ATLANTIS PRESS

Advances in Computer Science Research, (ACSR), volume 52

2016 International Conference on Computer Engineering and Information Systems (CEIS-16)

Robust Face Recognition Based on DCNN and CRC

Li-Na Yuan, Feng Cen

College of Electronics and Information Engineering

Tongji University

Shanghai, China

E-mail: 1433177@tongji.edu.cn, feng.cen@tongji.edu.cn

***Abstract*-Collaborative representation based classification (CRC) has gained popularity in recent years, but the conventional features could not effectively handle the variations of pose and occlusion such as sunglass. In this paper, the deep convolution neural network (DCNN) is introduced in the feature extraction to tackle the facial deformation, particularly, such as pose variation. After analysis on the feasibility of the dictionary construction on DCNN based features, a DCNN feature based occlusion dictionary computing algorithm is then presented to tackle the face recognition with occlusion. Experiments on representative face databases with variations of pose and occlusion demonstrated the effectiveness of the proposed algorithm scheme.**

Keywords-face recognition;sparse representation; collaborative representation;deep learning

I. INTRODUCTION

Due to the emerging demand in surveillance and security, face recognition (FR) becomes an important research topic in pattern recognition. Various factors such as background illumination, pose and facial corruption/disguise can easily affect the performance of the proposed face recognition method. Moreover, some intrapersonal variations, such as occlusion, tend to be larger than interpersonal variations. Thus, developing a robust and practical face recognition system that could handle all the possible facial transformation remains a very challenging task.

In dealing with the above mentioned problems, many face recognition techniques have been proposed during the past years. However, most existing face recognition technologies are still far from the perfection in uncontrolled cases.

In the aspect of improving the tolerance of the corruption, John Wright et al. [1] introduced the sparse representation classification (SRC) to the face recognition. It has been proved to be robust to the occlusion. But it is under the assumption that the face image has been well aligned. Zhang Lei et al. [2] put forward the improved the method called Collaborative Representation Classification (CRC). Its improvement mainly focuses on the computational complexity. Its running speed is 20 times quicker than SRC. Both the above methods could not handle the unconstrained circumstance [3].

However, on unconstrained occasions, Deep Convolution Neural Network (DCNN) technique have been proved efficient and dominated in the field of face recognition. The work in [4] has shown that the deep learned features are much more robust to face deformation, such as pose variation. For the face images of different angles are not in the same linear subspace, the image vectors can't be directly used in SRC. In that way, we could use DCNN to extract the high-dim features ensuring them in the same subspace while retaining the recognition-related information.

In this paper, we aim to develop an algorithm based on CRC and DCNN to address the problem of face recognition with pose variation and occlusion. We use DCNN to extract the feature vectors from face images. The feature vectors extracted from the training face images are stacked as the training gallery, and then the CRC algorithm is followed to achieve face recognition. We test the performance of the proposed algorithm in AR and FERET database under the circumstances both with and without occlusion images and pose variations. In condition of the occlusion of sunglass, we introduce an occlusion dictionary which stacked by the extracted features using DCNN from the images with sunglass. These images do not belong to the identities in the training and testing set. All the above experiments got relatively high recognition rates.

The rest of the paper is organized as follows. Section 2 briefly reviews related work of face recognition method based on sparse representation classifications and deep learning which aim to resolve the occlusion and pose variations. Section 3 illustrates the method in this paper, and the feasibility of the occlusion dictionary based on the features extracted by DCNN. Section 4 conduct experiments and Section 5 conclude the paper.

II. RELATED WORK

The sparse representation based classification methods have a good performance for face recognition in the occlusion condition. Most of improved methods based on SRC focus on handling with unconstrained environment with pose variation or illumination variation while not damaging the performance in occlusion. Yang Meng et al. [5] has introduced Gabor feature into the SRC. This method has a little improvement in the misalignment circumstance, but it failed to obtain a good recognition result in the pose variation at 25 degrees. More recently, Liansheng Zhuang et al.[6] has introduced a sparse illumination learning and transfer (SILT) technique into SRC to address the situation of single sample regime and fewer restrictions. But it has just considered the circumstance of rotation and scale, failed to take pose variation into account. Particularly, Xin Zhang et al. [7] have proposed an improved method called mixed-

norm sparse representation which is customized to solve the problem of multi-view in face recognition. And it has achieved a relatively good result but not consider the circumstance of occlusion. Thus, most of the proposed method specialized in solving one aspect in the field of face recognition

