Applied Mathematics & Information Sciences
*An International Journal*

http://dx.doi.org/10.12785/amis/070645

# Robust Sparse 2D Principal Component Analysis for Object Recognition

*Jicheng Meng*[1,*] *and Xiaolong Zheng*[2]

[1] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, P.R. China
[2] School of Electronic and Information Engineering, Hangzhou Dianzi University, Hangzhou, 310018, P.R. China

Received: 11 Apr. 2013, Revised: 15 Aug. 2013, Accepted: 17 Aug. 2013
Published online: 1 Nov. 2013

**Abstract:** We extensively investigate robust sparse two dimensional principal component analysis (RS2DPCA) that makes the best of semantic, structural information and suppresses outliers in this paper. The RS2DPCA combines the advantages of sparsity, 2D data format and L1-norm for data analysis. We also prove that RS2DPCA can offer a good solution of seeking spare 2D principal components. To verify the performance of RS2DPCA in object recognition, experiments are performed on three famous face databases, i.e. Yale, ORL, and FERET, and the experimental results show that the proposed RS2DPCA outperform the same class of algorithms for face recognition, such as robust sparse PCA, L1-norm-based 2DPCA.

**Keywords:** L1-norm, robust sparse two dimensional principal component analysis (RS2DPCA), object recognition.

## 1 Introduction

As a class of traditional data analysis approach, subspace learning is an active direction in pattern recognition and computer vision. Most of these approaches, e.g. principal component analysis (PCA), independent component analysis (ICA), and manifold learning algorithms, when combined with sparsity, they have shown state-of-the-art results [1]. This is due to the fact that the sparse representation can uncover the semantic information more than compact high-fidelity information of data. The semantic information, especially for image data, maybe has more important than the compact high-fidelity information in pattern recognition [2].

PCA is a traditional and popular data analysis approach which is widely used in computer vision, and pattern recognition, etc. [3]. However, traditional PCA for image data has three disadvantages. First, it suffers from the so-called curse of dimensionality and tracking with high-dimensional samples directly is also computationally expensive, especially in face recognition and gait recognition etc. Thanks to two dimensional PCA(2DPCA) and MPCA etc. [4,5], this problem can be dealt with efficiently. Furthermore, directly operating 2D data with the matrix algebra methods is much simpler and structural information of the original data can be preserved sufficiently, bringing more important and efficient features for image object recognition [4,6].

Second, the abovementioned semantic information can not be extracted in traditional PCA. The key idea behind PCA is to reduce the dimensionality of the high-dimensional data consisting of a larger number of interrelated variables, while holding as much as possible the variation present in these data. So, in the low dimensional linear subspace expanded by the principal components, the data structure of the original data space can preserve the compact high-fidelity information of data. Each component of PCA is a linear combination of all the original data variables. All loadings in this combination are general non-zeros. This is due to the fact that the traditional PCA lack semantic information. For example, in gene expression analysis, each variable might express a specific gene, and the interpretation of principal components will be easy if the components have many zero loadings[7]. Therefore, sparse PCA (SPCA) and its variants is an attractive topic [8,9,10]in virtue of its component has fewer non-zero loadings. Furthermore, for the underlying principle of traditional PCA, it is based on L2-norm that definitely suffers from the outliers. Fortunately, this robust problem can be alleviated with the use of L1-norm [11,12], leading to the L1-norm-based PCA methods(PCAL1) or called robust PCA (RPCA)

* Corresponding author e-mail: jcmeng@uestc.edu.cn

methods. In nature, L1-norm is also an appropriate measure to a model, where the noise follows the Laplace distribution [3]. Moreover, in some special applications like cellular automata model in which translations can only occur along unit direction, using L1-norm to measure the fit of a subspace is more natural [3].

