https://doi.org/10.1016/j.jvcir.2019.102590

Learning Sparse Discriminant Low-Rank Features for Low-Resolution Face Recognition

M. Saad Shakeel*, Kin-Man Lam, and Shun-Cheung Lai

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

ABSTRACT

In this paper, we propose a novel approach for low-resolution face recognition, under uncontrolled settings. Our approach first decomposes a multiple of extracted local features into a set of representative basis (low-rank matrix) and sparse error matrix, and then learns a projection matrix based on our proposed sparse-coding-based algorithm, which preserves the sparse structure of the learned low-rank features, in a low-dimensional feature subspace. Then, a coefficient vector, based on linear regression, is computed to determine the similarity between the projected gallery and query image's features. Furthermore, a new morphological pre-processing approach is proposed to improve the visual quality of images. Our experiments were conducted on five available face-recognition datasets, which contain images with variations in pose, facial expressions and illumination conditions. Experiment results show that our method outperforms other state-of–the-art low-resolution face recognition methods in terms of recognition accuracy.

Keywords—Face recognition, feature fusion, local features, low rank approximation, linear regression, sparse coding.

1. Introduction

Practical face recognition systems often deal with low-resolution (LR) or poor-quality images, with huge variations in pose, facial expressions and lighting conditions. These make the low-resolution (LR) face recognition task very challenging. Conventional face recognition methods [1-3] work well for images which are of good quality and taken under controlled environments. Recently, a recognition rate higher than 99% has been achieved in recognizing high-resolution (HR) face images, taken under unconstrained environments, by using deep learning. However, for LR face recognition, the face images contain limited information that can distinguish one from the other. This results in a significant decline in recognition rates, when the face resolution is low. Moreover, the face images enrolled in a gallery

* Corresponding author
Email addresses: 15902620r@connect.polyu.hk (M. Saad Shakeel)
enkmlam@polyu.edu.hk (K.-M. Lam)
shun-cheung.lai@connect.polyu.hk (Shun-Cheung Lai)

set are usually of high resolution, so a dimension mismatch problem occurs. The lack of resolution-robust features, misalignment of faces, and noise effects are the major challenges in LR face recognition. This makes LR face recognition a challenging and hot research topic for researchers in the field of computer vision and pattern recognition. Before investigating the LR face recognition problem, it is necessary to determine the minimum resolution at which conventional face recognition methods can retain their performance, but will drop rapidly when the resolution is further decreased. We have found that this minimum resolution is about 32×24. A lot of research is being conducted to tackle the abovementioned challenges, and handling all of them simultaneously is very challenging and requires significant attention. One straightforward solution is to reconstruct a HR image from a given LR probe image using a super-resolution (SR) technique, and the super-resolved face image is then used for recognition. Another way is to learn a common low-dimensional subspace for the gallery and probe images, where accurate classification can be performed. The third possible solution is to downsample the HR gallery images to the same resolution as the LR probe image for matching in the LR domain. Fig. 1 summarizes these three possible approaches for solving the LR face recognition problem.

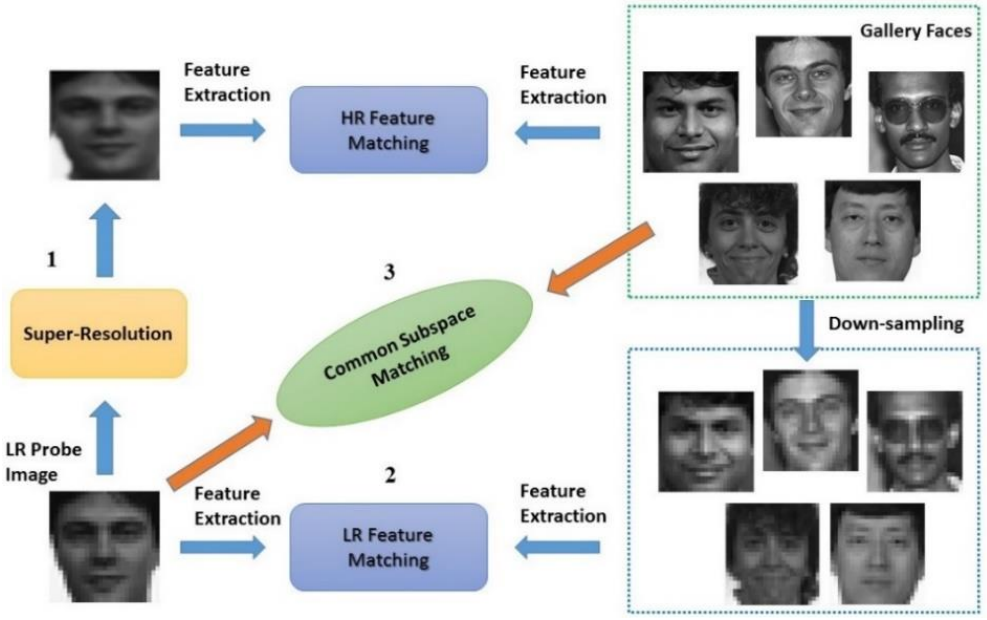


Fig. 1. Three possible approaches for LR face recognition.

In the next section, we will first give an overview of existing methods for LR face recognition. Then, the problems to be tackled by our proposed algorithm are presented and the way to solve them is briefly explained. Finally, the motivation and justification of the techniques used in our algorithm are elaborated.

2. Related Works

2.1 Super-resolution techniques

Super-resolution (SR) has been considered as one of the possible solutions, which increases the image resolution before performing recognition. For face images, the SR process is also known as face hallucination, which was first proposed in [4]. The method decomposes a face image into a pyramid of features by utilizing the Gaussian and Laplacian pyramids, and then reconstructs the corresponding high-resolution (HR) image. In [5], the limitations of SR were discussed and some possible solutions to breaking them were given. A SR method based on sparse coding was proposed in [6], which first computes the sparse representation coefficients of each LR image patch from a dictionary, and then uses those coefficients to generate the corresponding HR image patch. The similarity between the HR and LR image pairs is enhanced by jointly training the dictionaries, by using LR and HR image patch pairs. Wang et al. [7] utilizes principal component analysis (PCA) to linearly represent LR test image in terms of similar LR training images. The HR image is constructed, by replacing the LR training images with the corresponding HR training images. The SR methods in [8, 9] assume that the LR and HR images of the same person have some intrinsic correlation. Another face-hallucination framework presented in [10] assumes that two face images of the same identity have similar local-pixel structures. The approach learns the local-pixel structure for reconstructing a HR image by searching a face database for similar HR faces using the LR input image.

It was reported in [11, 12] that super-resolved images contain distortion and artifacts, which reduce the recognition accuracy, and hence are not a feasible solution to LR face recognition. Hennings-Yeomans et al. [13] proposed an objective function for performing hallucination and recognition simultaneously. This approach is computationally expensive, because optimization is required for each test image. Huang et al. [14] presented a SR method, based on non-linear mappings of coherent features. The method used Canonical Correlation Analysis (CCA) to create a coherent subspace between the PCA-based features of HR samples and LR samples. Radial basis function (RBF) is then used to learn the nonlinear mappings between the coherent features, and the super-resolved coherent features of a LR image are determined by using a trained RBF model. Zou et al. [15] proposed a framework for face hallucination, which learns a mapping function between the HR and LR image spaces by utilizing a new discriminative constraint. Jian et al. [16] proposed an improved method for performing hallucination and recognition simultaneously by utilizing the singular value properties of images at different resolutions. To recognize LR face images, only the super-resolved HR

features are required. Another way is to extract features from super-resolved HR images. However, these images, which are distorted versions of the true HR face images, are generated by estimation. Pong et al. in [17] proposed an approach, which directly estimates the HR features for recognition by performing super-resolution in the feature space. The method also fuses the features from different resolutions, so as to further improve the recognition accuracy.

2.2. Subspace learning methods

Another approach for recognizing LR face images is based on coupled mappings. Li et al. [11] proposed a coupled-mapping technique, which projects the LR and HR images into a unified feature subspace, which facilitates the ultimate classification. Zhou et al. [18] presented a method based on simultaneous discriminant analysis, which preserves the discriminative power of the HR and LR samples in the learned common subspace. Ren et al. [12] learned a common subspace for LR and HR samples using coupled kernel embedding, which uses a new similarity measure to compare the multimodal data. Biswas et al. [19] used multidimensional scaling for LR face recognition, which projects LR and HR samples into a common subspace in such a way that the geometrical structure of the samples is preserved. This approach also ensures that the distance between two LR images is nearly the same as that of its HR counterparts. Siena et al. [20] proposed a method to project HR gallery images and LR probe images into a common subspace based on their local structure's relationship. Similarly, Shi et al. [21] projects HR gallery images and LR probe images into a unified latent subspace by incorporating the geometrical structure of each given sample with respect to its neighboring points. The approach combines all the local optimizations to construct a global structure, which preserves the discriminant information of the samples in the learned subspace. Zhang et al. [22] proposed to learn a projection matrix, which maximizes the margin between inter-class and intra-class distances in the common subspace. Wang et al. [23] used CCA to determine the correlation between HR and LR image pairs, such that a pair of transform matrices are computed for the HR and LR face images, respectively. Jiang et al. [24] addressed the LR face recognition problem by proposing a coupled discriminant method based on multi-manifold analysis. The approach learns both the local structure and the neighborhood information about the manifold subspace covered by the image samples. After that, two mapping functions are learned to project the HR and LR samples, respectively, into a unified feature subspace. Chu et al. [25] proposed a cluster-based method, based on simultaneous discriminant analysis. This method learns the cluster-based scatter matrices, which is then used to regularize the between-class and within class scatter matrices. This enhances the discriminability of the feature space. Xing et al. [26] proposed a method based on coupled mappings,

which projects face-image samples into a unified feature subspace using the topology of a generalized bipartite graph. The approach also preserves the local geometrical structure of the samples when they are projected into a new subspace. Recently, Yang et al. [27] proposed a discriminative multi-dimensional scaling (DMDS) method for LR face recognition, which considers the intra-class, as well as inter-class, distances, while projecting the HR and LR data samples into a unified feature subspace.

2.3. Deep learning-based methods

Recently, deep learning has gained a lot of attention due to its excellent performance in many computer-vision tasks, such as image classification, object detection, recognition, etc. For face recognition, various deep learning models [28-32] have been established, and provided excellent performance. Sun et al. [28] proposed a deep convolutional neural network (CNN), namely DeepID, which learns high-level deep features from the patches of face regions for identification. Another deep face model, namely FaceNet [29], which uses a large network trained by distance constraints, was proposed. This model achieves a very high recognition accuracy of 99.60% on the challenging dataset, namely Labeled Faces in the Wild (LFW). Recently, Liu et al. [32] proposed an angular softmax loss function, which learns discriminative features using a CNN based on the ResNet architecture. The method achieves 99.42% accuracy on the LFW dataset. However, images captured by surveillance cameras are usually of low resolution, and require significant upscaling to serve as an input for the deep-CNN models. The upscaled face images have poor quality, because the original LR face images contain less amount of information. In this paper, we have also investigated the performance of LR face recognition, by using deep-learning-based face recognition methods, for face images at different resolutions.

2.4 Problem statement

Performances of face-recognition methods heavily depend on the amount of discriminant, robust features that can be extracted from face images, which exhibits as high-frequency components. However, when the image resolution decreases, the information available for distinguishing faces becomes less, and therefore SR becomes ineffective under unconstrained variations. This is because reconstruction from low-frequency components is an ill-posed problem and creates artifacts in the super-resolved images. Performances of the recognition-based and feature-based super-resolution methods depend on the feature extraction technique and reconstruction regularization model being used, but it is still unclear which regularization methods are optimal from the recognition perspectives. In addition to that, the

choice of features, which need to be able to handle large amount of variations in unconstrained environments, is critical. The previous methods, based on coupled mapping, do not extract the robust features from face images, but directly projecting the face samples into a unified feature subspace. This reduces the recognition accuracy in uncontrolled settings, especially under pose and illumination variations. Almost all the learning-based approaches consider using a subspace formed by linear projection to determine the similarity between the gallery and query samples, without examining some of the important properties of the samples.

Local features are more robust to pose and illumination variations, as compared to global features. Bereta et al. [33] presented a detailed comparative analysis of the performances of different local feature descriptors for face recognition. The advantages of using feature-level fusion for improving face recognition accuracy were also discussed. Lei et al. [34] proposed a novel texture descriptor for recognizing face images using the Fourier transformation of local binary pattern (LBP). For the descriptor, discriminant image filters are used for learning, and dominant patterns are constructed. Yoo et al. [35] proposed to decompose the LBP feature into a number of bit-planes, which not only contain the local structure information of face images, but also the method is proven to be robust to illumination variations. All the decomposed bit-planes are then combined to form a high-dimensional feature vector for face recognition. Recently, a robust local feature descriptor [36], based on shearlet transform, was proposed, which extracts useful information from face images by analyzing their singular structures. The estimated shearlet coefficients are then combined with LBP for face representation.

