



Improved image representation and sparse representation for image classification

Shijun Zheng¹ · Yongjun Zhang¹ · Wenjie Liu¹ · Yongjie Zou¹

Published online: 10 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

It seems that for multiple available images of the same object, the pixel values at the same image position are almost always different, which is especially obvious for the deformable object. This implies that it will be not easy to correctly classify the deformable object. In order to extract salient features of images and improve the performance of image classification, a novel image classification algorithm is proposed in this paper. The algorithm can effectively preserve the large-scale information and global features of the original image, reduce the difference in different images of the same object, and significantly improve the accuracy of image classification. Firstly, the virtual image is generated by the new image representation procedure. Secondly, the image classification algorithm is used to obtain the corresponding classification scores of the original image and the virtual image, respectively. Finally, the ultimate classification score is obtained by a simple and efficient score fusion scheme. A large number of experiments on three widely used image databases show that the proposed algorithm outperforms other state-of-the-art algorithms in classification accuracy. At the same time, the algorithm has the advantages of simple implementation and high computational efficiency.

Keywords Image representation · Image classification · Sparse representation

1 Introduction

In the image classification task, applying multiple representations to an object can effectively improve the classification accuracy. In particular, it is almost a common way to exploit multiple training samples of each object for classifying a test sample if they are available. In the field of face recognition, face images show great differences because of different facial expressions, lighting and postures [1–3], which causes

great difficulties in face recognition. In order to improve the accuracy of face recognition, many researchers have proposed different methods. For example, the expression invariant face recognition algorithm can effectively improve the recognition accuracy of face images with different expressions [4]. Jian et al. proposed a face recognition method based on illumination compensation [5]. Sharma et al. proposed a face recognition method based on position-invariant virtual classifiers [6]. Considering the symmetry of face images, Xu et al. proposed a method to generate “symmetrical” face images using original images. Combining original images and symmetrical face images can better reduce the impact of image appearance changes and improve the classification accuracy [7]. In addition, using the original image to generate a virtual image and providing multiple representation methods for the same face image can well reduce the error rate of face recognition [8]. Similarly, if an original face image is corrupted by noisy, it can also be used as a virtual image [9, 10]. In recent years, dictionary learning has been widely used in image classification and face recognition. In order to improve the robustness of face recognition, Wang et al. proposed a method of Discriminative and Common hybrid Dictionary Learning (DCDL) [11]. Xu et al. proposed a new dictionary learning framework

✉ Yongjun Zhang
zyj6667@126.com

Shijun Zheng
shijunzheng@163.com

Wenjie Liu
Udellliu@163.com

Yongjie Zou
yongjiezou@126.com

¹ Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, College of Computer Science and Technology, Guizhou University, Guiyang, China

that can effectively represent face images and enhance the diversity of face images [12]. Robust, Discriminant and Comprehensive Dictionary Learning Method (RDCDL) is a new dictionary learning method recently proposed by Lin et al. [13], which can also effectively improve the ability of image classification. The multi-resolution dictionary learning method can well enhance the robustness to noise of the dictionary by virtue of different resolutions [14].

After the naive sparse representation classification (SRC) algorithm was proposed, it has been widely used in the fields of image processing and face recognition. For example, in face recognition, Xu et al. proposed a sparse representation method based on l_2 regularization [15], which can achieve a noticeable precision improvement. Sparse representation is increasingly applied to image classification [15–19], image super-resolution [20, 21], image denoising [22], and image alignment [23]. At the same time, various sparse representation algorithms have been proposed one after another. Moreover, dictionary learning, as an important kind of methods directly related with sparse representation, is getting more and more attention from researchers. We generally divide the sparse representation algorithms into two categories, the first category is based on the original training sample, and the second category is based on the dictionary. The sparse representation based on the original training sample uses the training sample to linearly represent the test sample, and the dictionary-based sparse representation uses a dictionary to represent the test sample, which is generated from the set of the original training samples [17, 24–30]. The sparse representation algorithm based on the original training samples contains a large number of examples, such as the orthogonal matching pursuit (OMP) [31], L1-regularized least squares (L1LS) [32], the primal augmented Lagrangian method (PALM) [33], the dual augmented Lagrangian method (DALM) [33], fast iterative shrinkage and thresholding algorithm (FISTA) [34], two-phase test sample sparse representation (TPTSR) [35], collaborative representation (CRC) [36], etc. These algorithms have good results in face recognition.

In this paper, we propose a novel image classification algorithm based on the idea of sparse representation. The algorithm can better preserve the large-scale information and holistic features of the original image, and reduce the difference in different images of the same object, which is greatly beneficial to the image classification task. The improved algorithm has the following important property: if the original image is a gray image, when its pixel value varies between 0 and 255, the pixels of the virtual image generated by the improved algorithm have symmetrical values. In particular, the pixel values are symmetrical with respect to $\sqrt[3]{127 * 128}$. The improved algorithm combines generated new image representations and original representations to perform classification. We

conducted experiments on several facial image databases. The experimental results effectively verified that the improved image classification algorithm can achieve significant performance in image classification.

The other parts of the paper are organized as follows. Section 2 describes the principles and steps of the proposed algorithm and proposes two methods for generating virtual images. Section 3 explains the characteristics and advantages of the proposed algorithm intuitively. Section 4 presents a lot of experimental results and analyzes the results. Section 5 provides the conclusions of the paper.

2 Algorithm principle and steps

2.1 Algorithm steps

In this section, we will explain the steps of the algorithm in detail. The main steps of the algorithm are as follows.

Step 1: Select training samples and test samples. All original images are divided into two parts: training samples and test samples.

