IET Computer Vision

Research Article

Improved sparse representation method for image classification

ISSN 1751-9632
Received on 14th June 2016
Revised 15th December 2016
Accepted on 23rd December 2016
E-First on 22nd March 2017
doi: 10.1049/iet-cvi.2016.0186
www.ietdl.org

Shigang Liu[1,2]*, Lingjun Li*[1,2]*, Yali Peng*[1,2] ✉*, Guoyong Qiu*[1,2]*, Tao Lei*[3]

[1]*Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, People's Republic of China*
[2]*School of Computer Science, Shaanxi Normal University, Xi'an 710119, People's Republic of China*
[3]*College of Electronics and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, People's Republic of China*
✉ *E-mail: 15239228561@163.com*

Abstract: Among all image representation and classification methods, sparse representation has proven to be an extremely powerful tool. However, a limited number of training samples are an unavoidable problem for sparse representation methods. Many efforts have been devoted to improve the performance of sparse representation methods. In this study, the authors proposed a novel framework to improve the classification accuracy of sparse representation methods. They first introduced the concept of the approximations of all training samples (i.e., virtual training samples). The advantage of this is that the application of virtual training samples can allow noise in original training samples to be partially reduced. Then they proposed an efficient and competent objective function to disclose more discriminant information between different classes, which is very significant for obtaining a better classification result. The devised sparse representation method employs both the original and virtual training samples to improve the classification accuracy since the two kinds of training samples makes sample information to be fully exploited in a good way, also satisfactory robustness to be obtained. The experimental results on the JAFFE, ORL, Columbia Object Image Library (COIL-100) AR and CMU PIE databases show that the proposed method outperforms the state-of-art image classification methods.

1 Introduction

Image classification [1, 2], as a computer vision technology [3–7], has been developing rapidly. Meanwhile, it has also attracted considerable attention in recent years [8]. Achieving good image classification result is the basis for many socioeconomic and environmental applications. Therefore, researchers and scientists have made great efforts in devising advanced classification methods and technologies to improve classification accuracy [9–12].

Among all methods of image representation and classification, sparse representation has proven to be an extremely powerful tool [13–15]. The basic model of sparse representation indicates that the test sample can be represented approximately by using a weighted sum of the training samples. It's pointed out that the weight coefficients are sparse. In other words, the most coefficients are zero or close to zero in the weight coefficients vector. In general, sparse representation relates to an underdetermined system of linear equation $y = \boldsymbol{D}\alpha$, where y denotes the test sample, $\boldsymbol{D}$ is a dictionary (or training samples matrix) and α is a coefficient vector. Moreover, many variations and extensions of sparse representation have been proposed in the past few years [16–20]. Wright *et al.* [21] presented a sparse representation-based classification (SRC) method, which exploits the discriminative nature of sparse representation to perform classification. Zhang *et al.* [22] proposed an efficient image classification scheme, namely collaborative representation (CR)-based classification with regularised least square (CRC_RLS). Naseem *et al.* [23] gave a robust linear regression classification algorithm. He *et al.* [24] proposed a new multiple linear regression model which can regularise correntropy to enhance the robustness of pattern recognition. Xu *et al.* [25] proposed a novel transfer subspace learning method which integrates the methods of changing data's representation and classifier design. Deng *et al.* [26] extended SRC into commercial applications, these applications are designed primarily for the case of a single training sample of each class. Moreover, considering the non-linear relationship of samples and usage of different features with non-linear metrics, Wang *et al.* [27] proposed a kernel CR scheme for linear and non-linear representation-based approaches. Motivated by the fact the different features of a sample contribute to the object representation and classification differently, Wang *et al.* [28] proposed a novel relaxed CR (RCR) frame to exploit the similarity and distinctiveness of features effectively. A low-rank sparse coding method is proposed by Zhang *et al.* [29], which exploits local structure information among features of an image to achieve image classification. Yang *et al.* [30] proposed an extension of the spatial pyramid matching method. This method calculates a spatial-pyramid image representation based on Scale-invariant feature transform sparse codes. It can reduce the complexity of SVMs.

Intuitively, the sparsity of coefficient vector $\boldsymbol{\alpha}$ can be measured by the l_0-norm [31], which counts the number of non-zeros in vector $\boldsymbol{\alpha}$. However, the l_0-norm minimisation is a NP-hard problem [32–34]. To solve this problem, the l_1-norm minimisation, as a good convex approximation of l_0-norm minimisation, is widely employed in sparse coding [35, 36]. Even so, as reviewed in [22], the l_1-norm minimisation is still time consuming, hence many researchers strive to seek fast algorithms to solve the problem of time consumption. Yang *et al.* [37] summarised five representative fast l_1-norm minimisation methods, i.e. homotopy, proximal gradient, gradient projection, iterative shrinkage-thresholding and augmented Lagrange multiplier [38]. The researchers found that l_2-norm based representation method [39] can avoid over-fitting and improve the generalisation ability of the classification model. The closed-form solution can be derived by l_2-norm regularisation. In addition, l_1-norm tends to produce a small number of non-zero weight coefficients, while l_2-norm usually can obtain more non-zero weight coefficients. Therefore the l_2-norm minimisation can get a ‘limitedly-sparse’ representation solution. The solution has a property which is discriminative and distinguishable but not really sparse enough [40]. Nevertheless, we cannot deny the fact that the discriminative capability of the solution is helpful for image classification. Moreover, the authors of [41, 42] demonstrated that the influence of the sparsity on image classification is not strong by conducting lots of experiments. Zhang *et al.* [22] confirmed that the sparsity based on l_1-norm minimisation could not really make critical differences in classification, and so proposed a new method based on l_2-norm minimisation, namely collaborative representation classification method. It's classification accuracy is higher than SRC. Similarly, Shi *et al.* [42] proposed to use the simple l_2-norm minimisation to achieve more effective classification, and pointed out that the sparse approximation could not satisfy the needs of robustness and required performance. Accordingly, more methods based on l_2-norm minimisation have been proposed. Xu *et al.* [43] proposed a two-phase test sample representation method for face recognition. This method is an l_2 regularisation-based representation method. Liu *et al.* [44] proposed a novel representation-based palmprint recognition method, which also belongs to l_2-norm based method and classifies the test sample according to an approximate representation of the test sample. Xu *et al.* [45] proposed to use the classification procedure of sparse representation to improve the nearest neighbour classifier.

However, a sufficient number of available training samples are the basis for all the above mentioned methods [46–48]. Hence, various methods have been proposed for solving the problem of non-sufficient training samples in recent years. For example, a scheme was proposed in [49], which exploits the symmetry of the face to generate new samples, and devises a representation-based method to perform face recognition. Ryu *et al.* [50] proposed a method that adds virtual training samples into the training sample set. These virtual training samples are generated adaptively on the basis of the spatial distribution of each class's training samples. Xu *et al.* [51] proposed a novel method to improve the face recognition accuracy by synthesising virtual training samples. Wang and Yang [52] used the idea of perturbation to produce virtual training samples. A novel representation-based classification method, which integrates conventional and the inverse representation-based classification into face recognition, was also proposed by Xu *et al.* [53].

This paper aims to improve the classification accuracy of the sparse representation method in image classification. We proposed a novel method for image classification. Firstly, inspired by the prior work on virtual training samples, we introduced the concept of the approximations of all training samples, (i.e., virtual training samples [54]), which can expand the training sample set and solve the problem of non-sufficient training samples. Then we proposed an efficient and competent objective function to disclose more discriminant information between different classes, which is crucial and significant for obtaining good classification results. The devised sparse representation method simultaneously used the original and virtual training samples to perform classification. The use of virtual training samples allows noise in original training samples to be reduced partially and satisfactory robustness to be obtained. Simultaneously, the use of the original and virtual training samples makes sample information to be exploited in a more comprehensive way. In this method, we take advantages of the score level fusion, which has proven to be very competent and is usually better than the decision level and feature level fusion. Superficially, it is hard to set initialisation for the approximation of all training samples mentioned above. However, we can determine that the number of virtual training samples and original training samples are the same. When designing the objective function, we assumed that the initialisation of approximation of all training samples is the same as the training sample matrix. Then after the optimal coefficient vector is solved, this approximation will be updated. We will show the solution procedure in detail in Section 2. The experimental results show that the proposed method outperforms the state-of-art image classification methods. The main contributions of our work are as follows: (i) It proposed a simple and reasonable way to enhance the distinctiveness of different classes, besides, this method is compatible with the nature of sparse representation. (ii) The proposed method can obtain very accurate classification results by integrating the original and virtual training samples properly.

The remaining parts of this paper are organised as follows. Section 2 presents the description of the proposed method in detail. Section 3 analyses the advantages and rationale of the proposed method. Sections 4 and 5 offer the experimental results and conclusion, respectively.

2 Description of the proposed method

Let c denote the number of classes, each class provides n training samples. Then N ($N = c * n$) stands for the total number of training samples. Furthermore, let matrix $\boldsymbol{X}_i = \left[x_{n(i-1)+1}, \ldots, x_{ni}\right]$ $(0 < i \le c)$ denote the samples of the ith class, where $x_{n(i-1)+s}$ $(0 < s \le n)$, a column vector, stands for the sth training sample in the ith class. We define the training sample matrix as $\boldsymbol{X} = \left[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_c\right] = \left[x_1, \ldots, x_{n(m-1)+1}, \ldots, x_{nm}, \ldots, x_N\right]$. Let column vector$\boldsymbol{y}$ stand for a test sample. In addition, $x_1, \ldots, x_N$ and $\boldsymbol{y}$ are all D-dimensional column vectors. Hence $\boldsymbol{X}$ is a $D \times N$ matrix.

We define the objective function as

$$\min_{\beta, \boldsymbol{Z}} \lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j + \| y - \boldsymbol{Z}\beta \|_2^2 \quad (1)$$

where $\boldsymbol{Z}$ is the approximation of all the training samples, namely virtual training samples. λ_1 and λ_2, two small positive constants, are used to balance the effect of the three terms in the proposed objective function. β denotes a coefficient vector, i.e. $\beta_i = [b_{n(i-1)+1}, b_{n(i-1)+2}, \ldots, b_{ni}]^{\mathrm{T}}$, $\beta = [\beta_1, \ldots, \beta_c]^{\mathrm{T}} = [b_1, \ldots, b_{n(i-1)+1}, \ldots, b_{ni}, \ldots, b_N]^{\mathrm{T}}$. It should be noted that $\lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \mathbf{Z}_i^{\mathrm{T}} \mathbf{Z}_j \beta_j$ is a constraint on β. Meanwhile, it can be easily verified that the objective function in formula (1) is convex and differentiable. We will prove that this objective function is convex in subsequent Section 3.2. Hence the stationary point of the objective function is the optimal solution of (1). The derivative of the objective function is computed as follows.