Methods based on coupled mappings and super-resolution are robust to only one variation. Robustness can be improved by using a combination of local features. Now, the question is which features should be selected and combined? In this paper, we address this question by fusing two local feature descriptors, which are robust to various facial variations. Recently, low-rank approximation techniques [37,38] have achieved great success in solving the occluded face recognition problem, by decomposing face images into a set of representative bases (the low-rank matrix) and the corresponding sparse error matrix. It is observed that the learned low-rank matrix has better representation capability as compared to the original data matrix. Moreover, the low-rank matrix can also alleviate the effect of illumination variations, which improves the recognition accuracy. Inspired by this, we utilize the low-rank matrix-decomposition algorithm to convert the extracted fused features into a low-rank matrix and a corresponding error matrix (sparse in nature). For recognition, we only utilize the low-rank component, while discarding the sparse error

matrix. Since the discarded sparse error matrix represents noisy components, so the recovered low-rank matrix provides better feature representation for face recognition. Furthermore, our approach first down-samples gallery faces to the size of the query image, and perform recognition in the low-resolution domain. Although the super-resolved face image or feature contains more information, the estimated information may be incorrect and distorted. By downsampling the gallery faces, no prediction is necessary. In this paper, we will investigate all the drawbacks of existing LR face recognition methods, and then propose an effective solution for tackling all the variations simultaneously, by utilizing sparse coding on multiple robust low-rank local features for LR face recognition.

2.5 Motivation

Inspired by the applications of sparse representation [39,40] to pattern recognition, we propose a new approach for LR face recognition based on sparse coding of multiple low-rank local features. The major idea is to compute an optimum sparse matrix, which projects the gallery and query low-rank features onto a common low-dimensional subspace. Sparse coding provides natural discriminant power and represents face images in a compact manner. A linear relationship exists between a test sample and the other training samples of the same subject. Matching a HR gallery image with a LR probe image has the dimension-mismatch problem, which also produces noise when projecting the LR and HR samples into a unified subspace. Our proposed method first down-samples all HR gallery images to the same resolution as the LR probe (test) image, and performs recognition in the LR domain. In our proposed approach, we assume that two images of the same resolution have higher correlation as compared to having two different resolutions. This is because the low-dimensional features are effective in computing precisely both the within-class and between-class scattering matrices, as their dimension is lower than, or closer to, the total number of samples available. In [15], it has also been argued that downsampling both the training and testing images can increase the recognition rate, even for images of very low resolution, such as 6×6 pixels. Therefore, downsampling face images is a feasible approach to solve the dimension-mismatch problem.

According to the review discussed previously, SR algorithms are not feasible for recognition purposes. This is due to the generation of artifacts in the super-resolved images, which reduce the recognition accuracy. To overcome this problem, a new morphological pre-processing approach based on top and bottom-hat filtering is proposed, which improves image quality, without generating any kind of distortion or artifacts in the final processed image. To make our approach robust to variations in unconstrained environments, two local features, Gabor wavelets and Local Binary

Pattern Difference (LBPD), are extracted from both the training and testing face images. Extracted features are normalized and then concatenated to form a final normalized feature vector. The normalized, fused features are then used to learn a low-rank matrix and a sparse error matrix, using an augmented lagrangian method. The extracted low-rank matrix is then projected into a new low-dimensional feature subspace by computing a projection matrix based on our proposed algorithm, such that the sparsity of the learned features is preserved. After that, the similarity between the gallery and query features is determined by estimating a coefficient vector using linear regression [41]. Based on the coefficient vector, residuals are computed for feature matching. To increase the discriminability between the face images of two different subjects, class-label information is utilized. Furthermore, our proposed objective function does not need to tune the model parameters, especially the neighborhood size. Our method has less computational complexity than other linear and nonlinear mapping-based methods [42,43], and can estimate local structures of face images by utilizing the sparse prior knowledge. Extraction and fusion of multiple low-rank local features make our method effective for recognition in unconstrained environments.

This work is an extension of our preliminary work [44]. The new contributions of this paper are as follows:

- A more compact and discriminative feature representation is learned by fusing the Gabor features and the Local binary pattern difference feature (a numerical variant of LBP). The extracted local features are converted into a set of representative basis (low-rank) and a sparse error matrix. The learned low-rank features are then used to learn a projection matrix based on sparse coding, which increases the recognition accuracy up to a significant level, as reported in the experiment sections.
- Extensive experiments are conducted with detailed analysis. Compared to our previous paper [44], we evaluate our proposed method with respect to various parameters, such as image resolutions, feature dimensions, etc.
- Some new additional challenging datasets, namely the Multi-PIE [45], FERET [46], LFW [47], and Remote face dataset [48] are used to evaluate the performance.

The rest of the paper is structured as follows. Section 3 introduces our proposed morphological pre-processing method, then Gabor wavelets and LBPD features are described. Section 4 explains the concept of low-rank feature learning. Section 5 presents our proposed framework based on sparse coding. Section 6 introduces the linear regression model used for classification. Section 7 demonstrates the experimental results and compares our method with other LR face recognition methods. Finally, the paper is concluded in Section 8.

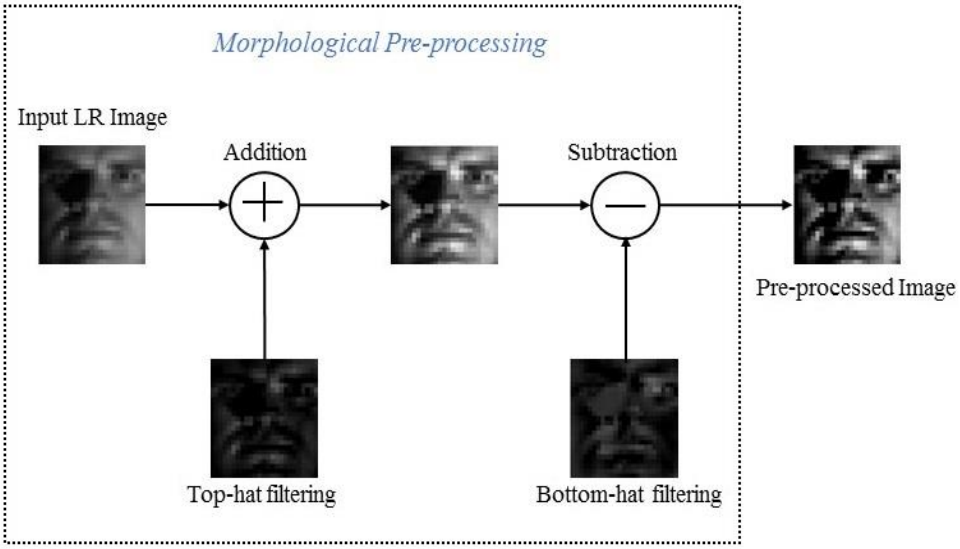


Fig. 2. The morphological pre-processing steps, on a LR face image.

3 Pre-Processing and Feature Selection

For face recognition, face images are first aligned and normalized to a specific resolution, and are usually pre-processed so that they become more standardized. This can help improve the recognition rate. In this section, we will describe our proposed morphological pre-processing method, and the features selected for our proposed algorithm.

3.1 Morphological pre-processing

To solve the problem of artifacts and distortion produced by SR algorithms, a novel morphological pre-processing method based on top and bottom-hat filtering is proposed. This method not only extracts useful information from face images, but also eliminates low-contrast features. It can also alleviate the effects of non-uniform illumination, so it is suitable for tackling variations in lighting conditions.

Let I be a grayscale image, and s be a disk-shaped structuring element with radius R. The first step is to apply the top-hat filtering which is defined as $T_t(I) = I - I \circ s$, where $\circ$ represents the opening operation. It can extract bright features from an image. Similarly, dark features are extracted by bottom-hat filtering defined as $T_b(I) = I \bullet s - I$, where $\bullet$ represents the closing operation. To enhance the local contrast for better image understanding, the face image is added to the difference between the two filtering outputs. Mathematically, it is given as follows:

$$I_{CE} = I + T_t(I) - T_b(I), \tag{1}$$

where $T_t(I)$ and $T_b(I)$ represent the top and bottom-hat filtering images, respectively, and I_{CE} is the contrast-enhanced

face image. Fig. 2 shows a LR face image with uneven illumination and the corresponding contrast-enhanced image generated by the proposed filtering operation.

3.2 Gabor Wavelets

Local features have been proven to be robust to variations in facial expressions, pose and illumination. From the biological point of view, Gabor functions can be used to model the responses of the cells in the visual cortex of mammalian brains. This feature exploits local regions to extract information at multiple scales and orientations. It is an efficient local feature descriptor, and has been used in many applications, such as object tracking, detection, and recognition. Gabor wavelets (GWs) can be used for spatial-frequency analysis because they have both the multi-orientation and multi-resolution properties, which enables it to provide valuable information about the local structure of an image. GWs achieve optimal representation in the frequency domains. It is defined as a complex exponential modulated by a Gaussian function, which is written as follows:

$$\phi_{\omega,\theta}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} . \left[e^{i(\omega x cos\theta + \omega y sin\theta)} - e^{-\frac{\omega^2\sigma^2}{2}} \right], \tag{2}$$

where (x, y) represents the pixel positions, ω is the frequency of the sinusoidal plane wave, θ represents the orientation, and σ is the standard deviation or scale corresponding to the Gaussian envelope. In our proposed method, we extract features at five scales and eight orientations.

3.3 Local Binary Pattern Difference Feature

To extract the local binary pattern (LBP) features [49], an image is first partitioned into a number of blocks, in order to retain the spatial information about the features. In each block, the LBP code at each pixel position is generated, by comparing the central pixel with its corresponding neighboring pixels residing on a circle of the radius R, centered at the pixel under consideration. If a neighboring pixel has its value smaller than the central one, then it is labeled as '0', otherwise as '1'. A string of the binary bits is used to form an LBP code for the pixel. Mathematically, the LBP code of the central pixel x_c with respect to the channel ϕ can be written as follows:

$$I_{LBP_{N,R}}(x_c, \phi) = \sum_{n=0}^{N-1} u\big(\phi(x_n) - \phi(x_c)\big) 2^n, \tag{3}$$

where x_n $(n = 0, \dots, N-1)$ represents the N neighboring pixels on the circle of radius R centered at pixel x_c; ϕ can be either intensity value or filter response of the image; and $u(x)$ is the step function, i.e. its value is '1' if $x \geq 0$, and '0' otherwise. There are many LBP variants [50-52]. A problem with LBP is that the LBP code is a non-numerical

representation, which is a discrete pattern rather than a numerical response. Therefore, LBP cannot combine with other features for recognition. Recently, a numerical variant of LBP [53], which is known as local binary pattern difference (LBPD), was proposed. To extract this feature, the mean LBP of a given region is computed, then the LBPD at a pixel position is computed as the difference between its LBP code and the mean LBP. The Karcher mean [54] is used to compute the mean LBP of a region, which plays a vital role in minimizing the sum of distances to all the points in a given image region. Each element of a binary vector $\hat{\boldsymbol{I}}_{LBP}$ represents a specific bit of the regular LBP. Specifically, the k^{th} bit of $\hat{\boldsymbol{I}}_{LBP}$ is given as follows:

$$\hat{\boldsymbol{I}}_{LBP}(k) = u\big(\phi(x_k) - \phi(x_c)\big). \tag{4}$$

Suppose that there are P LBPs in a region, represented as $L = \{\hat{l}_1, \hat{l}_2, \dots, \hat{l}_P\}$. The k^{th} element for $k = 0, \dots, K-1$ of its Karcher mean $\hat{\boldsymbol{m}}_I$ is defined as follows:

$$\hat{\boldsymbol{m}}_I(k) = \left\lfloor \frac{\sum_{p=1}^{P} \hat{I}_p(k)}{P} + 0.5 \right\rfloor, \tag{5}$$

where $\lfloor . \rfloor$ is the floor function, $\hat{\boldsymbol{m}}_I$ belongs to the set of the 2^K LBPs. To relax the constraint that the LBP mean is an LBP, the mean LBP vector can be a floating-point vector denoted by $\hat{\boldsymbol{m}}_f$, as follows:

$$\hat{\boldsymbol{m}}_f = \frac{\sum_{p=1}^{P} \hat{\boldsymbol{I}}_p}{P}. \tag{6}$$

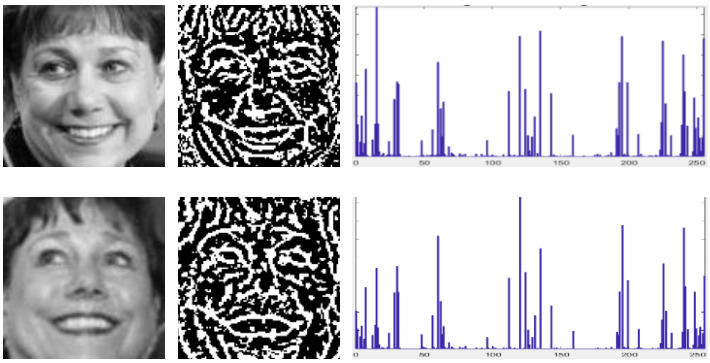


Fig. 3. LBPD feature and histogram representation of two face images.