Step 2: Obtain the virtual training samples using (15) or (16).

The original training sample is expressed as follows:

$$A = (A_1, A_2, \dots, A_C) \quad (1)$$

The virtual image generated by the training sample is recorded as V :

$$V = (V_1, V_2, \dots, V_C) \quad (2)$$

C is the number of classes, and each class has n training samples. The original and virtual training samples of the i -th class are denoted as A_i and V_i respectively, where $A_i = (a_{i1}, a_{i2}, \dots, a_{in})$, $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$. Each column vector a_{ij} in A_i represents the j -th image of the i -th class. The v_{ij} in V_i is also a column vector, which represents the virtual image generated by the j -th training sample of the i -th class.

Step 3: Let y denote the test sample, and y_v is the virtual test sample generated by y . According to CRC, we can express y by a linear combination of all training samples, as follows:

$$y = \sum_{k=1}^{Cn} a_k x_k \quad (3)$$

a_k denotes the k -th training sample, which is the k -th column in matrix A , and x_k is the coefficient of a_k . We rewrite (3) as follows:

$$y = Ax \quad (4)$$

We use the regularized least squares method to solve the equation. As a result, we can obtain the linear combination coefficient x :

$$x = (AA^T + \lambda I)^{-1} A^T y \tag{5}$$

Similarly, we can get the linear combination coefficient β of the virtual training samples:

$$\beta = (V V^T + \lambda I)^{-1} V^T y_v \tag{6}$$

λ is a constant, I is the identity matrix.

Step 4: Let d_o^i denote the distance (i.e. score) between the test sample and the original training sample of the i -th class, and d_v^i denote the distance between the test sample and the virtual training sample of the i -th class:

$$d_o^i = \| y - A_i x_i \|_2 \quad i = 1, 2, \dots, C \tag{7}$$

$$d_v^i = \| y_v - V_i \beta_i \|_2 \quad i = 1, 2, \dots, C \tag{8}$$

x_i and β_i are the coefficient vectors of the training samples and the virtual training samples of the i -th class, respectively, where $x^T = ((x_1)^T, (x_2)^T, \dots, (x_C)^T)$, $\beta^T = ((\beta_1)^T, (\beta_2)^T, \dots, (\beta_C)^T)$. $\| \cdot \|_2$ represents the L_2 norm.

Step 5: In this step we introduce the process of merging the results of the original training samples with the virtual training samples. We adopt a simple and efficient fusion method proposed in [37]. Let $S_o^1, S_o^2, \dots, S_o^C$ and $S_v^1, S_v^2, \dots, S_v^C$ denote the ascending ordering results of d_o^i and d_v^i , respectively. Then the fusion weights W_1 and W_2 are calculated by following equations:

$$w_{10} = S_o^2 - S_o^1 \tag{9}$$

$$w_{20} = S_v^2 - S_v^1 \tag{10}$$

$$W_1 = \frac{w_{10}}{w_{10} + w_{20}} \tag{11}$$

$$W_2 = 1 - W_1 \tag{12}$$

Finally, the ultimate fusion result is obtained as follows:

$$d_i = W_1 d_o^i + W_2 d_v^i \quad i = 1, 2, \dots, C \tag{13}$$

We classify the test sample by the following formula:

$$p_j = \underset{i}{\operatorname{arg\,min}} d_i \tag{14}$$

If p_j is the minimum value of d_i , then the test sample is classified into the j -th class.

2.2 Generating of virtual images

In this paper, we propose an improved image representation method to represent the original image. A transform of the original image is attained using the improved image

representation method. We call the transform virtual image. How to generate virtual images is introduced as follows.

We take the gray image as an example to illustrate how the original image generates its corresponding virtual image. The maximum pixel value in the gray image is 255, denoted as P_{max} . The pixel value at the r -th row and the c -th column of the original image is recorded as S_{rc} , the generated virtual image is represented by V , and the pixel value at the r -th row and the c -th column of the virtual image is recorded as V_{rc} . The representation of the original image to generate the virtual image is as follows:

$$V_{rc} = \sqrt[2]{S_{rc} \cdot (P_{max} - S_{rc})} \tag{15}$$

By analyzing the above formula, we can draw the following conclusions:

- (1) If S_{rc} is equal to 0 or the maximum pixel value of the image, the virtual image has a pixel value of 0 at the corresponding position.
- (2) If S_{rc} is closer to $\frac{P_{max}}{2}$, the pixel value of the corresponding position of the virtual image is larger, and the maximum value is $\sqrt[2]{127 * 128}$.
- (3) When the pixel value of the original image is S_{rc} or $(P_{max} - S_{rc})$, the pixel values of the corresponding positions in the virtual image are the same. In other words, the pixel values in the virtual image are symmetric with respect to $\sqrt[2]{127 * 128}$.

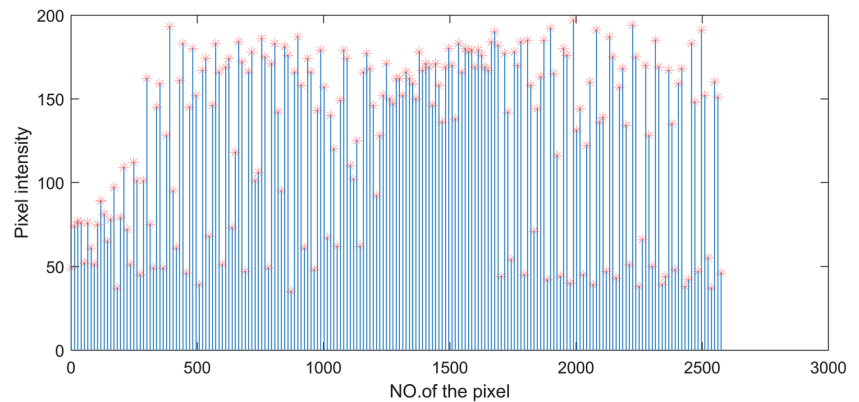
It has been proved in [37] that medium-intensity pixels have strong stability and are more conducive to image classification. Compared to other similar methods, the pixel value of the virtual image generated by our method is significantly reduced and better concentrated near the medium-intensity pixel. Moreover, two pixels with similar pixel values in the original training samples have smaller differences in the virtual image, which can greatly improve the performance of image classification.

