




# Robust and efficient face recognition via low-rank supported extreme learning machine

Tao Lu<sup>1,2</sup>  · Yingjie Guan<sup>1</sup> · Yanduo Zhang<sup>1</sup> ·  
Shenming Qu<sup>3</sup> · Zixiang Xiong<sup>2</sup>

Received: 18 April 2017 / Revised: 23 November 2017 / Accepted: 29 November 2017  
© Springer Science+Business Media, LLC, part of Springer Nature 2017

**Abstract** Recently, face recognition algorithms have made great progress in various real-world applications, e.g., authentication and criminal investigation. Deep-learning offers an end-to-end paradigm for vision recognition tasks and achieves good performance. However, designing and training the complex network architecture are time-consuming and labor-intensive. Moreover, under complex scenarios, illumination change, noise or occlusion in images degrade the performance of recognition algorithms. In order to ameliorate these issues, we propose an efficient three-layered low-rank supported extreme learning machine (LSELM) algorithm for face recognition which improves the recognition performance under complex scenarios with high efficiency. In the first layer, a given probe sample is clustered into certain training subspace as pre-clustering. In the second layer, with this subspace, a low-rank subspace of probe sample as robust feature which is insensitive to disguise, noise, variant expression or illumination will be recovered by low-rank decomposition. Furthermore, these low-rank discriminative features are coded to support training a forward neural network termed LSELM. Experimental results indicate that the proposed approach is on par with some deep-learning based face recognition algorithms on recognition performance but with less time complexity over some popular face datasets e.g., AR, Extend Yale-B, CMU PIE and LFW datasets.

---

✉ Tao Lu  
lutxyl@gmail.com

✉ Zixiang Xiong  
zx@ece.tamu.edu

<sup>1</sup> School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430073, China

<sup>2</sup> Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

<sup>3</sup> School of Software, Henan University, Kaifeng 475001, China

**Keywords** Face recognition · Robust feature · Low-rank matrix recovery · Extreme learning machine · Time complexity

## 1 Introduction

Face recognition (FR), which is widely used in information security [4] and public safety, has been always a popular topic in computer vision and pattern recognition fields [15, 36]. In recent years, it has gained a wide interest in both theoretical research and industrial application domains. Although several face recognition products e.g., Deep face [26], Face++, are released in last decade. It still seems a huge challenge in real practice due to aging, occlusion, pose, illumination and expression variations and there is a significant room for improvement [28, 32].

Generally speaking, there are two key steps of face recognitions, one is feature learning and the other one is classifier. Benefiting from theory of computer vision, various hand-crafted feature-extraction methods have been proposed and achieved good performances [20]. For instance, Scale Invariant Feature Transform (SIFT) [39], Histogram of Oriented Gradient (HOG) [35], Speed-Up Robust Features (SURF) [5], and Local Binary Patterns (LBP) [1] feature-extraction methods, they are widely applied in traditional face recognitions. Considering the attribute of sample identity, subspace-learning based methods e.g., Eigenface [33], Fisherface [6], and Locality Preserving Projections (LPP) [16] are developed to incorporate label information from samples. Subspace-learning based algorithms are efficient and easy to be applied to real applications. However, performance of abovementioned classical face recognition algorithms drops when the input query images are disturbed by noise occlusions or other outliers.

In order to promote the discriminative ability of classifier, by the merit of sparse representation theory, Wright et al. [37] proposed a robust facial recognition algorithm that achieved satisfactory results even in noisy and complex conditions. This classifier, called as “sparse representation classifier (SRC)”, showed a strong discriminative ability by class-specific reconstruction residual. Sparse representation randomly selects the dictionary atoms to represent the query input image, sometimes, results in unstable solution which degrades the recognition performance. Some variants of sparse representation based approaches e.g., extend SRC [10], discriminative SRC [41] were proposed to improve the stability of sparse representation. Which enhances the discrimination of face recognition indeed, sparse or collaborative representation? Some researchers argued that another kind of sparsity: locality often helps face recognition. Collaborative representation based classifier (CRC) [40], used multi-category atoms to collaboratively represent the query input image and distinguished its label information from both reconstruction residual and the norm of coefficients. By the high efficiency of  $\ell_2$ -norm, locality-constrained representation [24], probabilistic collaborative representation [7] have been developed to furtherly promote the recognition performance.

Abovementioned representation-based classifiers and hand-crafted feature learning schemes give impressive results. However, hand-crafted feature design schemes limit the representation ability of massive data. Recently, deep learning based face recognition approaches as emerging machine learning paradigms become popular due to their powerful representation ability. In order to close the gap between signal features and their semantics, deep learning is used to automatically learn intrinsically features from large amounts of data [30]. Different architectures of convolutional neural network (CNN) were investigated in [18]

to improve the performance of recognition, very deep networks were verified in [29]. Chan et al. [9] proposed a simple yet powerful deep network using multi-layered principal component analysis to enrich the discriminative ability during feature extracting. This algorithm was considered as a baseline for deep-learning based face recognition on some famous face databases, e.g., AR, extend Yale-B, CMU PIE and LFW datasets. These deep-learning based approaches provide end-to-end solutions for big data application scenario. However, there are two shortcomings for deep learning based face recognition algorithms: GPU cluster which is essential for deep-learning optimization and large-scale data resources are always not available for general researchers [2, 3, 17]. On the other hand, designing and optimizing these complex networks are always time-consuming and labor-intensive. Above two shortcomings limit the large scale extensive applications of deep-learning based approaches, especially in resource-limited scenarios, e.g., mobile computing, autonomous robots.