3.3.1 LBP Difference

Let us consider the LBP code $\hat{\boldsymbol{I}}$ and mean $\hat{\boldsymbol{m}}_f$ of a face image. The LBPD feature vector can be computed as $\hat{\boldsymbol{d}} = \hat{\boldsymbol{I}} - \hat{\boldsymbol{m}}_f$. Magnitude of the LBPD feature is given by:

$$I_{LBPD}{}^{s}(x, \phi) = \left\| \hat{\boldsymbol{I}} - \hat{\boldsymbol{m}}_f \right\|, \tag{7}$$

where $\|.\|$ can be of any type of norms. Its values are positive so it is also known as unsigned LBPD. To extract more discriminative information from an image, the sign is introduced by defining the LBPD feature as follows:

$$I_{LBPD}{}^{s}(x,\phi) = s(\|\hat{\boldsymbol{I}}\| - \|\hat{\boldsymbol{m}}_f\|)\|\hat{\boldsymbol{I}} - \hat{\boldsymbol{m}}_f\|\,, \tag{8}$$

where $s(x)$ is the signum function, whose value is 1 if $s(x) \geq 0$ and -1 if $s(x) < 0$. This will form an ordered LBP feature vector. It is not affected by the permutation of bits, so the representation is invariant to rotation. Fig. 3 shows two face images from the LFW database and their corresponding LBPD images and histograms.

3.4 Feature Selection

Selection of appropriate feature descriptor is quite important to achieve optimum performance in image classification and object recognition. In the last two decades, various global and local feature descriptors have been proposed. As discussed earlier, local features tend to outperform global features, and have been proven to be more robust against various geometric variations. Some of the state-of-the-art feature descriptors include SIFT [55], SURF [56], HOG [57], Gabor [58], and LBP [49]. To improve the performance, various extensions of these descriptors have been proposed. For facial image analysis, Gabor and LBP have been proved to be the best performing feature descriptors for face recognition [59]. For object recognition, the most widely used local feature descriptor is SIFT, which first extracts relevant keypoints from given images, and then represents the gradient information in the neighborhood of each keypoint. This feature exhibits both scale and rotation invariance. However, its major drawback is high computational complexity. Inspired by SIFT, a faster version known as SURF was proposed, whose performance strongly depends on the relative keypoints that can have a variable geometry. It is highly desirable that the selected feature descriptor has high discriminative power and low computational complexity. In comparison to SIFT, LBP is simple and fast to compute. It can efficiently describe the local texture information, while showing high robustness to monotonic gray-level transformations. Moreover, the features computed by using LBP are fixed relative to each other and can better distinguish between the curved surfaces, such as face images. For a difficult task, such as LR face recognition, using a single feature is unable to capture sufficient discriminative information from face images. In [59], it was argued that combining the LBP with Gabor features can enhance the recognition performance, up to a significant level. As discussed in Section 3.3, the LBP feature consists of discrete patterns or symbols, rather than a numerical response, so the LBP feature cannot combine with other features directly.

In our proposed method, we employ two efficient texture descriptors, namely Gabor wavelets (GWs) and LBPD, due to their supplementary natures, which makes them promising candidates for fusion. The effectiveness of combining these two features is explained, as follows. First, the Gabor features can encode the facial shape information at multiple

scales and orientations. Each GW may be viewed as a bandpass filter, which extracts features at a specific range of frequencies and orientations in the frequency domain. Okajima et al. [60] argued that Gabor wavelets can be used as a solution for the mutual information maximization problem. The Gabor-type receptive field has the capability to extract the maximum amount of information from local image regions. Motivated by the success of LBP in the computer-vision tasks, we employ a recently proposed numerical variant of LBP, namely LBPD. It encodes the variation of a LBP code from the average LBP of a local region. LBPD does not consider the intensity of pixels, because it utilizes the sign of comparisons between neighboring pixels, as in Equation (4). This makes LBPD invariant to lighting conditions. It is clear that image rotation affects the permutation of bits. According to Equation (3), LBP used the predetermined weights to weigh the bits, which results in the different LBP codes of the original image and its rotated version. Therefore, extra effort is necessary to achieve rotation invariance [61]. LBPD is inherently rotation-invariant, as the norm employed in Equations (7) and (8) can make sure that the code is independent of permutation of the bits. In comparison to LBP, LBPD also expresses the diversity of the local co-occurrence, instead of representing it directly as LBP does. Furthermore, LBPD consists of numerical responses, whereas LBP is a collection of discrete patterns. This numerical property of LBPD makes it attractive, in terms of texture analysis. According to the results shown in [53], LBPD outperforms various LBP variants and other feature descriptors, in terms of texture, as well as medical imaging analysis. Due to the abovementioned advantages, we utilize LBPD as a feature descriptor. In our algorithm, both features are first normalized and then concatenated to form a final normalized feature vector.

4 Low-rank Feature learning

Recently, low-rank matrix recovery has gained a lot of attention, due to the great number of applications it has for computer vision and pattern recognition, e.g. face recognition [37,38], data mining, image classification [62], etc. For face recognition, Wei et al. [38] decomposed the data matrix into two parts, known as the low-rank matrix and its corresponding sparse error matrix, and then utilized the low-rank matrix for recognition. Instead of using data matrices directly, various kinds of features can first be extracted from images for better representation. As discussed in [62], extracted local features may exhibit some noisy patterns, which can reduce the recognition performance. Motivated by this observation, we decompose the extracted fused feature vectors $\boldsymbol{F}$ into a low-rank feature matrix $\boldsymbol{A}$ and a corresponding sparse error matrix $\boldsymbol{S}$. The extracted low-rank feature matrix $\boldsymbol{A}$ has been proven to be more

discriminative for recognition, as it provides better feature representation, as shown in Fig. 4. It works by minimizing the rank of the matrix $\boldsymbol{A}$, while computing the l_0-norm of $\boldsymbol{S}$. It can be written as follows:

$$\min_{\boldsymbol{A},\boldsymbol{S}} \operatorname{rank}(\boldsymbol{A}) + \lambda\|\boldsymbol{S}\|_0 \ \text{ s.t. } \boldsymbol{F} = \boldsymbol{A} + \boldsymbol{S}. \tag{9}$$

The second term computes the non-zero elements in $\boldsymbol{S}$. For simplification, the first term of Equation (9) can be replaced with the nuclear norm, while the second one with l_1-norm. The resulting optimization problem can be written as:

$$\min_{\boldsymbol{A},\boldsymbol{S}} \|\boldsymbol{A}\|_* + \lambda\|\boldsymbol{S}\|_1 \ \text{ s.t. } \boldsymbol{F} = \boldsymbol{A} + \boldsymbol{S}. \tag{10}$$

It turns out to be a convex optimization problem, with two major constraints. First, the rank of the recovered low-rank matrix $\boldsymbol{A}$ should be as small as possible. Second, there should be a small number of non-zero elements in $\boldsymbol{S}$. In our method, we utilize Augmented Lagrange multiplier (ALM) to solve this optimization problem due to its low complexity. Let $\boldsymbol{F}$ be the fused features extracted from face images. Then the Lagrange function of Equation (10) is written as:

$$L_\mu(\boldsymbol{A},\boldsymbol{S},\boldsymbol{Y}) = \|\boldsymbol{A}\|_* + \lambda\|\boldsymbol{S}\|_1 + <\boldsymbol{Y}, \boldsymbol{F} - \boldsymbol{A} - \boldsymbol{S}> + \frac{\mu}{2}\|\boldsymbol{F} - \boldsymbol{A} - \boldsymbol{S}\|_F^2, \tag{11}$$

where $\boldsymbol{Y}$ and μ represents a Lagrange multiplier and a penalty parameter, respectively. The matrices $\boldsymbol{A}$ and $\boldsymbol{S}$ are updated alternatively until converged, as follows:

$$\left(\boldsymbol{A}^{j+1}, \boldsymbol{S}^{j+1}\right) = \arg\min_{\boldsymbol{A},\boldsymbol{S}} L_\mu\left(\boldsymbol{A},\boldsymbol{S},\boldsymbol{Y}^j\right), \tag{12}$$

$$\boldsymbol{Y}^{j+1} = \boldsymbol{Y}^j + \mu(\boldsymbol{F} - \boldsymbol{A}^{j+1} - \boldsymbol{S}^{j+1}), \tag{13}$$

where j is the iteration index.

1) Updating $\boldsymbol{A}_i$

To update the low-rank matrix ${\boldsymbol{A}_i}^{j+1}$ of class i at the $(j+1)$st iteration, all the variables except $\boldsymbol{A}_i$ are fixed, which leads to the following equation:

$$\begin{aligned}
{\boldsymbol{A}_i}^{j+1} &= \arg\min_{\boldsymbol{A}_i} L(\boldsymbol{A}_i, {\boldsymbol{S}_i}^j, {\boldsymbol{Y}_i}^j, \mu^j) \\
&= \arg\min_{\boldsymbol{A}_i} \|\boldsymbol{A}_i\|_* + <{\boldsymbol{Y}_i}^j, \boldsymbol{F}_i - \boldsymbol{A}_i - {\boldsymbol{S}_i}^j> + \frac{\mu^j}{2}\left\|\boldsymbol{F}_i - \boldsymbol{A}_i - {\boldsymbol{S}_i}^j\right\|_F^2 \\
&= \arg\min_{\boldsymbol{A}_i} \in \|\boldsymbol{A}_i\|_* + \frac{1}{2}\|\boldsymbol{X}_a - \boldsymbol{A}_i\|_F^2,
\end{aligned} \tag{14}$$

where $\in = (2\mu^j)^{-1}$ and $\boldsymbol{X}_a = 0.5(\boldsymbol{F}_i - \boldsymbol{S}_i{}^j + \frac{\mathbf{1}}{\mu^j}\boldsymbol{Y}_i{}^j)$.

According to Section 2.1 in [63], the closed form solution of the above equation is given as $\boldsymbol{A}_i{}^{j+1} = \boldsymbol{P}Z_\in[\boldsymbol{R}]\boldsymbol{Q}^T$, where $\boldsymbol{PRQ}^T$ is the singular value decomposition of $\boldsymbol{X}_a$, and $Z_\in[\boldsymbol{R}]$ is the elementwise thresholding of $\boldsymbol{R}$, i.e., $Z_\in[\boldsymbol{R}](i,j) = z_\in[\boldsymbol{R}(i,j)]$, where $z_\in[r]$ is defined as

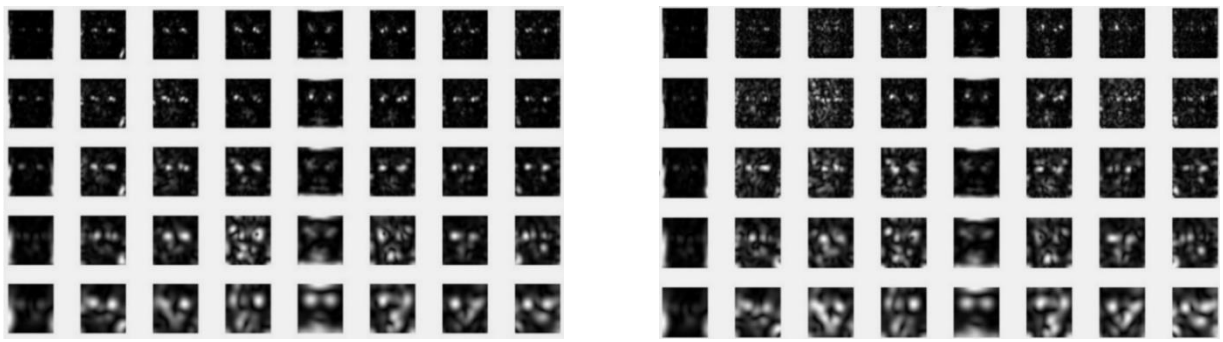
$$z_\in[r] = \begin{cases} r-\in, & \text{if } r > \in \\ r+\in, & \text{if } r < \in \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

2) *Updating* $\boldsymbol{S}_i$

$$\begin{aligned} \boldsymbol{S}_i{}^{j+1} &= \arg\min_{\boldsymbol{S}_i} L(\boldsymbol{A}_i{}^{j+1}, \boldsymbol{S}_i, \boldsymbol{Y}_i{}^j, \mu^j) \\ &= \arg\min_{\boldsymbol{S}_i} \lambda\|\boldsymbol{S}_i\|_1 + < \boldsymbol{Y}_i{}^j, \boldsymbol{F}_i - \boldsymbol{A}_i{}^{j+1} - \boldsymbol{S}_i > + \frac{\mu^j}{2}\left\|\boldsymbol{F}_i - \boldsymbol{A}_i{}^{j+1} - \boldsymbol{S}_i\right\|_F^2 \\ &= \arg\min_{\boldsymbol{S}_i} \in' \|\boldsymbol{S}_i\|_1 + \frac{1}{2}\|\boldsymbol{X}_s - \boldsymbol{S}_i\|_F^2, \end{aligned} \tag{16}$$

where $\in' = \left(\frac{\lambda}{\mu^j}\right)$ and $\boldsymbol{X}_s = \boldsymbol{F}_i - \boldsymbol{A}_i{}^{j+1} + \left(\frac{1}{\mu^j}\right)\boldsymbol{Y}_i{}^j$.

Similarly, the closed form solution of this optimization problem is given as $\boldsymbol{S}_i{}^{j+1} = Z_{\epsilon'}(\boldsymbol{X}_s)$. We set $\lambda = 0.001$ in our experiments. Furthermore, we discard the sparse error term $\boldsymbol{S}$, and use the low-rank approximation feature matrix $\boldsymbol{A}$ only for further processing. Fig.4 shows the recovered low-rank and the originally extracted Gabor features.



(a) (b)

Fig 4. Gabor features extracted from a face image. (a) Original Gabor features with 5 scales and 8 orientations, (b) Low-rank Gabor features.

