Research Article

Adaptive joint block-weighted collaborative representation for facial expression recognition

Zhe SUN[1], Zhengping HU[1,*], Meng WANG[1,2], Shuhuan ZHAO[1]

[1]Department of Information Science & Engineering, Faculty of Electronics & Communication, Yanshan University, Qinhuangdao, China

[2]Department of Physics & Electronics Engineering, Faculty of Electronics & Communication, Taishan University, Tai'an, China

Abstract: Facial expression recognition (FER) plays a significant role in human-computer interactions. Recently, regularized linear representation-based classification has achieved satisfying results in FER. Considering that different blocks in a sample should contribute differently to the representation and classification, we propose an adaptive joint block-weighted collaborative representation-based classification (JBW_CRC) method to effectively exploit the similarity and distinctiveness of different blocks. In JBW_CRC, samples are divided into different blocks and each block of the query sample is represented as a feature vector. Each feature vector is coded on its related block dictionary, which considers the similarity among the feature vectors. Additionally, the distinctiveness of different feature vectors is obtained by weighting its distance to other features, which addresses the distinctiveness in the different feature vectors. The proposed method is verified from the aspect of training samples, time complexity, and Gaussian noise variances on benchmark databases and the extensive experiments show that the proposed method is very competitive with some similar pattern classification methods.

Key words: Facial expression recognition, block-weighted collaborative representation, similarity, distinctiveness

1. Introduction

Facial expression recognition (FER) has attracted much attention and played an important role in human emotions due to its wide applications [1–4]. Moreover, FER relates to pattern recognition, image processing, computer vision, and other aspects [5]. Generally speaking, Paul Ekman proposes that there are seven kinds of expressions, i.e. anger, disgust, fear, happy, sad, surprise, and neutral [6–8], in the emotional state. Figure 1 shows some expression examples in the real world.

Many researchers have applied representation-based classification (RC) to FER and achieved lots of promising results. In the area of classification approaches, sparse representation classification (SRC) has achieved prior performance in FER [9–12]. Ma et al. proposed a block-LGBP sparse representation (SR) [10] that jointly extracted local binary pattern with Gabor (LGBP) features of each subblock and found the minimum residual vectors to achieve the recognition results of different expression classes via SR. Mery et al. [13] devised a new general approach, called adaptive sparse representation of random patches (ASR+), that addressed the problem of automated recognition of facial attributes. Despite the wide application of SRC, it is

---

*Correspondence: hzp_ysu@163.com

Figure 1. Some expression images in the real world.

always implemented with the constraint of the ℓ_1-norm minimization of the linear combination coefficients [14]. Therefore, some researchers referred to other RC methods with the constraint of the ℓ_2-norm minimization. Generally speaking, RC methods with the ℓ_2-norm minimization constraint can lead to similar FER results but speed up the algorithm. Shi et al. [15] demonstrated that an ℓ_2-norm based method can also obtain good results as an ℓ_1-norm constraint algorithm. Moreover, Zhang et al. [16] proposed collaborative representation-based classification with regularized least square (CRC_RLS), and showed that collaboration between classes can benefit classification accuracy and reduce complexity with a squared ℓ_2-regularization. Furthermore, kernel collaborative representation-based classification (KCRC) extended the CRC model to kernel space in a nonlinear way, which performed favorably against other linear or kernel sparse representation techniques with ℓ_1-optimization in terms of accuracy and time cost. For instance, Yang et al. [17] contributed a KCR approach, which extended CRC [18] model to a kernel space and reduced time cost. Wang et al. presented a KCR algorithm with squared ℓ_2-regularization that applies local binary patterns (LBPs) as well as the Hamming kernel (HK) [19] for face recognition and obtained favorable accuracy and speed. Besides CRC, linear regression

classification (LRC) [20] can be regarded as an RC method with the ℓ_2-norm minimization constraint as well. LRC exploited the training samples of each class to represent the query sample and used the representation residuals to perform the classification. In addition, the authors of [21] exploited the residuals of linear regression to define the distance from a test texture to a texture class and achieved satisfying performance. The paper [22] devised a linear regression-based SVM framework for the large-scale classification and obtained better performance among recent algorithm in most cases.

Similar to the results of RC methods, feature extraction methods for FER can also achieve satisfactory recognition results [23,24]. In [24], Nan et al. proposed a weighted multiclassifier optimization and sparse representation-based (WMSRC) method for face recognition and outperforms many existing block-based SRC algorithms. Though the methods above have achieved satisfying performance, they all ignore the fact that the feature vectors in a pattern not only share similarity but also have distinctiveness. Thus, the feature vectors should have similar representation coefficients to share similarity so that they can jointly represent similar features and have some diversity to reflect the distinctiveness at the same time (e.g., pixels in different blocks). Imagine that the pixels with moderate intensities of a sample should contribute more information and have very different coding coefficients compared to those of the remaining parts.

Based on this point, to make the representation flexible, we present an adaptive joint block-weighted collaborative representation-based classification method (JBW_CRC), which considers not only the similarity but also the distinctiveness in different feature vectors. Firstly, each feature vector is coded over its associated block dictionary, which obtains the coding vectors for different feature vectors. Then an adaptive weighted term is exploited to ensure that the coding vectors have a small variance, which considers the similarity in the feature vectors. Meanwhile, the weighting values are optimized in the coding process to address the distinctiveness of different feature vectors. Finally, the query samples are classified to the expression class that has the minimum weighted residual. To evaluate our algorithm, we conduct numerous experiments on several public databases and the experiments show that our method achieves competitive results against some state-of-the-art methods in both classification accuracy and computational complexity.

The paper is organized as follows. Section 2 introduces the proposed method. In Section 3, we evaluate the performance of our method through a number of experiments on several public databases. Section 4 presents analysis of the proposed method. Section 5 concludes the paper.