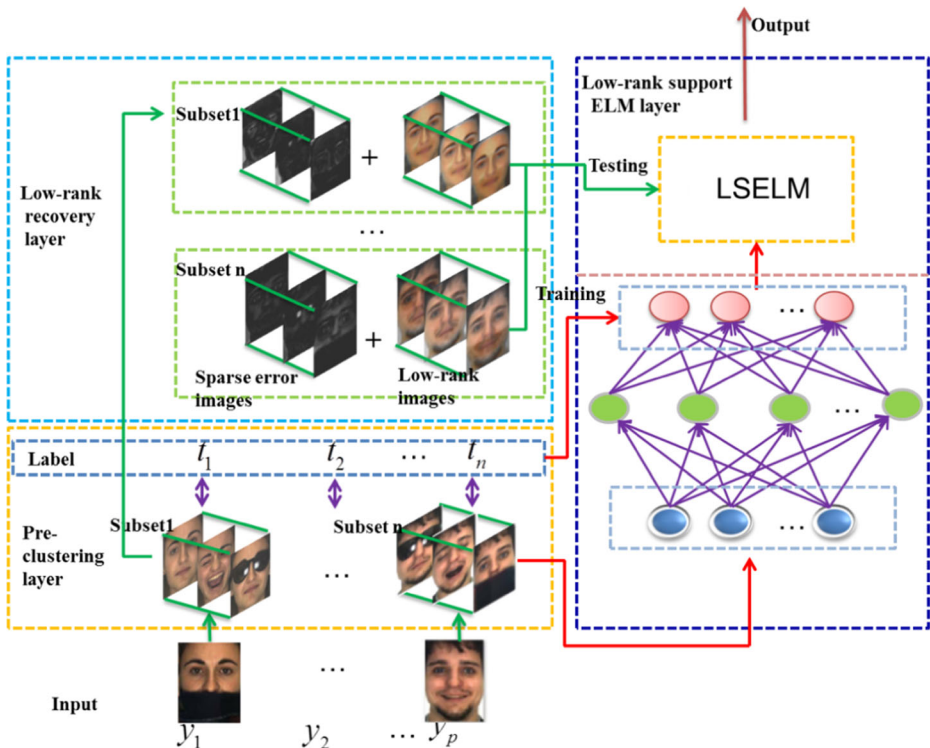
In fact, developing efficient and accurate face recognition algorithms are still challenging tasks. Especially, when the face images under complex conditions e.g., facial disguise, noise, and expression or illumination changes, performances of feature extraction and classification are both degraded. E. Candes et al. [8] theoretically proved that low rank minimization can be used to rectify the defects in an observation matrix by decomposing it into two parts: low-rank clear part and noise part. Thus this low-rank part of observation matrix would play more important role in recognition against noise. Du et al. [11] proposed a low-rank constrained sparse representation algorithm to enhance both feature representation and classification abilities. However, it suffered from high computational complexity of low-rank and sparse representation. Considering discriminations of traditional classifiers, e.g., K-Nearest Neighbor (KNN) [34] and Support Vector Machine (SVM) [12], they are in efficient computing manners, but just with limited accuracy. Some complex models including deep-learning based approaches, e.g. FaceIDs [29, 30] and PCANet [9], achieve impressive performance, but they are time-consuming to train their complex networks. Recently, a novel type of single hidden layer feed forward network, called “extreme learning machine (ELM)” offers a fast training paradigm for machine learning with strong generalization ability. ELM looks reasonable for accelerating training process in face recognition scenario. But [19] had pointed out that even though the ELM had advantages of low computational complexity and generalization ability, its prediction accuracy was sensitive to the noise in the input data. Thus, when the training or test data is noisy, the prediction performance of ELM drops dramatically.

To promote recognition performance with both high efficiency and accuracy [44], in this paper, we propose a novel low-rank supported extreme learning machine termed LSELM, aiming at extracting robust features, with fast learning manner and real application-oriented. By the merit of low-rank recovery, the contaminated input images can be exactly separated into clear inherent content and noise part. With those low-rank parts, the outliers would be transferred into their inherent contents which bring discriminative ability of feature representation. Furthermore, ELM is utilized to accelerate training phase to get fast training results. Consequently, by the low-rank content supported, the novel LSELM is robust to input noise and time-efficient for face recognition. The main contributions of the proposed approach can be summarized as follow: (1) We use low-rank recovery to decompose the input images into inherent part and noise part, including variable illumination, disguise, and expression changes. The low-rank part brings robustness of feature representation. (2) Low-rank supported extreme learning machine, not only promote the robust representation ability, but also in an efficient training manner. It ensures that the computational complexity of the proposed algorithm is lower than other traditional approaches. (3) The proposed three-layered architecture is easy to

deploy in real application scenario, with high performance which is on par with deep-learning based approaches, and lower time cost comparing with some-state-of-the-art approaches.

## 2 Robust FR via low-rank supported ELM

In real application scenarios, images always encounter different occlusions including noise, illumination changes and disguise. From the perspective of metric learning [42], occlusions in images result in pixel features having higher intra-variance than inter-variance, which degrade the recognition performance. In this paper, we propose a three-layered architecture of low-rank supported extreme learning machine algorithm to deal with the outliers in testing data. The details are shown in Fig. 1. In the first layer, for each given probe image, using the known label information from the training gallery dataset, we pre-cluster it into a certain sub-class as additional constraint matrix. The rationale behind this is that a testing image should lie in a sub-space for a linear representation. Then, each sub-class including testing samples is decomposed into a low-rank clear subspace without noise. This low-rank recovery step can be considered as the second layer. The last layer is ELM classifier, with the support of the clear low-rank subspace, clear context of features is fed into the ELM to exploit the complex relationship between representation coefficients and their labels to predict the final results.



**Fig. 1** Three-layered architecture of LSELM for robust face recognition. The green lines with arrow indicate testing process, and the red lines with arrow indicate training process

## 2.1 Pre-clustering layer

As we know, in scenario of face recognition task, every query sample should belong to certain subclass in the gallery dataset and it can be represented by a linear combination of its supported subspace (dictionary) [37]. If we use the subclass from gallery dataset, each input query image can be clustered into a subclass which includes clear facial images. Suppose that the input facial image  $y_m \in R^{d \times 1}$ ,  $m = 1, 2, 3, \dots, p$ ,  $p$  is the number of facial images,  $d$  is the dimension of input image in vector form. Given the gallery dataset as  $D = [D_1, D_2, \dots, D_n] \in R^{d \times l}$ , here  $D_i \in R^{d \times n_i}$  denotes the  $i$ -th subclass of gallery data,  $n_i$  denotes the sample number of this subclass,  $l = \sum_{i=1}^n n_i$ ,  $l$  is the number of gallery face images. Different from k-nearest neighbor, we use the nearest subclass as metric to cluster every input query image into certain subclass. The formula for the pre-clustering step can be written as follow:

$$pre-cluster(y_m) = \operatorname{argmin}_m \left\{ \left\| y_m - \frac{1}{n_i} \sum_{i=1}^{n_i} D_i \right\|_2 \right\}. \quad (1)$$

We compare the query image to every subspace of the gallery dataset, and find the class which is the closest to  $y_m$ . Once this subclass is selected, we plug the query image into the corresponding subclass matrix. For every query image  $y_m$ , we put it into its nearest subclass  $D_i$  to form a new augmented matrix  $X = [D_i, y_m] \in R^{d \times (l+1)}$ . It means every  $y_m$  should lie in a subclass and be linear represented by this subclass.

## 2.2 Low-rank matrix recovery layer

In this step, the new matrix  $X$  is decomposed into the sum of a low-rank sub-space matrix  $L$  and the sparse errors matrix  $E$  as follow:

$$\min_{L, E} \operatorname{rank}(L) + \lambda \|E\|_0 \text{ s.t. } X = L + E. \quad (2)$$

Where  $L \in R^{d \times (l+1)}$  is desired the clear sub-space matrix,  $E \in R^{d \times (l+1)}$  is the sparse errors matrix and value of  $\lambda$  controls the balance between the low-rank and the error term. However, this formula is a highly non-convex optimization problem. In order to solve this low-rank function, the nuclear norm is always a surrogate of this low-rank operation, and the  $\ell_0$ -norm can be replaced by  $\ell_1$ -norm. As a result, the above non-convex problem is relaxed into a convex one, given by:

$$\min_{L, E} \|L\|_* + \lambda \|E\|_1 \text{ s.t. } X = L + E, \quad (3)$$

where  $\|\cdot\|_*$  represents the nuclear norm, which is the sum of all singular values and  $\|\cdot\|_1$  represents the  $\ell_1$ -norm. In this paper, the Augmented Lagrange Multiplier (ALM) [22] method is adopted to solve the above minimization problem. Generally, the ALM algorithm is solved by following optimization function:

$$\min_X G(X) \text{ s.t. } F(X) = 0. \quad (4)$$

Where  $G$  is a convex function,  $F$  is a linear function. Equation (4) is solved by the Augmented Lagrange function:

$$L(X, Y, u) = G(X) + \operatorname{tr}[Y^T F(X)] + \frac{u}{2} \|F(X)\|_F^2. \quad (5)$$

