Neural Process Lett
DOI 10.1007/s11063-013-9333-6

Face Recognition Via Weighted Two Phase Test Sample Sparse Representation

Zhonghua Liu · Jiexin Pu · Meiyu Xu · Yong Qiu

Abstract Sparse representation (SR) for signals over an overcomplete dictionary fascinates a lot of researchers in the past decade. The two-phase test sample sparse representation method (TPTSSR) achieved an excellent performance in face recognition. However, TPTSSR exploits the global information and tends to lose local information. In this paper, the weighted two phase test sample sparse representation method (WTPTSSR) is proposed. WTPTSSR utilizes both data locality and linearity and it can be regarded as extensions of TPTSSR. Experiments on the face databases demonstrate that WTPTSSR is more effective than TPTSSR.

Keywords Face recognition · Sparse representation · Local structure · Linear presentation

1 Introduction

Face recognition has become one of the most active branches of biometrics [1–4]. The face recognition technology has the advantages of noncontact and might disturb the on-going activity of the people. A large number of methods have been proposed for face recognition. The face recognition technology has been applied to many scenarios such as airport, border controlling and access control.

A face recognition system should implement the following procedures: face image capture, feature extraction and face classification. Different literatures have proposed various face recognition methods. For example, Classical dimensionality reduction methods include

Z. Liu (✉) · J. Pu · M. Xu · Y. Qiu
Information Engineering College, Henan University of Science and Technology,
Luoyang, China
e-mail: lzhua_217@163.com

Z. Liu
China Airborne Missile Academy, Luoyang, China

Z. Liu
School of Electronic Engineering, Xidian University, Xi’an, China

Published online: 29 November 2013

principal component analysis (PCA) [5], general tensor discriminant analysis [6], Manhattan NMF (MahNMF) [7], robust principal component analysis (RPCA) [8], NeNMF [9], et al. The representation based face recognition techniques [10–19] have recently been attracting attention. The methods can be categorized into two kinds: the L_1 norm based representation method (1NBRM) and the L_2 norm based representation method (2NBRM). The methods proposed in [10–14] are typical examples of 1NBRM. SRC [11] casted the recognition problem as one of the multiple linear reconstruction models and argued that new theory from sparse signal representation offers the key to address this problem. SRC is robust to occlusion, illumination and noise, and achieves excellent performance. Gao et al. [12] introduced a kernelized version of SRC. Yang et al. [13] proposed a robust sparse coding method, by modeling the sparse coding as a sparsity constrained robust regression problem, and it is robust to outliers. Lai et al. [14] proposed a sparse local discriminant projections method which imposes a sparse constraint on local discriminant projections, and requires that the test sample be sparsely represented by the training samples. The methods proposed in [15–19] are typical examples of 2NBRM. Xu et al. [17] proposed a two-phase sparse representation method (TPTSR) which makes coarse to fine classification decisions for the face samples and receives good performances. Compared with 1NBRM, 2NBRM has remarkable advantages. Firstly, 2NBRM does not need to iteratively compute the solution. Secondly, as it obtains the solution by solving a linear system consisting of linear equations, it is usually computationally efficient. Thirdly, to use the L_1 norm is not the sole way to obtain the sparsity. Actually, a proper use of the L_2 norm can also lead to sparse representation (SR) [17]. Moreover, 2NBRM might obtain a higher accuracy than 1NBRM. For example, Zhang et al. [15] and Shi et al. [16] also illustrated that 2NBRM outperforms the well-know 1NBRMs proposed in [10,11] in classification accuracy.

In pattern recognition, it has been proved that the local structure of data is critical [20–22]. Nearest neighbor (NN) utilizes the local structure of data, however, NN classifies the query image by only using its NN in the training data. Therefore, it can easily be affected by noise. SR method uses the linearlity structure of data. Yu et al. [23,24] pointed out that under certain assumptions locality is more essential than sparsity. But two-phase test sample sparse representation (TPTSSR) does not guarantee to be local which lead to unstable. That is, TPTSSR reconstructs a test sample by training samples and these training samples may include samples which are far from the test sample and thus produce unstable classification results. Thus TPTSSR tends to lose local information. In order to overcome this problem, an extension of TPTSSR named weighted two-phase test sample sparse representation (WTPTSSR) is proposed. The key idea of the proposed WTPTSSR algorithm is that both the local structure and sparsity are taken into consideration.

