

# A face recognition algorithm based on collaborative representation



Zhengming Li<sup>a,b,\*</sup>, Tong Zhan<sup>a</sup>, Binglei Xie<sup>c</sup>, Jian Cao<sup>b</sup>, Jianxiong Zhang<sup>a</sup>

<sup>a</sup> Guangdong Industry Training Center, Guangdong Polytechnic Normal University, Guangzhou, China

<sup>b</sup> Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, China

<sup>c</sup> Shenzhen Key Laboratory of Urban Planning and Decision-Making Simulation, Shenzhen, China

## ARTICLE INFO

### Article history:

Received 3 September 2013

Accepted 5 April 2014

### Keywords:

Face recognition

Sparse coding

Neighbor matrix

Collaborative representation

## ABSTRACT

In this paper, we propose a face recognition algorithm by incorporating a neighbor matrix into the objective function of sparse coding. We first calculate the neighbor matrix between the test sample and each training sample by using the revised reconstruction error of each class. Specifically, the revised reconstruction error (RRE) of each class is the division of the  $l_2$ -norm of reconstruction error to the  $l_2$ -norm of reconstruction coefficients, which can be used to increase the discrimination information for classification. Then we use the neighbor matrix and all the training samples to linearly represent the test sample. Thus, our algorithm can preserve locality and similarity information of sparse coding. The experimental results show that our algorithm achieves better performance than four previous algorithms on three face databases.

© 2014 Elsevier GmbH. All rights reserved.

## 1. Introduction

In the last decade, a number of sparse representation based classification (SRC) algorithms have been proposed for face recognition. The basic idea is that the test sample can be represented as a linear combination of all the training samples with sparsity constraint, and then can be classified by exploiting the reconstruction error. Huang [1] presents a theoretical framework for signal classification with sparse representation, which sparsely codes a signal over a set of redundant bases and classifies the signal based on its coding vector. Because to minimize the  $l_0$ -norm is an NP hard problem, we usually formulate the sparse coding problem as the minimization of the  $l_1$ -norm of the reconstruction coefficients or minimization of the  $l_2$ -norm of the reconstruction coefficients.

In the past several years, many face recognition algorithms based on the minimization of the  $l_1$ -norm of the reconstruction coefficients have been proposed. For example, Wright [2] uses sparse representation for robust face recognition. A test image is first sparsely coded over the template images, and then the classification is performed by checking which class yields the least coding error. Moreover, there are many variations of the

SRC. Hui [3] exploits a  $k$ -nearest neighbor (KNN) method to classify a test sample using sparse representation, and which can reduce the computational complexity. Kang [4] presents a kernel sparse representation classification framework and utilizes the local binary pattern descriptor in the framework for robust face recognition. Mairal [5] proposes a joint dictionary learning and classifier construction framework. Deng [6] proposes an extended sparse representation-based classifier (ESRC) algorithm, and applies an auxiliary intra class variant dictionary to represent the possible variation between the training and testing images. Gabor features [7] and Markov random fields [8] are also used to further improve the accuracy of SRC. In addition, Ji [9] proposes an improved sparse representation classification algorithm based on non-negative constraint of sparse coefficient. Other nonnegative sparse representation algorithms can be found in Refs. [10–12]. Although SRC and its variations significantly improve the robustness of face recognition, they still need to solve the  $l_1$ -minimization problem on the whole dataset, which makes the computation expensive for large-scale datasets. Yang [13] presents a review of iterative shrinkage-threshold based sparse representation methods for robust face recognition. More sparse representation for computer vision and pattern recognition applications can be found in Ref. [14].

Recently, the collaboration representation used in SRC has shown very powerful classification capability, which is based on minimization of the  $l_2$ -norm of the reconstruction coefficients. Zhang [15] analyzes the working mechanism of sparse representation based classification (CRC), and indicates that it is the

\* Corresponding author at: Guangdong Industry Training Center, Guangdong Polytechnic Normal University, Guangzhou 510665, China. Tel.: +86 20 38256428; fax: +86 20 38256428.

