

Sample group and misplaced atom dictionary learning for face recognition

Meng WANG^{1,2}, Zhengping HU^{1,*}, Zhe SUN¹, Mei ZHU², Mei SUN²

¹Department of Information Science & Engineering, Faculty of Electronics & Communication, Yanshan University, Qinhuangdao, P.R. China

²Department of Physics & Electronics Engineering, Faculty of Electronics & Communication, Taishan University, Tai'an, P.R. China

Received: 04.02.2017

Accepted/Published Online: 01.06.2017

Final Version: 05.10.2017

Abstract: Latest research results have demonstrated the effectiveness of both sparse (or collaborative) representation and dictionary learning for problem solving in face recognition and other signal classification. Considering the fact that an informative dictionary helps a lot in sparse coding, a novel model that consists of group dictionary learning and high-quality joint kernel collaborative representation was proposed in this paper, where rich information from original and virtual space was mined and constructed as a sample group space to improve classification accuracy. Meanwhile, joint kernel collaborative representation with an ℓ_2 -regularization-based classifier was used to capture more nonlinear structure and minimize the time cost. Experiments showed that the proposed method outperformed several similar state-of-the-art methods in terms of accuracy and computational complexity.

Key words: Sample group, misplaced atom, dictionary learning, joint kernel collaborative representation, face recognition

1. Introduction

As an emerging convenient technology, face recognition (FR) advances fast along with the latest development of intelligent devices and computer vision. It exists in every aspect of our life owing to its touch-free and easily acquired features, such as blink face payment, intelligent access control, attendance management, augmented reality, behavior prediction, and suspect tracking [1–3]. Numerous algorithms have been proposed for the goal of having our identity and actions comprehended by machines in a human way. However, due to various appearance changes (e.g., expression, pose, illumination, occlusion, makeup, and age) in real-world situations, FR has been constantly confronted with more and more challenges and difficulties in both single-view learning and multiview learning [4].

Recently, researchers are particularly interested in sparse coding, dictionary learning, and deep learning (DL) [5–7] for FR. Although DL has made some achievements that show promise, it is still limited by hardware requirements such as large-scale training sets and mass storage, and it lacks a characteristic model for FR. This is not the case for sparse coding, which can still obtain satisfactory results without the aforementioned restrictions. With this in mind, Wright et al. proposed sparse representation-based classification (SRC) [8], which made use of ℓ_1 -minimization techniques for FR and obtained satisfying success in occlusion and noise environments. Later SRC inspired many excellent sparse representation-based classification methods. Yang et al. proposed robust sparse coding [9], which sought the maximum likelihood estimation solution for sparse regression problems.

*Correspondence: hzp_ysu@163.com

Along with their theory, they presented a robust regularized coding (RRC) [10], which can efficiently explore outliers and reconstruct intact face images. Later they further came up with a structured RRC [11], which added structured information into the RRC. Low-rank and sparsity constraints were also combined to mine discriminative components of facial images by Zhang et al. [12]. Deng et al. presented extended SRC [13] and superposed SRC [14] to deal with face recognition under uncontrolled conditions by considering the prototype and variation separately. Despite the wide use of these vibrant studies based on ℓ_1 -norm sparse representation, researchers have still brought new ideas to keep improving. Zhang et al. innovatively proposed collaborative representation-based classification with regularized least squares (CRC_RLS) [15] and showed that collaboration between classes could give similar FR results to ℓ_1 -regularization, but this could significantly reduce time cost with a squared ℓ_2 -regularization. Since then the CR model has also made significant progress as an SR model.

Furthermore, kernel methods have seen wide applications in support vector machines [16], principal component analysis (PCA) [17], linear discriminant analysis [18], locality-preserving projection [19], SRC, and CRC_RLS to enhance the ability of linear feature extraction and classification, as they can capture the nonlinear structure of samples (e.g., face images) effectively. Kernel sparse representation [20] was proposed for FR and object classification and it performed superiorly against other linear sparse representation algorithms, but it is a time-consuming process due to the use of complicated ℓ_1 -optimization techniques. Therefore, Yang et al. contributed a kernel collaborative representation (KCR) [21,22] model, which extended the CR [15] model to a kernel space and minimized time cost. Liu et al. constructed a locality-constrained dictionary for KCR [23] that used a simple yet effective unified similarity measure to enhance the robustness of the approach. A KCR algorithm with squared ℓ_2 -regularization presented by Wang et al. obtained favorable results, where local binary patterns (LBP) as well as the Hamming kernel [24] were applied for FR. Meanwhile, efforts have been made in combining different weights as different kernel functions.