Firstly, we assume that β is known, and derive the partial derivative of the objective function w.r.t $\mathbf{Z}$, i.e.

$$\frac{\partial}{\partial \mathbf{Z}} \left(\lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \mathbf{Z}_i^{\mathrm{T}} \mathbf{Z}_j \beta_j + \| y - \boldsymbol{Z}\beta \|_2^2 \right).$$

Let

$$H(\beta, \mathbf{Z}) = \lambda_1 \sum_{i=1}^{c} \| \mathbf{X}_i - \mathbf{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \mathbf{Z}_i^{\mathrm{T}} \mathbf{Z}_j \beta_j + \| y - \mathbf{Z}\beta \|_2^2,$$

then

$$\frac{\partial}{\partial \mathbf{Z}} \left(\| y - \mathbf{Z}\beta \|_2^2 \right) = -2(y - \mathbf{Z}\beta)\beta^{\mathrm{T}}. \quad (2)$$

Next, assuming

$$h_1(\mathbf{Z}) = \lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \mathbf{Z}_i^{\mathrm{T}} \mathbf{Z}_j \beta_j,$$

we take the derivative of function $h_1(\mathbf{Z})$ w.r.t $\mathbf{Z}$. However, function $h_1(\mathbf{Z})$ does not explicitly contain $\mathbf{Z}$. Hence we firstly seek partial derivatives $\partial h_1 / \partial \mathbf{Z}_k$ $(k = 1, \ldots, c)$, then we obtain derivative $\partial h_1 / \partial \mathbf{Z}$ according to $\partial h_1 / \partial \mathbf{Z}_k$. In addition

$$\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j = (\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j = \frac{1}{2}\left(\| \mathbf{Z}_i\beta_i + \mathbf{Z}_j\beta_j \|_2^2 - \| \mathbf{Z}_i\beta_i \|_2^2 - \| \mathbf{Z}_j\beta_j \|_2^2\right). \tag{3}$$

Based on the above analyses, $h_1(\mathbf{Z})$ is redefined as (see (4) and (5)) Thus, the partial derivative over $\mathbf{Z}_k$ of $h_1(\mathbf{Z})$ is (see (5) and (6)) then (see (6)) Let

$$\boldsymbol{M}_1 = \begin{pmatrix} \beta_1\beta_1^{\mathrm{T}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_c\beta_c^{\mathrm{T}} \end{pmatrix},$$

then

$$\frac{\partial h_1}{\partial \boldsymbol{Z}} = -2\lambda_1(\boldsymbol{X} - \boldsymbol{Z}) + 2\lambda_2\boldsymbol{Z}\beta\beta^{\mathrm{T}} - 2\lambda_2\boldsymbol{Z}\boldsymbol{M}_1. \tag{7}$$

By combining (2) and (7), the derivative over $\boldsymbol{Z}$ of function $H(\beta, \boldsymbol{Z})$ is

$$\frac{\partial H}{\partial \boldsymbol{Z}} = -2\lambda_1\boldsymbol{X} + 2\lambda_1\boldsymbol{Z} + 2\lambda_2\boldsymbol{Z}\beta\beta^{\mathrm{T}} - 2\lambda_2\boldsymbol{Z}\boldsymbol{M}_1 - 2y\beta^{\mathrm{T}} + 2\boldsymbol{Z}\beta\beta^{\mathrm{T}}. \tag{8}$$

To obtain the optimal solution of the objective function, we let $(\partial H/\partial \boldsymbol{Z}) = 0$, i.e. $\boldsymbol{Z}\left(\lambda_1\boldsymbol{I} + \lambda_2\beta\beta^{\mathrm{T}} - \lambda_2\boldsymbol{M}_1 + \beta\beta^{\mathrm{T}}\right) = \lambda_1\boldsymbol{X} + y\beta^{\mathrm{T}}$. Hence, under the condition that variation β is known, the optimal value of variation $\boldsymbol{Z}$ is

$$\hat{\boldsymbol{Z}} = \left(\lambda_1\boldsymbol{X} + y\beta^{\mathrm{T}}\right)\left(\lambda_1\boldsymbol{I} + \lambda_2\beta\beta^{\mathrm{T}} - \lambda_2\boldsymbol{M}_1 + \beta\beta^{\mathrm{T}}\right)^{-1}. \tag{9}$$

Secondly, we assume that variation $\boldsymbol{Z}$ is known, and derive the partial derivative of the objective function w.r.t β, i.e.

$$\frac{\partial}{\partial\beta}\left(\lambda_1\sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{Z}_j\beta_j + \| y - \boldsymbol{Z}\beta \|_2^2\right)$$

.

$$\begin{aligned} h_1(\boldsymbol{Z}) &= \lambda_1\sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j \\ &= \lambda_1\left(\left(\sum_{\substack{i=1,\dots,c\\ i\neq k}} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2\right) + \| \boldsymbol{X}_k - \boldsymbol{Z}_k \|_2^2\right) + \frac{\lambda_2}{2}\left[\sum_{\substack{i=1,\dots,c\\ i\neq k}}\left(\| \boldsymbol{Z}_i\beta_i + \boldsymbol{Z}_k\beta_k \|_2^2 - \| \boldsymbol{Z}_i\beta_i \|_2^2 - \| \boldsymbol{Z}_k\beta_k \|_2^2\right)\right. \\ &\quad \left. + \sum_{\substack{j=1,\dots,c\\ j\neq k}}\left(\| \boldsymbol{Z}_k\beta_k + \boldsymbol{Z}_j\beta_j \|_2^2 - \| \boldsymbol{Z}_k\beta_k \|_2^2 - \| \boldsymbol{Z}_j\beta_j \|_2^2\right) + \sum_{\substack{i=1,\dots,c\\ i\neq k}}\sum_{\substack{j=1,\dots,c\\ j\neq k}}\left(\| \boldsymbol{Z}_i\beta_i + \boldsymbol{Z}_j\beta_j \|_2^2 - \| \boldsymbol{Z}_i\beta_i \|_2^2 - \| \boldsymbol{Z}_j\beta_j \|_2^2\right)\right] \\ &= \lambda_1\left(\left(\sum_{\substack{i=1,\dots,c\\ i\neq k}} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2\right) + \| \boldsymbol{X}_k - \boldsymbol{Z}_k \|_2^2\right) + \lambda_2\sum_{\substack{i=1,\dots,c\\ i\neq k}}\left(\| \boldsymbol{Z}_i\beta_i + \boldsymbol{Z}_k\beta_k \|_2^2 - \| \boldsymbol{Z}_i\beta_i \|_2^2 - \| \boldsymbol{Z}_k\beta_k \|_2^2\right) \\ &\quad + \frac{\lambda_2}{2}\sum_{\substack{i=1,\dots,c\\ i\neq k}}\sum_{\substack{j=1,\dots,c\\ j\neq k}}\left(\| \boldsymbol{Z}_i\beta_i + \boldsymbol{Z}_j\beta_j \|_2^2 - \| \boldsymbol{Z}_i\beta_i \|_2^2 - \| \boldsymbol{Z}_j\beta_j \|_2^2\right). \end{aligned} \tag{4}$$

$$\begin{aligned} \frac{\partial h_1}{\partial \mathbf{Z}_k} &= \frac{\partial}{\partial \mathbf{Z}_k}\left(\lambda_1\sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j\right) \\ &= \frac{\partial}{\partial \mathbf{Z}_k}\left(\lambda_1\| \boldsymbol{X}_k - \boldsymbol{Z}_k \|_2^2 + \lambda_2\sum_{\substack{i=1,\dots,c\\ i\neq k}}\left(\| \mathbf{Z}_i\beta_i + \mathbf{Z}_k\beta_k \|_2^2 - \| \mathbf{Z}_i\beta_i \|_2^2 - \| \mathbf{Z}_k\beta_k \|_2^2\right)\right) \\ &= -2\lambda_1(\boldsymbol{X}_k - \mathbf{Z}_k) + \lambda_2\sum_{\substack{i=1,\dots,c\\ i\neq k}}\left(2(\mathbf{Z}_i\beta_i + \mathbf{Z}_k\beta_k)\beta_k^{\mathrm{T}} - 2\mathbf{Z}_k\beta_k\beta_k^{\mathrm{T}}\right) \\ &= -2\lambda_1(\boldsymbol{X}_k - \mathbf{Z}_k) + \lambda_2\sum_{\substack{i=1,\dots,c\\ i\neq k}}\left(2\mathbf{Z}_i\beta_i\beta_k^{\mathrm{T}}\right) \\ &= -2\lambda_1(\boldsymbol{X}_k - \mathbf{Z}_k) + 2\lambda_2\left[\left(\sum_{i=1,\dots,c}\mathbf{Z}_i\beta_i\beta_k^{\mathrm{T}}\right) - \mathbf{Z}_k\beta_k\beta_k^{\mathrm{T}}\right] \\ &= -2\lambda_1(\boldsymbol{X}_k - \mathbf{Z}_k) + 2\lambda_2\boldsymbol{Z}\beta\beta_k^{\mathrm{T}} - 2\lambda_2\mathbf{Z}_k\beta_k\beta_k^{\mathrm{T}}, \end{aligned} \tag{5}$$

$$\begin{aligned} \frac{\partial h_1}{\partial \boldsymbol{Z}} &= \left[\frac{\partial h_1}{\partial \boldsymbol{Z}_1}, \dots, \frac{\partial h_1}{\partial \boldsymbol{Z}_c}\right] \\ &= \left[-2\lambda_1(\boldsymbol{X}_1 - \mathbf{Z}_1) + 2\lambda_2\boldsymbol{Z}\beta\beta_1^{\mathrm{T}} - 2\lambda_2\mathbf{Z}_1\beta_1\beta_1^{\mathrm{T}}, \dots, -2\lambda_1(\boldsymbol{X}_c - \mathbf{Z}_c) + 2\lambda_2\boldsymbol{Z}\beta\beta_c^{\mathrm{T}} - 2\lambda_2\mathbf{Z}_c\beta_c\beta_c^{\mathrm{T}}\right] \\ &= -2\lambda_1(\boldsymbol{X} - \mathbf{Z}) + 2\lambda_2\boldsymbol{Z}\beta\beta^{\mathrm{T}} - 2\lambda_2\boldsymbol{Z}\begin{pmatrix} \beta_1\beta_1^{\mathrm{T}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_c\beta_c^{\mathrm{T}} \end{pmatrix}. \end{aligned} \tag{6}$$

We can obtain

$$\frac{\partial}{\partial \beta}\left(\| y - \boldsymbol{Z}\beta \|_2^2\right) = -2\boldsymbol{Z}^{\mathrm{T}}(y - \boldsymbol{Z}\beta). \qquad (10)$$

Let $h_2(\beta) = \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j$. Similarly, $h_2(\beta)$ dose not explicitly contain β, so we must seek partial derivatives $\partial h_2 / \partial \beta_k$ $(k = 1, \ldots, c)$. Then we get derivative $\partial h_2 / \partial \beta$ according to all $\partial h_2 / \partial \beta_k$.

