

Research Article

Occluded Face Recognition Based on Double Layers Module Sparsity Difference

Shuhuan Zhao and Zheng-ping Hu

School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

Correspondence should be addressed to Zheng-ping Hu; hzp_ysu@163.com

Received 14 May 2014; Revised 5 August 2014; Accepted 6 August 2014; Published 18 August 2014

Academic Editor: Durga Misra

Copyright © 2014 S. Zhao and Z.-p. Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image recognition with occlusion is one of the popular problems in pattern recognition. This paper partitions the images into some modules in two layers and the sparsity difference is used to evaluate the occluded modules. The final identification is processed on the unoccluded modules by sparse representation. Firstly, we partition the images into four blocks and sparse representation is performed on each block, so the sparsity of each block can be obtained; secondly, each block is partitioned again into two modules. Sparsity of each small module is calculated as the first step. Finally, the sparsity difference of small module with the corresponding block is used to detect the occluded modules; in this paper, the small modules with negative sparsity differences are considered as occluded modules. The identification is performed on the selected unoccluded modules by sparse representation. Experiments on the AR and Yale B database verify the robustness and effectiveness of the proposed method.

1. Introduction

Image recognition, especially face recognition, has attracted a lot of researchers due to its wide application. And many methods have been proposed to solve this problem, including PCA, LDA, SVM, and other related methods. Recently, sparse representation- (or coding-) based classification (SRC) is attracting more and more attention [1–3] and has gained great success in face recognition. Based on sparse representation, Qiao et al. propose sparsity preserving projections (SPP) [4] for unsupervised dimensionality reduction. It can preserve the sparse reconstructive weights and the application on the face recognition verifies the effective SPP. Although these methods perform well under some controlled conditions, they fail to perform well in the situation when test data is corrupted due to occlusion. To solve this problem, Wanger et al. proposed to solve the problem in paper [5] by extending training samples using the difference between samples. And paper [6] used the image Gabor-features for SRC, which can get a more compact occlusion dictionary; as a result, the computation complexity and the number of atoms were reduced. In addition, paper [7] proposed a novel low-rank matrix approximation algorithm with structural incoherence

for robust face recognition. In this paper the raw training data was decomposed into a low-rank matrix and the sparse error matrix. Besides, it introduced structural incoherence between low-rank matrices which promoted the discrimination between different classes, and thus this method exhibits excellent discriminating ability. However, all these methods have a common condition that they need occluded samples in the training set. If there are no occluded samples in the training set, how to solve the problem? Paper [8] modeled the sparse coding as a sparsity-constrained robust regression problem and proposed a new scheme, namely, the robust sparse coding (RSC) that seeks the MLE (maximum likelihood estimation) solution of the sparse coding problem, which makes it much more robust to outliers than SRC. And paper [9] proposed WGSR (modular weighted global sparse representation) for robust face recognition; in this paper image was first divided into modules and each module was processed separately by SRC; then the modular reliability was determined by the modular sparsity and residual jointly. Finally a reconstructed image from the modules weighted by their reliability was formed for the robust recognition. Paper [10] proposed a new approach. It detected the presence of sunglasses/scarves firstly and then processed the

nonoccluded facial regions only. The occlusion detection problem was approached by PCA and improved support vector machines (SVM), while the nonoccluded facial part was identified by block-based weighted local binary patterns (LBP). Paper [11–13] also divided image into modules but changed the strategy of division into randomized modules, multilevel modules, and horizontal modules.

To address the problem of face recognition with occlusion when there is no occlusion in the training samples, we proposed a new method named occluded face recognition based on double layer module sparsity difference (FR_DLMSD). Firstly, we partition the images into four blocks and sparse representation is performed on each block, so the sparsity of each block can be obtained; secondly, each block is partitioned again into two small modules. Sparsity of each small module is calculated as the first step. Finally, the difference of small module's sparsity with the corresponding block is used to detect the occluded modules; in this paper, the modules with negative sparsity differences are considered as occluded modules. The identification is performed on the selected unoccluded modules by sparse representation.

