Learning a Sparse Representation for Robust Face Recognition

Weihua Ou[1,2], Xinge You[1], Pengyue Zhang[1], Xiubao Jiang[1], Ziqi Zhu[1], and Duanquan Xu[1]

[1] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
{ouweihuahust,penryzhang,jiangxiubao}@gmail.com, youxg@mail.hust.edu.cn, xuduanquan@126.com, ziqi_zhu@163.com
[2] Department of mathematics, Huaihua University, Huaihua 418008, China

Abstract. Based on the assumption that occlusions have sparse representation on the nature pixel coordinate, Sparse Representation based Classification (SRC) [9] adopts an identity matrix as occlusion dictionary to deal with the occlusions or noises. However, this assumption is often violated in real applications, such as the faces are occluded by scarf. In this paper, we present an approach to learn an occlusion dictionary from the data. Thus, the occlusions have sparse representation on the learned occlusion dictionary and can be effectively separated from the occluded face images. Experimental results show our approach achieves better performance than SRC, while the computational cost is much lower.

Keywords: Face recognition, occlusions, dictionary learning, learning sparse representation.

1 Introduction

Though current face recognition techniques have reached a certain level of maturity in controlled settings, the complex intra-class variations, such as pose, illumination, expression and occlusions, are difficult to model and lead to recognition failures. Among them, occlusions are one of the most challenging problems. In real applications, the use of accessories, such as sunglasses, scarves, hats, or objects placed in front of the face can be viewed as occlusions. Moreover, violations of an assumed model for face appearance may act like occlusions, e.g., shadows due to extreme illumination violate the assumption of a low-dimensional linear illumination model [4]. Many methods are proposed to deal with occlusions, such as, localized non-negative matrix factorization [6], Local Binary Patterns [2] and Gabor wavelets [7]; however, these methods only operate on the non-occluded regions, and aim to circumvent the occluded regions, rather than to recover the occluded parts of face image, which might be essential for recognition.

Unlike the above methods, Sparse Representation based on Classification (SRC) [9] proposed by Wright et al. aims to eliminate the occlusions. Based on the assumption that occlusions have sparse representation on nature pixel coordinate, SRC adopts an identity matrix as the occlusion dictionary, and seeks the sparse representation over the

M. Lee et al. (Eds.): ICONIP 2013, Part III, LNCS 8228, pp. 233–240, 2013.

expanded dictionary which consists of the training sample dictionary and the occlusion dictionary. If the sparse representation is recovered correctly, the occlusion will be effectively eliminated and classification can be performed based on the reconstruction error only using the sparse representation coefficients over the training sample dictionary. The experiments shows that SRC has achieved the best performance for random noises. However, the experiments on the AR database also shows that, SRC is not nearly as robust to contiguous occlusion as it is to random pixel corruption. For sunglasses and scarf, it achieves only 87% and 59.5% respectively. The main reasons of this is that the assumption is violated and the representation is not sparse.

In this paper, we learn an occlusion dictionary from data to obtain the sparse representation and conduct recognition based on the learned structural sparse representation, which we call SSRC. First, we model the occluded faces as the summation of the non-occluded faces and the occlusions. Then we present a fast algorithm to learn an occlusion dictionary from data. Compared to the identity matrix occlusion dictionary used in SRC, the occlusions can be sparsely represented by the prototype of occlusion atoms. At the same time, the size of the expanded dictionary is significantly reduced with respect to the identity matrix occlusion dictionary in SRC. This will accelerate the speed of sparse coding for each test sample. Fig. 1 presents the illustration of the proposed

