Facial Gender Recognition via Low-rank and Collaborative Representation in An Unconstrained Environment

Ning Sun, Hang Guo, Jixin Liu, Guang Han
Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[e-mail: sunning@njupt.edu.cn]
Corresponding author: Ning Sun

Received January 4, 2017; revised April 23, 2017; accepted May 25, 2017; published September 30, 2017

Abstract

Most available methods of facial gender recognition work well under a constrained situation, but the performances of these methods have decreased significantly when they are implemented under unconstrained environments. In this paper, a method via low-rank and collaborative representation is proposed for facial gender recognition in the wild. Firstly, the low-rank decomposition is applied to the face image to minimize the negative effect caused by various corruptions and dynamical illuminations in an unconstrained environment. And, we employ the collaborative representation to be as the classifier, which using the much weaker l_2-norm sparsity constraint to achieve similar classification results but with significantly lower complexity. The proposed method combines the low-rank and collaborative representation to an organic whole to solve the task of facial gender recognition under unconstrained environments. Extensive experiments on three benchmarks including AR, CAS-PERL and YouTube are conducted to show the effectiveness of the proposed method. Compared with several state-of-the-art algorithms, our method has overwhelming superiority in the aspects of accuracy and robustness.

Keywords: Facial gender recognition, low-rank decomposition, collaborative representation, unconstrained environment

https://doi.org/10.3837/tiis.2017.09.018 **ISSN : 1976-7277**

1. Introduction

Over the past few decades, facial image analysis had always been a hot topic in the field of computer vision. Facial gender recognition, which was a technique of judging the person's sex from images or videos, was an important open question in the facial image analysis. The appearance of face difference between man and woman is always vague and subtle, which occurs a significant change according to various ages and ethnic. Accurately recognizing human gender from a single face image was a challenge task.

Many efforts had been made in the field of facial image gender recognition, which can be broadly categorized into appearance-based and geometric-based methods. Appearance-based methods used appearance information globally or locally extracted from the pixel of face images to classify the gender. The early work was presented by Gollomb et. al [1] who trained a multi-layer neural network, named SEXNET, to classify gender in 90 image samples of men and women. SVM-based gender classification methods were investigated by Moghaddan et al [2], which could achieve more than 96% accuracy on "Thumbnails" face images from FERET database. Baluja et al [3] presented a method based on AdaBoost algorithm to identify the sex of a person from a low resolution grayscale picture of their faces. Furthermore, Subspace Learning is a popular-used global feature extraction method for facial image analysis. Buchala et al [4] used Principal Component Analysis (PCA) to encode the properties of face image such as gender, ethnicity, age, and identity efficiently. Other Subspace Learning-based methods had been studied for facial gender recognition including Linear Discriminant Analysis (LDA) [5] and Independent Component Analysis (ICA) [6]. Several local appearance descriptors were used to represent the facial information for gender recognition. Sun et al [7] presented a novel approach for gender classification by boosting local binary pattern (LBP) based classifiers. Alexandre et al [8] combined LBP features with intensity and shape in a multi-scale fusion approach. In their method, classifiers were trained for each feature and at different image sizes. In literature [9], a novel face representation combining dense SIFT descriptors and shape contexts was proposed for recognizing gender from face images.

Geometric-based methods were depended on the measurements of facial landmarks. Fellous [10] selected 40 points to calculate 22 normalized vertical and horizontal fiducial distances. These points were extracted manually from images of frontal faces by a human operator. From these distances, five dimensions were derived using discriminant analysis and used to classify gender. Saatci et al [11] presented an active appearance model (AAM) based geometric-approach for recognizing gender and expression (using a SVM classifier with a radial basis kernel) from face images. Cao et al [12] used a metrology-based method for gender classification, they achieved 86.83% accuracy on the MUCT database and 90.63% accuracy on the XM2VTS database.

Recently, researches on deep learning have tremendous success in many competitions and applications of computer vision. In 2013, Ng et al [13] proposed a discriminatively-trained convolutional neural network (CNN) for gender classification of pedestrians. Levi et al [14] designed a seven layer CNN model, which included one input layer, three convolutional layers, two fully connected layers and one softmax output layer, for age and gender classification. Furthermore, many more methods [15-17] based on deep learning model were proposed for facial gender recognition.

As mentioned above, the research of facial gender recognition has made a great progress in recent years, and many high-quality methods have been presented. However, most of these methods are trained and tested in the laboratory environment where the conditions such as illumination, pose and occlusion are well controlled. The performances of these facial gender recognition method will be decreased significantly when they are implemented in an unconstrained environment. In this paper, we present a facial gender recognition method via Low-Rank decomposition and Collaborative Representation(LRCR). Firstly, we apply a low-rank decomposition to align the input images, which alleviate the negative effects of disalignment, illumination changes and noise caused by the unconstrained environment. The effectiveness of this pre-processing is verified with extensive experiments on uncontrolled face images. It is well-known that the sparse representation based classification (SRC) is less sensitive to variations of illumination, expression and pose than the traditional holistic feature learning methods such as PCA, LDA. And, the SRC is a special case of collaborative representation based classification (CRC), which has various instantiations by applying different norms to the coding residual and coding coefficient [18]. In order to further enhance the robustness of the facial gender recognition methods in an unconstrained situation, the CRC is used as the step of feature extraction and classifier. In a word, the face images captured in the unconstrained environment are pre-processed by low-rank decomposition algorithm, which makes the training images to be fit for a linear model of CRC. And, by means of CRC, the precise choice of the feature space is no longer indispensable to make the gender recognition system more reliable.

The rest of this paper is organized as follows: In Section 2, facial gender classification method based on low-rank decomposition algorithm and collaborative representation is introduced in details. Section 3 presents the experimental results and discussions of the comparison between our method and other 6 state-of-the-art methods on three benchmark dataset. At last, we make the conclusion in Section 4.

2. Facial Gender Recognition via the Low-rank Decomposition and Collaborative Representation

The SRC framework provides a new insight in facial gender recognition, but it still has one major drawback that this approach could not handle misaligned facial images, which are usual in an unconstrained environment. The performance of the facial gender recognition system will be significantly decreased if we directly use original input images without any preprocessing. We use the low-rank decomposition [19] for robustly aligning linearly correlated input images with large occlusions and corruptions. The collaborative representation rather than the sparse representation [20] is used to speed up the procedure of l_1-norm optimization through l_2-norm optimization with regularized least square and keep the ability of pattern discrimination. The pipeline diagram of our method and standard sparse representation is given in **Fig. 1**.