2. Related Works

In this section, we briefly introduce some works on the image representation. Some most popular face recognition methods, including sparse coding-based methods and nearest feature-based classifiers (NFCs), are reviewed.

In general, NFCs aim to find a representation of the query image and classify it to the class with the lowest residual. According to the mechanism of representing the query image, NFCs include NN, nearest feature line (NFL) [14], nearest feature plane (NFP) [15], and nearest feature subspace (NFS). Among these methods, NN is the simplest one with no parameters which classifies the query image to its nearest neighbor. The performance of NN can be easily affected by noises, for NN adopts only one sample to represent the query image. NFL classifier proposed by Li et al. forms a line by every two training samples of the same class and classifies the query image to its nearest line. Chen et al. proposed the NFP classifier which uses at least three training samples of the same class to form a plane rather than a line to determine the label of the query image. Instead of using a subset of the training samples with the same label to represent the query image like NN, NFL, and NFP, NFS represents the query image by all training samples of the same class. In general, the more the samples are used for representation, the more stable a method is supposed to be [16]. Hence, NFS is assumed to perform better than the other NFCs. Besides, NFCs are not robust in real-world face recognition applications because of various occlusions.

To overcome these problems, Wright et al. introduced the SRC method to represent the query image. Sparse representation for classification (SRC) seeks the sparsest representation of query sample y in the overcomplete dictionary D . Suppose that there exist k subjects and each training sample can be represented as vector $v_{i,n}$, so all the samples from the i th class construct a matrix $D_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times n_i}$, where

$m = w \times h$ means the dimensions of training sample and n_i means the number of training sample from the i th class. It supposes that samples from the same class construct a linear subspace, so any test sample can be represented as linear combination of samples from the same class; for example, test sample $y \in R^m$ from class i can be represented as

$$y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \dots + \alpha_{i,n}v_{i,n}, \quad (1)$$

where $\alpha_{i,j} \in R, \quad j = 1, 2, \dots, n_i$.

Since we do not know which class the test sample belongs to, it identifies a dictionary as $D = [D_1, D_2, \dots, D_k] = [v_{1,1}, v_{1,2}, \dots, v_{k,n}] \in R^{m \times N}$, where m denotes the dimension of training sample, k denotes the number of objects, and $N = \sum_{i=1}^k n_i$ refers to the total number of training samples. The test sample y can be represented by the linear combination of all training samples as

$$y = Dx_0 \in R^m, \quad (2)$$

where $x_0 = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n}, 0, \dots, 0]^T \in R^N$ denotes the vector of coefficients. In x_0 the nonzero atoms correspond to the training sample from the same class as the test sample. In image recognition, (2) is usually underdetermined; that is, $m < N$, so there will be many solutions. Since we know that x is sparse, we can restrain the equation by ℓ_0 norm:

$$(\ell_0): \quad \hat{x}_0 = \operatorname{argmin} \|x\|_0 \quad \text{s.t. } Dx = y. \quad (3)$$

Because (3) is a NP-hard problem, according to sparse representation and compressive sensing, we can replace it by ℓ_1 norm only if x_0 is sparse enough, so there is

$$(\ell_1): \quad \hat{x}_1 = \operatorname{argmin} \|x\|_1 \quad \text{s.t. } Dx = y. \quad (4)$$

Since there is noise in image, the linear combination of training samples cannot represent test sample accurately, so it permits the existence of error and defines the limit of error ε . Equation (4) can be converted into the following form:

$$(\ell_1): \quad \hat{x}_1 = \operatorname{argmin} \|x\|_1 \quad \text{s.t. } \|Dx - y\|_2 \leq \varepsilon. \quad (5)$$

Finally, it computes the residuals of each class and classifies the test sample into the class with lowest residual:

$$\operatorname{identify}(y) = \operatorname{argmin}_i r_i(y), \quad (6)$$

where $r_i(y) = \|y - D\delta_i(x)\|_2$, $i = 1, 2, \dots, k$, and $\delta_i(x) = [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n}]^T$ denotes the vector of coefficient of i th class.