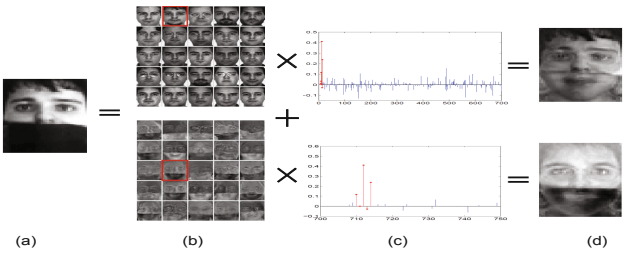

Fig. 1. Illustration of SSRC: an occluded face image (a) is represented as a sparse linear combination of the training sample dictionary and the occlusion dictionary in (b). The decomposed sparse coefficients in (c) correspond to the dictionary. The non-occluded face image and the occlusion in (d) can be jointly estimated.

approach. The lower images in Fig. 1(b) and 1(c) show the learned occlusion dictionary and the coefficients on this learned occlusion dictionary. It can clearly be seen that the atoms of the learned occlusion dictionary closely resemble the occlusion by scarf, and the coefficients are very sparse. As shown in the top image of Fig. 1(d), the occlusions are successfully separated and the recovered non-occluded face image is perfect except for the edge of the occlusion. Thus, the classification can be efficiently conducted on the recovered non-occluded face image.

The remainder of the paper is organized as follows. In Section 2, we describe the occlusion dictionary learning algorithm. In Section 3, we present the recognition algorithm based on the learned sparse representation. Finally, we conduct experiments in Section 4 and conclude this paper in Section 5.

2 Learning Sparse Representation

2.1 The Model of Occlusions

We denote the training samples of the i-th class as $\mathbf{A}_i = [\boldsymbol{s}_{i,1}, \boldsymbol{s}_{i,2}, \cdots, \boldsymbol{s}_{i,p_i}] \in \mathbb{R}^{m \times p_i}$ and each column of the matrix $\mathbf{A}_i$ denotes a sample. Suppose we have k classes and gather the training samples of all classes to build the training sample dictionary $\mathbf{A}_0 = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k] \in \mathbb{R}^{m \times p}$, where $p = p_1 + p_2 + \cdots + p_k$ is the number of training samples and p_i is the training sample number of i-th class. For the occluded face $\boldsymbol{y}$, we model it as follows:

$$\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{y}_d + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$, $\boldsymbol{y}_0$, $\boldsymbol{y}_d$ denotes the occluded face, the non-occluded face, and the occlusions, respectively, $\boldsymbol{e}$ is an error term that compensates the noise. Based on the assumption that the training samples from a single class lie on a subspace, $\boldsymbol{y}_0$ can be represented sparsely over the training sample dictionary $\mathbf{A}_0$, i.e., $\boldsymbol{y}_0 = \mathbf{A}_0\boldsymbol{\alpha}_0$, where $\boldsymbol{\alpha}_0$ is the sparse representation coefficients. In the next subsection, we present an approach to learn an occlusion dictionary, which can sparsely represent $\boldsymbol{y}_d$.

2.2 Learning Occlusion Dictionary

Given a data set $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n] \in \mathbb{R}^{m \times n}$ for occlusion dictionary learning, we first project them onto the corresponding class and utilize the associated projection residuals $\mathbf{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n] \in \mathbb{R}^{m \times n}$ to train the occlusion dictionary. For each sample $\boldsymbol{y}_r, r = 1, 2, \cdots, n$, suppose it belongs to class i, the projection residual is $\boldsymbol{p}_r = \boldsymbol{y}_r - \mathbf{A}_i(\mathbf{A}_i^T\mathbf{A}_i)^{-1}(\mathbf{A}_i^T\boldsymbol{y}_r)$, where $\mathbf{A}_i$ is the training sample of class i in $\mathbf{A}_0$.

