Face Recognition Based on Discriminative Dictionary with Multilevel Feature Fusion

Hongjun Li[1,2], Nicola Nobile[2], and Ching Y. Suen[2]

[1] School of Electronic Information Engineering, Nantong University, Nantong, 226019, China
[2] Centre for Pattern Recognition and Machine Intelligence, Concordia University, Montreal, Quebec H3G 1M8, Canada
hongjun@encs.concordia.ca

Abstract. In order to alleviate the influence of illumination, pose, expression and occlusion variations in face recognition, in this paper, an effective face recognition method based on discriminative sparse representation is proposed. To solve the problem of these variations, we extract discriminative features which represent for each of the training images, and propose a novel dictionary by learning discriminative features. Firstly, we decompose a test image by using nonsubsampled contourlet transform (NSCT), and then fuse the information according to the features from each subband and their contributions. Finally, we obtain the discriminative features of training images and construct a discriminative dictionary. Fuse these multiple features can improve the efficiency and effectiveness of face recognition, especially when training samples are limited and the dimension of feature vector is low. Experimental results on two widely used face databases are presented to demonstrate the efficiency of the proposed approach.

Keywords: Face recognition, Sparse representation, Discriminative, NSCT.

1 Introduction

Face recognition is always an attractive topic in computer vision and pattern recognition [1, 2]. For face recognition, image features are firstly extracted and then matched to those features in a gallery set. The task is to find the class to which a test sample belongs by given training samples from multiple classes. Recently, there has been an increasing interest in classification problem where the data across multiple classes comes from a collection of low dimensional linear subspace. An important method that deals with data on multiple subspaces relies on the notion of sparsity. During the past few years, sparse representation theories have been applied to face recognition and were paid much attention. It has been one of the most successful applications of image analysis and understanding. Although many technologies have been proposed to perform tasks of classifying facial images well, face recognition problem is still challenging. Wright et al. [3] used sparse representation for face recognition and the performance is impressive. This method constructs dictionary by training all kinds of samples, however, when the training samples are limited and the dimension of feature

N. El Gayar et al. (Eds.): ANNPR 2014, LNAI 8774, pp. 252–263, 2014.

vector is low, it becomes less efficient. Yang et al. [4] further extended SRC based framework to a sparsity constrained robust regression problem for handling outliers such as occlusions in facial images. In addition, low-rank representation method [5, 6] which has been established recently is based on the hypothesis that the data is approximately spanned by some low-rank subspaces. Liu and Yan [5] further proposed a latent low-rank representation approach. In latent LRR, the hidden data can be regarded as the input data matrix after being transposed. This idea has been used to design a classifier for image classification [6]. Recently, some works considering multi-resolution information have been proposed for face recognition [7, 8]. Most of these methods realize multi-resolution face recognition by extracting Gabor features in different scales and orientations, which are fused to form multi-resolution features.

Inspired by previous works, we aim to find a discriminative dictionary for face recognition. Most existing discriminative dictionary learning algorithms try to learn a common dictionary shared by all classes, as well as a classifier of coefficients for classification. However, the shared dictionary loses the correspondence between the dictionary atoms and the class labels, and hence performing classification based on the reconstruction error associated with each class is not allowed. Different from these works, Yang et al. [4] proposed a discriminative dictionary learning framework which employs fisher discrimination criterion to learn a structured dictionary. This method uses the reconstruction error associated with each class as the discriminative information for classification, but it does not enforce discriminative information analysis in dictionary construction.

Our method considers NSCT features of facial images in different scales and orientations. NSCT is shift invariant and can reduce the effect of posture variations in the process of face recognition. In addition, in order to obtain robust features, we use contribution criterion calculated by structure similarity when fusing features from different scales and orientations. The structural similarity criterion is imposed on the latent feature images to make them discriminative. We try to propose a new discriminative dictionary by learning fused discriminative features to improve the classification efficiency.