Where  $u$  represents a positive scalar,  $Y$  represents Lagrange Multiplier. The corresponding Augmented Lagrange function is written as follows:

$$P(L, E, Y, u) = \|L\|_* + \lambda \|E\|_1 + \text{tr}[Y^t(X-L-E)] + \frac{u}{2} \|X-L-E\|_F^2, \quad (6)$$

where  $\lambda$  is used to control the entry of low-rank structure and its proportional relationship with the sparse error term.  $Y$  is Lagrange Multiplier, and  $\text{tr}$  represents the trace of the matrix. Alternating iterative method is used to solve the objective function. In the first step, we optimize  $L$  for a fixed  $E$ , and next, we solve  $E$  for the fixed  $L$ . We update  $L$  and  $E$  as follows:

$$\begin{aligned} L &= \arg \min_L \|L\|_* + \text{tr}[Y^t(X-L-E)] + \frac{u}{2} \|X-L-E\|_F^2 \\ &= \arg \min_L \frac{1}{u} \|L\|_* + \frac{1}{2} \left\| L - \left( \frac{1}{u} Y + X - E \right) \right\|_F^2. \end{aligned} \quad (7)$$

This function has an analytical solution through soft threshold shrinkage method [23]. After obtaining the value of  $L$ , we use the following method to solve the error term  $E$ :

$$\begin{aligned} E &= \arg \min_E \lambda \|E\|_1 + \text{tr}[Y^t(X-L-E)] + \frac{u}{2} \|X-L-E\|_F^2 \\ &= \arg \min_E \frac{\lambda}{u} \|E\|_1 + \frac{1}{2} \left\| E - \left( \frac{1}{u} Y + X - L \right) \right\|_F^2. \end{aligned} \quad (8)$$

The above objective function is a typical method of solving  $\ell_1$ -norm problems. In this paper, we use the SPAMS package [13] to solve it. The Lagrange multiplier operator  $u$  is updated by the following formula:

$$u = \min(u^* \rho, u_{\max}), \quad (9)$$

where  $\rho$  satisfies the condition  $\rho > 1$  to control the rate of increase of the value of  $u$ .  $u_{\max}$  is the upper bound of  $\rho$ .

The above low-rank matrix minimizes recovery can be summarized as follows:

---

**Algorithm 1** Minimization of the low-rank matrix by Augmented Lagrange Multiplier.

---

Input: Input data matrix  $X$ , balance parameter  $\lambda$ , initialize errors matrix  $E=0$ ,  $u>0$ ,  $\varepsilon = 10^{-7}$ .

Output: Low-rank matrix  $L$ , sparse error matrix  $E$ .

Do while

Step 1: Compute  $L_{k+1}$  according to formula (7).

Step2: Compute  $E_{k+1}$  according to formula (8).

Step3: Update  $u$  according to formula (9).

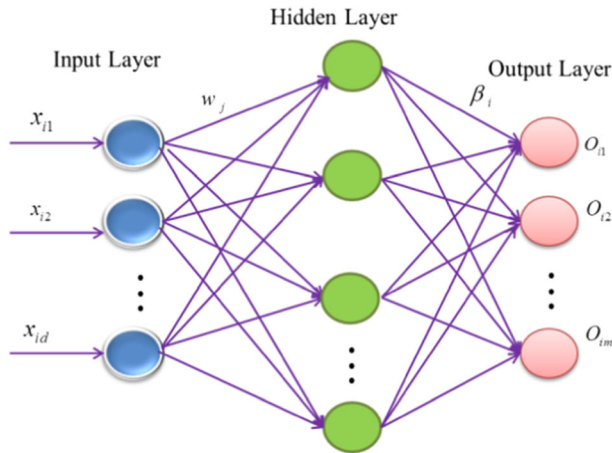
Check  $\|X-L-E\|_\infty < \varepsilon$

Repeat until convergence.

---

## 2.3 Robust FR via LSELM

ELM is different from the traditional neural network model, which can randomly assign input weights and hidden layer nodes, and then use the output weights to solve the least squares algorithm. As the Fig. 2 shown, the ELM avoids the iterative training process, leads to a



**Fig. 2** Network structure of extreme learning machine

significant improvement in the training speed of the algorithm compared to the traditional back-propagation (BP) and support vector machine (SVM) algorithms.

For the training dataset:  $L = \{(x_i, t_i) | i = 1, 2, \dots, N\}$ ,  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ ,  $t_i = (t_{i1}, t_{i2}, \dots, t_{im})^T$ , where  $N$  is the number of the dataset,  $d$  is the dimension of the input response signal,  $m$  is the number of classes, and  $t_i$  is the label information corresponding to input signal  $x_i$ . The activation function of  $k$ -hidden nodes  $f(x)$  is written as follows:

$$\sum_{j=1}^K \beta_j f_j(x_i) = \sum_{j=1}^K \beta_j f(w_j \cdot x_i + b_j) = o_i, i = 1, 2, \dots, N, \quad (9)$$

where  $w_j = (w_{j1}, w_{j2}, \dots, w_{jd})$  is the input weight vector corresponding to the  $j$ -th hidden node connected to the input nodes.  $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jm})^T$  is the weighted vector connecting the  $j$ -th hidden node to the output nodes.  $o_i = (o_{i1}, o_{i2}, \dots, o_{im})^T$  is the  $i$ -th data corresponding to the target vector.  $b_j$  is the deviation of the  $i$ -th hidden node,  $w_j \cdot x_i$  represents the inner dot product of  $w_j$  and  $x_i$ .

The objective of the single hidden layer neural network learning is to minimize the output error. It can be expressed as

$$\sum_{i=1}^{\tilde{N}} \|o_i - t_i\| = 0. \quad (10)$$

We rewrite above function with  $\beta_i$ ,  $w_j$ ,  $b_j$  as:

$$\sum_{j=1}^{\tilde{N}} \beta_j f(w_j \cdot x_i + b_j) = t_j, j = 1, 2, \dots, N. \quad (11)$$

This function can also be written in the matrix form as  $\beta^T H = T$ , where:

$$H = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \cdots & f(w_1 \cdot x_N + b_1) \\ f(w_2 \cdot x_1 + b_2) & \cdots & f(w_2 \cdot x_N + b_2) \\ \vdots & \vdots & \vdots \\ f(w_K \cdot x_1 + b_K) & \cdots & f(w_K \cdot x_N + b_K) \end{bmatrix}_{N \times K}, \beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_K^T \end{bmatrix}_{K \times m}, T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix}_{m \times N}.$$



Once the weights of the input layer and the hidden layer are randomly determined, the output matrix  $H$  of the hidden layer can be uniquely determined. Output weights  $\tilde{\beta}$  can be obtained using the Least Square method as:

$$\beta = \arg \min_{\beta} \|\beta^T H - T\|_2^2 = H^* T, \quad (12)$$

where  $H^*$  is the Moore-Penrose inverse of the matrix  $H$ . We summarize the proposed LSELM in algorithm 2.