If the original image is a gray image, the maximum value of the virtual image generated by our method is $\sqrt[2]{127 * 128}$, which not only ensures that the pixel value of the virtual image is distributed near the medium intensity, but also



Fig. 1 The first face image of the first subject in the ORL face database

Fig. 2 Original pixel value of the first sample of the first subject in the ORL face database



is beneficial to preserve the large-scale information of the original image. To some extent, the virtual image properly highlights the global features of the image, which will be very beneficial for image classification.

Based on the principle of the image representation method proposed above, we also proposed another new scheme to generate virtual images. It is shown by a large number of experiments that the method can significantly improve the image classification accuracy. The equation for generating a virtual image is as follows:

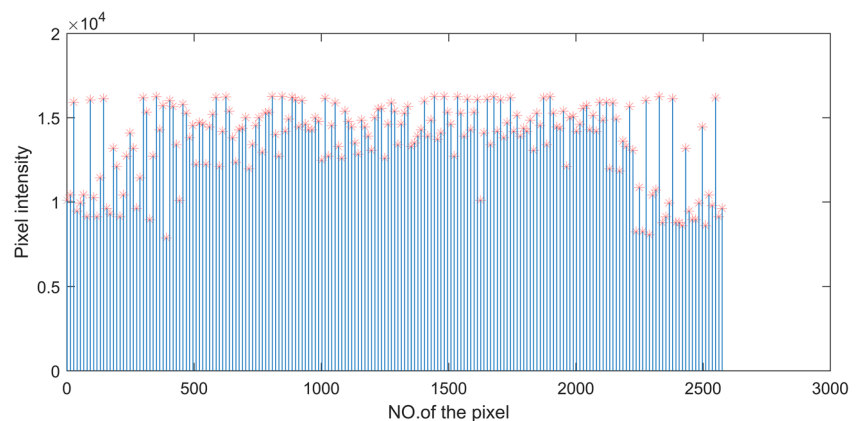
$$V_{rc} = \sqrt[3]{S_{rc} \cdot (P_{max} - S_{rc})} \quad (16)$$

Compared to (15), the virtual image generated by (16) has a smaller pixel value and the difference between pixels is smaller, which seems to be easier to obtain global information of the image.

3 Algorithm analysis

In this section, we mainly analyze the characteristics and advantages of the proposed algorithm. By comparison with the algorithm proposed in [37], the face recognition experiment is taken as an example to intuitively explain the principle of the algorithm on basis of (15). We select the first face image of the first subject in the ORL face database

Fig. 3 Pixel value of the first virtual image of the first subject in the ORL face database attained using the algorithm in [37]



as an example for analysis. This is a gray image. The face image is shown in Figs. 1 and 2 shows the distribution of the original pixels of the sample.

The pixel value distribution of the virtual image generated by the algorithm in [37] is shown in Fig. 3. When the pixel value of the original training sample changes from 0 to 255, the symmetry in pixel values of the virtual image are shown in Fig. 4.

Figure 5 shows the pixel value distribution of the virtual image generated by the proposed algorithm, and Fig. 6 shows the symmetry in pixel values of the virtual image as the pixel value of the original image changes from 0 to 255.

Figure 7 shows the normalized data distribution of the same sample under the original image, the algorithm in [37], and the proposed algorithm. Normalization means converting an image vector into a unit vector with a norm of 1.

According to Figs. 3 and 4, it is intuitively reflected that the virtual image generated by the algorithm in [37] has a very large pixel value, far exceeding the pixel range of the conventional gray image. Moreover, the two pixels with similar pixel values of the original image have great differences in the virtual image, which causes large-scale information of the original image to be lost. These problems are not conducive to image classification. By comparing Figs. 5 and 6, we see that the pixel value of the virtual image generated by our algorithm is significantly reduced, and

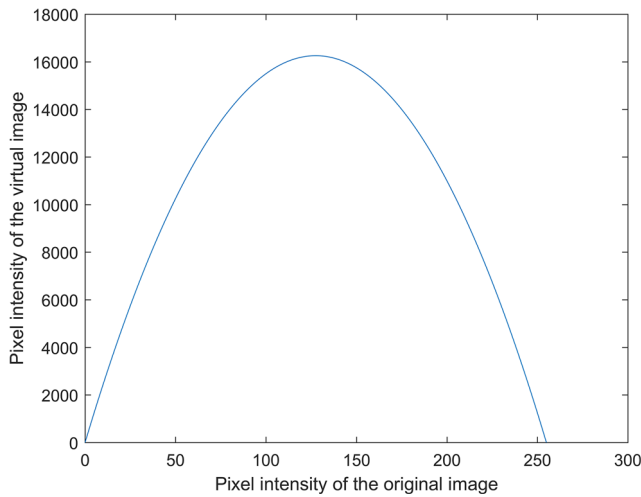


Fig. 4 Symmetry in pixel values of the virtual images attained using the algorithm in [37]

the maximum pixel value is $\sqrt[3]{127 * 128}$. In addition, the difference between two pixels in a virtual image is greatly reduced. The large-scale information of the original image is well preserved in the virtual image. Meanwhile, in the proposed algorithm, for gray images, pixels of intensity i and $(255 - i)$ have the same intensity in the virtual image, and the pixels whose values are closer to the medium-intensity play a more important role.

