Weighted average integration of sparse representation and collaborative representation for robust face recognition

Shaoning Zeng¹ (✉), Yang Xiong¹

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Sparse representation is a significant method to perform image classification for face recognition. Sparsity of the image representation is the key factor for robust image classification. As an improvement to sparse representation-based classification, collaborative representation is a newer method for robust image classification. Training samples of all classes collaboratively contribute together to represent one single test sample. The ways of representing a test sample in sparse representation and collaborative representation are very different, so we propose a novel method to integrate both sparse and collaborative representations to provide improved results for robust face recognition. The method first computes a weighted average of the representation coefficients obtained from two conventional algorithms, and then uses it for classification. Experiments on several benchmark face databases show that our algorithm outperforms both sparse and collaborative representation-based classification algorithms, providing at least a 10% improvement in recognition accuracy.

Keywords sparse representation; collaborative representation; image classification; face recognition

1 Introduction

Feature extraction and classification are two key steps in face recognition [1, 2]. Extraction of features is the basic of mathematical calculation performed in classification methods. Only if sufficient and proper

features are extracted, can a classification method produce good recognition results. One prevailing paradigm is to use statistical learning approaches based on training data to determine proper features to extract and how to construct classification engines. Nowadays, many successful algorithms for face detection, alignment, and matching are learning-based algorithms. Representation-based classification methods (RBCM), such as PCA [3, 4] and LDA [5, 6], have significantly improved face recognition techniques. Such linear methods can be extended by use of nonlinear kernel techniques (kernel PCA [7] and kernel LDA [8]). The basic process in these methods is as follows: first all training samples are coded to obtain a representation matrix, then this matrix is used to evaluate each test sample and determine new lower-dimensional representation coefficients, and finally classification is performed based on these coefficients [2, 9]. Therefore, the robustness of face recognition is determined by suitability of the representation coefficients.

Sparse coding or representation has recently been proposed as an optimal representation of image samples. Sparse representation-based classification (SRC) for face recognition [2, 9, 10] first codes the test sample using a linear combination on the training samples, and then determines the differences between the test sample and all training samples using the representation coefficients. Consequently, the test sample can be classified as belonging to the class with minimal distance. SRC has been widely applied to face recognition [11–13], image categorization [14, 15], and image super-resolution [9, 16]. Indeed, SRC can be viewed as a global representation method [17], because it uses all

1 Huizhou University, Guangdong 516007, China. E-mail: S. Zeng, zsn@outlook.com (✉); Y. Xiong, xyang.2010@hzu.edu.cn.

Manuscript received: 2016-05-05; accepted: 2016-09-22

training samples to represent the test sample. On the contrary, collaborative representation-based classification (CRC), proposed as an improvement to SRC, considers the local features in common for each class in its representation. The training samples as a whole are used to determine the representation coefficients of a test sample. CRC considers the collaboration between all classes in the representation as the underlying reason it is possible to make a powerful image classification method [18–20]. However, we believe the collaborative contribution from local classes can also be used to refine the sparse representation, and that it is possible to improve the robustness of image classification by integrating both types of representation. Zhang et al. [17] integrated the globality of SRC with the locality of CRC for robust representation-based classification, and Li et al. [21] also combined sparse and collaborative representations for hyperspectral target detection with a linear operation. Further similar integrative methods have been proposed for other domains.

In this paper, we propose to use a slightly more sophisticated mathematical operation performing weighted averaging of sparse and collaborative representations for classification, which we call WASCRC. Firstly, it determines the sparse representation coefficients β for the test sample via l_1 -norm minimization on all training samples. Secondly, it determines the collaborative representation coefficients α for the same test sample via l_2 -norm minimization on all training samples. Thirdly, it calculates the new representation coefficients as a weighted average of these two groups of coefficients: $\beta' = a\alpha + b\beta$. Finally, the distance between the test sample and each training sample is determined as $\text{res}_{\text{WASCRC}} = \|y - X_i\hat{\beta}'_i\|_2 / (a + b)$, allowing the test sample to be classified as belonging to the nearest class. Usually, we can let $a = 1$ for simplicity and vary b appropriately to a specific application. We conducted various experiments on several benchmark face databases, which showed that our WASCRC algorithm could decrease the failure rate of classification by up to 17% and 26% relative to SRC and CRC respectively.