Based on (3), $h_2(\beta)$ is redefined as (see (11) and (12)) Next, the partial derivative over β_k of $h_2(\beta)$ is (see (12)) Then, derivative $\partial h_2 / \partial \beta$ is

$$\begin{aligned} \frac{\partial h_2}{\partial \beta} &= \begin{bmatrix} \dfrac{\partial h_2}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial h_2}{\partial \beta_c} \end{bmatrix} = \begin{bmatrix} 2\lambda_2 \boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{Z}_1 \beta_1 \\ \vdots \\ 2\lambda_2 \boldsymbol{Z}_c^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{Z}_c^{\mathrm{T}} \boldsymbol{Z}_c \beta_c \end{bmatrix} \\ &= 2\lambda_2 \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \begin{pmatrix} \boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{Z}_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_c^{\mathrm{T}} \boldsymbol{Z}_c \end{pmatrix} \beta. \end{aligned} \qquad (13)$$

Let

$$\boldsymbol{M}_2 = \begin{pmatrix} \boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{Z}_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_c^{\mathrm{T}} \boldsymbol{Z}_c \end{pmatrix},$$

thus

$$\frac{\partial h_2}{\partial \beta} = 2\lambda_2 \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{M}_2 \beta. \qquad (14)$$

According to (10) and (14), we obtain the following derivative over β of function $H(\beta, \boldsymbol{Z})$:

$$\frac{\partial H}{\partial \beta} = 2\lambda_2 \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{M}_2 \beta - 2\boldsymbol{Z}^{\mathrm{T}}(y - \boldsymbol{Z}\beta). \qquad (15)$$

Let $(\partial H / \partial \beta) = 0$, i.e. $\left(\lambda_2 \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{Z} - \lambda_2 \boldsymbol{M}_2 + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{Z}\right)\beta = \boldsymbol{Z}^{\mathrm{T}} y$. Hence, under the condition that variation $\boldsymbol{Z}$ is known, the optimal value of variation β is

$$\hat{\beta} = \left(\lambda_2 Z^{\mathrm{T}} Z - \lambda_2 M_2 + Z^{\mathrm{T}} Z\right)^{-1} Z^{\mathrm{T}} y. \qquad (16)$$

To eventually determine the optimal solutions of $\boldsymbol{Z}$ and β, the training samples matrix $\boldsymbol{X}$ is considered as the initial value of $\boldsymbol{Z}$, then the initial value of β is obtained in terms of (16). Likewise, the latest value of $\boldsymbol{Z}$ can be calculated by exploiting the initial value of β and (9). These processes are implemented iteratively, until the results meet the final qualification.

In addition, our method employs the original training samples to represent the test sample. According to the research of the reference literature, the objective function based on the original training samples can be written as

$$\min_{\beta} \lambda_3 \sum_{i=1}^{c} \sum_{j=1}^{c} \| \boldsymbol{X}_i \beta_i + \boldsymbol{X}_j \beta_j \|_2^2 + \| y - \boldsymbol{X}\beta \|_2^2. \qquad (17)$$

Hence, we can obtain the optimal solution $\hat{\beta}_{\text{original}}$ of the original training samples

$$\hat{\beta}_{\text{original}} = \left((1 + 2\lambda_3)\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + 2\lambda_3 c \boldsymbol{M}\right)^{-1} \boldsymbol{X}^{\mathrm{T}} y. \qquad (18)$$

$$\begin{aligned} h_2(\beta) &= \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j \\ &= \frac{\lambda_2}{2} \left[\sum_{\substack{i=1,\ldots,c \\ i \neq k}} \left(\| \boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_k \beta_k \|_2^2 - \| \boldsymbol{Z}_i \beta_i \|_2^2 - \| \boldsymbol{Z}_k \beta_k \|_2^2\right) \right. \\ &\quad + \sum_{\substack{j=1,\ldots,c \\ j \neq k}} \left(\| \boldsymbol{Z}_k \beta_k + \boldsymbol{Z}_j \beta_j \|_2^2 - \| \boldsymbol{Z}_k \beta_k \|_2^2 - \| \boldsymbol{Z}_j \beta_j \|_2^2\right) \\ &\quad \left. + \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \sum_{\substack{j=1,\ldots,c \\ j \neq k}} \left(\| \boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_j \beta_j \|_2^2 - \| \boldsymbol{Z}_i \beta_i \|_2^2 - \| \boldsymbol{Z}_j \beta_j \|_2^2\right) \right] \\ &= \lambda_2 \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \left(\| \boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_k \beta_k \|_2^2 - \| \boldsymbol{Z}_i \beta_i \|_2^2 - \| \boldsymbol{Z}_k \beta_k \|_2^2\right) \\ &\quad + \frac{\lambda_2}{2} \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \sum_{\substack{j=1,\ldots,c \\ j \neq k}} \left(\| \boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_j \beta_j \|_2^2 - \| \boldsymbol{Z}_i \beta_i \|_2^2 - \| \boldsymbol{Z}_j \beta_j \|_2^2\right). \end{aligned} \qquad (11)$$

$$\begin{aligned} \frac{\partial h_2}{\partial \beta_k} &= \frac{\partial}{\partial \beta_k} \left(\lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j \right) \\ &= \frac{\partial}{\partial \beta_k} \left(\lambda_2 \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \left(\| \boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_k \beta_k \|_2^2 - \| \boldsymbol{Z}_i \beta_i \|_2^2 - \| \boldsymbol{Z}_k \beta_k \|_2^2\right) \right) \\ &= \lambda_2 \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \left(2\boldsymbol{Z}_k^{\mathrm{T}}(\boldsymbol{Z}_i \beta_i + \boldsymbol{Z}_k \beta_k) - 2\boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}_k \beta_k\right) = \lambda_2 \sum_{\substack{i=1,\ldots,c \\ i \neq k}} \left(2\boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}_i \beta_i\right) \\ &= 2\lambda_2 \left[\left(\sum_{i=1,\ldots,c} \boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}_i \beta_i \right) - \boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}_k \beta_k \right] = 2\lambda_2 \boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{Z}_k^{\mathrm{T}} \boldsymbol{Z}_k \beta_k. \end{aligned} \qquad (12)$$

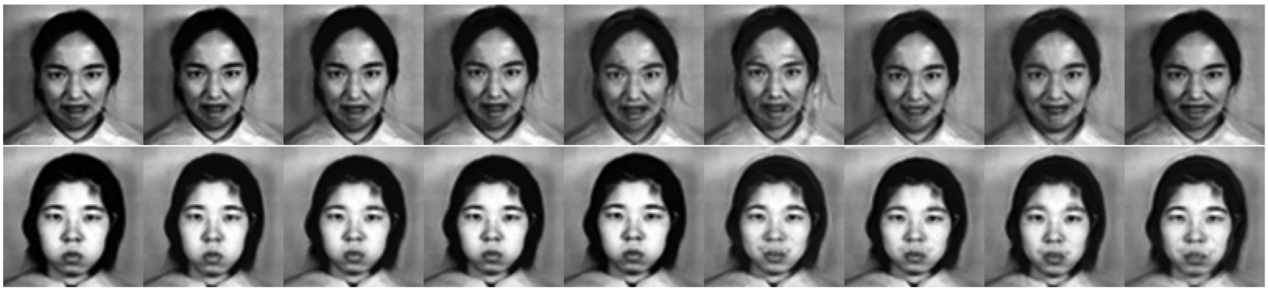


Fig. 1 *Some original training samples in the JAFFE face database and their corresponding virtual training samples*

We summarise the main steps of the proposed method as follows:

Step 1: We set the initial value of $\boldsymbol{Z}$ as training sample matrix $\boldsymbol{X}$. Let $\boldsymbol{Z}^0$ denote the initial value of $\boldsymbol{Z}$. According to (16), we get the initial value of β, and let β^0 stand for the initial value of β.

Step 2: We update $\boldsymbol{Z}^0$ by exploiting β^0 and (9), and let $\boldsymbol{Z}^1$ denote the latest value of $\boldsymbol{Z}$. Then we update β^0 by using $\boldsymbol{Z}^1$ and (16). β^1 stands for the latest value of β.

Step 3: Repeat step 2. The iterative updating is not terminated until one of the following two conditions is satisfied. (a) The number of iterations is greater than the predefined maximum value. (b) $\| \boldsymbol{Z}^{t+1} - \boldsymbol{Z}^t \| < \varepsilon$ and $\| \beta^{t+1} - \beta^t \| < \varepsilon$. $\boldsymbol{Z}^t$ and β^t denote the value of $\boldsymbol{Z}$ and β at time t, respectively. ε stands for a small positive constant.

Step 4: After obtaining the optimal $\boldsymbol{Z}^t$ and β^t, we use original sample matrix $\boldsymbol{X}$ to perform image classification. The distance between the test sample and the ith class is obtained using the class-specific residual, i.e. $u_i = \| \boldsymbol{X}_i\beta_i - y \|_2^2$ $(i = 1, \ldots, c)$. Let s_i^1 denote the score of test sample y with respect to the ith class.

Step 5: We use virtual training samples $\boldsymbol{Z}$ to perform image classification. The distance between the test sample and the ith class is obtained by using the class-specific residual, i.e. $v_i = \| \boldsymbol{Z}_i\beta_i - y \|_2^2$ $(i = 1, \ldots, c)$. Let s_i^2 denote the score of test sample y with respect to the ith class.

