Multi-class differentiation feature representation guided joint dictionary learning for facial expression recognition

Zhe Sun (sunzhe_ysu@163.com)
Yanshan University
Jiatong Bai
Yanshan University
Hehao Zhang
Yanshan University

Research Article

Keywords: facial expression recognition, multi-class differentiation feature representation, joint dictionary learning

Posted Date: April 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1591388/v1

Multi-class differentiation feature representation guided joint dictionary learning for facial expression recognition

Zhe Sun[1*], Jiatong Bai[1] and Hehao Zhang[1]

[1*]Information Science and Engineering, Yanshan University, Qinhuangdao, 066000, China.

*Corresponding author(s). E-mail(s): sunzhe_ysu@163.com;

Abstract

The limitation of the small-scale expression samples generally causes the performance degradation for facial expression recognition-based methods. Also, the correlation between different expression is always ignored when performing feature extraction process. Given above, we propose a novel approach that develops multi-class differentiation feature representation guided joint dictionary learning for FER. The proposed approach mainly includes two steps: firstly, we construct multi-class differentiation feature dictionaries corresponding to different expressions of training samples, aiming to enlarge inter-expression distance to mitigate the problem of nonlinear distribution in training samples. Secondly, we joint learn the multiple feature dictionaries by optimizing the resolutions of each feature dictionary, aiming to establish the strong relationship and enhance the representation ability among multiple feature dictionaries. To sum up, the proposed approach has more discriminative ability from the representation perspective. Comprehensive experiments carried out using three public datasets, including JAFFE, CK+, and KDEF datasets, demonstrate that the proposed approach has strong performance for small-scale samples compared to several state-of-the-art methods.

Keywords: facial expression recognition, multi-class differentiation feature representation, joint dictionary learning

1 Introduction

Facial expression recognition (FER) has been developed as an important research topic owing to its wide application [1-4]. Despite deep learning models achieves comparative performance in FER, they mostly leverage vast annotated samples during the training stage. Thus, it is a challenging and meaningful task to learn discriminative expression features from limited datasets.

To resolve the problem of expression recognition with scanty samples, most existing FER methods can be roughly partitioned into two major aspects: feature descriptor-based methods [5-16] and dictionary learning-based methods [22-33]. The former mainly focuses on extracting discriminative features to perform FER, including handcrafted descriptors and deep descriptors. The latter aims to learn the compacted atom representations for FER, including single dictionary learning and multi-dictionary learning.

Many handcrafted descriptors for FER have been successfully proposed to extract low-level expression features [5-11]. For example, Chen et al. [5] developed a novel method which utilized the expression samples based on l1-norm sparse representation to reduce the problem of cross-domain mismatch. Liu et al. [6] proposed a main directional mean optical-flow (MDMO) feature that can learn the discriminative dictionary from micro-expression samples. Furthermore, Yan et al. [7] presented an effective image filter learning method to acquire expressions representation. Besides the handcrafted features-based methods above, deep descriptors-based methods have made a breakthrough in the field of FER [12-18]. For instance, Yu et al. [12] proposed a multi-task method that used the channel module and spatial module to obtain co-attention scores on FER tasks. Zhu et al. [13] constructed convolutional relation network for FER, which adopted a feature similarity comparison among the enough expression images to identify new categories with fewer images. Also, Xue et al. [15] employed the vision transformers model to obtain the feature representation between different facial regions adaptively and achieved satisfying performance. However, considering the deep learning models have a complex structure that heavily reply on large number of training samples and consequently consumes a large amount of computational cost, some researchers [19-21] turned to deep subspace learning based methods that only used two filter layers to perform the feature representation procedure. For example, Sun et al. [20] proposed an extended dictionary consisting of feature and variation dictionaries to resolve the problem caused by limited training samples. Also, they [21] proposed an effective classification method based on PCANet and LDANet to learn abstract expression features. Although deep descriptors-based methods have shown many advantages in extracting abstract features, the performance on the small-scale datasets will degrade because of the deep model overfitting.

Dictionary learning (DL) methods [22-28] also have shown the good performance for FER tasks as well. From the perspective of single dictionary learning, Tanfous et al. [23] employed sparse coding and dictionary learning to obtain the 2D facial landmark sequences for expression recognition. Yan et al. [25] proposed an unsupervised domain adaptive dictionary method to connect source domain and target domain, which the coding of the two domains were embedded on each other to achieve comparative performance. Zhang et al. [26] presented the Fisher discrimination dictionary learning (FDDL) that added the Fisher discrimination criteria to make the dictionary atoms correspond to class labels. Given the fact that face similarities may confuse the expression recognition process, some multi-dictionary learning methods have been presented to address this problem [29-33]. For example, Moeini et al. [28] extracted comprehensive features by respectively learning identity and expression dictionaries to gain optimal expression classification for FER. Besides, Luo et al. [29] presented multi-resolution dictionary learning method that supplied the dictionary for each resolution to alleviate the problem of facial images resolution diversity. Zhang et al. [33] presented a cost-sensitively joint feature and dictionary learning approach, which considered the separate misclassification cost objectives during the feature and dictionary learning stages to achieve a minimum overall recognition loss.