---

**Algorithm 2** Robust FR via Low-rank supported ELM

---

Input: Gallery dataset  $D$ , ELM node number  $K$ , probe dataset  $y_m$ .

Output: Label information  $t_m$ .

**Training Phase**

Step1: Prepare the training sample  $D$  and the corresponding category label.

Step2: Establish the extreme learning machine network, optimize output weights  $\tilde{\beta}$  by formulate (12).

**Testing Phase**

Step 1: Obtain the testing sample by Algorithm1.

Step2: Extract the representation of facial image.

Step3: Output its label information  $t_m$  by trained network.

---

## 2.4 Time complexity

Time complexity of the algorithm quantifies the amount of time taken by an algorithm [27]. Let symbol  $O$  represents the time complexity in the algorithm, which excludes coefficients and lower order terms. Time consumption includes pre-clustering step, low-rank recovery step and ELM step. First, time cost of pre-clustering layer is  $O(n)$ , and time cost of low-rank matrix recovery layer is tracing norm computation in Eq. (7) and (8), with  $O(d^2(l+1))$  for  $L \in R^{d \times (l+1)}$ . Generally,  $d$  is the same order of magnitude with  $l$ . In the ELM algorithm step, weights of the input layer and the hidden layer are randomly determined, the output matrix  $H$  of the hidden layer having time complexity equal to  $O(KNd)$ , and time-cost comes from calculation of the network for output weight matrix  $\tilde{\beta}$  through (12), it has  $O(K^3 + K^2N + KNm)$  time complexity. For training the ELM network, its time cost is equal to  $O(K^3 + K^2N + KN(m+d)) \approx O(K^3 + K^2N)$ . The total cost of testing a sample is  $O(d^2(l+1) + K^3 + K^2N)$ .

## 3 Experimental results

We conduct experiments on four famous facial databases: AR [25], Extended Yale-B [14], CMU PIE [43] and Labeled Faces in the Wild (LFW) [21] datasets. To show the effectiveness of the proposed method, we compare recognition performance with three state-of-the-art methods: LSRC [11], CRC [40], SRC [37], three common-use classifiers including ELM [19], SVM [12] and NN [34]. Additionally, we choose other three famous face recognition algorithms including PCANet [9] (state-of-the-art deep learning based face recognition algorithm, since FaceIDs use large-scale dataset and GPU base deep-learning optimization which are not available for our experiments, for fairly comparing, we do not compare our algorithm



with them.), the LBP [1] (typical feature based face recognition algorithm) and P-LBP [31]. All the parameters are used in their papers for their best performance.

### 3.1 Databases

**AR database** has 100 individuals, and each individual having about 26 images, in totally 2600 facial images. These images are taken under conditions with variations in illumination, expression and disguise. The size of image is cropped into  $160 \times 120$  pixels. In the experiments, we randomly choose 8 samples for each individual for training, and the rest are for testing.

**Extended Yale-B database** contains 16,128 facial images of 38 subjects under 9 pose and 64 illumination conditions. The size of image is cropped  $192 \times 168$  pixels. In our experiments, we choose 60 front facial images with illumination changes to form a total of 2280 facial images. A random subset with 20 samples per individual is selected for training, and the rest are for testing.

**CMU PIE database** contains over 40,000 facial images of 68 individuals. Images of each individual were acquired across 13 different poses, under 43 different illumination conditions, and with 4 different expressions. In experiments, we chose 57 people a total of 2565 different light changes under the positive face, each person has 45 images, and all images are  $64 \times 64$  pixels. In addition, we randomly choose 12 instances per individual for training, the rest are for testing.

**LFW** is a dataset for studying unconstrained face recognition, which contains more than 13,000 images of faces collected from the web. Among them, 1680 of the people images have two or more distinct photos in the database. Here we only select 136 people who own more than 10 images and totally as 2886 facial images, and we randomly choose 10 samples per individual for training, and the rest are for testing in this experiment. All images are detected and cropped by the Viola-Jones face detector into  $64 \times 64$  pixels [38].

The experiments are conducted with the same paradigm that uses the same gallery of training and testing dataset for face recognition without overlap. The experiments are carried out to testify three aspects: for robust facial feature extraction, performance of classification and the case study with deep-learning algorithm. We also test the performance of the proposed algorithm under complex conditions.

### 3.2 Low-rank matrix recovery

In order to show the effectiveness of low-rank matrix recovery against outlier, we use the pre-clustered matrix to remove the noisy parts from original image. Benefiting from the label information of the facial training samples, the testing image is pre-clustered under an assumption that the class information of face recognition is known in advance. We randomly select the training sample and all the rest of the images are used as the probe dataset. Figures 3, 4, 5 and 6 shows the low-rank matrix recovery results, with  $\lambda = 0.005$ .

We show the subjective results from low-rank recovery by subclass augment matrix from Figs. 3, 4, 5 and 6. It is obvious that low-rank subspace successfully reduces the variations in expression, disguise, illumination and posture. Without the nuisance comes from above outliers, the clear low-rank subspace should be more discriminative than the original images.

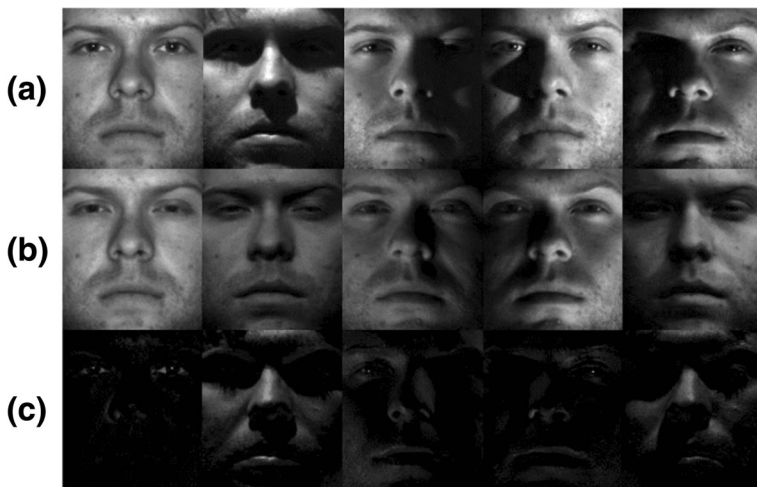


**Fig. 3** Low-rank recovery images from the AR database, **a** Original images; **b** Low-rank images; **c** Error images. Variations in expression and disguise (sunglass and scarf) are effectively removed by low-rank constraint

### 3.3 Face recognition performance

#### 3.3.1 The relationship between pre-clustering and recognition rate

Intuitively the performance of the low-rank matrix recovery depends on the accuracy of the pre-clustering. To investigate the relationship between the pre-clustering accuracy and the final recognition rate, we designed a further experiment. For this experiment we randomly choose 8



**Fig. 4** Low-rank recovery images from the extended Yale-B database, **a** Original images; **b** Low-rank images; **c** Error images. Variations in illumination are removed, and low-rank parts of subspace seem clearer than original images