The rest content of this paper is organized as follows. Section 2 introduces related work on sparse representation for robust face recognition. Section 3

describes our proposed algorithm and the rationale behind it. Section 4 presents experimental results on several benchmark face databases. Section 5 gives our conclusions.

2 Related work

2.1 Sparse representation

The sparse representation-based classification (SRC) algorithm was proposed by Wright et al. [2]. The basic procedure involves two steps, first representing the test sample as a linear combination of all training samples, and then identifying the closest class based on the minimal deviation.

Assume that there are C subjects or pattern classes with n training samples $x_1, x_2, \dots, x_n$ and the test sample is y . Let the matrix $X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n_i}] \in I^{m \times n_i}$ denote n_i training samples from the i th class. By stacking all columns from the vector for a $w \times h$ gray-scale image, we can obtain the vector for this image: $x \in I^m$ ($m = w \times h$). Each column of A_i then represents the training images of the i th subject. Any test sample $y \in I^m$ from the same class can be described by a linear formula as

$$y = a_{i,1}x_{i,1} + a_{i,2}x_{i,2} + \dots + a_{i,n}x_{i,n} \quad (1)$$

where $a_{i,j} \in I, j = 1, 2, \dots, n_i$.

The n training samples of C subjects can be denoted by a new matrix: $X = [X_1, X_2, \dots, X_C]$. Thus, Eq. (1) can be rewritten more simply as

$$y = X\beta \in I^m \quad (2)$$

where $\beta = [0, \dots, 0, a_{i,1}, a_{i,2}, \dots, 0, \dots, 0]^T$ is the sparse coefficient vector in which only entries for the i th class are non-zero. This vector of coefficients is the key factor which affects the robustness of classification. Note that SRC uses the entire set of training samples to find these coefficients. This is a significant difference from one-sample-at-one-time or one-class-at-one-time methods such as nearest neighbor (NN) [22] and nearest subspace (NS) [23] algorithms. These local methods can both identify objects represented in the training set and reject samples that do not belong to any of the classes present in the training set.

The next step in SRC is to perform l_1 -norm minimization to solve the optimization problem to find the sparsest solution to Eq. (2). This result is

used to identify the class of the test sample y . Here we use:

$$\hat{\beta} = \arg \min_{\beta} \|\beta\|_1 \quad (3)$$

Next, SRC computes the residuals for this representative coefficient vector for the i th class:

$$\text{res}_{\text{SRC}}(y) = \|y - X_i \hat{\beta}_i\|_2 \quad (4)$$

Finally the identity of y is output as

$$\text{identity}(y) = \arg \min_i \{\text{res}_{\text{SRC},i}\} \quad (5)$$

There are five prevailing fast l_1 -minimization approaches: gradient projection, homotopy, iterative shrinkage-thresholding, proximal gradient, and augmented Lagrange multipliers (ALM) [15]. As we know, it is more efficient to use first order l_1 -minimization techniques for noisy data, e.g., SpaRSA [9], FISTA [24], and ALM [13], while homotopy [25], ALM, and l_1 -ls [26] are more suitable for face recognition because of their accuracy and speed. Other SRC algorithms are implemented using l_0 -norm, l_p -norm ($0 < p < 1$), or even l_2 -norm minimization. Xu et al. [26] exploited $l_{1/2}$ -norm minimization to constrain the sparsity of representation coefficients; further descriptions of various norm minimizations can be seen in Ref. [22]. Yang et al. [13] proposed fast l_1 -minimization algorithms called augmented Lagrangian methods (ALM) for robust face recognition. Furthermore, many researchers proposed different SRC implementations and improvements, such as kernel sparse representation by Gao et al. [15], an algorithm by Yang and Zhang [27] that uses a Gabor occlusion dictionary to significantly reduce the computational cost when dealing with face occlusion, l_1 -graphs for image classification by Cheng et al. [28], sparsity preserving projections by Qiao et al. [29], combination of sparse coding with linear pyramid matching by Yang et al. [30], and a prototype-plus-variation model for sparsity-based face recognition [31]. Classification accuracy can be further improved by using virtual samples [32–34]. All these methods attempt to improve the robustness of image classification for face recognition—it is clear that sparsity plays a paramount role in robust classification for face recognition.