In this paper, a sample group and misplaced atom dictionary for a joint kernel collaborative representation (SGMA_JKCR)-based classification model is presented. The main contributions of this paper are as follows:

1. The sample groups, which were composed of original and virtual (affine-transformed and mirror symmetry) samples, were constructed and combined as a misplaced dictionary in a certain order;
2. To benefit classification accuracy and minimize the time cost, a joint kernel collaborative representation-based classification method with a squared ℓ_2 -regularization was applied;
3. The effectiveness of the misplaced atom dictionary for classification accuracy of our model was evaluated by an instance.

Our method was a practical model when the training sets were of small size and defective features. To evaluate our algorithm, we conducted numerous experiments on the Georgia Tech (GT) [25], Labeled Faces in the Wild (LFW) [26], and Caltech face databases [27]. Experimental results show that our approach achieves competitive results against some similar state-of-the-art methods and is worthy of further study.

The rest of this paper is organized as follows. In Section 2, we provide a brief review of related works. The main principle of the proposed approach is described in Section 3. The extensive experimental results of the face recognition problem are described in Section 4. Finally, conclusions are drawn in Section 5.

2. Brief review of related works

In this section, we summarize and review the related collaborative representation-based classification (CRC) methods and put forward our work straightforwardly. Suppose that there are L classes of subjects and $A = [A_1, A_2, \dots, A_L] \in R^{M \times N}$ is the set of original training samples, where $A_i = [x_{i,1}, x_{i,2}, \dots, x_{i,N_i}]$ is the set of training samples from class i ($i = 1, 2, \dots, L$), $x_{i,n}$ represents the n th sample from the i th class, M is the dimension of training samples, and $N = \sum_{i=1}^L N_i$ is the number of training samples from L classes. For a given valid testing sample y , we can collaboratively code it over A by a linear representation $y \approx A\alpha$, where $\hat{\alpha} = [\hat{\alpha}_1; \hat{\alpha}_2; \dots; \hat{\alpha}_L]$ and $\hat{\alpha}_i$ is the coding coefficients vector associated with the i th class.

2.1. Collaborative representation-based classification with regularized least squares (CRC_RLS)

In [15], it was found that CRC_RLS had a performance as good as SRC but with excellent running speed owing to the application of the regularized least squares method with ℓ_2 -minimization, as in Eq. (1):

$$\hat{\alpha} = \arg \min_{\alpha} \{ \|y - A\alpha\|_2 + \lambda \|\alpha\|_2^2 \}, \quad (1)$$

where λ is a trade-off parameter to balance the sparsity and fidelity. The regularized residuals e_i of each i th class can be computed and used to classify y as in Eqs. (2) and (3):

$$e_i = \|y - A_i \hat{\alpha}_i\|_2 / \|\hat{\alpha}_i\|_2, \quad (2)$$

$$Identify(y) = \arg \min_i \{ e_i \}. \quad (3)$$

This simple yet effective CR technique had shown very powerful classification capability, so it inspired a lot of later CR works using ℓ_2 -regularization constraint.

2.2. Kernel collaborative representation (KCR)

KCR [24] extended the CR model to kernel space in a nonlinear way, which performed more favorably than other sparse representation techniques with ℓ_1 -optimization in terms of accuracy and time cost. In addition, different kernel functions had been selected to improve the classification performance, such as single kernel functions (e.g., linear, log, Gaussian, Laplacian, polynomial, perceptron, Hamming kernel) and multiple kernel functions (e.g., MKCR). The whole framework of the KCR is as in Eq. (4):

$$\hat{\alpha} = \arg \min_{\alpha} \{ \|\phi(y) - \phi(A)\alpha\|_2 + \lambda \|\alpha\|_2^2 \}, \quad (4)$$

where $\phi(y)$ and $\phi(A)$ are mapped high-dimensional features of test sample y and dictionary A by the kernel function $\kappa(u, v) = \langle \phi(u), \phi(v) \rangle = \phi(u)^T \phi(v)$.