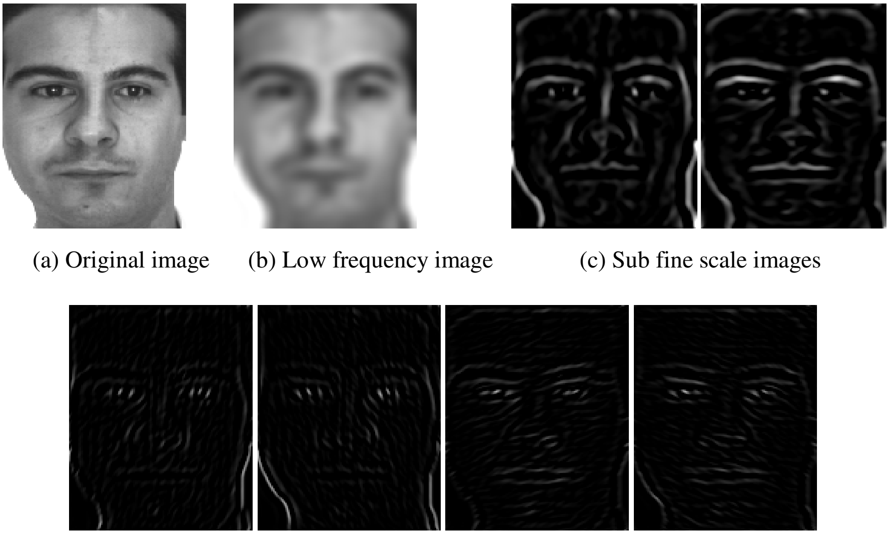
The remainder of the paper is organized as follows. Preliminary works are presented in Section 2. Section 3 presents the proposed method. Section 4 is devoted to experimental results and analysis, and Section 5 concludes the paper.

2 Multi-scale Geometry Analysis

Contourlet transform offers a high directionality but due to the up-sampling and down-sampling, it is shift-variant. However, image analysis applications such as edge detection, image enhancement etc. requires a transform which is shift-invariant. This requirement is fulfilled by nonsubsampled contourlet transform (NSCT) [9, 10]. The NSCT consists of nonsubsampled pyramid (NSP) that ensures multiscale decomposition and nonsubsampled directional filter bank (NSDFB) that offers directionality. NSP is constructed by using a low pass filter and a high pass filter, and it can decompose an image into a low frequency subband and a high frequency subband. NSDFB is then applied to high frequency subband whereas the low frequency subband is used for next finer pyramidal decomposition.

NSCT is shift-invariant so that each pixel of the transformed subbands corresponds to that of the original image in the same spatial location. Therefore we gather the geometrical information pixel by pixel from NSCT coefficients. There are three classes of pixels: strong edges, weak edges and noise. Strong edges correspond to those pixels with large magnitude coefficients in all subbands. Weak edges correspond to those pixels with large magnitude coefficients in some directional subbands but small magnitude coefficients in other directional subbands within the same scale. Noise corresponds to those pixels with small magnitude coefficients in all subbands.

We decompose a facial image under 3 scales by NSCT. In **Fig. 1**, there is an obvious difference between **Fig. 1** (c) and **Fig. 1** (d). In subimages from different scales, the amounts of coefficients are different. According to the characteristic of NSCT, we know that strong coefficient keeps a big value and do not varies when the scale changes, the value of weak coefficient varies as the scale changes and holds different value in different subbands. However, noise signal keeps a small value in all subbands. As a result, the recognition by strong coefficients is more robust than that by other coefficients. Combine coefficients from different scales and directions, acquire discriminative features may improve the efficiency of face recognition.