The rest of the paper is organized as follows: Sect. 2 introduces TPTSSR method. We propose WTPTSSR method and discuss the relationships between WTPTSSR and TPTSSR in Sect. 3. Section 4 demonstrates the detailed experimental results on face databases. The final section gives our conclusions.

2 Two-Phase Test Sample Representation (TPTSR) Method

A two phase test sample representation (TPTSR) method [19] is generally described as two steps. First, the TPTSSR method represents the test sample as a linear combination of all the training samples, and determines the M nearest neighbors for the test sample according to the training samples that are the best suited to represent the test sample. Then the TPTSSR method represents the test sample as a linear combination of all the M nearest neighbors and uses the representation result to perform classification. Suppose that we have N training samples for c classes and each class includes n training samples, $A_k = [a_{k,1}, \ldots, a_{k,n}] \in R^{m \times n} (1 \leq k \leq c)$, where m is the dimension of samples.

Step 1 Let y denote the test sample. Use $y = \frac{y}{||y||}$ and $a_{k,i} = \frac{a_{k,i}}{||a_{k,i}||}$ $(1 \leq i \leq n)$ to convert the test sample and training samples into unit vectors. It first assumes that the test sample y can be approximately represented as the linear combination of all training samples.

$$y = AX \ = a_{1,1}x_{1,1} + a_{1,2}x_{1,2} + \cdots + a_{c,n}x_{c,n} \tag{1}$$

where $A = [A_1, A_2, \ldots, A_c] = [a_{1,1}, a_{1,2}, \ldots, a_{c,n}]$ is a matrix composed of all the N training samples of c classes and $x_{i,j}(i = 1, \ldots, c, j = 1, \ldots, n)$ is the coefficient. The objective is to minimize $||y - AX||_2$. According to the Least Square method, X should be computed using Eq. (2).

$$\widehat{X} = \left(A^T A + \sigma I\right)^{-1} A^T y \tag{2}$$

where I and σ denote the identity matrix and a small positive constant, respectively. Equation (1) shows that every training sample makes its own contribution to representing the test sample. The contribution that the training sample $a_{i,j}$ makes is $a_{i,j}x_{i,j}(1 \leq i \leq c, 1 \leq j \leq n)$. We use the following equation to evaluate the contribution of the training samples $a_{i,j}$ in representing the test sample.

$$e_{i,j} = \left\| y - a_{i,j}x_{i,j} \right\|_2 \tag{3}$$

A small $e_{i,j}$ means a great contribution. Based on this fact, the M training samples that have the M greatest contributions are selected and denoted by $a_1, \ldots a_M (1 \leq M \leq N)$. We refer to these samples as the M nearest neighbors of the test sample.

Step 2 Suppose that test sample y can be approximately represented as the linear combination of the M training samples with great contribution in representing the test sample y:

$$y = \tilde{A}\tilde{X} = a_1x_1 + a_2x_2 + \cdots + a_Mx_M \tag{4}$$

where $\tilde{A} = [a_1, \ldots, a_M]$, $\tilde{X} = [x_1, \ldots, x_M]$. We can solve Eq. (4) by using

$$\breve{X} = \left(\tilde{A}^T \tilde{A} + \gamma I\right)^{-1} \tilde{A}^T y \tag{5}$$

where γ is a positive constant and I is the identity matrix.