2. The proposed method

2.1. Adaptive joint block-weighted collaborative representation

Let $\varphi = [\varphi_1, \varphi_2, \cdots, \varphi_C]$ be the training set with C expression classes, where φ_i is the training set of the i-th class. Let y be a query sample to be classified. As shown in Figure 2, the training samples can be divided into K blocks, i.e. $K = 4$. Thus, the training dictionary φ can be divided into K block dictionaries as well. Denote by A_k the block dictionary of the k-th $(k = 1, 2, \ldots, K)$ modality of feature. Similarly, the query image y also can be divided into K blocks and y_k is the k-th modality of the feature vector to be coded and α_k is the coding vector of y_k over A_k. Additionally, Algorithm 1 summarizes the proposed method.

It is reasonable to assume that the different features y_k may share some similarity with their corresponding block dictionary A_k. Thus, the representation coefficients α_k should be similar, which makes the representation stable. Furthermore, the different features y_k can be very distinctive from each other, and so their representation coefficients α_k also have distinctiveness. This makes the representation flexible. Generally speaking, a balance between similarity and difference will result in a more accurate representation for FER.
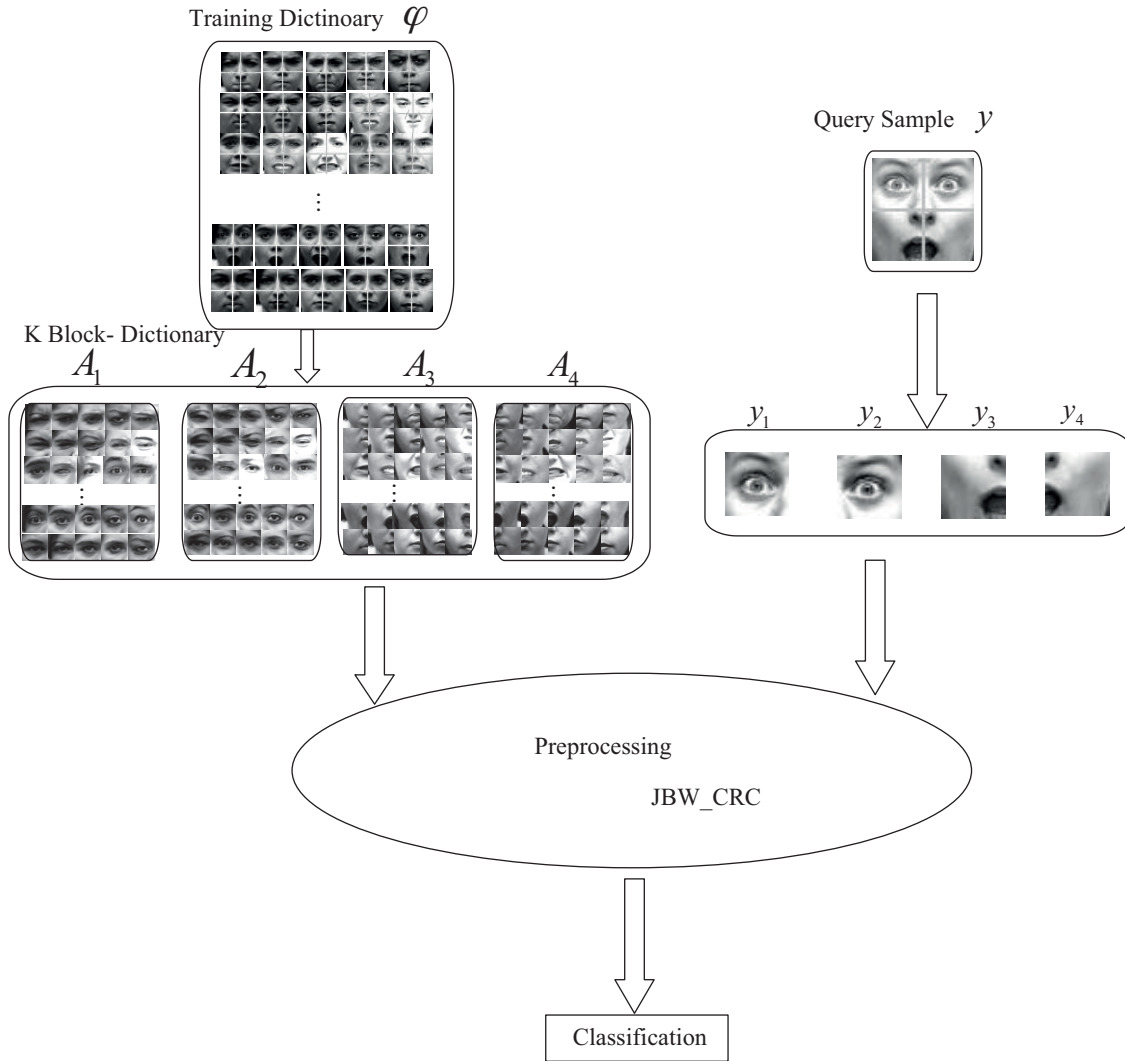
Figure 2. Diagram of the proposed method.

In order to achieve the purpose above, we use the following term to adjust the coding vector α_k of different features over their corresponding block dictionary:

$$\min_{\alpha_k} \sum_{k=1}^{K} w_k \left\| \alpha_k - \bar{\alpha} \right\|_2^2, \tag{1}$$

where $\alpha_k, k = 1, 2, \cdots, K$ means the coding vector of the k-th feature vector y_k over the k-th block dictionary A_k. In addition, $\bar{\alpha}$ is the mean vector of all α_k and w_k is the weighting value to the k-th feature.

With the regularization in Eq. (1), the adaptive joint block-weighted collaborative representation can be converted as

$$\min_{\alpha_k, w_k} \sum_{k=1}^{K} \left(\left\| y_k - A_k \alpha_k \right\|_2^2 + \gamma \left\| \alpha_k \right\|_2^2 + \xi w_k \left\| \alpha_k - \bar{\alpha} \right\|_2^2 \right) s.t. prior \left\{ w_k \right\}, \tag{2}$$

where γ and ξ are the positive constants and $prior\left\{w_k\right\}$ is the optimized weighting value of w_k. In (1), α_k and

w_k are regularized with ℓ_2-norm since it can make the computational time low. Therefore, the optimization of w_k is the key issue of the proposed method and will be discussed in the following part.