The analytical and optimized solution of Eq. (4) can be denoted as in Eq. (5):

$$\hat{\alpha} = (K_{AA} + \lambda I)^{-1} K_A(y), \quad (5)$$

where K_{AA} is an $N \times N$ matrix with $[K_{AA}]_{ij} = \kappa(x_i, x_j)$ and $K_A(y)$ is an $N \times 1$ vector with $[K_A(y)]_i = \kappa(x_i, y)$.

After obtaining collaborative sparse coefficients α , y can be appropriately classified by computing minimum regularized reconstruction residuals as in Eq. (6):

$$r_i = \|\phi(y) - \phi(A_i)\alpha_i\| / \|\alpha_i\|_2 \quad (6)$$

We know that a sparser α is more able to represent mapped test sample $\phi(y)$ over mapped dictionary $\phi(A)$. In other words, $\|\phi(y) - \phi(A_i)\alpha_i\|$ should be small and α_i should be large if y belongs to class i , so the regularized residual r_i is more effective and reasonable to benefit satisfactory classification by considering both the reconstruction error and the energy of the sparse coefficient.

2.3. Affine transform model

In [28], Wang et al. proposed a novel likelihood function:

$$p(Y^i | x^i) = \exp[-(\|W^i \odot (Y^i - UZ^iV^T)\|_F^2 + \beta \sum (1 - W^i))] \quad (7)$$

where $p(Y | x)$ is used for estimating the hidden state $x_t = \{x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t\}$ (i.e. affine transform model parameters: translations, rotation angle, scale, aspect ratio, and skew respectively), Y is the test set, W reflects the nonzero elements of error matrix $E = Y - UZV^T$ (computed by PCA), and \odot and β denote the Hadamard product and penalty term. This model was applied to track objects in most cases, but it could also align images well due to affine transform participation.

3. The proposed pattern model

The proposed model is composed of two parts: an extended SGMA dictionary combination and joint kernel structure formation. Figure 1 explains the principle and implementation of the model ($E_i (i = 1, 2, 3)$ is the error matrix of Y_i in Figure 1). We can see that E_1 is very close to zero, but both E_2 and E_3 have dense representation because of misaligned facial images. This may lead to misclassification. To obtain minimum error matrix images, we used an affine transform model to get transformed (virtual) images in an adaptive affine transform parameter way. Then sample groups (each of which contains 1 original, 1 affine-transformed, and 2 mirror symmetrical facial images) were used to form the group dictionary for the purpose of obtaining more useful features of samples when the training set is small. Meanwhile, a joint kernel function was applied to optimize the single kernel structure and enhance the robustness of the classifier.

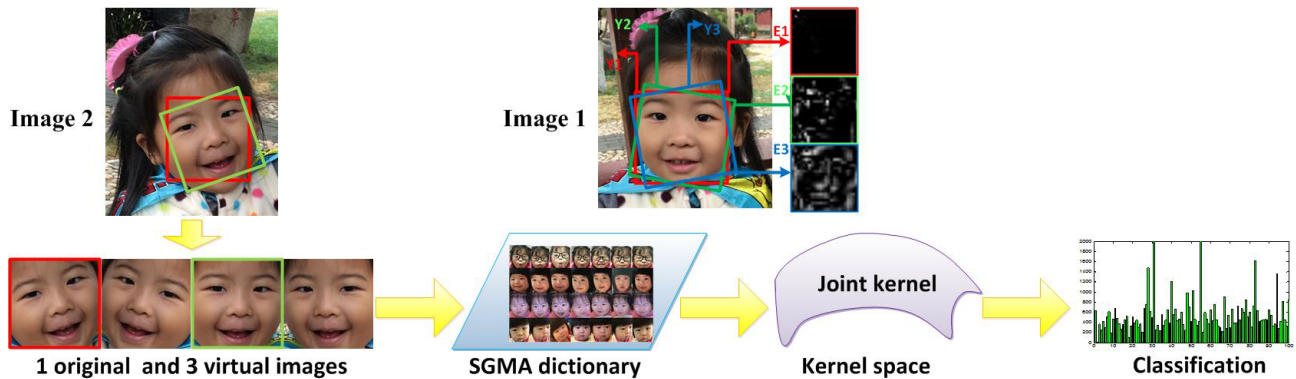


Figure 1. The principle and implementation of our method.