In the **Fig. 1**, the green line represents our method via low-rank decomposition and collaborative presentation, and red line shows standard sparse representation. In the LRCR, the input images are decomposed into two matrix: low-rank matrix and error matrix. Face images processed by low-rank decomposition are denoised and well aligned, and the dictionary generated by these low-rank matrices has the more robust discriminative power. In standard sparse representation, the original face images are employed to build sparse dictionary without the pre-processing of alignment and denoising. The corrupted face images will do harm to accurately represent the probability distribution of residual, which will seriously reduce the performance of the facial gender recognition method. In the phase of recognition, we use CRC instead of SRC to be as the classifier, using the much weaker l_2 -norm sparsity constraint to achieve similar classification results but with significantly lower complexity.

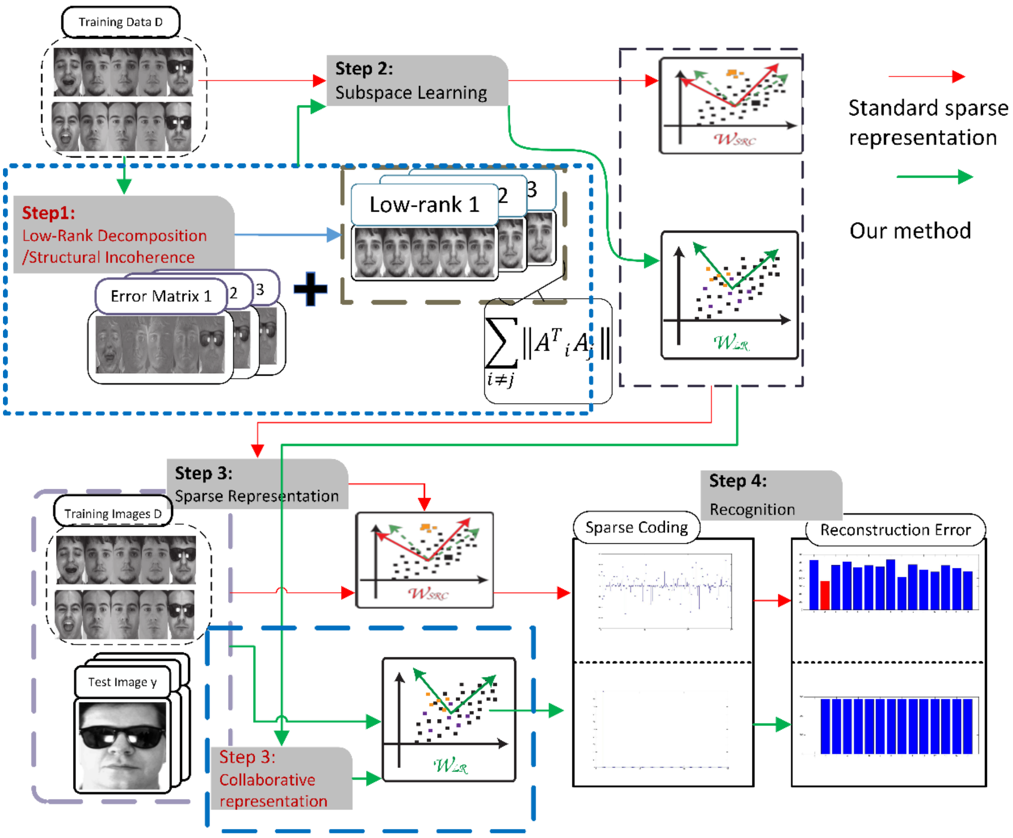


Fig. 1. The pipeline diagram of LRCR. The red line indicates the standard processing of sparse representation, the green line shows the one of LRCR.

2.1 Facial features extracting with low-rank decomposition

According to the facial gender recognition system based on SRC, the misaligned face images always lead to a poor accuracy of recognition. The proposed method aligns the face images by the low-rank decomposition, which also works well for dealing with a batch of misaligned face images.

Suppose we are given n well-aligned gray-scale images $I_1^0,...,I_n^0 \in R^{\omega\times h}$ of some objects, and the original corrupted images can be written as $I_1 = I_1^0 + e_1,..., I_n = I_n^0 + e_n$, e_i is the noise of ith image. In many situations of interest, these well-aligned images are linearly correlated. More precisely, if we let $vec : R^{\omega\times h} \to R^m$, which denotes the operator that selects an m-pixel region of interest from an image and stacks it as a vector, then as a matrix

$$D = \left[vec\left(I_1\right)|...|vec\left(I_n\right)\right] = A + E \tag{1}$$

Where $A = [vec(I_1^0)|...|vec(I_n^0)] \in R^{m\times n}$ is a low-rank matrix that models the common linear structure in the batch of images, and $E = \left[vec\left(e_1\right)|\ldots|vec\left(e_n\right)\right] \in R^{m\times n}$ is a matrix of large-but-sparse errors that models the facts of corruption, occlusion, shadows, and specularities etc. It is more practical to assume that we observe $I_1 = \left(I_1^0 + e_1\right) \circ \tau_1^{-1}, ..., I_n = \left(I_n^0 + e_n\right) \circ \tau_n^{-1}$. We can also model practical misalignment as a certain transformation $\tau_1, \ldots, \tau_n \in G$, acting on the two-dimensional domain of images $I_1^0, \ldots, I_n^0$, respectively. And, suppose G is a finite-dimensional group that has a parametric representation. Instead of observing the original images I_i^0, we observe misaligned images $\left(I_i^0 + e_i\right) \circ \tau_i^{-1}$.

In order to solve this problem, it iteratively searches for a set of transformations $\tau = \left\{\tau_1, \ldots, \tau_n\right\}$, which forces the rank of transformed images to be as small as possible. Formally, writing $D \circ \tau$ as shorthand for $\left[vec\left(I_1 \circ \tau_1\right)|\ldots|vec\left(I_n \circ \tau_n\right)\right] \in R^{m \times n}$, the object function can be represented as:

$$\min_{A,E,\tau} rank\left(A\right) + \gamma \left\|E\right\|_0 \;\; s.t. \;\; D \circ \tau = A + E \tag{2}$$

where the l_0-norm $||.||_0$ counts the number of nonzero entries in the error matrix *E*. $\gamma > 0$ is a penalty parameter that makes a trades off between the rank of the solution and the sparsity of the error.