2.2. Optimization and classification

The objective function in Eq. (2) can be solved by alternatively optimizing coding vector α_k and weighting value w_k, i.e. updating α_k by fixing w_k and updating w by fixing α_k. The processes are iterated until α_k and w_k converge to some minimum.

Since there is no information about w_k of different feature vectors, we adjust the entropy of w_k using the following formula:

$$-\sum_{k=1}^{K} w_k \ln w_k > \sigma \tag{3}$$

Then the objective function in (1) can be reduced to

$$\min_{w_k} \sum_{k=1}^{K} \xi w_k \|\alpha_k - \bar{\alpha}\|_2^2 + \rho w_k \ln w_k, \tag{4}$$

where $\rho > 0$ is the Lagrange multiplier. According to a close-form solution for $k = 1, 2, \cdots, K$, α_k and $\bar{\alpha}$ in Eq. (2) can be derived by

$$\alpha_k = \alpha_{0,k} + \xi \frac{w_k}{\sum_{\eta=1}^{K} G_k H \sum_{\eta=1}^{K} w_\eta \alpha_{0,\eta}} \tag{5}$$

$$\bar{\alpha} = \sum_{k=1}^{K} w_k \alpha_k \bigg/ \sum_{k=1}^{K} w_k, \tag{6}$$

where $G_k = \left(A_k^T A_k + I\left(\gamma + \xi w_k\right)\right)^{-1}$, $\alpha_{0,k} = G_k A_k^T y_k$, $H = \left(I - \sum_{\eta=1}^{K} \bar{w}_\eta G_\eta\right)^{-1}$, $\bar{w}_\eta = \frac{\xi w_\eta^2}{\sum_{k=1}^{K} w_k}$. Then the weighting value w_k can be updated as

$$w_k = \exp\left\{-1 - \xi \|\alpha_k - \bar{\alpha}\|_2^2 \bigg/ \rho\right\} \tag{7}$$

Finally, the residual for each class can be computed by

$$r_i(y) = \sum_{k=1}^{K} w_k \left\|y_k - A_k^i \alpha_k^i\right\|_2^2 \quad , \tag{8}$$

where α_k^i represents the coefficients vector α_k corresponding to the i-th class and output the identity of y as

$$identify(y) = \arg\min_i r_i(y) \tag{9}$$

2.3. Experiments

In this part, the proposed method is evaluated on several public databases; some databases meet the kind for the seven emotions but some do not. To assess the performance, we use leave one-subject-out (LOSO) cross-validation in all experiments. This means we choose each subject at a time for testing and train the other subjects. In other words, we perform a separate experiment for each of the subjects: we leave out all images of the left subject for testing and train all images of other subjects. We repeat this scenario for all subjects.

Algorithm 1 JBW_CRC algorithm for facial expression recognition

---

Input: Normalize the training dictionary φ and y to have unit ℓ_2-norm.

1: Divide φ and y into block dictionary A_k and feature vector y_k, respectively.
2: Solve the Eq. (2) by optimizing α_k and w_k.
3: Update the coefficients of i-th class α_k and the mean coefficient vector $\bar{\alpha}$ via Eq. (5) and Eq. (6).
4: Obtain the optimal adaptive weighted value w_k via (7).
5: Compute the residuals via (8).
6: Identify y via (9)

Output: The label of y.

(e.g., JAFFE database has 10 subjects in total. LOSO selects 9 out of the 10 subjects for training and uses the remaining subject for testing. This procedure is repeated for all the 10 subjects.) Afterwards, the input facial images from all the databases are all cropped to the size of 64 × 64 based on two eye locations [25] and downsampled to 20 × 20 pixels and normalized to consider as the feature vector with 400 elements.

2.4. Parameter setting

In this paper, SVM (linear kernel), SRC-ℓ_1, KCRC-ℓ_2 (LBP+HK), and LRC are used to compare with the proposed JBW_CRC. Furthermore, parameter λ_1 in SRC is set to 0.1 and the parameters λ_2, λ_3 used in the KCRC, LRC are set to 0.005 and 0.1, respectively. Additionally, parameters γ, ξ, and ρ used in JBW_CRC are set to 0.0005, 0.005, and 0.1, respectively. In addition, all the experiments are conducted on a PC (Intel Core i5-4460 CPU, 3.20 GHz) with MATLAB R2012a software.

2.5. Facial expression databases

2.5.1. JAFFE database

The JAFFE database includes 10 females of 213 facial expression images, including anger, disgust, fear, happy, sadness, surprise, and neutral, the numbers of which are 30, 29, 32, 31, 31, 30, and 30, respectively.

2.5.2. Extended Cohn–Kanade database

The extended Cohn–Kanade (CK+) database contains 593 image sequences from 123 subjects. We select 118 subjects since they meet the kinds of seven emotions, i.e. anger, disgust, fear, happy, sadness, surprise, and neutral, whose numbers are 135, 177, 75, 207, 84, 246, and 314, respectively. We select the first frame from each sequence as neutral images and use the last three frames from each sequence as the facial expression images in our experiments. Some expression images are shown in Figure 3.

2.5.3. KDEF database

The KDEF database records facial expression images of 70 subjects at 5 different viewing angles. We only select the 980 frontal facial expression images, including anger, disgust, fear, happy, sadness, surprise, and neutral.