Step 6: The weighted score level fusion is conducted by combining the scores obtained in the fourth and fifth steps. For the test sample y, we use $s_i = w_1 s_i^1 + w_2 s_i^2$ to calculate the ultimate score with respect to the ith class. w_1 and w_2 are the weights. Let $w_1 + w_2 = 1$ and w_2 be smaller than w_1. If $j = \arg\min_i s_i$, then test sample y is assigned to the jth class.

3 Analysis of the proposed method

This section analyses the advantages and rationalities of the proposed method and gives meaningful conclusions.

3.1 Advantages of the proposed method

The first advantage is that the proposed method put forward a novel way to generate virtual samples. That is, the proposed method of generating the virtual training samples is based neither upon the symmetrical structure of the face, nor upon the mirror image, but upon the objective function. Concretely, we first assume that the approximate values of the original training samples are in existence. Then we construct an objective function based on the approximate values of the original training samples. We can obtain the virtual training samples in the process of solving the optimal value of the objective function. Moreover, the virtual training samples obtained by our method reflect possible variation of illuminations and facial expressions. According to the first term of the objective function (i.e., $\lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2$), virtual training samples matrix $\boldsymbol{Z}$ is the approximate value of all original training samples matrix $\boldsymbol{X}$, by definition, the number of virtual training samples is equal to that of original training samples. In addition, we can observe that test sample y is involved in the generation of virtual training samples from (9) (i.e., $\hat{\boldsymbol{Z}} = (\lambda_1 \boldsymbol{X} + y\beta^{\mathrm{T}})(\lambda_1 \boldsymbol{I} + \lambda_2 \beta\beta^{\mathrm{T}} - \lambda_2 \boldsymbol{M}_1 + \beta\beta^{\mathrm{T}})^{-1})$, which is helpful for representation methods to better represent and recognise the test sample.

The first four face images of each subject in the JAFFE face database are used as training samples, the remaining images are considered as test samples. The first column in Fig. 1 shows the original training samples, the subsequent columns show the corresponding virtual training samples which are produced by combining test samples from the same class. Similarly, the first two face images of each subject in the ORL face database are considered as training samples, the remaining images are taken as test samples in the ORL face database. In Fig. 2*a*, the first column shows the original training samples, the subsequent columns show the corresponding virtual training samples which are produced by combining test samples from the same class. On the contrary, in Fig. 2*b*, the first column shows the original training samples, but the subsequent columns show the corresponding virtual training samples which are produced by combining test samples from different classes. It is clear that test samples have an effect on only the training samples which are from the same class, while test samples from different classes almost have no influence on training samples. In other words, the virtual training samples generated by exploiting the same class of the test sample would be more close to the test sample. Experimental results presented later also show that the proposed method could achieve more accurate classification.

The second advantage is that our method proposed an effective way to enhance the distinctiveness of different classes. This advantage will be amply explained in the second paragraph of Section 3.2. We explain some advantages of the l_2-norm minimisation as follows. The proposed l_2 regularisation-based method has satisfactory performance. As shown later, the proposed method obtains more higher classification accuracy than CRC in some databases, which is a typical example of l_2 regularisation-based representation method. Furthermore, the proposed method also illustrates that collaboration plays important roles in sparsity representation methods. It is beneficial to decrease correlation of the approximate representation of the test sample generated from different classes. Moreover, our method can also enhance the distinctiveness of different classes, which is helpful for representation methods to obtain discriminative class-special residuals and to achieve excellent classification accuracy.

As mentioned before, the l_2-norm minimisation can obtain a ‘limitedly-sparse’ representation solution, but the solution of the proposed method is discriminative. Representation coefficients obtained using the proposed method, CRC and 1/-Regularized Least Squares [55] methods which are a regularisation-based representation methods are shown in Fig. 3. It is intuitive that the representation coefficients of training samples from the same class as the test sample have distinct differences from other coefficients. Hence, we can directly determine the labels of the test sample according to the representation coefficient distribution shown in Fig. 3. Moreover, we can also see that representation coefficients of these three methods have somewhat similar distributions.

3.2 Rationalities of the proposed method

In general, the previous representation-based classification methods can work under the premise that the test sample is represented by training samples. However, the dimension of the

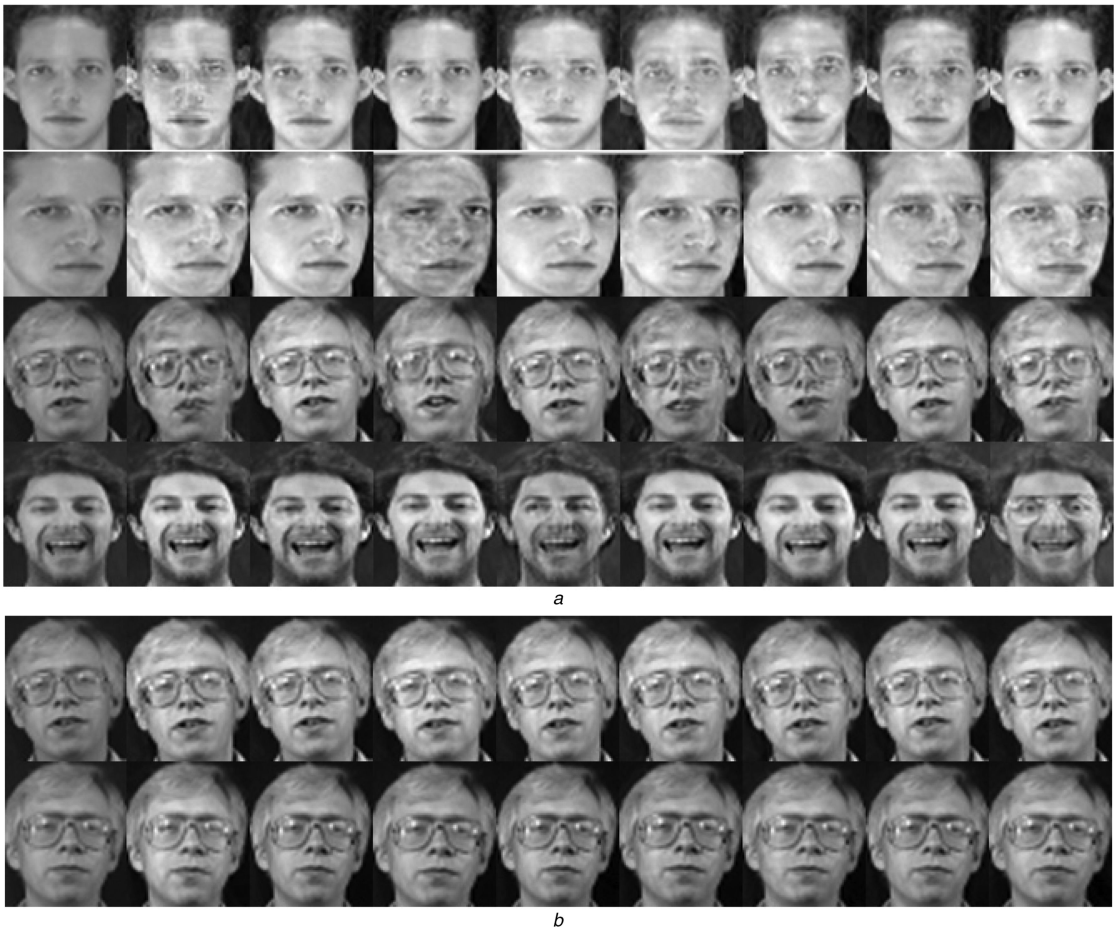


Fig. 2 *Some original training samples in the ORL face database and their corresponding virtual training samples*
(a) the first column shows the original training samples, and the subsequent columns show the corresponding virtual training samples which are produced by combining test samples from the same class, ***(b)*** the first column shows the original training samples, and the subsequent columns show the corresponding virtual training samples which are produced by combining test samples from different classes

sample vector is always larger than the number of the training samples in image classification. Meanwhile, affected by the variation of illumination conditions, facial expressions and poses, the linear combination of training samples is just an approximation value of the test sample and it cannot accurately represent the test sample. These factors are called small size sample (SSS) problem. Thus it is the key to solve the SSS problem and to improve the precision of the representation of the test sample by using the training samples. The proposed method designs an objective function based on virtual samples, meanwhile virtual samples can be obtained by solving the optimal value of the objective function. The test sample is involved in the process of generating virtual samples. In other words, the generated virtual samples are similar to the test sample in some extent, which is beneficial to the improvement of the accuracy of image classification. For example, we conduct an experiment on the JAFFE face database by taking the first four images of each subject as training samples and the rest of images as test images. Fig. 4 shows a case that our method outperforms both the CRC and L1LS methods. According to the experimental results, we concluded that the test sample from the fifth class shown in Fig. 4*a* is erroneously classified to the first class in Fig. 4*b* by CRC and L1LS, respectively. However, our method can correctly classify the test sample to the fifth class in Fig. 4*c*.

In Second, we designed a subjective function

$$\min_{\beta,\boldsymbol{Z}} \lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j + \| y - \boldsymbol{Z}\beta \|_2^2 .$$

Next, we will prove that it is a convex function. In [56], a theorem and an example are given as follows.

Theorem 1: Assume that f is twice differentiable, i.e. its Hessian or second derivative $\nabla^2 f$ exists at each point in $\boldsymbol{dom} f$, which is open. Then f is convex if and only if $\boldsymbol{dom} f$ is convex and its Hessian is positive semidefinite: for all $x \in \boldsymbol{dom} f$, $\nabla^2 f(x) \succeq 0$.

Example 1: Consider the quadratic function $f: \boldsymbol{R}^n \to \boldsymbol{R}$, with $\boldsymbol{dom} f = \boldsymbol{R}^n$, given by

$$f(x) = (1/2) x^{\mathrm{T}} \boldsymbol{P} x + q^{\mathrm{T}} x + r,$$

with $\boldsymbol{P}$ is the set of symmetric $n \times n$ matrix, $q \in \boldsymbol{R}^n$, and $r \in \boldsymbol{R}$. Since $\nabla^2 f(x) = \boldsymbol{P}$ for all x, f is convex if and only if $\boldsymbol{P} \succeq 0$ (and concave if and only if $\boldsymbol{P} \preceq 0$).