The literatures above have shown the promising performance, while the nonlinear distribution in expression samples and the correlation between different expressions are still crucial for FER. Hence, in this paper, we proposed a multi-class differentiation feature representation guided joint dictionary learning approach which focused on extracting the discriminative features from the limited expression images. More specifically, we first constructed the multi-class differentiation feature dictionaries to enlarge inter-expression distance, aiming to increase the linear separability among expression samples. Furthermore, our approach jointly learned the multiple feature dictionaries by adding a relatively constraint to establish the strong relationship and enhance the representation ability among multiple feature dictionaries. To sum up, we presented a novel feature learning approach with high discriminability based on the multi-class differentiation representation and joint dictionaries learning. The major contributions include:

- We provided a universal multi-class feature representation approach, which is benefit for increasing inter-expression distance to make the class-wise information discriminative.
- We further jointly learned the multiple feature dictionaries to form a more robust and comprehensive dictionary, which can be extended to application of other multi-class recognition tasks.
- We conducted extensive experiments in different scenarios, such as different datasets, to demonstrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 describes the proposed approach in detail. We compare our approach with several state-of-the-art methods on three public expression datasets in Section 3. Section 4 shows the visualization study and analyzes the proposed approach. Finally, we draw the conclusion in Section 5.

2 The proposed approach

In this section, the proposed approach is illustrated in **Fig. 1**. As is shown in the figure, the proposed approach includes two main steps. Firstly, we construct multi-class differential feature dictionary. Secondly, we jointly learn the multiple feature dictionaries by optimizing the resolutions of each feature dictionary. The detail of multi-class differentiation feature representation model is introduced in Section 2.1, and the process of joint feature dictionary learning as well as classification criterion are given in Section 2.2.

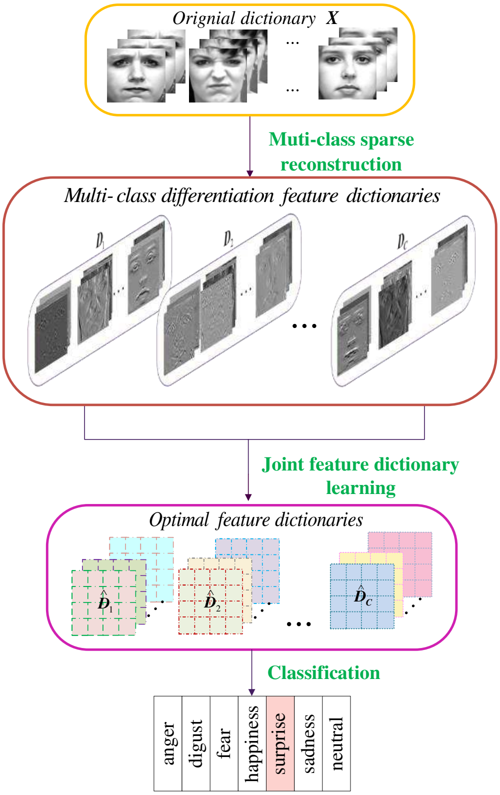


Fig. 1 An illustration of the proposed approach

2.1 Multi-class differentiation feature representation model

In this subsection, we present the proposed multi-class differentiation feature representation model, aiming to highlight class-related features corresponding to each expression while suppressing class-unrelated expression features. Now we will depict the process of this model in detail. Let

$X = [X_1, \ldots, X_i, \ldots, X_C] \in R^{m*N}$ $(i = 1, 2, \ldots, C)$ with C expression classes denote the original dictionary, where $X_i = [x_{i,1}, \ldots, x_{i,j}, \ldots, x_{i,n_i}] \in R^{m*n_i}$ is the training subset of the i^{th} class and n_i is the number of samples from class i. Also, m represents the dimensionality of training sample and N is the total number of training samples. For a random training sample $x_{i,j}$, we first generate the intra-class sample $f_{i,j}$ obtained by the sparsely linear reconstruction of X_i:

$$f_{i,j} = X_i \cdot \delta \tag{1}$$

Where $\alpha = [\delta_1, \ldots\ldots, \delta_{n_i}]^T \in R^{n_i*1}$ is the representation coefficient that corresponds to the i^{th} class and can be obtained by employing the sparse representation based l_1-norm minimization given that it shows promising sparsity for reconstruction [34]:

$$(\hat{\alpha}) = \arg\min_{\alpha} \|\alpha\|_1, s.t. \|x_{i,j} - f_{i,j}\|_2 < \varepsilon \tag{2}$$

