Front. Comput. Sci., 2016, 10(6): 1118–1129
DOI 10.1007/s11704-016-5024-6

RESEARCH ARTICLE

Face recognition by decision fusion of two-dimensional linear discriminant analysis and local binary pattern

Qicong WANG, Binbin WANG, Xinjie HAO, Lisheng CHEN, Jingmin CUI, Rongrong JI, Yunqi LEI (✉)

Department of Computer Science, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

Abstract To investigate the robustness of face recognition algorithms under the complicated variations of illumination, facial expression and posture, the advantages and disadvantages of seven typical algorithms on extracting global and local features are studied through the experiments respectively on the Olivetti Research Laboratory database and the other three databases (the three subsets of illumination, expression and posture that are constructed by selecting images from several existing face databases). By taking the above experimental results into consideration, two schemes of face recognition which are based on the decision fusion of the two-dimensional linear discriminant analysis (2DLDA) and local binary pattern (LBP) are proposed in this paper to heighten the recognition rates. In addition, partitioning a face non-uniformly for its LBP histograms is conducted to improve the performance. Our experimental results have shown the complementarities of the two kinds of features, the 2DLDA and LBP, and have verified the effectiveness of the proposed fusion algorithms.

Keywords face recognition, global feature, local feature, linear discriminant analysis, local binary pattern, decision fusion

1 Introduction

Having a long history dated back to the late 1960s, the face recognition has been a hot research topic in the fields of computer vision and pattern recognition for several decades, and now it ever has more emerging applications in bioinformatics, video surveillance, security, human-computer interaction etc. Along with the significant progresses made in the past decades, the various features of geometry, global statistic (principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], shape and texture (local binary pattern (LBP) [3], Gabor [4]), have been proposed, and two-dimensional (2D) face recognition in controlled environment has become more or less mature. However, mainly due to the image's variations in illumination, facial expression and posture, the performances of recognizing a face under uncontrolled conditions, such as in photo album, surveillance videos, or social network image corpus, are not satisfactory. There are still challenging problems that are retained to be studied further.

To tackle the problems induced by the impacted negative factors, many approaches for face recognition of 2D images have been proposed, as described by the surveys [5, 6]. To handle varying illumination problems, the self quotient image was used in Ref. [7], while the gradient faces were used in [8]. In Ref. [9], by considering simultaneously both rows and columns, the two-directional 2DPCA was developed for efficient face representation and recognition, which can reduce much the coefficient set while keeping the recognition accuracy at the same level or even higher. Following the same idea in [9], [10] proposed a two-directional 2DLDA. An expression-invariant face recognition method based on

Received January 17, 2015; accepted November 12, 2015

E-mail: yqlei@xmu.edu.cn

distributed compressed sensing theory was presented in Ref. [11]. Sharma et al. [12] proposed a generic discriminative coupled latent subspace to deal with pose changes, their learned set of coupled subspaces projects the images of the same person under different poses to close locations in the latent space to make recognition possible by using simple learning. Also, the three-dimensional (3D) face models have been utilized to overcome the impacts of changes in posture and illumination. The work [13] proposed a way of extracting geometric feature from 3D face and conducted face recognition by PCA. Lei et al. [14] investigated 3D face recognition by using fisher linear discriminant (FLD) method on depth image and the speeded up robust features (SURF) of the 2D gray image. In Ref. [15], the Gaussian and mean curvatures of a 3D face were classified to achieve a surface classification image (SCI) and the SCI was then input to the process of PCA to obtain the SCI eigenfaces to recognize the 3D face. However, considering the difficulties of acquiring specific apparatus to collect 3D face data as well as the heavy computational load, the face recognition methods that are based on 2D images still have their own advantages.

It should be noted that a single feature cannot respond to many varieties under complex situations, it has been shown that both holistic and local information is crucial for the perception and recognition of 2D faces. Corresponding to the holistic information, the global features of a face refer to all the dimensions of its feature vector that contains the information reflecting the overall properties of the face, such as its contour, the relative locations of facial structure, and the form and size of a specific area. Complementarily, each dimension of the local features corresponds to a limited area on the human face image, such as the characteristics of facial moles, scars and dimples. This means that a local feature focuses on the details extracted from a human face. Those two kinds of features play different roles when describing the contents of a face, and both of them are necessary for recognition. Generally, the global features are used to identify the class that a face belongs to, while the local features are employed to discriminate the specific personal within a class. Many researchers have proposed methods of taking advantages of those two features to represent a human face. By directly fusing local and global features and applying Fisher's linear discriminate (FLD) method and the support vector machine (SVM), Chowdhury et al. [16] used reduced feature vector for face classification. Kim et al. [17] combined global and local features by applying an LDA based method, respectively, to the whole and the parts of a face image. A multilayer framework for high resolution face recognition was proposed by Ref. [18], they used multilevel PCA followed by regularized LDA to model global appearance and facial organs, and the discriminative multiscale texton features and the scale-invariant feature transform (SIFT) activated pictorial structure were exploited to describe the skin and subtle details. In Ref. [19], the global features were extracted from the whole face images by using Fourier transform, and the local features were emphasized on some spatially divided face patches by using Gabor wavelets. Geng et al. [20] integrated global and local features in a cascaded way, in which they firstly used the holistic approach of eigenfeature regularization and extraction to retrieve some candidate images from the whole gallery set, then a SIFT-based local feature extraction and matching algorithm was performed on the remains. A learning-based shape descriptor was designed in Ref. [21], where a strategy for feature division was established when encoding feature pooling and vocabulary learning, so a more discriminative descriptor was constructed by incorporating both global and local information while keeping high time efficiency. Zhang et al. [22] proposed newly an effective algorithm of multi-label learning with Label specIfic FeaTures (LIFT) to discriminate different class labels, where the clustering analysis on positive and negative instances were conducted, and training and testing by querying the clustering results then were performed. Using a kernel combination, Zhang et al. [23] combined three modalities of biomarkers to discriminate between Alzheimer's disease (or mild cognitive impairment) and healthy controls, their multimodal classification method performed considerably better compared to the case of using an individual modality. On the whole, the experimental results in above literatures indicate that the fusion methods of global and local features generally can improve recognition rates. It should be noted that our previous works [13–15] do not consider the fusion scheme when the work was done, anyway the findings there provide us the bases to look for other ways to improve performances, as indicated in this paper.