Let

$$H(\beta, \boldsymbol{Z}) = \lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2 + \lambda_2 \sum_{i=1}^{c} \sum_{j=1}^{c} \beta_i^{\mathrm{T}} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{Z}_j \beta_j + \| y - \boldsymbol{Z}\beta \|_2^2,$$

when variation β is known, one derivative is $\nabla^1 H(\boldsymbol{Z}) = -2\lambda_1 \boldsymbol{X} + 2\lambda_1 \boldsymbol{Z} + 2\lambda_2 \boldsymbol{Z}\beta\beta^{\mathrm{T}} - 2\lambda_2 \boldsymbol{Z}\boldsymbol{M}_1 - 2y\beta^{\mathrm{T}} + 2\boldsymbol{Z}\beta\beta^{\mathrm{T}}$ according to formula (8). Furthermore, second derivative is $\nabla^2 H(\boldsymbol{Z}) = 2\lambda_2 \boldsymbol{I} + (2\lambda_2 + 2)\beta\beta^{\mathrm{T}} - 2\lambda_2 \boldsymbol{M}_1$, where $\boldsymbol{I}$ is a unit matrix of size $N \times N$, and

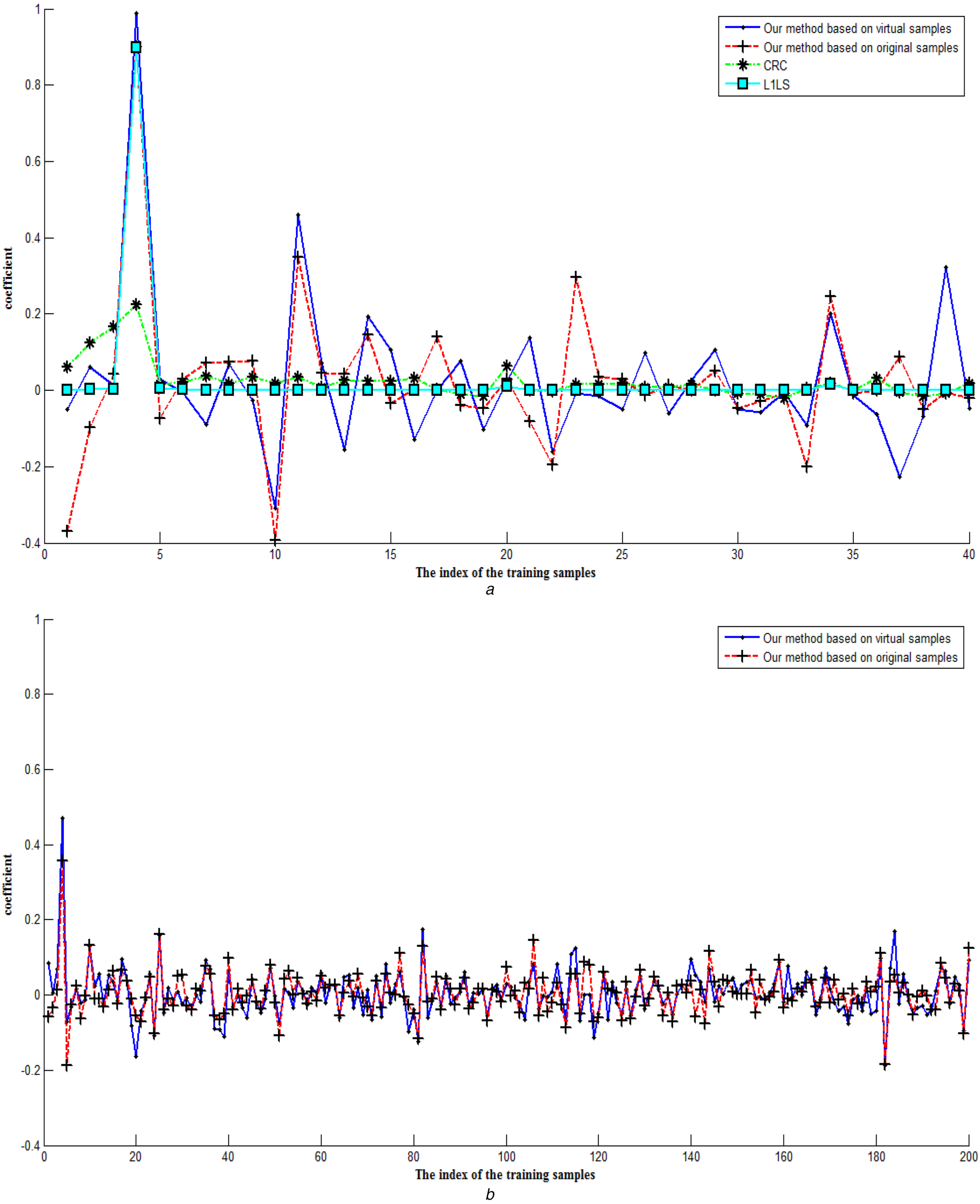
$$\boldsymbol{M}_1 = \begin{pmatrix} \beta_1\beta_1^{\mathrm{T}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_c\beta_c^{\mathrm{T}} \end{pmatrix},$$

$\nabla^2 H(\boldsymbol{Z})$ is a symmetric matrix of size $N \times N$. All principal minors of $\nabla^2 H(\boldsymbol{Z})$ are greater than or equal to zero, hence $\nabla^2 H(\boldsymbol{Z})$ is positive semidefinite matrix, $\nabla^2 H(\boldsymbol{Z}) \succeq 0$. We can draw a conclusion that $H(\beta, \boldsymbol{Z})$ is convex if β is known. Similarly, when variation $\boldsymbol{Z}$ is known, one derivative is $\nabla^1 H(\beta) = 2\lambda_2 \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\beta - 2\lambda_2 \boldsymbol{M}_2\beta - 2\boldsymbol{Z}^{\mathrm{T}}(y - \boldsymbol{Z}\beta)$ according to formula (15). The second derivative is $\nabla^2 H(\beta) = (2\lambda_2 + 2)\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} - 2\lambda_2 \boldsymbol{M}_2$, where

$$\boldsymbol{M}_2 = \begin{pmatrix} \boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{Z}_c^{\mathrm{T}}\boldsymbol{Z}_c \end{pmatrix},$$

$\nabla^2 H(\beta)$ is also a symmetric matrix of size $N \times N$, all the principal minors of $\nabla^2 H(\beta)$ are greater than or equal to zero, so $\nabla^2 H(\beta)$ is a positive semidefinite matrix, $\nabla^2 H(\beta) \succeq 0$. We can draw a conclusion that $H(\beta, \boldsymbol{Z})$ is convex when variation $\boldsymbol{Z}$ is known.

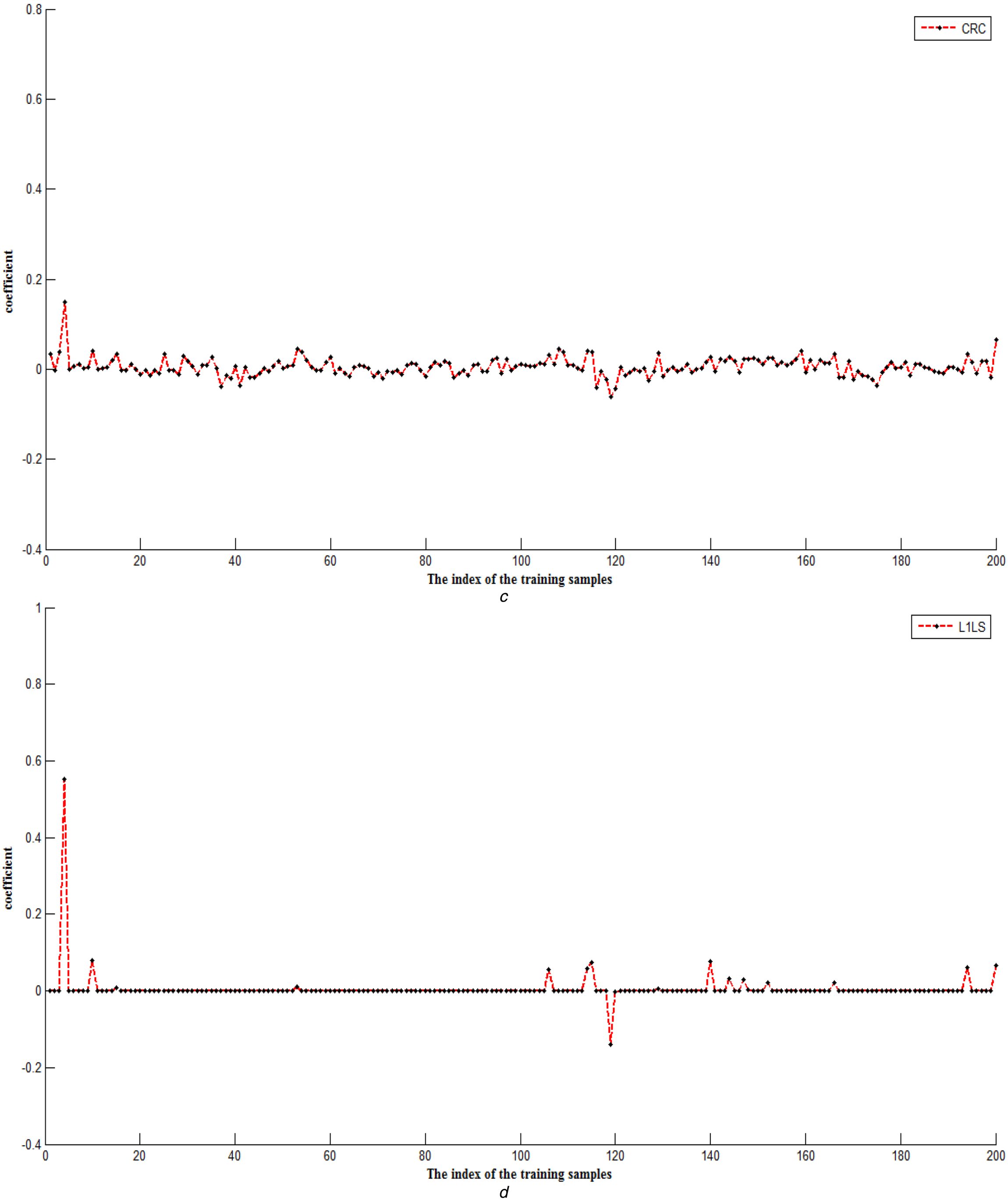
Next, we will analyse the effect of each term of the objective function of the proposed method in detail. The first term of the objective function is $\lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2$. Since $\boldsymbol{Z}$ is the approximation of training sample matrix $\boldsymbol{X}$, it can also be understood that $\boldsymbol{Z} \simeq \boldsymbol{X}$. $\min\left\{\lambda_1 \sum_{i=1}^{c} \| \boldsymbol{X}_i - \boldsymbol{Z}_i \|_2^2\right\}$ is designed to obtain the minimum residual between the original and virtual training sample. From Fig. 1, we can observe that though the



a

b

Fig. 3 *Continued*



c

d

Fig. 3 *(**a**) Representation coefficients on the first test sample of the JAFFE face database obtained by using the proposed method, CRC and L1LS methods. The first four face images of each subject are used as training samples and the remaining images are used for testing, (**b**)–(**d**) Representation coefficients on the first test sample of the ORL face database obtained by using the proposed method, CRC and L1LS, respectively. The first five face images of each subject are used as training samples and the remaining images are used for testing*