2.2 Collaborative representation

Collaborative representation-based classification (CRC) was proposed as an improvement to and

replacement for SRC by Zhang et al. [18, 19] and Chen and Ramadge [20]. Much literature on SRC, including Ref. [2], overemphasizes the significance of l_1 -norm sparsity in image classification, while the role of collaborative representation (CR) is downplayed [18]. CR involves contributions from every training sample to represent the test sample y , because different face images share certain common features helpful for classification. It is thus based on nonlocal samples. CRC can use this nonlocal strategy to output more robust face recognition results.

Setting $X = [X_1, X_2, \dots, X_C] \in I^{m \times n}$, and the test sample $y \in I^m$ can be represented as

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (6)$$

Using a regularized least square approach [35] we can collaboratively represent the test sample using X with low computational burden:

$$(\hat{\alpha}) = \arg \min_{\alpha} \{\|y - X \cdot \alpha\|_2^2 + \lambda \|\alpha\|_2^2\} \quad (7)$$

where λ is a regularization parameter, which makes the least square solution stable and introduces better sparsity in the solution than using l_1 -norm. Thus, the CR in Eq. (5) now becomes:

$$\hat{\alpha} = (X^T \cdot X + \lambda \cdot I)^{-1} X^T \quad (8)$$

Let $P = (X^T \cdot X + \lambda \cdot I)^{-1} X^T$, so we can just simply project the test sample y onto P :

$$\hat{\alpha} = P \cdot y \quad (9)$$

We may then compute the regularized residuals by

$$\text{res}_{\text{CRC}} = \|y - X_i \cdot \hat{\alpha}_i\|_2 / \|\hat{\alpha}_i\|_2 \quad (10)$$

Finally, we can output the identity of the test sample y as

$$\text{identity}(y) = \arg \min_i \{\text{res}_{\text{CRC},i}\} \quad (11)$$

In this way, CRC involves all training samples to represent the test sample. We consider this collaboration to be an effective approach, giving a better sparse representation result.

3 Our method

We believe that sparse representation (SR) still makes a significant contribution to robust classification, while the real importance of collaborative representation (CR) is to refine the sparse representation but not to negate it. Recent literature has proposed novel approaches which integrate both algorithms in pursuit of more robust results. Zhang et al. [17] integrated

the globality of SRC with the locality of CRC for robust representation-based classification. In this method, integration was performed in the residual calculation in the representation, rather than in the sparse vector for the test sample. Li et al. [21] proposed a method to combine sparse and collaborative representations for hyperspectral target detection. This combination also happened at the step of computing the distance after the sparse vector had been determined. We compute a weighted average of the representation coefficients produced by SRC and CRC algorithms, as well as the computation of residuals, in an approach we call WASCRC. WASCRC works as follows. In the first stage, we obtain two kinds of coefficients from SRC and CRC. We use a conventional SRC algorithm to find the sparse representation coefficients β for the test sample, using Eq. (3). We also find the collaborative representation coefficients α using a conventional CRC algorithm, as in Eq. (8).

Next, we integrate them by means of a weighted average, denoted by $y = (ax_1 + bx_2)/(a + b)$. Our algorithm obtains new coefficients by imposing different weights on the two kinds of coefficients found by the two algorithms as follows:

$$\beta' = a\alpha + b\beta \quad (12)$$

where a and b indicate the weights of two algorithms.