Some recent researches focus on learning based feature design, targeting at providing data-driven distance metric toward invariance measurement from varying illumination, facial expression and postures, or, in other words, capturing underlying face manifold from the original feature space. Recently, proposed in Ref. [24], to enhance the information content, features were extracted and combined at different resolutions to form a face recognition system. In Ref. [25], Gabor entropy weighted features and local normalization were considered for face identification. A neighborhood discriminant hashing (NDH) was provided to implement approximate similarity search by exploiting local discriminative information and by using the maximum entropy principle [26], where the learned hashing function was compact and the bits were highly informative. A novel robust structured subspace learning (RSSL) algorithm was proposed by integrating image understanding and feature learning into a joint learning framework [27]. An unsupervised algorithm for feature selection named clustering-guided sparse structural learning (CGSSL) was proposed in Ref. [28] by combining cluster analysis and sparse structural analysis into a joint framework. The multi-feature fusion and dictionary learning were integrated to construct a framework for face recognition [29]. In Ref. [30], based on local phase quantization, a blur-robust face image descriptor was provided and extended to a multiscale framework to achieve the robustness to blurring, and by using kernel fusion, the framework was also combined with a descriptor of multiscale local binary pattern to increase the robustness to illumination.

Notably in recent years, various deep learning methods have also been studied to apply for face recognition. In Ref. [31], a deep convolutional neural network which was trained layer-by-layer was used to help the network to converge, and a sample transformation method was proposed to avoid overfitting. In Ref. [32], learning a set of high-level feature representations through deep learning was studied for face verification, where the face features were extracted from various face regions to form complementary and over-complete representations. In Ref. [33], a deep architecture named AU-inspired deep networks (AUDN) was constructed and a multi-layer learning process was employed to build group-wise sub-networks for face higher-level representations. Yet, when using deep learning methods, there exist three issues that should be considered properly, these are: 1) collecting of representative and rich training samples, 2) the long time taken during training, and 3) avoiding of over-fitting.

However, despite the existed progress, the selection of global and local features that are relatively complementary to each other to enhance the face recognition performance is still a topic worthy to be studied further. Extending our earlier work [34] by providing more details in the algorithms, and adding more experiments and analysis, in this paper, we analyse and compare the effects of several typical types of the global features that are extracted based on PCA, LDA, 2DPCA, 2DLDA, DFT (discrete Fourier transform) and the local features that are based on LBP and Gabor, under various illumination, facial expression and posture conditions. After the advantages and disadvantages of the various types of the extracted features are discussed, the complementarity of the two kinds of features are demonstrated experimentally. We further demonstrate that, a way of non-uniform spatial partition of a face before LBP feature extraction is indeed beneficial and can further improve the recognition performance. Based on those results, we then propose two schemes for face recognition using the decision fusion of global feature 2DLDA and local feature LBP. Our experimental results show that the proposed decision fusion schemes can achieve better performance than those without the fusion. Besides, compared with the methods of Refs. [16–20], where only one or two global features and one local feature were considered, here we have analyzed and compared five global features and two local features, then the performance is improved further. Furthermore, since the performances of the proposed schemes are obtained without specifying the illumination, facial expression and posture, then the fusion schemes are suitable to be used under uncontrolled environment.

The rest of this paper is organized as follows. Section 2 describes the module diagram of face recognition and the preprocessing of face images. Section 3 provides the feature extraction based on 2DLDA and LBP, respectively, and presents two decision fusion ways for the recognition. The construction of experiment datasets, the experimental results of comparison and analysis for the seven algorithms and fusion methods are shown in Section 4. The conclusion follows in Section 5.

2 The modules of face recognition

A typical face recognition system is composed of four modules, including preprocessing of face image, feature extraction, feature mapping and face classification, as shown in Fig. 1.

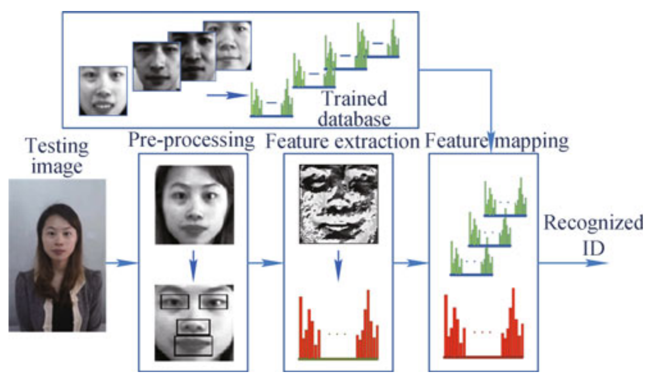


Fig. 1 Modules of face recognition

For face recognition, there are two main negative factors that affect the performance: illumination and posture. The variations of illumination and posture change the image matrix greatly and may lead the matrix features to be non-discriminant in the sense of face classification. Some examples are shown in Fig. 2 (how the illumination and posture could vary in a large scope). The tasks of illumination balance and posture registration should be fulfilled at the best effort in preprocessing.

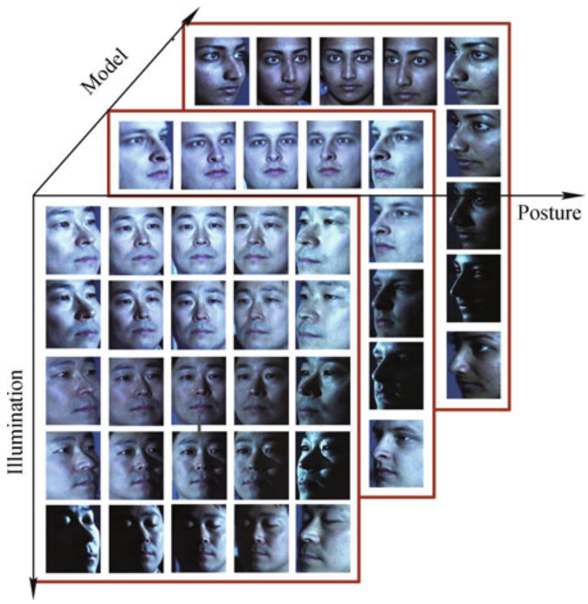


Fig. 2 Variations of illumination and posture

In the preprocessing, there are several steps that need to be done before its output is feeded into the module of feature extraction. Those steps include face detection, key points localization, pose identification and adjustment, equalization of grey scale histogram, and geometric normalization and registration, etc. When doing face recognition from a video, a step of face tracking is also necessary. A properly preprocessed face image is critical to improve the overall performance of recognition.