Because both rank and l_0-norm are nonconvex and discontinuous, and the constraint $D \circ \tau = A + E$ is highly nonlinear, the optimization of Equation (2) is not directly tractable. To address this issue, the low-rank matrix can be recovered from sparse errors. As long as the rank of the matrix A to be recovered is not too high and the number of errors is not too large, minimizing natural convex surrogate for $rank\left(A\right) + \lambda \left\|E\right\|_0$ can exactly recover matrix A. This convex relaxation replaces $rank\left(.\right)$ with nuclear norm or sum of the singular values: $||A||_* = \sum_{i=1}^{m} \sigma_i\left(A\right)$, and replaces the l^0-norm $||E_0||$ with the l^1-norm $\sum_{ij}\left|E_{ij}\right|$. Applying the same relaxation to the problem (2) yields a new optimization problem:

$$min_{A,E,\tau} \left\|A\right\|_* + \lambda \left\|E\right\|_1 \;\; s.t. \;\; D \circ \tau = A + E \tag{3}$$

Theoretical considerations in literature [19] suggest that the weighting parameter λ should be form $C / \sqrt{m}$ where C is a constant, typically set to be $C \approx 1$. The new objective function is non-smooth, but now continuous and convex.

To solve the nonlinearity of the constraint $D \circ \tau = A + E$, we can approximate this constraint with $D \circ \left(\tau + \Delta\tau\right) \approx D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau_i \epsilon_i^T$, where $J_i = \frac{\partial}{\zeta} vec\left(I_i \circ \zeta\right)|_{\zeta = \tau_i} \in R^{m \times n}$ is the Jacobian of the *i*th image with respect to the transformation parameters τ_i and ϵ_i denotes the standard basis for R^n .This leads to a convex optimization problem in unknown $A, E, \Delta\tau$:

$$\min_{A,E,\tau} \left\|A\right\|_* + \lambda \left\|E\right\|_1 \;\; s.t. \;\; D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau_i \;\; \in_i^T = A + E \tag{4}$$

2.2 The collaborative representation based classification

In this paper, the collaborative representation classification with the minimum mean square error criterion is introduced to improve the sparse representation model. The SRC algorithm [20] is summarized as:

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \quad s.t. \ \|y - X\alpha\|_2 < \varepsilon \tag{5}$$

In this formula, y is test sample, X is dictionary, and α is coding vector. We can estimate the class of the object according to residual: $e_i = \|y - X_i\hat{\alpha}_i\|_2^2$.

A general model of CRC is: ,

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \|y - X\alpha\|_{l_q} + \lambda \|\alpha\|_{l_p} \right\} \tag{6}$$

where λ is the regularization parameter and $p, q \in \{1, 2\}$. Different settings of p and q lead to different instantiations of the collaborative representation model. For example, in SRC, p is set as 1 while q is set as 1 or 2 to handle face recognition with and without occlusion/corruption, respectively. By using l_1-norm or l_2-norm to characterize the coding vector α and the coding residual $e = y - X\alpha$. We can have different instantiations of CRC, while S-SRC(coding e by using l_2-norm) and R-SRC(coding e by using l_1-norm) are special cases of CRC. The l_1- or l_2-norm characterization of e is related to the robustness of CRC to outlier pixels, while the l_1- or l_2-norm characterization of α is related to the discrimination of facial feature y.

After the processing of collaborative representation with all classes, SRC classifies y individually. For the simplicity of analysis, we remove the l_1-norm sparsity term in Equation (5), and then the representation becomes a least square problem: $\hat{\alpha} = \arg\min_{\alpha} \left\{ \|y - X\alpha\|_2^2 + \gamma \|\alpha\|_2^2 \right\}$. The associated representation $\hat{y} = \sum_i X_i\hat{\alpha}_i$ is actually the perpendicular projection of y onto the space spanned by X .In SRC, the reconstruction error by each class $e_i^* = \|y - X_i\hat{\alpha}_i\|_2^2$ is used for classification. It can be readily derived that

$$e_i(y) = \|y - X_i\hat{\alpha}_i\|_2 = \|y - \hat{y}\|_2^2 + \|\hat{y} - X_i\hat{\alpha}_i\|_2^2 \tag{7}$$

Obviously, it is the amount $e_i^* = \|\hat{y} - X_i\hat{\alpha}_i\|_2^2$ that works for classification because $\|y - \hat{y}\|_2^2$ is constant for all classes. Suppose that $\chi_i = X_i\hat{\alpha}_i$ and $\bar{\chi}_i = \sum_{j \neq i} X_j\hat{\alpha}_j$,we can readily have

$$\frac{\|\hat{y}\|_2}{\sin(\chi_i, \bar{\chi}_i)} = \frac{\|\hat{y} - X_i\hat{\alpha}_i\|_2}{\sin(\hat{y}, \chi_i)} \tag{8}$$

where $\langle \bar{\chi}_i, \chi_i \rangle$ is the angle between $\bar{\chi}_i$ and χ_i. Finally, the representation error can be represented by

$$e_i^* = \frac{\sin^2(\hat{y}, \chi_i)\|\hat{y}\|_2^2}{\sin^2(\chi_i, \bar{\chi}_i)} \tag{9}$$

Equation (9) shows that by using collaborative representation. When judging whether y belongs to class i, we need to satisfy two constraints, one is that the angle between $\hat{y}$ and χ_i should be small, the other is that the angle between $\bar{\chi}_i$ and χ_i should be big. Such a "double checking" makes the CRC [18] more effective and robust.

Finally, the algorithm of the LRCR is summarized in **Table 1**.