We denote the occlusion dictionary as $\mathbf{A}_d = [\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_q] \in \mathbb{R}^{m \times q}$, where each atom is of unit length, i.e., $\boldsymbol{d}_j^T\boldsymbol{d}_j = 1, j = 1, 2, \cdots, q$. According to above discussions, we expect that $\mathbf{A}_d$ can sparsely represent $\boldsymbol{y}_d$ associated with the occlusion part, and at the same time "bad" for the non-occluded face image $\boldsymbol{y}_0$. We formulate the objective function for learning occlusion dictionary as follows:

$$\begin{aligned} &\min_{\mathbf{A}_d, \mathbf{\Lambda}} \|\mathbf{P} - \mathbf{A}_d\mathbf{\Lambda}\|_F^2 + \lambda_1\|\mathbf{\Lambda}\|_1 + \lambda_2\|\mathbf{A}_0^T\mathbf{A}_d\|_F^2 \\ &s.t. \quad \boldsymbol{d}_j^T\boldsymbol{d}_j = 1, \quad j = 1, 2, \cdots, q, \end{aligned} \tag{2}$$

where $\mathbf{A}_0$ is the training sample dictionary, $\mathbf{\Lambda} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n] \in \mathbb{R}^{q \times n}$ contains sparse coefficients of the projection residuals $\mathbf{P}$ on dictionary $\mathbf{A}_d$, λ_1 and λ_2 are the regularization parameters. The regularization term $\|\mathbf{A}_0^T\mathbf{A}_d\|_F^2$ encourages the incoherence between $\mathbf{A}_0$ and $\mathbf{A}_d$, and thus the learned occlusion dictionary $\mathbf{A}_d$ prefers to be as independent as possible to the training sample dictionary $\mathbf{A}_0$. As a result, the non-occluded face image can be efficiently obtained by seeking the sparse representation of the test face image on the expanded dictionary, i.e., $\mathbf{A} = [\mathbf{A}_0, \mathbf{A}_d]$.

Formulation.(2) is a joint optimization problem of the occlusion dictionary $\mathbf{A}_d$ and sparse coefficients $\mathbf{\Lambda}$. Although it is not jointly convex on both, it is convex on $\mathbf{A}_d$ (or $\mathbf{\Lambda}$) given fixed $\mathbf{\Lambda}$ (or $\mathbf{A}_d$). Similar to dictionary learning algorithm [1], we optimize

$\mathbf{A}_d$ and $\mathbf{\Lambda}$ alternatively, i.e., we optimize $\mathbf{A}_d$ given fixed $\mathbf{\Lambda}$, and then optimize $\mathbf{\Lambda}$ given fixed $\mathbf{A}_d$. The two steps are conducted iteratively until convergence. The whole algorithm is shown below.

Algorithm 1. Occlusion Dictionary Learning

Input: Training data $\mathbf{Y}$, training sample dictionary $\mathbf{A}_0$, λ_1, λ_2
Output: The occlusion dictionary $\mathbf{A}_d$ and the sparse coefficients $\mathbf{\Lambda}$
Initialization:We initialize each column of $\mathbf{A}_d$ as a random vector with unit l_2-norm
Step 1: Compute the projection residuals $\mathbf{P}$
Step 2: Fix $\mathbf{A}_d$ and optimize $\mathbf{\Lambda}$

$$\min_{\mathbf{\Lambda}} \|\mathbf{P} - \mathbf{A}_d\mathbf{\Lambda}\|_F^2 + \lambda_1\|\mathbf{\Lambda}\|_1 \tag{3}$$

Step 3: Fix $\mathbf{\Lambda}$ and optimize $\mathbf{A}_d$. We update each atom $\boldsymbol{d}_l$ of the dictionary $\mathbf{A}_d$ separately with all the other atoms $\boldsymbol{d}_{j\neq l}$ fixed, sweep through the columns and always use the most updated atoms that emerge from the preceding step. The update rule is given by

$$\begin{aligned} \boldsymbol{d}_l &= \left[(\boldsymbol{\beta}_l\boldsymbol{\beta}_l^T - \gamma)\mathbf{I} + \lambda_2\mathbf{A}_0\mathbf{A}_0^T\right]^{-1}\mathbf{Z}\boldsymbol{\beta}_l^T \\ \boldsymbol{d}_l &= \boldsymbol{d}_l/\|\boldsymbol{d}_l\|_2 \end{aligned} \tag{4}$$

Conduct steps 2 and 3 iteratively until the maximum number of iterations is reached or the values of the adjacent objective functions are sufficiently close.