Finally, we compute the residuals between the test sample and training samples with an l_2 -norm operation. Unlike conventional SRC and CRC, after performing the normalization, we need to divide by the sum of the two weights:

$$\text{res}_{\text{WASCRC}} = \|y - X_i\beta'_i\|/(a + b) \quad (13)$$

In this way, this new residual incorporates the weighted average, producing a refined solution. We can use it to identify the test sample y as

$$\text{identity}(y) = \arg \min_i \{\text{res}_{\text{WASCRC}, i}\} \quad (14)$$

In practice, we use $a = 1$ for simplicity and vary b to adjust the contribution of the two algorithms. We used two values, $b = 4$ and $b = 300$, in our experiments.

4 Results

We conducted comprehensive experiments on several mainstream benchmark face databases to compare the robustness of our WASCRC and conventional SRC and CRC algorithms. The benchmarks chosen

include ORL [36], Georgia Tech [37], and FERET [38] face databases. We ran experiments with different numbers of training samples for each face database. We now explain the samples, steps, and results for each experiment, as well as an analysis and comparison of the results. The experimental results indicate that WASCRC produces a lower classification failure rate than the SRC and CRC algorithms, reaching a 10% improvement in some cases.

4.1 Experiments on the ORL face database

The ORL face database [36] is a small database that includes only 400 face images with 10 different face images taken from 40 distinct subjects. The face images were captured under different conditions for each subject at varying time, with varying lighting, facial expressions (open or closed eyes, smiling or not smiling), and facial details (glasses or no glasses). These images were captured against a dark homogeneous background while the subjects were in an upright, frontal position. To reduce the computational complexity, we resized all face images to 56×46 pixels. Figure 1 presents the first 20 images from the ORL database.

We calculated the improvements provided by WASCRC over both SRC and CRC in each case. We set the weighted average factor $b = 1$ in these cases. The best case for SRC was the one using two training samples, in which WASCRC diminished the classification failure by 27%. The best case for CRC reached 23% reduction in failure when using five training samples. On average, the improvements to SRC and CRC by WASCRC were 17% and 18% respectively. In the one-training-sample case, which

Fig. 1 The first two subjects in the ORL face database.

is typical of real applications, WASCRC gained 1% and 17% in accuracy, respectively.

In order to further understand the cause of these improvements, we added a step to analyze the change in representation coefficients in all three algorithms. We picked a single test sample that WASCRC succeeded in classifying for which both SRC and CRC failed. We selected the first two samples for all 40 subjects as training samples, so that 80 training samples in total were used to determine the representation coefficients for the test sample. In our experiment, we found that the 214th test sample, the 6th sample for the 26th subject, was not recognized correctly by either SRC or CRC, while WASCRC succeeded in classifying it. We thus carefully analyzed the representation coefficients of this test sample, shown in Fig. 3. It is clear that

every single coefficient used by WASCRC (green) was between the values used by SRC (pink) and CRC (yellow): the new coefficients were smoother than the original ones, due to the weighted average calculation. The distance between the test sample and each class is calculated by the sum of entries for all training samples belonging to a class. We believe that if the curve is smoother, which means the values are relatively smaller and closer to zero, the resulting distances will be closer to zero and have smaller differences. On one hand, more entries close to zero produce a sparser representation vector; on the other hand, smaller differences help output a more precise comparison. Therefore, this had a positive effect on the representation and made a better sparse representation than conventional representation-based methods, leading to higher classification accuracy.

4.2 Experiments on the Georgia Tech face database

The next group of experiments used the Georgia Tech face database [37, 39]. This database has 750 face images captured from 50 individuals in several sessions; all images are in 15-color JPEG format. Each subject shows frontal and/or tilted faces as well as different facial expressions, lighting conditions, and scale. The original face images all have cluttered backgrounds, and a resolution of 640×480 pixels, which is too large for efficient representation. For the experiments, we programmatically removed the background and resized the images to 30×40 pixels to reduce computing load. However, this image preprocessing did not negatively affect on our results. Figure 4 shows the first subject with 15 samples in the face database. It is not necessary to use all three dimensions of color data in these colored images: we only used two dimensions of gray-scale data from these RGB images. Again, this did not affect our experimental results.