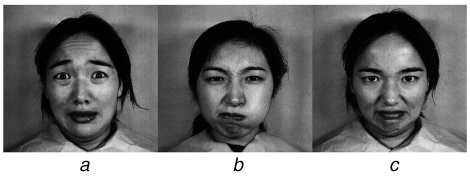


Fig. 4 *Test sample that is erroneously and correctly classified by different methods*

*(**a**)* Test sample from the fifth class, *(**b**)* and *(**c**)* Respectively give each one sample from the first and the fifth classes

virtual training samples seem to be similar to the original training sample, they indeed reflect possible variation of face in illuminations and facial expressions. Meanwhile they are similar to the test sample from the same class to some extent. As far as the second term of the objective function is concerned, $\lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j$ can be rewritten as $\lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j = \lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}(\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j$. Therefore, the minimisation of $\lambda_2\sum_{i=1}^{c}\sum_{j=1}^{c}\beta_i^{\mathrm{T}}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_j\beta_j$ indeed means that the minimisation of $(\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j$, where $\mathbf{Z}_i\beta_i$ stands for the representation result of the test sample obtained by the virtual training samples of the ith class, and it is a column vector of size $D \times 1$. Hence $(\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j$ equates to an inner product of two column vectors, i.e. $\min\left((\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j\right) = \min\left(\| \mathbf{Z}_i\beta_i \|_2 \| \mathbf{Z}_j\beta_j \|_2 \cos\theta\right)$, where $\| \mathbf{Z}_i\beta_i \|_2$ is the length of vector $\mathbf{Z}_i\beta_i$, θ is an angle between $\mathbf{Z}_i\beta_i$ and $\mathbf{Z}_j\beta_j$. So the minimisation of $(\mathbf{Z}_i\beta_i)^{\mathrm{T}}\mathbf{Z}_j\beta_j$ is similar to the minimisation of $\cos\theta$. According to the principle of mathematical trigonometric

Fig. 5 *Some face images from the JAFFE database*

Table 1 Rates of the classification accuracies of different methods on the JAFFE face database

Training samples per class	2, %	3, %	4, %	5, %	6, %
proposed method	82.22	83.53	95.00	96.00	98.57
CRC	78.89	77.65	86.88	88	90.71
L1LS	81.67	80.00	85.63	87.33	91.43
Homotopy	81.11	79.41	87.50	88.00	92.14
DALM	81.11	79.41	87.50	88.00	92.14
LRC	77.78	77.65	83.13	82.00	86.43
FISTA	81.11	79.41	87.50	88.00	92.14
MI_SRC	73.89	78.57	80.00	74.00	87.50

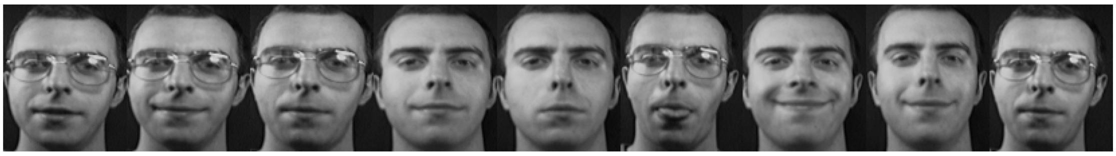


Fig. 6 *Some face images from the ORL face database*

function, a bigger the angle θ means a smaller value of $\cos\theta$. With the increase of θ, the distance between $\boldsymbol{Z}_i\beta_i$ and $\boldsymbol{Z}_j\beta_j$ becomes larger, which implies that the correlation between $\boldsymbol{Z}_i\beta_i$ and $\boldsymbol{Z}_j\beta_j$ is reduced. So minimisation of the sum of $(\boldsymbol{Z}_i\beta_i)^{\mathrm{T}}\boldsymbol{Z}_j\beta_j$ can achieve the de-correlation effect for different classes and representation results of different classes would be very discriminative. Hence, the proposed method is able to classify the test sample with a higher accuracy. The third term of the objective function is $\| y - \boldsymbol{Z}\beta \|_2^2$, which aims to achieve the minimum residual between the test sample and representation result of all virtual samples. Since real data always contains noises, representation noise is unavoidable in most cases. So this residual can also be regarded as noises. The alleviation of noises is also helpful to improve the classification accuracy. It should be noted that the role of β is similar to ρ in the CRC_RLS method (i.e., $\hat{\rho} = \arg\min_\rho \{\| y - \boldsymbol{X}\rho \|_2^2 + \lambda \| \rho \|_2^2\}$). Based on the above analyses, minimising $\cos\theta$ can make sure that β is conducive to classification. Finally, we obtain the optimal values of $\boldsymbol{Z}$ and β by exploiting an iterative method.

4 Experimental results

In this section, we use the ORL [57], JAFFE [58], Columbia Object Image Library (COIL-100) [59], AR [60] and CMU PIE [61] databases to conduct face recognition and image classification experiments. In addition, we compare the proposed method with other methods, including CRC, L1LS, Homotopy, FISTA [62], mirror image and the representation-based classification method (MI_SRC) [63], dual augmented Lagrangian method (DALM) [64], RCR and linear regression classification (LRC) [65].

4.1 Experiments on the JAFFE database

The JAFFE face database contains 213 images of seven facial expressions (six basic facial expressions and one neutral) posed by ten Japanese female models. Each image has been rated on six emotion adjectives by 60 Japanese subjects. The database is planned and assembled by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba. The photos are taken at the Psychology Department in Kyushu University. The size of an image is 256 × 256 pixels. In our experiments, each face image is resized to 64 × 64 image, we only use a consisting of 200 images from 10 subjects with each subject providing 20 images. Fig. 5 shows some face images from the IAFFE face database. We, respectively, take the first two, three, four, five and six face images of each subject as the training samples and regard the rest face images as the test samples. The experimental results are shown in Table 1.

We introduce the implementation details of the compared methods. Firstly, for the proposed method, parameters λ_1, λ_2 and λ_3, respectively, are assigned to 0.01, 0.001, and 0.001. The number of the maximum iterations is assigned to 10, and the parameter of CRC is 0.1. For L1LS method, the optimal value of the regularisation parameter is assigned to 0.1. The parameters of the DALM, FISTA and Homotopy are assigned to 0.01.

As shown in Table 1, it is obvious that the proposed method obtains the best recognition rate of 98.57%. The recognition rate is greatly improved by using the proposed method. For instance, when the number of training samples is four, the classification accuracy rate of our method can greatly outperform the CRC, L1LS, Homotopy, DALM, LRC, FISTA and MI_SRC methods by a list of 86.88, 85.63, 87.50, 87.50, 83.13, 87.50 and 80.00%, respectively. Hence, the proposed method can achieve higher classification accuracy than other methods on the JAFFE face database.

4.2 Experiments on the ORL database

We perform experiments on the ORL face database which includes 400 face images taken from 40 subjects, and each subject provides 10 images. For some subjects, the images are taken at different times, varying the illumination conditions, facial expressions (open or closed eyes, smiling or not smiling) and facial details (glasses or no glasses). All images are taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The original size of an image is 92 × 112 pixels. Each image is resized and cropped to 46 × 56 pixels. Some face images from the ORL face database are shown in Fig. 6. In our experiments, we treat the first one, two, three, four, five and six face images of each subject as original training samples and take the rest of face images as test samples. The experimental results have been shown in Table 2.

In our method, parameters λ_1, λ_2 and λ_3 are assigned to 0.0001, 0.00001 and 0.001, respectively. The number of the maximum iterations is assigned to 10. For CRC, parameter is assigned to 0.1. In addition, the parameters of the L1LS, DALM and Homotopy are assigned to 0.01. The parameter of FISTA is assigned to 0.001.

Table 2 Rates of the classification accuracies of different methods on the ORL face database

Training samples per class	1, %	2, %	3, %	4, %	5, %
proposed method	71.11	87.50	88.93	94.17	94.00
CRC	71.67	83.75	86.07	91.25	90.50
L1LS	71.94	86.25	88.57	92.08	92.50
Homotopy	72.78	86.56	89.64	91.43	92.14
DALM	72.78	86.11	88.33	91.11	92.22
LRC	67.50	79.37	81.43	86.25	88.00
FISTA	72.78	86.11	88.33	91.11	92.22
MI-SRC	71.11	82.19	87.86	87.92	88.50

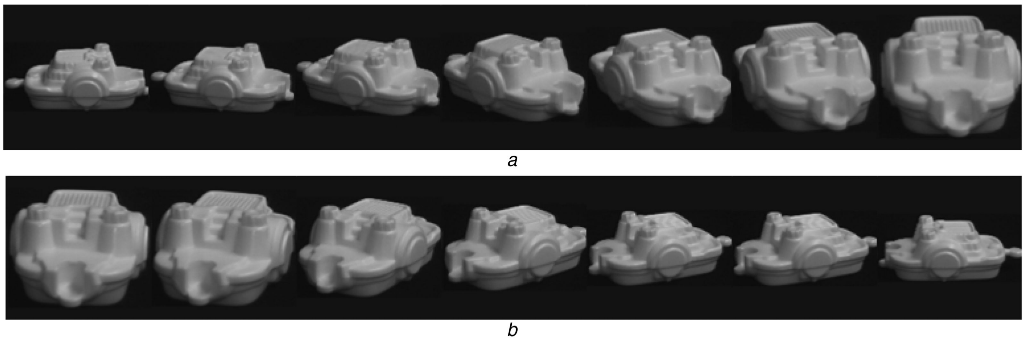


Fig. 7 *Some images from the COIL-100 database*
(a) Case 1, ***(b)*** Case 2

Table 3 Rates of the classification accuracies of different methods on Case 1 of COIL-100 database

Training samples per class	3, %	4, %	5, %	6, %	7, %	8, %	9, %
proposed method	67.33	67.68	66.92	67.29	67.27	68.00	67.50
CRC	61.67	61.25	61.54	63.12	64.09	64.00	65.28
L1LS	63.50	64.46	64.81	65.00	64.55	65.00	65.00
Homotopy	63.33	62.50	61.00	59.83	58.17	57.33	63.83
DALM	63.33	62.50	61.00	59.83	58.17	57.33	63.83
LRC	60.33	61.96	62.50	63.54	64.77	65.00	66.67
FISTA	63.33	62.50	61.00	59.84	58.17	57.33	63.83
MI_SRC	63.17	63.04	64.42	65.00	66.36	67.00	68.00

As can be seen from Table 2 that our method is superior to the other methods. Our method obtains the best recognition accuracy of 94.17%, when the number of training samples is four, and has 2.09, 3.06 and 3.06% higher than the L1LS, DALM and FISTA, respectively. Moreover, when we take the first five face images of each subject as training samples and the rest face images as test samples, the classification accuracies of our method, CRC, L1LS, Homotopy, DALM, LRC, FISTA and MI_CSR methods are 94.00, 90.50, 92.50, 92.14, 92.22, 88.00, 92.22 and 88.50%, respectively. Thus, experimental results show that our method can outperforms other methods.