5 Sparse Coding of Multiple Low-Rank Features

Sparse representation has gained much attention in the last decade, due to its state-of-the-art performance in pattern recognition tasks. Wright et al. [39] proved the effectiveness of sparse theory for recognition of face images taken in uncontrolled environments. According to the representation, a linear relationship exists between each test sample and

the other training samples from the same subject, which is sparse in nature. A face image $\boldsymbol{y}$ can be expressed as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{k}$, where $\boldsymbol{k}$ represents the sparse coefficient vector, and $\boldsymbol{X}$ is a data matrix whose columns represent the training data. It should be known that the samples of the same subject are highly correlated, while the correlation between samples of different subjects is weak. For the samples in $\boldsymbol{X}$ belonging to the same subject of $\boldsymbol{y}$, the corresponding coefficients in the sparse coefficient vector $\boldsymbol{k}$ should have non-zero values, while the rest of the coefficients are zero. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_M]$, where $\boldsymbol{x}_j$ is the j^{th} training sample and M is the total number of training samples. Let us consider that there are c classes in the training set, and n samples for each class, i.e. $M = nc$. According to sparse theory, each training sample can be linearly reconstructed by the remaining $M - 1$ samples, with most of the weights of the samples being zero. Our major objective is to project the features of training and testing samples into a low-dimensional feature space, such that their sparsity is preserved. Let us assume that the sparse coefficient vectors of the training samples are denoted as $\boldsymbol{K} = [\boldsymbol{k}_1, \boldsymbol{k}_2, \dots, \boldsymbol{k}_M]$, where $\boldsymbol{k}_j \in R^M$ is the sparse vector of the j^{th} training sample, computed using the l_1-minimization technique.

5.1 Feature Representation based on Sparse Coding

The purpose of sparse representation is to represent test images by using the minimum number of training samples. Mathematically, it can be written as follows:

$$\min\|\boldsymbol{k}\|_0 \text{ s.t. } \boldsymbol{y} = \boldsymbol{X}\boldsymbol{k}. \tag{17}$$

If it contains enough sparsity, then the solution of Equation (17) is equivalent to the l_1-minimization problem, i.e.

$$\min\|\boldsymbol{k}\|_1 \text{ s.t. } \boldsymbol{y} = \boldsymbol{X}\boldsymbol{k}. \tag{18}$$

In an ideal situation, a test image $\boldsymbol{y}$ from the j^{th} class can be linearly represented in terms of all the training samples, which can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{k} = \boldsymbol{X}\boldsymbol{k}_{1j} + \boldsymbol{X}\boldsymbol{k}_{2j} +, \dots, +\boldsymbol{X}\boldsymbol{k}_{nj}, \tag{19}$$

where n is the number of training samples in the j^{th} class, and $\boldsymbol{k}_{ij}$ is the sparse coefficient vector whose entries are zero, except those associated with the j^{th} class and the i^{th} training sample in the class.

In our algorithm, we propose the following objective function to compute the optimal projection matrix $\boldsymbol{P}$, which preserves the sparse structure when projecting the extracted multiple low-rank features $\boldsymbol{F}$ onto a new feature subspace.

$$\boldsymbol{P} = \arg\min_{\boldsymbol{P}} \sum_{j=1}^{M} \left\| \boldsymbol{P}^T \boldsymbol{f}_j - \boldsymbol{P}^T \boldsymbol{F}\, \boldsymbol{k}_j \right\|^2, \tag{20}$$

where $\boldsymbol{f}_j$ is the low-rank feature vector of the j^{th} training sample. By using simple algebraic formulation, Equation (20) can be written as:

$$\min \boldsymbol{P}^T \left(\sum_{j=1}^{M} (\boldsymbol{f}_j - \boldsymbol{F}\boldsymbol{k}_j)(\boldsymbol{f}_j - \boldsymbol{F}\boldsymbol{k}_j)^T \right) \boldsymbol{P}. \tag{21}$$

Assume that the low-rank feature vectors are projected onto an m dimensional vector space. Let $\boldsymbol{u}_j$ be the m-dimensional unit vector with the j^{th} element equal to 1, and 0 otherwise. Equation (21) can then be written as follows:

$$\begin{aligned} &\min \boldsymbol{P}^T \left(\sum_{j=1}^{M} (\boldsymbol{F}\boldsymbol{u}_j - \boldsymbol{F}\boldsymbol{k}_j)(\boldsymbol{F}\boldsymbol{u}_j - \boldsymbol{F}\boldsymbol{k}_j)^T \right) \boldsymbol{P} \\ &= \min \boldsymbol{P}^T \boldsymbol{F} \left(\sum_{j=1}^{M} (\boldsymbol{u}_j - \boldsymbol{k}_j)(\boldsymbol{u}_j - \boldsymbol{k}_j)^T \right) \boldsymbol{F}^T \boldsymbol{P} \\ &= \min \boldsymbol{P}^T \boldsymbol{F} \left(\sum_{j=1}^{M} (\boldsymbol{u}_j \boldsymbol{u}_j{}^T - \boldsymbol{u}_j \boldsymbol{k}_j{}^T - \boldsymbol{k}_j \boldsymbol{u}_j{}^T + \boldsymbol{k}_j \boldsymbol{k}_j{}^T) \right) \boldsymbol{F}^T \boldsymbol{P} \\ &= \min \boldsymbol{P}^T \boldsymbol{F} (\boldsymbol{I} - \boldsymbol{K} - \boldsymbol{K}^T + \boldsymbol{K}^T \boldsymbol{K}) \boldsymbol{F}^T \boldsymbol{P}. \end{aligned} \tag{22}$$

We set the constraint $\boldsymbol{P}^T \boldsymbol{F} \boldsymbol{F}^T \boldsymbol{P} = 1$. Then, the objective function is converted into the following optimization problem

$$\min_{\boldsymbol{P}} \frac{\boldsymbol{P}^T \boldsymbol{F} (\boldsymbol{I} - \boldsymbol{K} - \boldsymbol{K}^T + \boldsymbol{K}^T \boldsymbol{K}) \boldsymbol{F}^T \boldsymbol{P}}{\boldsymbol{P}^T \boldsymbol{F} \boldsymbol{F}^T \boldsymbol{P}}. \tag{23}$$

To solve (23), the Lagrange method is used, which provides the following equation:

$$L(\boldsymbol{P}, \lambda) = \boldsymbol{P}^T \boldsymbol{F} (\boldsymbol{I} - \boldsymbol{K} - \boldsymbol{K}^T + \boldsymbol{K}^T \boldsymbol{K}) \boldsymbol{F}^T \boldsymbol{P} - \lambda (\boldsymbol{P}^T \boldsymbol{F} \boldsymbol{F}^T \boldsymbol{P} - 1), \tag{24}$$

where λ is a lagrange multiplier, and $\boldsymbol{I}$ represents the identity matrix. To compute the optimum sparse projection matrix $\boldsymbol{P}$, we set the derivative to zero, i.e. $\frac{\partial L}{\partial \boldsymbol{P}} = 0$, which gives the following equation:

$$\boldsymbol{F} (\boldsymbol{I} - \boldsymbol{K} - \boldsymbol{K}^T + \boldsymbol{K}^T \boldsymbol{K}) \boldsymbol{F}^T \boldsymbol{P} = \lambda \boldsymbol{F} \boldsymbol{F}^T \boldsymbol{P}. \tag{25}$$

Finally, it becomes an eigen-decomposition problem in which we select the m eigenvectors of the matrix $(\boldsymbol{F}\boldsymbol{F}^T)^{-1} \boldsymbol{F} (\boldsymbol{I} - \boldsymbol{K} - \boldsymbol{K}^T + \boldsymbol{K}^T \boldsymbol{K}) \boldsymbol{F}^T$, with the smallest eigenvalues to construct a new low-dimensional feature subspace. Our proposed algorithm builds the sparse coefficient matrix by utilizing all the training data, so no search for nearest neighbors is required during testing. For visualization of the learned low-rank sparse features, we randomly selected 10 face samples from each of the 10 different classes. The low-rank sparse features are first extracted and then visualized using t-Distributed stochastic neighbor embedding (t-SNE) [64], as shown in Fig. 5. It can be observed that

the discriminablity of the learned features is enhanced, as samples from different classes are well separated in the feature space.

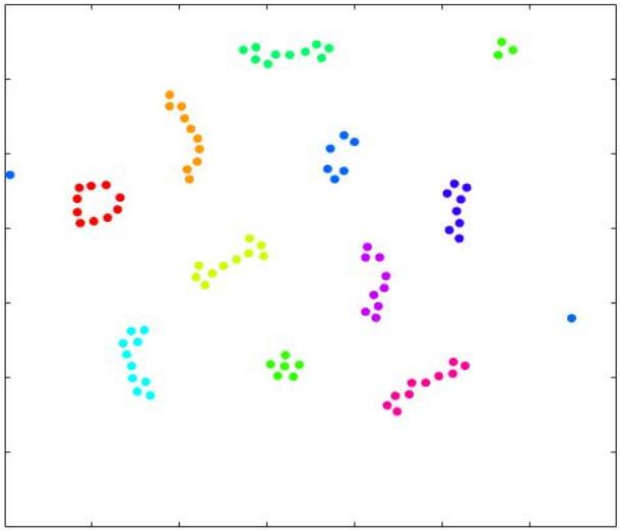


Fig 5. Visualization of the learned low-rank sparse features using t-SNE.

5.2 Geometrical and Mathematical properties of a Generalized Eigenvalue Problem

In this section, we explain our proposed formulation from geometrical and mathematical points of view. The proposed formulation involves two major steps, which are: (1) construction of the sparse coefficient matrix $\boldsymbol{K}$, and (2) determination of the projection matrix $\boldsymbol{P}$. To understand this eigenvalue problem, attention needs to be paid to these two major steps. Firstly, sparse coefficient vector $\boldsymbol{k}_j$ is determined for each sample $\boldsymbol{x}_j$ using l_1 minimization. Each column of $\boldsymbol{K}$ represents the sparse coefficient vector of a sample. Now, we will analyze the effectiveness of this computed sparse coefficient matrix $\boldsymbol{K}$. Geometrically, each sparse coefficient vector $\boldsymbol{k}_j$ is invariant to scaling and rotation of the data samples. It is also invariant to translation due to the constraint $1 = \mathbf{1}^T\boldsymbol{k}_j$, where $\mathbf{1}$ is an identity vector. Therefore, the sparse coefficient matrix $\boldsymbol{K}$ remains unchanged whenever data samples are translated and rotated, which is one of its important geometrical properties. In our proposed method, we construct a coefficient matrix, using whole training data, instead of using the k-nearest neighbors. This helps in preserving the global structure of the data, while projecting them into a new sparse feature subspace. The constructed sparse coefficient matrix also has the capability to preserve the discriminant information. To understand this, let us take an example of face recognition. The major assumption in face recognition is that the face images, belonging to the same class, lie on a linear subspace. Let $\boldsymbol{x}_j$ be a face image belonging to the j^{th} class, $\boldsymbol{x}_j$ can be represented as a linear combination of the other face images

from the same j^{th} class, and the computed coefficient vector $\boldsymbol{k}_j$ is sparse. This shows that $\boldsymbol{k}_j$ naturally contains discriminant information, so it can easily distinguish the face images of two different classes. Our proposed method uses l_1 regularization, which enables it to encode the prior knowledge of sparsity. It performs sparse reconstruction only in the training process. Having determined the projection matrix $\boldsymbol{P}$, sparse reconstruction is no longer necessary and our algorithm is therefore efficient.

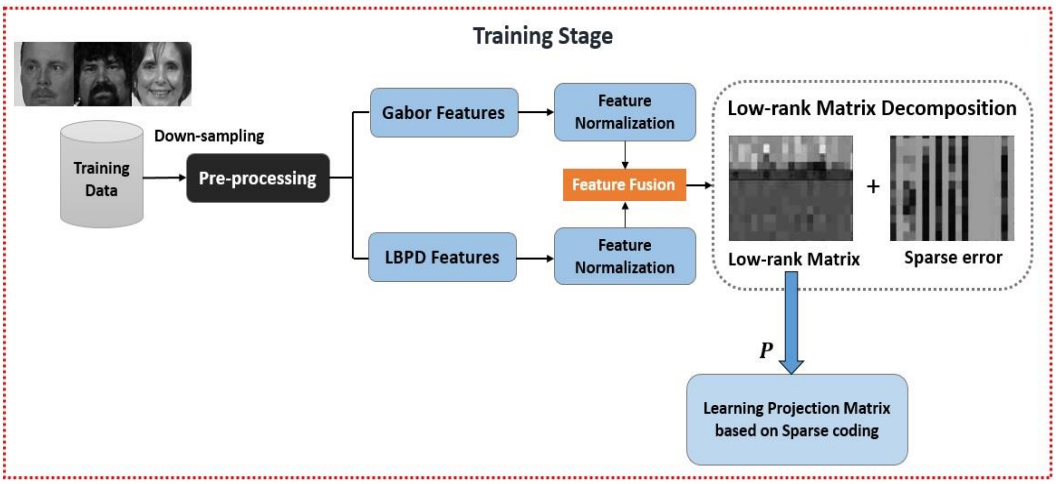


Fig 6. Training stage of our proposed framework.