The feature extraction is the core of face recognition system. Here we demand to take out the features which should be sufficient to discriminate faces that belong to different people, while keeping the features that are belong to the same person robust. Generally, the dimension of feature is high and we need to reduce the dimension to a lower one, so as to not only speed up the recognition but also increase the correctly identified rate.

In feature mapping module, a classifier is designed to make decision about the identity of the input face image, where the selection of classifier is closely related to the properties of extracted features. There exist many classification methods, such as the nearest neighborhood classifier, and the SVM classifier based on kernel transformation.

The feature extraction and classification will be stated in the next section, here we provide the preprocessing module in some detail, and its diagram is shown in Fig. 3.

For an input image that may contain a face, the AdaBoost algorithm is used to cut out the area of face roughly, and the grey level of the area is adjusted by balancing its histogram to eliminate the illumination effects. The locations of the left and right pupils are determined by a dynamic threshold method. Then for the vertical strip below the eyes, horizontally we sum the grey values and project the values to the vertical axis to get a histogram. In this histogram, two valleys can be found and they are the centre locations of nose and mouth, henceforth the areas of nose and mouth can be determined. As long as the areas of eyes, nose and mouth are known, the outer strips that are irrelevant to face features, like hair or ears, can be deleted to get an area that contains the most part of the face.

In the procedures conducted above, the determination of the locations of pupils has great impact on the certainty for cutting out the face area. The dynamic threshold method we chose can be done fast simply for its less computation. It is based on the fact that, generally, the grey levels of pupils are lower than that of the inner corner of eye or that of the face, thus the pupils are easy to be identified. We have also used AdaBoost algorithm to determine the pupils, however, at the same level of assurance, AdaBoost algorithm takes much longer time to get a result.

3 Feature extraction algorithms

The global feature, 2DLDA, and the local feature, LBP, are

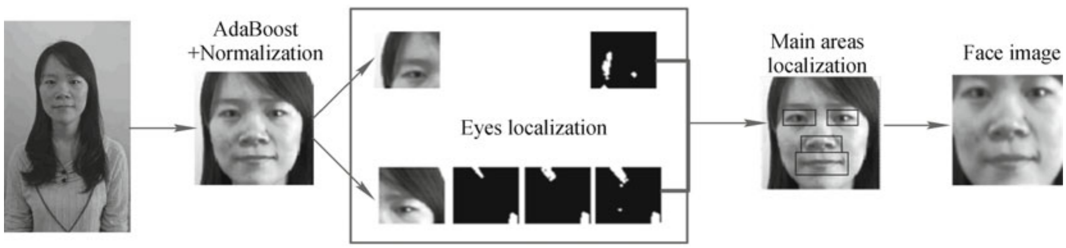


Fig. 3 The diagram of face preprocessing

discussed in this section in some detail, which will be used in our decision fusion schemes later.

3.1 2DLDA global feature extraction

The two dimensional linear discriminant analysis (2DLDA) algorithm [35–37], which originates from linear discriminant analysis, overcomes the singularity problem effectively. By taking the lower dimensional features, that are more discriminant, from the high dimensional feature space, it works on the matrix representation of face image directly and extracts the most discriminating facial features with the characteristic that the ratio of the between-class distance to the within-class distance reaches the maximum value.

When applying LDA to face recognition, it needs to rearrange the image matrix to form a high-dimensional vector, generally row by row. Different from LDA, 2DLDA is directly based on the image matrix and fundamentally it compresses the image matrix by row and then extracts the features. 2DLDA reduces or eliminates the correlation between columns while still keeps the correlation between rows, which are suitable for face recognition. Mainly, 2DLDA has the advantages of higher efficiency and faster speed when extracting the features.

The facial feature extraction process of the 2DLDA based on the lateral compression is as follows.

3.1.1 Preprocessing of the face images

We use preprocessing strategies first, such as geometric normalization and histogram equalization, to the input face data to eliminate the impact of various external factors on the images. The preprocessing follows the procedures stated in Section 2.

3.1.2 Computing of the optimal projection matrix

Denote T as the training data,

$$T = \{X_1^1, \ldots, X_{n_1}^1, X_1^2, \ldots, X_{n_2}^2, \ldots, X_1^C, \ldots, X_{nC}^C\}, \quad (1)$$

where C is the number of pattern classes, n_i is the number of samples in class i ($i = 1, 2, \ldots, C$), and X_j^i denotes an $m \times n$ image matrix of the jth sample of ith class. Suppose the mean image of ith class is represented by $\overline{X^i}$,

$$\overline{X^i} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^i, \quad (2)$$

and the mean image of all training samples is written by $\overline{X}$,

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{n_i} X_j^i, \quad (3)$$

where $N = \sum_{k=1}^{C} n_k$, then the within-class scatter matrix S_w and the between-class scatter matrix S_b are defined as:

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{n_i} \left(X_j^i - \overline{X^i}\right)^{\mathrm{T}} \left(X_j^i - \overline{X^i}\right), \quad (4)$$

$$S_b = \sum_{i=1}^{C} n_i \left(\overline{X^i} - \overline{X}\right)^{\mathrm{T}} \left(\overline{X^i} - \overline{X}\right). \quad (5)$$

Apparently, S_w reflects the overall scatter distributions that are relative to each class centre, and S_b reflects the scatter distance from each class centre to the centre of all classes. The purpose of 2DLDA is to find the optimal discriminant vector that can maximize a ratio, that is the between-class scatter distance over the within-class scatter distance. To find the vector, the Fisher criterion is applied, which is to conduct the following maximization:

$$J(W) = \arg\max_{W} \left(\frac{W^{\mathrm{T}} S_b W}{W^{\mathrm{T}} S_w W}\right), \quad (6)$$

where W is a matrix to be determined. In fact, the normalized eigenvector corresponding to the maximum eigenvalue of $S_w^{-1} S_b$ is just the optimal projection vector for simple classification. However, one projection vector is not enough for complicated situations when many classes exist, it is necessary to find a group of vectors to present the discriminative features for identifying multi-classes.