Under unconstrained environment which include pose variation, deep learning based methods are prevailing in recent years and have shown immensely impressive results [8][9]. Mostafa Mehdipour Ghazi et al. [10] have analyzed two most popular and advanced deep learning based approaches, namely VGG-Face [11] and Lightened CNN [12]. The result has shown that the two deep learning models are benefit for the performance of face recognition when they have been utilized in the preprocessing. Moreover, when used for pose and illumination normalization, they could help to achieve better performance for face recognition. However, the experiments in [10] have shown that it has got a bad performance in AR sunglass image set, only achieving less than 50% recognition rate. Wael AbdAlmageed et al. [13] use several pose specific deep convolution neural network (CNN) models to generate multiple pose-specific features. Jun-Cheng Chen et al [14] present an algorithm for unconstrained face verification based on deep convolution features, and obtain a good recognition result on the occasion with pose variations.

In addition, Liang H et al. [15] has used Using Sparse Representations of Convolution Neural Network Features in the field of image classification and has achieved a good result, which indicates the feasibility of this kind of integration.

III. PROPOSED METHODS

The section is divided into two parts. The first part shows the algorithm framework for pose variation. The method adopts the DCNN method to extract the features to avoid additional alignment or other preprocesses. The second part presents the improved method for occlusion such as sunglass. Particularly, the occlusion dictionary which consists of the DCNN features of the extra occlusion images of the same occlusion pattern. The occlusion dictionary is introduced to collaboratively represent the test features, for the reason that the features in train dictionary can't linearly represent the occlusion part of test features, while the occlusion dictionary can offset occluded part.

A. DCNN-feature Based CRC

The first part mainly talks about the situation when pose variation includes. Some algorithm will add extra process of alignment to address such kind of problem, or introduce additional pose dictionary. However, the deep convolution neural network can extract the pose-invariant features to help the unconstrained face recognition. For example, using VGG [11] to extract the features could handle pose variations of up to 67.5 degree. So the method proposed in this paper chooses the DCNN to extract the features to exploit its pose-invariance, and then uses the features to collaboratively represent the feature of test image. The method is illustrated ad figure 1.

The DCNN-based feature extraction includes several processes of convolution and pooling which followed by several layers of full connection. The images should be first transformed to the vector before the step of feature extraction. The DCNN feature extraction can be illustrated as follows.

$$f : \mathbb{R}^{h \times w} \to \mathbb{R}^{z} \tag{1}$$

where the original image's size is $h \times w$, after the feature extraction (which is represented by f), the image feature whose dimension is z is obtained. Then DCNN extraction of train and test image vectors could be expressed as follows.

$$\chi = f\left(I_{train}\right) \tag{2}$$

$$y = f\left(I_{test}\right) \tag{3}$$

where I_{train}, I_{test} represent, respectively, original train and test images. f, which represents the process of DCNN based feature extraction, specifically includes several layers of deep convolution neural network. This process should retain the identity-related information for recognition. The DCNN-extracted face feature vector χ appears to cluster semantic topics more readily than conventional features, and more robust to the variations like pose, expressions and illumination. So we propose to use χ to replace conventional face features in the CRC framework. y represents the DCNN extracted features of the test images.

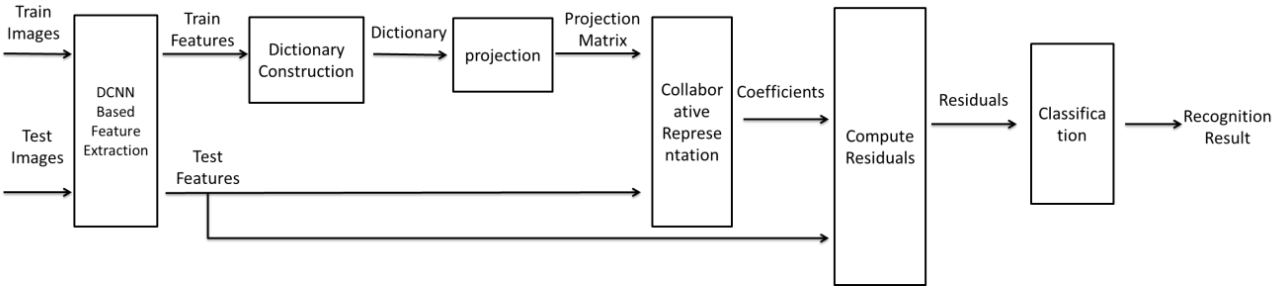

Figure 1. The basic algorithm framework.