3. Occluded Face Recognition Based on Double Layer Module Sparsity Difference

For the problem of image recognition with occlusion when there is no occlusion in the training samples, we proposed a new method named occluded face recognition based on

TABLE 1: Comparison of recognition rate when image occluded with scarf.

Method	Dim						
	80	120	320	500	720	980	1280
FR_DLMSD	0.5182	0.4818	0.5994	0.6737	0.7117	0.7171	0.6779
SRC	0.2129	0.2507	0.3347	0.3656	0.3599	0.3936	0.4048
SPP	0.1807	0.2157	0.1401	0.1625	0.4398	0.5071	0.5546
NN	0.1917	0.2306	0.3042	0.3375	0.3278	0.3722	0.3819
NFS	0.2000	0.2334	0.2917	0.3194	0.3222	0.3528	0.3667

double layer module sparsity difference (FR_DLMSD). First, the occluded test image y and the training samples are partitioned into 4 blocks as Figure 1(a) shows; then y can be rewritten as $y = [y_1, y_2, y_3, y_4]$ and D is also composed by 4 parts as $D = [D_1, D_2, D_3, D_4]$. For each block, sparse representation is performed; then we have $y_i = D_i x_i + \varepsilon_i$, so each block's sparsity can be computed as follows:

$$S_i = \frac{\sqrt{C}}{\sqrt{C} - 1} \left(\frac{\|x_i\|_2}{\|x_i\|_1} - \frac{1}{\sqrt{C}} \right), \quad (7)$$

in which C denotes the number of training classes. From (7) we know that when $S_i = 1$, the coefficient x_i is the sparsest; when $S_i = 0$, the coefficient x_i 's sparsity is the minimum. The sparsity of the four blocks is represented as $Sb = [Sb_1, Sb_2, Sb_3, Sb_4]$.

To detect the occluded modules, each block is partitioned into 2 modules as Figure 1(b) shows. Then test sample y and the dictionary D can be rewritten as $y = [y_{1,1}, y_{1,2}, y_{2,1}, y_{2,2}, y_{3,1}, y_{3,2}, y_{4,1}, y_{4,2}]$ and $D = [D_{1,1}, D_{1,2}, D_{2,1}, D_{2,2}, D_{3,1}, D_{3,2}, D_{4,1}, D_{4,2}]$, respectively. The sparse representation coefficient on the corresponding sub-dictionary is $x = [x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}, x_{3,1}, x_{3,2}, x_{4,1}, x_{4,2}]$. Similarly, the sparsity of each module calculated by (7) can be denoted as $Ss = [Ss_{1,1}, Ss_{1,2}, Ss_{2,1}, Ss_{2,2}, Ss_{3,1}, Ss_{3,2}, Ss_{4,1}, Ss_{4,2}]$. According to the theory of sparse representation, we know that when there is no occlusion in the test sample, then the image can be reconstructed well by the training samples and the corresponding coefficient has higher sparsity; when the image is occluded, then it cannot be represented well by the dictionary, so the corresponding coefficient's sparsity should be lower. Figure 1(c) shows the sparsity of different blocks and modules. It verifies that when the module is occluded its sparsity is lower than the corresponding block, such as Sb_3 versus $Ss_{3,2}$ and Sb_4 versus $Ss_{4,2}$. According to the above analysis, the sparsity difference of the module and its corresponding block is used to estimate the occluded parts: if the difference is negative then the module is viewed as occluded; if it is positive then the module can be used for identification. Suppose k modules are preserved and we have $y_f = [y_{f1}, y_{f2}, \dots, y_k]$ and $D_f = [D_{f1}, D_{f2}, \dots, D_k]$ in which y_{fi} and D_{fi} mean the module with positive sparsity difference. Finally the identification is performed on y_{fi} and D_{fi} by SRC.

