

# Discriminative collaborative representation for multimodal image classification

Dawei Sun, Shicheng Wang, Dongfang Yang and Yongfei Li

## Abstract

Sparse representation has been widely researched for image-based classification. However, sparse representation classification directly treats training samples as a dictionary, so it needs a large training set and is time consuming, especially for a large training set. To derive a small dictionary, many dictionary learning algorithms are researched. Thus, object recognition problem is transformed to optimize the sparse representation errors on the compact dictionary. The sparse representation optimization is constraint by  $l_0$ -norm, which is NP-hard problem. Though we can use  $l_1$ -norm minimization instead to work effectively, it is still time consuming for optimization. To make the algorithm discriminative and simultaneously decrease the computational burden, we proposed a fast discriminative collaborative representation-based classification algorithm. The new algorithm incorporated the within-class scatter and the linear classification error terms into the objective function to derive a more discriminative dictionary and simultaneously added collaborative representation mechanism to cut off the time consuming. At the end of this article, we designed two experiments to validate our method using near-infrared and AR visible databases for multimodal face recognition. The results showed that our algorithm outperformance sparse representation-based, collaborative representation-based, and discriminative-KSVD classification algorithms.

## Keywords

Object recognition and classification, collaborative representation, sparse representation, sparse coding, multimodal image

Date received: 23 February 2017; accepted: 4 May 2017

Topic: Special Issue - Multi-Modal Fusion for Robotics

Topic Editor: Huaping Liu

Associate Editor: Huaping Liu

## Introduction

In the past few years, sparse representation, also called sparse coding, were widely researched by computer community. The technology of sparse representation was used for image denoising<sup>1,2</sup> image analysis,<sup>3</sup> image super resolution,<sup>4</sup> and especially image-based classification such as face recognition,<sup>5–11</sup> automatic target recognition,<sup>12</sup> and traffic sign recognition.<sup>13</sup> Wright et al.<sup>6</sup> used sparse representation for classification task and proposed a sparse representation-based classification (SRC) in face recognition task. In SRC framework, training samples of all classes are arranged as columns of a matrix usually called dictionary, and the query image is considered linearly represented by the atoms of the

dictionary. Most entries of the linear representation coefficient coding vector are zeros or approximate zeros, that is, the coefficient vectors are sparse. The minimal reconstruction errors are used for classification. However, when the training data set is larger, SRC is time consuming for all training samples are regarded as a dictionary.

---

High-Tech Institute of Xi'an, Xi'an, Shaanxi, People's Republic of China

### Corresponding author:

Dawei Sun, High-Tech Institute of Xi'an, Xi'an, Shaanxi 710025, People's Republic of China.

Email: sundawei988@163.com



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License

(<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

In order to cut down the time consuming, many researchers proposed dictionary learning (DL) algorithms<sup>3,5,11,14–18</sup> to derive a smaller dictionary. K-SVD algorithm<sup>14</sup> is the representative one of DL algorithms. An overcomplete dictionary can be learned from a training data set by K-SVD, which works well for signal representation but not for classification tasks. To address the image classification issue, Mairal et al.<sup>3</sup> added a discriminative reconstruction constraint and optimized both class discrimination and sparse reconstruction components based on K-SVD. In a study by Pham and Venkatesh,<sup>16</sup> a dictionary construction and joint learning algorithm was proposed for classification task. Later, a discriminative K-SVD (D-KSVD) was proposed for face recognition by Zhang and Li et al.<sup>5</sup> The classification error was incorporated into the objective function based on extending K-SVD algorithm. The linear classifier was utilized for obtaining the query image's label in D-KSVD method. Jiang et al.<sup>17</sup> introduced a new discriminative sparse coding error constraint to jointly learn a single dictionary and a linear classification, the algorithm is called label consistent K-SVD (LC-KSVD). The common character of the methods mentioned above is that a shared dictionary is learned during DL scheme. Different from the above methods, Yang et al.<sup>18</sup> proposed to learn a structured dictionary. Fisher discrimination dictionary learning method, which made the coding coefficients having big between-class scatter but small within-class scatter, was used in the classification scheme. Xu et al.<sup>11</sup> synthesized the advantages of the methods mentioned above. They proposed supervised within-class similar discriminative DL method, which incorporates linear classification error term and within-class representation coefficients constraint into objective function for face recognition.