3.1. Sample group and misplaced atom (SGMA) dictionary

The SGMA dictionary is a comprehensive and regrouped dictionary. First, we formed four kinds of image sets: an original (ORI) facial images set and their mirror (ORIM) images set, an affine-transformed (AFT) images set, and a corresponding mirror (AFTM) images set. Then we regrouped the above four training sets to form the SGMA dictionary, as shown in Figure 2. In this process, the atoms of each set were misplaced in a certain order, which not only expands the size but also enhances the ability of the dictionary. Thus, our dictionary with rich variation information can yield desired classification results.

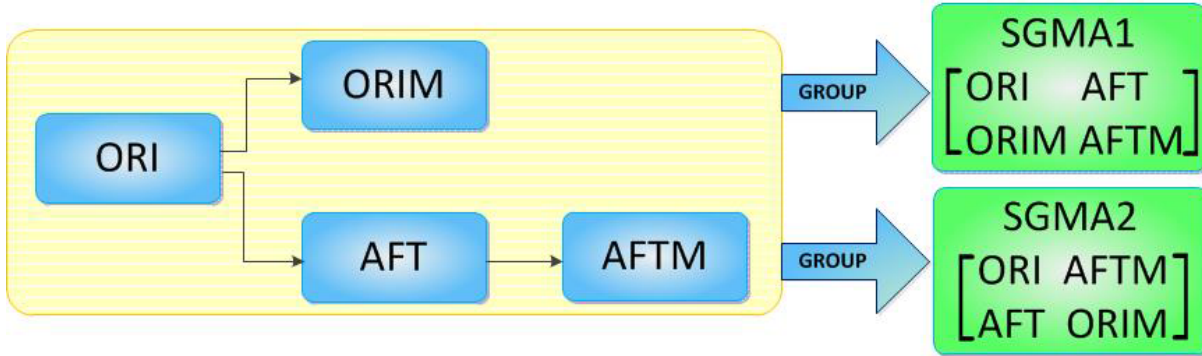


Figure 2. The procedure of SGMA dictionary formation.

We mainly used two regroup dictionaries, the SGMA1 and SGMA2 (S1 and S2) dictionaries, to proceed with our following work. Letting each class have n facial images for training in each image set, $x_{i,p}^{ORI}, x_{i,p}^{AFT} \in \mathbb{R}^{M \times 1}$ represents the p th sample from the i th class of the ORI and AFT sets, $x_{i,q}^{ORIM}, x_{i,q}^{AFTM} \in \mathbb{R}^{M \times 1}$ represents the q th sample from the i th class of the ORIM and AFTM sets, M is the dimension of training samples, and $x_i^{SGMA1}, x_i^{SGMA2} \in \mathbb{R}^{2M \times 2n^2}$ represents the samples from the i th class of SGMA1 and SGMA2 dictionaries. With different values of p and q , two SGMA dictionaries were obtained in the misplaced atom scheme, as in Eqs. (8) and (9) :

$$x_i^{SGMA1} = \begin{bmatrix} x_{i,p}^{ORI} & x_{i,p}^{AFT} \\ x_{i,q}^{AFTM} & x_{i,q}^{ORIM} \end{bmatrix} \quad (1 \leq p, q \leq n) \quad , \quad (8)$$

$$x_i^{SGMA2} = \begin{bmatrix} x_{i,p}^{ORI} & x_{i,p}^{AFTM} \\ x_{i,q}^{AFT} & x_{i,q}^{ORIM} \end{bmatrix} \quad (1 \leq p, q \leq n) \quad . \quad (9)$$

In this way, the test set was also regrouped to cooperate with the SGMA dictionaries, only with the ORI and AFT sets in the test group.

3.2. Joint kernel collaborative representation (JKCR) model

The proposed model derived from the KCR, which improved the discrimination performance of CRC_RLS in a high-dimensional kernel space and reduced the time cost comparing with other ℓ_1 -regularization based methods (e.g., KSRC). However, it was different from the conventional KCR due to the application of the joint kernel function, which had been demonstrated as effective in many works. For example, Yang et al. generalized a multiple kernel collaborative representation (MKCR) [29], where three multiple kernel ideas were adopted in the MKCR model. We also attempted to use different weights to combine different kernel functions.