4. Experiments

We evaluate the performance of the proposed method on AR database and Yale B database. SRC, SPP, NN, and NFS are selected as benchmark. The parameter λ of methods is set as 0.1.

4.1. AR Database. The AR database consists of over 4000 frontal-face images from 126 subjects, among which 26 pictures were taken in two separate sessions for each subject. In this experiment, 119 subjects are selected. And, for each subject, we choose 7 images without occlusion for training and 3 images with scarf and 3 images with sunglasses for testing. All the chosen images have a little difference in expression, hair type, occlusion, and so forth. Figure 2 shows some images of AR database, in which images on the first line are training samples and images on the second line are test images. Before the experiment, all the images are downsampled to reduce dimensions. Since the pictures are taken in two separate sessions, we do experiments on each session, respectively, and the mean values are taken as the final result.

The recognition result when the test image wears sunglasses is shown in Figure 3. Compared with other methods, our method FR_DLMSD has no obvious advantages, but FR_DLMSD can get a little higher recognition rate than others. One reason maybe that the occlusion rate of sunglasses is low, so SPP, NN, NFS, and FR_DLMSD can capture the discriminant information, so their recognition rates are much higher than SRC. While SRC is a global method and its recognition rate is much lower than other local methods, so when the image is occluded by sunglasses the SRC can not achieve satisfying performance.

The average accuracy when the query image wears scarf is shown in Table 1. It can be seen from Table 1 that the recognition rates of methods are much lower than that of image with sunglasses. The occlusion rate of scarf is large. The performance of SPP, SRC, NN, and NFS is much poorer, which means that these methods are much sensitive to occlusion. Although NN and NFS are local methods, their recognition rates are very low. With the increasing of the occlusion rate, more discriminate information is occluded and the recognition rate declines rapidly. The occlusion has changed the structure of the data and processing image as a whole is not appropriate. But our method partitioned image into modules and can estimate the occlusion part, so it can obtain much better result and the recognition rate is

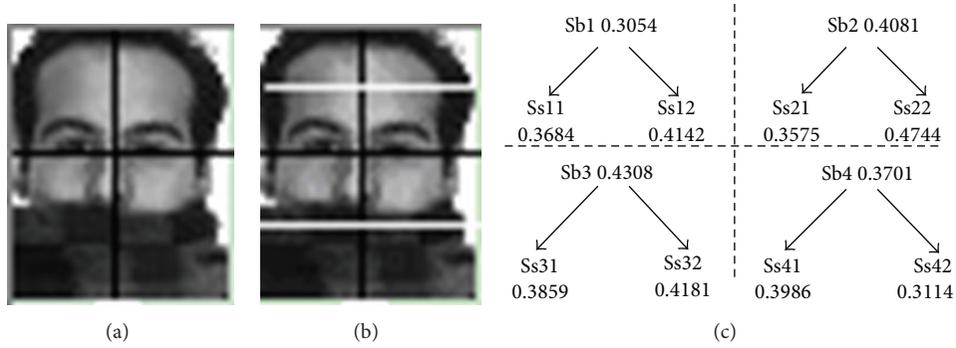


FIGURE 1: Introduction of FR_DLMSD. (a) Partition on the first layer; (b) partition on the second layer; (c) images sparsity on different layers.



FIGURE 2: Sample images on AR database.