Since the neighbors might be from different classes, we calculate the sum of the contribution to represent the test sample of the neighbors from each class and exploit the sum to classify the test sample. Suppose that there are t_i from the ith class and they are denoted by $a_1^i, \ldots, a_{t_i}^i$, and the corresponding coefficients are denoted by $x_1^i, \ldots, x_{t_i}^i$. We use the following equation to evaluate the contribution, of the training samples of the ith class, in representing the test representation.

$$con_i = \left|\left| y - \sum\nolimits_{j=1}^{t_i} a_j^i x_j^i \right|\right|_2 \tag{6}$$

A smaller deviation con_i means a greater contribution to representing the test sample. If $h = \arg\min_i con_i$, we will assign test sample y into the hth class.

3 Weighted Two Phase Test Sample Sparse Representation (WTPTSSR) Method

The TPTSSR method reconstructs a test sample by the training samples and the training samples may include some samples which are far from the test sample. This may lead to unstable classification results. In addition, The TPTSSR method exploits linear structure of data and tends to lose local information. It has been shown that in some case locality is very essential and it provides a high recognition ratio [22]. In order to overcome this problem, a novel method called WTPTSSR method is proposed in this paper. Both the local structure and sparsity are taken enough consideration in WTPTSSR. Similar to TPTSSR, the test sample is represents as a linear combination of all the training samples. At the same time, we impose the locality on the l_2 regularization. WTPTSSR solves the following weighted l_2-minimization problem:

$$\hat{X} = \arg\min \|WX\|_2 \ \ subject\ to\ y = AX \tag{7}$$

where W is a diagonal matrix which is the local adaptor that penalizes the distance between y and each training data. The weighted W can be computed by Eq. 8.

$$W = diag\left(\left[dist(y, a_{1,1}), dist(y, a_{1,2}), \ldots, dist(y, a_{c,1}), \ldots, dist(y, a_{c,n})\right]\right) \tag{8}$$

where $dist(y, a_{i,j}) = \left\|y - a_{i,j}\right\|^k$, k is the local adaptor parameter. When k=0, WTPTSSR degenerates to TPTSSR. A larger $dist(y, a_{i,j})$ $(1 \leq i \leq c, 1 \leq j \leq n)$ indicates a farther distance between y and $a_{i,j}$, it can well characterize the similarity between the test sample and training data. Therefore, the coding coefficient of WTPTSSR tends to be local in linear representation.

By the mechanics of Lagrange multiplicity t, we can rewrite (7) into the following equation:

$$\begin{aligned}\min\left(\|AX - y\|_2^2 + t * \|WX\|_2^2\right) &= (X^T A^T - y^T)(AX - y) + t(X^T W^T)WX \\ &= X^T A^T AX - X^T A^T y - y^T AX + y^T y + tX^T W^T WX \\ &= X^T(A^T A + tW^T W)X - 2X^T A^T y + y^T y \end{aligned} \tag{9}$$

where t(t > 0) is a regularization parameter. This Tikhonov regularization problem has the analytical solution

$$X = (A^T * A + t * W^T * W)^{-1} * A^T * y \tag{10}$$

The main steps of the WTPTSSR are as follows.

1 Input: matrix of training samples $A \in R^{m\times N}$, a test sample $y \in R^m$.
2 Normalize the columns of A and y to have unit l_2-norm.
3 Use Eq. (10) to determine the M nearest neighbors for the test sample.
4 Exploit the M nearest neighbors of the test sample to construct (4) and solve this equation.
5 Use Eq. (6) to compute deviation con_i which is generated from the ith class, and classify the test sample into the class that has minimum deviation.

WTPTSSR combines both linear information and local information for improving recognition. Therefore, it can get high recognition performance. From the perspective of linearity, TPTSSR is a direct extension of WTPTSSR.