4.3 Experiments on the Columbia Object Image Library database

Columbia Object Image Library (COIL-100) is a database of colour images of 100 objects. The objects are placed on a motorised turntable against a black background. The turntable is rotated through 360° to vary object pose with respect to a fixed colour camera. Images of the objects are taken at pose intervals of 5°. This corresponds to 72 poses per object. The images are size-normalised. The size of an image is 128 × 128 pixels. We construct two cases to evaluate different methods. In the first case (Case 1), we use the first 18 images of each subject as samples, and select 40 subjects. In the second case (Case 2), we use the 19–36th images of each subject as samples, and take 40 subjects. In addition, we take the first three, four, five, six, seven, eight and nine images of each subject as training samples and use the remaining images as test samples. We simply crop all images and resize them to 64 × 64 pixels. Some face images from the COIL-100 database are shown in Fig. 7.

For the proposed method, parameters λ_1, λ_2 and λ_3 are assigned to 0.001, 0.0000001 and 0.001, respectively. The number of the maximum iterations is assigned to 10. The parameters of CRC, L1LS, Homotopy, DALM and FISTA are set assigned to 0.001.

Tables 3 and 4, respectively, show the rates of classification accuracies of different methods. From the results, it is observed that the proposed method obtains better classification accuracy than the other methods. For instance, in Table 3, when the number of training samples is eight, the classification accuracy of the proposed method greatly outperforms the CRC, L1LS, Homotopy, DALM, LRC, FISTA and MI_SRC methods by a series of 64.00, 65.00, 57.33, 57.33, 65.00, 57.33, 67.00%, respectively. In Table 4, when we consider the first seven face images of each subject as training samples and the rest of the face images as test samples, the classification accuracy of the proposed method, CRC, L1LS, Homotopy, DALM, LRC, FISTA and MI_SRC methods are 66.82, 61.36, 64.77, 63.67, 66.14, 62.50, 64.79 and 63.41%, respectively.

4.4 Experiments on AR database

In the experiment, we use 3120 Gray images from 120 subjects, each providing 26 images. These images are taken at different times, with varying facial expressions, illuminations and facial details (glasses/no glasses, scarves/no scarves). Each image is normalised to 40 × 50 pixels. We take the first two, four, six, eight and ten face images of each subject as the training samples, respectively, and treat the remaining face images as the test samples. The experimental results are shown in Table 5.

Table 4 Rates of the classification accuracies of different methods on the Case 2 of COIL-100 database

Training samples per class	3, %	4, %	5, %	6, %	7, %	8, %	9, %
proposed method	57.17	61.61	66.15	66.67	66.82	69.50	71.67
CRC	51.83	56.61	57.88	57.71	61.36	63.00	65.28
L1LS	54.67	58.04	62.12	63.33	64.77	66.00	68.89
Homotopy	54.83	57.50	61.83	63.83	63.67	64.83	66.00
DALM	54.83	58.04	62.32	64.29	66.14	66.59	66.59
LRC	53.50	55.18	57.31	60.42	62.50	65.50	71.22
FISTA	54.83	57.50	61.83	65.63	64.79	65.63	65.83
MI_SRC	53.33	55.89	56.92	60.21	63.41	66.75	71.58

Table 5 Rates of the classification accuracies of different methods on the AR database

Training samples per class	2, %	4, %	6, %	8, %	10, %
proposed method	69.90	67.58	68.54	65.79	61.51
CRC	58.61	57.46	58.51	59.17	56.29
L1LS	66.28	64.92	65.88	63.43	59.27
Homotopy	65.63	64.17	65.29	63.01	58.49
DALM	65.63	64.17	65.04	63.01	58.49
LRC	59.62	58.14	62.54	59.54	54.74
FISTA	65.59	64.17	65.04	63.01	58.54
MI_SRC	55.14	53.30	57.92	55.42	50.47
RCR	65.83	67.23	66.13	65.60	58.59

Table 6 Rates of the classification accuracies of different methods on the CMU PIE database

Training samples per class	5, %	10, %	15, %	20, %	25, %	30, %
proposed method	25.00	26.75	28.40	29.12	30.96	31.99
CRC	19.92	22.65	24.05	25.53	27.49	30.39
L1LS	17.67	21.01	22.62	24.13	26.34	27.14
Homotopy	15.75	19.61	20.90	22.04	24.24	25.33
DALM	15.75	19.61	21.39	22.04	24.24	25.33
RCR	22.03	18.96	27.39	28.66	30.52	31.29
FISTA	15.74	19.89	21.55	22.04	24.24	25.33
MI_SRC	18.04	22.13	23.89	25.64	28.53	29.69

For our proposed method, parameters λ_1, λ_2 and λ_3 are assigned to10,000, 0.00001 and 0.0001, respectively. The number of the maximum iterations is assigned to 10. From Table 5, we can see that the proposed achieves a higher rate of classification accuracies than all the other methods. For instance, when the first two face images of each subject are used for the training samples and the remaining face images are taken as the test samples, the rates of classification accuracy of our method, CRC, L1LS, Homotopy, DALM, LRC, FISTA, MI_SRC and RCR are 69.90, 58.61, 66.28, 65.63, 65.63, 59.62, 65.59, 55.14 and 65.83%, respectively.

4.5 Experiments on CMU PIE database

In this experiment, the CMU PIE database has 11,560 face images of 68 subjects captured in different sessions with different pose, expression and illumination. Each image is normalised to 32×32 pixels. The first 5, 10, 15, 20, 25 and 30 face images of each subject are treated as the training samples, the remaining face images are taken as the test samples. Table 6 shows the experimental results.

For our proposed method, parameters λ_1, λ_2 and λ_3 are assigned to10,000, 0.00001 and 0.0001, respectively. The number of the maximum iterations is assigned to 5. From Table 6, we can see that the proposed achieves a relatively higher rate of classification accuracies than all the other methods.

5 Conclusion

In this paper, we proposed a novel framework of sparse representation method. The proposed method not only exploits original training samples to perform sparse representation, but also uses virtual training samples to perform classification. Moreover, according to the objective function, we know that the test sample is involved in the generation of virtual training samples. That is, virtual samples may be more consistent with the test sample. It is helpful for the representation method to better represent and recognise the test sample. In addition, we proposed an efficient and competent objective function to enhance the distinctiveness between different classes, that is, the proposed method can reduce the correlation of the representations of the test sample generated from different classes. The proposed method can also improve the robustness of the sparse representation method. Moreover, we take advantages of the score level fusion, which has proven to be very competent and is usually better than the decision level and feature level fusion. The effectiveness of the proposed method has been demonstrated by extensive image classification experiments including face recognition experiments.

6 Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61402274, 31501228, 61461025, 61202314, 41471280, 61501286, U1504610), the Key Science and Technology Program of Shaanxi Province (no. 2016GY-081), the Pivot Science and Technology Innovation Team of Shaanxi Province of China (no. 2014KTC-18), China Postdoctoral Special Science Foundation (no. 200902594), Fundamental Research Funds for the Central Universities (no. GK201402040) and Interdisciplinary Incubation Project of Learning Science of Shaanxi Normal University.

7 References

[1] Xu, Y., Zhang, B., Zhong, Z.: 'Multiple representations and sparse representation for image classification', *Pattern Recognit. Lett.*, 2015, **68**, pp. 9–14

[2] Lu, D., Weng, Q.: 'A survey of image classification methods and techniques for improving classification performance', *Int. J. Remote Sens.*, 2007, **28**, (5), pp. 823–870

[3] Paisitkriangkrai, S., Shen, C., Zhang, J.: 'Performance evaluation of local features in human classification and detection', *IET Comput. Vis.*, 2008, **2**, (28), pp. 236–246

[4] Xiong, H., Liu, T., Tao, D., *et al.*: 'Dual diversified dynamical Gaussian process latent variable model for video repair', *IEEE Trans. Image Process.*, 2016, **25**, (8), pp. 3626–3637

[5] Liu, T., Tao, D.: 'On the performance of MahNMF Manhattan non-negative matrix factorization', *IEEE Trans. Neural Netw. Learn. Syst.*, 2016, **27**, (9), pp. 1851–1863

[6] Duan, C., Meng, X., Tu, C., *et al.*: 'How to make local image features more efficient and distinctive', *IET Comput. Vis.*, 2008, **2**, (3), pp. 178–189

[7] Walia, E., Suneja, A.: 'Fragile and blind watermarking technique based on Weber's law for medical image authentication', *IET Comput. Vis.*, 2013, **7**, (1), pp. 9–19

[8] Fei, L., Xu, Y., Tang, W., *et al.*: 'Double-orientation code and nonlinear matching scheme for palmprint recognition', *Pattern Recognit.*, 2015, **49**, (C), pp. 89–101

[9] Xu, C., Liu, T., Tao, D., *et al.*: 'Local Rademacher complexity for multi-label learning', *IEEE Trans. Image Process.*, 2016, **25**, (3), pp. 1495–1507

[10] Luo, Y., Liu, T., Tao, D., *et al.*: 'Multiview matrix completion for multilabel image classification', *IEEE Trans. Image Process.*, 2015, **24**, (8), pp. 2261–2274