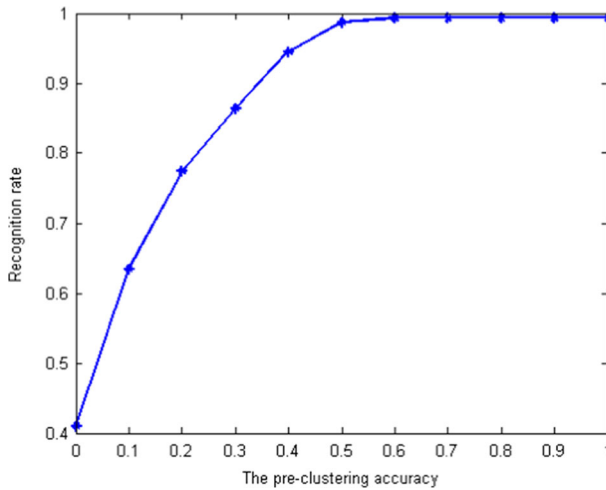
**Fig. 5** Low-rank recovery images from the CMU PIE dataset, **a** Original images; **b** Low-rank images; **c** Error images. Illumination and posture in this configuration are well removed

faces from each category of AR database as the gallery dataset. We use the rest of the 1800 images for pre-clustering. After pre-clustering, images are performed using low-rank matrix recovery combined with ELM classifier. First, two extreme cases are selected: that 1) the pre-clustering images are all wrong and 2) they are all in pairs. Further experiments showed that even the pre-clustering accuracy is not high (just at 40%), the final recognition result is still acceptable, shown in Fig. 7.

From Fig. 7, we can see that for these extreme cases the recognition rate is 0.415 when the pre-clustering is completely wrong, and the recognition rate is 1 when the pre-clustering images are all in pairs. When the correct pre-clustering rate is 50%, the recognition rate is 0.988. After pre-clustering, the classifier is used to classify the data without applying a low-rank matrix recovery. In this case, when the correct pre-clustering rate is 100% the recognition rate is 0.924, and when the correct pre-clustering rate is 50%, the recognition rate drops to



**Fig. 6** Low-rank recovery images from the LFW dataset, **a** Original images; **b** Low-rank images; **c** Error images. Variations in posture are well aligned in the experiments

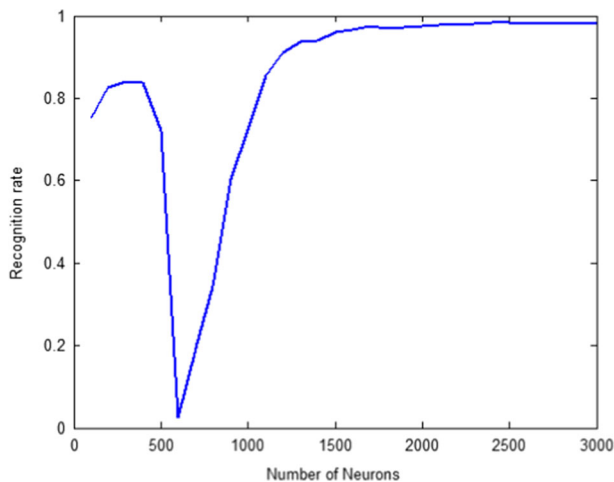


**Fig. 7** The relationship between pre-clustering accuracy and recognition rate. X axis indicate the pre-cluster accuracy from 0 to 100%

0.487. We can observe that low rank matrix recovery significantly improves the result, and the final recognition is not sensitive to the pre-clustering. In the presence of low-rank matrix recovery, the recognition rate becomes 0.945 when the pre-clustering correct rate is 40%. This indicates that low-rank matrix recovery after pre-clustering plays a key role in improving the classification accuracy. In the case of subclass misclassification, low-rank recovery is still very effective to demonstrate the effectiveness of the proposed algorithm.

### 3.3.2 The influence of neurons number on recognition rate

ELM is a typical neural network, the number of neurons effect the performance of the training networks. The goal of this experiment is to obtain the best number of neurons to achieve the



**Fig. 8** The relationship between recognition rate and the number of neurons

best performance of the classifier. We perform experiments on AR database. The recognition rate varies with the number of neurons from 50 to 3000, as shown in Fig. 8.

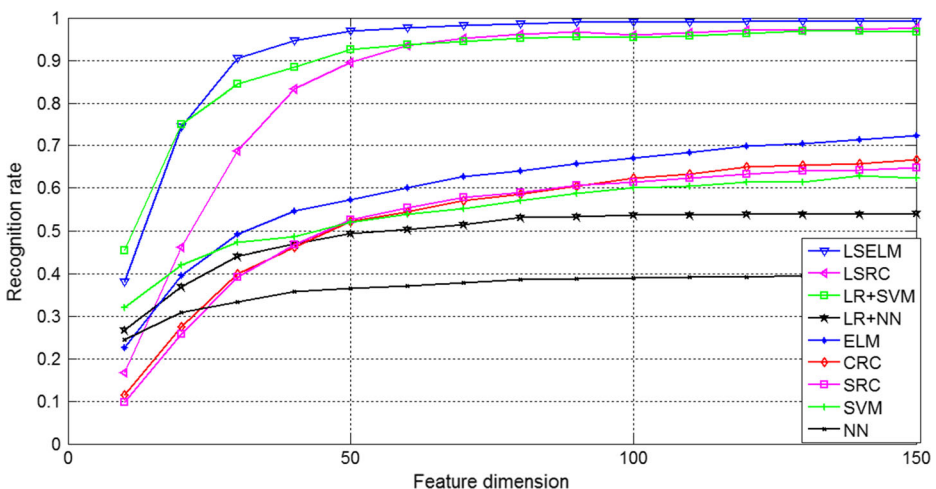
From Fig. 8, we observe that when the number of neurons is too high or too low, it will affect the recognition rate of the classifier. The best performance of ELM classifier is obtained when the number of neurons is at least more than 1600. However, as the number of neurons increased, the time complexity of training also increases. Therefore, in this paper, we set the number of neurons as 2000 to balance the time cost and recognition performance.

### 3.3.3 Recognition rate of different classifiers on different face datasets

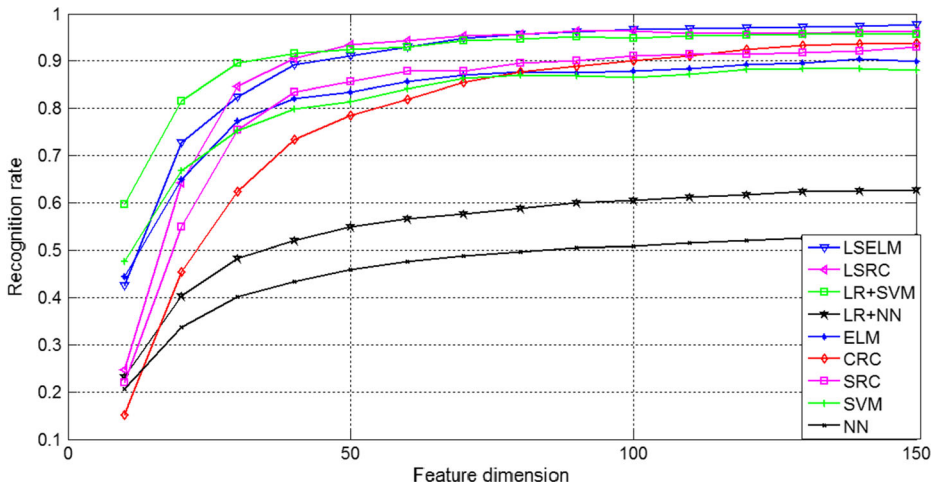
We set up the gallery dataset with the original samples and the probe dataset with the low-rank recovery parts of the testing samples. The sizes of each image are  $45 \times 50$  pixels in AR dataset and Extend Yale-B dataset. In the CMU-PIE database, the sizes of each image are  $32 \times 28$  pixels. In the FLW database, images are  $64 \times 64$  pixels. Several classifiers are cascaded with the Low-rank Recovery (LR) as benchmarks. The list of classifiers and their combinations are as follows: 1) LR + NN, 2) LR + SVM, 3) LSRC [11], 4) the original NN [34], 5) SVM [12], 6) SRC [37], 7) CRC [40]. We vary the dimension of by eigenface to exact feature representation, and we compare the recognition performance between different methods. Results of the classifiers are shown in Figs. 9, 10, 11 and 12 with different databases.