Figure 7 shows that the virtual image generated by our proposed algorithm has a low correlation with the original image, which indicates that the original image and the obtained virtual image are complementary.

In summary, the proposed algorithm can obtain more abundant large-scale information, and to some extent, more information corresponding to the global feature of the image. As we know, large-scale and global information is more important for recognition of object appearances. Therefore, our proposed algorithm has greater precision for image classification.

Similarly, the algorithm for generating virtual images using (16) also has the above characteristics and advantages. Figure 8 shows the pixel value of the virtual image

Fig. 5 Pixel value of the first virtual image of the first subject in the ORL face database using the proposed algorithm

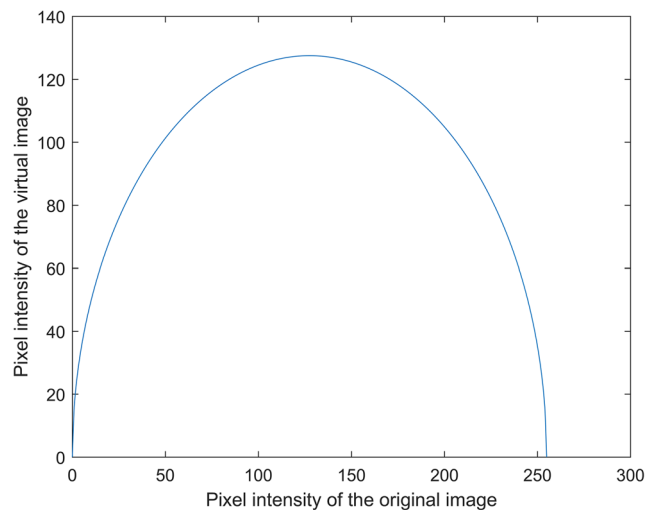
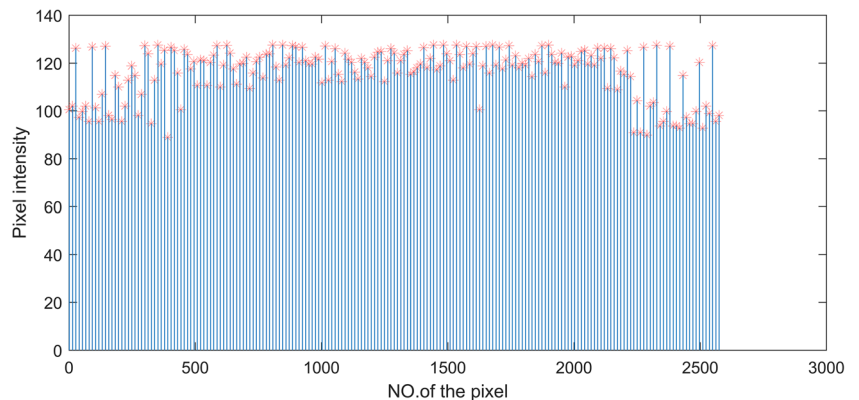


Fig. 6 Symmetry in pixel values of the virtual images attained using the proposed algorithm

generated by the algorithm, and Fig. 9 shows the symmetry in pixel values of the virtual image when the pixel of the original image changes between 0 and 255.

Figure 10 shows eight original face images (line 1) and virtual images generated by the algorithm in [37] (line 2) and virtual images generated by the proposed algorithm (line 3) of a subject in the Georgia Tech face database.

We can find that the virtual image generated by the algorithm in [37] or the proposed algorithm is a relatively natural face image. Although there are some great differences between virtual image and original image in appearance, the fusion of the virtual image and the original image can provide multiple representation methods for the same face image, which is beneficial to improve the accuracy of face classification.

4 Experimental and results

In this section, we verify the feasibility and rationality of the proposed algorithm through a large number of experiments.