Then we obtain the proposed differentiation feature $d_{i,j}$ by:

$$d_{i,j} = x_{i,j} - f_{i,j} \tag{3}$$

For a random query sample y, we can obtain the corresponding differentiation feature vectors $[y_1, \ldots, y_i, \ldots, y_C]$ in the similar way with $d_{i,j}$. Then, all differentiation features can form multi-class differentiation feature dictionaries $[D_1, \ldots, D_i, \ldots, D_C]$, which can be written as:

$$\begin{bmatrix} D_1 \\ \vdots \\ D_{\mathrm{i}} \\ \vdots \\ D_{\mathrm{C}} \end{bmatrix} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,\mathrm{j}} & \cdots & d_{1,\mathrm{N}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{\mathrm{i},1} & \cdots & d_{\mathrm{i},\mathrm{j}} & \cdots & d_{\mathrm{i},\mathrm{N}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{\mathrm{C},1} & \cdots & d_{\mathrm{C},\mathrm{j}} & \cdots & d_{\mathrm{C},\mathrm{N}} \end{bmatrix} \tag{4}$$

2.2 Joint feature dictionary learning

To establish the strong relationship and enhance the representation ability among multi-class differentiation feature dictionaries $[D_1, \ldots, D_i, \ldots, D_C]$, we propose to jointly learn the multiple feature dictionaries by optimizing the resolutions of each feature dictionary. We assume that the multi-class difference feature dictionaries above can be incorporated into a framework, namely the joint feature dictionary learning model that can be defined as:

$$\left\langle \hat{D}_1, \ldots, \hat{D}_i, \ldots, \hat{D}_C, A \right\rangle = \arg\min_{D_1, \ldots\ldots, D_C, A} \sum_{i=1}^{C} \|X_i - D_i A\|_2^2 + \gamma \|A\|_2^2 \tag{5}$$

Where $A = [\mu_1, \ldots, \mu_i, \ldots, \mu_N]^T$ is the coding coefficient matrix and γ is the regularization parameter. It's worth mentioning that $\gamma\|A\|_2^2$ stabilizes the least square solution and prevents the model from over-fitting of the training data. Also, $\gamma\|A\|_2^2$ generates a certain degree of sparsity to the solution $\left\langle \hat{D}_1, \ldots, \hat{D}_i, \ldots, \hat{D_C}, A \right\rangle$, aiming to reduce the number of feature vectors and the complexity of the model.

Obviously, Eq. (5) is a typical norm minimization problem and the initialization of the dictionary is vital for learning the ideal dictionary. In this paper, we initialize $[D_1, \ldots, D_i, \ldots, D_C]$ and obtain the coding coefficient matrix A for the first time:

$$A = (\sum_{i=1}^{C} {D_i}^T D_i + \gamma I)^{-1} \cdot \sum_{i=1}^{C} {D_i}^T X_i \tag{6}$$

Where I is an identity matrix. Assuming the coefficient matrix A and sub-dictionaries $D_1, \ldots, D_{i-1}, D_{i+1}, \ldots, D_C$ were fixed, we subsequently update D_i to obtain the optimized feature dictionary $\hat{D}_i$ corresponding to D_i by:

$$\hat{D}_i = \underset{D_j,(i \neq j)}{\arg\min} \|X_i - D_j A\|_F^2 = X_i A^+ \tag{7}$$

Where $A^+ = A^T(AA^T)^{-1}$. Then the optimal feature dictionaries $\left[\hat{D}_1, \ldots, \hat{D}_i, \ldots, \hat{D_C}\right]$ can be obtained. The summation of $[y_1, \ldots, y_i, \ldots, y_C]$ is defining by:

$$s = \sum_{i=1}^{C} y_i \tag{8}$$

In the classification step, we code s on the dictionaries $\hat{D}_1, \ldots, \hat{D}_i, \ldots, \hat{D_C}$ respectively. Assuming θ_m is the coefficient vector corresponding to $\hat{D_m}$, and the query sample can be assigned to the label of the minimized reconstruction error of m^{th} class:

$$K(m) = \underset{m}{\arg\min} \left\{ \|s - \hat{D_m}\theta_m\|_2^2 \right\}, (m = 1, 2, \ldots, C) \tag{9}$$