The sparsity of the sparse representation coefficients is measured by  $l_0$ -norm, which is NP-hard. Using  $l_1$ -norm minimization instead is popular. It is still time consuming, though  $l_1$ -norm minimization is more efficient. Zhang et al.<sup>19</sup> indicated that it was the collaborative representation (CR) mechanism that improved the face recognition accuracy. Motivated by this idea, we proposed a discriminative collaborative representation-based classification (DCRC), which incorporated the within-class scatter and the linear classification error terms into the objective function to derive a more discriminative dictionary, and simultaneously added CR mechanism to decrease the time cost. Besides, recently researchers proposed algorithms for multisource image classification.<sup>20,21</sup> Liu et al.<sup>20</sup> proposed a joint sparse coding model to solve the room-level localization using multiple sets. Li et al.<sup>22</sup> proposed multimodal fusion method for object recognition. Our algorithm is also effective for multimodal image sets.

This article was organized as follows. In the second section, we described the related works about classification on sparse representation and CR, respectively. In the third section, we described the methodology of the DCRC

algorithm. Experiments were performed in the fourth section using different well-known databases to prove the validity of the proposed method. The conclusions were given in the fifth section.

## Related works

### Brief introduction of sparse representation-based classification

SRC was first proposed for face recognition by Wright et al.<sup>6</sup> The SRC framework contains two procedures, sparse coding and classification. Suppose the training samples from  $K$  different classes are denoted as  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_i, \dots, \mathbf{D}_K] \in \mathbb{R}^{d \times n}$ ,  $i = 1, \dots, K$ , where  $\mathbf{D}_i \in \mathbb{R}^{d \times n_i}$  is  $n_i$  training samples subset from class  $i$ , and  $n = \sum_{i=1}^K n_i$ . Given a query image  $\mathbf{z} \in \mathbb{R}^d$ , SRC considers the query image as a sparse linear combination of the training data set, and the sparse coding can be described as follows

$$\mathbf{x} = \arg \min_x \|\mathbf{z} - \mathbf{D}\mathbf{x}\|_2^2, \quad s.t. \|\mathbf{x}\|_0 \leq \tau \quad (1)$$

where  $\mathbf{x}$  is the code coefficient vector,  $\tau$  is a sparsity constant, and  $l_0$ -norm counts the number of nonzero elements of  $\mathbf{x}$ . This optimization is an NP-hard problem, and it is generally transformed to an  $l_1$ -norm minimization instead

$$\mathbf{x} = \arg \min_x \|\mathbf{z} - \mathbf{D}\mathbf{x}\|_2^2, \quad s.t. \|\mathbf{x}\|_1 \leq \tau \quad (2)$$

or

$$\mathbf{x} = \arg \min_x \{\|\mathbf{z} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1\} \quad (3)$$

where  $\lambda$  is a scalar constant to balance the sparsity and reconstruction error terms. Equation (3) is equivalent to equation (2).

The coding vector  $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_K]$  is obtained, and then we do classification via reconstruction errors minimization:

$$\begin{aligned} \text{Label}(\mathbf{z}) &= \arg \min_i r_i(\mathbf{z}) \\ &= \arg \min_i \|\mathbf{z} - \mathbf{D}_i \mathbf{x}_i\|_2, \quad i = 1, \dots, K \end{aligned} \quad (4)$$

where  $\mathbf{x}_i$  is the representation coefficient vector corresponding to class  $i$ .

### Dictionary learning model

The SRC framework directly used training data set as the dictionary. In order to learn a smaller dictionary, DL model was proposed. DL model used the training data set to learn a corresponding compact dictionary  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m] \in \mathbb{R}^{d \times m}$  with  $m$  atoms ( $m < n$ ). Then, DL framework was written as follows<sup>11</sup>

$$\begin{aligned} \langle \mathbf{D}, \mathbf{X} \rangle &= \arg \min_{\mathbf{D}, \mathbf{X}} \{ \|\mathbf{Y} - \mathbf{DX}\|_F^2 + \lambda \|\mathbf{X}\|_1 \} \\ \text{s.t. } \|\mathbf{d}_j\|_2^2 &= 1, \quad j = 1, 2, \dots, m \end{aligned} \quad (5)$$

where  $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_i, \dots, \mathbf{Y}_K] \in \mathbb{R}^{d \times n}$ ,  $i = 1, \dots, K$  is the training data set,  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$  is the sparse representation coefficients matrix of training set  $\mathbf{Y}$  on dictionary  $\mathbf{D}$ ,  $\|\cdot\|_F$  is the Frobenius norm, and  $\|\mathbf{X}\|_1 = \sum_{i=1}^n \|\mathbf{x}_i\|_1$ ,  $\lambda$  is a scalar constant to balance the sparsity and reconstruction error terms as equation (3).