6 Linear-Regression-based Classification

After projecting gallery and probe features onto the sparse feature subspace, a linear regression framework is developed to model the similarity between them. The main idea of this model is to compute a linear mapping function between the gallery and the probe face images. According to the developed model, there exists a linear relationship between a probe image and each of the samples in a gallery. If a query face image fits to the i^{th} class in the gallery set, it can be represented as a linear combination of the gallery-images features from the same class. Therefore, we have

$$\boldsymbol{I}_R = \boldsymbol{X}_i \boldsymbol{\alpha}_i \,, \tag{26}$$

where $\boldsymbol{I}_R$ is the reconstructed probe image based on the gallery images from the i^{th} class; $\boldsymbol{X}_i = [\boldsymbol{x}_{i,1}, \dots, \boldsymbol{x}_{i,n_i}]$ are the training samples from the i^{th} class, which has n_i samples; and $\boldsymbol{\alpha}_i$ represents the image coefficient vector of the probe image, estimated by the least-squares algorithm. The next step is to find the residual values for each class, based on the computed coefficient vectors. The query face image $\boldsymbol{y}$ is assigned to the class j with the minimum residual value, i.e.

$$j = \min_i \|\boldsymbol{y} - \boldsymbol{X}_i \boldsymbol{\alpha}_i\|. \tag{27}$$

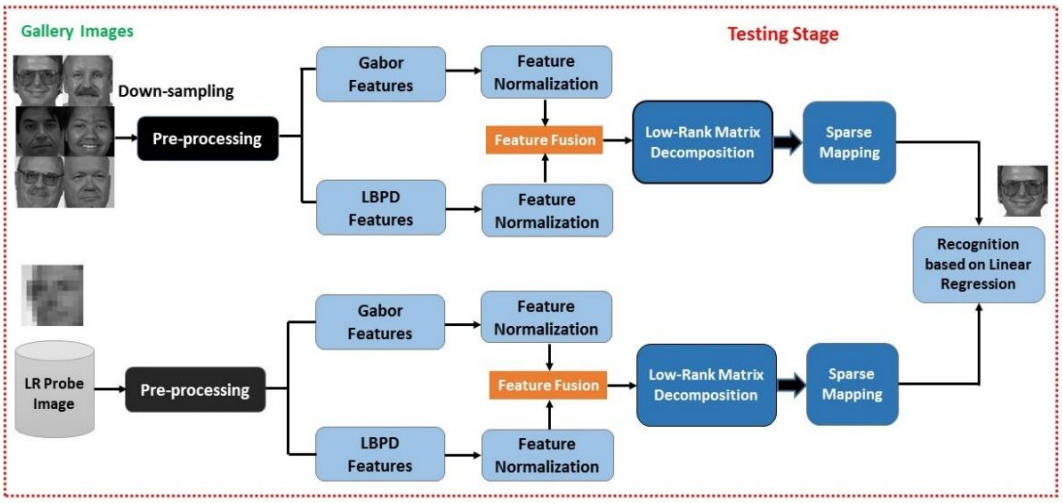


Fig 7. Testing stage of the proposed framework.

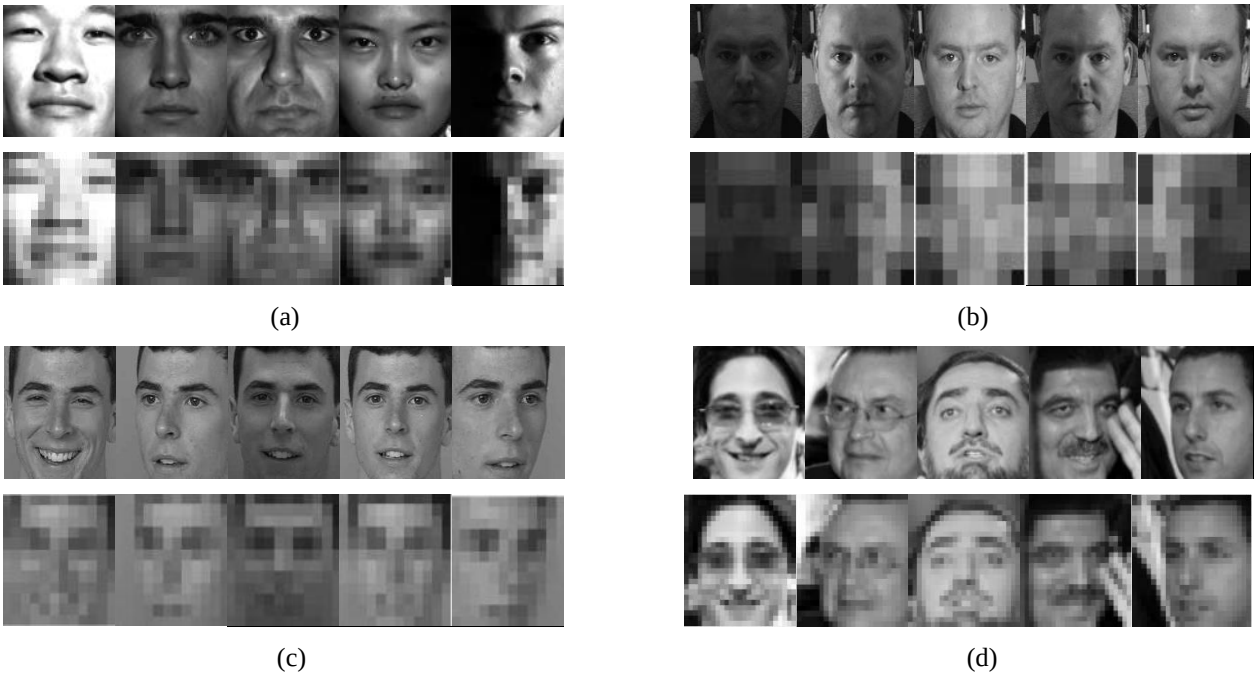


Fig. 8. Original images and the corresponding LR images: (a) Extended Yale-B, (b) Multi-PIE, (c) FERET, and (d) LFW databases. The first rows show the original face images, while the second rows show the downsampled images.

7 Experiments

We evaluate the performance of our proposed framework by conducting the experiments on five face datasets, including the Extended Yale-B, Multi-PIE [45], FERET [46], LFW [47], and Remote face [48] databases. These databases contain images with variations in pose, facial expressions, and lighting conditions. In the pre-processing stage, face images are first detected and aligned using MTCNN [65]. For the different databases, we followed the other papers on LR face recognition to down-sample face images to a specific size. This allows our algorithm to be directly compared to other algorithms.

7.1 Experimental Results on the Extended Yale-B Database

The Extended Yale-B database consists of 2,432 images from 38 subjects with 64 images per subject, which were captured under different lighting conditions. In our experiments, all 64 images per subject with different illumination conditions are utilized. For training, we randomly select 10, 20, and 30 images per subject. LR probe images of size 12×12 are generated using a down-sampling operation. The HR and LR face images of five individuals from the Extended Yale-B database are shown in Fig. 8(a). In Fig. 9, the recognition rate shows an increasing trend and remains stable as feature dimension increases. Our proposed method outperforms other LR face recognition methods, and achieves the highest recognition rate of 94.74%, when the feature dimension is higher than 70. The highest recognition rates of CLPM [11], DSR [15], NMCF [14], MDS [19], CCA [23], CDMMA [24], and CMDA [26] are 89.35%, 62.3%, 81.42%, 77.01%, 77.90%, 88.2%, and 89.8%, respectively, at their corresponding optimal feature dimensions.

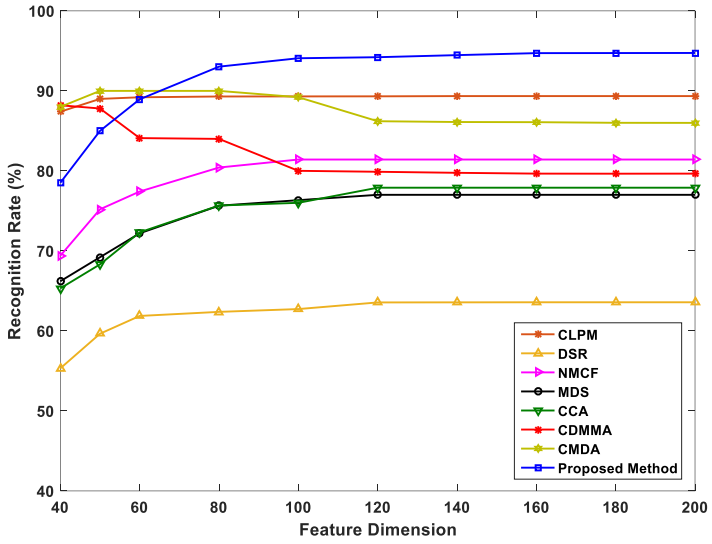
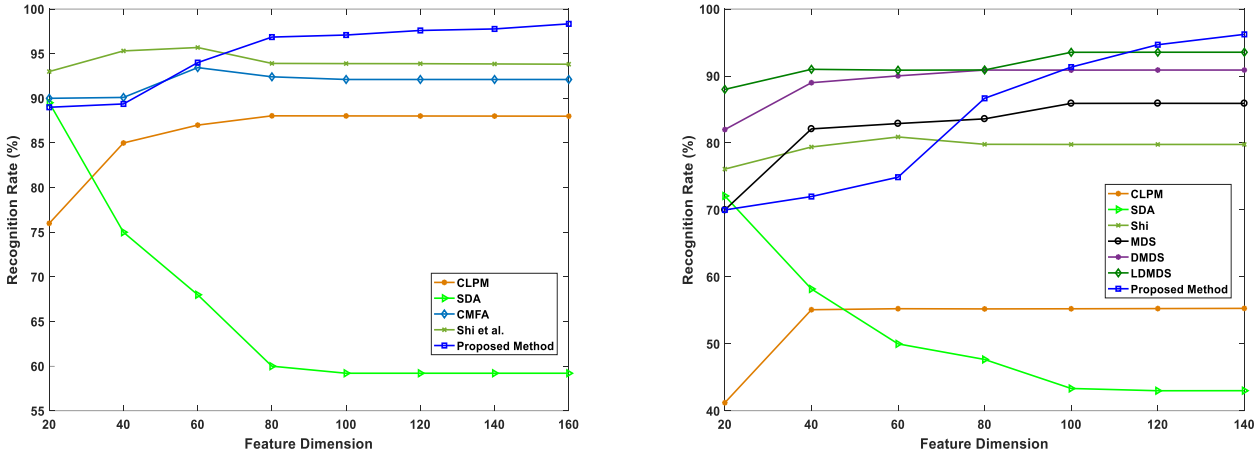


Fig.9. Recognition rates of different methods at different feature dimensions on the Extended Yale-B database (LR: 12×12).

7.2 Experimental Results on the Multi-PIE Database

The Multi-PIE dataset consists of more than 750,000 face images of 337 subjects. Images were captured in four different sessions. Following the protocol used in [27], we conducted experiments on a subset of session 04, which contain images with frontal pose under 20 different illumination conditions. The camera and the recording numbers, used in our experiments, are 05-1 and 01, respectively. For training, we randomly selected 50 subjects, while the remaining subjects were used for testing. To construct a gallery set, 6 images of each subject were selected, while the remaining 14 images were included in a probe set. LR probe images of size 8×8 are generated using a downsampling operation. The HR and the corresponding LR images from the Multi-PIE dataset are shown in Fig. 8(b). Table 2 shows

the comparative results in terms of Rank-1 recognition accuracy. Fig. 10(a) shows the recognition rate at different feature dimensions. Our method performs better than the other LR face recognition methods and achieves the highest recognition accuracy of 97.41%.



(a) (b)

Fig. 10. Recognition rates with different feature dimensions: (a) Multi-PIE database (LR: 8 × 8) and (b) the FERET database (BaBe) (LR: 8 × 8).

7.3 Experimental Results on the FERET Database

The FERET dataset consists of more than 13,000 face images from 1,565 subjects. Each subject has images with large variations in pose, expressions, and illumination conditions. The dataset consists of one standard gallery set (*Fa)*, and three probe sets (*Fb, Dup1*, and Dup2). The HR and LR face images of five individuals from the FERET database are shown in Fig. 8(c). In our experiments, we selected the group, namely *Fa*, which consists of 994 frontal face images with one image per subject, and used it as a gallery set, while *Fb*, which consists of 994 images with expression variations, was used as a probe set. Throughout this paper, we called this subset 'FERET (Fa)'.We carried out our experiments by selecting a training set that consists of only one image per subject. LR probe images of size 12 × 12 are generated by using a down-sampling operation. Our method performs better than other LR face recognition methods, and achieves the highest recognition rate of 89.66%, with a feature dimension of 200, whereas the recognition accuracy achieved by CLPM [11], CMFA [20], SDA [18] and C-RSDA [25] are 82.46%, 75.40%, 71.77%, and 86.29%, respectively, with their optimal feature dimensions. Table 1 shows the comparative results at different resolutions.

Table 1
Comparative results on the FERET (Fa) dataset, in terms of Rank-1 Recognition accuracy, at different resolutions with optimal feature dimensions.

Algorithm	8*8	12*12	16*16
CLPM [11]	79.94%	82.46%	84.48%
CMFA [20]	72.08%	75.40%	75.60%
SDA [18]	68.75%	71.77%	72.08%
C-RSDA [25]	82.36%	86.29%	86.29%
Proposed method	**84.71%**	**89.66%**	**95.22%**

Table 2
Comparative results for FERET (BaBe) and Multi-PIE datasets in terms of Rank-1 Recognition accuracy at optimal feature dimensions (Probe image resolution:8×8).