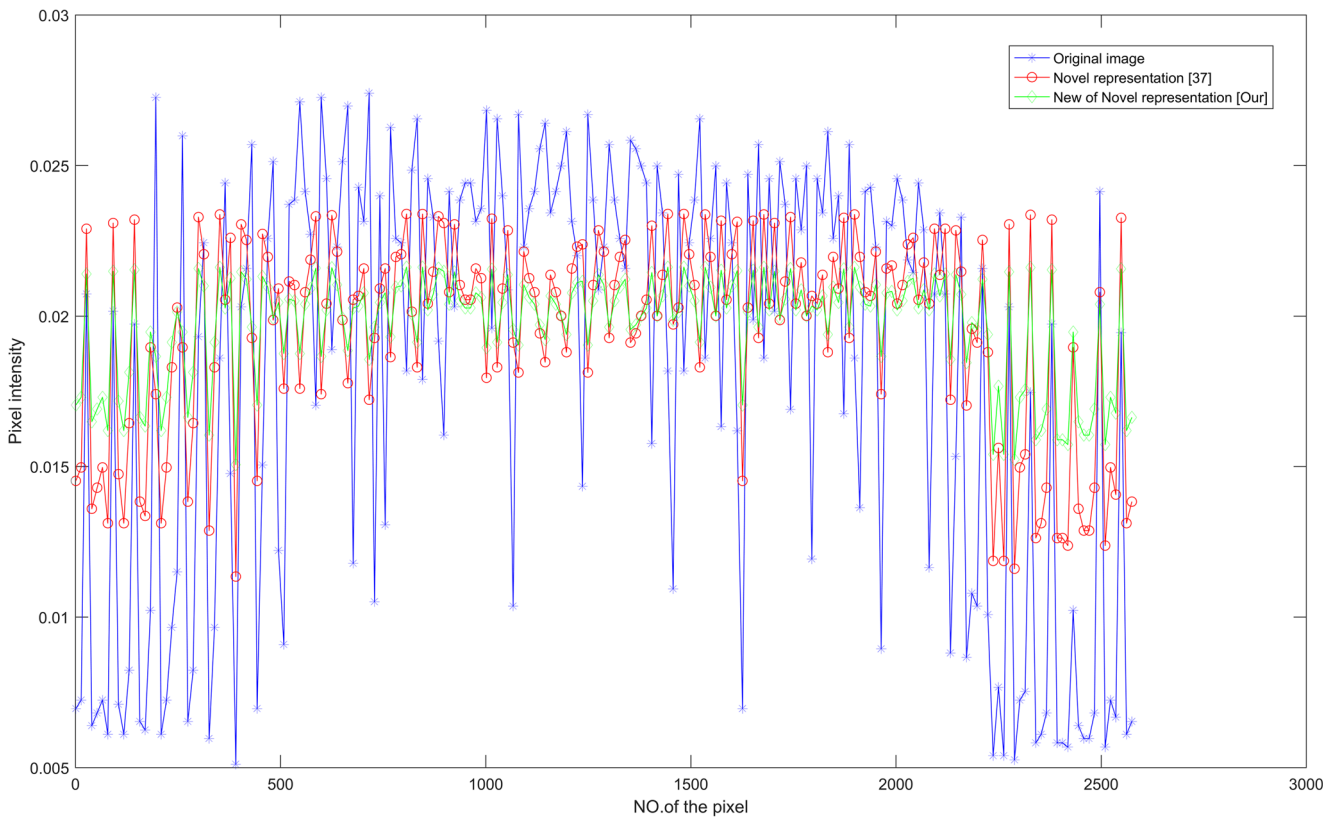


Fig. 7 Normalized data distribution of the first sample of the first subject in the ORL face database under the original image, the algorithm in [37], and the proposed algorithm

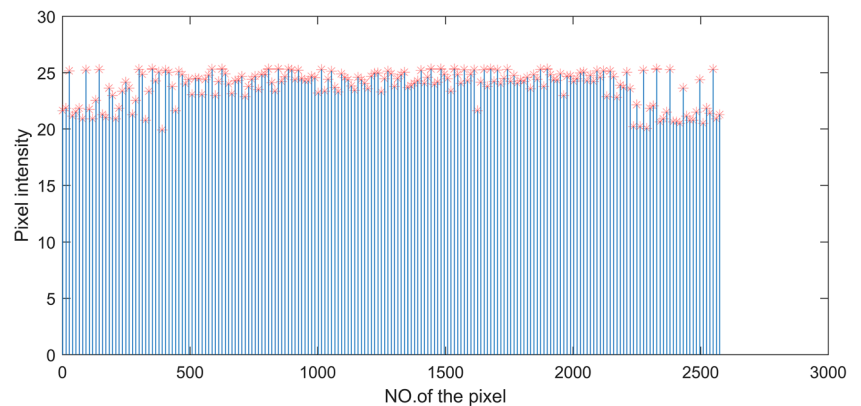
We conducted experiments on three face databases, namely ORL face database, Georgia Tech face database and FERET face database. The experimental results show that the proposed algorithm has a greater precision improvement than other similar algorithms in face image classification.

In order to better reflect the advantages of the proposed algorithm, we compared it with the typical sparse representation algorithms, such as L1-regularized least squares (L1LS), the primal augmented Lagrangian method (PALM), fast iterative shrinkage and thresholding algorithm (FISTA). Then, it compares with the new algorithm proposed in

recent years. For example, the multi-resolution dictionary learning method proposed [14], Robust Sparse Linear Discriminant Analysis (RSLDA) [38], block-diagonal low-rank representation (BDLRR) [39] and the improved collaborative representation [37]. In addition, the methods of applying collaborative representation, PALM, L1LS, and FISTA directly on the original image are called naive collaborative representation, naive PALM, naive L1LS, naive FISTA, respectively.

In all experiments, the face images of each subject were divided into two parts: the training set and the test set. The

Fig. 8 Pixel value of the first sample of the first subject in the ORL face database attained using (16)



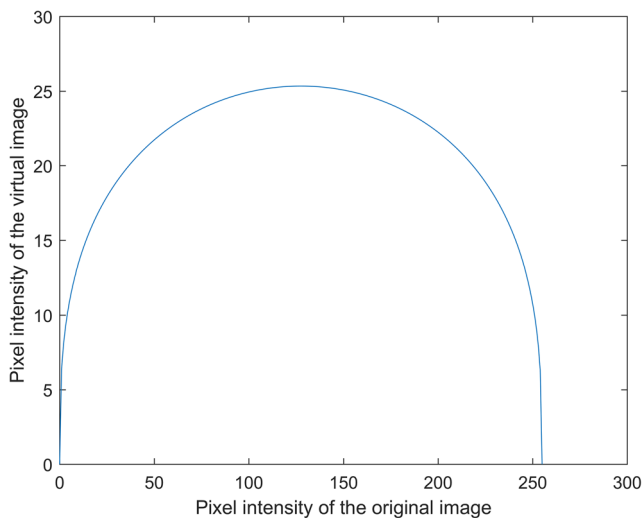


Fig. 9 Symmetry in pixel values of the virtual image when using (16)

training set and test set are mutually exclusive, and the sum of the two is the total face images of the subject.