The features vectors χ of the training images are then stacked to a matrix called the dictionary A, which are used to collaboratively represent the test target. Assume that there are K subjects, and each subject has H images in training gallery, i.e. $A \in \mathbb{R}^{m \times n}$, where $n = H \times K$, and m is the feature vectors' dimension after DCNN-based feature extraction and PCA feature dimension reduction. The dictionary construction can be expressed as follows.

$$A_i = [\chi_{i1}, \chi_{i2}, \cdots, \chi_{iH}] \tag{4}$$

$$A = [A_1, A_2, \cdots, A_K] \tag{5}$$

To code y with dictionary A, a projection is needed to make the obtained projection matrix independent of y, just as described in [2],i.e.,

$$P = (A^T A + \lambda I)^{-1} A^T \tag{6}$$

Because P is independent of y, it can be pre-calculated. Once a query sample y comes, we can just simply project y via Py, which is shown as,

$$\hat{p} = Py \tag{7}$$

The residual of collaborative representation of the features for identity $r_i(y)$ can be obtained.

$$r_i(y) = \|y - A_i \hat{p}_i\|_2 / \|\hat{p}_i\|_2 \quad for \quad i = 1, \cdots, K, \tag{8}$$

where $\hat{p}_i$ is the coefficients vectors associated with class i. The $\|\hat{p}_2\|$ can also bring some discrimination information for classification.

Classifying y is achieved on the basis of these approximations by assigning it to the object class that minimizes the residual $r_i(y)$, i.e.,

$$identity(y) = Identity(min(r_i(y))) \tag{9}$$

B. Occlusion Dictionary

When occlusion is included, especially, under the condition of wearing sunglass, the occlusion dictionary is introduced in this method to achieve a good recognition result. However, on the situation of wearing scarf, occlusion dictionary is not specially introduced. The reason is that DCNN based feature extraction is not sensitive to the region of mandibles, while the region of eyes plays an overwhelming role in DCNN for face recognition.

To illustrate the rationality of this conjecture, we first visualize the shared features of the images of different individuals for normal faces and wearing sunglass or scarf, to further examine the role of the region of eyes and mandibles in DCNN based feature extraction. Part of AR database [16] is used and divided into 3 groups according to deformations of the images. In Figure 2, the solid dot indicates the shared elements of the feature vectors among different persons. To make the illustration clearer, we reshape the 4096×1 vector into 64×64 matrix.

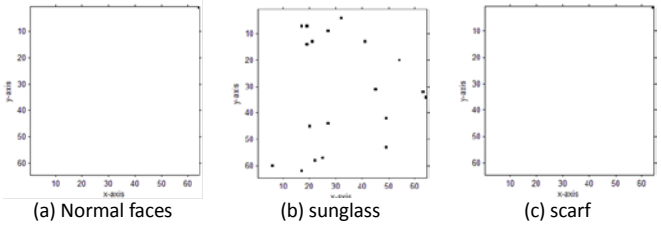

Figure 2. The similarity of the feature vector in 2-dim plot. The dots indicate the shared features across the images of different individuals.

As illustrated in figure 2, DCNN based features of the images with sunglass have some share entries and with scarf have almost no shared entries. So, DCNN based feature extraction is really sensitive to the region of eyes, while the region of mandibles does not have an important influence. For the DCNN based features of images with sunglass share some feature elements, the feature vectors of different persons can be used to reduce the residuals for collaborative representation with the train dictionary to improve the recognition rate under the condition of wearing sunglass.

Then, t-SNE [17] is used to show why the occlusion dictionary should be used for the recognition of the images wearing sunglasses. By finding a 2-dimensional embedding of the high-dimensional feature space for the condition of wearing sunglass, and plotting them as points colored depending on their semantic category in a particular hierarchy, the visualization result is as shown in Figure 3.

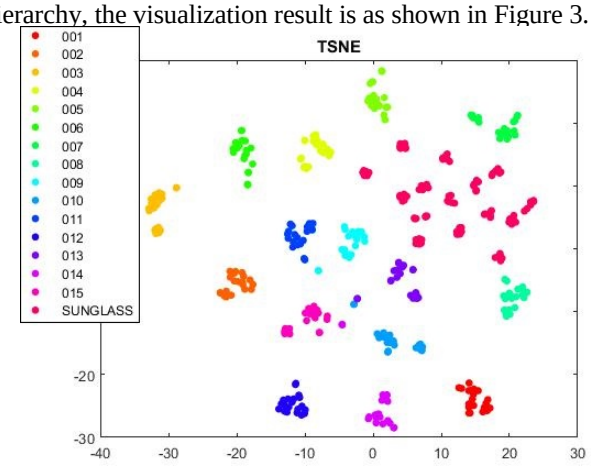

Figure 3. The visualization of the high-dimensional features in 2-dim plot.