(a) Original image (b) Low frequency image (c) Sub fine scale images

(d) Fine scale images

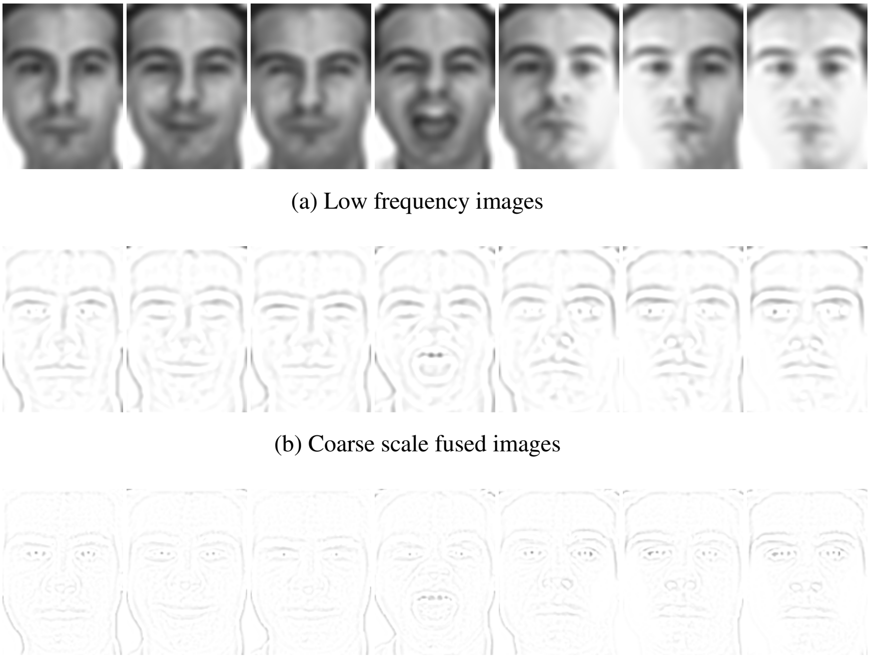
Fig. 1. Image decomposed by NSCT

3 Face Recognition Algorithm

3.1 Discriminative Feature Fusion

Unsupervised feature extraction is a fundamental and critical step in face recognition. Raw facial images usually contain a high dimensional data structure. It is desirable to utilize feature extraction for the sake of avoiding the curse of dimensionality.

We propose a novel approach based on discriminative sparse representation. Firstly, we decompose an image sparsely by NSCT and acquire subimages in different scales and directions. As shown in **Fig. 2**, we can see that subimages in different scales contain different information, some information from key points of the face mainly concentrates on fine scale. As a result, extracting key information from different subimages to construct discriminative features is very important.



(a) Low frequency images

(b) Coarse scale fused images

(c) Fine scale fused images

Fig. 2. Facial feature images of different training samples

The features in different scales can be easily extracted by multi-scale geometry analysis. If we use a simple and effective feature fusion method to combine these features, the efficiency and the performance of face recognition should be improved. Images which are used to reconstruct should be similar to the original image, so we use Structural SIMilarity (SSIM) as a criterion [11] to calculate the contribution of each subimage over each scale and fuse discriminative features according to their contributions. SSIM is easy to calculate and applicable to various image processing applications. SSIM is defined as:

$$\mathrm{SSIM}(u,v)=\frac{\sigma_{uv}}{\sigma_u\sigma_v}\cdot\frac{2\overline{uv}}{(\overline{u})^2+(\overline{v})^2}\cdot\frac{2\sigma_u\sigma_v}{\sigma_u{}^2+\sigma_v{}^2} \qquad (1)$$

The first component is the correlation coefficient between u and v. Even if u and v are linearly related, there still might be relative distortions between them, which are evaluated in the second and third components. The second component, with a value range in [0, 1], measures how close the mean luminance is between u and v. It equals 1 if and only if $\overline{u}=\overline{v}$. σ_u and σ_v can be viewed as the estimation of contrast of u and v, so the third component measures how similar the contrasts of the image are. The values range from 0 to 1, where the best value 1 is achieved if and only if $\sigma_u=\sigma_v$.