3 Face Recognition via the Learned Sparse Representation

By concatenating the learned occlusion dictionary with the training sample dictionary, we obtain a structured dictionary $\mathbf{A} = [\mathbf{A}_0, \mathbf{A}_d]$. Given a test sample$\boldsymbol{y}$, we formulate the structured sparse recovery problem as follows:

$$\{\hat{\boldsymbol{\alpha}}_0, \hat{\boldsymbol{\alpha}}_d\} = \arg\min_{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_d} \|\boldsymbol{y} - \mathbf{A}_0\boldsymbol{\alpha}_0 - \mathbf{A}_d\boldsymbol{\alpha}_d\|_2^2 + \xi_1\|\boldsymbol{\alpha}_0\|_1 + \xi_2\|\boldsymbol{\alpha}_d\|_1. \tag{5}$$

We use the l_1-l_s [5] to solve it. Once the sparse solution is computed, denote the recovered face as $\hat{\boldsymbol{y}}_0 = \boldsymbol{y} - \mathbf{A}_d\hat{\boldsymbol{\alpha}}_d$, then the reconstruction error is computed with respect to the recovered face as follows:

$$r_i(\boldsymbol{y}) = \|\hat{\boldsymbol{y}}_0 - \mathbf{A}_0\delta_i(\hat{\boldsymbol{\alpha}}_0)\|_2, \quad i = 1, 2, \cdots, k. \tag{6}$$

Finally, the identity is the class corresponding to the minimum reconstruction error. The whole procedure is presented in algorithm 2.

Algorithm 2. Structured Sparse Representation based Classification (SSRC)

Input: Test sample $\boldsymbol{y} \in \mathbb{R}^m$, ξ_1, ξ_2
Output: Identity of test sample $\boldsymbol{y}$
Initialization: Training sample dictionary $\mathbf{A}_0$ and the learned occlusion dictionary $\mathbf{A}_d$
Step 1: Compute the sparse representation of $\boldsymbol{y}$ on the structured dictionary $\mathbf{A}$

$$\{\hat{\boldsymbol{\alpha}}_0, \hat{\boldsymbol{\alpha}}_d\} = \arg\min_{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_d} \|\boldsymbol{y} - \mathbf{A}_0\boldsymbol{\alpha}_0 - \mathbf{A}_d\boldsymbol{\alpha}_d\|_2^2 + \xi_1\|\boldsymbol{\alpha}_0\|_1 + \xi_2\|\boldsymbol{\alpha}_d\|_1$$

Step 2: Compute the residuals

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - \mathbf{A}_d\hat{\boldsymbol{\alpha}}_d - \mathbf{A}_0\delta_i(\hat{\boldsymbol{\alpha}}_0)\|_2, \quad i = 1, 2, \cdots, k \tag{7}$$

Step 3: Output the identity: $\text{Identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

4 Experiments

In this section, we conduct experiments to evaluate the performance of SSRC. We compare to following algorithms: (1) sparse representation based classification (SRC) [9], in which an identity matrix is utilized as the occlusion dictionary; (2) Gabor feature based sparse representation for face recognition (GSRC) [8], in which a Gabor occlusion dictionary is learned; (3) robust sparse coding for face recognition (RSC) [10], which needs several iterations of sparse coding for each test sample; and (4) Extended SRC: undersampled face recognition via intra-class variant dictionary(ESRC) [3], in which an intra-class variant dictionary is constructed by subtracting the class centroid of images from the same class. For SSRC, we learn one occlusion dictionary with the mutual incoherence regularization term and the other occlusion dictionary without. We denote them as SSRC1 and SSRC2, respectively.