Table 1. algorithm of facial gender recognition based on LRCR

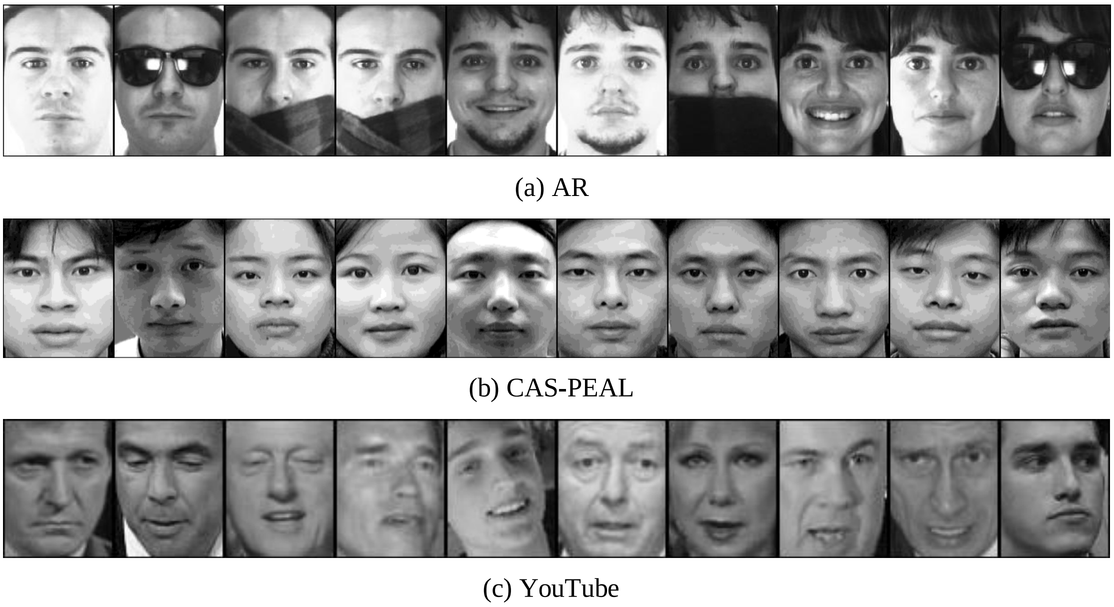
```
Algorithm1:Facial gender recognition via LRCR algorithm

Step 1:Low-rank decomposition for initial images: $D \circ \tau = A + E$
    where $D$ counts initial images, $\tau$ counts transformation parameter, $A$ counts low-rank matrix,
    $E$ counts error matrix
Step 2:normalize each column of low-rank matrix $A$: $norm(a_1, a_2, \ldots, a_n)$
Step 3:generate the dictionary $X$ according to low-rank $A$;
Step 4:compute the coefficient vector of collaborative representation:
    $\hat{\rho} = \left(X^T X + \lambda \cdot I\right)^{-1} X^T y$
Step 5:output result of classification:
    $Class = Identity(y) = \mathrm{argmin}_i \{\|y - X_i \hat{\rho}_i\|_2 / \|\hat{\rho}_i\|_2\}$
```

3. Experiments and Discussions

In this section, we evaluate the performance of the proposed facial gender recognition algorithm on several benchmark datasets including AR [21], CAS-PEAL [22], and YouTube [23]. **Fig. 2** shows some image samples of three benchmark databases used for the following experiments.

The AR and CAS-PEAL database are both obtained under the laboratory-controlled conditions. Images in AR dataset are all well aligned frontal face view with different facial expressions, illumination conditions, and occlusions (e.g., glasses and scarf). The CAS-PEAL dataset provides large-scale Chinese face images with different sources of variations, including pose, expression and lighting. Face images in AR dataset are aligned more strictly than those in the CAS-PEAL dataset. Images in YouTube Faces database are automatically obtained from videos of YouTube website without any manual filtering, so these images are captured with extreme variations in head pose, lightning conditions quality, and more under unconstrained environment. Owing to those facts that the sparse representation based recognition method is always robust to the disguise and corruption, but sensitive to misalignment, we think that the facial gender recognition experiments based on YouTube will be the most challenge task for the proposed method.



(a) AR

(b) CAS-PEAL

(c) YouTube

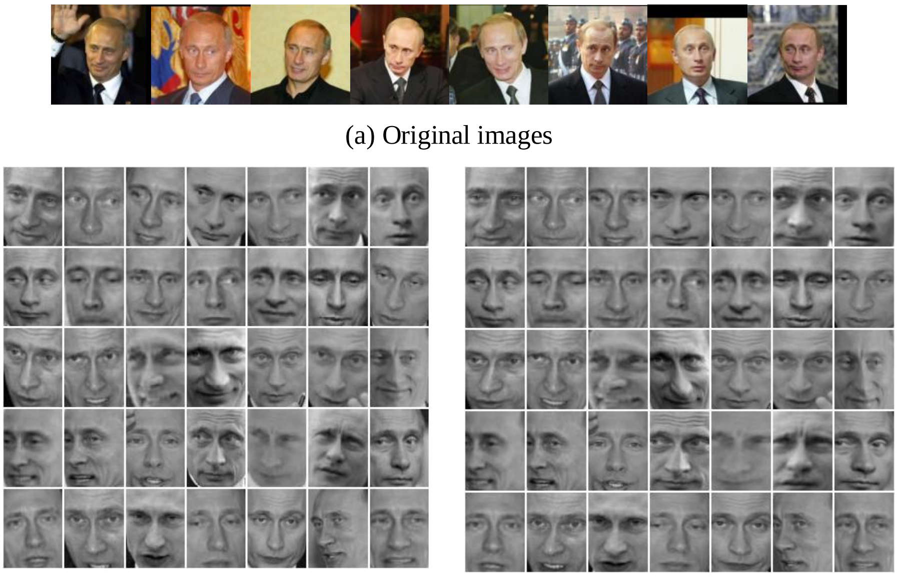
Fig. 2. Some face images of the three benchmark dataset used in our experiments (a) AR (b) CAS-PEAL (c) YouTube

In the following sections, three parts of facial gender recognition experiments are conducted to comprehensively evaluate the proposed method. The first part of experiments focuses on the low-rank decomposition. We show the result of low-rank decomposition on the face alignment task and compare the classification accuracy by collaborative representation with or without low-rank decomposition. These tests all run on the YouTube dataset. In the second part of experiments, the comparison of facial gender recognition between collaborative representation and other two typical SRC methods is performed on the three benchmark dataset. At last, we show the superior performance of the LRCR method for facial gender recognition under an unconstrained environment on comparison with 6 the state-of-art algorithms.

For each experiment, the images are separated into two subsets for training and testing, in which the training set dominates 90% and the testing set in the rest 10%. Randomly choosing the training set ensures that our results and conclusions will not depend on any special choice of the training data. Every experiment runs 5 times based on above-mentioned image samples partition and the average will be the final result. All experiments are developed in the Matlab platform and run on an image processing workstation with an Intel Xeon 2.4 GHz 8 core CPU, 128 GB memory.

3.1 The experiments based on low-rank decomposition

In this section, we firstly visually show the effect of low-rank decomposition through comparing the result face images aligned by the low-rank decomposition and the original face images. These face images are all collected from the YouTube dataset, which are captured with obvious misalignment and variations of poses, expression and illumination. And, facial gender recognition experiments with or without the preprocessing of the low-rank decomposition based on the YouTube dataset are performed to demonstrate the performance of the facial image analysis using the low-rank decomposition under the unconstrained environments.



(a) Original images

(b) Face images cropped by face detector

(c) Aligned face images

Fig. 3. Face images with or without the processing of low-rank decomposition from YouTube dataset. (a) Original images (b) face images cropped by face detector (c) alignment results by low-rank decomposition

The comparison results between face images with or without processing of the low-rank decomposition are shown in **Fig. 3**. **Fig. 3**(a) is original images collected from YouTube dataset. The face region of every original images cropped automatically using face detection tool is shown in the **Fig. 3**(b). We can find that the cropped face images have obvious difference in head poses. **Fig. 3**(c) shows the well aligned face images processed by low-rank decomposition. For example, the face images shown in the third row of **Fig. 3**(b) and (c), the position of the eyes in the unprocessed face are not leveled, but the position of eyes in the processed face images are well aligned.