2.5.4. CAS-PEAL database

The CAS-PEAL database records 99,594 images of 1040 individuals (595 males and 445 females) with varying pose, expression, accessory, and lighting (PEAL). We only select the 2256 frontal facial expression images of

Anger   Disgust   Fear   Happy   Sadness   Surprise   Neutral
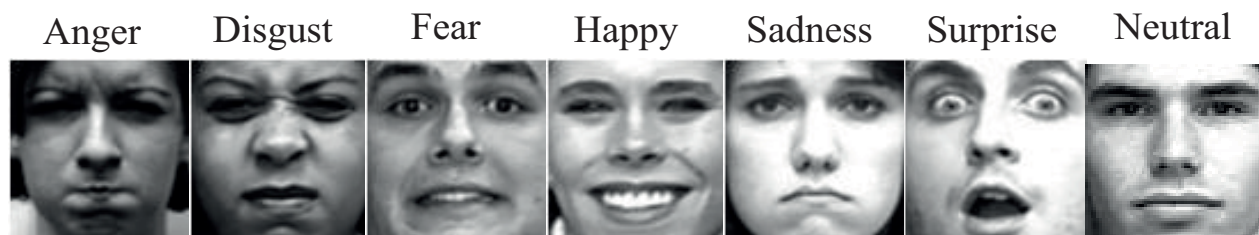
Figure 3. Some facial images in the CK+ database.

376 subjects, including closed eyes, frown, smile, open mouth, surprise, and neutral.

2.5.5. AR database

The AR database contains over 4000 images for 100 subjects. We only select the 800 frontal facial expression images, which include neutral, happy, anger, and surprise.

2.6. Comparison of different blocks

In this subsection, we perform experiments on several possibilities about the selection of K. As observed from Table 1, when the samples are divided into 1×4, the average accuracies on all databases are higher than those of other cases. Thus, K is set to 4 in this paper.

**Table 1.** Average accuracy under different selection of parameter K (%).

| Databases | Selection of parameter K | | | | |
| | 1×4 | 2×4 | 3×3 | 3×4 | 4×4 |
	($K=4$)	($K=8$)	($K=9$)	($K=12$)	($K=16$)
JAFFE	64.97	63.65	61.57	62.13	59.49
CK+	82.51	76.90	75.65	74.11	73.05
KDEF	75.92	72.65	70.41	70.10	70.00
CAS-PEAL	75.35	72.85	71.14	68.66	69.68
AR	83.75	83.37	83.15	82.63	81.97

2.7. Comparison of different classification algorithms

In this part, we respectively exploit SVM, SRC, KCRC, and LRC algorithm-based classification for comparison. Figure 4 shows the results of each expression class under different algorithms. From these figures, we can observe that JBW_CRC performs better than other classification algorithms. Additionally, Table 2 (corresponding to Figure 4) summarizes the average accuracy and running time (per sample) of different algorithms. It can be seen that JBW_CRC achieves the best performance in both each expression class and average accuracy. Although the time cost of JBW_CRC is not the lowest, it also has superiority over other algorithms such as SRC and SVM.

2.8. Comparison of different training samples per subject

Figure 5 shows the average accuracies of different training samples per subject on JAFFE and AR databases, respectively. For instance, for the experimental results on JAFFE database shown in Figure 5a, when the first 2 images of each subject are used as training samples, JBW_CRC can achieve the best recognition rate of 45.81%
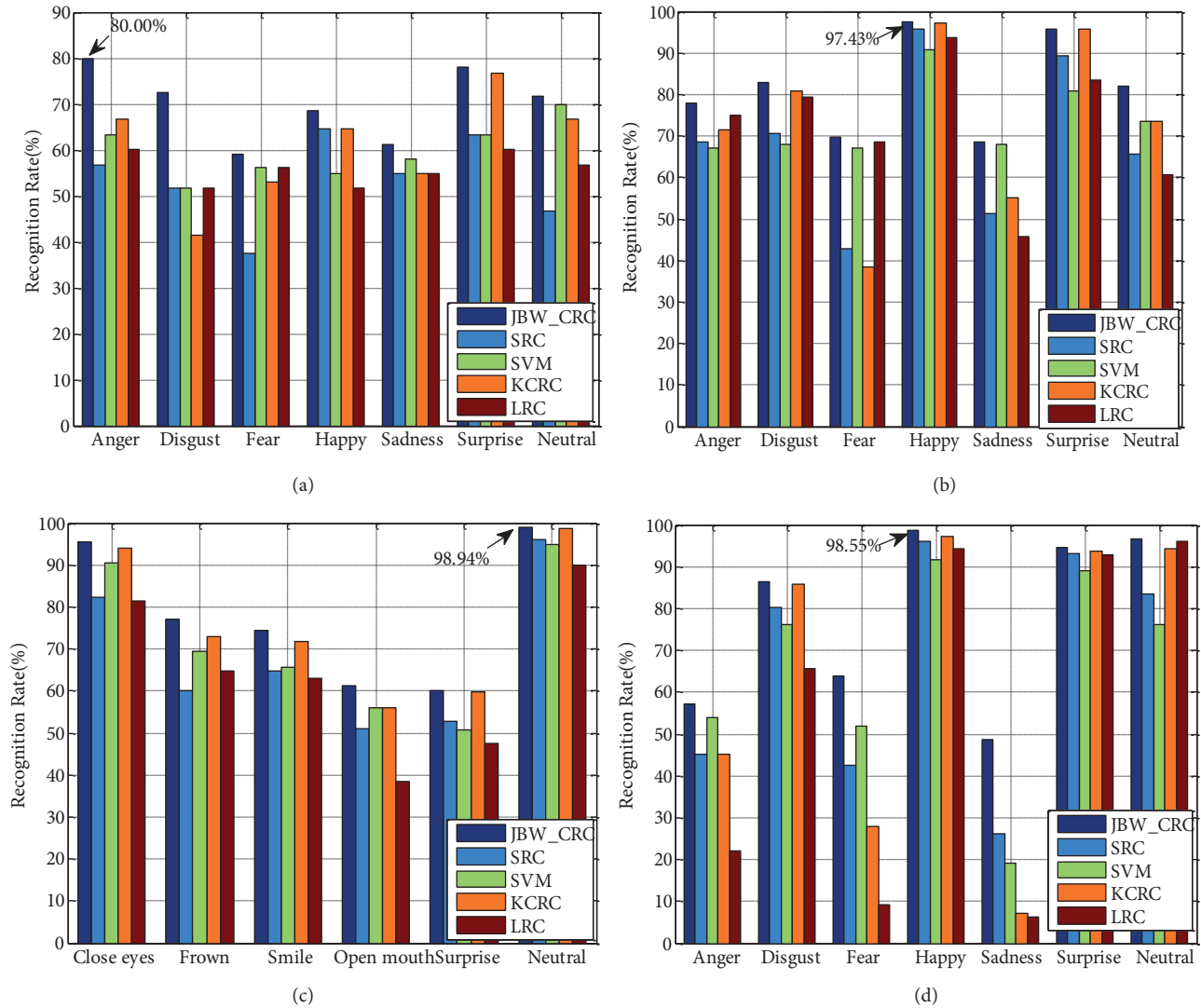
Figure 4. Comparison of test accuracies under five types of classifiers with different databases. (a) JAFFE database, (b) KDEF database, (c) CAS-PEAL database, (d) CK+ database.