3 Experiments and Results

Three public datasets containing basic seven facial expressions were used in experiments, including the Japanese Female Facial Expression (JAFFE) [35], the Extended Cohn-Kanade (CK+) [36], and the Karolinska Directed Emotional Faces (KDEF) [37] datasets. To be convenient, the expressions anger, disgust, fear, happiness, sadness, surprise, and neutral were abbreviated to "An", "Di", "Fe", "Ha", "Sa", "Su", and "Ne", respectively. We adopted Leave-One-Subject-Out (LOSO) cross-validation in the experiments. The facial images were first cropped to a size of 64*64 and subsequently down-sampled to 48*48 pixels. We empirically set the parameter γ to 0.0001 in the proposed joint feature dictionary learning. In this paper, we compared the proposed approach with some handcrafted-based and the deep-based methods. Also, we conducted the experiment with different block occlusion and random corruption to verify the robustness of our approach.

3.1 Comparison with state-of-the-art methods

We compared the performance of our approach with some state-of-the-art handcrafted methods (e.g. LTeP+SVM [10], LPQ+SLPM+NN [11], and HOG+ SRC [44] et al.) and some deep methods (e.g. E-PCANet [20], K-PCANet [21], and DCNN [38] et al.). **Tables 1-3** showed the average accuracies obtained by different methods on the three datasets. From these tables, we see that our approach achieved the best performance. Our approach achieved the highest average accuracy of 80.28% and 96.31% on the JAFFE and CK+ datasets, respectively. Also, the proposed approach outperforms all other methods with a significant advantage on the KDEF dataset. To wrap up, our approach is superior to some handcrafted methods and even better than some of the deep learning methods.

Table 1 State-of-the-art FER accuracies on the JAFFE dataset

Methods	Type	Avg accuracy (%)
LTeP+SVM [10]	Handcrafted	67.14
LPQ+SLPM+NN [11]	Handcrafted	67.61
E-PCANet [20]	Deep	69.40
K-PCANet [21]	Deep	68.80
Feature fusion [39]	Handcrafted	70.00
Kas et al. [40]	Handcrafted	77.62
Proposed	Handcrafted	**80.28**

3.2 Comparison with different DL methods

We also compare our approach with some different DL methods including SRC [47], SVGDL [48], and DPL [49]. The average accuracies under different number of atoms are shown in **Fig. 2**, respectively. From **Fig. 2**, we observe

Table 2 State-of-the-art FER accuracies on the CK+ dataset

Methods	Type	Avg accuracy (%)
CNN+AFM [41]	Deep	89.84
WPLBP [42]	Handcrafted	91.72
LPQ+SLPM+NN [11]	Handcrafted	94.61
E-PCANet [20]	Deep	85.66
DCNN [38]	Deep	94.44
MSCNN [43]	Deep	95.54
Proposed	Handcrafted	**96.13**

Table 3 State-of-the-art FER accuracies on the KDEF dataset

Methods	Type	Avg accuracy (%)
HOG+SRC [44]	Handcrafted	78.00
DFD [45]	Handcrafted	82.24
MobileNet [46]	Deep	73.74
E-PCANet [20]	Deep	80.61
K-PCANet [21]	Deep	80.20
Proposed	Handcrafted	**84.69**

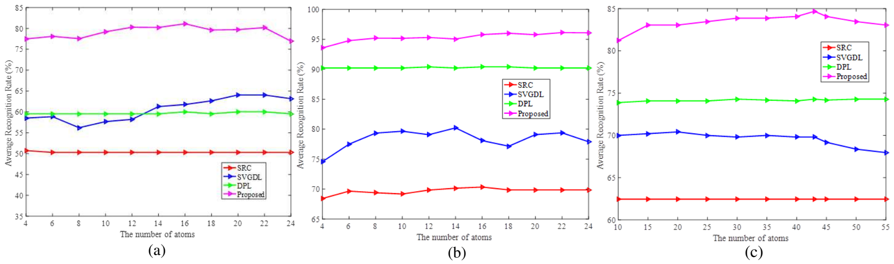
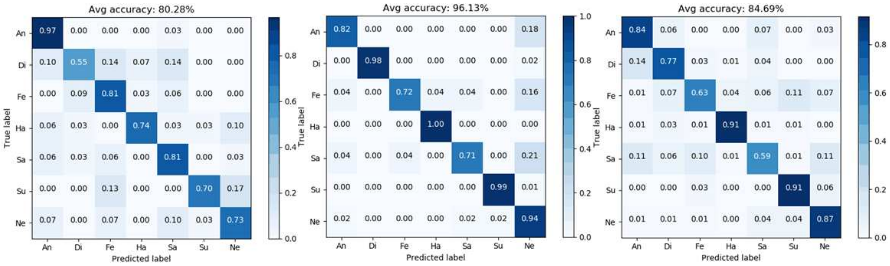