E-mail address: [lizhengming2004@126.com](mailto:lizhengming2004@126.com) (Z. Li).

collaborative representation but not the  $l_1$ -norm sparsity that makes SRC powerful for face classification. Therefore, many face recognition algorithms based on collaborative representation have been proposed. For example, Lee [16] presents an efficient sparse coding algorithm that is based on iteratively solving the  $l_1$ -regularized least squares problem and the  $l_2$ -constrained least squares problem, which can significantly enhance the speed of sparse coding. Huang [17] proposes a face recognition algorithm based on collaborative image similarity assessment. Moreover, several variants of collaborative representation have been proposed in recent years by adding some additional regularization and/or constraints. Jadoon [18] proposes a collaborative neighbor representation algorithm for multi-class classification based on the  $l_2$ -minimization approach with the assumption of locally linear embedding. Naseem [19] presents a linear regression classification (LRC) algorithm by formulating the pattern recognition problem as a linear regression problem. Yang [20] proposes a regularized robust coding (RRC) model, which could robustly regress a given signal with regularized regression coefficients. Moreover, the recently proposed two-phase test sample representation algorithm uses a novel representation based classification algorithm to perform face recognition (TPTSR) [21,22]. He [23] proposes a two-stage sparse representation (TSR) for robust face recognition on a large-scale database. Based on the divide and conquer strategy, TSR decomposes the procedure of robust face recognition into outlier detection stage and recognition stage.

However, in sparse coding, local features are dealt separately. The mutual dependence among local features is ignored, which results in the sparse codes may vary a lot even for similar features. To overcome this drawback, a number of face recognition algorithms based on local features have been proposed. Chen [24] proposes a nonnegative local coordinate factorization (NLCF) for feature extraction. NLCF adds a local coordinate constraint into the standard NMF objective function. In this way, each data point can be represented by a linear combination of only a few nearby basis vectors, which leads to sparse representation. Yu [25] assumes that each data point can be locally approximated by a linear combination of its nearby anchor points, and the linear weights become its local coordinate coding. Wang [26] utilizes the locality constraints to project each descriptor into its local-coordinate system, and the projected coordinates are integrated by max pooling to generate the final representation. Arpit [27] imposes a locality constraint to choose the training samples that are in the vicinity of the test sample. Chao [28] uses both group sparsity and data locality to formulate a unified optimization framework, which produces a locality and group sensitive sparse representation (LGSR) for improved recognition. Gao [29] incorporates Laplacian matrix into the objective function of sparse coding to preserve the consistence in sparse representation of similar local features. Those algorithms demonstrate that locally information can improve the performance of face recognition.

In this paper, we propose a face recognition algorithm based on the neighbor matrix, which can preserve locality and similarity information of sparse coding. The main idea is to calculate the neighbor matrix between the test sample and each training sample by using the revised reconstruction error of each class. Then we use the neighbor matrix and all the training samples to linearly represent the test sample. Furthermore, we propose a new optimization function that can preserve locality and similarity information of sparse coding. The experimental results show that our algorithm is very competitive on the ORL, YALEB and PIE face databases.

The remaining of this paper is organized as follows. Section 2 describes our proposed face recognition algorithm. Section 3 describes the neighbor coefficients assessment between images.

The experiments and performance evaluation are reported in Section 4. Finally, the conclusion is presented in Section 5.

## 2. The proposed algorithm

In this section we describe the proposed algorithm, we assume that there are  $C$  classes of training samples  $D = [x_1, \dots, x_n]^T$ ,  $n$  is the number of training samples. If a test sample  $y$  belongs to one of the labeled classes in the training samples set  $D$ , we can use all the training samples to represent the test sample  $y$ . Then, the linear representation of a test sample  $y$  can be written as:

$$y = \alpha_1 x_1 + \dots + \alpha_n x_n \quad (1)$$

where  $A = [\alpha_1, \dots, \alpha_n]$  is a coefficient vector. In general, the sparse representation based algorithms reconstruct a test sample using a sparse linear combination of training samples. But they do not consider the underlying neighbor relation between the test sample and each training sample. Therefore, we assume that the following equation is approximately satisfied:

$$y = m_1 x_1 \alpha_1 + \dots + m_n x_n \alpha_n \quad (2)$$

where  $P = [m_1, \dots, m_n]$  is the neighbor coefficient vector,  $m_i$  ( $i = 1, \dots, n$ ) represents that the neighbor relation between the  $i$ th training sample and the test sample. A bigger  $m_i$  means that the  $i$ th training sample is more close to the test sample. Eq. (2) also shows that if a training sample is far away from the test sample, it has smaller contribution to represent the test sample. Thus, the test sample can be represented better by using Eq. (2) than Eq. (1). Then, we can rewrite Eq. (2) as:

$$y = AMD \quad (3)$$

where  $M = \text{diag}(m_1, \dots, m_n)$  is the neighbor matrix whose only nonzero diagonal entries represents the neighbor coefficients between the test sample and each training samples. In general, in collaborative representation based classifier algorithm (CRC), we usually obtain the sparse solution of Eq. (1) by solving the following optimization function.

$$A = \arg \min_A \|y - AD\|_2^2 + \lambda \|A\|_2^2 \quad (4)$$

Therefore, the sparse solution of Eq. (2) can be

$$A = \arg \min_A \|y - AMD\|_2^2 + \lambda \|A\|_2^2 \quad (5)$$

where  $\lambda$  is a regularization parameter. In order to better preserve locality and similarity information of sparse coding, we incorporate the neighbor information into the objective function of Eq. (5). Therefore, we propose the following optimization function:

$$A = \arg \min_A \left\{ \frac{1}{2} \left( \|y - AMD\|_2^2 + \delta \sum_{i=1}^n \alpha_i^2 \beta_i^2 + \lambda \|A\|_2^2 \right) \right\} \quad (6)$$

where  $\beta_i = 1 - m_i$  ( $i = 1, \dots, n$ ), a smaller  $\beta_i$  means that the  $i$ th training sample is more close to the test sample,  $\delta$  and  $\lambda$  are regularization parameters having small positive values. The rationale of Eq. (6) is as follows: we first calculate the distances between the test sample and all the training samples. If the test sample is the same class as the training sample, then the distance will be small. It is reasonable to assume that the class close to the test sample has small distance, and has its own contribution to minimize the object function of sparse code. So, we incorporate the distance information into the objective function of sparse code, which can preserve locality and similarity information of sparse code. Furthermore, it can eliminate the side-effect on the classification decision of the

class that is far from the test sample. As a result, we may obtain higher accuracy. Then Eq. (6) can be rewrite as:

$$A = \operatorname{argmin}_A \left\{ \frac{1}{2} (\|y - AMD\|_2^2 + \delta AL^T + \lambda \|A\|_2^2) \right\} \quad (7)$$

where  $L = \operatorname{diag}(\beta_1^2, \dots, \beta_n^2)$ . The closed form solution of Eq. (7) can be derived by taking partial derivatives and setting them equal to zero as:

$$-yD^T M^T + AMDD^T M + \delta AL + \lambda A = 0 \quad (8)$$

The solution is the following equation:

$$A = yD^T M^T (MDD^T M + \delta L + \lambda I)^{-1} \quad (9)$$

Therefore, we get the collaborative neighbor coefficients vector  $A$ , which can better linearly reconstructed the test sample  $y$  from the neighborhood. In order to increase the discrimination information, we use the regularized reconstruction error (RRE) for classification [18].

$$RRE_k = \frac{\|y - D_k C_k\|_2^2}{\|C_k\|_2^2} (k \in [1, \dots, C]) \quad (10)$$

where  $C_k = [0, \dots, 0, a_{k_1}, \dots, a_{k_j}, 0, \dots, 0]$  is a coefficients vector, whose only nonzero entries are the coefficients in  $A$  that are corresponding to the  $k$ th class,  $j$  is the number of training samples in each class,  $y$  is the test sample,  $D_k$  is the training samples of  $k$ th class,  $RRE_k$  is the regularized reconstruction error of the  $k$ th class. In general, a lower value of  $RRE_k$  means that the  $k$ th class has a more important contribution for representing the test sample. Therefore, we use Eq. (11) to perform face recognition.

$$s = \operatorname{argmin}_i (RRE_i), (i = 1, \dots, C) \quad (11)$$

If the  $s$ th class has the lowest  $RRE$  value among of all the  $C$  classes, and then the test sample is classified to the  $s$ th class.