Table 2. Average rate (R) and running time (T) of different methods on different databases.

RC methods	JAFFE		KDEF		CAS-PEAL		CK+	
	R(%)	T(s)	R(%)	T(s)	R(%)	T(s)	R(%)	T(s)
JBW_CRC	64.97	0.0075	75.92	0.0947	75.35	0.9879	82.51	0.1608
SRC	53.26	0.2065	69.18	1.2785	67.77	5.3095	76.18	2.2481
SVM	59.24	0.0097	73.57	0.3344	71.37	13.978	75.90	0.5689
CRC_RLS	60.37	0.0023	73.16	0.0240	76.20	1.0561	79.90	0.0556
LRC	55.67	0.0020	72.35	0.0205	69.19	0.2204	82.51	0.1608

while the second best method, i.e. SVM, only obtains a rate of 40.61%. Therefore, JBW_CRC is superior in both recognition rates and computational time. Furthermore, Tables 3 and 4 show the accuracy and time cost on the KDEF and CK+ databases. We use the abbreviations 'An', 'Di', 'Fe', 'Ha', 'Sa', and 'Su' to stand for the expressions anger, disgust, fear, happiness, sadness, and surprise, respectively. As can be seen, in these

figures and tables, our method is superior to the other comparison algorithms from the aspect of time cost and test accuracy.
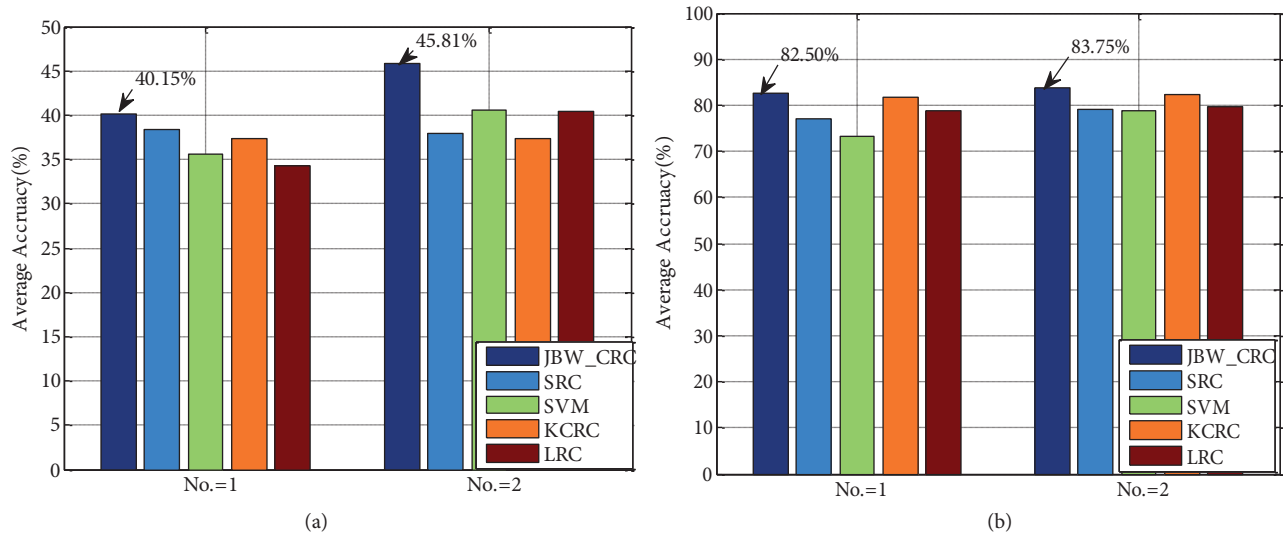
Figure 5. Comparison of test accuracies under different training samples per subject with different databases. (a) JAFFE database, (b) AR database.

Table 3. Recognition rate and running time of different training samples on KDEF database. (No.=1)

RC methods	Recognition rate (%)								Running
	An	Di	Fe	Ha	Sa	Su	Ne	Overall	time (s)
JBW_CRC	74.29	80.71	58.57	94.86	52.86	93.57	72.86	70.31	0.0237
SRC	63.57	66.43	35.71	92.14	42.86	82.86	62.75	62.96	0.3592
SVM	57.14	65.71	51.43	82.14	47.86	72.86	57.14	62.76	0.0701
KCRC	70.14	80.00	31.43	94.14	47.14	92.14	59.29	68.67	0.0068
LRC	60.71	76.43	40.00	92.86	40.00	81.29	45.71	67.45	0.0079

Table 4. Recognition rate and running time of different training samples on CK+ database. (No.=1)

RC methods	Recognition rate (%)								Running
	An	Di	Fe	Ha	Sa	Su	Ne	overall	time (s)
JBW_CRC	56.30	85.88	56.00	97.58	36.90	94.37	96.17	80.14	0.0191
SRC	44.67	76.27	38.67	93.72	23.81	92.12	76.36	74.09	0.3287
SVM	53.33	77.40	44.00	87.92	15.48	86.59	68.69	71.92	0.0448
KCRC	43.70	85.31	20.33	95.87	8.79	92.75	90.73	77.28	0.0059
LRC	21.19	63.76	11.06	92.68	6.19	90.34	88.22	67.97	0.0067

2.9. Comparison of different gaussian noise variances

To verify the proposed algorithm's robustness to noise, all the query samples are added with gaussian random noise, with different variances. Figure 6 shows some examples of the images from the CK+ database with different noise variances. When the variance becomes 0.4, we can hardly distinguish the expression information.