We successively picked the first 1 to 10 face images as training samples and the rest as test samples. In this case, we set the weighted average factor $b = 300$ and recorded the classification results for all test samples given by all three algorithms. The resulting failure rates are shown in Fig. 5. The results from the SRC algorithm (blue) unexpectedly outperformed the CRC algorithm (green), while our WASCRC algorithm (yellow) not surprisingly gave best results.

Fig. 2 Classification results for all 3 algorithms for the ORL face database.

Fig. 3 Coefficients determined by the 3 algorithms for the ORL face database.

Fig. 4 The first subject, with 15 samples in the Georgia Tech face database.

Fig. 5 Classification results for all 3 algorithms for the Georgia Tech face database.

Furthermore, as the number of training samples increased, the failure rates dropped dramatically.

We again determined the improvement by WASCRC over SRC and CRC, which were slightly lower than those for the ORL face database. The conventional SRC algorithm still performed well and the best case over SRC generated only 4% improvement when using 7 training samples. As conventional CRC algorithm underachieved, WASCRC outperformed it by up to 20% when using 8 training samples. On average, WASCRC outperformed SRC and CRC by 2% and 11% respectively. WASCRC outperformed CRC by 9% in the one-training-sample case.

4.3 Experiments on the FERET face database

The last group of experiments was performed on one of the largest public benchmark face databases, the FERET database [38]. It is much bigger than the Georgia Tech and ORL face databases. Each subject has five to eleven images with two frontal views (fa and fb) and one more frontal image with a different

facial expression. Our experiments used 200 subjects in total, with 7 samples for each. Figure 7 shows the first three subjects in the database; images 1–7 belong to the first subject, while 8–14 belong to the second subject and image 15 belongs to the third subject (who has 6 more images not shown in the figure).

We used images 1–5 as training samples and the remaining images as test samples, and again set the weighted average factor $b = 300$. The resulting classification failure rates from all three algorithms are shown in Fig. 8. As for the experiments on other databases, our WASCRC algorithm (yellow) still outperformed both SRC (blue) and CRC (green) algorithms in all test cases. Even in the one-training-sample case, WASCRC also produced the highest classification accuracy.

WASCRC outperformed SRC and CRC by up to 26% and 49% respectively when using 5 training samples. On average, WASCRC outperformed SRC and CRC by 12% and 26% respectively. WASCRC outperformed SRC and CRC by 5% and 10% in the one-training-sample case.

Fig. 6 Coefficients determined by the 3 algorithms for the Georgia Tech face database.

Fig. 7 The first fifteen face images from the FERET face database.

Fig. 8 Classification results for all 3 algorithms for the FERET face database.

Fig. 9 Coefficients determined by the 3 algorithms for the FERET face database.

The representation coefficients used by WASCRC were always smoother, as shown in Fig. 9. The weighted average operation worked well, as expected. Figure 9 only shows the first half of all 200 training samples, from one test sample classified correctly by WASCRC and incorrectly by both SRC and CRC. This result validates that our proposed algorithm is a more robust classifier.

5 Conclusions

Sparsity of a representation is the key to successful sparse representation-based classification, while collaboration from all classes in the representation is the key to promising collaborative representation-based classification. We have shown how to integrate these approaches in a method that performs a weighted average operation on sparse

and collaborative representations for robust face recognition. Such integration can lower the failure rate in face recognition. Our experiments demonstrated that our new approach can outperform both sparse and collaborative representation-based classification algorithms for face recognition, decreasing the recognition failure rate by about 10%. It is possible to achieve higher accuracy still in some specific cases by altering the factor used for weighted averaging.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61502208), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20150522), the Scientific and Technical Program of City of Huizhou (Grant No. 2012-21), the Research Foundation of Education Bureau of Guangdong Province of China (Grant No. A314.0116), and the Scientific Research Starting Foundation for Ph.D. in Huizhou University (Grant No. C510.0210).