The main steps of our proposed algorithm can be summarized as follows:

- Step 1. Calculate the neighbor matrix  $M$  by using Eq. (15).
- Step 2. Use all the training samples and the neighbor matrix  $M$  to represent the test sample, and solve Eq. (6) by using Eq. (9).
- Step 3. Calculate the  $RRE$  value of each class using Eq. (10).
- Step 4. Identify the class label of the test sample using Eq. (11).

### 3. Neighbor coefficients assessment between images

To quantitatively measure the neighbor coefficient between the test sample and each training sample, we use the linear representation method to obtain the neighbor matrix. Firstly, the test sample  $y$  can be represented as a linear combination of the training images from the same class.

$$y = D_i \theta_i, \quad (i \in [1, \dots, C]) \quad (12)$$

where  $D_i$  is the training samples of  $i$ th class,  $D_i = [x_{i_1}, \dots, x_{i_j}]$ , ( $i \in [1, \dots, C]$ ),  $j$  is the number of training samples in each class,  $\theta_i = [\theta_{i_1}, \dots, \theta_{i_j}]$ , ( $i \in [1, \dots, C]$ ) is the representation coefficient vector of  $i$ th class, Then  $\theta_i$  can be obtained by solving Eq. (13).

$$\theta_i = (D_i^T D_i)^{-1} D_i^T y \quad (13)$$

Then the reconstructed image of the test sample is:

$$\tilde{y}_i = D_i \theta_i \quad (14)$$



Fig. 1. Some face images from the ORL face database.



Fig. 2. Some face images from the YALEB face database.

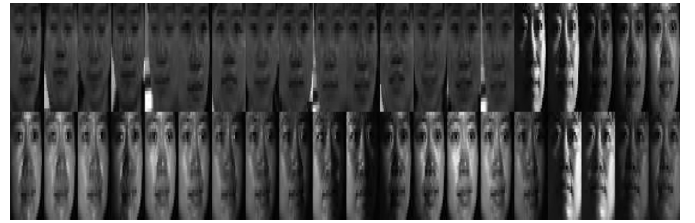


Fig. 3. Some face images from the PIE face database.

We calculate the regularized reconstruction error between the test sample and the  $i$ th class training samples.

$$d_i = \frac{\|y - \tilde{y}_i\|_2^2}{\|\theta_i\|_2^2}, \quad (i \in [1, \dots, C]) \quad (15)$$

If the training samples are belongs to the  $i$ th class, then we use the  $m_i = 1 - d_i$  to represent the neighbor coefficient between the test sample with the each training sample of  $i$ th class. Thus, we can obtain the neighbor matrix  $M$ .

### 4. Experimental results

In order to test the efficiency of our algorithm presented above, we perform a series of experiments using the ORL [30], YALEB [31], and PIE [32] face databases. Sample images from the three face databases are displayed in Figs. 1–3.

The ORL face database contains 400 images of 40 persons (10 images per person). The resolution of ORL images is  $46 \times 56$ . If  $s$  samples of the  $n$  samples per class are used for training and the remaining samples are used for testing, there are  $C_n^s = \frac{n!}{s!(n-s)!}$  possible combinations. We use six images of each person as the training samples, and there are 210 training and test sample sets. The extended Yale Face Database B contains 2414 images of 38 persons and around 64 near frontal images under different illuminations per person. For this database, we simply use the cropped images and resize them to  $32 \times 32$  pixels. A random subset with 20 images per person was taken with labels to form the training set, and the rest of the database was considered to be the testing set.<sup>1</sup> There are 50 random splits in our experiments. The Multi PIE database contains 11554 images of 68 persons, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. For this

<sup>1</sup> <http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>.