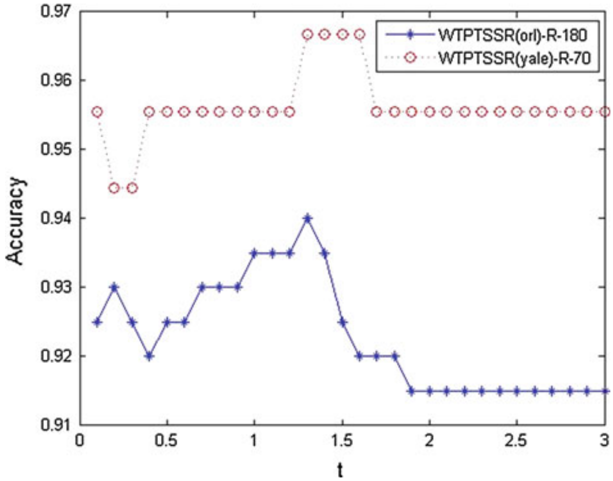

Fig. 1 Performance of WTPTSSR for varying t

4 Experiments

In this section, experiments are designed to evaluate the efficacy of our WTPTSSR algorithm. Firstly we determine the procedure parameters. The parameter δ and γ in TPTSSR algorithm were respectively set to 0.01, and for the parameter M (the number of the NN samples for the test sample) in TPTSSR and WTPTSSR algorithms, we select the optimal value in the experiments.

4.1 Selection of Parameter

There is still a regularization parameter t to be determined in WTPTSSR algorithm. In order to evaluate the parameter t, we use Olivetti Research Laboratory (ORL) face database and Yale database to test it. We randomly choose five different images per individual to form the training set. The rest of each person is used for test. The experiment is repeated for ten times on each database. The reduced dimension in experiments is set to 180 for ORL face database and to 80 for Yale face database. The parameter t is varied from 0.1 to 3 and Fig. 1 plots the average recognition accuracy of WTPTSSR with respect to t.

It can be seen from Fig. 1 that the recognition rates of WTPTSSR vary along with the change of t. The average recognition rates of WTPTSSR in two face databases reach the highest recognition rate when t is 1.3. Therefore, we generally select t = 1.3 in the next experiments.

4.2 Face Recognition Experiments

In this section, we use four data sets which are publicly available for the experiments. For the sake of completeness, we compare the performance of PCA [5], SRC [11], WSRC [22], TPTSSR and the proposed WTPTSSR.

Fig. 2 Sample images of one person on ORL database

Table 1 The maximal average recognition rates and standard deviations of the various methods on ORL database across ten runs when five images per individual are randomly used for training

Method	PCA	SRC	WSRC	TPTSSR	WTPTSSR
Recognition rate	88 % ± 3.0	92 % ± 1.0	93 % ± 2.0	92.5 % ± 1.5	93.5 % ± 2.0
Dimension	50	40	70	40	80

4.2.1 Experiment on ORL Face Database

The ORL face database [25] contains 400 images of 40 individuals (ten samples per person). The images are captured at different times and have different variations including expressions (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20°. The size of each face image is 112×92 pixels. Figure 2 shows sample images of one person.

In the experiment, five images per individual are randomly chosen for training and the remaining five images are used for testing. The experiment is repeated for ten times. The parameter M in TPTSSR and WTPTSSR algorithms is set as 35. Figure 3 shows the average recognition rates versus the dimensions. Table 1 lists the maximal average recognition rate and the standard deviation of each method across ten runs. From Fig. 3, we can see two main points. First, our WTPTSSR outperforms other four algorithms irrespective of the variation of dimensions. Second, WTPTSSR always outperforms TPTSSR. From Table 1, it can be seen that the maximal average recognition rate of WTPTSSR achieves 93.5 %.

4.2.2 Experiments on AR Face Database

The AR face database [26] contains over 4,000 color face images of 126 people, including 26 frontal views of faces with different facial expressions, lighting conditions, and occlusions for each people. The pictures of 120 individuals were taken in two sessions (14 days apart) and each session contains 13 color images. Fourteen face images (each session containing

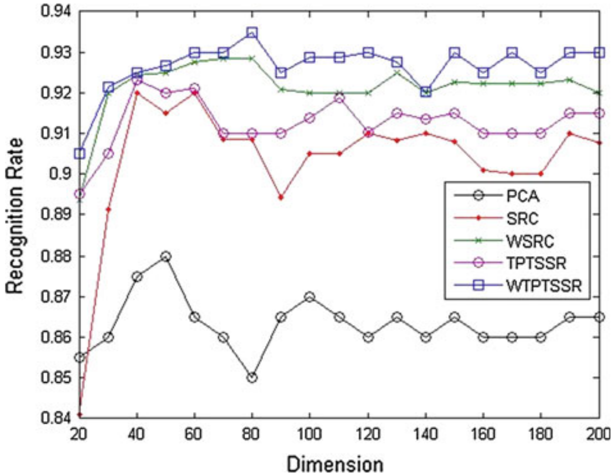

Fig. 3 The average recognition rates versus the reduced space dimension for the ORL database

