Original Article

A syncretic representation for image classification and face recognition

Zhongli Ma [a,*], Quanyong Liu [a], Kai Sun [b], Sui Zhan [c]

[a] *College of Automation, Harbin Engineering University, Harbin, China*
[b] *Shenzhen Sunwin Intelligent Corporation, Shenzhen, China*
[c] *College of Computer Science and Electronic Engineering, Hunan University, Changsha, China*

Abstract

For representation based image classification methods, it is very important to well represent the target image. As pixels at same positions of training samples and test samples of an object usually have different intensities, it brings difficulty in correctly classifying the object. In this paper, we proposed a novel method to reduce the effects of this issue for image classification. Our method first produces a new representation (i.e. virtual image) of original image, which can enhance the importance of moderate pixel intensities and reduce the effects of larger or smaller pixel intensities. Then virtual images and corresponding original images are respectively used to represent a test sample and obtain two representation results. Finally, this method fuses these two results to classify the test sample. The integration of original image and its virtual image is able to improve the accuracy of image classification. The experiments of image classification show that the proposed method can obtain a higher accuracy than the conventional classification methods.

Copyright © 2016, Chongqing University of Technology. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Image syncretic representation; Pixel intensity; Image classification; Face recognition

1. Introduction

Image classification is a significant branch of computer vision. In this branch, the representation based classification methods have attracted considerable attention. A good representation for target images is greatly beneficial to improve the performance of image classification [1,2]. An object can be distinguished from the others when its image is well represented by the other images from this object. The combination of multiple representations of images is an effective method to improve the performance of representation based methods [3,4]. Therefore, it is an important and meaningful topic to find a proper representation for representation based image classification methods.

At present, face recognition has been studied widely and many useful methods have been presented [5−9]. However, we still face with some great challenges. Different poses and expressions, various intensities of illuminations and insufficient training samples seriously influence the recognition effects. In order to address these challenges, people have made many efforts. For various illuminations, by handling the original images to enhance pixels with moderate intensities of the original images and reduce the importance of other pixels, Xu et al. [10] obtained the complementary images to improve the accuracy of image classification. Producing the mirror image of the face and integrating the original face image and its mirror image are also useful to improve the recognition accuracy of representation-based face recognition [11]. For the problem of insufficient training samples, Huang et al. [12] proposed a robust kernel collaborative representation classification method based on virtual samples for face recognition to reduce the influence of insufficient training samples. The use of symmetrical face images generated from original face images is very useful to overcome the problem of varying appearances of faces [13,14]. Until now, many works focus on generating virtual or synthesized face images to enhance the recognition accuracy

---

* Corresponding author.
E-mail address: mazhongli@hrbeu.edu.cn (Z. Ma).
Peer review under responsibility of Chongqing University of Technology.

[15−19]. The simultaneous use of original face images and their virtual face images can improve the accuracy of face recognition. What is more, several works have shown that virtual image obtained by exploiting the adjacent rows of original image are also useful for image classification [20−24].

Wright et al. [25] proposed the sparse representation classification (SRC) algorithm which can reach satisfactory result. There are many SRC algorithms [26−30]. However, the original SRC algorithm with the constraint of l_1 minimization is time consuming. Zhang et al. [31] proved that the essence to obtain the satisfactory performance of the SRC algorithm is the collaborative representation but not the sparsity, and proposed a collaborative representation classification (CRC) method with the constraint of l_2 minimization. CRC methods can obtain comparative performance to SRC algorithm, but is much faster than SRC algorithm. Various representation methods with the constraints of l_2 minimization are also proposed, such as linear regression classification (LRC) [32], and two phase sparse representation [33−35]. They not only used simple constraint conditions but also achieved satisfactory recognition accuracy.