The specific implementation process is: Firstly, we use the improved image representation method to generate the virtual image of the original image, and then apply the sparse representation algorithm to the original image and the virtual image respectively, and obtain the classification scores corresponding to the test image through collaborative representation [36]. The score fusion scheme is used to fuse scores obtained by the original image and the virtual image respectively, and the ultimate classification score of the test sample is obtained. The following is an experimental analysis of different algorithms on various databases.

4.1 Experiments on the ORL database

In this section, we experimented with the proposed algorithm on the ORL face database [40]. The ORL face database contains a total of 40 subjects, each with 10 images

Fig. 10 The first line is the original face image, the second line is the virtual image generated by the algorithm in [37], and the third line is the virtual image generated by the proposed algorithm



and a total of 400 gray face images. These images have different angles, lighting and facial expressions. The facial expressions include different details such as smiling and not smiling, eyes open and closed, glasses and no glasses. In our algorithm, all face images in the ORL database are first adjusted to an image of 56×46 pixels. Figure 11 shows an example of images of two subjects in the database.

On ORL face database, we conducted experimental comparison of different algorithms, including sparse representation algorithm and the newly proposed algorithm. The experimental results are shown in Table 1, which shows the classification error rates of various algorithms on this database.

From Table 1, we can clearly see that our algorithm achieves the best performance when the number of training samples per subject is 2, 3, 4, and 5. For example, when the number of training samples per subject is 5, the error rate of our proposed algorithm (15) is 7.50%. However, the classification error rates of Original collaborative representation [37] and Multi-resolution dictionary learning [14], RSLDA [38], and BDLRR [39] are 8.5%, 9.55%, 8.00%, 8.00%, respectively. When the number of training samples per subject is 4, the error rate of our proposed algorithm (16) is 7.08%. However, the classification error rates of Original collaborative representation [37] and Multi-resolution dictionary learning [14], RSLDA [38], and BDLRR [39] are 8.75%, 16.62%, 10.83%, 10.42%, respectively. Through the experiment of ORL face database, it is verified that our proposed algorithm can significantly improve the accuracy of image classification.

4.2 Experiments on the georgia tech database

In this section, we experiment with the proposed algorithm on Georgia Tech face database [41, 42]. Georgia Tech face database has a total of 50 subjects, each subject has 15 JPEG format color images, a total of 750 color face images. The



Fig. 11 Image examples of two subjects in the ORL database

background of the image in the database is messy, and the resolution of the image is 640x480 pixels. The image of each subject contains a frontal face image and a tilted face image of the subject with different expressions, lighting and scales. Each image is manually labeled to determine the position of the face in the image. In our improved algorithm, the image in the database is processed first, using the face image with the background removed, and each face image is 40×30 pixels. Figure 12 shows the face images of three subjects in the Georgia Tech face database.

The comparison results of image classification error rates of different algorithms on Georgia Tech face database are shown in Table 2.

From Table 2, we can see that when the number of training samples per subject is 1, 2, 3, our algorithm performs better than other algorithms. For example, when the number of training samples per subject is 2, the error rate of our proposed algorithm (15) is 51.69%. However, the classification error rates of Original collaborative representation [37] and Multi-resolution dictionary learning [14], RSLDA [38], and BDLRR [39] are 52.15%, 65.86%, 51.70%, 56.77%, respectively. Similarly, when the number of training samples per subject is 3, the classification error rate of our algorithm (16) is 48.17%, which is 4.00%, 13.08%, 2.49%, and 1.00% lower than Original collaborative representation [37], Multi-resolution dictionary learning [14], RSLDA [38], and

BDLRR [39], respectively. The results show that our proposed algorithm greatly improves the image classification accuracy.

4.3 Experiments on the FERET database

The proposed algorithm was tested on the FERET face database [43, 44]. FERET face database is one of the most widely used face databases in the field of face recognition. It is to collect the images of the subjects under different lighting conditions. The face images of each subject show the characteristics of different posture and facial expressions. In this section, experiments were performed using the "ba", "bj", "bk", "be", "bf", "bd", and "bg" subsets of the FERET face database. There are 1400 gray face images of 200 subjects, and each subject has seven gray face images. Figure 13 shows the face images of three subjects in the FERET face database.

In the experiment, we adjust all human face images to 40×40 pixels. The classification error rate comparison results of different algorithms on the FERET face database are shown in Table 3.

From Table 3, we can see that when the number of training samples per subject is 1, 2, 5, our proposed algorithm has a lower classification error rate than other algorithms. For example, when the number of training samples per

Table 1 Rate of classification errors (%) on the ORL dataset

Number of training samples per class	2	3	4	5
Our proposed algorithm (16)	11.25	10.71	7.08	7.00
Our proposed algorithm (15)	10.94	11.07	6.67	7.50
Improved collaborative representation [37]	12.50	11.43	8.75	8.50
naive collaborative representation	16.56	13.93	10.83	11.50
naive PALM	19.69	18.57	13.57	13.00
naive L1LS	19.69	18.93	14.58	13.50
naive FISTA	18.44	16.79	12.08	13.50
PALM	20.44	11.93	9.88	8.20
Multi-resolution dictionary learning [14]	23.15	18.21	16.62	9.55
RSLDA [38]	24.37	12.85	10.83	8.00
BDLRR [39]	21.87	13.93	10.42	8.00