Fig. 4 Sample face images from AR database

seven) of these 120 individuals are selected in our experiment. The images are converted to grayscale. Figure 4 shows sample images of one person.

In our experiment, we select 120 × 7 non included face images of 120 persons from the first session for the training data, and the testing data are composed of 120 × 7 non included face images from the second session. The parameter M in TPTSSR and WTPTSSR algorithms is set as 150. Figure 5 shows the recognition rates versus the variation of the dimensions. Table 2 lists the maximal recognition rate and the corresponding dimension of each classification method. From Table 2, we can also see that the maximal recognition rates of WTPTSSR and TPTSSR are close. It can be known from Fig. 5 that our WTPTSSR consistently outperforms other four classification methods irrespective of the variation of the dimensions.

4.2.3 Experiments on Yale Face Database

The Yale face database contains 165 images of 15 individuals, and each person has 11 images under various facial expressions and lighting conditions. In our experiments, each image is

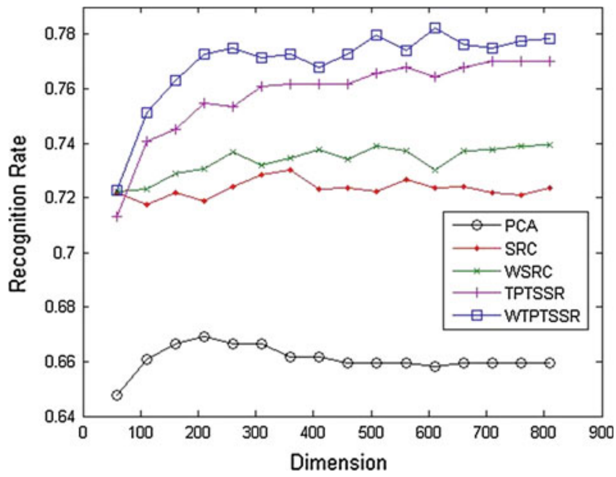

Fig. 5 The average recognition rates versus the reduced space dimension for AR database

Table 2 The maximal recognition rates of the various methods on AR database

Method	PCA	SRC	WSRC	TPTSSR	WTPTSSR
Recognition rate	66.9 %	73 %	73.92 % ± 2.0	77.02 %	78.21 %
Dimension	210	360	260	710	610

Fig. 6 Sample face images from the Yale database

manually cropped and resized to 100 × 80 pixels. Figure 6 shows sample images of one person.

For this experiment, five images per individual are randomly chosen for training and the remaining six images are used for testing. The experiment is repeated for ten times. Figure 7 illustrates the recognition rates of each method versus the number of the nearest neighbors. Table 3 lists the maximal recognition rate of each method and the corresponding dimension or the number of the nearest neighbors. It can be seen from Figure 7 that both WTPTSSR and TPTSSR have a significantly high recognition rate. From Table 3, it can be known that

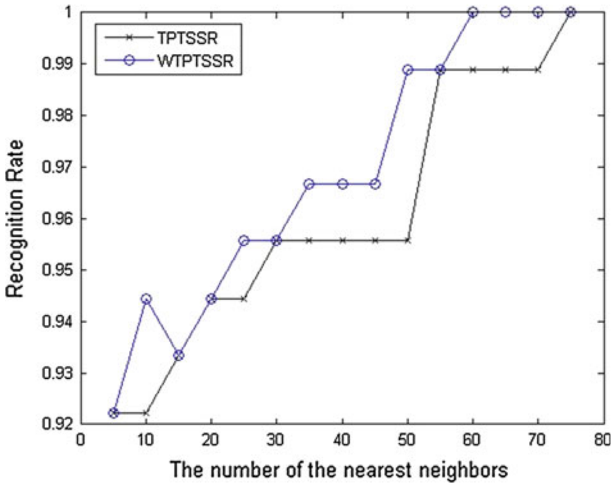

Fig. 7 The recognition rates of TPTSSR and WTPTSSR versus the variation of the nearest neighbor numbers on Yale database