L1-norm-based 2DPCA (2DPCAL1) [13] can overcome the first and third disadvantages theoretically. It not only is robust to outliers because of using L1-norm, but also suppresses the "dimensionality curse" by virtue of directly operating 2D date. Recently, a robust sparse PCA (RSPCA) based on L1-norm has been proposed in [7]. RSPCA seeks the directions of feature space, where the L1-norm dispersion of the input data is maximized as large as possible, instead of the L2-norm dispersion employed by traditionally aforementioned SPCA methods. RSPCA combine the merits of sparsity and L1-norm to subdue the second and third disadvantages . Inspired by these, we proposed the robust sparse 2DPCA(RS2DPCA) in our early work [14]. Though the proposed RS2DPCA can efficiently overcome the weaknesses of traditional PCA by virtue of sparsity, 2D data format, and L1-norm, the theory analysis is not investigated and the large scale experiments are still lacking. In this paper, we give more theory analysis and more experimental results for object recognition.

Object recognition is now becoming a very active research subject. It includes many fields, such as face recognition and gait recognition. Various types of object recognition approaches have been presented [15,16,17]. Among these, subspace methods have received incessant attention for a long time. While subspace methods have been studied intensively recently, updated research results are emerging from the areas of tensor analysis(including 2D data based), sparse representation, manifold learning, matrix factorization, and matrix completion [18,19,20, 21,5,22,23]. Here we apply RS2DPCA to face recognition, and show its inspiring application prospect in pattern recognition and computer vision.

The rest of this paper is organized as follows. In section 2, RS2DPCA is described, and we give the theory analysis about RS2DPCA in detail. Furthermore, we introduce it to object recognition in this section. The experimental results on true data, i.e. three famous public face databases, are shown in section 3. The last section concludes RS2DPCA and discusses the future works.

## 2 Robust sparse 2DPCA

In this section, we extend RSPCA to RS2DPCA and analyze it in detail[14].

**Algorithm Description** RS2DPCA can be depicted as the following optimization problem.

$$W^* = \arg\max_{w} \sum_{i=1}^{N} ||A_i W||_1 = \sum_{i=1}^{N}\sum_{j=1}^{m} |a_j^i W| \qquad (1)$$

subject to $W^T W = I_m, ||W||_1 < t$.

Where $|\cdot||_1$ denotes the $\ell_1$-norm of a matrix or vector, and $N$ is the size of the given data respectively. $I_m$ denotes the m rank unit matrix.

The data $A_i = \left[(a_1^i)^T \cdots (a_j^i)^T \cdots (a_m^i)^T\right]^T \in R^{m\times n}$( $m\times n$)is the image size). For face data, $A_i (i = 1,2,\cdots,N)$ means the image data.

Obtaining the optimal solution of (1) is a very hard work. So, as PCA-L1 did, we simply replace (1) by some easier problem to solve as below:

$$w^* = \arg\max_{w} \sum_{i=1}^{N} ||A_i w||_1 = \sum_{i=1}^{N}\sum_{j=1}^{m} |a_j^i w| \qquad (2)$$

subject to $W^T W = I_m, ||W||_1 < t$.

It is obvious that solving (2) is much easier than (1). Unfortunately, it is difficult to find optimal solution because it contains sparsity constraint and absolute value operation. Now, we firstly introduce a good solution for (2), then give the successive greedy solutions of (2) to give a good approximation for (1). The algorithms for solving the optimization problem (2) are described as Algorithm a and b.

Algorithm a: RS2DPCA for one sparse PC

1.Initialize $w(0)$. Normalize $w(0)$ , i.e. set $w(0) = \frac{w(0)}{||w(0)||_2}$ and $t = 0$.

2.Set $u = (u_1, u_2, \cdots, u_m)^T = \sum_{i=1}^{N}\sum_{j=1}^{n} p_{ij}(t) a_j^{iT}$, where $p_i(t) = \begin{cases} 1, if\ a_j^i w(t) \geq 0 \\ -1, if\ a_j^i w(t) < 0 \end{cases}$.