Figure 6. Some images on CK+ database with different noise variance from left to right of 0, 0.01, 0.1, 0.2, 0.3, 0.4, and 0.5.

The first column of Figure 7 shows the average accuracies on different databases. Moreover, we compare the time cost of different methods with the proposed JBW_CRC as shown in the second column. Intuitively, we can see that the proposed approach achieves better performance than the other comparison algorithms and has a relative advantage in the aspect of the time cost as well.

3. Analysis of the proposed method

It should be pointed out that the proposed algorithm considers both the similarity and distinctiveness among the feature vectors that existed in different blocks. Note that the term $\sum_{k=1}^{K} \left(\|y_k - A_k \alpha_k\|_2^2 + \gamma \|\alpha_k\|_2^2 \right)$ in Eq. (3) ensures all features y_k have the similar coding vector α_k over their associated block dictionary A_k. However, it ignores the fact that different features in a module also have distinctiveness. Thus, the third term $\xi w_k \|\alpha_k - \bar{\alpha}\|_2^2$ is added to Eq. (3) in our method, which considers the fact that different features y_k should have similar coding vector α_k so that they can jointly represent the query sample, while α_k should also have distinctiveness to reflect the distinctive property of different features. Here we take a query sample from the CK+ database for example and show the adaptive weights obtained by our method, which can be seen in Figure 8. We can see that w_1 is the smaller and it means the feature vector y_1 is less similar to other features. Thus, we exploit the adaptive weighted value w_1 to adjust the coding vector α_1. Overall, if the feature vector y_k is similar to other features, w_k should be larger to enforce α_k approaching $\bar{\alpha}$. Otherwise, w_k should be smaller so that α_k can vary more from others.

Additionally, the proposed method has a lower time complexity. For instance, the time complexity of all the coding procedures is $O\left(\sum_{k=1}^{K} 3m_k^2 + m_k N\right)$ if A_k has a size of $m_k \times N$. Meanwhile, the time complexity of computing $G_k A_k^T y_k$ is $O\left(\sum_{k=1}^{K} m_k^2 + m_k N\right)$ and obtaining $G_k H \sum_{\eta=1}^{K} w_\eta \alpha_o$ is $O\left(\sum_{k=1}^{K} 2m_k^2\right)$, respectively. Moreover, the weighting value w_k is optimized with the iteration of q times. Thus, the whole time complexity of JBW_CRC is $O\left(q \sum_{k=1}^{K} 3m_k^2 + m_k N\right)$. A large number of experimental results shown in Section 3 also verify the effectiveness of our method from the aspect of running time per image.