Fig. 2 The average recognition rates with different numbers of atoms on the **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively

that our approach produced the highest average recognition rate on the three datasets, and the overall performance of our approach can be improved by the increasing numbers of dictionary atoms. This is mainly because our approach jointly learns the multiple feature dictionaries, whereas comparing DL methods are only adopting single dictionary learning that ignored the representation diversities among multi-dictionaries. Also, our proposed approach performs a global optimization of the dictionary, which can obtain the overall optimal solutions. Besides, the number of dictionaries obtained by our approach is seven times more than the original dictionary, which not only achieves the diversity of the feature dictionary but also provides the diversity of dictionary atoms during the dictionary update process. Based on the results, we can conclude that the performance of our approach is superior to the other DL methods.

3.3 Confusion matrix graphs

The confusion matrices for our approach on the three datasets are shown in **Fig. 3**. From the diagonal values in **Fig. 3**, we observe that our approach provides promising performance in most classes. For example, **Fig. 3a** showed that our approach obtained an average accuracy of 80.28% on the JAFFE dataset. The proposed approach also achieves better accuracies in the other two datasets (see **Fig. 3b** and **3c**). Despite the small number of incorrect classifications caused by the similar movements of facial muscles (e.g., anger, fear, and neutral expressions), the diagonal results across each matrix in **Fig. 3** indicate that most expressions on the three datasets are classified correctly,



(a) (b) (c)

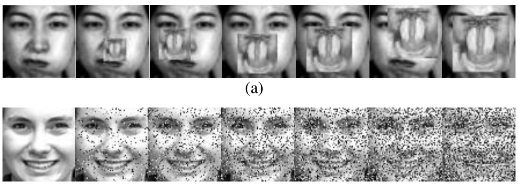
Fig. 3 The confusion matrix on the **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively

thus verifying that our approach is indeed reliable and performs better in terms of classification.

3.4 Effects of block occlusion and random corruption

To verify the robustness of our approach, block occlusion and random corruption were added to all testing samples, with the ratios ranging from 0 to 0.6. **Fig. 4** showed the sample images with different block occlusion and random corruption variances. We can observe that the images are increasingly obscured with the increase of the ratio and we can hardly distinguish the exact expression information when the samples were damaged severely (e.g., ratio $\geq$ 0.4).

Due to block occlusion and random corruption were randomly distributed in the facial images, we did experiments on the three datasets for ten times to obtain the final results (average accuracy $\pm$ standard deviation), as respectively shown in **Fig. 5** and **Fig. 6**. From these figures, we can intuitively see that the average accuracies decrease with the increasing of the block occlusion ratio and random corruption ratio. The reason here is that the ratio increases but the sparsity gets worse. Even in this case, our approach still has a significant advantage than the baseline method (that only use the original dictionary). This is because the proposed multi-class differentiation features better filters the occluded parts that benefit for the subsequent joint dictionary learning. Overall, our proposed approach has stronger robustness than the baseline method.



(a)

(b)

Fig. 4 Sample images with **(a)** block occlusion variances, and **(b)** random corruption variances ranging from (left to right) 0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6, respectively

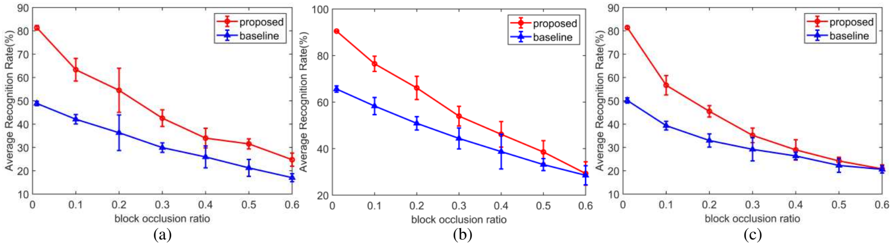


Fig. 5 Comparing the average recognition rate of different approaches under different ratios of block occlusion on the **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively

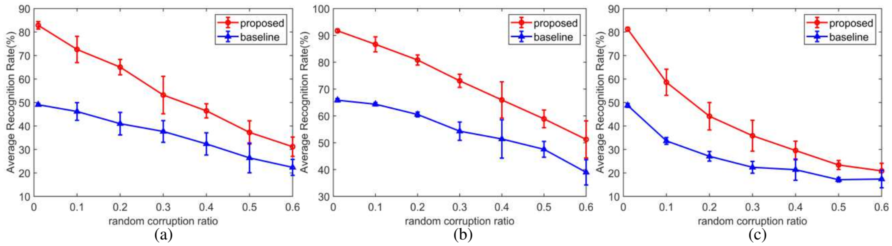


Fig. 6 Comparing the average recognition rate of different approaches under different ratios of random corruption on the **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively

4 Visualization analysis of proposed approach

To better analyze our proposed approach, we first conduct the visualization study from the perspective of the feature representation. The leftmost column (bounded by green rectangles) of **Fig. 7** represents a query image from the JAFFE, CK+ and KDEF datasets, respectively. The right column (bounded by blue rectangles) stands for images obtained by using seven classes to reconstruct the query sample. From the right column and the images bounded by red rectangles are from the true class in **Fig. 7**, we can observe that our approach can make the true class sensitive. Consequently, we can conclude that only the class-related dictionaries can be sensitive, which stand out the significant advantage of our approach during the feature representation step.

We also use **Fig. 8** to show the visualization of the sample distribution from two representation spaces (original space and optimal feature space, respectively). From the upper row of **Fig. 8**, we can observe that samples were non-linear and cluttered distribution because seven facial expressions had not been separated in the original space. From the lower row of **Fig. 8**, the atoms from the optimal feature space gradually form different separated clusters. This is reasonable that dictionary atoms from optimal feature dictionary can correspond to true class labels. Thus, our method can learn the dictionary with strong representation ability and effectively discriminate facial expression features.

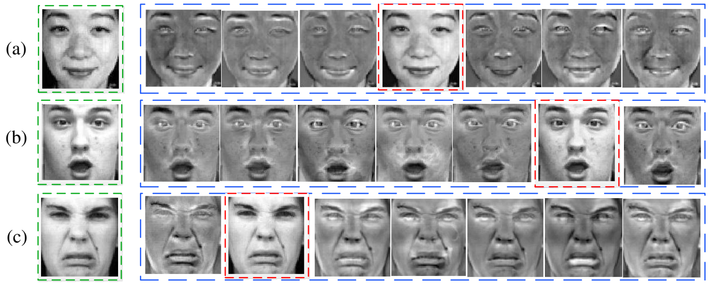


Fig. 7 A query image (bounded by green rectangles) from **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively. Images obtained by using seven classes to reconstruct the query samples in the blue rectangles. Images bounded by red rectangles are from the true class

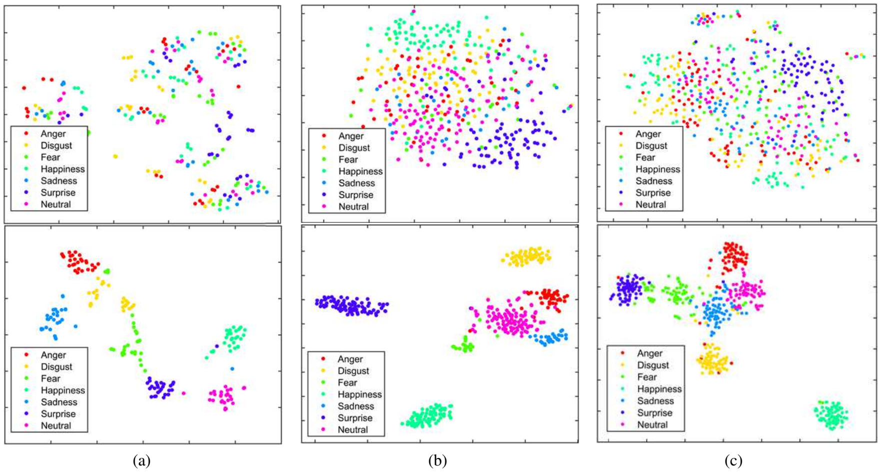


Fig. 8 A visualization of the sample distribution under two representations on the **(a)** JAFFE, **(b)** CK+, and **(c)** KDEF datasets, respectively. Top to bottom in each subgraph: samples from the original space and atoms from the optimal feature space

5 Conclusion and future works

In this paper, we proposed a novel feature learning approach that presented the multi-class differentiation representation guided joint feature dictionaries learning for FER. Firstly, we obtained the multi-class differentiation feature dictionaries aiming to increase the linear separability among expression samples. Secondly, we further jointly learned the multiple feature dictionaries by optimizing the resolutions of each feature dictionary, aiming to establish the strong relationship and enhance the representation ability among multiple feature dictionaries. Comprehensive experiment conducted on the JAFFE, CK+, and KDEF datasets showed that our approach can be used for small-scale datasets and have the promising performance. Although our approach obtained higher accuracies, it was only applicable to static facial images in a laboratory environment and in fact the real facial expressions were complex and real-time changing. In the future, we plan to research a more universal model to analyze expression recognition in realistic scenarios as well as in dynamic video sequences.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grants 62001413, the Natural Science Foundation of Hebei Province under Grants F2020203064, and the Doctoral Foundation of Yanshan University under Grant BL18033.