Here two kernel functions are applied: Gaussian kernel $\kappa_G(x, y) = \exp(-\|x - y\|^2/t)$ and Hamming kernel $\kappa_H(x, y) = 1 - \frac{1}{mN} \sum_{i=1}^m D(x_i, y_i)$, where x and y are two sample sequences, t is a variable parameter, m is the number of image pixels, N is the length of the sequence, x_i and y_i are the i th pixels of two samples, and $D(\cdot, \cdot)$ denotes the Hamming distance between two sequences. Then a simple yet effective joint kernel function framework is presented:

$$\kappa = w \times \kappa_G + (1 - w) \times \kappa_H \quad , \quad (10)$$

where w is the weight parameter and can be tuned automatically in practical use. The joint kernel function can optimize the kernel structure property and improve classification accuracy to some extent.

4. Experimental results

In this section, we conduct experiments on popular face databases to evaluate the performance of our method. All the experiments were carried out using MATLAB 2013 on a 3.70 GHz computer with 4.0 GB RAM. First, we compare test results with the conventional CRC_RLS [15], RRC_L₂ [10], KCRC (Gaussian), KCR (Hamming) [21,22], and KCR- ℓ_2 (LBP + Hamming (LH)) [24] to demonstrate the excellent performance of our method on different databases in Sections 4.1–4.3. Then a similar instance is analyzed and compared with our method on different databases in Section 4.4. Finally, advantages of the misplaced scheme are expounded in Section 4.5. It is worth mentioning that there is no relevance between all the experiments due to the random selection of training samples.

In order to improve algorithm efficiency, we carefully adjusted the regularization parameters, which had small positive values in our algorithm. Parameters λ were set as 0.005 in our method, CRC_RLS, KCRC (Gaussian), KCR- ℓ_2 (LH), and RRC_L₂ uniformly. Other parameters of RRC_L₂ continued to be used as their original best versions.

4.1. Georgia Tech (GT) database

The GT database contains 750 frontal face images of 50 people with irregular poses and expressions, and the size of images is cropped to 20×20 . A random subset with 3, 4, or 5 samples per person were picked with their labels as original training samples, and the rest forms the corresponding original test samples. Then we established the SGMA (S1, S2) dictionary and test group sets. Here we compared the recognition error rates (RERs) of different methods using the original (ORI) dictionary and group dictionary (S1, S2), as listed in Table 1. We can see that our method performs better than other comparative methods. The RERs of our method are the smallest over all three dictionaries, even 14% lower than CRC_RLS when training number (TRN) is 3 on the SGMA1 dictionary. The performance of the two dictionaries (S1, S2) is similarly good, and both are better than the ORI dictionary. Meanwhile, our JKCR method is effective from the aspect of the results on the same dictionary.

4.2. Labeled Faces in the Wild (LFW) database

The whole LFW database contains more than 13,000 images of faces collected from the web with irregular poses, expressions, and illumination. We used a subset of 1000 images from 100 persons (10 random images per person), and the size of images was cropped to 20×20 . Three, 4, or 5 samples per person were picked with their labels as original training samples, and the rest formed the corresponding original test samples. Then the

Table 1. RERs (%) of different methods using different dictionaries on the GT database.

Methods \ TRN	3			4			5		
	ORI	S1	S2	ORI	S1	S2	ORI	S1	S2
CRC_RLS	46.83	35.83	33.67	41.64	30.73	27.27	38.80	27.40	23.60
RRC_L ₂	49.17	31.33	29.50	43.27	27.45	24.36	38.60	21.00	22.00
KCR (Gaussian)	44.67	28.33	28.50	40.00	23.45	22.18	37.20	20.60	21.40
KCR (Hamming)	43.17	22.67	23.50	39.82	18.18	18.36	35.60	17.00	16.40
KCR- ℓ_2 (LH)	44.26	23.37	24.67	41.25	18.92	18.36	36.20	18.30	17.60
Our method	41.50	21.83	23.17	37.82	17.82	16.73	33.00	16.20	14.40

SGMA dictionary and test group sets were constructed and the results are shown in Table 2. We can see that our method is still stable and leading, and it obtains similar good results on SGMA1 and SGMA2. In addition, the RERs reduced with the training sample increasing, even being below 47% on SGMA2 in the 5 training sample test.

Table 2. RERs (%) of different methods using different dictionaries on the LFW database.