In this paper, we propose a novel representation method which can enhance the significances of pixels with moderate intensities of images. The proposed method has the following merits. (1) The novel representation method can classify images effectively. (2) It highlights the regions which has great difference of pixel intensities, such as edges. Edges have been shown to be beneficial to improve the recognition accuracy. The method also exploits the correlation of adjacent rows in a good way. (3) It increases the number of training samples, and represents a test sample effectively by combining original and virtual images.

The remainder of this paper is organized as follows. Section 2 presents the proposed novel representation method of images. Section 3 describes the underlying rationale of the proposed method. Section 4 shows the experimental results. Section 5 provides the conclusions of this paper.

2. The proposed method

The method proposed in [10] generated virtual images which can enhance pixels with moderate intensities of the original images. This method used the following scheme to generate virtual images.

$$J_{ij} = I_{ij} \cdot (m - I_{ij}) \tag{1}$$

where J stands for virtual image and J_{ij} represents the pixel intensity at the i-th row and j-th column of J. Then, it used the difference between adjacent sorted results to automatically decide syncretic coefficient.

Our method mainly includes two procedures. By the first procedure, we obtain the novel representation of original images as follows. Let I stand for an original image which has been converted into a gray image, and I_{ij} represents the pixel intensity at the i-th row and j-th column of I. Let m be the maximum pixel intensity of the gray image. For conventional gray images, $m = 255$. Then the novel representation of I is denoted as

$$J_{ij} = (255 - I_{ij})*I_{(i+1)j} \tag{2}$$

Because J_{ij} may be greater than 255, the normalization measure is applied to J_{ij}.

By using Eq. (2), we obtain virtual images of corresponding original images. The second procedure of our method can be described as the method of fusing original and virtual images. After obtaining virtual images, a classification algorithm is respectively applied to the original and virtual images. Then, the residual between the test sample and train samples is calculated. Let $r_o^j (j = 1,2, … ,c)$ stand for the residual between the test sample and original training samples of the j-th class. Here, c is the number of classes in a database. Let r_v^j ($j = 1,2, … ,c$) stand for the residual between the test sample and virtual training samples of the j-th class. After that, the two residuals are fused based on their weight. The syncretic formula is

$$r_f^j = \alpha*r_o^j + (1 - \alpha)*r_v^j \tag{3}$$

where $r_f^j (j = 1, … ,c)$ stands for the ultimate residual. a is the syncretic coefficient and it is a number between 0 and 1. In the end, the test sample will be classified into the r-th class based on the following formula

$$r = \operatorname{argmin}_j r_f^j \tag{4}$$

The main steps of our method are presented as follows.

Step 1. Separate the database into two sets, a set of training samples and a set of test samples.
Step 2. Obtain virtual images of all original images using Eq. (2). Then all images are converted into unit column vectors with the norm of 1.
Step 3. A classification algorithm is applied to the original and virtual images to obtain corresponding residuals r_o^j and r_v^j, respectively.
Step 4. Obtain residual r_f^j using Eq. (3).
Step 5. Use Eq. (4) to classify the test sample.

3. Rationale of the proposed method

Different intensities of pixels have different importance in image classification and we can exploit a subset of all image pixels for image classification [36]. Hence, it seems that setting different weights to different pixels is reasonable. Furthermore, the important features of an image are mainly concentrated on the moderate intensities of pixels, so it should emphasize the moderate intensities of pixels [10]. Fortunately, our method has the idea that different pixel intensities play different roles in representing the object.

3.1. Analysis of the proposed method

In Section 2, we know that the obtained virtual image is very different from its original image. From Eq. (2), we have the following propositions.

Proposition 1. If I_{ij} is very large, the computational result is quite small no matter what $I_{(i+1)j}$ is large or small.

Proposition 2. If I_{ij} is very small, the computational result changes over $I_{(i+1)j}$. Moreover, the result is large when I_{ij} and $I_{(i+1)j}$ have great difference. The result makes the edges prominent.

Proposition 3. . If I_{ij} is moderate, the result is moderate or small no matter what $I_{(i+1)j}$ is.