4.1 Datasets and Parameter Setting

For SRC and ESRC, we implement the error-constrained model with the same error tolerance used in the original paper [9,3], i.e., $\varepsilon = 0.05$. For RSC and GSRC, we use the programs provided by the authors[1]. Since SSRC performed stably for wide ranges of model parameters, we set $\lambda_1 = 0.02$, $\lambda_2 = 0.5$, and $\xi_1 = \xi_2 = 0.001$. For SSRC, the occlusion dictionary size q is set to 50.

We conduct experiments on the AR database by choosing a subset with 50 men and 50 women. We resize each image into different resolutions 12×10, 26×20, 31×24, 42×30, and 51×40, which correspond to feature dimensions 120, 520, 744, 1,260 and 2,040, respectively. We consider following three scenarios.

Sunglasses. In this scenario, one image with sunglasses in session one for each of the 100 subjects is randomly chosen as a training sample for training the occlusion dictionary, while the test set consists of seven images without occlusion in session two and

[1] `http://www4.comp.polyu.edu.hk/~cslzhang/code.htm`

the remaining images with sunglasses in both sessions for each subject. In total, we have twelve test images for each subject.

Scarf. Similar to the sunglasses scenario, one image with scarf in session one for each of the 100 subjects is randomly chosen for training the occlusion dictionary, and the remaining images with scarf in both sessions and the seven images without occlusion in session two for each subject are used for testing.

Sunglasses+Scarf. The third scenario is that two occluded images (one with sunglasses and one with scarf) for each of the 100 subjects are randomly chosen from the session one are used for training the occlusion dictionary, and the rest seventeen images for each subject are used for testing. The challenge of this case is that there are two kinds of disguise with illumination variations, expression variations and whether the faces are occluded or not is unknown to the algorithm. In all three scenarios, the training sample dictionary is the same and consists of seven images without occlusion in session one.

4.2 Recognition Rate

The recognition results are shown in Fig.2. It shows that SSRC1 and SSRC2 greatly outperform SRC, RSC and GSRC. In these three scenarios, SSRC1 achieves the maximal recognition rate at 92.99%, 92.74%, 92.59% and outperforms SRC by 12.01%, 26.35%, and 27.47%, respectively. Both SRC and GSRC have a much worse recognition rate in dealing with "Sunglasses + Scarf", while RSC performs better than SRC and GSRC in almost all dimensions.

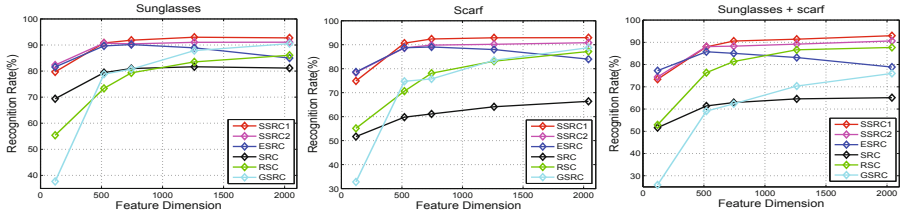

Fig. 2. Recognition rates of different methods on a subset of the AR database with the feature dimensions varying from 120 to 2,040: (a) Sunglasses, (b) Scarf, (c) Sunglasses +scarf

4.3 Comparison for Representation Coefficients

The assumption in SRC is that the occlusions can be sparsely coded by the identity matrix occlusion dictionary. This assumption might be violated when there are severe occlusions. Fig. 3 shows such an example, in which roughly 40% of the whole face is occluded by the scarves. It is obvious that the occlusion of scarves can not be sparsely represented by the identity matrix occlusion dictionary as shown in Fig. 3(c). The reconstructed face using SRC is completely different to the original face, as presented in the middle subfigure of Fig. 3(a); severe shadows are found in the mouth area and the

outline of the forehead is severely distorted, which influence classification correctness. As presented in Fig. 3(d), the red bar indicates that SRC fails to identify the subject. However, our learned occlusion dictionary represents the scarf occlusion sparsely as shown in Fig. 3(o), many of the coefficients are zeros. Obviously, the reconstructed face image using SSRC1 is almost perfect except for a light scarf mark which does not obscure the true identity. The green bar in Fig. 3(p) shows the correct class. Despite ESRC and SSRC2 identify the subject correctly, the reconstruction residuals is larger than that of SSRC1, which means that they are less robust than SSRC1.