References

[1] T. S. Ma, W. H. Tian, Y. L. Xie, Multi-level knowledge distillation for low-resolution object detection and facial expression recognition, Knowledge-Based Systems, 240 (2022) 108136.

[2] D. Poux, B. Allaert, N. Ihaddadene, I. M. Bilasco, C. Djeraba, M. Bennamoun, Dynamic Facial Expression Recognition Under Partial Occlusion With Optical Flow Reconstruction, IEEE Transactions on Image Processing, 31 (2022) 446-457.

[3] A. H. Reddy, K. Kolli, Y. L. Kiran, Deep cross feature adaptive network for facial emotion classification, Signal, Image and Video Processing, 16 (2022) 369–376.

[4] I. Chaturvedi, Q. Chen, E. Cambria, Landmark calibration for facial expressions and fish classification, Signal, Image and Video Processing, 16 (2022) 377–384.

[5] D. Chen, P. Song, W. Zheng, Learning Transferable Sparse Representations for Cross-corpus Facial Expression Recognition, IEEE Transactions on Affective Computing (2021) 3077489.

[6] Y. Liu, B. Li, Y. Lai, Sparse MDMO: Learning a Discriminative Feature for Micro-Expression Recognition, IEEE Transactions on Affective Computing, 12 (2021) 254-261.

[7] Y. Yan, Z. Z. Zhang, S. Chen, H. Z. Wang, Low-resolution facial expression recognition: A filter learning perspective, Signal Processing, 169 (2020) 107370.

[8] N. Y. Zeng, H. Zhang, B. Y. Song, W. B. Liu, Y. R. Li, Abdullah M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing, 273 (2018) 643-649.

[9] W. S. Chu, F. De la Torre, J. F. Cohn, Selective Transfer Machine for Personalized Facial Expression Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2017) 529-545.

[10] F. Bashar, A. Khan, F. Ahmed, M. H. Kabir, Robust facial expression recognition based on median ternary pattern (MTP), International Conference on Electrical Information and Communication Technology, (2014) 1-5.

[11] C. Turan, K. M. Lam, Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study, Journal of Visual Communication and Image Representation, 55 (2018) 331-341.

[12] W. M. Yu, H. Xu, Co-attentive multi-task convolutional neural network for facial expression recognition, Pattern Recognition, 123 (2022) 108401.

[13] Q. Zhu, Q. R. Mao, H. J. Jia, O. Elias N. Noi, J. J. Tu, Convolutional relation network for facial expression recognition in the wild with few-shot learning, Expert Systems with Applications, 189 (2022) 116046.

[14] J. Y. Han, L. Du, X. Q. Ye, L. Zhang, J. F. Feng, The devil is in the face: Exploiting harmonious representations for facial expression recognition, Neurocomputing, 486 (2022) 104-113.

[15] F. Xue, Q. Wang, G. Guo, TransFER: Learning Relation-aware Facial Expression Representations with Transformers, International Conference on Computer Vision, (2021) 3581-3590.

[16] Z. Zhao, Q. Liu, S. Wang, Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild, IEEE Transactions on Image Processing, 30 (2021) 6544-6556.

[17] Q. H. Huang, C. Q. Huang, X. Z. Wang, F. Jiang, Facial expression recognition with grid-wise attention and visual transformer, Information Sciences, 580 (2021) 35-54.

[18] N. Sun, Q. Li, R. Z. Huan, J. X. Liu, G. Han, Deep spatial-temporal feature fusion for facial expression recognition in static images, Pattern Recognition Letters, 119 (2019) 49-61.

[19] T. H Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, PCANet: A Simple Deep Learning Baseline for Image Classification? IEEE Transactions on Image Processing, 24 (2015) 5017-5032.

[20] Z. Sun, R. Chiong, Z. P. Hu, An extended dictionary representation approach with deep subspace learning for facial expression recognition, Neurocomputing, 316 (2018) 1-9.

[21] Z. Sun, Z. P. Hu, R. Chiong, M. Wang, W. He, Combining the Kernel Collaboration Representation and Deep Subspace Learning for Facial Expression Recognition, Journal of Circuits Systems and Computers, 27

(2018) 1850121.

[22] Z. Sun, R. Chiong, Z. P. Hu, S. Dhakal, A dynamic constraint representation approach based on cross-domain dictionary learning for expression recognition, Journal of Visual Communication and Image Representation, 85 (2022) 103458.