It is very easy to prove the above propositions. We can also know that the importance of regions with moderate pixel will be enhanced in the novel representation. Otherwise, the regions with very large or small pixels except for edge regions have a relative small value in the novel representation. The virtual image is a nonlinear transform for the pixels of the original image, and this can be known intuitively from Fig. 2 and Fig. 3. Moreover, when I_{ij} is closed to $I_{(i+1)j}$, Eq. (2) can be regarded as a first and second order entry of the pixel value of original image. Compared with virtual images obtained by linear transform, virtual images obtained by nonlinear transform have more complementarity with corresponding original images. That's to say, the integration of original and corresponding virtual images can reflect more image information than any one of them. For the deformable original image, the pixel with moderate intensity may be more stable, so our method is reasonable. Moreover, in order to exploit the information contained in original and virtual images, our method simultaneously uses them to perform image classification.

When fusing the two residuals in Eq. (3), α is selected according to the classification results of original and virtual images. When the classification result of original images is more reliable than virtual images, α is set to be a large value. Otherwise, α is set to be a small value.

3.2. Insight into the proposed method

In this part, we give an intuitive explanation to the rationale of our method. We take the ORL database as an example to illustrate the difference between original and virtual images. Fig. 1 shows five original images and their corresponding virtual images of the first subject in the ORL database. From these images, we know that virtual images are directly associated with the corresponding original images but they also

(a) Five original face images.

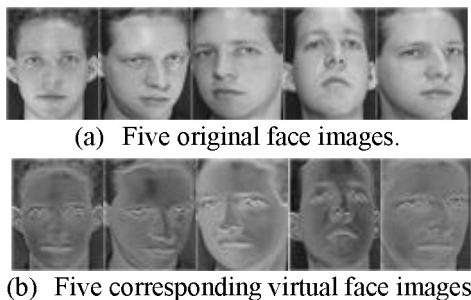(b) Five corresponding virtual face images.

Fig. 1. The original and corresponding virtual images of the first subject in the ORL database.
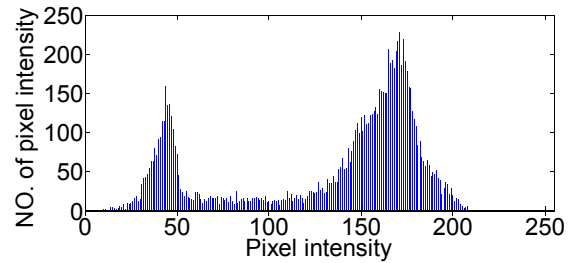

Fig. 2. The gray histogram of original image of the first sample of the first subject in the ORL database.

have clear differences in image appearance. Since original and virtual images contain different features in different aspects for a same subject, the simultaneous use of them allows the image to be better recognized.

Fig. 2 shows original pixel intensities of the first sample of the first subject in the ORL database. Fig. 3 shows pixel intensities of reconstitution images came from different methods of the same sample. From these figures, we see that most of small and large intensities in Fig. 2 are converted into small intensities (almost zero) in Fig. 3 (a). In the virtual image in Fig. 3 (a), pixel intensities mainly concentrate on central area and only a few distribute in high intensity area (such as the edges). Compared with Fig. 3 (b) and (c), most pixels in Fig. 3 (a) are concentrated on regions with moderate intensity. Though Fig. 3 (d) also has this merit, some large pixels in Fig. 3 (a) are distributed in the regions which can easily reflect image features. These illustrate that our method has more advantage in representing images.

4. Experiments and results

We conduct image classification and face recognition experiments to test our method. As shown later, the recognition accuracy obtained by using our method is satisfactory. Three databases including a non-face image database, the COIL20 database, are used in these experiments. Moreover, in order to balance the classification results of different training samples, α is set to be different value in CRC and LRC.