Algorithm	FERET (BaBe)	Multi-PIE
CLPM [11]	55.22%	88.04%
CMFA [20]	75.98%	93.44%
SDA [18]	72.09%	89.51%
Shi et al. [21]	80.90%	95.69%
MDS [19]	85.91%	91.78%
LMCM [22]	90.00%	--------
DMDS [27]	90.89%	93.88%
LDMDS [27]	93.55%	95.81%
Proposed Method	**96.22%**	**97.41%**

We further perform experiments on another challenging subset of the FERET dataset, which contains images of 200 subjects (including $ba, bd, be, bf, bg, bj\ and\ bk$). Throughout this paper, we called this subset 'FERET (BaBe)'. In this subset, 7 images per subject are available. We randomly selected 50 subjects for training, while the remaining 150 subjects were used to construct a testing set. In other words, different subjects were used for training and testing. During training, all of the 7 images per subject were used. For the testing set, the first four images were used to construct a galley set, while the remaining 3 images were used as a probe set. For evaluation, we downsample the probe images to the size of 8×8. Our proposed method outperforms other state-of-the-art LR face recognition methods, and achieves the highest recognition rate of 96.22%, with a feature dimension of 140. The comparative results are shown in Table 2. Recognition rates with different feature dimensions are shown in Fig. 10(b). The face images in this subset contain large pose and illumination variations. Therefore, very low-dimensional features are not effective in achieving optimal performance. However, our method still outperforms the other methods, when the feature dimension is higher than 120.

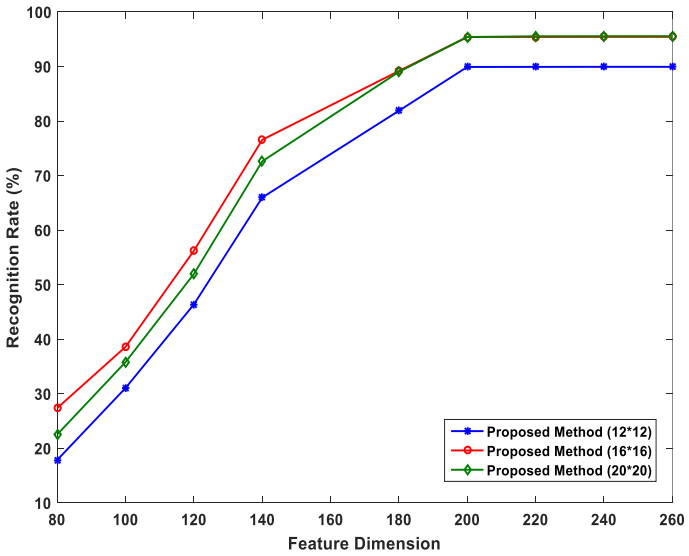


Fig. 11. Recognition rates of our proposed method on the LFW database, with different feature dimensions and at different probe image resolutions.

7.4 Experimental Results on the LFW Database

We also evaluate our proposed method on LR face images taken under uncontrolled environments by conducting experiments on the LFW dataset. All the images in this dataset were captured in the wild, having large variations in expression, pose, make-up, lighting condition, etc. It consists of 13,233 face images from 5,749 individuals. Out of these, 1,680 individuals contain more than two images, and 610 of them contain more than four images in the dataset. We randomly selected 4 images from each of the 610 individuals. For training, we randomly selected 150 subjects with 10 images each from the CASIA-Web face dataset [66]. For testing, two images per subject from the selected LFW images were used to construct the gallery set, and the other two were used for the probe set. In other words, the datasets used for training and testing are different. LR probe images of resolutions 12×12, 16×16, and 20×20 were generated using a downsampling operation. Our method shows a promising result by achieving 88.23% accuracy on LR images of size 12×12. For the LFW dataset, we compare our results with a recently proposed deep-learning-based method [32], which will be shown in the next section. The HR and LR face images of five individuals from the LFW database are shown in Fig. 8(d). The recognition rate of our proposed method, with different feature dimensions and at three different probe image resolutions, is shown in Fig. 11.

7.5 Comparison with Deep-Learning Methods

Convolutional neural networks (CNNs) have revolutionized pattern-recognition research by providing state-of-the-art performances in various computer-vision tasks. One of the main reasons for its success is the availability of a large

amount of training data and the networks are trained for feature extraction and recognition from end to end. As discussed earlier, deep-learning methods have achieved more than 99% recognition accuracy on the LFW dataset. However, deep learning is still finding a way to make its mark in solving the LR face recognition problem. Schroff et al. [29] reported around 50% decline in validation rate, when the size of face images was reduced from 256×256 to 40×40. In this section, we analyze the performance of our proposed method in comparison to the deep-learning-based method. In this regard, we have performed several experiments to evaluate whether current deep face models are good enough for recognizing LR face images.

For deep-learning-based experiments, we used three different models of SphereFace (SF) [32] (a deep CNN model, trained on the CASIA-Web Face [66] dataset, to perform face recognition). It should be noted that the size of the training data used by SphereFace is small, compared to the datasets used in VGG-Face [30], FaceNet [29], and Deep Face [28]. The model achieves excellent performance on both the LFW and Youtube face (YTF) datasets. First, we used a pretrained model (Model no. 1) of SphereFace for LR face recognition. Our experimental results show that using a pretrained model for LR face recognition gives the worst performance. The reason for this is that SphereFace was originally trained on HR images, so fine-tuning or retraining is necessary to achieve good performance on LR face images. It should be noted that the Multi-PIE and LFW datasets contain color face images. Therefore, we fine-tune the model with a small learning rate of 0.01, which linearly decays to zero. We also randomly downsample the input face images to different sizes between 8×8 and 95×95, and report the best performance for the two datasets in Tables 3 and 4, respectively. On the other hand, the Extended Yale-B and FERET datasets consist of grayscale images, so fine-tuning the original model with color images might not give the optimal performance on these two datasets. Therefore, we retrain the original model using gray-scale images instead of color face images. In this case, the input size of the first convolutional layer is changed to 1, rather than 3. Similar to the fine-tuned models, the input face images are downsampled to the sizes between 8×8 and 95×95. The initial learning rate is set to 0.1, which linearly decays to zero. At last, we fine-tune the model with the LR version of both CASIA-Web Face and the images we used for training our proposed algorithm with the same configuration, as explained before. This fine-tuned model is denoted as SF-LR_C. For LFW dataset, our algorithm is already trained with CASIA-Web Face, so no further fine-tuning is required for the LFW dataset. However, this fine-tuned model (SF-LR_C) does not perform well, when the image resolution is less than 20×20. The CASIA-Web Face dataset consists of large number of images per subjet with variations in pose,

expressions, and illumination. This allows deep models to learn rich level of image representations for recognition. Therefore, adding more images into the training set does not increase the recognition rate. The optimal recognition performances are reported in Tables 5, 6 , and 7, respectively.

Table 3
Recognition results of deep-learning-based methods and the proposed method for LR face recognition at different resolutions, on the Multi-PIE Dataset.

Method	8 × 8	12 × 12	16 × 16	20 × 20
SF-HR	84.39%	90.59%	96.03%	99.02%
SF-LR	96.94%	98.41%	99.96%	**100.00%**
SF_LR_C	90.33%	97.31%	99.84%	**100.00%**
Proposed Method	**97.41%**	97.94%	97.94%	98.06%

Table 4
Recognition results of deep-learning-based methods and the proposed method for LR face recognition at different resolutions, on the LFW database.

Method	12 × 12	16 × 16	20 × 20
SF-HR	18.93%	19.67%	41.64%
SF-LR	52.70%	78.85%	89.92%
Proposed Method	**88.36%**	**93.28%**	**95.41%**

Table 5
Recognition results of deep-learning-based methods and the proposed method for LR face recognition at different resolutions, on the Extended Yale-B database.

Method	12 × 12	16 × 16	20 × 20
SF-HR	61.22%	54.64%	57.12%
SF-LR	63.08%	68.03%	73.61%
SF-LR_C	62.12%	67.98%	73.61%
Proposed Method	**94.74%**	**95.13%**	**94.87%**

Table 6
Recognition results of deep-learning-based methods and the proposed method for LR face recognition at different resolutions, on the FERET (Fa) database.

Method	8 × 8	12 × 12	16 × 16	20 × 20
SF-HR	52.73%	65.07%	72.30%	77.64%
SF-LR	59.29%	49.28%	86.76%	94.22%
SF-LR_C	34.51%	51.31%	73.94%	86.32%
Proposed Method	**84.71%**	**89.66%**	**95.22%**	**96.55%**

Table 7
Recognition results of deep-learning-based methods and the proposed method for LR face recognition at different resolutions, on the FERET (BaBe) database.

Method	8 × 8	12 × 12	16 × 16	20 × 20
SF-HR	42.67%	65.56%	80.22%	88.44%
SF-LR	55.11%	54.67%	94.22%	99.56%
SF-LR_C	39.33%	44.22%	73.56%	90.44%
Proposed Method	**96.22%**	**99.33%**	**98.67%**	**99.78%**

It can be seen that fine-tuning and retraining result in significant improvements in the recognition of LR face images. Throughout our experiments, we follow the SphereFace's implementation to align the face images using MTCNN [65]. In the testing stage, deep features are extracted from the output of the FC1 layer. In the experiments, the final feature of a probe face image is obtained by concatenating the original features and its mirrored features. Finally, the similarity score is computed using the cosine similarity. In [32], the performance of the SphereFace model was evaluated using different numbers of layers (10, 20, 36, 40 and 64). However, a minor improvement is reported in the recognition rate when the number of layers increases from 20 to 64.

7.6 Experimental Results on the Remote Face Database

In order to evaluate the performance of our proposed method under more challenging conditions, we conducted experiments on the Remote Face dataset [48], which contains images taken under unconstrained outdoor maritime environments. The images were taken at different distances, ranging from 10-250 m. There are 2,102 face images in total from 17 subjects. Each subject has a number of images, ranging between 29 and 306.

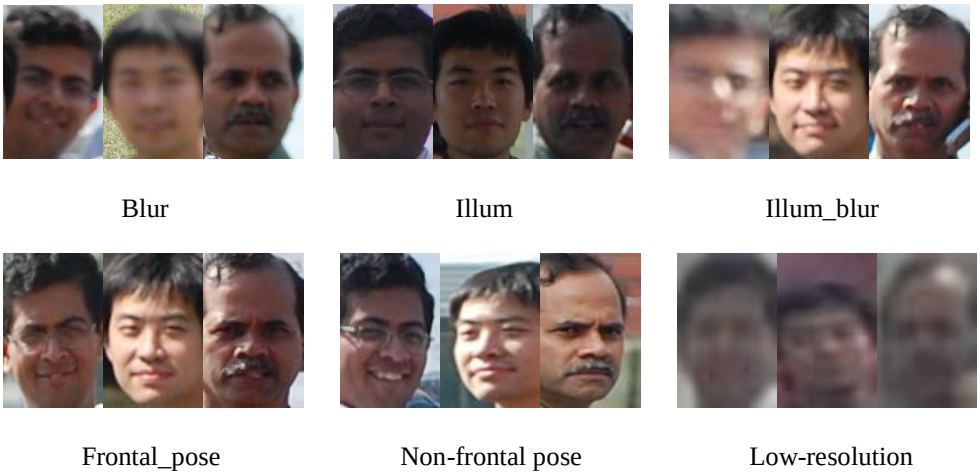


Fig. 12. Sample face images from all the six subsets of the Remote Face database.

The dataset consists of six subsets, denoted as *blur*, *illum*, *illum_blur*, *frontal_pose*, *Nf_pose*, and *low_res*, respectively. The *blur* subset contains 75 face images with blurring effects. The *illum* subset consists of 561 face images with different lighting conditions. The *illum_blur* subset consists of 128 images, with both lighting and blurring effects. The *low_res* subset is the most challenging one, as it contains 90 face images of very low resolution. The *frontal_pose* and non-frontal pose (*Nf_pose*) subsets include images having frontal and non-frontal poses, with 1,166 and 846 face images, respectively. The gallery set consists of five HR face images of each subject. We conducted experiments on all of the six subsets and achieved competitive results. For the subsets with a blur and lighting effects, all the methods

can achieve promising results, as these effects do not have severe impact on the image's apperance. All the methods can also achieve satisfactory performance on the frontal-pose subset. For the *Nf_pose* subset, most of the methods have their performance drop significantly, while our method can achieve the best performance, with the recognition rate of 96.4%. The *low_res* subset is the most challenging one, because the images are of a small size and suffered from blurring. The resolution of the images in this subset is 20×30 only. The face images of three different subjects from all the six subsets are shown in Fig. 12. No training set is provided by the Remote Face dataset [48], so we used 16,028 frontal face images from the Face Recognition Grand Challenge (FRGC) dataset for training. It can be seen that all the other methods do not perform well on the LR face images, as it is very difficult to extract useful information from such low-quality, low-resolution images. Similar to [36], we compare the performance of our method with different local feature descriptors, including binarized statistical image features [67], DFD [34], LSF [36], and two deep-learning-based methods (DL [68], and SphereFace (SF) [32]). Experimental results are tabulated in Table 8.

Table 8
Comparative results on remote face dataset, in terms of Rank-1 Recognition rate on all the six subsets.