Methods \ TRN	3			4			5		
	ORI	S1	S2	ORI	S1	S2	ORI	S1	S2
CRC_RLS	65.57	63.14	63.29	63.33	58.67	57.33	59.40	55.80	53.80
RRC_L ₂	64.57	64.57	63.29	62.83	58.83	58.83	60.20	55.40	56.20
KCR (Gaussian)	66.43	58.14	58.29	62.33	53.33	53.67	59.00	50.60	49.60
KCR (Hamming)	66.71	60.00	60.00	62.33	55.17	55.00	58.40	52.20	52.20
KCR- ℓ_2 (LH)	64.57	57.99	58.22	59.72	53.00	53.00	57.60	49.50	49.40
Our method	62.86	55.43	55.14	58.33	48.83	49.17	51.20	45.60	45.00

4.3. Caltech face database

There are 450 frontal face images of 27 persons with irregular poses and expressions in the Caltech face database, and the size of images is cropped to 20×20 . A random subset with 1, 2, or 3 samples per person was picked with their labels as original training samples, and the rest formed the corresponding original test samples. Then we formed the SGMA dictionary and test group set. From Table 3, it can be seen that our method still beats the other algorithms. The results were satisfying with the TRN increasing, and the best RER was achieved at less than 2%. Meanwhile, we can conclude that useful features of samples are increased by the effective group dictionary, and our method does well under the “small training size (STS)” condition, even for a single sample size (SSS).

Table 3. RERs (%) of different methods using different dictionaries on the Caltech face database.

Methods \ TRN	1			2			3		
	ORI	S1	S2	ORI	S1	S2	ORI	S1	S2
CRC_RLS	14.85	8.66	9.9	7.14	3.44	3.97	4.55	2.27	2.56
RRC_L ₂	16.09	9.16	8.91	9.26	6.35	5.56	7.39	3.12	2.84
KCR (Gaussian)	14.36	7.92	8.66	7.41	3.7	3.97	4.83	2.27	2.27
KCR (Hamming)	16.83	11.88	11.39	13.49	6.88	7.41	9.37	5.68	5.4
KCR- ℓ_2 (LH)	15.52	10.38	9.16	11.54	5.22	5.56	8.03	4.17	4.36
Our method	13.61	6.93	7.43	7.67	3.7	3.97	4.26	2.27	1.99

4.4. Comparison of similar approaches

Xu et al. proposed a method that used the original facial image and symmetrical facial image to improve recognition accuracy due to the symmetrical structure of common human faces [30]. Our method was inspired by theirs (simply, Xu’s) with many improvements. In particular, we added affine-transformed facial images and corresponding symmetrical facial images, and we combined these four kinds of facial images in a particular order as the SGMA. Table 4 summarizes the advantage of our method over three databases. For example, if we used Xu’s dictionary and our classified method, our result was 13% better than Xu’s result, or, we used our SGMA1 dictionary and classifier, our result was almost 26% lower than theirs on the GT database. It can be seen that our dictionary and classification method have more competitive advantage than Xu’s, especially under the STS condition.

Table 4. RERs (%) of different methods using different dictionaries on (a) GT (TRN = 3); (b) LFW (TRN = 5); (c) Caltech (TRN = 1) face databases.

Scenarios	Dictionary		Classification methods		RERs		
	Xu’s	SGMA1	Xu’s	Ours	(a)	(b)	(c)
1	√	×	√	×	47.67	64.40	22.28
2	×	√	√	×	41.83	58.30	18.07
3	√	×	×	√	34.17	48.60	8.17
4	×	√	×	√	21.83	46.60	6.93

4.5. Misplaced atom scheme evaluation

In order to address the distinctiveness of different samples, we applied the misplaced atom scheme when regrouping the SGMA dictionary. Next, we evaluated the effectiveness of our scheme under different dimensions (256, 400, 1024, 4096) on the GT and LFW databases when $TRN = 5$, as shown in Figure 3. Compared with the no-misplaced atom dictionary, our misplaced atom dictionary has a good performance in classification