Fig. 12 Image examples of three subjects in the GT dataset

Table 2 Rate of classification errors (%) on the GT dataset

Number of training samples per class	1	2	3
Our proposed algorithm (16)	63.00	51.23	48.17
Our proposed algorithm (15)	63.29	51.69	48.83
Improved collaborative representation [37]	63.86	52.15	52.17
naive collaborative representation	66.57	57.54	54.67
naive PALM	68.14	61.38	58.33
naive L1LS	68.00	61.08	58.50
naive FISTA	65.14	54.92	53.50
PALM	65.26	55.71	52.29
Multi-resolution dictionary learning [14]	\	65.86	61.25
RSLDA [38]	70.00	51.70	50.66
BDLRR [39]	67.43	56.77	49.17

Fig. 13 Image examples of three subjects in the FERET dataset



Table 3 Rate of classification errors (%) on the FERET dataset

Number of training samples per class	1	2	5
Our proposed algorithm (16)	52.67	39.80	27.25
Our proposed algorithm (15)	52.08	40.10	28.75
Improved collaborative representation [37]	53.08	39.80	30.25
naive collaborative representation	55.67	41.60	31.50
L1LS	73.71	58.88	32.93
FISTA	73.48	59.29	33.33
PALM	77.85	69.60	70.25
Multi-resolution dictionary learning [14]	\	56.12	49.02
RSLDA [38]	83.16	55.30	30.25
BDLRR [39]	71.42	54.20	29.75

subject is 5, the error rate of our proposed algorithm (15) is 28.75%. However, the classification error rates of Original collaborative representation [37] and Multi-resolution dictionary learning [14], RSLDA [38], and BDLRR [39] are 30.25%, 49.02%, 30.25%, 29.75%, respectively. Similarly, when the number of training samples per subject is 2, the proposed algorithm (16) has a classification error of 39.8%, which has higher classification accuracy than other algorithms, such as RSLDA [38] and BDLRR [39]. The above experiments show that the proposed algorithm can effectively improve the accuracy of image classification.

5 Conclusions

In order to improve the accuracy of image classification, especially image classification on deformable objects, such as human faces, this paper proposes an image classification algorithm. The experimental results show that our algorithm has better classification performance than other similar algorithms, such as Multi-resolution dictionary learning, RSLDA, BDLRR, L1LS, FISTA, PALM and other sparse representation algorithms. At the same time, the proposed algorithm has the advantages of high computational efficiency, simple implementation, and complete automation. In addition, the two new image representation procedures that we propose are complementary to the original image when representing the object. The original image and the virtual image are used to perform multiple representations on the same object, which makes our algorithm very versatile. The above experiments also prove the feasibility and effectiveness of the algorithm.

References

- Pishchulin L, Gass T, Dreuw P, Ney H (2012) Image warping for face recognition: from local optimality towards global optimization. *Pattern Recogn* 45(9):3131–3140
- Kautkar SN, Atkinson GA, Smith ML (2012) Face recognition in 2D and 2.5D using ridgelets and photometric stereo. *Pattern Recogn* 45(9):3317–3327
- Wang J, You J, Li Q, Xu Y (2012) Orthogonal discriminant vector for face recognition across pose. *Pattern Recogn* 45(12):4069–4079
- Patil HY, Kothari A, Bhurchandi KM (2015) Expression invariant face recognition using semidecimated DWT, Patch-LDSMT, feature and score level fusion. *J. Appl Intell* 127(5):1–18
- Jian M, Lam KM, Dong J (2011) Illumination compensation and enhancement for face recognition. In: *Proceedings of Asia-Pacific signal and information processing association annual summit conference (APSIPA ASC'2011)*, paper Wed-AM.RS6
- Sharma A, Dubey A, Tripathi P, Kumar V (2010) Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces. *Neurocomputing* 73(10–12):1868–1880
- Xu Y, Zhu X, Li Z, Liu G, Lu Y, Liu H (2013) Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition. *Pattern Recogn* 46(4):1151–1158
- Xu Y, Li X, Yang J, Lai Z, Zhang D (2014) Integrating conventional and inverse representation for face recognition. *IEEE Trans Cybern* 44:1738–1746
- Tang D, Zhu N, Yu F et al (2014) A novel sparse representation method based on virtual samples for face recognition. *Neural Comput Appl* 24:513–519
- Sanderson C, Paliwal KK (2003) Noise compensation in a person verification system using face and multiple speech feature. *Pattern Recogn* 36:293–302
- Peng WC, Bing SH, She ZJ et al (2017) Robust face recognition via discriminative and common hybrid dictionary learning. *J. Applied Intelligence* 2017(5500):1–10
- Xu Y, Li Z, Zhang B, Yang J, You J (2017) Sample diversity, representation effectiveness and robust dictionary learning for face recognition. *Inform Sci* 375:171–182
- Lin G, Yang M, Yang J, Shen L, Xie W (2018) Robust, discriminative and comprehensive dictionary learning for face recognition. *Pattern Recogn* 81:341–356
- Luo X, Xu Y, Yang J (2019) Multi-resolution dictionary learning for face recognition. *Pattern Recogn* 93:283–292
- Xu Y, Zhong Z, Yang J, You J, Zhang D (2017) A new discriminative sparse representation method for robust face recognition via $l(2)$ regularization. *IEEE Trans Neural Netw Learn Syst* 28(10):2233–2242
- Zhang H, Zhang Y, Huang TS (2013) Pose-robust face recognition via sparse representation. *Pattern Recogn* 46(5):1511–1521
- Yang J, Chu D, Zhang L, Xu Y, Yang J (2013) Sparse representation classifier steered discriminative projection with applications to face recognition. *IEEE Trans Neural Netw Learn Syst* 24(7):1023–1035