**Table 1**

The average classification errors rates (%) of SRC, LRC, CRC, CNRC and our algorithm on the ORL face database.

| Dim           | 50   | 100  | 150  | 180  |
|---------------|------|------|------|------|
| SRC(11-ls)    | 5.11 | 4.90 | 5.00 | 4.95 |
| CRC           | 5.69 | 4.09 | 4.49 | 4.84 |
| LRC           | 3.29 | 3.24 | 3.44 | 3.56 |
| CNRC          | 4.41 | 3.67 | 3.62 | 3.12 |
| Our algorithm | 3.19 | 2.97 | 3.08 | 3.77 |

**Table 2**

The average classification errors rates (%) of SRC, LRC, CRC, CNRC and our algorithm on the YALEB face database.

| Dim           | 50    | 100   | 150   | 180  |
|---------------|-------|-------|-------|------|
| SRC(11-ls)    | 16.67 | 11.70 | 10.07 | 9.54 |
| CRC           | 21.20 | 10.05 | 7.29  | 6.31 |
| LRC           | 11.58 | 9.4   | 8.78  | 8.59 |
| CNRC          | 21.16 | 11.67 | 8.72  | 7.88 |
| Our algorithm | 10.59 | 6.81  | 5.73  | 5.39 |

**Table 3**

The average classification errors rates (%) of SRC, LRC, CRC, CNRC and our algorithm on the PIE face database.

| Dim           | 50    | 100   | 150   | 180   |
|---------------|-------|-------|-------|-------|
| SRC(11-ls)    | 47.36 | 42.07 | 40.27 | 39.70 |
| CRC           | 46.46 | 34.54 | 30.46 | 29.14 |
| LRC           | 53.80 | 50.85 | 49.91 | 49.62 |
| CNRC          | 48.50 | 39.43 | 36.04 | 34.98 |
| Our algorithm | 40.12 | 33.27 | 31.08 | 30.44 |

database, we simply use the cropped images and resize them to  $32 \times 32$  pixels. A random subset with 5 images per person is taken with labels to form the training set, and the rest of the database is considered to be the testing set.<sup>2</sup> There are 50 randomly splits in our experiments.

To show the superior performance of our algorithms, we use four state-of-the-art algorithms for comparison. They are the CNRC [18], SRC [2], LRC [19] and CRC [15]. Moreover, The Eigen face is used as the face feature in these experiments.<sup>3</sup>

Table 1 shows the average classification errors rates (%) of the SRC, LRC, CRC, CNRC and our algorithm on the ORL face database. Table 2 shows the average classification errors rates (%) of the SRC, LRC, CRC, CNRC and our algorithm on the YALEB face database.

Table 3 shows the average classification errors rates (%) of the SRC, LRC, CRC, CNRC and our algorithm on the PIE face database.

Tables 1–3 show that the average classification errors rates of our proposed algorithm is lower than SRC (11-ls), CNRC, LRC and CRC in most cases on the ORL, YALEB and PIE face databases. Most important of all, the experimental results show that if we use a suitable dimension of training samples to represent the test sample, the average classification errors rates of our algorithm can achieve 2.97%, 5.39% and 30.44% on the ORL, YALEB and PIE face databases, respectively. The experimental results show that the neighbor matrix can preserve locality and similarity information of sparse code. Furthermore, it can be used to improve the performance of the sparse coding, and increase the classification performance.

Moreover, we notice that as the dimensions of training samples increase, the improvement of our algorithm decreases compared with four comparison algorithms in Tables 1–3. The reason may be that the neighbor matrix can make similar features close to each other in the sparse codes space by reducing the variance. Those

results in the images belonging to the same class are finally represented similarly. As a result, the features of the same class are enhanced. This will facilitate the classification performance. As the dimension of training samples increases, the feature from the same class is dense enough. So the improvement compared with sparse coding without neighbor matrix becomes smaller.

## 5. Conclusion

This paper proposes a neighbor matrix based face recognition algorithm. A new optimization function is proposed, which can obtain a better linearly representation results for the test sample. Moreover, our proposed algorithm can preserve locality and similarity information of sparse code. Three standard face databases including the ORL, YALEB and PIE are selected to evaluate our proposed algorithm. Experimental results demonstrate that our proposed algorithm is very competitive. Furthermore, our algorithm is based on the  $l_2$ -norms and it is computationally more efficient than the  $l_1$ -norm based sparse representation algorithms.