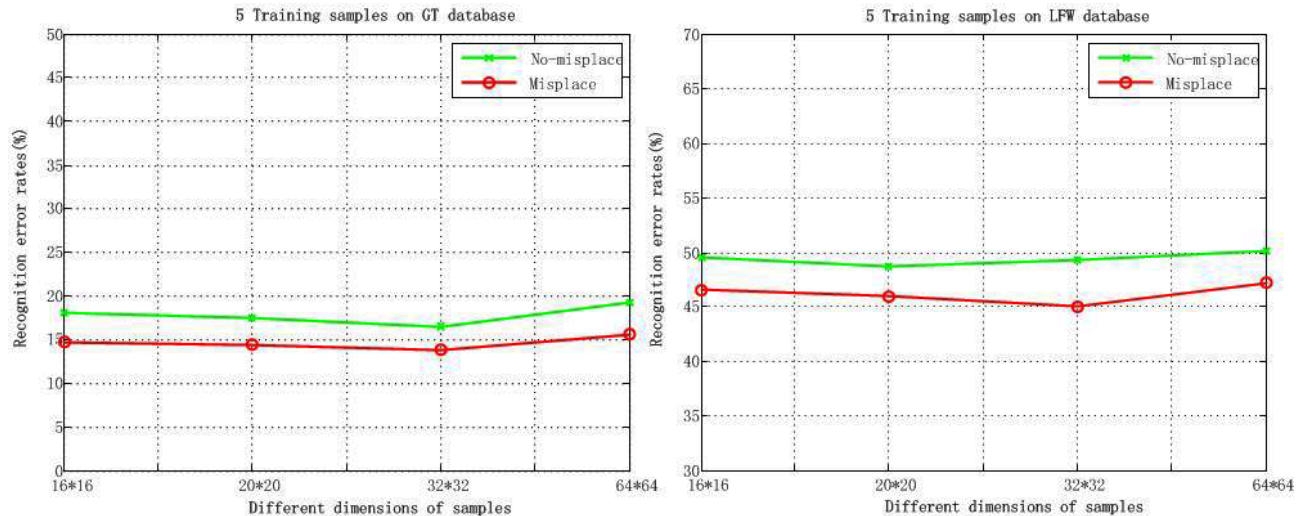


Figure 3. RERs(%) of no-misplaced and misplaced strategies using different dimensions on two databases.

processing through the comparison curves. This validates that our misplaced atom scheme can more effectively enhance the robustness of dictionary and benefit classification.

5. Conclusion

In this paper, we presented a sample group and misplaced atom dictionary for a joint kernel collaborative representation (SGMA_JKCR)-based classification method, which combined a regrouped dictionary and joint kernel function together. This framework could capture more linear and nonlinear sample feature information when the training set was of small scale. The group dictionaries, which consisted of sample groups (one original and three virtual images), made up for some missing information, and the misplaced atom strategy also reflects more variations of samples. In addition, we applied the joint kernel function to optimize the single kernel structure and enhanced the robustness of the KCR model, which can benefit both accuracy and speed for FR. As a whole, our method was a practical model when the training sets were of small scale and had defective features. Experiments on different databases showed that the proposed method was superior to several similar state-of-the-art methods in terms of accuracy and speed, especially on STS cases, and had more robust performance owing to reasonable structure combination.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 61071199, the Natural Science Foundation of Hebei Province of China under Grant No. F2016203422, the Postgraduate Innovation Project of Hebei (China) under Grant No. 00302-6370011, the Foundation of State Key Laboratory of Acoustics (China) under Grant No. SKLA201602, and the Natural Science Foundation of Shandong Province under Grant No. ZR2016AL10.

References

- [1] Zhang H, Li F, Deng H, Li Z, Yan K, Xie C, Wang K. Adjusting samples for obtaining better l2-norm minimization based sparse representation. *J Vis Commun Image R* 2016; 39: 93-99.
- [2] Zhang H, Wang F, Chen Y, Zhang W, Wang K, Liu J. Sample pair based sparse representation classification for face recognition. *Expert Syst Appl* 2016; 45: 352-358.
- [3] Zheng C, Hou Y, Zhang J. Improved sparse representation with low-rank representation for robust face recognition. *Neurocomputing* 2016; 198: 114-124.
- [4] Wu F, Jing XY, You X, Yue D, Hu R, Yang JY. Multi-view low-rank dictionary learning for image classification. *Pattern Recogn* 2016; 50: 143-154.
- [5] Sun Y, Wang X, Tang X. Deep learning face representation by joint identification-verification. *Adv Neur In* 2014; 27: 1988-1996.
- [6] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust. In: *IEEE 2014 Conference on Computer Vision and Pattern Recognition*; 24-27 July 2014; Columbus, OH, USA. New York, NY, USA: IEEE. pp. 2892-2900.
- [7] Zhang Z, Luo P, Chen CL, Tang X. Learning deep representation for face alignment with auxiliary attributes. *IEEE T Pattern Anal* 2016; 38: 918-930.
- [8] Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE T Pattern Anal* 2009; 31: 210-227.
- [9] Yang M, Zhang L, Yang J, Zhang D. Robust sparse coding for face recognition. In: *IEEE 2011 Computer Vision and Pattern Recognition*; 20-25 July 2011; Colorado Springs, CO, USA. New York, NY, USA: IEEE. pp. 625-632.