After doing eigenvalue decomposition on $S_w^{-1} S_b$, commonly, the eigenvectors, $v_1, v_2, ..., v_d$, that are corresponding to the largest d eigenvalues respectively, are selected and normalized to form the optimal projection matrix W_{opt}, that is,

$$W_{opt} = [v_1, v_2, ..., v_d]. \quad (7)$$

3.1.3 Computing of the feature matrix of face image

According to the linear transformation formula:

$$Y_j^i = X_j^i * W_{opt}, \quad (8)$$

where ($i = 1, 2, \ldots, C$), and ($j = 1, 2, \ldots, n_i$), we can obtain the feature matrix Y_j^i by projecting the face image X_j^i onto the optimal projection matrix W_{opt}. Consequently, the 2DLDA feature vector can be formed by concatenating each column of Y_j^i.

3.2 LBP local feature extraction

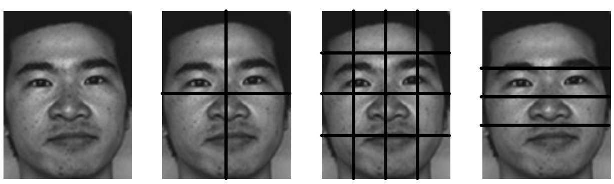
LBP has been widely used in texture analysis for image retrieval, face recognition, video processing, etc. The core idea of LBP is to set the center pixel value as the threshold, then, in comparison with the value of its neighbors, a binary code which describes the local texture feature is obtained. The facial feature extraction based on LBP operator works as follows.

3.2.1 Encoding of face image with LBP operator

The original LBP operator defines a 3×3 neighborhood window. For the central pixel in a single scanned window, its grey value is compared with all of its neighbors sequentially. When the grey value of a neighborhood pixel is greater than that of the central pixel, we mark it as "1", otherwise, "0". Arranging the marked values of all eight neighborhoods clockwise or counter-clockwise forms a binary string, this string then can be interpreted as an eight-bit binary integer. The decimal expression of the eight-bit binary integer is then defined as the LBP value of the central pixel. Scanning a given image pixel by pixel, its LBP feature map can be yielded after the above process is executed.

3.2.2 The spatially partitioned LBP histograms

After the calculation of LBP feature map, usually its histogram is used to describe the distribution of various patterns in the facial image, such as its edge, bright spots, dark spots and smooth areas. Nevertheless, there are different ways to consider the formation of histograms, either we can calculate the histogram by taking the feature map as a whole, or we can compute correspondingly the histograms by dividing the map into several blocks. When conducting the similarity measurement for classification, the impacts of the formation of histograms are certainly different. In order to express the spatial information of face's texture, we further propose to divide, non-uniformly, the LBP feature map into several local regions and then extract the regional LBP histograms from each of them. As the result, we extend the normal LBP histogram to the spatially partitioned LBP histograms.



(a) (b) (c) (d)

Fig. 4 Four allocation ways of face blocks

We take a variety of division modes into account when partitioning. Figure 4 shows four kinds of the blocks. Figures 4(a)–4(c) represent three uniform allocation ways, respectively. Figure 4(d) divides the face horizontally into four uneven parts ($\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}$) for the consideration of the areas of forehead, eyes, nose and mouth, respectively. Table 1 shows the rank 1 recognition rates for the four allocations (the number of bins in histogram is chosen as 128). The nonuniform division mode (d) achieves a higher recognition rate than the uniform division mode (b) which has the same feature dimension of 512 with (d). Compared with division mode (c), mode (d) gets a similar recognition rate, but its dimension is much less than the counterpart. Since the lower dimension of mode (d) improves the speed of feature extraction and the similarity calculation, naturally we adopt the division mode (d) in our algorithm.

Table 1 Comparisons of recognition rates for different partitioning on ORL

Division mode	Number of blocks	Feature dimension	Recognition rate (n images for training)		
			$n = 1$	$n = 3$	$n = 5$
(a)	1×1	128	0.550	0.750	0.785
(b)	2×2	512	0.701	0.900	0.905
(c)	4×4	2048	0.770	0.936	0.965
(d)	4	512	0.767	0.943	0.955

3.3 Decision fusion of 2DLDA and LBP

None of the single kind feature has the characteristics to handle the recognition task in all cases. Instead, we use two kinds of decision-level fusion, the weighted fusion and the score fusion, to utilize the 2DLDA feature and the LBP feature. It should be noted that for the 2DLDA feature we use Euclidean distance to measure the similarity between test samples and training samples, while for the LBP feature we use the histogram intersection. For LBP, we have conducted the recognition by three similarity measures: the Euclidean distance, χ^2 distance and the histogram intersection. Their definitions are, respectively,

$$E(X, Y) = \sqrt{\sum_i (x_i - y_i)^2}, \tag{9}$$

$$\chi^2(X, Y) = \sum_i (x_i - y_i)^2 / (x_i - y_i), \tag{10}$$

$$H(X, Y) = \sum_i \min(x_i, y_i), \tag{11}$$

where X and Y are two LBP histograms.

Table 2 shows the results of identification accuracy on the Olivetti research laboratory (ORL) dataset [38] when using the above three different measurements. Clearly, the Euclidean distance has less advantage, while the other two demonstrate similar results. However, the results of histogram intersection are still much better, and also it is more efficient in computation, so we choose the histogram intersection $H(X, Y)$ as our similarity measure for the LBP feature.