The figure 3 shows the group of features of all above 15 identities with sunglass showing a clear semantic clustering. So the deep convolution neural network captures the semantic difference between the non-occluded and occluded images.

Thus, for the situation of sunglass, the DCNN features based occlusion dictionary is introduced in this paper. The occlusion dictionary $A_e \in \mathbb{R}^{m \times n_e}$ (where n_e is the size of the occlusion dictionary) consists of the DCNN extracted features of the images. The images used in occlusion dictionary belong to the identities that are not included in the train or test images. Then occlusion dictionary is stacked after the train dictionary $A \in \mathbb{R}^{m \times n}$ to form dictionary matrix $B \in \mathbb{R}^{m \times (n+n_e)}$, just illustrated as below.

$$B = [A \quad A_e] \tag{10}$$

The projection matrix P is changed to

$$P = (B^T B + \lambda I)^{-1} B^T \tag{11}$$

The coefficients of collaborative representation by the projection matrix are computed as follows.

$$[\hat{p} \quad \hat{p}_e] = Py, \tag{12}$$

where p_e is the coefficients vector of occlusion dictionary.

The class specific representation residual for the target $r_i(y)$ becomes

$$r_i(y) = \|y - A_i \hat{p}_i - A_e \hat{p}_e\|_2 / \|\hat{p}_i\|_2 \text{ , for } \quad i = 1, \cdots, k \tag{13}$$

Then we classify y based on the approximations by assigning it to the object class that minimizes the residual $r_i(y)$, as shown by (9).

The whole algorithm is described in the following Algorithm 1.

Algorithm 1

1. For each training and testing image, DCNN is used to extract the features.
2. Use PCA to achieve dimensionality reduction
3. Normalize the columns of A (in the case of non-occlusion) or B (in the case of occlusion) to have the L2-normwhere $B = [A \quad A_e]$, A_e is the occlusion dictionary which is composed by the occluded features.
4. Code y over A or B by

$$\hat{p} = Py \tag{5}$$

Where $P = (A^T A + \lambda I)^{-1} A^T$

$$[\hat{p} \quad \hat{p}_e] = Py \tag{10}$$

or $P = (B^T B + \lambda I)^{-1} B^T$

5. Compute the residuals

$$r_i(y) = \|y - A_i \hat{p}_i\|_2 / \|\hat{p}_i\|_2 \text{ , for } \quad i = 1, \cdots, k \tag{7}$$

Or

$$r_i(y) = \|y - A_i \hat{p}_i - A_e \hat{p}_e\|_2 / \|\hat{p}_i\|_2 \text{ , for } \quad i = 1, \cdots, k \tag{12}$$

6. Output that $identity(y) = Identity(min(r_i(y)))$

IV. EXPERIMENT

In this section, we perform experiments on benchmark face databases to demonstrate the improvement of our method over CRC. To evaluate more comprehensively the performance of the method proposed in this paper, in section 4.1, we first test face recognition under constrained conditions with illumination and slight expression variation only, and then in section 4.2, we test the proposed algorithm against face recognition with pose variations. Finally in section 4.3, we demonstrate the robustness and efficiency of the method with occlusion dictionary in face recognition with disguise occlusion.

Here we should also note that the regularization parameters in sparse coding are tuned by experience (Actually, how to adaptively set the regularization parameters is still an open problem). In addition, all the face images are cropped, and, the specific method of deep convolution neural network used in this work is VGG-face [11]. It is an open-source framework of Deep Convolution Neural Network. The network involves 16 convolution layers, five max-pooling layers, three fully connected layers. VGG-Face takes color image patches of size 224 × 224 pixels as the input and utilizes dropout regularization in the fully-connected layers. Moreover, it applies ReLU activation to all of its convolution layers. The first two FC layers' output is 4,096 dimensional and the last FC layer has either N = 2622 or L = 1024 dimensions, depending upon the loss functions used for optimization. In this work, we use the first FC layer's output as the extracted features.