Table 3 The maximal recognition rates of the various methods on Yale database

Method	PCA	TPTSSR	WTPTSSR
Recognition rate	87.78 % ± 2.5	100 % ± 2	100 % ± 1.5
Dimension or the number of the neighbors	41(D)	75(N)	60(N)

Fig. 8 Sample face images from the FERET database

the maximal recognition rates of TPTSSR and WTPTSSR achieve 100 % when the number of the NNs is respectively 60 and 75.

4.2.4 Experiments on FERET Face Database

The FERET database [27] contains a total of 13,539 face images of 1,565 subjects. The images vary in size, pose, illumination, facial expression, and age. We selected 1400 images of 200 individuals (each one has seven images). Each image was cropped to 80 × 80 pixels. Figure 8 shows images of one individual.

In our experiment, we select the first three images per individual for training and the rest for testing. PCA is used to feature extraction. The recognition results are illustrated in Figure 9. Table 4 lists the maximal recognition rate of each method and the corresponding dimension or the number of the NNs. From Fig. 9, it can be seen that our WTPTSSR consistently

Table 4 The maximal recognition rates of the various methods on FERET database

Method	PCA	TPTSSR	WTPTSSR
Recognition rate	44.87 %	57.62 %	59.25 %
Dimension or the number of the neighbors	67(D)	10(N)	20(N)

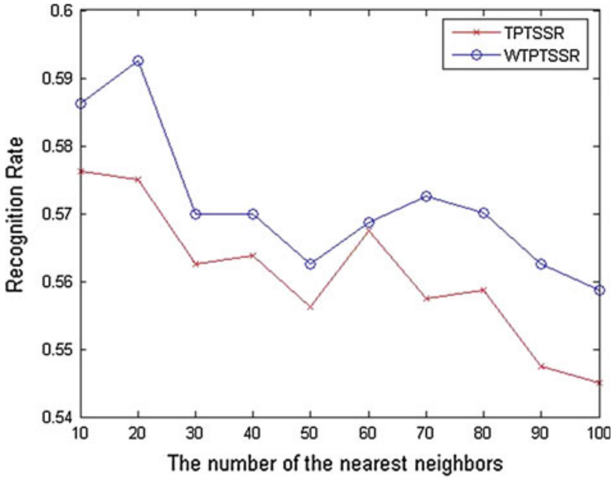

Fig. 9 The recognition rates of TPTSSR and WTPTSSR versus the number of the nearest neighbors on FERET database

outperforms TPTSSR irrespective of the variation of the dimensions. Table 4 shows that WTPTSSR has the best performance.

5 Conclusion

In this paper, we first review the two phase test sample sparse representation (TPTSSR) method which used an elaborate scheme to first determine the training samples that are the best suited to represent the test sample and depends on a weighted sum of these training samples to classify the test sample. TPTSSR exploits global information and tends to lose local information. However, it has been proved that the local information is effective for image classification. Therefore, as a direct extension of TPTSSR, the WTPTSSR method is proposed, which integrates the locality to TPTSSR method. Experiments on ORL, Yale, FERET and AR face databases have been demonstrated the effectiveness of our method.

Acknowledgments This work was partly supported by National Natural Science Foundation of China (31201801, 2011DFR10480, 2011AA041001), China Postdoctoral Science Foundation (2013M542001) and the Natural Science Foundations of Henan Province (102300410113, 102300410113).