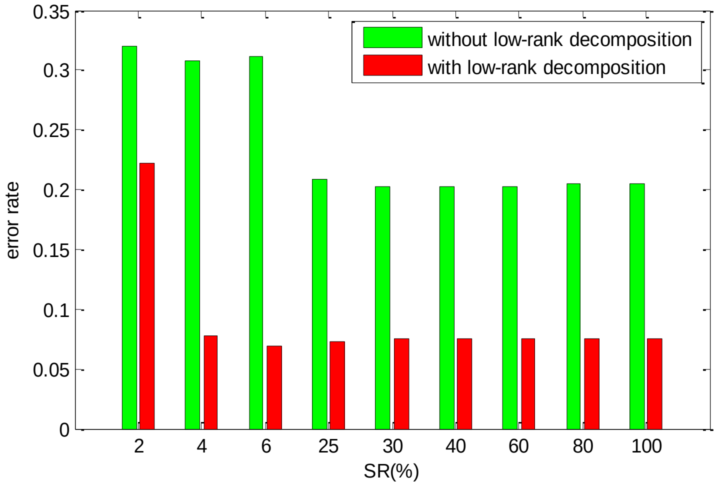


Fig. 4. The recognition results of facial gender recognition with or without low-rank decomposition.

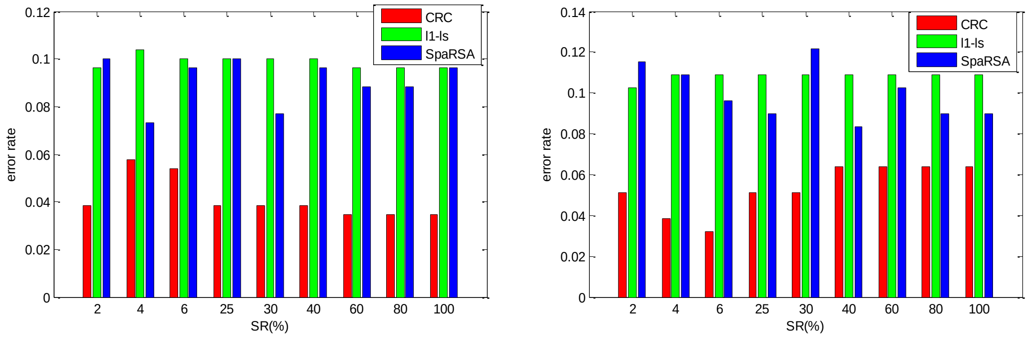
We present the results of facial gender recognition with or without low-rank decomposition in **Fig. 4**, where the horizontal coordinate SR represents the ratio of the input image to the original image size, and the vertical coordinate show the error rate of the algorithms. The red bar in the graph indicates the result of facial gender recognition without the low-rank decomposition, and the green bar is the error rate of the method with the low-rank decomposition preprocessing.

It can be observed that the using of the low-rank decomposition can effectively reduce the error rate of facial gender recognition. When SR is 4%, there is the most significant difference between the results by the method with or without low-rank decomposition. The error rate of the method without the low-rank decomposition is 30.8%, and that of the method with the low-rank decomposition is 7.78%, which is about a quarter of the former. When the SR is 25%, the method with low-rank decomposition achieves the best recognition error rate, which is 6.94%. Considering the results shown in the **Fig. 3** and **Fig. 4**, we can draw the conclusion that projecting the original image matrix to the direction of low-rank according to Equation(2) can effectively improve the correlation of vectors in the image matrix and achieve the purpose of image alignment and denoising.

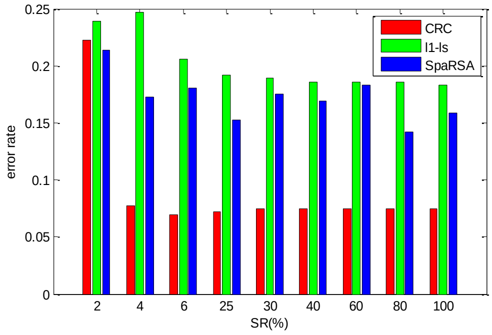
3.2 The facial gender recognition experiments based on collaborative representation

We evaluate the performance of the proposed facial gender recognition method compared with the SRC-based facial gender recognition methods. $l_1 l_s$ [24] and SpaRSA [25] are chosen as the rivals, which can achieve high accuracy and robust in the field of face recognition. Gradient project is used to be the l_1 -min in algorithm $l_1 l_s$,while iterative shrinkage-thresholding method is used in algorithm SpaRSA. Specifically, truncated Newton interior-point method which is an efficient interior-point method for solving large-scale l_1 -regularized logistic regression problems is used in algorithm $l_1 l_s$. Iterative shrinkage-thresholding method, which can be viewed as an extension of the classical gradient algorithm, is attractive due to its simplicity and thus is adequate for solving large-scale problems even with dense matrix data.

All experiments are conducted on three benchmark datasets including AR, CAS-PEAL and YouTube. As mentioned, the face images in AR and CAS-PEAL datasets have been well aligned. Experiments on AR and CAS-PEAL are carried out without the low-rank decomposition preprocessing. Due to the serious misalignment and heavy head pose changing, the face images in the YouTube dataset are all aligned by the low-rank decomposition before learning a dictionary using collaborative representation or sparse representation.



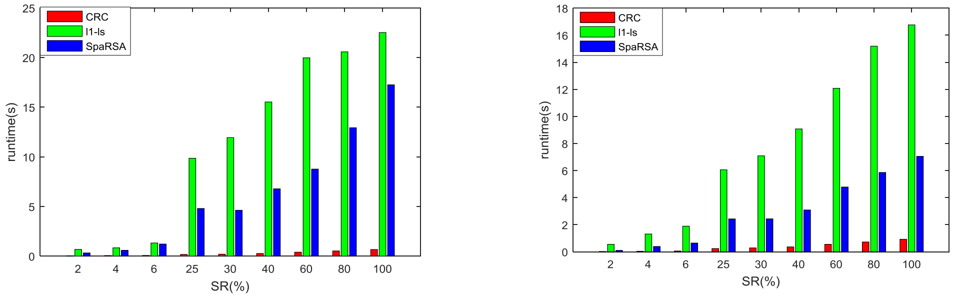
(a) Recognition results on AR dataset (b) Recognition results on CAS-PEAL dataset