3.Set be the $(k+1)-th$ largest entry of $|u|$ . Then, let $t = t+1$. $w_l(t) = \mathrm{sgn}(u)(|u_l| - v)_+$for $l = 1,2,\cdots,m$ . Here $(x)_+ = \begin{cases} x, x > 0 \\ 0, x \leq 0 \end{cases}$ and $\mathrm{sgn}(x) = \begin{cases} 1, x > 0 \\ 0, x = 0 \\ -1, x < 0 \end{cases}$ denote the threshold and sign function respectively. Normalize $w(t)$ , i.e. set $w(t) = \frac{w(t)}{||w(t)||_2}$ .

4. (a)If $w(t) \neq w(t-1)$ , go to 2.
   (b)Else if there exists $i$ such that $a_j^i w(t) = 0$ and $a_j^i$ is not a zero vector, set $w(t) = \frac{w^T(t)+\Delta w}{||w^T(t)+\Delta w||}$ , and go to 2. Here $\Delta w$ is a small nonzero random vector.
   (c)Set $w^* = w(t)$ and stop.

Algorithm a only can compute the first sparse PC. To obtain $w_d (d > 1)$, the input training data should be updated. This algorithm is depicted as follows.

Algorithm b: RS2DPCA for $d(d > 1)$ sparse PCs

For $l = 2,3,\cdots,d$ do

1.Update $a_j^{il} = a_j^{il-1} - a_j^{il-1} w_{l-1} w_{l-1}^T (i = 1,2,\cdots,m; j = 1,2\cdots,n)$, here if $l = 2$, then $\mathrm{a_j^{i1}} = \mathrm{a_j^i}$.

2.Apply algorithm 1 to $a_j^{il}$ $(i = 1,2,\cdots,m; j = 1,2\cdots,n)$.

**Algorithm Analysis** To analysis the convergence of Algorithm 1, we review Lemma 1 in [4]. Lemma 1. Given the vector $v = (v_1, v_2, \cdots, v_d)^T$ , the solution of the following optimization problem ,

$$\max_{w} w^T v \tag{3}$$

subject to $w^T w = 1, ||w||_1 < t$. is of the following form

$$w^* = \frac{\beta}{\underset{w}{||\beta||_2}} \tag{4}$$

where $\beta = (\beta_1, \beta_2, \cdots, \beta_d)^T$ and

$$\beta_i = \text{sgn}(v_i)(|v_i| - \gamma)_+, i = 1, 2, \cdots, d \tag{5}$$

Furthermore, if the sparsity of the solution $w^*$ is known to be $k$ beforehand, then $\gamma = \theta_{k+1}$ where $\theta_k$ denotes the $k-th$ largest element of $|v|$.

Based on Lemma 1, we can prove the convergence of $w(t)$ by verifying the nondecreasing property of $\sum_{i=1}^{N} ||A_i w(t)||_1$ with respect to $t$ as follows:

$$\begin{aligned}
\sum_{i=1}^{N} ||A_i w(t)||_1 &= \sum_{i=1}^{N}\sum_{j=1}^{m} |a_j^i w(t)| \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n} p_{ij}(t) a_j^{iT} w(t) \\
&\geq \sum_{i=1}^{N}\sum_{j=1}^{n} p_{ij}(t-1) a_j^{iT} w(t) \\
&\geq \sum_{i=1}^{N}\sum_{j=1}^{n} p_{ij}(t-1) a_j^{iT} w(t-1) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n} |a_j^{iT} w(t-1)| \\
&= \sum_{i=1}^{N} ||A_i w(t-1)||_1
\end{aligned}$$

According to the fact that $p_{ij}(t) a_j^{iT} w(t) \geq 0$ hold for all $i$ and $j$, the first inequality is evident. According to Lemma 1, for any $t$ , $w(t)$ is the $k$-sparse unit vector which maximizes the inner product of $w(t)^T v(t-1) = w(t) v(t-1)^T$ (Here $v$ can be regarded as $u$ in Algorithm 1. So the second inequality holds. From the above results, the value of $w$ is updated towards a good solution.