18. Gao S, Chia L-T, Tsang IW-H (2011) Multi-layer group sparse coding—For concurrent image classification and annotation, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp 2809–2816
19. Yang J, Zhang L, Xu Y, Yang J-Y (2012) Beyond sparsity: The role of L1-optimizer in pattern classification. *Pattern Recogn* 45(3):1104–1118
20. Yang J, Wright J, Huang T, Ma Y (2008) Image super-resolution as sparse representation of raw image patches. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp 1–8
21. Jing G, Shi Y, Kong D, Ding W, Yin B (2014) Image super-resolution based on multi-space sparse representation. *Multimedia Tools Appl* 70(2):741–755
22. Li H, He X, Tao D, Tang Y, Wang R (2018) Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recogn* 79:130–146
23. Wagner A, Wright J, Ganesh A, Zhou Z, Mobahi H, Ma Y (2012) Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Trans Pattern Anal Mach Intell* 34(2):372–386
24. Rubinstein R, Bruckstein AM, Elad M (2010) Dictionaries for sparse representation modeling. *Proc IEEE* 98(6):1045–1057
25. Yang M, Zhang L, Feng X, Zhang D (2011) Fisher discrimination dictionary learning for sparse representation. In: Proc. Int Conf. Comput. Vis. (ICCV), pp 543–550
26. Li Z, Lai Z, Xu Y, Yang J, Zhang D A locality-constrained and label embedding dictionary learning algorithm for image classification, *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. <https://doi.org/10.1109/TNNLS.2015.2508025>
27. Mazhar R, Gader PD (2008) EK-SVD: Optimized dictionary design for sparse representations. In: Proc. IEEE Conf. Pattern Recognit. (ICPR), pp 1–4
28. Aharon M, Elad M, Bruckstein A (2006) K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311–4322
29. Jiang Z, Lin Z, Davis LS (2013) Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Trans Pattern Anal Mach Intell* 35(11):2651–2664
30. Liu Z, Pu J, Huang T, Qiu Y (2013) A novel classification method for palmprint recognition based on reconstruction error and normalized distance. *Appl Intell* 39(2):307–314
31. Needell D, Vershynin R (2010) Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit. *IEEE J Sel Topics Signal Process* 4(2):310–316
32. Boyd SP (2008) https://web.stanford.edu/~boyd/l1_ls/, (accessed 04.08)
33. Yang AY, Zhou Z, Balasubramanian AG, Sastry SS, Ma Y (2013) Fast l_1 -minimization algorithms for robust face recognition. *IEEE Trans Image Process* 22(8):3234–3246
34. Gong P-H, Zhang C-S, Lu Z-S, Huang J-H, Ye J-P (2013) A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: Proceedings of international conference on machine learning, ICML, pp 37–45
35. Xu Y, Zhang D, Yang J, Yang J-Y (2011) A two-phase test sample sparse representation method for use with face recognition. *IEEE Trans Circuits Syst Video Technol* 21(9):1255–1262
36. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition?. In: Proceedings of IEEE international conference on computer vision, pp 471–478
37. Xu Y, Zhang B, Zhong Z (2015) Multiple representations and sparse representation for image classification. *Pattern Recogn Lett* 68:9–14. <https://doi.org/10.1016/j.patrec.2015.07.032>
38. Wen J, Fang X, Cui J, Fei L, Yan K, Chen Y, Xu Y (2019) Robust Sparse Linear Discriminant Analysis. *IEEE Trans Circuits Syst Video Technol* 29(2):390–403
39. Zhang Z, Xu Y, Shao L, Yang J (2018) Discriminative block-diagonal representational learning for image recognition. *IEEE Trans Neural Netw Learn Syst* 29(7):3111–3125
40. Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE workshop on applications of computer vision, IEEE Comput. Soc. Press, pp 138–142
41. Xu Y, Li XL, Yang J, Lai ZH, Zhang D (2014) Integrating conventional and inverse representation for face recognition. *IEEE Trans Cybern* 44(10):1738–1746
42. Fang XZ, Xu Y, Li XL, Lai ZH, Wong WK (2015) Learning a nonnegative sparse graph for linear regression. *IEEE Trans Image Process* 24(9):2760–2771
43. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intell* 22(10):1090–1104
44. Phillips PJ The facial recognition technology (FERET) database [Online]. Available: <https://www.nist.gov/programs-projects/face-recognition-technology-feret>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Shijun Zheng is currently pursuing the M.S. degree in software engineering at Guizhou University, Guiyang, China. His research interests include face recognition, object detection, and deep learning.



Yongjun Zhang received the Master and Doctoral degree in software engineering from Guizhou University, Guiyang, China, in 2010 and 2015, respectively. Now he is an associate professor in Guizhou University. From 2012 to 2015, He is a joint training doctoral student of Peking University and Guizhou University, and studying in the key laboratory of Integrated Microsystems of Shenzhen Graduate School of Peking University. His research inter-

ests mainly focus on the intelligence image algorithms of computer vision, such as scene target detection, extraction, tracking, recognition, and behavior analysis.



Wenjie Liu was born in Dingxi, Gansu province in China, in 1995. He received the B.S. degree in information and computing science in 2018. Then, he majored in software engineering in school of computer science and technology, Guizhou University, and his research includes machine learning, computer vision as well as remote sensing image processing based on deep learning.



Yongjie Zou was born in Qianjiang, Hubei province in China, in 1993. He received the B.S. degree in computer science and technology from Guizhou Education University in 2018. his research includes machine learning, computer vision as well as remote sensing image processing based on deep learning.