[11] Xu, Y., Fei, L., Zhang, D.: 'Combining left and right palmprint images for more accurate personal identification', *IEEE Trans. Image Process.*, 2015, **24**, (2), pp. 549–559

[12] Luo, Y., Liu, T., Tao, D., *et al.*: 'Decomposition-based transfer distance metric learning for image classification', *IEEE Trans. Image Process.*, 2014, **23**, (9), pp. 3789–3801

[13] Yang, M., Zhang, L., Yang, J., *et al.*: 'Robust sparse coding for face recognition'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2011, pp. 625–632

[14] Zhang, Z., Xu, Y., Yang, J., *et al.*: 'A survey of sparse representation: algorithms and applications', *Access IEEE*, 2015, **3**, pp. 490–530

[15] Fei, L., Xu, Y., Zhang, B., *et al.*: 'Low-rank representation integrated with principal line distance for contactless palmprint recognition', *Neurocomputing*, 2016, **218**, pp. 264–275

[16] Mei, X., Ling, H.: 'Robust visual tracking and vehicle classification via sparse representation', *IEEE Trans. Softw. Eng.*, 2011, **33**, (11), pp. 2259–2272

[17] Yang, J., Wright, J., Huang, T., *et al.*: 'Image super-resolution via sparse representation', *IEEE Trans. Image Process.*, 2010, **19**, (11), pp. 2861–2873

[18] Kreutzdelgado, K., Murray, J., Rao, B., *et al.*: 'Dictionary learning algorithms for sparse representation', *Neural Comput.*, 2003, **15**, (2), pp. 349–396

[19] Mairal, J., Elad, M., Sapiro, G.: 'Sparse representation for color image restoration', *IEEE Trans. Image Process.*, 2008, **17**, (1), pp. 53–69

[20] Zhang, T., Liu, S., Ahuja, N., *et al.*: 'Robust visual tracking via consistent low-rank sparse learning', *Int. J. Comput. Vis.*, 2014, **111**, (2), pp. 171–190

[21] Wright, J., Ma, Y., Mairal, J., *et al.*: 'Sparse representation for computer vision and pattern recognition', *Proc. IEEE*, 2010, **98**, (6), pp. 1031–1044

[22] Zhang, L., Yang, M., Feng, X.: 'Sparse representation or collaborative representation: Which helps face recognition?'. Proc. Int. Conf. on Computer Vision, 2012, pp. 471–478

[23] Naseem, I., Togneri, R., Bennamoun, M.: 'Robust regression for face recognition', *Pattern Recognit.*, 2012, **45**, (1), pp. 104–118

[24] He, R., Zheng, W., Hu, B.: 'Maximum correntropy criterion for robust face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **33**, (8), pp. 1561–1576

[25] Xu, Y., Fang, X., Wu, J., *et al.*: 'Discriminative transfer subspace learning via low-rank and sparse representation', *IEEE Trans. Image Process.*, 2015, **25**, (2), pp. 1–1

[26] Deng, W., Hu, J., Guo, J.: 'Extended SRC: undersampled face recognition via intraclass variant dictionary', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (9), pp. 1864–1870

[27] Wang, D., Lu, H., Yang, M.: 'Kernel collaborative face recognition', *Pattern Recognit.*, 2015, **48**, (10), pp. 3025–3037

[28] Wang, S., Zhang, D., Zhang, L., *et al.*: 'Relaxed collaborative representation for pattern classification'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2012, pp. 2224–2231

[29] Zhang, T., Ghanem, B., Liu, S., *et al.*: 'Low-rank sparse coding for image classification'. Proc. Int. Conf. on Computer Vision, 2013, pp. 281–288

[30] Yang, J., Yu, K., Gong, Y., *et al.*: 'Linear spatial pyramid matching using sparse coding for image classification'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2009, pp. 1794–1801

[31] Mancera, L., Portilla, J.: 'L0-norm-based sparse representation through alternate projections'. Proc. Int. Conf. on Image Processing, 2006, pp. 2089–2092

[32] Wright, B., Member, S., Yang, A., *et al.*: 'Robust face recognition via sparse representation', *J. Inner Mongolia Agricultural Univ.*, 2010, **31**, (2), pp. 210–227

[33] Yang, J., Zhang, L., Xu, Y., *et al.*: 'Beyond sparsity: the role of L1-optimizer in pattern classification', *Pattern Recognit.*, 2012, **45**, (3), pp. 1104–1118

[34] Donoho, D.: 'For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution', *Commun. Pure Appl. Math.*, 2006, **59**, (6), pp. 797–829

[35] Candès, E., Romberg, J., Tao, T.: 'Stable signal recovery from incomplete and inaccurate measurements', *Commun. Pure Appl. Math.*, 2005, **19**, (5), pp. 410–412

[36] Candes, E., Tao, T.: 'Near-optimal signal recovery from random projections: universal encoding strategies?', *IEEE Trans. Inf. Theory*, 2007, **52**, (12), pp. 5406–5425

[37] Yang, A., Zhou, Z., Balasubramanian, A., *et al.*: 'Fast l1-minimization algorithms for robust face recognition', *IEEE Trans. Image Process.*, 2010, **22**, (8), pp. 3234–3246

[38] Yang, J., Zhang, Y.: 'Alternating direction algorithms for l1-problems in compressive sensing', *SIAM J. Sci. Comput.*, 2009, **33**, (1), pp. 250–278

[39] Chen, S., Chen, G., Gu, R.: 'An efficient L2-norm regularized least-squares temporal difference learning algorithm', *Knowl.-Based Syst.*, 2013, **45**, (3), pp. 94–99

[40] Zhang, Z., Wang, L., Zhu, Q., *et al.*: 'Noise modeling and representation based classification methods for face recognition', *Neurocomputing*, 2015, **148**, pp. 420–429

[41] Rigamonti, R., Brown, M., Lepetit, V.: 'Are sparse representations really relevant for image classification?'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2011, pp. 1545–1552

[42] Shi, Q., Eriksson, A., Hengel, A., *et al.*: 'Is face recognition really a Compressive Sensing problem?'. Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2011, pp. 553–560

[43] Xu, Y., Zhang, D., Yang, J., *et al.*: 'A two-phase test sample sparse representation method for use with face recognition', *IEEE Trans. Circuits Syst. Video Technol.*, 2011, **21**, (9), pp. 1255–1262

[44] Liu, Z., Pu, J., Huang, T., *et al.*: 'A novel classification method for palmprint recognition based on reconstruction error and normalized distance', *Appl. Intell.*, 2013, **39**, (2), pp. 307–314

[45] Xu, Y., Zhu, Q., Chen, Y., *et al.*: 'An improvement to the nearest neighbor classifier and face recognition experiments', *Int. J. Innov. Comput. Inf. Control*, 2013, **9**, (2), pp. 543–554

[46] Kim, S.: 'On using a dissimilarity representation method to solve the small sample size problem for face recognition'. Proc. Int. Conf. on Advanced Concepts for Intelligent Vision Systems, 2006, pp. 1174–1185

[47] Xu, Y., Fang, X., Li, X., *et al.*: 'Data uncertainty in face recognition', *IEEE Trans. Cybern.*, 2014, **44**, (10), pp. 950–1961

[48] Thian, N., Marcel, S., Bengio, S.: 'Improving face authentication using virtual samples'. Proc. Int. Conf. on Acoustics, 2003, pp. III-233–236

[49] Xu, Y., Zhu, X., Li, Z., *et al.*: 'Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition', *Pattern Recognit.*, 2013, **46**, (4), pp. 1151–1158

[50] Ryu, Y.-S., Oh, S.-Y.: 'Simple hybrid classifier for face recognition with adaptively generated virtual data', *Pattern Recognit. Lett.*, 2002, **23**, (7), pp. 833–841

[51] Xu, Y., Zhang, Z., Lu, G., *et al.*: 'Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification', *Pattern Recognit.*, 2016, **54**, pp. 68–82

[52] Wang, W., Yang, J.: 'Quadratic discriminant analysis method based on virtual training samples', *Acta Autom. Sin.*, 2008, **34**, (34), pp. 400–407

[53] Xu, Y., Li, X., Yang, J., *et al.*: 'Integrating conventional and inverse representation for face recognition', *IEEE Trans. Cybern.*, 2014, **44**, (10), pp. 1738–1746

[54] Poggio, T., Vetter, T.: '*Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries*' (Massachusetts Inst of Technology Cambridge Artificial Intelligence Lab, 1992), p. 1347

[55] Kim, S., Koh, K., Lustig, M., *et al.*: 'An interior-point method for large-scale l 1 -regularized least squares', *IEEE J. Sel. Top. Signal Process.*, 2007, **1**, (4), pp. 606–617

[56] Boyd, S., Vandenberghe, L.: '*Convex functions*', in Boyd, S. (Ed.): '*Convex optimization*' (Cambridge University Press, 2004, 3rd edn.), pp. 67–71

[57] Samaria, F., Harter, A.: 'Parameterisation of a stochastic model for human face identification'. Proc. Int. Conf. the Second IEEE Workshop on Applications of Computer Vision, 1995, pp. 138–142

[58] 'The JAFFE database', http://www.kasrl.org/jaffe.html, accessed 6 December 2016

[59] 'The COIL-100 database', http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php, accessed 6 December 2016

[60] 'The AR database', http://www2.ece.ohio-state.edu/~aleix/ARdatabase.htmlS, accessed 6 December 2016

[61] Gross, R., Matthews, I., Cohn, J., *et al.*: 'Multi-PIE', *Image Vis. Comput.*, 2010, **28**, (5), pp. 807–813

[62] Beck, A., Teboulle, M.: 'A fast iterative shrinkage-thresholding algorithm for linear inverse problems', *Siam J. Imaging Sci.*, 2009, **2**, (1), pp. 183–202

[63] Xu, Y., Li, X., Yang, J., *et al.*: 'Integrate the original face image and its mirror image for face recognition', *Neurocomputing*, 2014, **131**, (7), pp. 191–199

[64] Shia, V., Yang, A., Sastry, S., *et al.*: 'Fast L1-minimization and parallelization for face recognition'. Proc. Int. Conf. 2011 Conf. Record of the Forty Fifth Asilomar Conf. on Signals, Systems and, 2011, pp. 1199–1203

[65] Naseem, I., Togneri, R., Bennamoun, M.: 'Linear regression for face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (11), pp. 2106–2112