Two points about RS2DPCA should be noted. On one hand, RS2DPCA can only offer the local maximum of $w$ , and the global optimal solution may not be obtained. On the other hand, the projected vectors $\{w_l\}_{l=1}^{d}$ , obtained by Algorithm 2, are not orthogonal to each other. Despite of these weaknesses, the proposed algorithm show its efficiency in seeking good robust sparse 2D PCs for face recognition.

**Object Recognition using RS2DPCA** In object recognition, a feature matrix $\mathbf{Y}_i = [y_{i1} y_{i2} \cdots y_{id}]$ for each training sample can be obtained by $y_{ik} = \mathbf{X}_i w_l, l = 1, 2, \cdots, d$ . Here, $\{w_l\}_{l=1}^{d}$ denotes the projected vectors in RS2DPCA.

In the same way, we can also get a feature matrix $\mathbf{Y}_t = [y_{t1} y_{t2} \cdots y_{td}]$ for each testing sample after the transformation by RS2DPCA. Then, a nearest neighbor classifier based on the matrix distance is used for classification.

$$\begin{aligned}
c &= \arg\min d(\mathbf{Y}_t, \mathbf{Y}_i) \\
&= \arg\min \sum_{k=1}^{d} \| \mathbf{y}_{tk} - \mathbf{y}_{ik} \|_2
\end{aligned} \tag{6}$$

where $d(\mathbf{Y}_t, \mathbf{Y}_i) = \sum_{k=1}^{\text{d}} ||y_{tk} - y_{ik}||_2 \quad c \in [1, 2, \cdots, N]$ , and the distance between $\mathbf{Y}_c$ and $\mathbf{Y}_t$ is minimal. Then, $\mathbf{Y}_t$ belongs to the class where $\mathbf{Y}_c$ belongs to. This classification measure is based on the Yang distance[10].

## 3 Experimental results

In this section,we use three well-known face databases, i.e. the YALE face database [24], and the ORL face database [25], and the FERET face database [26], to verify RS2DPCA in face recognition. On YALE and ORL face database, we examined the system performance by average recognition rates (ARR) and standard deviations(SD, ), while on FERET face database, we investigated the system performance by the recognition rate(first hit). For comparison, the classical PCA, PCAL1[12], RSPCA, 2DPCA, and 2DPCAL1 have also been utilized. All programs were implemented under the Matlab 7.0 platform.

**Results on YALE Database** The YALE face database includes 11 different images of each of 15 individuals, and the images vary in different light conditions and facial expressions. All images were grayscaled and we cropped and normalized them to a resolution of 66x56 pixels in our experiments. We gave the average recognition rates (ARR) and standard deviations(SD,$\sigma$) on the above mentioned six algorithms. We randomly chose 3 samples per person for training, and the others for testing. The ARR and over 10 runs were calculated, and they were described in Fig.1, and Fig. 2.

The classification results using 3 samples per person are shown in Fig.1 and Fig. 2. These two figures indicate that generally the performance of 2D-based algorithms is better than that of 1D-based, and the RS2DPCA is the best among these shown algorithms.

**Results on ORL Database** The ORL face database includes ten different images of each of 40 individuals, and the images vary in different light conditions, facial expressions, and sampling time etc. All images are