## References

- [1] K. Huang, S. Aviyente, Sparse representation for signal classification, in: Conference on Neural Information Processing Systems, 2006, pp. 609–616.
- [2] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
- [3] K. Hui, C. Li, L. Zhang, Sparse neighbor representation for classification, Pattern Recognit. Lett. 33 (5) (2012) 661–669.
- [4] C. Kang, S. Liao, S. Xiang, C. Pan, Kernel sparse representation with local patterns for face recognition, in: IEEE International Conference on Image processing, 2011, pp. 3009–3012.
- [5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Discriminative learned dictionaries for local image analysis, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 24–26.
- [6] W. Deng, J. Hu, J. Guo, Extended SRC: under sampled face recognition via intra class variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
- [7] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, in: European Conference on Computer Vision, 2010, pp. 1–4.
- [8] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, Y. Ma, Face recognition with contiguous occlusion using Markov random fields, in: IEEE International Conference on Computer Vision, 2009, pp. 1050–1057.
- [9] Y. Ji, T. Lin, H. Zha, Mahalanobis distance based non-negative sparse representation for face recognition, in: 26th International Conference on Machine Learning, 2009, pp. 41–46.
- [10] R. He, W. Zheng, B. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1561–1576.
- [11] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 625–632.
- [12] N. Vo, B. Moran, S. Challa, Non-negative least-square classifier for face recognition, Lecture Notes Comput. Sci. Adv. Neural Netw. 5553 (2009) 449–456.
- [13] A.Y. Yang, S.S. Sastry, A. Ganesh, Y. Ma, Fast L1-minimization algorithms and an application in robust face recognition: a review, in: IEEE International Conference on Image processing, 2010, pp. 1849–1852.
- [14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.
- [15] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: IEEE International Conference on Computer Vision, 2011, pp. 471–478.
- [16] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: Conference on Neural Information Processing Systems, 2006, pp. 801–808.
- [17] D. Huang, T. Liu, W. Hu, M. Chen, Face recognition based on collaborative image similarity assessment, in: International Conference on Information Security and Intelligence Control, 2012, pp. 254–257.
- [18] W. Jadoon, Y. Zhang, L. Zhang, Collaborative neighbor representation based classification using  $l_2$ -minimization approach Pattern Recognit. Lett. 34 (2013) 201–208.
- [19] I. Naseem, R. Togner, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
- [20] M. Yang, L. Zhang, J. Yang, D. Zhang, Regularized robust coding for face recognition, IEEE Trans. Image Process. 22 (5) (2013) 1753–1766.
- [21] Y. Xu, D. Zhang, J. Yang, J. Yang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1255–1262.
- [22] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, Z. Lai, Using the idea of the sparse representation to perform coarse to fine face recognition, Inf. Sci. 238 (2013) 138–148.

<sup>2</sup> <http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>.

<sup>3</sup> <http://www4.comp.polyu.edu.hk/~cslzhang/>.



- [23] R. He, W. Zheng, B. Hu, X. Kong, Two-stage nonnegative sparse representation for large-scale face recognition, *IEEE Trans. Neural Netw. Learning Syst.* 24 (1) (2013) 35–46.
- [24] Y. Chen, J. Zhang, D. Cai, W. Liu, X. He, Non-negative local coordinate factorization for image representation, *IEEE Trans. Image Process.* 22 (3) (2013) 969–978.
- [25] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: *Conference on Neural Information Processing Systems*, 2009, pp. 2223–2231.
- [26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3360–3367.
- [27] D. Arpit, G. Srivastava, Y. Fu, Locality-constrained low rank coding for face recognition, in: *21st International Conference on Pattern Recognition*, 2012, pp. 1687–1690.
- [28] Y. Chao, Y. Yeh, Y. Chen, Y. Lee, Y. Wang, Locality-constrained group sparse representation for robust face recognition, in: *IEEE International Conference on Image Processing*, 2011, pp. 761–764.
- [29] S. Gao, I. Tsang, L. Chia, P. Zhao, Local features are not lonely – Laplacian sparse coding for image classification, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3555–3561.
- [30] Available at: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [31] A.S. Georgiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (6) (2001) 643–660.
- [32] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (2) (2003) 1615–1618.