References

- [1] Brunelli, R.; Poggio, T. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 15, No. 10, 1042–1052, 1993.
- [2] Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 31, No. 2, 210–227, 2009.
- [3] Xu, Y.; Zhang, D.; Yang, J.; Yang, J.-Y. An approach for directly extracting features from matrix data and its application in face recognition. *Neurocomputing* Vol. 71, Nos. 10–12, 1857–1865, 2008.
- [4] Turk, M.; Pentland, A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* Vol. 3, No. 1, 71–86, 1991.
- [5] Park, S. W.; Savvides, M. A multifactor extension of linear discriminant analysis for face recognition under varying pose and illumination. *EURASIP Journal on Advances in Signal Processing* Vol. 2010, 158395, 2010.
- [6] Lu, J.; Plataniotis, K. N.; Venetsanopoulos, A. N. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks* Vol. 14, No. 1, 195–200, 2003.
- [7] Debruyne, M.; Verdonck, T. Robust kernel principal component analysis and classification. *Advances in Data Analysis and Classification* Vol. 4, No. 2, 151–167, 2010.
- [8] Muller, K.-R.; Mika, S.; Ratsch, G.; Tsuda, K.; Scholkopf, B. An introduction to kernel-based learning

- algorithms. *IEEE Transactions on Neural Networks* Vol. 12, No. 2, 181–201, 2001.
- [9] Yang, J.; Wright, J.; Huang, T. S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* Vol. 19, No. 11, 2861–2873, 2010.
 - [10] Xu, Y.; Zhang, D.; Yang, J.; Yang, J.-Y. A two-phase test sample sparse representation method for use with face recognition. *IEEE Transactions on Circuits and Systems for Video Technology* Vol. 21, No. 9, 1255–1262, 2011.
 - [11] Zhong, D.; Zhu, P.; Han, J.; Li, S. An improved robust sparse coding for face recognition with disguise. *International Journal of Advanced Robotic Systems* Vol. 9, 126, 2012.
 - [12] Xu, Y.; Zhu, Q.; Zhang, D. Combine crossing matching scores with conventional matching scores for bimodal biometrics and face and palmprint recognition experiments. *Neurocomputing* Vol. 74, No. 18, 3946–3952, 2011.
 - [13] Yang, A. Y.; Zhou, Z.; Balasubramanian, A. G.; Sastry, S. S.; Ma, Y. Fast l_1 -minimization algorithms for robust face recognition. *IEEE Transactions on Image Processing* Vol. 22, No. 8, 3234–3246, 2013.
 - [14] Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Discriminative learned dictionaries for local image analysis. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2008.
 - [15] Gao, S.; Tsang, I. W.-H.; Chia, L.-T. Kernel sparse representation for image classification and face recognition. In: *Computer Vision—ECCV 2010*. Daniilidis, K.; Maragos, P.; Paragios, N. Eds. Springer Berlin Heidelberg, 1–14, 2010.
 - [16] Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, 2008.
 - [17] Zhang, Z.; Li, Z.; Xie, B.; Wang, L.; Chen, Y. Integrating globality and locality for robust representation based classification. *Mathematical Problems in Engineering* Vol. 2014, Article No. 415856, 2014.
 - [18] Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In: *Proceedings of IEEE International Conference on Computer Vision*, 471–478, 2011.
 - [19] Zhang, L.; Yang, M.; Feng, X.; Ma, Y.; Zhang, D. Collaborative representation based classification for face recognition. *arXiv preprint arXiv:1204.2358*, 2012.
 - [20] Chen, X.; Ramadge, P. J. Collaborative representation, sparsity or nonlinearity: What is key to dictionary based classification? In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 5227–5231, 2014.
 - [21] Li, W.; Du, Q.; Zhang, B. Combined sparse and collaborative representation for hyperspectral target detection. *Pattern Recognition* Vol. 48, No. 12, 3904–3916, 2015.