Subset	Algorithm	Rate (%)	Subset	Algorithm	Rate (%)	Subset	Algorithm	Rate (%)
blur	BSIF	62.2	*frontal_pose*	BSIF	70.1	*Nf_pose*	BSIF	49.2
	DFD	63.5		DFD	78.6		DFD	52.5
	DL	48.6		DL	80.3		DL	49.8
	Shearlet	62.5		Shearlet	77.8		Shearlet	51.7
	LSF	67.3		LSF	83.8		LSF	57.2
	SF-FT	89.4		SF-FT	95.4		SF-FT	90.1
	SF-FT_C	85.9		SF-FT_C	95.0		SF-FT_C	89.1
	SF	**93.8**		SF	97.8		SF	94.2
	Ours	90.5		Ours	**98.8**		Ours	**96.4**
illum	BSIF	79.3	*illum_blur*	BSIF	74.8	*low_res*	BSIF	11.2
	DFD	83.4		DFD	75.2		DFD	14.5
	DL	80.4		DL	71.8		DL	11.5
	Shearlet	81.6		Shearlet	74.3		Shearlet	13.8
	LSF	92.5		LSF	76.0		LSF	19.9
	SF-FT	**99.2**		SF-FT	96.9		SF-FT	51.4
	SF-FT_C	98.0		SF-FT_C	96.9		SF-FT_C	61.4
	SF	**99.4**		SF	98.5		SF	66.0
	Ours	98.8		Ours	**98.8**		Ours	**81.1**

It can be seen that the deep-learning-based model SF [32] can achieve a state-of-the-art performance on the five subsets, except the *low_res* subset. This is because the original model was trained on HR face images. In our experiments, we have also fine-tuned the model using two schemes. In the first scheme, our model is fine-tuned with the LR images from the CASIA-Web Face dataset [66], with the sizes between 8×8 and 95×95, for performance

evaluation. In the experiment results, this fine-tuned model is denoted as SF-FT. In the second scheme, our model is fine-tuned with the combination of LR images from both the CASIA-Web Face and the FRGC dataset (16,028 frontal face images), with the sizes between 8×8 and 95×95. This fine-tuned model is denoted as SF-FT_C. After including the FRGC dataset for fine-tuning, the recognition rate is increased by 10% on the *low_res* subset. However, the original SF model performs better than the two fine-tuned models (SF-FT and SF-FT_C), because the distribution of the downsampled faces used for fine-tuning is different from the native LR faces. For those subsets other than *low_res*, no considerable improvement is observed in terms of recognition rate by including the FRGC dataset in the training set. It can be observed that the performance of the deep learning models declines when the presented probe images are of low resolution. However, our method performs better than all the other methods on this subset, as well as on *frontal_pose*, *illum_blur*, and *Nf_pose*. For the *blur* and *illum* subsets, the performance of our method is comparable to SF and SF-FT.

Our proposed method also provides state-of-the-art performance on the *low_res* subset, due to the two effective stages of learning (low-rank feature learning and sparse mapping). LR images contain very limited amounts of information, which cannot be fully extracted by using only multiple local features. Furthermore, the LR images may contain some artifacts or noises, which can greatly degrade the recognition performance. The low-rank feature learning part in our algorithm generates a clean low-rank feature matrix and excludes the distorted components, which helps in revealing the correct identity information. In Section 7.8, the effectiveness of low-rank feature learning will be further explained. Furthermore, computing the sparse-coefficient matrix using the whole training set leads to the projection of the learned features into a discriminant sparse feature subspace, which can further increase the inter-personal variations. This can also improve the recognition performance.

7.7 Feature Fusion

Most of the existing face recognition methods utilize only one feature descriptor. However, in difficult tasks, such as LR face recognition, no single feature is good enough to extract all the relevant information from LR face images. Combining multiple efficient features is a promising way to bring major improvement in recognition accuracy. In our previous approach [44], we used only one local feature descriptor (Gabor wavelets) for extracting facial details from LR images, which provided satisfactory performance. However, it does not perform well under large pose variations and poor lighting conditions. To overcome this problem, we fuse two efficient local feature descriptors, i.e. Gabor

wavelets and LBPD, which can achieve much better performance even under large unconstrained environments. Table 9 shows the recognition rates on the four databases (in comparison to previously proposed approach [44]), with and without performing feature fusion at the corresponding optimal feature dimensions. It can be observed that learning and fusing the low-rank features brings significant improvements, in terms of recognition accuracy.

Table 9
Recognition rates in comparison to the preliminary work [44], recorded at the optimal feature dimensions.

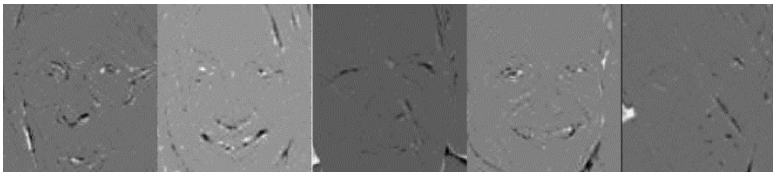
Dataset	Previous Method [44]	Proposed Method
Extended Yale-B (LR: 12×12)	49.21%	**94.74%**
FERET (fa) (LR: 12×12)	70.08%	**89.66%**
FERET (BaBe) (LR: 8×8)	46.44%	**96.22%**
LFW (LR: 12×12)	44.34%	**88.36%**
Multi-PIE (LR: 8×8)	56.40%	**97.41%**

7.8 Effectiveness of Low-rank Feature Learning

Low-rank feature learning has been proven to be an efficient technique to handle large amounts of noise variations in data samples. It aims to learn a low-rank dictionary, by optimizing the dictionary atoms, and removing sparse errors from the data samples. Furthermore, it can reveal the global structural information of the data samples, which helps in reconstructing a given test sample using a discriminative low-rank dictionary. The learned low-rank feature of a sample has high correlation with that of the same class, which enhances the recognition performance. It can effectively reveal the identity of the data samples by incorporating the global structural information.

As discussed before, we first decompose the extracted local features into a low-rank feature matrix, and a sparse error matrix. After that, only the low-rank component is utilized for identification. In this section, we evaluate the performance of our proposed method with and without including the estimated sparse error matrix in the final feature representation. Firstly, we visualize the estimated low-rank and sparse error components by applying the low-rank matrix decomposition algorithm on some face images from the LFW dataset. Fig. 13 shows the low-rank representation of face images and their corresponding sparse errors.

The low-rank representation of face images from LFW dataset.



Corresponding Sparse error images

Fig. 13. Low-rank representation of face images and the corresponding sparse error images.

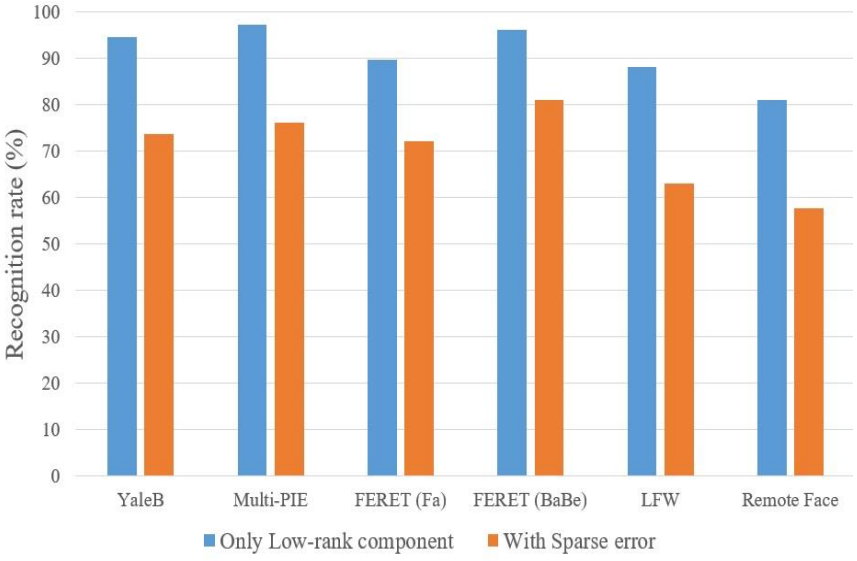


Fig. 14. Recognition rates with and without using the sparse error matrix for all the five datasets at optimal feature dimensions.

In our experiments, we found that by including the sparse error matrix in the final feature representation, recognition rate drops by a significant level. We repeat our experiments by following the protocol discussed before, and report the recognition rates for all of the five datasets at optimal feature dimensions. Fig. 14 shows the recognition rates with and without using the sparse error matrix. It can be observed that the recognition rate increases by 20-25%, when the sparse error matrix is discarded. Furthermore, we also evaluate the performance by excluding low-rank feature learning from the recognition pipeline. This means that the extracted fused features are directly projected into the sparse feature subspace for classification. The results are reported for all the five datasets with optimal feature dimensions. For the datasets consisting of images taken in constrained environments, a 10-20% decline in recognition rate is observed. In the next part, we evaluate the performance without projecting the extracted low-rank features into the sparse feature subspace. The dimension of the low-rank features is reduced directly using PCA, which is then fed to a linear-regression-based classifier. Fig. 15 shows the recognition rates obtained after conducting various ablation studies. It can be seen that, by excluding low-rank feature learning and sparse mapping, the recognition rate drops down to a significant level. This shows the effectiveness of our proposed method. The recognition results based on our method

on the four datasets, at the corresponding optimal feature dimensions and with different numbers of training images per subject, are shown in Tables 10-13, respectively.

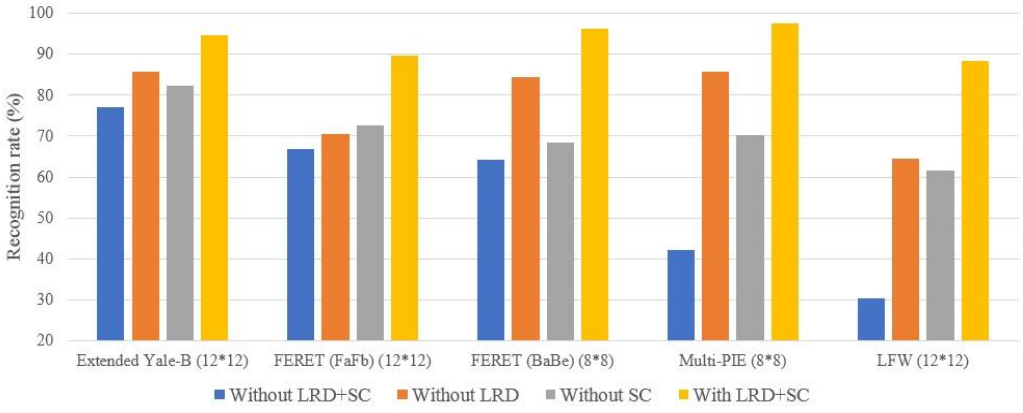


Fig. 15. Recognition rates of our proposed method, with and without using low-rank decomposition and sparse mapping, on the ??? dataset.

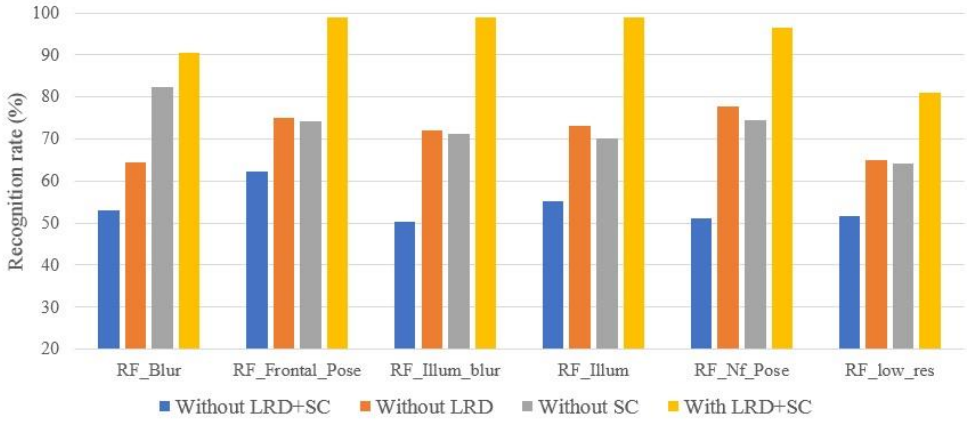


Fig. 16. Recognition rates of our proposed method, with and without using low-rank decomposition and sparse mapping, on the Remote Face dataset.

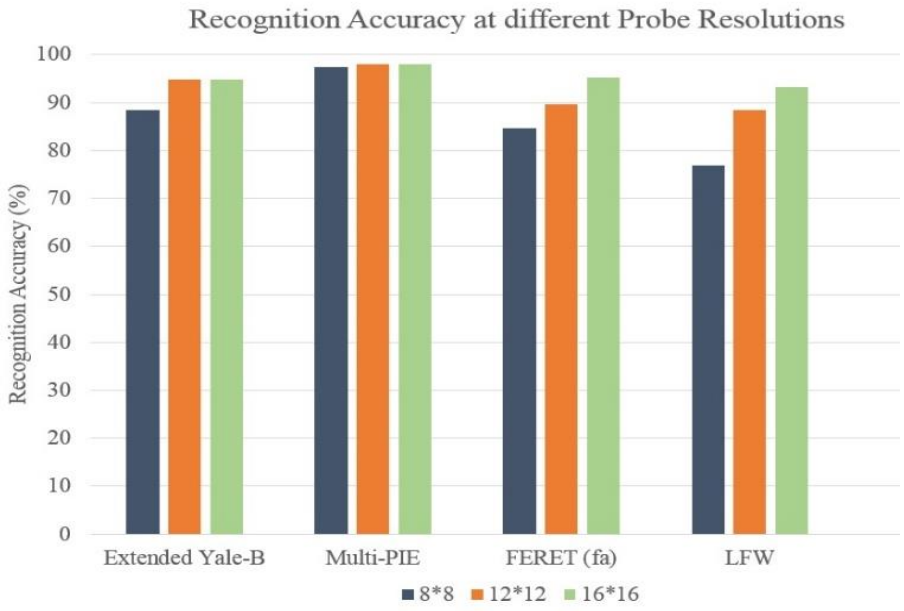


Fig. 17. Recognition rates of the proposed method on four datasets at different probe image resolutions.

