processing, Digital Signal Processing, APSIPA Trans. on Signal and Information Processing, and EURASIP International Journal on Image and Video Processing. He is also an Editor of HKIE Transactions. His current research interests include human face recognition, image and video processing, and computer vision.