(c) Recognition results on YouTube dataset

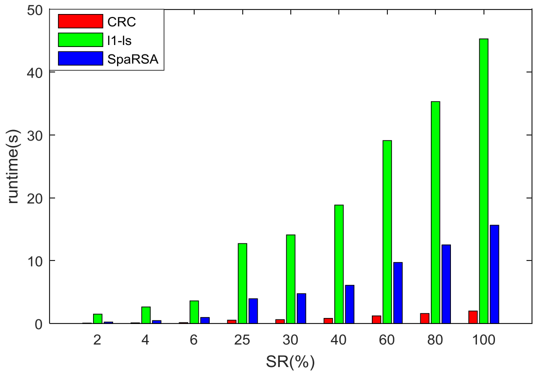
Fig. 5. Recognition results of three methods on three benchmark dataset

The results of facial gender recognition obtained by collaborative representation-based method and two sparse representation-based methods on three benchmark datasets are illustrated in **Fig. 5**. Firstly, the error rate of three methods achieved on the AR dataset is the lowest, which is all below 10%. The collaborative representation-based method hold overwhelming superiority, and the error rate is always 1/3 of that of sparse representation-method. The results demonstrate that collaborative representation-based method can achieve a better performance in the task of facial gender recognition, and show a better robustness to the face images with different sources of variations. From **Fig. 5** (a) and (b), it can be observed that the recognition results on CAS-PEAL dataset is poorer than that on AR dataset. The results clearly confirm our view at the beginning of the experiments: the collaborative representation-based method is sensitive to the misalignment but robust to the occlusion and illumination, facial expression changes. The presence of heavy occlusions like sun glass and scarf and various changes of illumination and facial expression in AR dataset does not affect the recognition result, but slightly face misalignment in CAS-PEAL dataset lead to an obvious accuracy decrease. This phenomenon is even more significant in the experiments on the YouTube dataset. So, reviewing the objective function in Equation (9) and the results shown in **Fig. 5,** it can be seen that the double checking of collaborative representation makes the classification more effective and robust.



(a) Runtime on AR

(b) Runtime on CAS-PEAL



(c) Runtime on YouTube

Fig. 6. Runtime of three methods on three benchmark dataset

In the AR, CAS-PEAL and YouTube dataset, the size of face images is 165 x 120, 200 x 150 and 200 x 150, and the size of all images is normalized to 72*60 in these experiments. And the number of the column in the dictionary is 2340, 702 and 3240, respectively. **Fig. 6** presents runtime of three method on different benchmark dataset. From **Fig. 6**, we can find that the runtime of SRC-based methods are proportional to the amount of the column of the dictionary. In the **Fig. 6** (a) and (c), the runtime of SRC is also dependent on the size of the input image size, and the bigger input image, the more time consuming. The time-consuming of the CRC method is lowest in the CAS-PEAL dataset and less than method in the AR dataset. In the YouTube dataset, the CRC method is the most efficient method when SR is more than 60%. In a word, CRC has a notable advantage on small sample image dataset.

3.3 The experiments based on low-rank and collaborative representation

We compared the proposed LRCR method against other 7 state-of-the-art methods which including LBP+SVM, 2D-Gabor+SVM, LBP+Adaboost, 2D-Gabor+Adaboost, CRC without low-rank decomposition, DBN and CNN. From the view of feature extraction, LBP and Gabor are typical local feature representation methods, which have been successfully exploited in many computer vision applications. As far as classifier, SVM, Adaboost and sparse representation are three kinds of the best performance in the shallow classifier. DBN and CNN are the dominated learning model in recently booming deep neural networks. These two kinds of deep neural networks combine the procedure of feature extraction and classification together, which attempt to model high-level abstractions in the data by using multiple processing layers. DBN and CNN have been demonstrated to produce state-of-the-art results on natural language processing and large-scale visual recognition tasks. The specific parameters of 7 methods are listed in **Table 2**.

Table 2. the specific parameters of 7 methods

No.	Method	Feature	Classifier
1)	LBP+SVM	pattern for 3 different sizes divided face images.	SVM
2)	2D-Gabor+SVM	18 2D-Gabor features extracted by 3 different block sizes and 6 directions' kernel functions	SVM

3)	LBP+Adaboost	the same as 1)	Adaboost
4)	2D-Gabor+Adaboost	the same as 2)	Adaboost
5)	CRC without low-rank	as described in Section 3.2	
6)	DBN	use the same model as the literature [26]	
7)	CNN*	7 layers model for test 100% size of input face images.	

* the model design of CNN is closely relative to the size of the input image. And, CNN model always suffers from the problem of gradient diffusion by dealing with the small sample binary classification task just like facial gender recognition. Because of the above two reasons, we only report the result of 100% size of input face images

As the above-mentioned experimental protocol, 8 methods are run on the YouTube dataset for 5 times, and the recognition results are shown in the **Fig. 7**. The meaning of the legend in the lower right corner is: method name (average accuracy / highest accuracy). Primarily, it can be clearly seen that the proposed LRCR algorithm is remarkably superior to the other 6 methods in almost all possible input image size. Recognition accuracy of the LRCR method has quickly reached 92.2% when input data size is only 4% of original image. We obtain the best recognition accuracy 93.1% when SR is 6%. According to the results achieved by CRC without low-rank decomposition, the top accuracy is 86.94% and the average accuracy is 75.93%. It shows that the misalignment of the face image has a serious impact on the performance of CRC. Then the recognition accuracy of LRCR method has a slight decline along with the increase of input image size. The foremost reason is that the impact of misalignment and disturbance of the face images in YouTube dataset is significantly enhanced when the size of input images becomes bigger. In general, the results show that the proposed LRCR based method can effectively identify gender from the nature face images with large variations of pose, expression, illumination and other changes in the YouTube database.