7.9 Recognition across Different Probe Resolutions

In this section, the performance of our proposed method is evaluated using probe images of different resolutions, with and without using our proposed morphological preprocessing method. For all the four datasets, three different probe resolutions, 8×8, 12×12, and 16×16, were used. Experiments were conducted by selecting a fixed number of training samples per subject and the respective optimal feature dimensions, at different probe resolutions, and results are shown in Fig. 17. The results show that our proposed method can achieve a very good recognition performance even if image resolution is reduced to lower than 12×12. Table 14 shows the recognition rates, with and without using the proposed pre-processing method, at the corresponding optimal feature dimensions for the four datasets. It can be observed that 1% to 6% of improvement, in terms of recognition rate, can be achieved when the pre-processing step is employed.

Table 10
Recognition rates, using different numbers of training images per subject, on the Extended Yale-B database (LR:12×12).

Training images / subject	Recognition rate
10	0.9382
20	0.9474
30	0.9461

Table 11
Recognition rate, using different numbers of training images per subject, on the Multi-PIE database (LR:8 × 8).

Training images / subject	Recognition rate
10	0.9148
15	0.9720
20	0.9741

Table 12
Recognition rate, using different numbers of training images per subject, on the FERET (Fa) database (LR:12 × 12).

Training images / subject	Recognition rate
1	0.8966
2	0.9399

Table 13
Recognition rate, using different numbers of training images per subject, on the LFW database (LR:12 × 12).

Training images / subject	Recognition rate
5	0.6664
10	0.8823

Table 14
Recognition rates of our proposed method, with and without using the morphological pre-processing method.

Database	Recognition rate (without pre-processing)	Recognition rate (after pre-processing)
Extended Yale-B (160-D features)	0.882	**0.947**
Multi-PIE (160-D features)	0.922	**0.974**
FERET (Fa) (200-D features)	0.831	**0.896**
FERET (BaBe) (100-D features)	0.873	**0.915**
LFW (200-D features)	0.819	**0.883**

8 Conclusions

This paper addresses the problem of low-resolution face recognition by proposing a sparse-coding-based approach, which first extracts multiple local features (Gabor wavelets and LBPD) from face images and then decomposes them into a corresponding low-rank feature matrix and a sparse error matrix. The learned low-rank features are then projected into a new discriminative feature subspace using the proposed sparse-coding-based algorithm. It can be observed that sparsity plays an important role in discriminating face images of two different classes. Our proposed method performs sparse reconstruction in the training process, without searching for any nearest neighbors. The learned projection matrix also preserves the global structure of the data samples in the learned sparse feature subspace. For matching, a coefficient vector is computed to find the similarity between the training and testing image's features by using linear regression. Residual values are then computed based on the estimated coefficient vectors, which represents a testing feature, in terms of a set of training features. Finally, the LR query face image is then assigned to the class label with the least

residual value. Our objective function does not need to tune any model parameter. Experimental results demonstrate that our proposed method can achieve a better performance than other LR face recognition methods, in terms of recognition accuracy.

Acknowledgment

The work described in this paper was supported by the GRF Grant PolyU 152765/16E (project code: B-Q55J) of the Hong Kong SAR Government.

References

[1] M. Turk and A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3(1) (1991) 71-86.

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs Fisherfaces: Recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 17(7) (1997) 711-720.

[3] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3) (2005) 328-340.

[4] S. Baker, T. Kanade, Hallucinating faces, in: IEEE International Conference on. Automatic Face and Gesture Recognition, 2000, pp. 83-88.

[5] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9) (2002) 1167-1183.

[6] J. Yang, J. Wright, T. S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Transactions on. Image Processing 19(11) (2010) 2861-2873.

[7] X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Transactions on Systems, Man, and Cybernetics 35(3) (2005) 425-434.

[8] G. Qiu, A progressively predictive image pyramid for efficient lossless coding, IEEE Transactions on. Image Processing 8(1) (1999) 109–115.

[9] G. Qiu, Interresolution look-up table for improved spatial magnification of image, Journal of Visual Communication and Image Representation 11(4) (2000) 360–373.

[10] Y. Hu, K.-M. Lam, G. Qiu, T. Shen, From local pixel structure to global image super-resolution: A new face hallucination framework, IEEE Transactions on. Image Processing 20(2) (2011) 433–445.

[11] B. Li, H. Chang, S. Shan, X. Chen, Low-resolution Face Recognition via Coupled Locality Preserving Mappings, IEEE Signal Processing Letters 17(1) (2010) 20-23.

[12] C.-X. Ren, D.-Q. Dai, H. Yan, Coupled kernel embedding for low resolution face image recognition, IEEE Transactions on Image Processing 21(8) (2012) 3770-3783.

[13] P.H. Hennings-Yeomans, S. Baker, B. V. K. V. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: IEEE International conference on computer vision and pattern recognition, 2008, pp. 1-8.

[14] H. Huang, H. He, Super-Resolution Method for Face Recognition using Nonlinear Mappings on Coherent Features, IEEE Transactions on Neural Networks 22(1) (2011) 121-130.

[15] W. W. W. Zou, P. C. Yuen, Very low resolution face recognition problem, IEEE Transactions on Image Processing 21(1) (2012) 327-340.

[16] M.W. Jian, K.-M. Lam, Simultaneous Hallucination and Recognition of low-resolution faces based on singular value decomposition" IEEE Transactions on Circuits and Systems for Video Technology 25(11) (2015) 1761-1772.

[17] K.-H. Pong, K.-M. Lam, Multi-resolution feature fusion for face recognition, Pattern Recognition 47(2) (2014) 556-567.

[18] C. Zhou, Z. Zhang, D. Yi, Z. Lei, S. Li, Low-resolution face recognition via Simultaneous Discriminant Analysis , in: International joint conference on biometrics (IJCB), 2011, pp. 1-6.

[19] S. Biswas, K. W. Bowyer, P. J. Flynn, Multidimensional scaling for matching low-resolution Face images, IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10) (2012) 2019-2030.

[20] S. Siena, V.N. Bodetti, B.V. Kumar, Coupled marginal fisher analysis for low-resolution face recognition, in: European conference on computer vision, 2012, pp. 240-249.

[21] J. Shi, C. Qi, From Local Geometry to Global Structure: Learning Latent Subspace for Low-resolution Face Image Recognition, IEEE Signal Processing Letters, 22(5) (2015) pp. 554-558.

[22] J. Zhang, Z. Guo, X. Li, Y. Chen, Large Margin Coupled Mapping for Low Resolution Face Recognition, in: Pacific Rim International conference on Artifical Intelligence, 2016, pp. 661-672.

[23] Z. Wang, W. Yang, X. Ben, Low-resolution degradation face recognition over long distance based on CCA, Neural Computing and Applications 26(7) (2015) 1645-1652.

[24] J. Jiang, R. Hu, Z. Wang, Z. Cai, CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images, Signal Processing 124 (2016) 162-172.

[25] Y. Chu, T. Ahmad, G. Bebis, L. Zhao, Low-resolution face recognition with single image per person, Signal Processing, 141 (2017) 144-157.

[26] X. Xing, K. Wang, Couple manifold discriminant analysis with bipartite graph embedding for low-resolution face recognition, Signal Processing, 125 (2016) 329-335.

[27] F. Yang, W. Yang, R. Gao, Q. Liao, Discriminative multidimensional scaling for Low-resolution face recognition, IEEE Signal Processing Letters, 25(3) (2018) 388-392.

[28] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891-1898.

[29] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815-823.

[30] O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep Face Recognition, in: British Machine Vision Conference (BMVC), 2015, pp. 6.

[31] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: European conference on computer vision, 2016, pp. 499-515.

[32] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, SphereFace: Deep Hypersphere Embedding for Face Recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[33] M. Bereta, W. Pedrycz, M. Reformat, Local descriptors and similarity measures for frontal face recognition: A comparative analysis, Journal of Visual Communication and Image Representation, 24 (2013) 1213-1231.

[34] Z. Lei, M. Pietikainen, S. Z. Li, Learning Discriminant Face Descriptor, IEEE Transactions on Pattern Analysis and Machine Intelligence 36(2) (2014) 289-302.

[35] C.-H. Yoo, S.-W.Kim, J.-Y. Jung, S.-J. Ko, High-dimensional feature extraction using bit-plane decomposition of local binary patterns for robust face recognition, Journal of Visual Communication and Image Representation, 45 (2017) 11-19.

[36] J. Chen, V M. Patel, L. Liu, V. Kellokumpu, G. Zhao, M. Pietikainen, R. Chellappa, Robust local features for remote face recognition, Image and Vision computing, 64 (2017) 34-46.

[37] C.-F. Chen, C.-P. Wei, Y.-C.F. Wang, Low-Rank Matrix Recovery with Structural Incoherence for Robust Face Recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2618-2625.

[38] C.-P. Wei, C.-F. Chen, Y.-C.F. Wang, Robust Face Recognition With Structurally Incoherent Low-Rank Matrix Decomposition, IEEE Transactions on Image Processing 23(8) (2014) 4322-4334.

[39] J. Wright, A. Y. Yang, A. Ganesh, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2) (2008) 210-227.

[40] S. Zhang, X. Zhao, Locality-sensitive kernel sparse representation classification for face recognition, Journal of Visual Communication and Image Representation. 25 (2014) 1878-1885.

[41] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 32(11) (2010) 2106-2112.

[42] X. He and P. Niyogi, "Locality preserving projections, in: Conference on Neural Information Processing Systems, (2004), pp. 153.

[43] X. He, D. Cai, S. Yan, H. J. Zhang, Neighborhood preserving embedding, in: International conference on computer vision (ICCV), 2005, pp. 1208-1213.

[44] M.S. Shakeel, K.-M. Lam, Recognition of Low Resolution Face Images using Sparse Coding of Local Features, in: Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016, pp. 1-5.

[45] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-pie" Image Vis. Comput. 28(5) (2010) 807-813.

[46] P. J. Philips, H. Moon, S. A. Rizvi, P. J. Rauss, The FERET Evaluation Methodology for Face-Recognition Algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10) (2000) 1090-1104.

[47] G. B. Huang, M. Ramesh, T. Berg, E. L-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, 2007 (vol. 1, no. 2, p. 3), "Technical Report 07-49, UMass.

[48] R. Chellappa, J. Ni, V M. Patel, Remote identification of faces: Problems, prospects and progress, Pattern recognition letters, 33 (14) (2012) 1849-1859.

[49] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7) (2002) 971–987.

[50] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, IEEE Transactions on Image Processing 19(6) (2010) 1657–1663.

[51] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recognition 43(3) (2010) 706–719.

[52] G. Zhao, T. Ahonen, J. Matas, M. Pietikainen, Rotation-invariant image and video description with local binary pattern features, IEEE Transactions on Image Processing 21(4) (2012) 1465-1477.

[53] X. Hong, G. Zhao, M. Pietikainen, X. Chen, Combining LBP Difference and Feature correlation for texture description, IEEE Transactions on Image Processing 23(6) (2014) 2557-2568.

[54] H. Karcher, Riemannian center of mass and mollifier smoothing, Communications on Pure and Applied Mathematics 30(5) (1977) 509-541.

[55] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision (IJCV) 60 (2) (2004) 91-110.

[56] H. Bay, A. Ess, T. Tuytclaars, L-V. Gool, Speeded-Up Robust Features, Computer Vision and Image Understanding 110 (2008) 346-359.

[57] N. Dalal, B. Triggs, Histograms of Oriented Gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[58] T-S. Lee, Image Representation using 2D Gabor Wavelets, IEEE Transactions on Pattern Analysis and Machine Intelligence 18(10) (1996) 959-971.

[59] X. Tan, B. Triggs, Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions, IEEE Transactions on Image Processing 19(6) (2010) 1635-1650.

[60] K. Okajima, Two-dimensional Gabor-type receptive field as derived by mutual information maximization, Neural Networks, 11(3) (1998) 441-447.

[61] L. Nanni, A. Lumini, S. Brahnam, Local binary patterns variants as texture descriptors for medical image analysis, Artificial intelligence in medicine 49 (2) (2010) 117-125.

[62] L. Zhang, C. Ma, Low-Rank, Sparse matrix decomposition and group sparse coding for image classification, in: IEEE International conference on Image processing, 2012, pp. 669-672.

[63] Z. Lin, M. Chen, L. Wu, Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, UIUC Technical Report UILU-ENG-09-2215, Tech. Rep., 2009.

[64] V D Maaten, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2431-2456.

[65] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Processing Letters, 23(10) (2016) 1499-1503.

[66] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, arXiv preprint:1411.7923, 2014.

[67] J. Kannala, E. Rahtu, BSIF: binarized statistical image features, Proc. of International conference on pattern recognition, 2012.

[68] G. E. Hinton, R.R. Salukhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504-507.