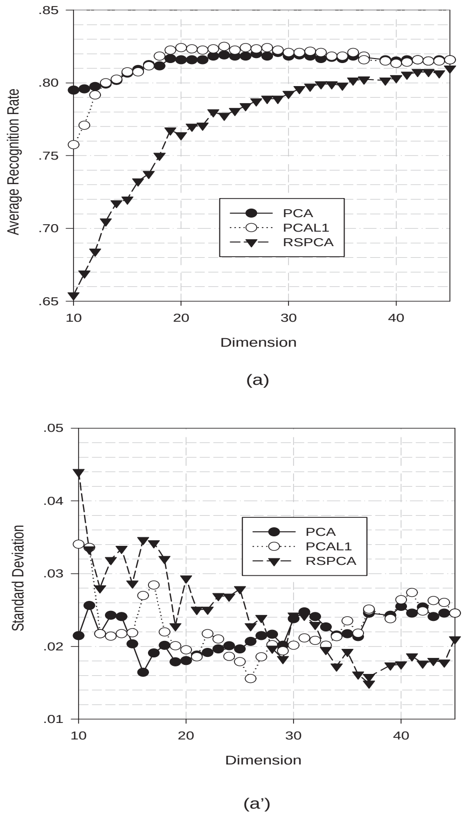


**Fig. 1** ARR and SD using 1D-based algorithms on YALE. (a) ARR. ( á) SD.

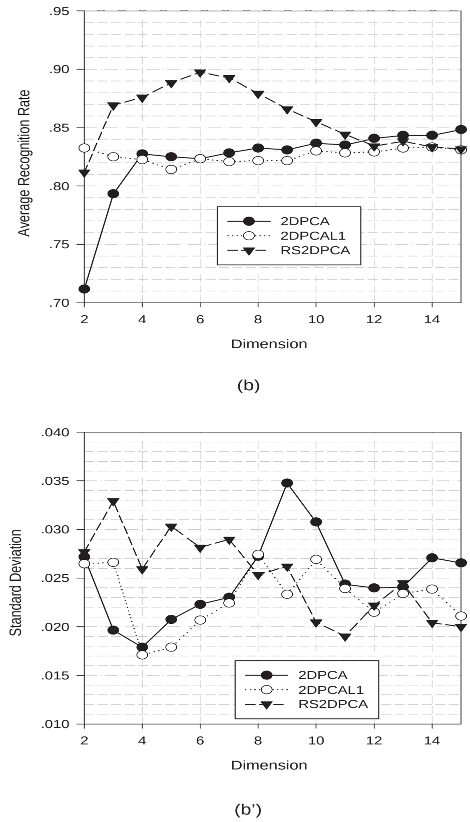


**Fig. 2** ARR and SD using 2D-based algorithms on YALE. (a) ARR. ( b́) SD.

grayscale and we normalized them to a resolution of $68 \times 58$ pixels in our experiments. We randomly chose 2 samples per person for training, and the others for testing. The ARR and $\sigma$ over 10 runs were calculated, and they are described in Fig.3 and Fig.4. These figures also indicate that the performance of RS2DPCA is the best among these six algorithms.

**Results on FERET Database**For the purpose of applications, we tested RS2DPCA on the FERET face database, which has been widely used to evaluate some face recognition methods [27]. In our experiments, two FERET image sets are used i.e. FA and FB. FA, consisting of frontal images of 1196 individuals, is a regular frontal face library, which was used as the gallery set. FB, consisting of frontal images of 1195 individuals, is an alternative frontal face library, taken seconds after the corresponding FA, which was used as the probe set. Before performing the experiments, all images in FA, FB sets were rectified using the positions of the eyes, provided by FERET, and then cropped and normalized to the $60 \times 50$ images. To further reduce the effect of illumination, we applied the histogram equalization method. The experimental results (the first hit) also suggest that the recognition rates of RS2DPCA are superior to those of the same class of algorithms.

**Table 1** Recognition Rates On FERET(%)

| Algorithm | The first hit |
|---|---|
| PCA | 79.83 |
| PCAL1 | 81.67 |
| RSPCA | 79.67 |
| 2DPCA | 81.34 |
| 2DPCAL1 | 82.01 |
| RS2DPCA | 82.26 |

## 4 Conclusion

We investigate the robust sparse 2DPCA (RS2DPCA) to make full use of semantic, structural information, as well as suppressing outliers. RS2DPCA employs the