References

1. Tian X, Tao D, Rui Y (2012) Sparse transfer learning for interactive video search reranking. ACM Trans Multimed Comput Commun Appl 8:26. doi:10.1145/2240136.2240139

2. Gao X, Wang N, Tao D, Li X (2012) Face sketch-phono synthesis and retrieval using sparse representation. IEEE Trans Circuits Syst Video Technol 22(8):1213–1216
3. Tao D et al (2008) Bayesian tensor approach for 3-D face modelling. IEEE Trans Circuits Syst Video Technol 18(10):1397–1410
4. Xu Y, Zhong A, Yang J, Zhang D (2010) LPP solution schemes for use with face recognition. Pattern Recognit 43:4165–4176
5. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: IEEE computer society conference on computer vision and, pattern recognition, 1991, pp. 586–591
6. Tao D, Li X, Maybank SJ (2007) General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans Pattern Anal Mach Intell 29(10):1700–1715
7. Guan N, Tao D, Luo Z, Shawe-Taylor J (2012) Manhattann non-negative matrix factorization. J Mach Learn Res. http://arxiv.org/abs/1207.3438
8. Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? J ACM 58(3):1–37
9. Guan N, Tao D, Luo Z, Yuan B (2012) NeNMF: an optimal gradient method for nonnegative matrix factorization. IEEE Trans Signal Process 60(6):2882–2898
10. Wright J, Ma Y, Mairal J et al. (2009) Sparse representation for computer vision and pattern recognition. In: Proceedings of IEEE, 2009, pp. 1–8
11. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31:210–227
12. Gao S, Tsang I, Chia L-T (2010) Kernel sparse representation for image classification and face recognition. In: Daniilidis K, Maragos P, Paragios N (eds) Computer vision ECCV 2010. Springer, Berlin, Heidelberg, pp 1–14
13. Yang M, Zhang L, Yang J, Zhang D (2011) Robust sparse coding for face recognition. In: IEEE international conference computer vision and pattern recognition, 2011, pp. 625–632
14. Lai Z, Jin Z, Yang J, Wong WK (2010) Sparse local discriminant projections for feature extraction. In: Proceedings of ICPR, 2010, pp. 926–929
15. Zhang L et al. (2011) Sparse representation or collaborative representation: which helps face recognition? In: ICCV 2011
16. Shi Q, Eriksson A, Hengel A, Shen C (2011) Is face recognition really a compressive sensing problem? In: CVPR 2011
17. Xu Y, Zhang D, Yang J, Yang J-Y (2011) A two-phase test sample sparse representation method for use with face recognition. IEEE Trans Circuits Syst Video Technol 21(9):1255–1262
18. Xu Y, Zuo W, Fan Z (2011) Supervised sparse presentation method with a heuristic strategy and face recognition experiments. Neurocomputing 79:125–131
19. Xu Y, Zhu Q (2013) A simple and fast representation-based face recognition method. Neural Comput Appl. doi:10.1007/s00521-012-0833-5
20. Tao D, Li X, Wu X, Maybank SJ (2009) Geometric mean for subspace selection. IEEE Trans Pattern Anal Mach Intell 31(2):260–274
21. Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y (2010) Locality-constrained linear coding for image classification. In: IEEE computer society conference on computer vision and, pattern recognition, 2010, pp. 3360–3367
22. Lu C-Y, Min H, Gui J, Zhu L, Lei Y-K (2010) Face recognition via weighted sparse representation. J Vis Commun Image Represent 24:111
23. Yu K, Zhang T, Gong Y (2009) Nonlinear learning using local coordinate coding. Adv Neural Inf Process Syst 22:2223–2231
24. Yu K, Zhang T (2010) Improved local coordinate coding using local tangents. In: International conference on machine learning, 2010, pp. 1215–1222
25. Samaria F, Harter A (1994) Parameterisation of a stochastic model for human face identification. In: Second IEEE workshop on applications of computer vision, Sarasota
26. Martinez A, Benavente R (1998) The AR face database. In: CVC technical report no. 24
27. Phillips PJ, Wechsler H et al (1998) The FERET database and evaluation procedure for face-recognition algorithms. Image Vis Comput 16(5):295–306