The first term in equation (5) denotes reconstruction errors, so the model is suitable for signal representation tasks but not for classification tasks. The model in equation (5) is an unsupervised DL framework, because the labels of the training set are not taken into account. A supervised DL framework is designed for classification tasks via adding discriminative terms into objective function. Xu et al.<sup>11</sup> synthesized the methods in literature<sup>5,17,18</sup> and proposed a supervised DL framework called SCDDL.

### Classification based on collaborative representation

The DL methods mentioned above are  $l_1$ -norm minimizing optimization problem. Though there are many speeding up methods proposed as reviewed in the study by Zhang et al.,<sup>22</sup> they are complicated and still time consuming. Zhang et al.<sup>19</sup> analyzed the mechanism of SRC and indicated that CR played an essential role for classification in SRC but not  $l_1$ -norm sparsity. To collaboratively represent the query image  $\mathbf{z}$ , they proposed to utilize the regularized least square method. The CR-based classification with regularized least square (CRC-RLS) model is as follows

$$\mathbf{x} = \arg \min_x \{ \|\mathbf{z} - \mathbf{Dx}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \} \quad (6)$$

where  $\lambda$  is the regularization parameter. It is easy to derive the analytical solution of equation (7) as

$$\mathbf{x} = \mathbf{Pz} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T \mathbf{z} \quad (7)$$

where  $\mathbf{P}$  is a projection matrix independent of  $\mathbf{z}$ . Matrix  $\mathbf{P}$  can be calculated offline, and this makes CR more effective.

It indicates that equation (6) makes the classification discriminative, while equation (7) makes the classification fast. Can we combine these models to make the classification procedure more discriminative and faster?

### Discriminative collaborative representation for image-based classification

Drawing inspiration from literatures,<sup>5,11,17-19</sup> we proposed a DCRC algorithm. The DCRC method was described as follows.

#### Discriminative collaborative representation-based dictionary learning

Suppose  $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_i, \dots, \mathbf{Y}_K] \in \mathbb{R}^{d \times n}$ ,  $i = 1, \dots, K$  is the set of training samples from  $K$  classes. Every training sample is a  $d$ -dimensional column vector of  $\mathbf{Y}$ . Discriminative collaborative representation-based DL is to derive a discriminative dictionary  $\mathbf{D} \in \mathbb{R}^{d \times m}$  and a classifier  $\mathbf{W} \in \mathbb{R}^{K \times m}$  for classification. Dictionary  $\mathbf{D}$  consists of  $m$  atoms, and  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_K] \in \mathbb{R}^{m \times n}$  is the coding coefficients matrix of training samples on dictionary  $\mathbf{D}$ . The discriminative collaborative representation-based DL model can be written as follows

$$\begin{aligned} \langle \mathbf{D}, \mathbf{W}, \mathbf{X} \rangle &= \arg \min_{\mathbf{D}, \mathbf{W}, \mathbf{X}} \left\{ \|\mathbf{Y} - \mathbf{DX}\|_F^2 + \lambda_1 \|\mathbf{X}\|_F^2 + \alpha \|\mathbf{H} - \mathbf{WX}\|_F^2 + \right. \\ &\quad \left. \beta \|\mathbf{W}\|_F^2 + \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2 + \eta \|\mathbf{X}_i\|_F^2) \right\} \\ \text{s.t. } \|\mathbf{d}_j\|_2^2 &= 1, \quad j = 1, \dots, m \end{aligned} \quad (8)$$

where  $\|\mathbf{X}\|_F^2 = \sum_{i=1}^K \|\mathbf{X}_i\|_F^2$  and  $\|\cdot\|_F$  is the Frobenius norm.

Then we can rewrite equation (8) as

$$\begin{aligned} \langle \mathbf{D}, \mathbf{W}, \mathbf{X} \rangle &= \arg \min_{\mathbf{D}, \mathbf{W}, \mathbf{X}} \left\{ \begin{aligned} &\|\mathbf{Y} - \mathbf{DX}\|_F^2 + \lambda_1 \sum_{i=1}^K \|\mathbf{X}_i\|_F^2 + \alpha \|\mathbf{H} - \mathbf{WX}\|_F^2 + \\ &\beta \|\mathbf{W}\|_F^2 + \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2 + \eta \|\mathbf{X}_i\|_F^2) \end{aligned} \right\} \\ &= \arg \min_{\mathbf{D}, \mathbf{W}, \mathbf{X}} \left\{ \begin{aligned} &\|\mathbf{Y} - \mathbf{DX}\|_F^2 + (\lambda_1 + \lambda_2 \eta) \sum_{i=1}^K \|\mathbf{X}_i\|_F^2 + \\ &\alpha \|\mathbf{H} - \mathbf{WX}\|_F^2 + \beta \|\mathbf{W}\|_F^2 + \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2) \end{aligned} \right\} \\ &s.t. \quad \|\mathbf{d}_j\|_2^2 = 1, \quad j = 1, \dots, m \end{aligned} \quad (9)$$