4. Conclusion

In this paper, we present a novel adaptive joint block-weighted collaborative representation-based classification (JBW_CRC) to effectively join the similarity and distinctiveness of different feature vectors. First, each feature vector can be flexibly represented on its associated block dictionary. Then an adaptive weighted value is exploited to ensure that the coding vectors have a small variance, which considers the similarity between feature vectors and simultaneously addresses the distinctiveness of different feature vectors in the coding stage. Finally, the
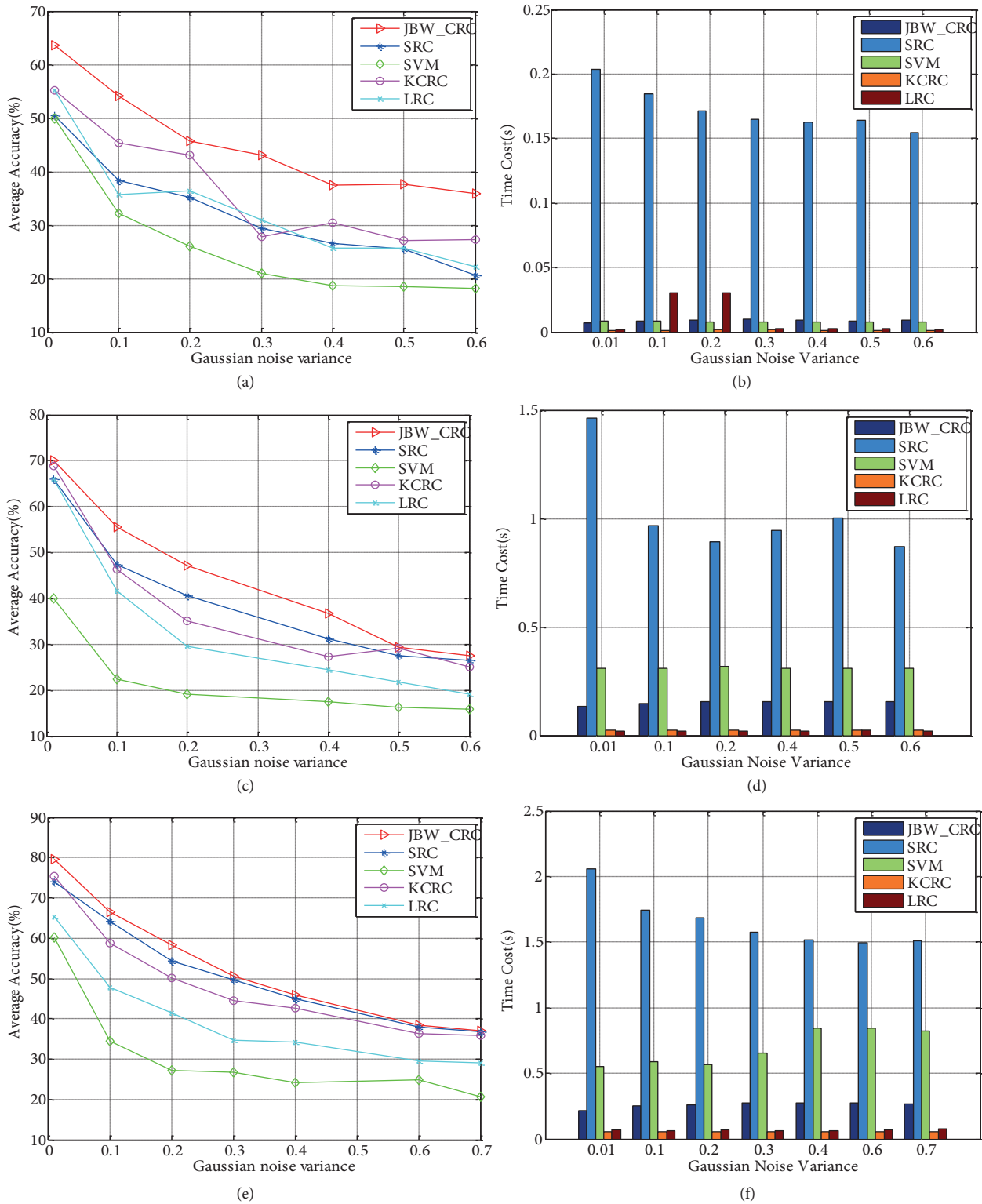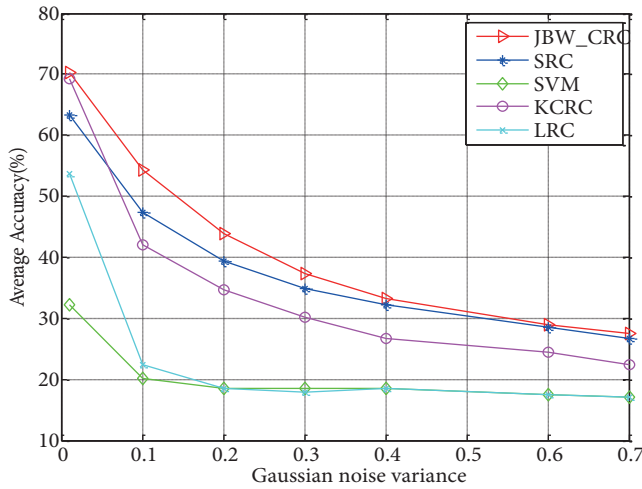
Figure 7. Comparison of average accuracies and time cost different methods under different Gaussian noise variances.
(a, b) JAFFE database, (c, d) KDEF database, (e, f) CK+ database, (g, h) AR database, (i, j) CAL-PEAL database.
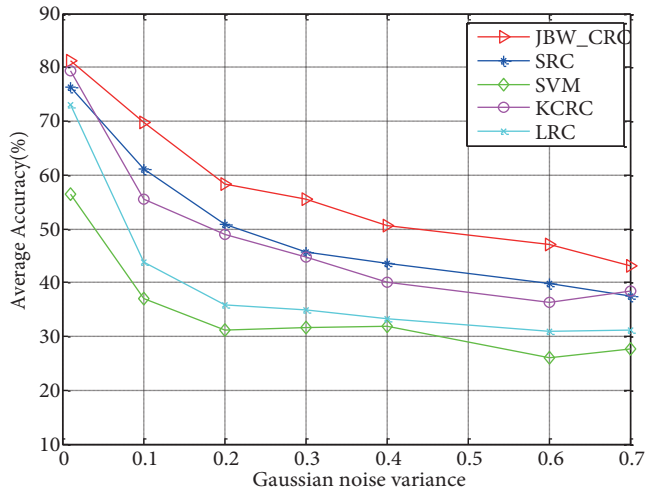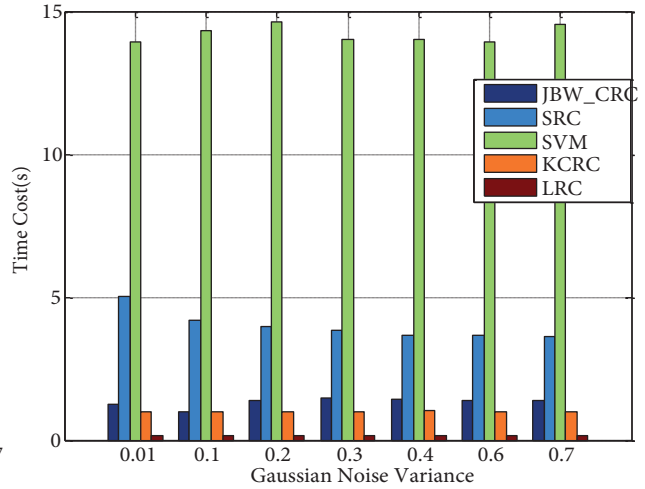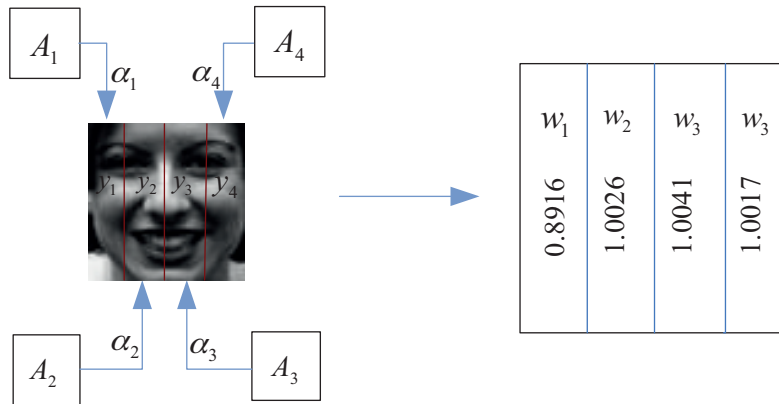
(g)
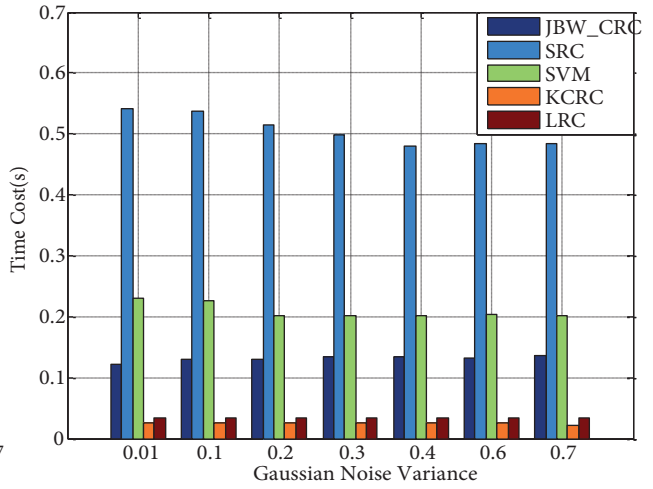
(h)