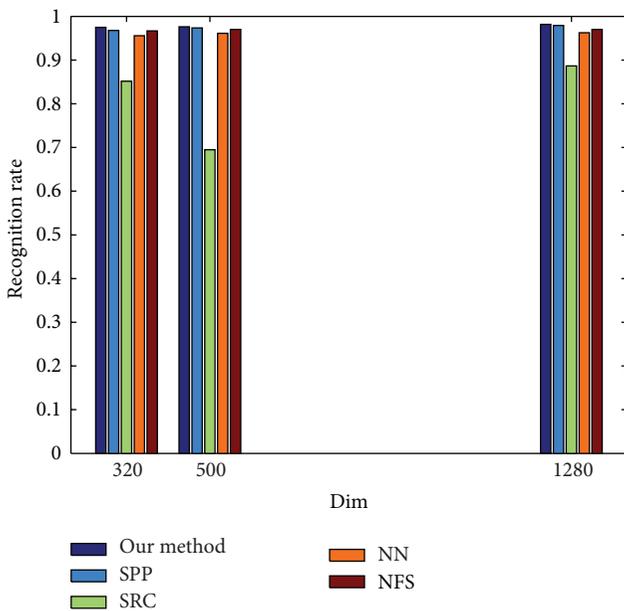


FIGURE 3: Recognition result when the image is with sunglasses.

improved about 30% than others. It also verifies that our method is effective in some extent.

4.2. Extended Yale B Database. The extended Yale B database consists of images of 38 subjects captured under different

lighting conditions. We select 7 images of each subject for training as Figure 4(a) shows. To detect the performance of our method, occlusion with different rates is added to the test image as Figures 4(b)–4(d) show. For different occlusion rates, 3 images of each subject are selected for testing. For each given occlusion rate and dimension, we repeat the experiment 5 times with different training samples and test images and the average accuracies are shown in Figures 5, 6, and 7.

In this experiment SRC, SPP, NN, and NFS are still selected as benchmark. Figure 5 to Figure 7 show the recognition rate comparison of different methods with different occlusion rates. All the figures show that our method can get higher recognition rate than others. And the higher the occlusion rate, the more obvious the advantage of our method than the others.

The results indicate that with the increasing occlusion rate, the performance of SRC declines seriously especially when the occlusion rate is 50%. When the occlusion rate is 30%, the performance of SPP is lower than our method, but much higher than SRC. When the occlusion rate increases to 50%, the recognition rate of SPP is still higher than SRC. But the recognition rate of SPP is much lower than SRC when the occlusion rate is 70%. So we can conclude that SPP and SRC are sensitive to occlusion. With the increasing of the occlusion rate, the highest recognition rate that our method can achieve is dropped, but it is still much higher than the other two methods.

It can be seen from Figure 5 that our method performs much better than others and SPP's performance is superior to

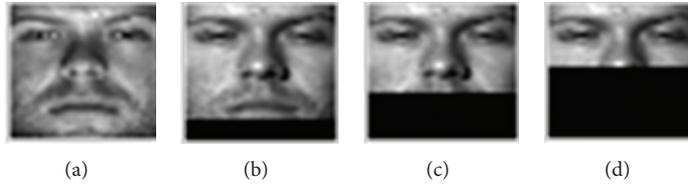


FIGURE 4: Sample images of Yale B database. (a) Training sample; (b) test image with 10% occlusion rate; (c) test image with 30% occlusion rate; (d) test image with 50% occlusion rate.

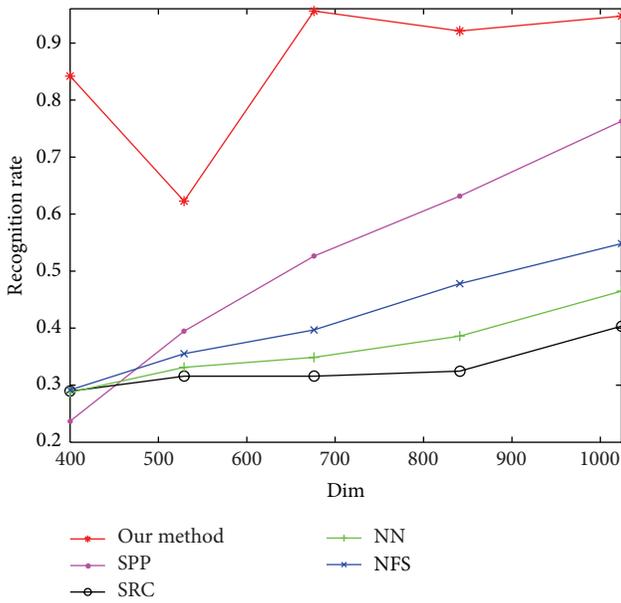


FIGURE 5: Recognition rates with 30% occlusion rate.