4.1. Experiment on the ORL database

In this section, we use the ORL database to test our method. The ORL database [37] includes 400 face images taken from 40 subjects each providing 10 face images. For some subjects, the images were taken at different times, with varying lighting, facial expressions, and facial details. Each image was resized to a 92 by 112 image matrix. They are all converted into gray images. In experiments of the ORL database, $\alpha = 0.3$ for CRC and $\alpha = 0.7$ for LRC. Fig. 4 shows examples in the ORL database.

We take the first 3, 5 and 7 images of each subject as original training samples and treat the remaining images as test samples. The experimental results on the ORL database are shown in Table 1. From it, we can see the rates of classification errors have been reduced effectively. When the

(a) The gray histogram of virtual image of our method

(b) The gray histogram of reconstitution image of CRC

(c) The gray histogram of reconstitution image of LRC

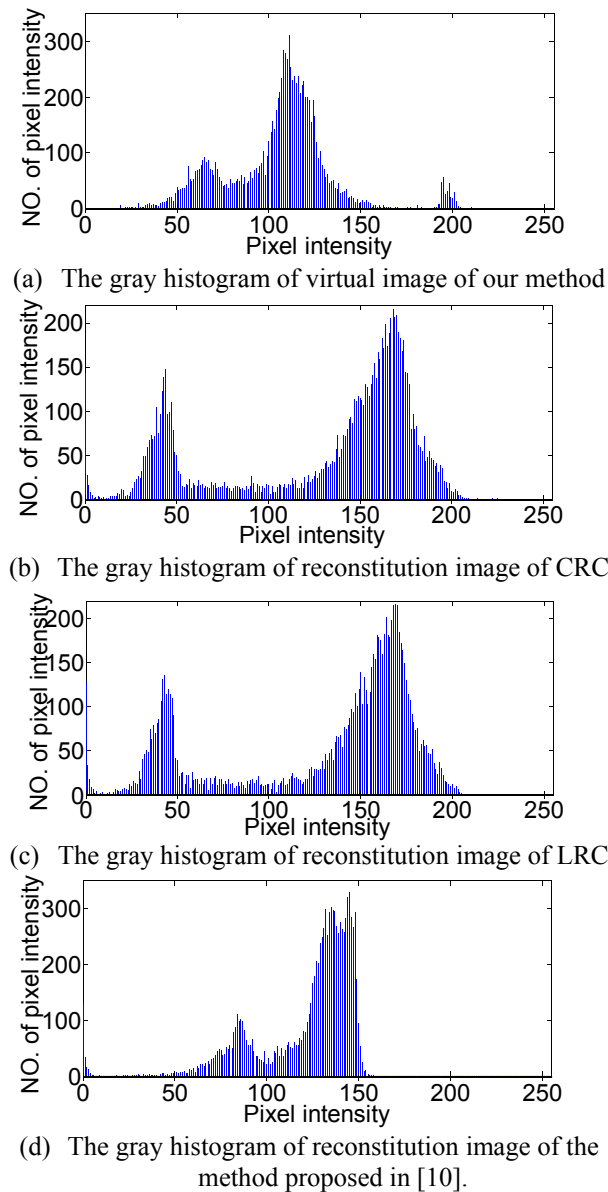(d) The gray histogram of reconstitution image of the method proposed in [10].

Fig. 3. The gray histogram of reconstitution image of different methods of the first sample of the first subject in the ORL database.

number of training samples is 3, the rates of classification errors of original CRC is 13.21% and our method's is 8.93%. That is, our method can improve 4.28% for the recognition accuracy. We also know that our method is better than the method proposed in [10].

Fig. 4. The examples in the ORL database.

Table 1
The rates of classification errors (%) of different methods on the ORL database.