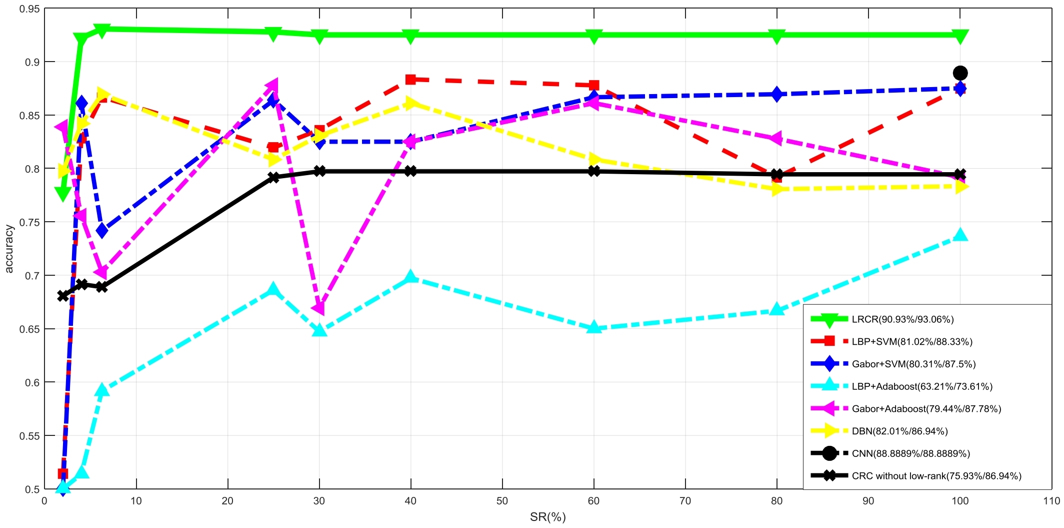


Fig. 7. Comparison of recognition accuracy of the 7 methods on YouTube dataset

Among the rest 6 methods, the CNN has achieved the best recognition accuracy rate of 88.8% when SR is 100%. This shows that the scheme of weight sharing and the multi-layer pooling in the CNN makes this kind of deep network to better adapt to mild deformation and misalignment of face images. On the other hand, the model design of CNN is closely relative to the size of the input image. And, CNN model always suffers from the problem of gradient diffusion when dealing with the small sample binary classification task just like facial gender recognition. So the overall recognition performance of the CNN model is less effective than that of the LRCR, which shows that the model of deep multi-layer convolutional neural network is more suitable to solve the problem of many categories classification of big data like that does in the ILSVRC competition rather than the problem of binary-classification with relatively small image samples. The recognition accuracy of another deep neural networks model--DBN is obviously lower than that of CNN and the LRCR. This suggests that the fully connected neural network lack of local awareness and pooling mechanisms is quite sensitive to the misalignment caused by the changing head pose. In a word, although deep learning-based methods have achieved state-of-the-art results in almost all the task of the computer vision, the training of deep neural networks relies mainly on large-scale and high-quality training images. Due to the large number of parameters that need to be trained, large-scale deep neural network is difficult to obtain a good result in a small sample, low-resolution binary-classification problem. The result shown in Fig.7 has confirmed this view.

In the four shallow learning method, the recognition accuracy of SVM based methods (LBP+SVM 81.0%/88.3%, 2D-Gabor+SVM (80.3%/87.5%)) is generally better than the one of AdaBoost based methods (LBP+Adaboost 63.2%/73.6%, 2D-Gabor+Adaboost (79.4%/87.8%), and the performance of SVM based methods is more stable than one of the Adaboost based method with the changing of the parameter SR. The main reason for this result is the difference of the essentials of the Adaboost and SVM. Training of Adaboost algorithm is a procedure to form a strong classifier for gender recognition by selecting several weak classifiers which have the most discriminative power in the weak classifiers candidates. The location of craniofacial organs is always not fixed in the misalignment face images, which has seriously damaged to the semantic information contained by the weak classifiers in Adaboost classifier, and leads to the decline of recognition accuracy. In the SVM algorithm, all the data of input features are used to solve the hyperplane, which is robust to the distortion and the misalignment in the face images. Moreover, LBP is a local feature representation approach, and 2D-Gabor belongs to the global features extraction method. Local features are more sensitive to the distortion and the misalignment of object parts in the images than global features. The recognition accuracy curves of the 2D-Gabor based methods and LBP based methods shown in the **Fig. 7** also verify this conclusion. LBP+Adaboost is the worst affected method by the distortion of face pose variation and misalignment, which can only achieve the recognition performance of the average precision 63.21% with the highest accuracy of 73.61%.

Table 3. The dimension of 5 features including LBP, Gabor, LRCR, DBN and CNN

SR(%) / Feature	**2**	**4**	**6**	**25**	**30**	**40**	**60**	**80**	**100**
LBP	826	826	826	826	826	826	826	826	826
Gabor	1140	3240	5130	20520	23616	31968	48024	63918	82080
LRCR	92	182	285	1140	1368	1824	2736	3648	4560
DBN	92	182	285	1140	1368	1824	2736	3648	4560
CNN	/	/	/	/	/	/	/	/	4560

Moreover, we list the feature dimension of 5 features including LBP, Gabor, LRCR, DBN and CNN and the runtime of 7 facial gender recognition methods in **Table 3** and **Table 4**, respectively. In **Table 3**, the feature dimensions of LBP and Gabor are the size of the features extracted by the two operators. The feature dimensions of LRCR, DBN and CNN are the sizes of images under the corresponding value of SR because these three methods deal with the input images without any manually feature extraction. From Table 4, the LBP+Adaboost is the most convenient method since using simple LBP feature and fast Adaboost algorithm. LRCR is the most time-consuming method because of its complex computational processing. DBN and CNN are faster than SVM when the input feature dimension is high but slower than SVM when the input feature dimension is small.

Table 4. The runtime of 7 facial gender recognition methods. The unit of time in this table is milliseconds

SR(%) Feature	2	4	6	25	30	40	60	80	100
LBP+Adaboost	0.69	0.72	0.84	0.87	0.89	0.92	0.96	1.13	1.35
Gabor+Adaboost	4.98	10.14	14.59	55.21	63.24	84.85	125.89	167.81	214.41
LBP+SVM	1.73	1.94	3.27	3.81	5.65	7.37	8.57	13.58	17.17
Gabor+SVM	13.23	16.56	26.68	71.64	81.48	107.71	159.46	207.80	264.18
LRCR	74.6	22.38	111.92	650.55	780.04	1019.54	1497.90	1937.65	2404.75
DBN	74.86	74.92	75.17	77.12	77.53	77.85	78.45	79.13	81.47
CNN	/	/	/	/	/	/	/	/	139.78

All in all, compared with the other the-state-of-the-art methods, the LRCR method can effectively identify the gender from the nature face images captured in the unconstrained environment. The proposed method maintains the stable capability of feature extraction and discrimination in the facial image analysis task when the input face images are suffering from the large variations of pose, expression, illumination and misalignment.