Table 2 Accuracy comparisons on ORL for three different similarity measures

Similarity measure	Measurement notation	Number of images used for training		
		$n = 1$	$n = 3$	$n = 5$
Euclidean Distance	$E(X, Y)$	0.744	0.904	0.920
χ^2 Distance	$\chi^2(X, Y)$	0.772	0.925	0.940
Histogram Intersection	$H(X, Y)$	0.767	0.943	0.955

• Weighted fusion Figure 5 shows the scheme of weighted fusion of 2DLDA global feature and LBP local feature. The decision fusion is defined as:

$$S = w * S_{2\mathrm{DLDA}} + (1 - w) * (1 - S_{\mathrm{LBP}}) \qquad (12)$$

where $S_{2\mathrm{DLDA}}$ and S_{LBP} stand for the normalized similarities of 2DLDA and LBP methods, respectively, and w is a weight factor between 0 and 1.

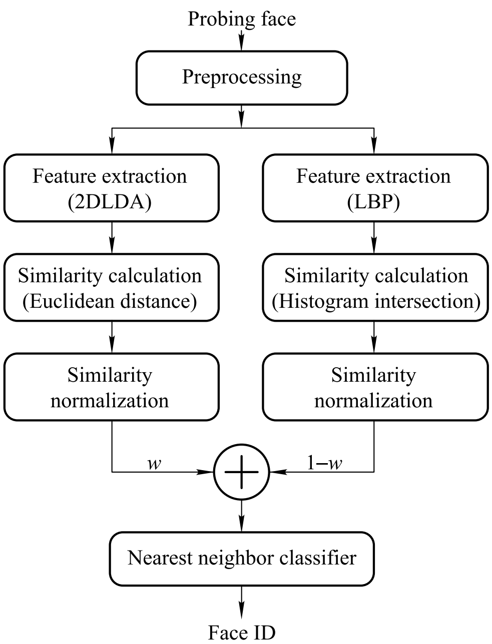


Fig. 5 Weighted fusion of 2DLDA and LBP

• Score fusion Figure 6 shows the scheme of score fusion. We calculate the similarities based on 2DLDA and LBP feature vectors respectively. Each method gets k most similar faces using the nearest neighbor (NN) classifier. Then we set the scores of the first k faces from k to 1 (k is chosen as 10 in our study). Finally, we accumulate the scores of the same face in both methods. The face with the highest score is considered to be the most similar face.

4 Experimental results

4.1 Face datasets

Firstly, we tested our approach on the ORL face dataset [38]. The dataset contains 10 different images of 40 distinct subjects. There are variations in illumination, facial expression (open/closed eyes, smiling/non-smiling), with/without glasses, and scale (variations of up to nearly 10%). All of the images were taken against a dark homogeneous background with the subjects in an upright, frontal position, with tolerance for some tilting and rotation of up to about 20°.

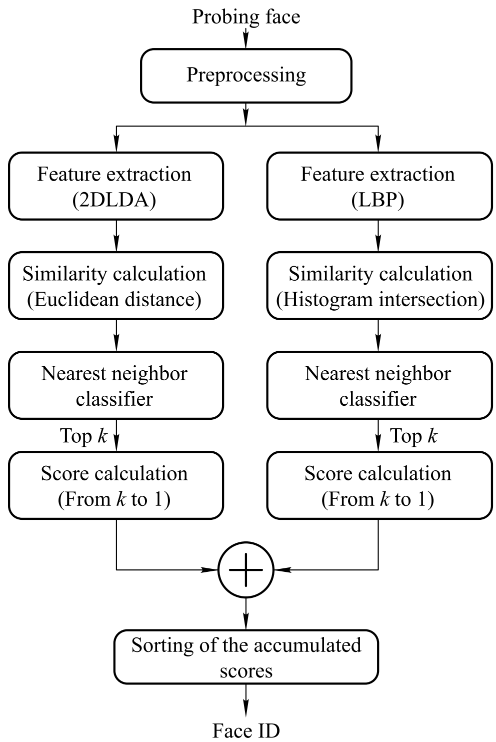


Fig. 6 Score fusion of 2DLDA and LBP

Additionally, to gain knowledge about the robustness of our methods against the complex illumination, facial expression and posture respectively, we have constructed three subsets: Illumination Subset S_1, Expression Subset S_2 and Posture Subset S_3. Illumination Subset S_1 is a subset of YaleB [39] face dataset, which contains 64 different illumination conditions of each subject in an upright, frontal position with neutral expression. Expression Subset S_2 is a subset of CASIA-FaceV1 [40] face dataset, which contains 11 different facial expressions in an upright, frontal position. We neglected the various illumination conditions in Expression Subset S_2 as it is tolerable. Posture Subset S_3 is a subset of PIE [41] face dataset, which contains seven different postures, using yaw angles (−67.5°, −45°, −22.5°, 0°, 22.5°, 45° and 67.5°), of each subject in an upright, frontal position with normal illumination. Here we have selected subsets from YaleB, CASIA-FaceV1 and PIE, instead of the whole databases, that those redundant face data are not included if the illumination, expression or posture are almost the same.

Actually, if we include the redundant data in our experiments, the performance is even a little bit better but it is not so clear to show the robustness of our fusions. For example, suppose we have ten samples to recognize and the recognition rate is nine out of ten, that is 90%. However, if we use eleven samples where two of them are very close to each other, then one of them should be excluded, otherwise, the recognition rate would be ten out of eleven, that is 90.9%, which is a result that does not reflect the actual performance.

Some examples from the four datasets are shown in Fig. 7, where the cutting and normalization have been done to all of the three subsets. The cutting can be obtained by automatic face detection methods like [42]. Table 3 shows the information about each of the four subsets.

4.2 Comparative experiments on various algorithms

The four experiments we conducted are described in the following. Seven algorithms, including PCA, LDA, 2DPCA, 2DLDA, DFT, LBP and Gabor, are considered.

Experiment 1 Compare the performances of the seven algorithms on ORL dataset (first five images per subject for training, the other five images per subject for testing).

Experiment 2 Compare the performances of the seven algorithms on Illumination Subset S_1. Randomly divide the 64 images of each subject into two even parts (one part as training set, the other part as testing set) for ten times, we calculate the average recognition rate of each algorithm in the ten rounds as the eventual performance.