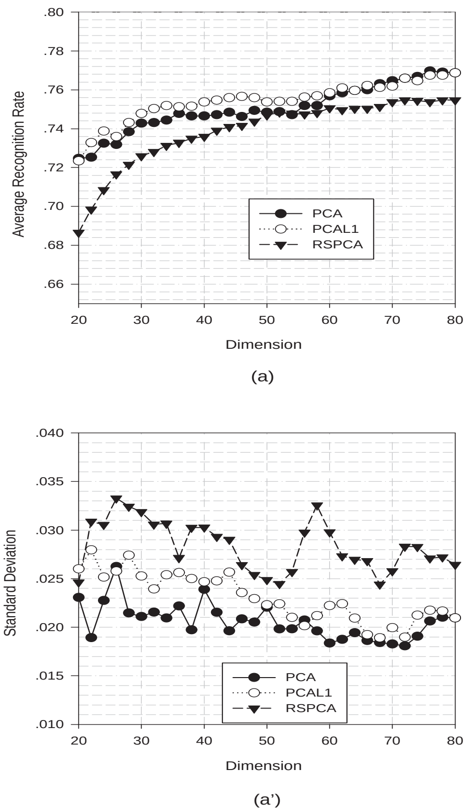


**Fig. 3** ARR and SD using 1D-based algorithms on ORL. (a) ARR. ( á) SD.

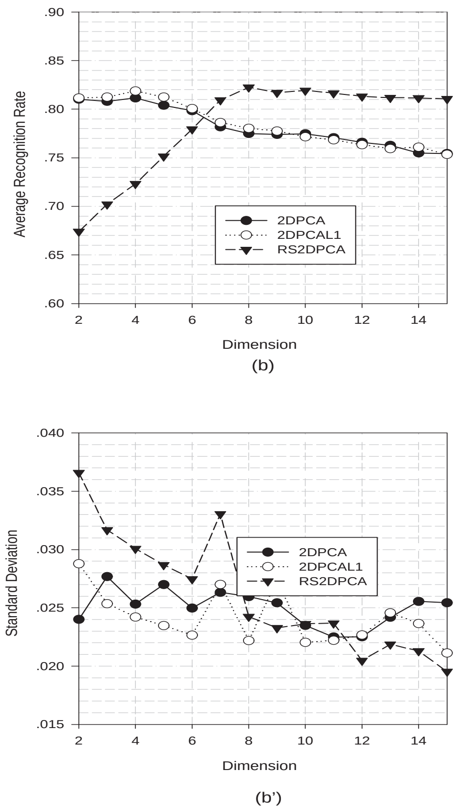


**Fig. 4** ARR and SD using 2D-based algorithms on ORL. (a) ARR. ( b́) SD.

advantages of sparsity, 2D data format and L1-norm at the same time. To evaluate the performance of RS2DPCA in object recognition, we use three famous face databases, i.e. Yale, ORL, and FERET. The experimental results suggest that the proposed RS2DPCA can outperform the same class of algorithms, such as RSPCA, 2DPCAL1.

Though RS2DPCA has shown promising applications to object recognition, there are some issues to be investigated. First, the L1-norm optimization (2) only achieves a local optimum and finding the best parameters (sparsity and feature dimension) is a stuffy work. Second, robust sparse PCA based on tensor data may be developed. Furthermore, the proposed RS2DPCA should be further evaluated by more practical applications, including image compression, reconstruction, and so on.

## Acknowledgement

This work is supported in part by the National Natural Science Foundation for Distinguished Young Scholars of China under Grant No.60905011. We appreciate the anonymous reviewers for their helpful comments, and we also would like to thank Dr. Deyu Meng for sharing the related matlab code at http://gr.xjtu.edu.cn:8080 /c/ portal /layout?p_l_id =PUB.3243.2.

## References

[1] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, Proceeding of the IEEE, **98**, 1031-1044 (2010).

[2] M. Elad, M. A. T. Figueiredo, and Y. Ma, Proceeding of the IEEE, **98** 972-982 (2010).

[3] J. P. Brooks, J. H. Dula, and E. L. Boone, Submitted. http://www.optimization-online.org/ DB_HTML/2010 / 0 1 /2513.html.