It can be seen from Fig. 9 that with an increase of the feature dimension, there is a gradual increase in the recognition rates. With the low-rank feature support, the recognition rate of ELM is higher than those of SVM, NN and LSRC algorithms. With feature dimension set 100, our method outperforms ELM with more than 25% improvement for recognition rate. In the case of small number of samples and fewer features, the ELM classifier shows a better recognition performance too. The low-rank subspace successfully reduces the negative impacts under disguise and illumination changes, so that recognition rate significantly improves.

In the extended Yale-B database, traditional FR techniques such as SRC and CRC can achieve good performance. We notice that in the AR database, Low-rank supported features boost the recognition rate more than in the Extended Yale-B dataset. Compared to the case of



**Fig. 9** Recognition rate for different FR algorithms on AR datasets

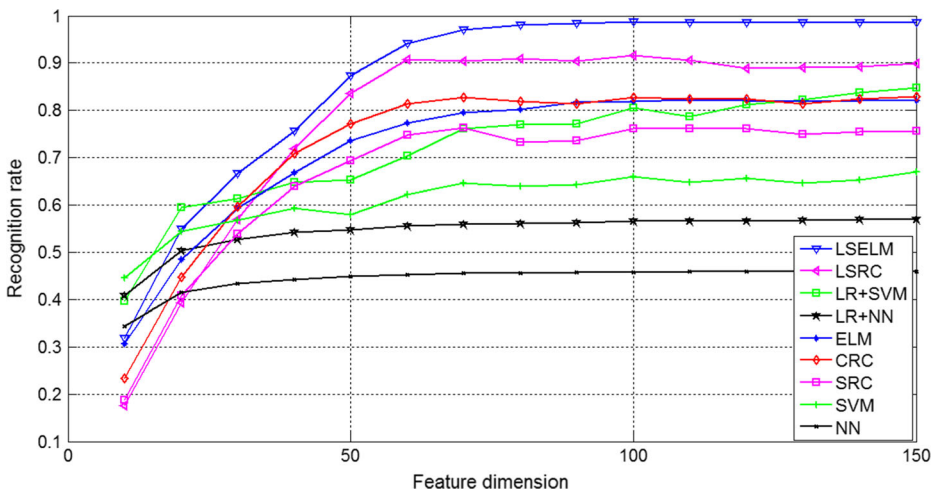


**Fig. 10** Recognition rate for different FR algorithms on Extend Yale-B datasets

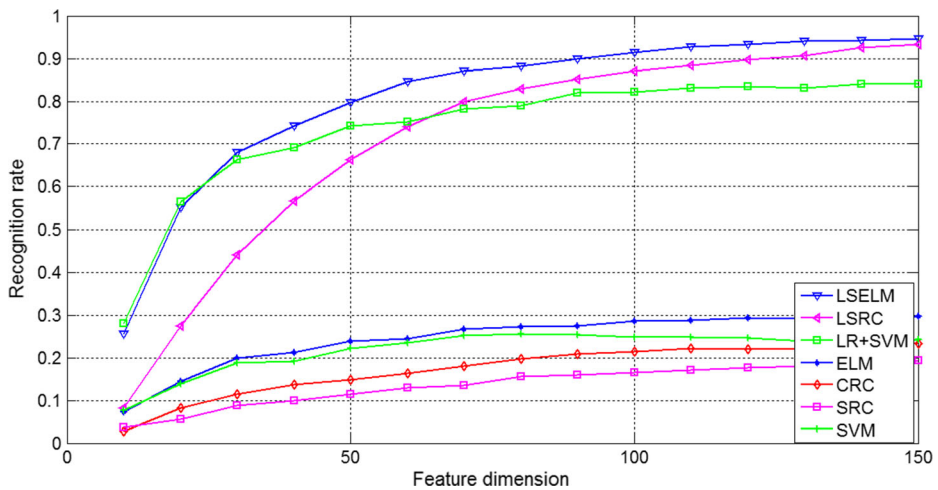
without the support of low-rank features, there is a significant improvement of these face recognition algorithms in the case of the low-rank feature support. From these results, we confirm that the low-rank subspace successfully reduces the negative impacts under illumination changes.

From Fig. 11, while LSRC based approaches obtains better results than other methods does (e.g., LR + SVM, LR + NN), our method achieved the highest recognition rates and outperformed all other approaches. For example, with a feature dimension of 100, our method outperforms than all the other methods with more than 6% improvement of recognition rate. From the experimental results on the CMU PIE database, the effectiveness of the proposed algorithm can be verified.

From Fig. 12, we can conclude that not all the mentioned methods have achieved impressive recognition results. LFW dataset is under unconstraint environment, which needs



**Fig. 11** Recognition rate for different FR algorithms on CMU PIE datasets



**Fig. 12** Recognition rate for different FR algorithms on FLW datasets

more priors or regularizations to be incorporated into learning framework. The proposed method is superior to other state-of-the-art methods. It also shows that its robustness to noise, occlusion, which is critical to improve recognition performance.

From above figures, with an increase in the feature dimension, there is a gradual increase in the recognition rates. With the low-rank feature support, the recognition rate of ELM is higher than those of SVM, NN and SRC algorithms. In case of without the support of low-rank features, the original ELM still achieves acceptable results. Hence with or without low-rank feature support, there is a significant improvement of these face recognition algorithms, thus it proves that both low-rank feature support and ELM are favorable for improving the recognition performance.

### 3.3.4 Time efficiency and recognition rate of different classifiers on different face datasets

Furthermore, in order to investigate the efficiency of these algorithms, we report the time-cost of all listed algorithms in Table 1. In this experiment, the feature dimension is 100 for all the

**Table 1** Time efficiency comparison for different FR algorithms

Databases	Methods	LSELM	LSRC	LR + SVM	LR + NN
AR	IT (s)	0.1403	114.125	27.536	5.918
	TT (s)	0.0698	—	34.214	—
	RR	0.9890	0.9716	0.9540	0.4865
Extended Yale-B	IT(s)	0.1701	155.495	3.9512	8.053
	TT(s)	0.1271	—	13.6496	—
	RR	0.9670	0.9630	0.95	0.605
CMU PIE	IT(s)	0.1149	135.9426	4.927	6.182
	TT(s)	0.1315	—	15.7801	—
	RR	0.9885	0.9160	0.806	0.4727
LFW	IT(s)	0.2820	121.5315	54.2710	20.341
	TT(s)	0.4865	—	132.4849	—
	RR	0.9187	0.8702	0.8202	0.7321





**Fig. 13** The cropped images of one person of the AR dataset. The single natural image is used for training, while the other 12 images with several variations are used for testing

databases. We compare the training time (TT) and the identification times (IT) and the recognition rate (RR).