A. AR with Illumination And Expression Variation

As in [1], a subset (with only illumination and expression changes) that contains 50 male subjects and 50 female subjects was chosen from the AR database [16] in our experiments. For each subject, the first seven images without occlusion were used for training, with the latter seven images without occlusion for testing. The images are transformed into gray images, and resized to 60×43, just as the experiment in [2]. And for our method, the processed images are then resized to 224×224 and transformed to the three-channel pseudo-color image for the reason that the VGG only accepts this specification of image. The comparison of methods is given in figure 4.

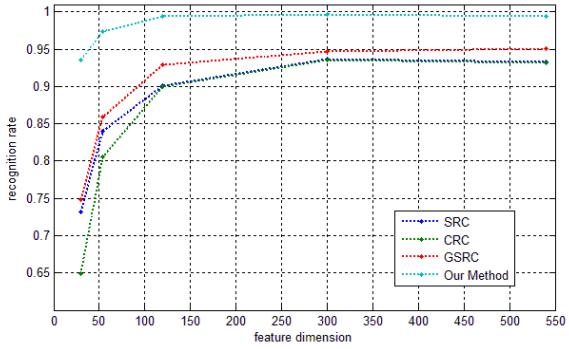

Figure 4. Face recognition rate.

From figure 4, we can achieve the best result when dimensionality between 100 and 300. The recognition rate of our method is about 5% higher than SRC, CRC_RLS and GSRC. Even when feature dimension is small, our method can also achieve a good recognition result. This shows that our method has made some improvement.

B. FERET Pose Database

Here we used the pose subset of the FERET database [18], which includes 1365 images from 195 subjects (seven pictures each). This subset is composed of the images marked 'ba','bd','be','bf','bg','bj', and 'bk'. Figure.5 compares our method (feature dimension=200 for best result) with GSRC (feature dimension=380 for best result) for different poses. The λ in both methods is 0.01. Specifically, in our Method, each image has resized to 224×224 form the size of 80×80 and transform the single-channel gray image to three-channel pseudo-color image, for VGG only accept the color image as input. Some sample images of one person are shown in the Figure.5

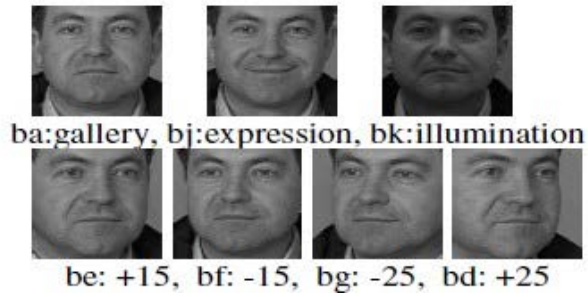

Figure 5. Samples of one subject.

Five tests with different pose angles were performed. When the testing pose angle is zero degree, images marked with 'ba' and 'bj' were used as training set, and images marked with 'bk' were used as testing set. While testing the pose variation of 25 and 15 degree left/right, we used images marked with 'ba', 'bj' and 'bk' as gallery, and used the images with 'bg', 'bf ', 'be' and 'bd' as probes.

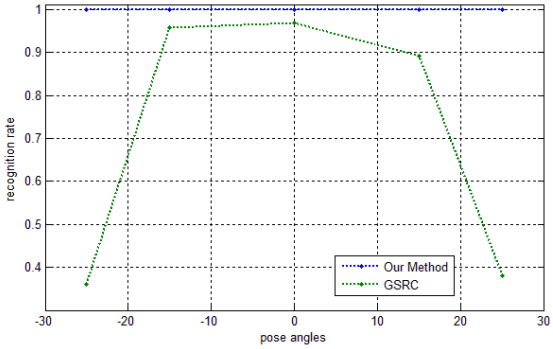

Figure 6. The recognition rate.

As shown in figure 6, the GSRC could not achieve a good recognition rate in the angel of -25 and 25 degree, while our method could achieve a good face recognition rate in the pose variation range from -25 to 25, with 97% recognition rate in -25 and 25 degree. The result shows that our method is more efficient for face recognition with pose variation.

C. AR with Real Face Disguise

As in [2], AR database consists of 1400 images from 100 subjects, 50 females and 50 males, we separate it into two parts, 800 images (about 8 samples per person) of non-occluded frontal views with various facial expressions were used for training, while the others with sunglasses and scarves (as shown in Figure. 7) were used for testing. The images are resized to 83×60 , just as the corresponding experiments in [1],[2],[5]. The results are shown in table 1.