Experiment 3 Compare the performances of the seven algorithms on Expression Subset S_2 (first two images per subject for training, the other nine images per subject for testing).

Experiment 4 Compare the performances of the seven algorithms on Posture Subset S_3 (seven rounds for cross validation).

• Recognition rate comparison Figures 8–11 show the cumulative match curves (CMC) of the four experiments. The curves indicate that none of the algorithms outperforms others under conditions of severe illumination, facial expression and posture variations.

The DFT algorithm outperforms others in Experiment 1 (Fig. 8). As the slight variation of illumination and facial expression in the ORL dataset affects only the local light intensity, which can be regarded as the high part of frequencies of the image, then eliminating of the high frequencies in the DFT algorithm performs better than others. The results of Experiment 2 (Fig. 9) and Experiment 3 (Fig. 10) show the extraordinary robustness of Gabor local feature on variations of illumination and expression. Contrary to the Gabor, the algorithms based on space projection (PCA and LDA) and global frequency domain features (DFT) are less robust.

• Posture variation problem As shown in Fig. 11, except for LBP and 2DLDA, all of the other algorithms show poor

Fig. 7 Samples of face datasets

Table 3 Information of datasets used in experiments

Face dataset	Source	Number of subjects	Number of images per subject	Image size
ORL	ORL	40	10	92×112
Illumination Subset S_1	YaleB	38	64	84×96
Expression Subset S_2	CASIA-FaceV1	123	11	96×96
Posture Subset S_3	PIE	68	7	96×96

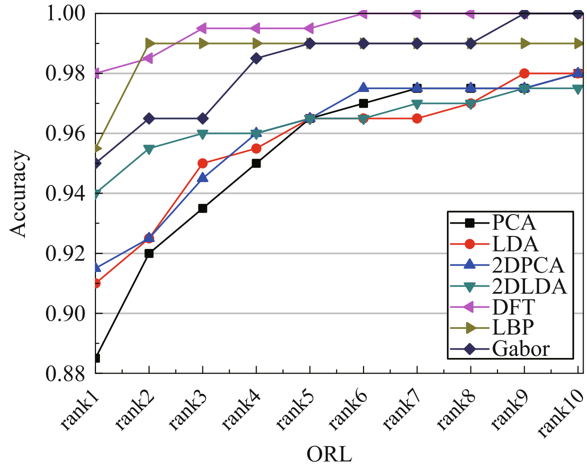


Fig. 8 CMC curves of the seven algorithms for Experiment 1

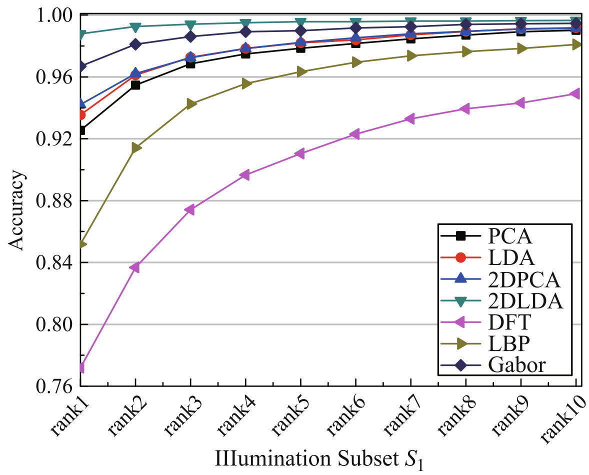


Fig. 9 CMC curves of the seven algorithms for Experiment 2

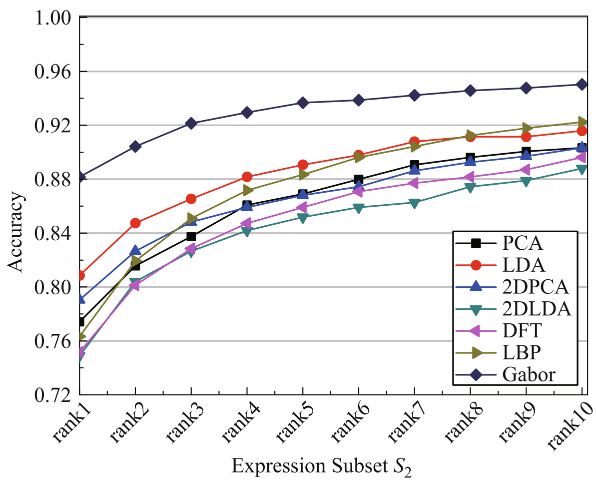


Fig. 10 CMC curves of the seven algorithms for Experiment 3

performances under the condition of severe posture variations. We own the relative better performances of LBP and 2DLDA to the retaining of spatial information through partitioning for LBP and working on matrix for 2DLDA.

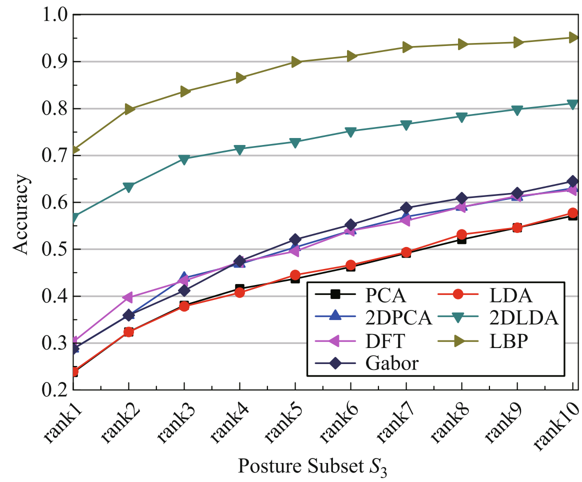


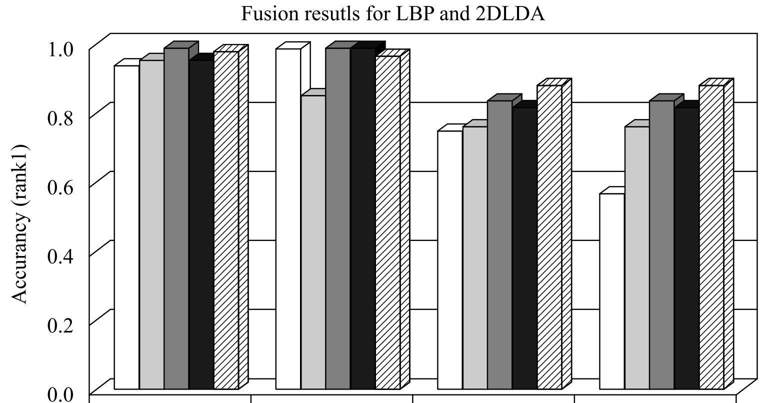
Fig. 11 CMC curves of the seven algorithms for Experiment 4