From Table 1, LSELM algorithm has the lowest computational complexity among all the competitive algorithms in AR database, Extended Yale-B database, CMU PIE database and LFW database. The training times of LSELM are 0.1403s, 0.1701s, 0.1149s and 0.2820 s. Coupled with the identification time, it is still shorter than other classifiers, it is clear that LSELM has advantage of efficiency in the classification. Notice that all these algorithms are based on the same low-rank features so the running time for low-rank recovery is not listed. LSELM has the best recognition rate and the lowest computational complexity among all the competitive algorithms. Considering the time-consumption for feature extraction, we observe that the running times for low-rank recovery and LBP, are 0.29 s and 1.37 s per image respectively. So, we can conclude that the efficiency from using the ELM classifier during training and testing is beneficial to real-world applications.

### 3.4 LSELM vs PCANet

Deep-learning is widely applied to face recognition in recent years. To investigate the robustness of our algorithm for the face recognition task, we compare the proposed algorithm with state-of-the-art face recognition algorithms such as the P-LBP, PCANet, and the LBP. All the configurations used for these experiments are the same as in PCANet. We conduct

**Table 2** Recognition rate for different FR algorithms under different testing conditions

Testing sets	AR dataset				Extend Yale-B dataset	CMU PIE dataset	FLW dataset
	Illumination changes	Expression changes	Disguise	Disguise& Illumination	Illumination changes	Illumination changes	Posture & Illumination
LBP	0.9383	0.8133	0.9125	0.7963	0.7576	0.7977	0.7547
P-LBP	0.9750	0.8033	0.9300	0.8858	0.9613	—	—
PCANet-1	0.9850	0.8567	0.9575	0.9275	0.9777	0.9870	0.8255
PCANet-2	0.9950	0.8500	0.9700	0.9500	0.9958	1	0.8628
LSELM	0.9975	0.8467	0.9725	0.9587	0.9701	0.9885	0.9187

**Table 3** Time consumption of different FR algorithms

Testing databases	LSELM(s)	LBP(s)	PCANet(s)
AR	1320.73	6629.5	9136.35
Extend Yale-B	1135.56	5687.1	11,665.92
CMU PIE	1166.32	5701.3	2515.11
FLW	900.87	5523.5	2031.58

experiments on different databases, with frontal illumination and neutral expression facial images as the training set, the testing set is divided into the following several ways: expression changes, illumination variation, disguise and disguise and illumination for testing (on the AR dataset), illumination variation for testing (on the Extended Yale-B and CMU PIE datasets) and posture & illumination as the testing set (on the FLW dataset), as shown in Fig. 13.

Table 2 shows the result of recognition rate for different FR algorithms under different testing conditions. LBP features can achieve good performance while dealing with illumination changes and the conditions of disguise and illumination changes. However, PCANet which applies deep-learning outperforms both LBP and P-LBP algorithms. We also observe that our algorithm is better than PCANet under the condition of disguise, disguise and illumination changes (on the AR dataset) and posture and illumination changes (on the LFW dataset), but its recognition rate is lower than PCANet under the conditions of expression changes (on the AR dataset) and illumination variation (on the Extend Yale-B dataset and CMU PIE dataset). However, our proposed algorithm still outperforms than LBP and P-LBP algorithm. For a fair comparison of time efficiency, we compare our result (as shown in Table 3) with PCANet just for the training and testing (without feature learning). Those preliminary results illustrate that the LSELM is on par with the PCANet algorithm in terms of recognition performance but has lower time complexity. The running time is about 1/9 of PCANet in AR database, 1/10 of PCANet in Extend Yale-B database, 1/4 of PCANet in CMU PIE database and 1/2 of PCANet in FLW database. When our recognition rate is 1% lower than PCANet, our time complexity is 1/4 of PCANet. Therefore, it is shown that the LSELM in this paper obtain better performance to the state-of-the-art deep-learning algorithm. In a word, LSELM is efficient and robust to disguise, and illumination variations.

## 4 Conclusion

In real-world FR scenarios, illumination variations, disguise and noise often degrade the recognition performance. Moreover, training complex networks are time-consuming, which limits their applications on source-limited scenarios, e.g., mobile computing or battery-dependent robot. In this paper, we propose a novel low-rank supported ELM algorithm to balance a tradeoff between the computational burden and recognition performance. Low-rank recovery is verified to be effective to noisy inputs. Extensive experimental results prove that the proposed model is not only on par with PCANet (deep-learning based approach) over the AR, Extended Yale-B, CMU PIE and FLW databases in term of recognition performance, but also has lower time complexity. Our method has robust ability to eliminate the outliers in input images and fast training speed by the merit of ELM. Fast training manner is naturally suitable for resource-limited applications. On the other hand, the proposed method uses low-rank recovery to extract features which is in hand-craft manner, mechanisms for assigning weights

randomly in ELM may result in unstable prediction performance. In future, we plan to learn the robust feature from training data automatically. Furthermore, the analysis of the observed relationship between the low-rank support matrix and the final accuracy will be fully investigated.

**Acknowledgements** This work is supported by the grant of China Scholarship Council, the National Natural Science Foundation of China (61502354, 61501413, 61671332, 41501505), the Natural Science Foundation of Hubei Province of China (2012FFA099, 2012FFA134, 2013CF125, 2014CFA130, 2015CFB451), Hubei Chenguang Talented Youth Development Foundation, Scientific Research Foundation of Wuhan Institute of Technology(K201713). Thanks for the reviewers' valuable comments for the paper.

## References

1. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
2. Al-Ayyoub M, AlZu'bi S, Jararweh Y, Shehab MA, Gupta BB (2016) Accelerating 3D medical volume segmentation using GPUs. *Multimedia Tools and Applications*. Springer New York LLC, New York, pp. 1–20
3. Alsmirat MA, Jararweh Y, Al-Ayyoub M, Shehab MA, Gupta BB (2017) Accelerating compute intensive medical imaging segmentation algorithms using hybrid CPU-GPU implementations. *Multimedia Tools and Applications* 76(3):3537–3555
4. Atawneh S, Almomani A, Al Bazar H, Sumari P, Gupta B (2017) Secure and imperceptible digital image steganographic algorithm based on diamond encoding in DWT domain. *Multimedia Tools and Applications* 76(18):18451–18472
5. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). *Comput Vis Image Underst* 110(3):346–359
6. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
7. Cai S, Zhang L, Zuo W, Feng X (2016) A probabilistic collaborative representation based approach for pattern classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2950–2959
8. Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? *Journal of the ACM (JACM)* 58(3):11
9. Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y (2015) PCANet: a simple deep learning baseline for image classification? *IEEE Trans Image Process* 24(12):5017–5032
10. Deng W, Hu J, Guo J (2012) Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans Pattern Anal Mach Intell* 34(9):1864–1870
11. Du HS, Hu QP, Qiao DF, Pitas I (2015) Robust face recognition via low-rank sparse representation-based classification. *Int J Autom Comput* 12(6):579–587
12. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16(10):906–914
13. Furrer R, Sain SR (2010) Spam: a sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *J Stat Softw* 36(10):1–25
14. Georgiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans Pattern Anal Mach Intell* 23(6):643–660
15. Gupta B, Agrawal DP, Yamaguchi S (2016) Handbook of research on modern cryptographic solutions for computer and cyber security. IGI Global, Hershey. <https://doi.org/10.4018/978-1-5225-0105-3>
16. He X, Niyogi P (2004) Locality preserving projections. In: *Advances in Neural Information Processing Systems 16 - Proceedings of the 2003 Conference, NIPS 2003, Vancouver, BC, Canada*, pp. 153–160
17. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507