The number of training samples per subject	3	5	7
CRC with original images	13.21	8.00	8.33
CRC with virtual images	13.57	9.00	6.67
The method in [10] with CRC	11.43	8.50	7.50
Our method with CRC	8.93	6.00	6.67
LRC with original images	14.29	7.00	4.17
LRC with virtual images	16.79	8.00	1.67
Our method with LRC	13.93	6.00	2.50

4.2. Experiment on the FERET database

In this section, we use FERET database to test our method. The used the FERET database [38] consists of 1400 images from 200 subjects each providing 7 images. We resized each image to a 40 by 40 image matrix. They are all converted into gray images. In experiments of the FERET database, $\alpha = 0.7$ for CRC and LRC. Fig. 5 shows examples in the FERET database.

We take the first 4 to 6 images of each subject as original training samples and treat the remaining images as test samples. Table 2 shows the rate of classification errors on the FERET database. From it, we can see the rates of classification errors have been reduced effectively. When the number of training samples is 4, the rates of classification errors of original CRC are 44.67% and our method's is 39.17%. That is, our method can improve 5.50% for the recognition accuracy. Moreover, it is better than the method proposed in [10]. This means that our method is greatly useful for representing the images.

4.3. Experiment on the COIL-20 database

In this section, we use the COIL-20 database [39] to test our method. The used database contains 360 images taken

Fig. 5. The examples in the FERET database.

Table 2
The rates of classification errors (%) of different methods on FERET database.

The number of training samples per subject	4	5	6
CRC with original images	44.67	31.50	25.50
CRC with virtual images	45.17	38.75	36.50
The method in [10] with CRC	43.83	30.25	27.50
Our method with CRC	39.17	29.00	23.00
LRC with original images	21.50	16.50	21.50
LRC with virtual images	30.67	25.00	34.50
Our method with LRC	21.77	14.00	17.50

Fig. 6. The examples in the COIL-20 database.

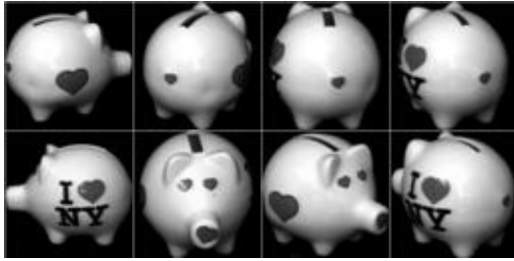Fig. 7. The image examples of two classes in the AR database.

from 20 classes and each class has 18 images. Images were taken from several angles and we take an image for every 20° for a subject. We resized each image to a 128 by 128 image matrix. They are all converted into gray images. In experiments of COIL-20 database, $\alpha = 0.3$ for CRC and LRC. Fig. 6 shows examples in the COIL-20 database.

We take the first 9, 11 and 13 images of each subject as original training samples and treat the remaining images as test samples. Table 3 shows the rates of classification errors in the COIL-20 database. From it, we can see the rates of classification errors have been reduced effectively. When the number of training samples is 11, the rates of classification errors of original CRC is 20.71% and our method's is 13.57%. That is, our method can improve 7.14% for the recognition accuracy. Compared with the recognition accuracy of the method proposed in [10], our method is better.

4.4. Experiment on the AR database

In this section, we use the AR database [40] to test our method. The AR database contains over 4000 color face images of 126 people. In this paper, we only choose 120 people of them. Hence, the used database contains 3120 images taken from 120 people and each people has 26 images. We resized each image to a 50 by 40 image matrix. In experiments of the AR database, $\alpha = 0.7$ is set for CRC and LRC. Fig. 7 shows image examples in the AR database.

We take the first 7 to 9 images of each subject as original training samples and treat the remaining images as test samples. Table 4 shows the rate of classification errors on the AR database. From it, we can see the rates of classification errors have been reduced effectively. When the number of training samples is 9, the rate of classification errors of original LRC is 42.84% and our method's is 37.21%. That is, our method can

Table 4
The rates of classification errors (%) of different methods on the AR database.