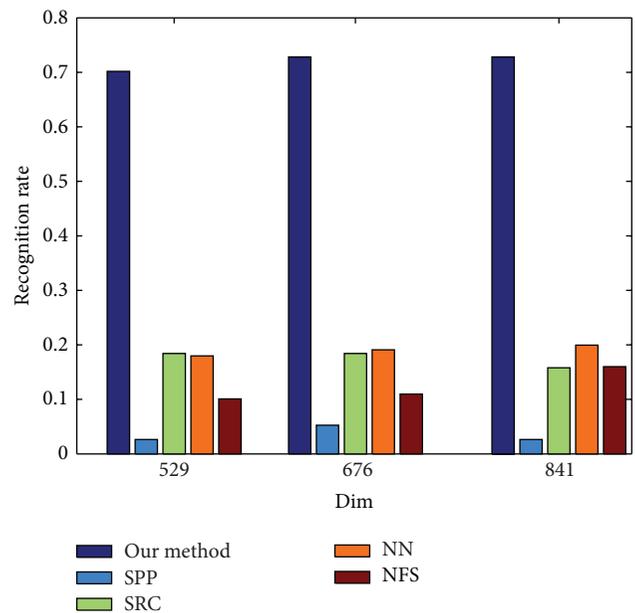


FIGURE 7: Recognition rate with 70% occlusion rate.

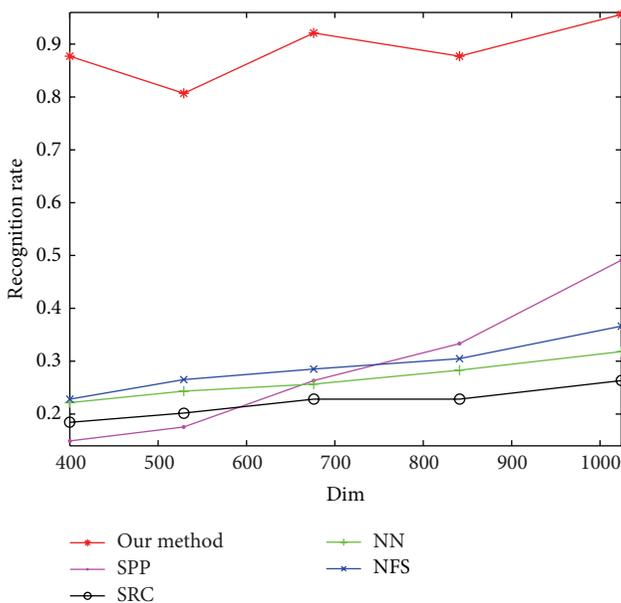


FIGURE 6: Recognition results with 50% occlusion rate.

SRC, NN, and NFS. NFS can obtain higher recognition rate than SRC and NN. So we can conclude that the global method SRC is not suitable for the image recognition with occlusion.

Even at 70% occlusion, the recognition rate of our method is still over 70%. The recognition rate of SRC and NN is about 20% which is much lower than our method. SPP has the poorest performance.

Based on the results shown in Figures 5–7 we can draw the conclusion that our method is more effective and robust to occlusion on the Yale B database.

The run time of our method, SPP, and SRC is shown in Figure 8. It notes that SRC costs the least time. When the dimension is less than 600, our method and SPP have similar running time. With the increase of dimensions, the running time of our method increases considerably. For our method partitioning image twice, although the low dimension makes the subblock run faster, the total run time is twice compared to SRC. But in the case of no special requirements for running time, our method has strong desirability for it can obtain higher recognition rate and be robust to occlusion.