[4] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, IEEE Trans. Pattern Anal. and Mach. Intell., **26**, 131-137 (2004).

[5] J. Lu, and Y. Tan,IEEE Trans. Syst. Man Cybern. B, Cybern, **40**, 958-963 (2010).

[6] W. M. Zuo, D. Zhang, and K. Wang, IEEE Trans. Syst. Man Cybern. B, Cybern, **36**, 863-871 (2006).

[7] D. Y. Meng, Q. Zhao, and Z. Xu, Pattern Recognition, **45**, 487-497 (2012).
[8] H. Zou, T. Hastie, and R. Tibshirani, Journal of Computational and Graphical Statistics, **15**, 265-286 (2006).
[9] R. Zass, and A. Shashua, Proc. Neural Information and Processing Systems, 1561-1568 (2006).
[10] R. Jenatton, G. Obozinski, and F. Bach, Technical report, arXiv:0909. **1440**, (2009).
[11] A. Baccini, P. Besse, and A.D. Falguerolles, Proceedings of the International Conference on Ordinal and Symbolic Data Analysis, 359-368 (1996).
[12] N. Kwak, IEEE Trans. Pattern Anal. and Mach. Intell., **30**, 1672-1680 (2008).
[13] X. Li, Y. Pang, and Y. Yuan, IEEE Trans. Syst. Man Cybern. B, Cybern, **40**, 1170-1175 (2010).
[14] J. Meng, and X. Zheng, International Symposium on Photonics and Optoelectronics, (2012).
[15] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, ACM Comput. Surv., **35**, 399-458 (2003).
[16] T. Ahonen, A. Hadid, and M. Pietikainen, IEEE Trans. Pattern Anal. and Mach. Intell., **28**, 2037-2041 (2006).
[17] W. Xue-guang, C. Shu-hong, An improved image segmentation algorithm based on two-dimensional Otsu method, Information Sciences Letters, **1**, 77-83 (2012).
[18] S. Yan, D. Xu, Q. Yang, L. Yang, X. Tang, and H. Zhang, IEEE Trans. Image Processing, **16**, 212-220 (2007).
[19] H.P. Shen, and J.Z. Huang, Journal of Multivariate Analysis, **99**, 1015-1034 (2008).
[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, IEEE Trans. Pattern Anal. and Mach. Intell., **31**, 210-227 (2009).
[21] E. Candes, X. Li, Y. Ma, and J. Wright, Journal of ACM, **58**, 1-37 (2009).
[22] Y. Peng, A. Ganesh,J. Wright,W. Xu ,Y. Ma, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 763-770 (2011).
[23] G. Zhang, and Z. Cui, Applied Mathematics & Information Sciences, **6**, 93S-97S (2012).
[24] Yale University, http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
[25] AT&T, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.htm.
[26] Defense Advanced Research Products Agency (DARPA), http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.
[27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, IEEE Trans. Pattern Anal. and Mach. Intell., **22**, 1090-1104 (2000).

**Jicheng Meng** received the B.S. and M.S. degrees in Computer Science from University of Electronic Science & Technology of China(UESTC), Cheng du, China, in 1997 and 2000, respectively, and the Ph.D. degree in electronic engineering from UESTC, Chengdu, China, in 2005. He is currently an associate professor at School of Automation Engineering, UESTC. His research interests include biometrics recognition, computer vision, and machine learning. He is a member of the IEEE.

**Xiaolong Zheng** received his B.E.(2002) degree in Electronic Engineering from Hang zhou Institute of Electronic Engineering. He obtained his M.S. (2005) degrees in Circuit and System from University of Electronic Science and Technology of China. In 2008, he got his Ph.D degree in Pattern Recognition and Intelligent System from Institute of Automation, Chinese Academy of Sciences. He is currently a lecturer in Hangzhou Dianzi University.His research interests include image processing, biometrics and pattern recognition.