Here, we set  $\eta = 1$  as in Xu et al.<sup>11</sup> for simplicity. Then, equation (9) is rewritten as

$$\begin{aligned} \langle \mathbf{D}, \mathbf{W}, \mathbf{X} \rangle &= \arg \min_{\mathbf{D}, \mathbf{W}, \mathbf{X}} \left\{ \begin{aligned} &\|\mathbf{Y} - \mathbf{DX}\|_F^2 + (\lambda_1 + \lambda_2) \|\mathbf{X}\|_F^2 + \\ &\alpha \|\mathbf{H} - \mathbf{WX}\|_F^2 + \beta \|\mathbf{W}\|_F^2 + \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2) \end{aligned} \right\} \\ &s.t. \quad \|\mathbf{d}_j\|_2^2 = 1, \quad j = 1, \dots, m \end{aligned} \quad (10)$$

where  $\|\mathbf{Y} - \mathbf{DX}\|_F^2$  is the reconstruction errors term of training data set  $\mathbf{Y}$  on dictionary  $\mathbf{D}$ ,  $(\lambda_1 + \lambda_2) \|\mathbf{X}\|_F^2$  is the coding coefficients restriction term,  $\alpha \|\mathbf{H} - \mathbf{WX}\|_F^2 + \beta \|\mathbf{W}\|_F^2$  is the classification errors term, and  $\lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2)$  is the within-class similar term as in the study by Xu et al.<sup>11</sup>  $\mathbf{W} \in \mathbb{R}^{K \times m}$  is the linear classifier, and  $\mathbf{H} \in \mathbb{R}^{K \times n}$  is the label matrix corresponding to the training samples. Each column of  $\mathbf{H}$  is a vector with the form  $[0, \dots, 0, 1, 0, \dots, 0]^T \in \mathbb{R}^K$ , where 1 locates the training samples to the corresponding class.  $\mathbf{X}_i$  is the coding coefficients corresponding to class  $i$ .  $\mathbf{M}_i$  is the same type matrix

of  $\mathbf{X}_i$ , and each column of  $\mathbf{M}_i$  is the mean vector of  $\mathbf{X}_i$ . Parameters  $\lambda_1$ ,  $\lambda_2$ ,  $\alpha$ , and  $\beta$  are all nonnegative constants. This framework simultaneously derives the dictionary  $\mathbf{D}$  and classifier  $\mathbf{W}$ . Once the classifier is obtained, the query image can be easily classified by the classifier.

### Optimization scheme

The dictionary  $\mathbf{D}$  and classifier  $\mathbf{W}$  can be optimized simultaneously. We can synthesize  $\mathbf{D}$  and  $\mathbf{W}$  to derive an extended matrix as in literature,<sup>5,11,17</sup> and equation (10) can be transformed to

$$\langle \mathbf{D}, \mathbf{W}, \mathbf{X} \rangle = \arg \min_{\mathbf{D}, \mathbf{W}, \mathbf{X}} \left\{ \begin{aligned} &\left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha} \mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha} \mathbf{W} \end{pmatrix} \mathbf{X} \right\|_F^2 + (\lambda_1 + \lambda_2) \|\mathbf{X}\|_F^2 + \\ &\beta \|\mathbf{W}\|_F^2 + \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_F^2) \end{aligned} \right\} \quad (11)$$

If  $\left(\frac{\mathbf{D}}{\sqrt{\alpha}\mathbf{W}}\right)$  is normalized column-wise, the regularization penalty can be dropped.<sup>5</sup> Denote  $\left(\frac{\mathbf{Y}}{\sqrt{\alpha}\mathbf{H}}\right)$  as  $\mathbf{Y}^*$  and

$$\langle \mathbf{D}^*, \mathbf{X} \rangle = \arg \min_{\mathbf{D}^*, \mathbf{X}} \left\{ \begin{array}{l} \|\mathbf{Y}^* - \mathbf{D}^* \mathbf{X}\|_{\text{F}}^2 + (\lambda_1 + \lambda_2) \|\mathbf{X}\|_{\text{F}}^2 + \\ \lambda_2 \sum_{i=1}^K (\|\mathbf{X}_i - \mathbf{M}_i\|_{\text{F}}^2) \end{array} \right\} \quad (12)$$

*s.t.*  $\|\mathbf{d}_j^*\|_2^2 = 1, \quad j = 1, \dots, m$

We can alternatively update  $\mathbf{X}$  and  $\mathbf{D}^*$ , that is, fix one of them, and update the other.