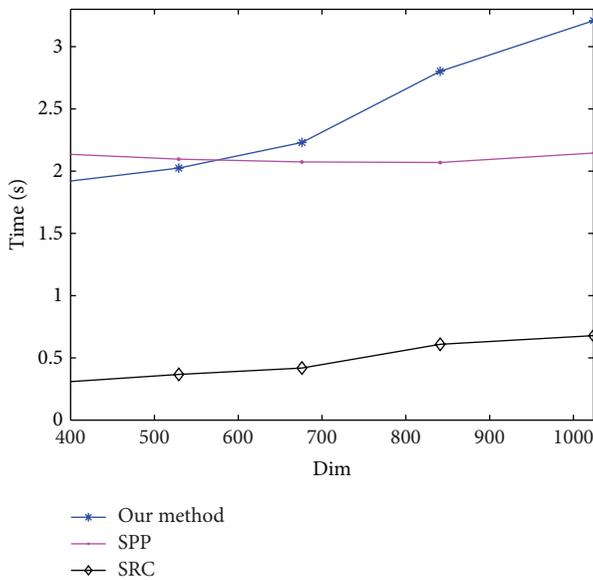


FIGURE 8: Run time of different methods.

5. Conclusion

To perform the identification of occluded image using unoccluded training samples, this paper proposed a new method to detect occlusion. FR_DLMSD partitions the image into 4 blocks and 8 modules in two layers. The sparsity difference of modules and its corresponding block is used for occlusion detection. The modules with negative difference are viewed as occlusion. The final identification is performed on the unoccluded modules by sparse representation. Experimental results on real-world face image data sets show that our method outperforms the face recognition methods, such as SPP, SRC, NN, and NFS in terms of recognition precision and robustness. So our method has certain value. In the future work, we will pay more attention to the improvement of our method on running time and recognition rate.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [2] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 471–478, Barcelona, Spain, November 2011.
- [3] Y. Xu, W. Zuo, and Z. Fan, "Supervised sparse representation method with a heuristic strategy and face recognition experiments," *Neurocomputing*, vol. 79, pp. 125–131, 2012.
- [4] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.
- [5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a practical face recognition system: robust registration and illumination by sparse representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 597–604, Miami, Fla, USA, 2009.
- [6] M. Yang, L. Zhang, and D. Zhang, "Efficient misalignment-robust representation for real-time face recognition," in *Proceedings of the European Conference on Computer Vision (ECCV '12)*, pp. 850–863, Firenze, Italy, 2012.
- [7] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, pp. 448–461, Heraklion, Greece, 2010.
- [8] C. Chen, C. Wei, and Y. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2618–2625, June 2012.
- [9] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 625–632, Providence, RI, USA, June 2011.
- [10] J. Lai and X. Jiang, "Modular weighted global sparse representation for robust face recognition," *IEEE Signal Processing Letters*, vol. 19, no. 9, pp. 571–574, 2012.
- [11] Z. Chen, T. Xu, and Z. Han, "Occluded face recognition based on the improved SVM and block weighted LBP," in *Proceeding of the International Conference on Image Analysis and Signal Processing (IASP '11)*, pp. 118–122, Hubei, China, October 2011.
- [12] J. S.-C. Yuk, K.-Y. K. Wong, and R. H.-Y. Chung, "A multi-level supporting scheme for face recognition under partial occlusions and disguise," in *Proceedings of the Asian Conference on Computer Vision (ACCV '10)*, pp. 690–701, Queenstown, New Zealand, 2010.
- [13] A. M. Andres, S. Padovani, M. Tepper, M. Mejail, and J. Jacobo, "Randomized face recognition on partially occluded images," in *Progress in Pattern Recognition, Image Analysis and Applications (CIARP)*, Lecture Notes in Computer Science, pp. 90–97, 2012.
- [14] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439–443, 1999.
- [15] J. Chien and C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1644–1649, 2002.
- [16] J. Wang, C. Lu, M. Wang et al., "Robust face recognition via adaptive sparse representation," *IEEE Transactions on Cybernetics*, 2014.