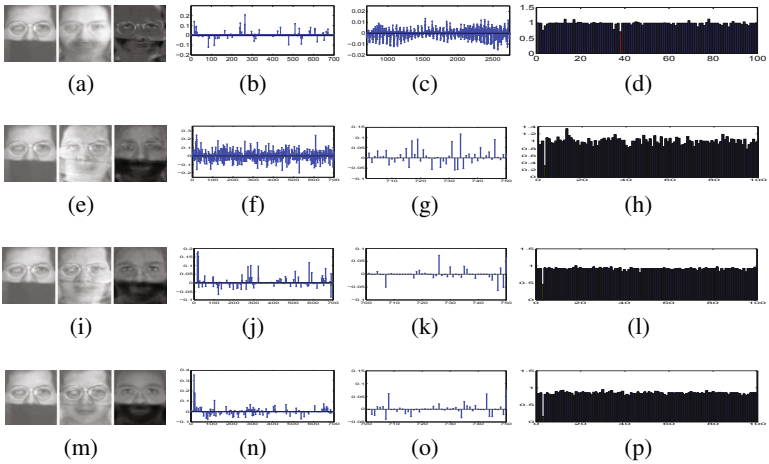

Fig. 3. Comparison between SRC, ESRC, SSRC1 and SSRC2 on the AR database with scarf: the upper row is the SRC result, the second row is the ESRC result, the third row is the SSRC2 result, and the lower row is the SSRC1 result. (a)(e)(i)(m) A test face image from subject 3 in the AR database: the left, middle and right subfigures correspond to the occluded image, reconstructed image and reconstruction error image, respectively. (b)(f)(j)(n) Sparse coefficients associated with the training sample dictionary. (c)(g)(k)(o) Sparse coefficients associated with the occlusion dictionary. (d)(h)(l)(p) Reconstruction residuals with respect to the coefficients for different classes; the green bar indicates the correct class.

5 Conclusion

In this paper, we present an approach to learn a sparse representation for robust face recognition. A occlusion dictionary is learned by mutual incoherence regularization. The learned occlusions dictionary can sparsely represent the occluded parts of faces. Apart from the improved recognition rate, an important advantage of SSRC is its compact occlusion dictionary, which has many fewer atoms than used in SRC [9]. This greatly speed the sparse coding. We evaluate the proposed method on different scenarios, including extreme variations of illumination, expressions and real disguises. The experimental results clearly demonstrate the proposed method achieves better performance than existing sparse representation methods.

Acknowledgements. This work was supported partially by the National Natural Science Foundation of China (Grant No.61272203), the International Scientific and Technological Cooperation Project (Grant No. 2011DFA12180), National Science & Technology R&D Program (Grant No.2012BAK31G01, Grant No. 2012BAK02B06), the Ph.D Programs Foundation of Ministry of Education of China (Grant No. 20110142110060), and the Key Science and Technology Program of Wuhan (Grant No.201210121021).

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54(11), 4311–4322 (2006)
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on PAMI 28(12), 2037–2041 (2006)
3. Deng, W., Hu, J., Guo, J.: Extended src: Undersampled face recognition via intraclass variant dictionary. IEEE Transactions on PAMI 34(9), 1864–1870 (2012)
4. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on PAMI 23(6), 643–660 (2001)
5. Kim, S., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing 1(4), 606–617 (2007)
6. Lee, D., Seung, H., et al.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791 (1999)
7. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
8. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 448–461. Springer, Heidelberg (2010)
9. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on PAMI 31(2), 210–227 (2009)
10. Yang, M., Zhang, L., Yang, J., Zhang, D.: Robust sparse coding for face recognition. In: Proc. of CVPR, pp. 625–632. IEEE (2011)