18. Hu G, Yang Y, Yi D, Kittler J, Christmas W, Li SZ, Hospedales T (2015) When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. In Proceedings of the IEEE international conference on computer vision workshops, pp. 142–150
19. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B Cybern* 42(2):513–529
20. Jararweh Y, Al-Ayyoub M, Fakirah M, Alawneh L, Gupta BB (2017) Improving the performance of the needleman-wunsch algorithm using parallelization and vectorization techniques. *Multimedia Tools and Applications*. Springer New York LLC, New York, pp. 1–17
21. Learned-Miller E, Huang GB, RoyChowdhury A, Li H, Hua G (2016) Labeled faces in the wild: a survey. In *Advances in face detection and facial image analysis*, pp. 189–248
22. Lin Z, Chen M, Ma Y (2010) The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*
23. Lingling Z, Huaxiang W, Yanbin X, Da W (2011) A fast-iterative shrinkage-thresholding algorithm for electrical resistance tomography. *WSEAS Trans Circuits Syst* 10(11):393–402
24. Lu T, Xiong Z, Zhang Y, Wang B, Lu T (2017) Robust face super-resolution via locality-constrained low-rank representation. *IEEE Access* 5:13103–13117
25. Martinez AM, Kak AC (2001) PCA versus LDA. *IEEE Trans Pattern Anal Mach Intell* 23(2):228–233. <https://doi.org/10.1109/34.908974>
26. Milborrow S, Nicolls F (2008) Locating facial features with an extended active shape model. In *European conference on computer vision*. Springer, Berlin, pp. 504–513
27. Pan Z, Zhang Y, Kwong S (2015) Efficient motion and disparity estimation optimization for low complexity multiview video coding. *IEEE Trans Broadcast* 61(2):166–176
28. Sun Y, Wang X, Tang X (2015) Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2892–2900
29. Sun Y, Liang D, Wang X, Tang X (2015) Deepid3: face recognition with very deep neural networks. <http://arxiv.org/abs/1502.00873>
30. Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708
31. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process* 19(6):1635–1650
32. Tirkaz C, Albayrak S (2009) Face recognition using active appearance model. In *proceedings of the 17th signal processing and communications applications conference*. IEEE pp. 940–943
33. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In *proceedings of 1991 I.E. Computer Society Conference on Computer Vision and Pattern Recognition* pp. 586–591
34. Veenman CJ, Reinders MJ (2005) The nearest subclass classifier: a compromise between the nearest mean and nearest neighbor classifier. *IEEE Trans Pattern Anal Mach Intell* 27(9):1417–1429
35. Wang X, Han TX, Yan S (2009) An HOG-LBP human detector with partial occlusion handling. In *Proceeding of 2009 the 12th international conference on computer vision*, IEEE. pp. 32–39
36. Wang J, Li T, Shi YQ, Lian S, Ye J (2017) Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics. *Multimed Tools and Appl* 76(22):23721–23737
37. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227
38. Yu C, Li J, Li X, Ren X, Gupta BB (2017) Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram. *Multimedia Tools and Applications*. Springer New York LLC, New York, pp. 1–24
39. Yuan C, Xia Z, Sun X (2017) Coverless image steganography based on SIFT and BOF. *J Int Technol* 18(2): 435–442
40. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In *Computer vision (ICCV), 2011 I.E. international conference*, pp.471–478
41. Zhang Z, Liang Y, Bai L, Hancock ER (2016) Discriminative sparse representation for face recognition. *Multimedia Tools and Applications* 75(7):3973–3992
42. Zhang Y, Sun X, Wang B (2016) Efficient algorithm for k-barrier coverage based on integer linear programming. *China Communications* 13(7):16–23
43. Zheng M, Bu J, Chen C, Wang C, Zhang L, Qiu G, Cai D (2011) Graph regularized sparse coding for image representation. *IEEE Trans Image Process* 20(5):1327–1336
44. Zhou Z, Yang CN, Chen B, Sun X, Liu Q, QM J (2016) Effective and efficient image copy detection with resistance to arbitrary rotation. *IEICE Trans Inf Syst* 99(6):1531–1540



**Tao Lu** received the B.S and M.S. degrees from School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China in 2003 and in 2008 respectively, and the Ph.D. degree from National Engineering Research Center For Multimedia Software, Wuhan University, in 2013. Now he is an associate professor in School of Computer Science and Engineering, Wuhan Institute of Technology and a research member in Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology). He was a Post-doc in Department of Electrical and Computer Engineering, Texas A&M University (Wuhan Institute of Technology). His research interests include image/video processing, computer vision and artificial intelligence.



**Yingjie Guan** received the B.S degree in Communication Engineering from the College of Post and Telecommunication of WIT, Wuhan, China in 2015. She is currently pursuing the M.S. degree in Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China. Her research interests include image processing and computer vision.



**Yanduo Zhang** received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 1999. He is a Professor with the Wuhan Institute of Technology, Wuhan, China. He is currently the Vice President with the Wuhan Institute of Technology, and the Director of the Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology). He was as Visiting Professor with the Center for Vision, Cognition, Learning, and Art, University of California, Los Angeles between 2012 and 2013. He is a member of the IEEE and ACM. He has coauthored 2 books and published more than 100 research papers. His research interests include computer vision, robot control, and intelligent computing.



**Qu Shenming** received the B.S degree in computer science from Hebei University, Baoding, China, in 2004, the M.S. degree in computer application technology from Henan University, Kaifeng, China, in 2010. He is currently pursuing the Ph.D. degree in National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China. Currently he is a lecturer at Henan University. His research interests include image processing, computer vision.



**Zixiang Xiong** received the Ph.D. degree in Electrical Engineering in 1996 from the University of Illinois at Urbana-Champaign. From 1995 to 1997, he was with Princeton University, first as a visiting student, then as a research associate. From 1997 to 1999, he was with the University of Hawaii. Since 1999, he has been with the Department of Electrical and Computer Engineering at Texas A&M University, where he is a professor. During Spring 2010, he spent his sabbatical leave at Stanford University. He received an NSF Career Award in 1999, an ARO Young Investigator Award in 2000 and an ONR Young Investigator Award in 2001. He also received the 2006 I.E. Signal Processing Magazine best paper award and top 10% paper awards at the 2011 and 2015 IEEE. Multimedia Signal Processing Workshop. He is a fellow of the IEEE. His research interests include image processing, source-channel coding and network information theory.