Now, we fix the value of  $\mathbf{D}^*$  and then update  $\mathbf{X}$ . The function in equation (12) is turned to a coding problem

$$\langle \mathbf{X} \rangle = \arg \min_{\mathbf{X}} \left\{ \begin{array}{l} \|\mathbf{Y}^* - \mathbf{D}^* \mathbf{X}\|_{\text{F}}^2 + (\lambda_1 + \lambda_2) \|\mathbf{X}\|_{\text{F}}^2 + \\ \lambda_2 \sum_{i=1}^k (\|\mathbf{X}_i - \mathbf{M}_i\|_{\text{F}}^2) \end{array} \right\} \quad (13)$$

We can derive  $\mathbf{X}$  class by class. For example, if we want to obtain  $\mathbf{X}_i$  of class  $i$ , the function in equation (13) can be written by class

$$\langle \mathbf{X}_i \rangle = \arg \min_{\mathbf{X}_i} \{ \|\mathbf{Y}_i^* - \mathbf{D}^* \mathbf{X}_i\|_{\text{F}}^2 + (\lambda_1 + \lambda_2) \|\mathbf{X}_i\|_{\text{F}}^2 + \lambda_2 \|\mathbf{X}_i - \mathbf{M}_i\|_{\text{F}}^2 \} \quad (14)$$

The objective function in equation (14) is differentiable, so we can obtain  $\mathbf{X}_i$  via derivative of objective function and setting to be zero. Denote  $f$  as the objective function and then derivative

$$\frac{\partial f}{\partial \mathbf{X}_i} = \left\{ \begin{array}{l} -2\mathbf{D}^{*T} \mathbf{Y}_i^* + 2\mathbf{X}_i + \\ 2(\lambda_1 + \lambda_2) \mathbf{X}_i + 2\lambda_2 (\mathbf{X}_i - \mathbf{M}_i) \end{array} \right\} \quad (15)$$

Set  $\frac{\partial f}{\partial \mathbf{X}_i}$  to be zero and denote the obtained variables from the  $k$ -th to  $(k+1)$ -th iteration using the subscripts  $(k)$  and  $(k+1)$ . Thus,  $\mathbf{X}_i$  can be calculated as follows

$$\mathbf{X}_{i(k+1)} = \frac{\lambda_2 \mathbf{M}_{i(k)} + \mathbf{D}_{(k)}^{*T} \mathbf{Y}_{i(k)}^*}{1 + \lambda_1 + 2\lambda_2} \quad (16)$$

After obtaining the coding coefficient  $\mathbf{X}$ , we fix it and update  $\mathbf{D}^*$ . Then the function in equation (12) is reduced to a DL problem

$$\langle \mathbf{D}^* \rangle = \arg \min_{\mathbf{D}^*} \|\mathbf{Y}^* - \mathbf{D}^* \mathbf{X}\|_{\text{F}}^2 \quad (17)$$

*s.t.*  $\|\mathbf{d}_j^*\|_2^2 = 1, \quad j = 1, \dots, m$

The objective function in equation (17) is differentiable, so we can derive  $\mathbf{D}^*$  as follows

denote  $\left(\frac{\mathbf{D}}{\sqrt{\alpha}\mathbf{W}}\right)$  as  $\mathbf{D}^*$ . Then, the objective function can be simply written as

$$\mathbf{D}_{(k+1)}^* = \mathbf{Y}_{(k)}^* \mathbf{X}_{(k+1)}^T (\mathbf{X}_{(k+1)} \mathbf{X}_{(k+1)}^T)^{-1} \quad (18)$$

$\mathbf{D}^* = \left(\frac{\mathbf{D}}{\sqrt{\alpha}\mathbf{W}}\right)$  contains  $\mathbf{W}$  and  $\mathbf{D}$ , so we can derive dictionary  $\mathbf{D}$  and classifier  $\mathbf{W}$  simultaneously via the lemma in the study by Zhang and Li.<sup>5</sup> Suppose  $\mathbf{D}^* = \left(\frac{\mathbf{d}_1^*, \mathbf{d}_2^*, \dots, \mathbf{d}_m^*}{\mathbf{w}_1^*, \mathbf{w}_2^*, \dots, \mathbf{w}_m^*}\right)$ , then we can obtain  $\mathbf{D}$  and  $\mathbf{W}$  as follows

$$\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m) = \left(\frac{\mathbf{d}_1^*}{\|\mathbf{d}_1^*\|_2}, \frac{\mathbf{d}_2^*}{\|\mathbf{d}_2^*\|_2}, \dots, \frac{\mathbf{d}_m^*}{\|\mathbf{d}_m^*\|_2}\right) \quad (19)$$