 - [22] Feng, Q.; Pan, J.-S.; Yan, L. Restricted nearest feature line with ellipse for face recognition. *Journal of Information Hiding and Multimedia Signal Processing* Vol. 3, No. 3, 297–305, 2012.
 - [23] Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 35, No. 11, 2765–2781, 2013.
 - [24] Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* Vol. 2, No. 1, 183–202, 2009.
 - [25] Zhang, Z.; Xu, Y.; Yang, J.; Li, X.; Zhang, D. A survey of sparse representation: Algorithms and applications. *IEEE Access* Vol. 3, 490–530, 2015.
 - [26] Xu, Z.; Zhang, H.; Wang, Y.; Chang, X.; Liang, Y. $L_{1/2}$ regularization. *Science China Information Sciences* Vol. 53, No. 6, 1159–1169, 2010.
 - [27] Yang, M.; Zhang, L. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In: *Computer Vision—ECCV 2010*. Daniilidis, K.; Maragos, P.; Paragios, N. Eds. Springer Berlin Heidelberg, 448–461, 2010.
 - [28] Cheng, B.; Yang, J.; Yan, S.; Fu, Y.; Shuang, T. S. Learning with l^1 -graph for image analysis. *IEEE Transactions on Image Processing* Vol. 19, No. 4, 858–866, 2010.
 - [29] Qiao, L.; Chen, S.; Tan, X. Sparsity preserving projections with applications to face recognition. *Pattern Recognition* Vol. 43, No. 1, 331–341, 2010.
 - [30] Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1794–1801, 2009.
 - [31] Deng, W.; Hu, J.; Guo, J. In defense of sparsity based face recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 399–406, 2013.
 - [32] Xu, Y.; Zhu, X.; Li, Z.; Liu, G.; Lu, Y.; Liu, H. Using the original and ‘symmetrical face’ training samples to perform representation based two-step face recognition. *Pattern Recognition* Vol. 46, No. 4, 1151–1158, 2013.
 - [33] Xu, Y.; Zhang, Z.; Lu, G.; Yang, J. Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recognition* Vol. 54, 68–82, 2016.
 - [34] Liu, Z.; Song, X.; Tang, Z. Fusing hierarchical multi-scale local binary patterns and virtual mirror samples to perform face recognition. *Neural Computing and Applications* Vol. 26, No. 8, 2013–2026, 2015.
 - [35] Lee, H.; Battle, A.; Raina, R.; Ng, A. Y. Efficient sparse coding algorithms. In: *Proceedings of Advances in Neural Information Processing Systems*, 801–808, 2006.
 - [36] AT&T Laboratories Cambridge. The database of faces. 2002. Available at <http://www.cl.cam.ac.uk/research/>

- dtg/attarchive/facedatabase.html.
- [37] Computer Vision Online. Georgia Tech face database. 2015. Available at <http://www.computervisiononline.com/dataset/1105138700>.
 - [38] NIST Information Technology Laboratory. Color FERET database. 2016. Available at <https://www.nist.gov/itl/iad/image-group/color-feret-database>.
 - [39] Xu, Y.; Zhang, B.; Zhong, Z. Multiple representations and sparse representation for image classification. *Pattern Recognition Letters* Vol. 68, 9–14, 2015.

Shaoning Zeng received his M.S. degree in software engineering from Beihang University, China, in 2007. Since 2009, he has been a lecturer at Huizhou University, China. His current research interests include pattern recognition, sparse representation, image recognition, and neural networks.

Xiong Yang received his B.S. degree in computer science and technology from Hubei Normal University, China, in 2002. He received his M.S. degree in computer science from Central China Normal University, China, in 2005 and Ph.D. degree from the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and

Technology, China, in 2010. Since 2010, he has been teaching in the Department of Computer Science and Technology, Huizhou University, China. His current research interests include pattern recognition and machine learning.

Open Access The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.