The number of training samples per subject	7	8	9
CRC with original images	29.74	30.05	31.08
CRC with virtual images	37.15	36.16	37.89
The method in [10] with CRC	28.06	27.81	28.65
Our method with CRC	27.46	27.13	28.33
LRC with original images	38.77	40.37	42.84
LRC with virtual images	40.31	41.53	43.24
Our method with LRC	34.17	35.46	37.21

improve 5.63% for the recognition accuracy. This means that our method is very helpful for image classification.

5. Conclusions

In this paper, we proposed a novel representation based classification method for image classification. This method can effectively enhance the recognition rate by fusing original and virtual images. Moreover, as original and virtual images are simultaneously used, it can improve the number of training samples for each subject and adequately exploit detail features of each target so as to improve the recognition accuracy. Compared with other algorithms which generate the virtual images, our method is extremely simple and the computation is quite efficient. The analyses and experimental results sufficiently show the rationales of the proposed method.

Acknowledgments

Table 3
The rates of classification errors (%) of different methods on the COIL-20 database.

The number of training samples per subject	9	11	13
CRC with original images	20.00	20.71	14.00
CRC with virtual images	15.56	17.14	10.00
The method in [10] with CRC	16.11	17.14	11.00
Our method with CRC	15.00	13.57	7.00
LRC with original images	11.67	8.57	8.00
LRC with virtual images	11.11	10.71	7.00
Our method with LRC	10.56	7.86	6.00

References

[1] J. Chen, S.-G. Shan, C. He, G.-Y. Zhao, M. Pietikäinen, X.-L. Chen, W. Gao, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1705−1720.
[2] X.-P. Hong, G.-Y. Zhao, M. Pietikäinen, X.-L. Chen, IEEE Trans. Image Process 23 (6) (2014) 2557−2568.
[3] A.F. Mansano, J.A. Matsuoka, Luis C.S. Afonso, João P. Papa, Fábio Augusto Faria, Ricardo da Silva Torres, Improving image

classification through descriptor combination, in: Proceedings of SIB-GRAPI, 2012, pp. 324−329.

[4] Z. Ma, J. Wen, Q. Liu, J. Mod. Opt. 62 (9) (2015) 745−753.

[5] Y. Xu, J. Yang, Z. Jin, Pattern Recognit. 36 (12) (2003) 3031−3033.

[6] L. Gan., Phys. Procedia 24 (C) (2012) 1689−1695.

[7] Y. Xu, IEEE Trans. Cybernet. 44 (10) (2013) 1738−1746.

[8] A. Eftekhari, M. Forouzanfar, H.A. Moghaddam, Inf. Process. Lett. 110 (17) (2010) 761−766.

[9] I. Naseem, R. Togneri, M. Bennamoun, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106−2112.

[10] Y. Xu, B. Zhang, Z. Zhong, Pattern Recognit. Lett. 68 (2015) 9−14.

[11] Y. Xu, X. Li, J. Yang, D. Zhang, Neurocomputing 131 (5) (2014) 191−199.

[12] W. Huang, X. Wang, Y. Ma, Opt. Eng. 54 (5) (2015) 053103.

[13] S. Wu, J. Cao, Optik 125 (2014) 3530−3533.

[14] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu, H. Liu, Pattern Recognit. 46 (2014) 1151−1158.

[15] B. Tang, S. Luo, H. Huang, High performance face recognition system by creating virtual sample, in: Proceedings of International Conference on Neural Networks and Signal Processing, 2003, pp. 972−975.

[16] H.-C. Jung, Authenticating corrupted face image based on noise model, in: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 272−277.

[17] N.P.H. Thian, S. Marcel, S. Bengio, Improving face authentication using virtual samples, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 2003, pp. III-233−III-236.

[18] Y.-S. Ryu, S.-Y. Oh, Pattern Recognit. Lett. 23 (7) (2002) 833−841.

[19] Y. Xu, X. Fang, X. Li, IEEE Trans. 44 (10) (2014) 1950−1961.