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) = \left(\frac{\mathbf{w}_1^*}{\sqrt{\alpha}\|\mathbf{d}_1^*\|_2}, \frac{\mathbf{w}_2^*}{\sqrt{\alpha}\|\mathbf{d}_2^*\|_2}, \dots, \frac{\mathbf{w}_m^*}{\sqrt{\alpha}\|\mathbf{d}_m^*\|_2}\right) \quad (20)$$

### Classification scheme

To collaboratively represent a given test sample  $\mathbf{z}$  on normalized  $\mathbf{D}$  with low computational burden, we utilize the regularized least square method via equation (6). The solution of equation (6) can be analytically and easily derived by equation (7).

The classification of  $\mathbf{z}$  is based on its CR coefficient  $\mathbf{x}$ , which contains most discriminative information for classification. We can easily use the linear classifier  $\mathbf{W}$  with  $\mathbf{x}$  to derive the label of the test sample  $\mathbf{z}$

$$\text{Label}(\mathbf{z}) = \arg \max_i (\mathbf{W}\mathbf{x})_i, \quad i = 1, 2, \dots, K \quad (21)$$

The maximum  $i$ -th element of  $\mathbf{W}\mathbf{x}$  locates the class label  $i$  of  $\mathbf{z}$ .

### Flow of our algorithm

The DL is proposed for one-dimensional signal processing, while image-based classification is usually two-dimensional signals. Therefore, an image preprocessing is needed. First, we need to arrange the image data in a line as a vector. The transformation makes the original data being high-dimensional data. Thus, a dimensionality reduction

**Table 1.** The flow of discriminative collaborative representation classification algorithm.**Algorithm 1.** Discriminative collaborative representation classification (DCRC)Task: Derive the label of the test sample of  $z$ Input: A set of  $n$   $h$ -dimensional training samples from  $K$  classes

$$T = [T_1, \dots, T_i, \dots, T_K] \in \mathbb{R}^{h \times n}, \text{ and test sample } z$$

Initialize:

1. Use PCA to initialize  $T$  and get the low  $d$ -dimensional training

$$\text{set } Y = [Y_1, \dots, Y_i, \dots, Y_K] \in \mathbb{R}^{d \times n}$$

2. Initialize constant parameters:  $\alpha, \lambda, \lambda_1, \lambda_2$ 3. Initialize  $D_{(0)}$  using random matrix, then  $X_{(0)} = \frac{1}{1+\lambda_1} D_{(0)}^T Y$ 4. Initialize labels of training samples  $H_{(0)}$ 5. Initialize  $W_{(0)} = H_{(0)} X_{(0)}^T (X_{(0)} X_{(0)}^T + I)^{-1}$ 6. Initialize  $D_{(0)}^* = \begin{pmatrix} D_{(0)} \\ \sqrt{\alpha} W_{(0)} \end{pmatrix}$ 

Dictionary Learning Procedure:

While not converged

1. Updating  $X$  via Eq. (16)2. Updating  $D^*$  via Eq. (18)3. Normalizing  $D$  and  $W$  via Eq. (20) and Eq. (21)

end while

Output:  $D, W, X$ 

Classification Procedure:

1. CR coding via Eq. (7)

2. Classification via Eq. (21)

Output:  $Label(z)$ 

(DR) processing is needed, for high-dimensional data lead to the inefficiency of data processing. In our method, the basic DR method called principal component analysis (PCA)<sup>23</sup> is used, for it is simple and widely used. The procedures of DCRC are summarized in Table 1.

## Experiments for multimodal databases

To verify the validity of DCRC algorithm, we design two experiments with two databases: near-infrared database<sup>24</sup> and AR visible database.<sup>25</sup> We compared our algorithm with SRC,<sup>6</sup> CRC,<sup>19</sup> D-KSVD<sup>17</sup> algorithms on accurate recognition rate and time for classifying one test sample.

All the experiments were run on Matlab R2011a. The PC is Lenovo E40-80 notebook computer with an Intel Core i5 2.30 GHz CPU and 8 GB RAM. The GPUs are AMD Radeon(TM) R5 M330 and Intel(R) HD Graphics 5500 with 2 GB RAM. We also employed sparse solver SPAMS<sup>26</sup> to optimize a standard sparse representation.