• Comparison of executing efficiencies To analyze the efficiencies of the seven algorithms, we have calculated the execution time taken by the three processes: feature extraction on training set, feature extraction on testing set and similarity computing. The training feature extraction of Gabor requires much more time than others, because the convolution operation on Gabor kernel and face image is inefficient. In the training process, PCA and 2DPCA need to do eigenroot decomposition on covariance matrix, the execution time of the former depends on the number of training subjects, while that of the latter depends on the width of the images. Therefore, as the training set grows, PCA executes much longer than 2DPCA. Also, compared with PCA, 2DPCA only operates one directional (horizontally or vertically) compression, which is low efficient in similarity computing process. The relationship between LDA and 2DLDA is similar to that of PCA and 2DPCA. Moreover, in the methods of DFT, LBP and Gabor, training and testing processes cost similar time as they take the same processes correspondingly.

4.3 Results of desicion fusion experiments

Based on the results stated above, we provide fusion experiments of LBP and 2DLDA. Like the processes of experiments in subsection 4.2, the weighted and score fusion algorithms of LBP and 2DLDA have been tested on each dataset. In the experiments, we have tested various values of weight w. When $w = 0.5$, the fusion algorithms on each dataset achieve better recognition results than other values.

Figure 12 shows the five recognition rates (rank 1) of LBP, 2DLDA, the two decision fusions, and the single feature with the highest recognition rate among all the seven algorithms, respectively. The results show that both weighted and score fusion schemes outperform the single method using either



	ORL	Illumination S_1	Expression S_2	Posture S_3
□ 2DLDA	0.94	0.988	0.75	0.569
□ LBP	0.955	0.852	0.763	0.712
■ Weighted fusion	0.99	0.99	0.837	0.771
■ Score fusion	0.955	0.992	0.818	0.769
▨ Highest rate	0.98	0.967	0.882	0.712

Fig. 12 Recognition rate (rank 1) of the feature fusion experiments

LBP or 2DLDA. Except for Expression Subset S_2, the weighted fusion algorithm gets a higher recognition rate than the best of all the seven algorithms that are with single feature. Moreover, the weighted fusion outperforms the score fusion in all of the datasets except for Illumination Subset S_1. The experimental results indicated above verify the complementarities of the two features, LBP and 2DLDA. Averagely considering the recognition rates on the four datasets (although the weighted fusion is a little bit better than the score fusion, it is not so reasonable to deny the later one), we would recommend to consider both fusions when doing recognition.

5 Conclusion

Two schemes of face recognition based on the decision fusion of global feature 2DLDA and local feature LBP are proposed in this paper. After considering the advantages and disadvantages of various types of algorithms on extracting global and local features, we have demonstrated the complementarity of the two features and the effectiveness of our schemes through the experiments on the ORL database and the other three subsets. For LBP, since the uneven partitions are helpful to retain the spatial information of a face, the partitions are beneficial to improve the recognition rate. Both the two fusion algorithms have better performances than the single method using LBP or 2DLDA. The weighted fusion algorithm gets the highest recognition rate among all the algorithms mentioned above on all the four datasets, except for Expression Subset S_2 on which Gabor performs best.

However, the proposed algorithms need offline training in 2DLDA, which may affect the scalability of the algorithms. The results in Fig. 11 show that looking for methods to improve the recognition rate and efficiency under the situation of complicated postures remains a task for further research, and using of 3D data should be a promising way.

Acknowledgements The research described in this paper was supported by the Key Project Funds for Industrial Technology of Fujian, China (2012H6024) and the XMU-NU Joint Strategic Partnership Fund. A preliminary version of this paper was appeared at the 4th IEEE International Conference on Information, Intelligence, Systems and Applications (IISA), Greece, July 10–12, 2013.

References

1. Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991, 3(1): 71–86
2. Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711–720
3. Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037–2041
4. Shen L L, Bai L. A review on Gabor wavelets for face recognition. Pattern Analysis and Applications, 2006, 9(2): 273–292
5. Makwana R M. Illumination invariant face recognition: a survey of passive methods. Procedia Computer Science, 2010, 2(1): 101–110
6. Murphy-Chutorian E, Trivedi M M. Head pose estimation in computer vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4): 607–626