We calculate the contribution of subimages in each scale and direction by using structural similarity theory. The contribution of each subimage can be defined as:

$$c_d=\frac{1}{\sum_{d=1}^{2^J}\mathrm{SSIM}(y,z_d)}\mathrm{SSIM}(y,z_d) \qquad (2)$$

where J is the decomposition scale, z_d is the subimage in each scale and direc tion, c_d is the contribution value of a subimage. So fused discriminative feature (DF) can be defined as:

$$DF=\sum_{d=1}^{2^J}c_d*z_d \qquad (3)$$

We use a multilevel decomposition method to extract features from different scales and directions, and fuse these features of each level according to their contributions to reconstruct an image with discriminative features which is more similar to the original image. Fuse these multiple features can improve the efficiency of face recognition to some extent. **Fig. 3** shows some images with discriminative features from the same person under different lights and expressions.

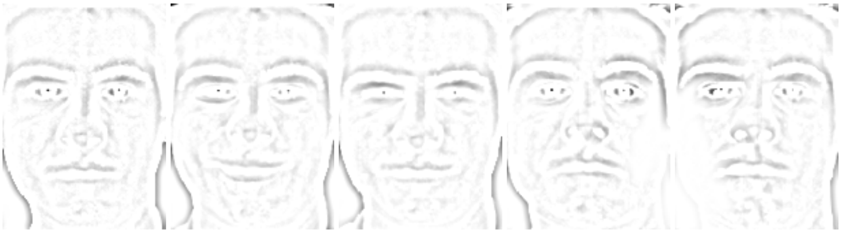


Fig. 3. Images with discriminative features

3.2 Multilevel Discriminative Dictionary Construction

In this section, we construct a multilevel discriminative dictionary by learning discriminative features. Define matrix Y as the entire training set which consists of n training samples from all c different classes: $Y = [Y_1, Y_2, Y_3, \ldots, Y_c]$ where $Y_i \in R^{d*n_i}$ is all the training samples from i-th class, d is the dimension of samples, and n_i is the sample from i-th class.

We construct dictionary by training set, using NSCT analysis method. The training samples are decomposed by NSCT, and then using formula (2) and (3) to obtain fused discriminative features. We construct sub-dictionaries class by class, and finally form a dictionary $D = [D_1, D_2, \ldots, D_k] \in R^{m \times N}$,where D_i is the vector of discriminative features of face images in each class.

```
Algorithm 1: Dictionary Construction
Input: Training Sample Y
Output: Dictionary D
1. Decompose training sample Y by using NSCT.
2. Obtain fused discriminative features by formula (2-3).
3. Construct sub-dictionary of each class and gain
dictionary D.
```

3.3 Face Recognition Algorithm

In practical, occlusion exists in both training and testing samples, the dictionary constructed via nonsubsampled contourlet transform can reduce this influence. **Fig. 4** shows the procedure of our face recognition algorithm. Constructing dictionary $D = [D_1, D_2, \ldots, D_N]$ via Algorithm 1 on face database, where D_i is the constructed sub-dictionary of i-th class. Algorithm 2 below summarizes the complete recognition procedure. Assume that y is a test sample, and we can get the sparse coefficients vector by solving:

$$x = \arg\min_x \{\|y - Dx\|_2^2 + \lambda \|x\|_1\} \tag{4}$$

Denoted by $x = [x_1; x_2; \ldots; x_c]$,where x_i is the vector of coefficients in sub-dictionary D_i. We can calculate the residual associated with i-th class by:

$$e_i = \| y - D_i x_i \|_2^2 \tag{5}$$

The identity of testing sample y is determined according to:

$$identify(y) = \arg\min_i \{e_i\} \tag{6}$$

```
Algorithm 2: Face Recognition
Input:  Dictionary D, a test sample $y$
Output: $identify(y) = \arg\min_i \{e_i\}$
1. Decompose training sample $y$ by using NSCT.
2. Obtain contributions of subimages in each scale by formula (2).
3. Fuse discriminative features by formula (3).
4. Minimize $x_i$ to problem (4).
5. Compute the residual $e_i = \| y - D_i x_i \|_2^2$ .
```