### Experiment with near-infrared database

The near-infrared database contains 50 distinct subjects and 10 different infrared images for each one. Each image is  $100 \times 80$  pixels. Figure 1 shows some images of the near-infrared database.

With this database, we tested SRC, CRC, D-KSVD, and our method DCRC. The samples from the database were

average divided into two groups: one group was used as training sample and the other group was used as test sample. In order to reduce the calculation cost, we used PCA<sup>20</sup> to reduce the dimension of the sample vector from  $\mathbb{R}^{100 \times 80}$  to  $\mathbb{R}^{100}$ . For the SRC and CRC methods, we used all the training samples as the dictionary atoms (total 250 atoms). For the D-KSVD method, we used 150 atoms for the dictionary learning. For the DCRC method, we used a dictionary with 70 atoms, and with the parameters  $\alpha = 1$ ,  $\lambda = 0.001$ ,  $\lambda_1 = 0.001$ ,  $\lambda_2 = 0.1$ .

Table 1 shows that the proposed DCRC algorithm contains two procedures: DL and classification. In the dictionary training procedure, the representation reconstruction errors are quickly convergent as shown in Figure 2.

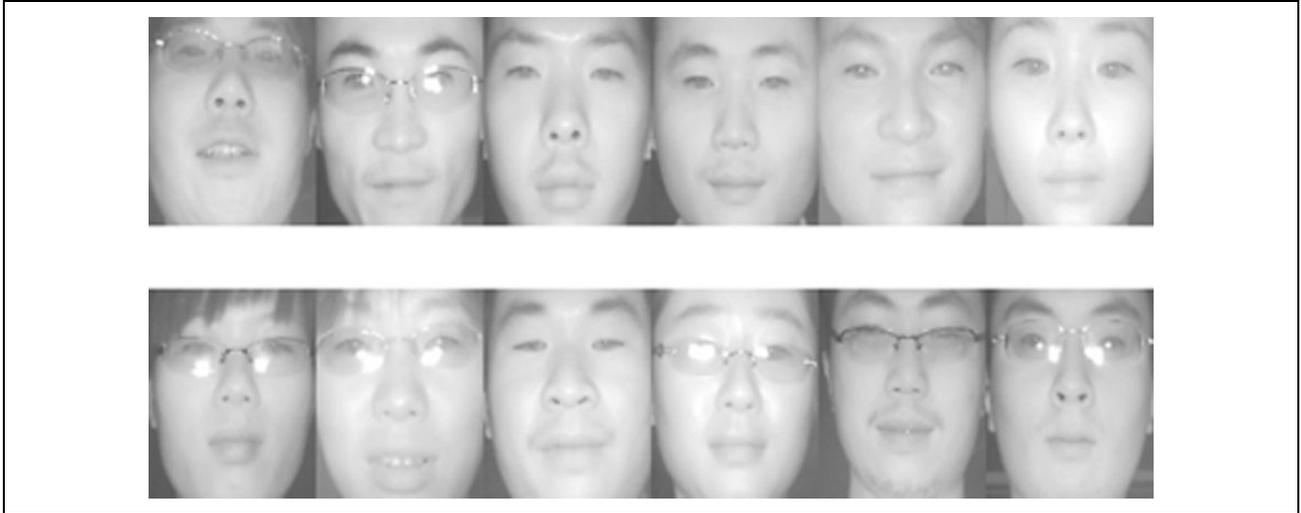
The effect on the accurate recognition rate versus dictionary size is shown in Figure 3. The result shows that the curve is boosting quickly with the increasing atoms forepart, and terminal recognition rates change smooth to 99%. Dictionary size is the main factor affecting the recognition rate. The parameters  $\lambda, \lambda_1, \lambda_2$  affect the recognition rate weakly. The parameter  $\lambda_1$  affects the sparsity, while the parameter  $\lambda_2$  affects the within-class errors. The coding coefficients of one test sample for the four methods are shown in Figure 4, and from top to bottom followed by the coefficient of SRC, CRC, D-KSVD, and DCRC. From the results, we can see that SRC and D-KSVD are strong sparse, while CRC and DCRC are weak sparse.

Compared with SRC, CRC, and D-KSVD, the proposed method DCRC has very competitive recognition rate but with significantly lower complexity. We recorded the time for classifying one test sample of the three methods, and the results were shown in Table 2.