4. Conclusion

In this paper, the LRCR method is proposed to recognize the gender from nature face images in the unconstraint environment. First of all, the low-rank decomposition is performed to align and de-noise the nature face images, and then the method uses collaborative representation for gender recognition instead of the sparse representation. Based on the benchmark datasets including AR, CAS-PERL and YouTube, the experimental results show that the proposed method hold overwhelming superiority to the other 7 methods in the aspect of recognition accuracy. In addition, the solution of l_1 -norm and l_2 -norm is quite time-consuming when the number of column atoms of the dictionary becomes larger. The runtime of the LRCR method will rise exponentially with the increase of the number of dictionary atoms. So we will focus on the study of the improvement of the regularization to speed up the procedure of recognition in the future.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (61471206 and 61401220), Natural Science Foundation of Jiangsu province (BK20141428 and BK20140884) and Science Foundation of Ministry of Education-China Mobile Communications Corporation (MCM20150504).

Reference

[1] Golomb, Beatrice A., David T. Lawrence, and Terrence J. Sejnowski, "Sexnet: a neural network identifies sex from human faces," *NIPS*, Vol.1, 1990. Article (CrossRef Link)

[2] Moghaddam, Baback, and Ming-Hsuan Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*24.5, 707-711, 2002. Article (CrossRef Link)

[3] Baluja, Shumeet, and Henry A. Rowley, "Boosting sex identification performance," *International Journal of computer vision* 71.1, 111-119, 2007. Article (CrossRef Link)

[4] Buchala, Samarasena, et al., "Principal component analysis of gender, ethnicity, age, and identity of face images," in *Proc. of IEEE ICMI* 7, 2005. Article (CrossRef Link)

[5] Bekios-Calfa, Juan, Jose M. Buenaposada, and Luis Baumela, "Revisiting linear discriminant techniques in gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.4, 858-864, 2011. Article (CrossRef Link)

[6] Jain, Amit, Jeffrey Huang, and Shiaofen Fang, "Gender identification using frontal facial images," in *Proc. of Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005. Article (CrossRef Link)

[7] Sun N, Zheng W, Sun C, et al., "Gender classification based on boosting local binary pattern," in *Proc. of International Symposium on Neural Networks*. Springer Berlin Heidelberg, 2006. Article (CrossRef Link)

[8] Alexandre, Luís A., "Gender recognition: A multiscale decision fusion approach," *Pattern Recognition Letters* 31.11, 1422-1427, 2010. Article (CrossRef Link)

[9] Wang, Jian-Gang, et al., "Boosting dense SIFT descriptors and shape contexts of face images for gender recognition," in *Proc. of Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010. Article (CrossRef Link)

[10] Fellous, Jean-Marc, "Gender discrimination and prediction on the basis of facial metric information," *Vision research* 37.14, 1961-1973, 1997. Article (CrossRef Link)

[11] Saatci, Yunus, and Christopher Town, "Cascaded classification of gender and facial expression using active appearance models," *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. IEEE, 2006. Article (CrossRef Link)

[12] Cao, Deng, et al., "Can facial metrology predict gender?," in *Proc. of Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011. Article (CrossRef Link)

[13] Ng, Choon-Boon, Yong-Haur Tay, and Bok-Min Goi, "A convolutional neural network for pedestrian gender recognition," in *Proc. of International Symposium on Neural Networks*. Springer Berlin Heidelberg, 2013. Article (CrossRef Link)

[14] Levi, Gil, and Tal Hassner, "Age and gender classification using convolutional neural networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015. Article (CrossRef Link)

[15] Kalansuriya, Thakshila R., and Anuja T. Dharmaratne, "Neural network based age and gender classification for facial images," *ICTer* 7.2, 2014. Article (CrossRef Link)

[16] Liew, Shan Sung, et al., "Gender classification: a convolutional neural network approach," *Turkish Journal of Electrical Engineering & Computer Sciences* 24.3, 1248-1264, 2016. Article (CrossRef Link)

[17] Zhang, Hao, Qing Zhu, and Xiaoqi Jia, "An Effective Method for Gender Classification with Convolutional Neural Networks," in *Proc. of International Conference on Algorithms and Architectures for Parallel Processing*. Springer International Publishing, 2015. Article (CrossRef Link)

[18] Zhang, Lei, et al., "Collaborative representation based classification for face recognition," *arXiv preprint arXiv:1204.2358*, 2012. Article (CrossRef Link)

[19] Peng, Yigang, et al., "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11 2233-2246, 2012. Article (CrossRef Link)

[20] Zhang, Lei, Meng Yang, and Xiangchu Feng., "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. of Computer vision (ICCV), 2011 IEEE international conference on*. IEEE, 2011. Article (CrossRef Link)

[21] Martinez, Aleix M., "The AR face database," *CVC technical report* 24, 1998. Article (CrossRef Link)

[22] Gao, Wen, et al., "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38.1, 149-161, 2008. Article (CrossRef Link)

[23] Wolf, Lior, Tal Hassner, and Itay Maoz., "Face recognition in unconstrained videos with matched background similarity," in *Proc. of Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011. Article (CrossRef Link)

[24] Kim, Seung-Jean, et al., "An Interior-Point Method for Large-Scale l1-Regularized Least Squares," *IEEE journal of selected topics in signal processing* 1.4, 606-617, 2007. Article (CrossRef Link)

[25] Wright, Stephen J., Robert D. Nowak, and Mário AT Figueiredo., "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing* 57.7, 2479-2493, 2009. Article (CrossRef Link)

[26] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov., "Reducing the dimensionality of data with neural networks," *science* 313.5786, 504-507, 2006. Article (CrossRef Link)

Ning Sun is an Associate Professor in Nanjing University of Posts and Telecommunications. He got the B.S., M.S. and Ph.D. degrees from Guilin University of Electronic Technology, Nanjing Institute of Electronic Technology and Southeast University, in 2000, 2004 and 2007. Prior to joining NUPT in 2012, he was a senior engineer at the 28th research institution of China Electronics Technology Group Corporation (CETC). His current research interests include deep learning, pattern recognition and embedded platform based video analysis.

Hang Guo received the B.S. degree from Jinling Institute of Technology in 2014. Currently, he is pursuing the M.S. degree in the Department of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include image processing and pattern recognition.

Ji-xin Liu received the Ph.D. degree in Pattern Recognition and Intelligence System from Nanjing University of Science and Technology (NUST), China, in 2013. He is an associate professor in the Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, at Nanjing University of Posts and Telecommunications (NUPT), China. His current interests include pattern recognition, compressed sensing, remote sensing information system, image processing.

Guang Han received the B.S. degree from Shandong University of Technology in 2004, and M.S. and Ph.D. degrees from Nanjing University of Science and Technology, in 2006 and 2010, respectively. Since 2010, he has been with Nanjing University of Posts and Telecommunications, Nanjing, China, where he is currently an Associate Professor in the Engineering Research Center of Wider and Wireless Communication Technology, Ministry of Education. His current research interests include pattern recognition, video analysis, computer vision and machine learning.