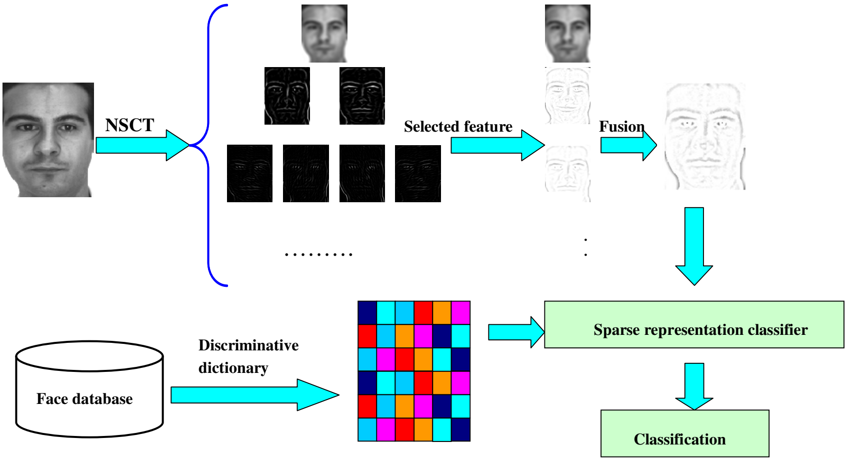


Fig. 4. Face recognition algorithm

4 Experiments and Analysis

In this section, we conduct experiments on AR Database [14] and Extended Yale B Database [15] which are widely used to test the efficacy of our method. The experiments are implemented by MATLAB R2013b on a computer with Intel(R) Xeon(R) CPU X3450@2.67GHz 2.66 GHz, and windows 7 operating system. We repeat each experiment 10 times and the accuracy is averaged. We examine the performance of our method when dealing with different illuminations, expressions, occlusions and different amounts of training samples. Compared with several state-of-the-art methods, such as SRC [3], LRC [12], and FDDL [13], our proposed approach achieves a better performance in terms of high recognition accuracy and robustness.

4.1 AR Face Database

The AR database consists of over 4000 frontal images from 126 individuals. For each individual, 26 pictures were taken in two separate sessions. A subset that contains males and females was chosen from AR database. For each class, seven images are randomly selected as training images, and seven images as testing images. **Fig. 5** shows some facial images from the AR database. We select four feature space dimensions: 30, 54, 120, and 300, according to the down sample ratios 1/24, 1/18, 1/12, and 1/8, respectively. **Table 1** shows the recognition rates of four algorithms when feature dimension changes. Our method gains the highest recognition accuracy of four algorithms in same feature dimension and it outperforms SRC about 3% on average. The experimental results illustrate the accuracy and the robustness of our proposed method when illumination or expression changes. The reason for this improvement is due to the fusion of discriminative features extracted by NSCT which can enhance the representation power.

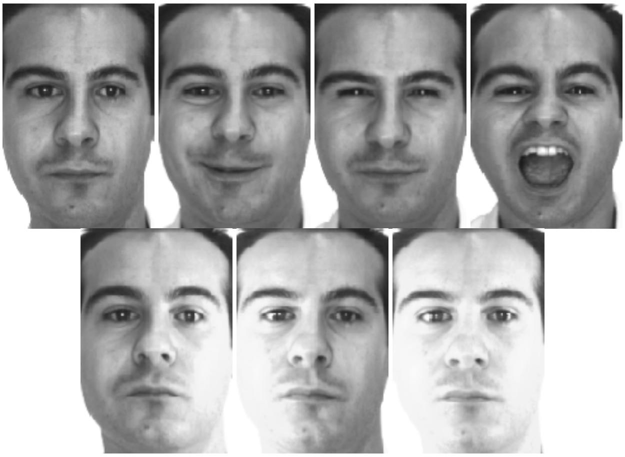


Fig. 5. Samples of different illuminations and expressions from the AR database

Table 1. Recognition rates of different feature dimensions

Algorithm	Feature dimension			
	30	54	120	300
LRC[12]	60.6%	65.8%	71.2%	75.2%
SRC[3]	66.8%	83.0%	88.6%	93.2%
FDDL[13]	64.6%	79.2%	86.5%	93.0%
Our method	69.5%	85.4%	91.5%	94.7%