[23] A. B. Tanfous, H. Drira, B. B. Amor, Sparse Coding of Shape Trajectories for Facial Expression and Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2020) 2594-2607.

[24] Z. Sun, R. Chiong, Z. P. Hu, Self-adaptive feature learning based on a priori knowledge for facial expression recognition, Knowledge-Based Systems, 204 (2020) 106124.

[25] K. Y. Yan, W. M. Zheng, Z. Cui, Y. Zong, T. Zhang, C. G. Tang, Unsupervised facial expression recognition using domain adaptation based dictionary learning approach, Neurocomputing, 319 (2018) 84-91.

[26] M. Yang, L. Zhang, X. Feng, Fisher discrimination dictionary learning for sparse representation, IEEE International Conference on Computer Vision, (2011) 543-550.

[27] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing, 54 (2006) 4311-4322.

[28] A. Moeini, K. Faez, H. Moeini, A. M. Safai, Facial expression recognition using dual dictionary learning, Journal of Visual Communication and Image Representation, 45 (2017) 20-33.

[29] X. Luo, Y. Xu, J. Yang, Multi-resolution dictionary learning for face recognition, Pattern Recognition, 93 (2019) 283-292.

[30] P. Song, X. Deng, J. F .C. Mota, N. Deligiannis, P. L. Dragotti, M. R. D. Rodrigues, Multimodal Image Super-Resolution via Joint Sparse Representations Induced by Coupled Dictionaries, IEEE Transactions on Computational Imaging, 6 (2020) 57-72.

[31] F. Juefei-Xu, D. K. Pal, M. Savvides, NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction, IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2015) 141-150.

[32] Y. L. Peng, L. P. Li, S. G. Liu, T. Lei, Space-frequency domain based joint dictionary learning and collaborative representation for face recognition, Signal Processing, 147 (2018) 101-109.

[33] G. Q. Zhang, F. Porikli, H. J. Sun, Q. S. Sun, G. Y. Xia, Y. H. Zheng, Cost-sensitive joint feature and dictionary learning for face recognition, Neurocomputing, 391 (2020) 177-188.

[34] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2009) 210-227.

[35] M. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21 (1999) 1357-1362.

[36] P. Lucey, F. C. Jeffrey, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion specified expression, Computer Vision and Pattern Recognition, (2010) 94-101.

[37] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, IEEE International Conference on Automatic Face and Gesture Recognition, (1998) 200-205.

[38] S. Ouellet, Real-time emotion recognition for gaming using deep convolutional network features, ArXiv, (2014) 1408.3750.

[39] A. Poursaberi, H. A. Noubari, M. Gavrilova, S. N. Yanushkevich, Gauss-Laguerre wavelet textural feature fusion with geometrical information for facial expression identification, EURASIP Journal on Image and Video Processing, 17 (2012) 1-13.

[40] M. Kas, Y. El merabet, Y. Ruichek, R. Messoussi, New framework for person-independent facial expression recognition combining textural and shape analysis through new feature extraction approach, Information Sciences, 549 (2021) 200-220.

[41] B .F. Wu, C. H. Lin, Adaptive feature mapping for customizing deep learning based facial expression recognition model, IEEE Access, 6 (2018) 12451-12461.

[42] L. S. Du, H. F. Hu, Weighted Patch-based Manifold Regularization Dictionary Pair Learning model for facial expression recognition using Iterative Optimization Classification Strategy, Computer Vision and Image Understanding, 186 (2019) 13-24.

[43] K. Zhang, Y. Huang, Y. Du, L. Wang, Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks, IEEE Transactions on Image Processing, 26 (2017) 4193-4203.

[44] K. Nurzynska, Emotion recognition: The influence of texture's descriptors on classification accuracy, Communications in Computer and Information Science, 716 (2017) 427-438.

[45] Z. Sun, Z. P. Hu, M. Wang, S. H. Zhao, Discriminative feature learning-based pixel difference representation for facial expression recognition, IET Computer Vision, 11 (2017) 675-682.

[46] S. Guo, L. Feng, Z.B. Feng, Y.H. Li, Y. Wang, S.L. Liu, H. Qiao, Multi-view laplacian least squares for human emotion recognition, Neurocomputing, 370 (2019) 78-87.

[47] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust Face Recognition via Sparse Representation, IEEE Trans. Pattern Anal. Mach. Intell, 31 (2009) 210–227.

[48] S. Cai, W. Zuo, L. Zhang, Support vector guided dictionary learning, Proceedings of the European Conference on Computer Vision, Springer International Publishing, (2014) 624–639.

[49] S. Gu, L. Zhang, W. Zuo, X. Feng, Projective dictionary pair learning for pattern classification, Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS), (2014) 793-801.