- [10] Yang M, Zhang L, Yang J, Zhang D. Regularized robust coding for face recognition. *IEEE T Image Process* 2013; 22: 1753-1766.
- [11] Yang M, Song T, Liu F, Shen L. Structured Regularized robust coding for face recognition. In: 2015 Chinese Conference on Computer Vision; 18–20 September 2015; Xi'an, China. pp. 80-89.
- [12] Zhang Q, Li B. Mining discriminative components with low-rank and sparsity constraints for face recognition. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 12–16 August 2012; Beijing, China. New York, NY, USA: ACM. pp. 1469-1477.
- [13] Deng W, Hu J, Guo J. Extended SRC: undersampled face recognition via intraclass variant dictionary. *IEEE T Pattern Anal* 2012; 34: 1864-1870.
- [14] Deng W, Hu J, Guo J. In defense of sparsity based face recognition. In: IEEE 2013 Conference on Computer Vision and Pattern Recognition; 23–28 June 2013; Portland, OR, USA. New York, NY, USA: IEEE. pp. 399-406.
- [15] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: which helps face recognition? In: IEEE 2011 International Conference on Computer Vision; 6–13 November 2011; Barcelona, Spain. New York, NY, USA: IEEE. pp. 471-478.
- [16] Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc* 1998; 2: 121-167.
- [17] Turk M, Pentland A. Eigenfaces for recognition. *J Cognitive Neurosci* 1991; 3: 71-86.
- [18] Etemad K, Chellappa R. Discriminant analysis for recognition of human face images. In: International Conference on Audio-Video-Based Biometric Person Authentication; pp. 1724-1733.
- [19] He X, Yan S, Hu Y, Niyogi P, Zhang HJ. Face recognition using Laplacian faces. *IEEE T Pattern Anal* 2005; 27: 328-340.
- [20] Kang C, Liao S, Xiang S, Pan C. Kernel sparse representation with local patterns for face recognition. In: IEEE 2011 International Conference on image Processing; 11–14 September 2011; Brussels, Belgium. New York, NY, USA: IEEE. pp. 3009-3012.
- [21] Wang Z, Yang W, Yin J, Sun C. Kernel collaborative representation with regularized least square for face recognition. In: 2012 10th International Conference on Service Oriented Computing; 12–15 November 2012; Jinan, China. pp. 130-137.
- [22] Yang W, Wang Z, Yin J, Sun C, Ricanek K. Image classification using kernel collaborative representation with regularized least square. *Appl Math Comput* 2013; 222: 13-28.
- [23] Liu W, Yu Z, Lu L, Wen Y, Li H, Zou Y. KCRC-LCD: Discriminative kernel collaborative representation with locality constrained dictionary for visual categorization. *Pattern Recogn* 2015; 48: 3076-3092.
- [24] Wang D, Lu H, Yang MH. Kernel collaborative face recognition. *Pattern Recogn* 2015; 48: 3025-3037.
- [25] Chen L, Man H, Nefian AV. Face recognition based on multi-class mapping of Fisher scores. *Pattern Recogn* 2005; 38: 799-811.
- [26] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. Available online at <http://vis-www.cs.umass.edu/papers/eccv2008-lfw.pdf>.
- [27] Li FF, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *Comput Vis Image Und* 2004; 106: 59-70.
- [28] Wang D, Lu H. Object tracking via 2DPCA and l_1 -regularization. *IEEE Signal Proc Let* 2012; 19: 711-714.
- [29] Liu W, Yu Z, Wen Y, Yang M. Multi-kernel collaborative representation for image classification. In: IEEE 2015 International Conference on Image Processing; 27–30 September 2015; Quebec City, Canada. New York, NY, USA: IEEE. pp. 21-25.
- [30] Xu Y, Li X, Yang J, Zhang D. Integrate the original face image and its mirror image for face recognition. *Neurocomputing* 2014; 131: 191-199.