4.2 Extended Yale B Face Database

The Extended Yale B database consists of 2414 frontal-face images of 38 individuals. The images were captured under various laboratory-controlled illumination conditions. For each class, there are about 64 images, half of the images are randomly selected as training images, and the rest as testing images. **Fig. 6** shows some facial images from the Extended Yale B database. We compute the recognition rates in the feature space dimensions 30, 56, 120, and 224, according to the down sampling ratios of 1/32, 1/24, 1/16, and 1/12, respectively. **Table 2** shows the recognition rates of different feature dimensions, our method outperforms SRC about 8% on average.

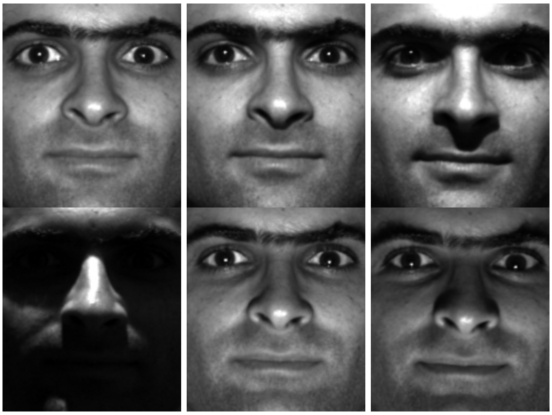


Fig. 6. Samples of different illuminations from the Extended Yale B database

Table 2. Recognition rates of different feature dimensions

Algorithm	**Feature dimension**			
	30	**56**	**120**	**224**
LRC[12]	49.6%	61.9%	71.2%	75.2%
SRC[3]	53.8%	67.0%	81.6%	88.2%
FDDL[13]	25.0%	40.0%	81.5%	92.7%
Our method	67.5%	77.4%	88.5%	91.7%

4.3 Robustness to Occlusion Samples

In this part of the experiments, we choose a subset of the AR database consisting of *both* neutral and corrupted images. We consider the following scenarios:

Sunglasses: We consider corrupted training images due to the occlusion of sunglasses. We use seven neutral images plus one image with sunglasses for training, and the remainder as test images. Note that the presence of sunglasses occludes about 20% of the facial image.

Scarf: We consider the training images corrupted by scarves, occluding roughly 40% of the facial images. We apply a similar training/test set choice, and have a total of eight training images (seven neutral plus one randomly selected image with scarf) and the remainder as test images.

Sunglasses+Scarf: We consider the case that both sunglasses and scarves are presented in the images during training. We choose all seven neutral images and two corrupted images (one with sunglasses and the other with scarf) for training. A total of two test images (one with sunglasses and the other with scarf) are available for this case.

Comparing our method with SRC and FDDL under three different scenarios, **Table 3** summarizes the comparison of three approaches. Our approach achieves the best result and outperforms other approaches more than 7% for the sunglass scenario, 2% for scarf scenario, and 10% for the mixed scenario when the sizes of dictionaries are the same.

Table 3. Recognition rates of different occlusion samples

Dimension300	**Sunglass**	**Scarf**	**Sunglass+scarf**
SRC[3]	82%	69%	76%
FDDL[13]	83%	60%	75%
Our method	90%	71%	86%

4.4 Influence of Training Samples

We conduct experiments according to different amounts of training samples, compare the recognition rate of our algorithm with that of SRC and FDDL. **Table 4** illustrates the results. The recognition rate of our method is higher than that of SRC and almost equal to FDDL by the same amount of training samples, but our method performs better when there are few samples.