(i)

(j)

Figure 7. Continued.

Figure 8. Analysis of the proposed method.

query samples are classified to the expression class that has the minimum weighted residual. The experiments on benchmark databases show that the proposed method is superior to several similar state-of-the-art methods from the aspect of recognition accuracy and speed.

Acknowledgments

References

[1] Lee SH, Baddar WJ, Yong MR. Collaborative expression representation using peak expression and intra class variation face images for practical subject-independent emotion recognition in videos. Pattern Recogn 2016; 54: 52-67.

[2] Happy SL, Routray A. Automatic facial expression recognition using features of salient facial patches. IEEE Transactions on Affective Computing 2015; 6: 1-12.

[3] Shao J, Gori I, Wan S, Aggarwal JK. 3D Dynamic facial expression recognition using low-resolution videos. Pattern Recogn Lett 2015; 65: 157-162.

[4] Yusuf R, Sharma DG, Tanev I, Shimohara K. Evolving an emotion recognition module for an intelligent agent using genetic programming and a genetic algorithm. Artificial Life and Robotics 2016; 21: 85-90.

[5] Pablos SM, García-Bermejo JG, Casanova EZ, Lopez J. Dynamic facial emotion recognition oriented to HCI applications. Interact Comput 2015; 27: 99-199.

[6] Ekman P, Friesen WV. A new pan-cultural facial expression of emotion. Motivation & Emotion 1986; 10: 159-168.

[7] Ekman P, Rolls ET, Perrett DI, Ellis HD. Facial expressions of emotion: an old controversy and new findings. Philosophical Transactions of the Royal Society of London 1992; 335: 63-69.

[8] Ekman P, Friesen WV, Ellsworth P. Sampling Emotion Words, Categories, or Dimensions in Judgment Studies. in: Ekman P, editor. Emotion in the Human Face. Emotion in the Human Face. Cambridge, UK: Cambridge University Press, 1998. pp. 45-47.

[9] Fang Y, Chang L. Multi-instance feature learning based on sparse representation for facial expression recognition. Lect Notes Comput Sc 2015; 8935: 224-233.

[10] Ma D, Li M, Nian F, Kong C. Facial expression recognition based on characteristics of block LGBP and sparse representation. Journal of Computational Methods in Sciences and Engineering 2015; 15: 537-547.

[11] Zhang S, Zhao X, Lei B. Facial expression recognition using sparse representation. WSEAS Transactions on Systems 2015; 11: 440-452.

[12] Guo Y, Zhao G, Pietikainen M. Dynamic facial expression with atlas construction and sparse representation. IEEE T Image Process 2016; 25: 1977-1992.

[13] Mery D, Bowyer K. Automatic facial attribute analysis via adaptive sparse representation of random patches. Pattern Recogn Lett 2015; 68: 260-269.

[14] Yan Q, Song N, Huang R. Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers. Neurocomputing 2015; 149: 71-78.

[15] Shi Q, Eriksson A, Hengel AVD, Shen C. Is face recognition really a Compressive Sensing problem? Computer Vision and Pattern Recognition 2011; 42: 553-560.

[16] Waqas J, Yi Z, Lei Z. Collaborative neighbor representation based classification using l2-minimization approach. Pattern Recogn Lett 2013; 34: 201-208.

[17] 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.

[18] Zhang L, Yang M, Feng X. Spares representation or collaborative representation: which helps face recognition? International Conference on Computer Vision 2011; 6669: 471-478.

[19] Wang D, Lu H, Yang MH. Kernel collaborative face recognition. Pattern Recogn 2015; 48: 3025-3037.

[20] Yang W, Ricanek K, Shen F. Image classification using local linear regression. Neural Comput Appl 2014; 25: 1913-1920.

[21] Dong Y, Tao D, Li X, Ma J, Pu J. Texture classification and retrieval using shearlets and linear regression. IEEE T Cybern 2015; 45: 358-369.

[22] Wu J, Yang H. Linear Regression-Based Efficient SVM Learning for Large-Scale classification. IEEE T Neural Networ 2015; 26: 2357-2369.

[23] Huang HF, Tai SC. Facial expression recognition using new feature extraction algorithm. Electronic Letters on Computer Vision and Image Analysis 2012; 11: 41-54.

[24] Nan D, Xu Z, Bian SQ. Face recognition based on multi-classifier weighted optimization and sparse representation. International Journal of Signal Processing Image Processing and Pattern Recognition 2013; 6: 423-436.

[25] Tian Y. Evaluation of face resolution for expression analysis. Conference on Computer Vision and Pattern Recognition Workshops 2004; 1: 82-82.