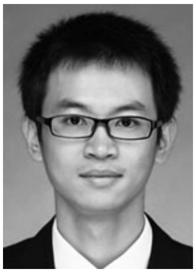
7. Wang H, Li S Z, Wang Y. Face recognition under varying lighting conditions using self quotient image. In: Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. 2004, 819–824
8. Zhang T, Tang Y Y, Fang B, Shang Z, Liu X. Face recognition under varying illumination using gradientfaces. IEEE Transactions on Image Processing, 2009, 18(11): 2599–2606
9. Zhang D, Zhou Z H. $(2D)^2$PCA: two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing, 2005, 69(1): 224–231
10. Noushath S, Kumar G H, Shivakumara P. $(2D)^2$LDA: an efficient approach for face recognition. Pattern Recognition, 2006, 39(7): 1396–1400
11. Nagesh P, Li B X. A compressive sensing approach for expression-invariant face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009, 1518–1525
12. Sharma A, Al Haj M, Choi J, Davis L S, Jacobs D W. Robust pose invariant face recognition using coupled latent space discriminant analysis. Computer Vision and Image Understanding, 2012, 116(11): 1195–1110
13. Lei Y Q, Lai H B, Li Q M. Geometric features of 3D face and recognition of it by PCA. Journal of Multimedia, 2011, 6(2): 207–216
14. Lei Y Q, Lai H B, Jiang X T. 3D face recognition by surf operator based on depth image. In: Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology. 2010, 240–244
15. Lei Y Q, Chen D J, Yuan M L, Li Q M, Shi Z X. 3D face recognition by surface classification image and PCA. In: Proceedings of the 2nd International Conference on Machine Vision. 2009, 145–149
16. Chowdhury S, Sing J K, Basu D K, Nasipuri M. Face recognition by fusing local and global discriminant features. In: Proceedings of the 2nd International Conference on Emerging Applications of Information Technology. 2011, 102–105
17. Kim C, Oh J O, Choi C H. Combined subspace method using global and local features for face recognition. In: Proceedings of the IEEE International Joint Conference on Neural Networks. 2005, 2030–2035
18. Lin D H, Tang X O. Recognize high resolution faces: from macrocosm to microcosm. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006, 1355–1362
19. Su Y, Shan S G, Chen X L, Gao W. Integration of global and local feature for face recognition. Journal of Software, 2010, 21(8): 1849–1862
20. Geng C, Jiang X. Fully automatic face recognition framework based on local and global features. Machine Vision and Applications, 2012, 24(3): 537–549
21. Bai X, Rao C, Wang X G. Shape vocabulary: a robust and efficient shape representation for shape matching. IEEE Transactions on Image Processing, 2014, 23(9): 3935–3949
22. Zhang M L, Wu L. Lift: multi-label learning with label-specific features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107–120
23. Zhang D Q, Wang Y P, Zhou L P, Yuan H, Shen D G. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage, 2011, 55(3): 856–867
24. Pong K H, Lam K M. Multi-resolution feature fusion for face recognition. Pattern Recognition, 2014, 47(2): 556–567
25. Cament L A, Castillo L E, Perez J P, Galdames F J, Perez C A. Fusion of local normalization and Gabor entropy weighted features for face identification. Pattern Recognition, 2014, 47(2): 568–577
26. Tang J H, Li Z C, Wang M, Zhao R Z. Neighborhood discriminant hashing for large-scale image retrieval. IEEE Transactions on Image Processing, 2015, 24(9): 2827–2840
27. Li Z C, Liu J, Tang J H, Lu H Q. Robust structured subspace learning for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2085–2098
28. Li Z C, Liu J, Yang Y, Zhou X F, Lu H Q. Clustering-guided sparse structural learning for unsupervised feature selection. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(9): 2138–2150
29. Wang D H, Wang X K, Kong S. Integration of multi-feature fusion and dictionary learning for face recognition. Image and Vision Computing, 2013, 31(12): 895–904
30. Chan C H, Tahir M A, Kittler J, Matti P, Pietikainen M. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5): 1164–1177
31. Wang W H, Yang J, Xiao J W, Li S, Zhou D X. Face recognition based on deep learning. Lecture Notes in Computer Science, 2015, 8944: 812–820
32. Sun Y, Wang X G, Tang X O. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 1891–1898
33. Liu M Y, Li S X, Shan S G, Chen X L. AU-inspired deep networks for facial expression feature learning. Neurocomputing, 2015, 159(1): 126–136
34. Wang B B, Hao X J, Chen L S, Cui J M, Lei Y Q. Face recognition based on the feature fusion of 2DLDA and LBP. In: Proceedings of the 4th International Conference on Information, Intelligence, Systems and Applications. 2013, 155–158
35. Ye J P, Janardan R, Li Q. Two-dimensional linear discriminant analysis. Advances in Neural Information Processing Systems, 2004, 17: 1569–1576
36. Wang X M, Huang C, Fang X Y, Liu J G. 2DPCA vs. 2DLDA: face recognition using two-dimensional method. In: Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence. 2009, 357–360
37. Yu W X, Wang Z Z, Chen W T. A new framework to combine vertical and horizontal information for face recognition. Neurocomputing, 2009, 72(4): 1084–1091
38. Samaria F S, Harter A C. Parameterisation of a stochastic model for human face identification. In: Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision. 1994, 138–142
39. Lee K C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 684–698
40. Xu C H, Tan T N, Wang Y H, Quan L. Combining local features for robust nose location in 3D facial data. Pattern Recognition Letters, 2006, 27(13): 1487–1494
41. Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression (PIE) database. In: Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition. 2002, 46–51
42. Stathopoulou I O, Tsihrintzis G A. Appearance-based face detec-

tion with artificial neural networks. Intelligent Decision Technologies, 2011, 5(2): 101–111

Qicong Wang received the PhD degree from Zhejiang University, China in 2007. He is currently an associate professor at the Department of Computer Science, Xiamen University, China. His research interests include pattern recognition, machine learning and computer vision.

Binbin Wang received the BE degree and the ME degree in 2010 and 2013, respectively, both from Xiamen University, China. She is currently a software engineer at the center of software development of Industrial and Commercial Bank of China, Guangdong.

Xinjie Hao received the BE degree and the ME degree in 2010 and 2013, respectively, both from Xiamen University, China. He is currently a postgraduate student majoring in computer science at Worcester Polytechnic Institute, USA. His research interests include machine learning and image processing.



Lisheng Chen received the BE degree and the ME degree in 2010 and 2013, respectively, both from Xiamen University, China. He is currently a technical manager at the department of Internet products of Hebao Financial Information Consultant Co. Ltd., Shenzhen, China.

Jingmin Cui received the BE degree and the ME degree in 2010 and 2013, respectively, both from Xiamen University, China. She is currently a system analyst (SA) at the Research and Development Center (Shenzhen), the Bank of Communications, China.

Rongrong Ji received the PhD degree from Harbin Institute of Technology, China. He had been a Postdoctoral Research Fellow in Columbia University, USA. Currently he is a Minjiang Chair Professor of Xiamen University, China. His research interests include machine learning and computer vision.

Yunqi Lei received the BE degree from the University of Science and Technology of China, China in 1982, and the PhD degree from the National University of Defense Technology, China in 1988. Currently, he is a professor at the Department of Computer Science, Xiamen University, China. His research interests include machine learning and computer vision.