Table 4. Recognition rates of different amounts of training samples

Algorithm	**1**	**2**	**3**	**4**	**5**	**6**	**7**
SRC[3]	45.0%	60.2%	68.5%	77.1%	79.9%	83.5%	87.0%
FDDL[13]	----	67.4%	73.4%	79.9%	88.4%	90.7%	92.0%
Our method	53.0%	68.9%	80.2%	83.8%	87.1%	90.0%	91.2%

In all the experiments, we can find that our method outperforms others. The dictionary constructed via NSCT can optimize the discriminative features and is more robust. Compared with SRC and FDDL, our method is higher in recognition rate, especially, when the training samples are limited. It can reduce the training time and also suitable for lack samples situation.

5 Conclusion

In this paper, we propose an efficient discriminative sparse representation face recognition method. The proposed method obtains fused discriminative features according to the characteristics of coefficients in different subbands and their contributions, and constructs a dictionary by learning discriminative features. The dictionary constructed by our method has two characteristics: Firstly, the subdictionary can optimize discriminative features in each class. Secondly, the dictionary has low-rank optimized features. The discriminative power of the dictionary comes from minimizing the characteristic of shift invariable, multi-direction and multi-scale of NSCT. It can reduce the influence of illumination, pose, expression and occlusion in both training and testing samples and performs well when there are few training samples. We apply our algorithm to face recognition and the experimental results clearly demonstrate its superiority to numerous other state-of-the-art methods.

Acknowledgment. We thank Wei Hu for the help and discussion. This work was supported by the National Natural Science Foundation of China (NO.61171077), University Science Research Project of Jiangsu Province (NO.12KJB510025), and the Natural Sciences and Engineering Research Council of Canada.

References

1. Yang, M., Zhang, L., Shiu, S.C.-K., Zhang, D.: Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition. IEEE Trans. on Information Forensics and Securit 7(6), 1738–1751 (2012)
2. Ma, L., Wang, C., Xiao, B., Zhou, W.: Sparse representation for face recognition based on discriminative low-rank dictionary learning. In: Proc. CVPR 2012, pp. 2586–2593 (2012)
3. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
4. Yang, M., Zhang, D., Yang, J.: Robust sparse coding for face recognition. In: Proc. CVPR, pp. 625–632 (2011)
5. Liu, G., Yan, S.: Latent low-rank representation for subspace segmentation and feature extraction. In: Proc. ICCV, pp. 1615–1622 (2011)
6. Bull, G., Gao, J.B.: Transposed low rank representation for image classification. In: Proc. DICTA, pp. 1–7 (2012)
7. Xu, Y., Li, Z., Pan, J.S., Yang, J.Y.: Face recognition based on fusion of multi-resolution Gabor features. Neural Computing and Application 23(5), 1251–1256 (2013)
8. Pong, K.H., Lam, K.M.: Multi-resolution feature fusion for face recognition. Pattern Recognitio 47(2), 556–567 (2014)
9. Cunha, A.L., Zhou, J., Do, M.N.: The nonsubsampled Contourlet transform: theory, design and applications. IEEE Transactions on Image Processin 15(10), 3089–3101 (2006)
10. Li, H.J., Zhao, Z.M., Yu, X.L.: Grey theory applied in non-subsampled Contourlet transform. IET Image Processin 6(3), 264–272 (2012)
11. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters 9(3), 81–84 (2002)

12. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. IEEE Trans. on Pattern Analysis and Machine Intelligenc 32(11), 2106–2112 (2010)
13. Yang, M., Zhang, D., Feng, X.: Fisher discrimination dictionary learning for sparse representation. In: Proc. ICCV, pp. 543–550 (2011)
14. Martinez, A., Benavente, R.: The AR face database. CVC Tech. Report No. 24 (1998)
15. Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. on Pattern Analysis and Machine Intelligence 27(5), 684–698 (2005)