Figure 7. The testing samples with sunglasses and scarves in the AR database.

TABLE I. THE RESULT OF FACE RECOGNITION WITH REAL DISGUISE USING AR DATABASE

Method	Sunglass	Scarf
SRC	87.0%	59.5%
CRC_RLS	68.5%	90.5%
GSRC	93.0%	79%
Our method without occlusion dictionary	67.5%	93%

Above experiments like CRC_RLS and Our method are not included the occlusion dictionary. Then we introduce DCNN extracted features based occlusion dictionary, and design the experiment to examine the validity of the introduced occluded dictionary for the frontal views with sunglass. We separate the database into two parts, the former 7 images of former 35 females and former 35 males without occlusion are used as train images, the former 3

images of above identities with sunglass are used as test images. Specifically, in our method, the whole images of the latter 5/10/15 females and latter 5/10/15 males with sunglass are used as the occlusion dictionary. The corresponding experiments are labeled as with OccDic size 10, with OccDic size 20, with OccDic size 30.

TABLE II. THE RESULT OF FACE RECOGNITION WITH OCCLUSION DICTIONARY

Occluded dictionary size	**30**	**54**	**120**	**300**
DCNN+Softmax	0.5143	0.5905	0.6333	0.6286
CRC	0.4096	0.3524	0.4619	0.6538
no OccDic	0.5667	0.7238	0.7191	0.6857
with OccDic size 10	0.6619	0.7714	0.8238	0.7571
with OccDic size 20	0.6571	0.7857	0.8381	0.7667
with OccDic size 30	0.7000	0.7619	0.8381	0.8000

By comparing the results of introducing and not introducing the occluded dictionary, table 2 shows that the recognition result could be improved by the introduced occlusion dictionary.

V. CONCLUSION

This paper integrates collaborative representations into deep learning techniques for face recognition, for deep learning technique has a strong ability of semantic representation of features. In addition, the DCNN based occlusion dictionary is introduced to tackle the occasion of sunglass. Experiments have shown that the method in this paper achieves good recognition result on the occasion with occlusion and poses variation. When considering the problem with pose variation, our method avoids extra alignment, which makes the recognition easier.

REFERENCES

[1] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(2): 210-227.

[2] Zhang, Lei, Meng Yang, and Xiangchu Feng. "Sparse representation or collaborative representation: Which helps face recognition?." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011

[3] Wagner A, Wright J, Ganesh A, et al. Toward a practical face recognition system: Robust alignment and illumination by sparse representation [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(2): 372-386

[4] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2892-2900.

[5] Yang Meng, and Lei Zhang. "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary." Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010. 448-461.

[6] Zhuang L, Chan T H, Yang A Y, et al. Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment[J]. International Journal of Computer Vision, 2015, 114(2-3): 272-287.

[7] Zhang X, Pham D S, Venkatesh S, et al. Mixed-norm sparse representation for multi view face recognition[J]. Pattern Recognition, 2015, 48(9): 2935-2946.

[8] Sun Y, Liang D, Wang X, et al. Deepid3: Face recognition with very deep neural networks[J]. arXiv preprint arXiv:1502.00873, 2015.

[9] Fan H, Cao Z, Jiang Y, et al. Learning deep face representation[J]. arXiv preprint arXiv:1403.2802, 2014.

[10] Ghazi M M, Ekenel H K. A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition[J]. arXiv preprint arXiv:1606.02894, 2016.

[11] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//British Machine Vision Conference. 2015, 1(3): 6

[12] Wu X, He R, Sun Z. A Lightened CNN for Deep Face Representation[J]. arXiv preprint arXiv:1511.02683, 2015.

[13] AbdAlmageed W, Wu Y, Rawls S, et al. Face recognition using deep multi-pose representations[C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016: 1-9.

[14] Chen J C, Patel V M, Chellappa R. Unconstrained face verification using deep cnn features[C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016: 1-9.

[15] Liang H, Li Q. Hyperspectral Imagery Classification Using Sparse Representations of Convolutional Neural Network Features[J]. Remote Sensing, 2016, 8(2): 99

[16] Martinez A M. The AR face database[J]. CVC Technical Report, 1998, 24.

[17] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2579-2605): 85.

[18] Phillips P J, Wechsler H, Huang J, et al. The FERET database and evaluation procedure for face-recognition algorithms[J]. Image and vision computing, 1998, 16(5): 295-306.