[20] T. Payne, M.C. Nicely . Non-rectangular and/or non-orthogonal arrangement of gambling elements in a gaming apparatus U.S. Patent 6,676,511[P]. 2004-1−13.

[21] T. Fukuoka, A.G. Engel, B. Lang, Ann. Neurology 22 (2) (1987) 193−199.

[22] H.P. Killackey, G. R Belford, J. Comp. Neurology 183 (2) (1979) 285−303.

[23] R.G. Hegg, M.J. Chern, A. Au. Virtual image display having a high efficiency grid beamsplitter U.S. Patent 5,383,053[P]. 1995-1−17.

[24] R.J. Hamers, R. M Tromp, J.E. Demuth, Phys. Rev. B 34 (8) (1986) 5343.

[25] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.-S. Huang, S.-C. Yan, Proc. IEEE 98 (2010) 1031−1044.

[26] Z. Zhang, Y. Xu, J. Yang, Access IEEE 3 (2015) 490−530.

[27] J. Yang, J. Wright, T. S Huang, Image Process. IEEE Trans. 19 (11) (2010) 2861−2873.

[28] J. Mairal, M. Elad, G. Sapiro, Image Process. IEEE Trans. 17 (1) (2008) 53−69.

[29] R. Rubinstein, A. M Bruckstein, M. Elad, Proc. IEEE 98 (6) (2010) 1045−1057.

[30] S. Zhao, Z. P Hu, Inf. Process. Lett. 115 (2015) 677−683.

[31] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition?, in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 471−478.

[32] I. Naseem, R. Togneri, M. Bennamoun, Pattern Recognit. 45 (1) (2012) 104−118.

[33] F. Dornaika, Y. Traboulsi, C. Hernandez, A. Assoum, Self-optimized two phase test sample sparse representation method for image classification, in: Proceedings of the 2nd International Conference on Advances in Biomedical Engineering, 2013, pp. 163−166.

[34] Y. Xu, D. Zhang, J. Yang, J.-Y. Yang, IEEE Trans. Circuits Syst. Video Technol. 25 (2011) 1255−1262.

[35] F. Dornaika, Y. Traboulsi, A. Assoum, Adaptive two phase sparse representation classifier for face recognition, in: Proceedings of Advanced Concepts for Intelligent Vision Systems, 2013, pp. 182−191.

[36] P.-C. Hsieh, P.-C. Tung, Neuro-computing 73 (13) (2010) 2708−2717.

[37] Available: http://www.cl.cam.ac.uk/research/dtg/attarchiv-e/face database. html (Online).

[38] P.J. Phillips, H. Moon, S. Rizvi, et al., Pattern Analysis Mach. Intell. IEEE Trans. 22 (10) (2000) 1090−1104.

[39] S.A. Nene, S.K. Nayar, H. Murase, Columbia object Image Library (COIL-20), Technical Report CUCS-005−96, 1996.

[40] A.M. Martinez, The AR Face Database, CVC Technical Report, 1998, p. 24.

Zhongli Ma received the M.S. and Ph.D degrees at Harbin Engineering University (China) in 2003 and 2006, respectively. She is an associate professor with the Control Science and Engineering in Harbin Engineering University. She is a visiting scholar of the University of Louisiana at Lafayette (American) and the Texas A&M University (American) in 2006−2007 and 2013−2013, respectively. She has published more than 50 journal and conference papers. Her research interests include target detecting, tracking and recognition, image and video enhancement.

Quanyong Liu is currently pursuing the M.S. degree with Control Science and Engineering at Harbin Engineering University, Harbin, China. His research interests include target tracking and recognition, image processing.

Kai Sun received his M.S. degree in Harbin Institute of Technology, Harbin, China in 2009. His research interests include background modeling, multi-object tracking, and pedestrian detection.

Sui Zhan is currently pursuing the M.S. degree with Communication Engineering at Hunan University, Changsha, China. Her research interests include target tracking and recognition, computer vision.