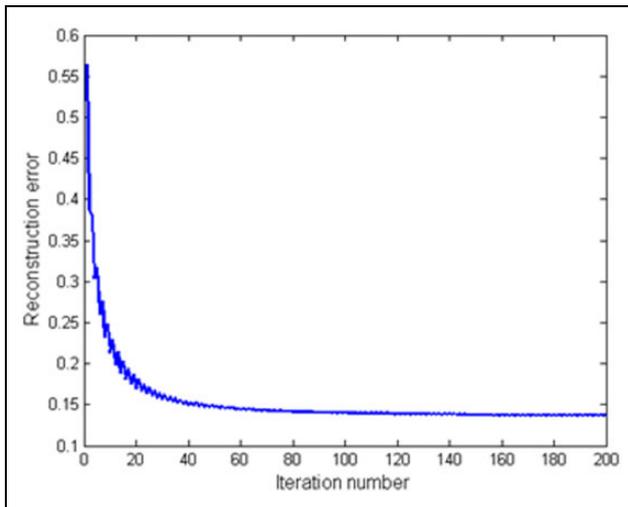
The results indicate that the proposed method DCRC has the approximate recognition rate with the SRC, CRC, and D-KSVD methods for this database. The recognition rates are near the same or almost the same as each other, maybe because the near-infrared database is without noise. SRC and CRC directly used the training sample as a dictionary, while D-KSVD and DCRC learned a compact dictionary from the training set. Therefore, the latter two methods are faster for object recognition, and DCRC is the fastest one of all. To validate our method being competitive for recognition accurate rate, we designed a new experiment in subsection "Experiment with AR database."

### Experiment with AR database

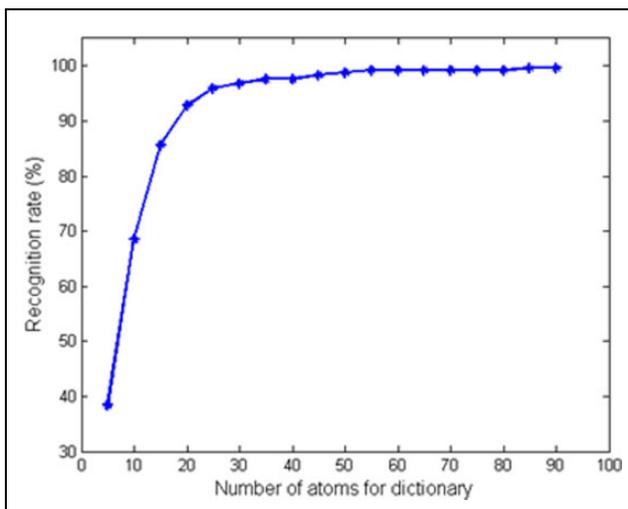
The AR database contains images for 126 persons, with 26 images for each one. Each image is  $165 \times 120$  pixels. This database is widely used for face recognition. Figure 5 shows some images of the AR database. We can see that these images are captured with different viewpoints, different illuminations, different facial expressions, and different disguises (sunglass and neckerchief). These interfering factors make it more difficult for face recognition. In order to



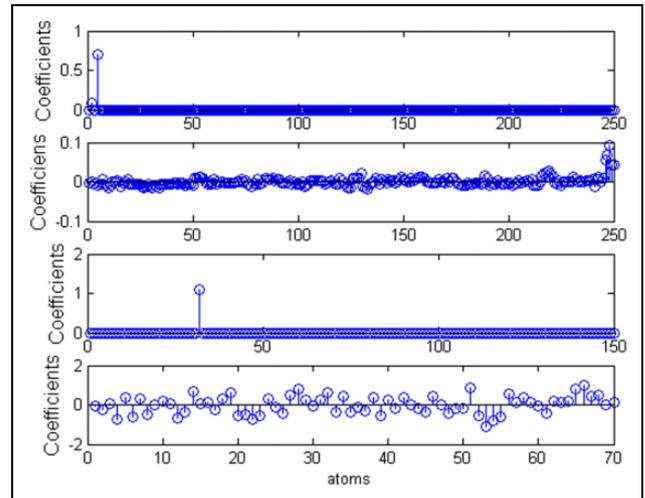
**Figure 1.** Samples from the near-infrared database.



**Figure 2.** The representative reconstruction error.



**Figure 3.** The accurate recognition rate versus dictionary size.



**Figure 4.** The coding coefficients of one test sample for SRC, CRC, D-KSVD, and DCRC. SRC: sparse representation-based classification; CRC: collaborative representation-based classification; D-KSVD: discriminative-KSVD; DCRC: discriminative collaborative representation-based classification.

validate our method is still effective for samples with noise, we designed an experiment with the AR database.

In our experiment, we chose 2600 images from 50 males and 50 females. For each person, there are 26 samples; we chose 20 samples for training and the other 6 samples for testing. First, we used PCA algorithm to reduce the character vector dimension from  $\mathbb{R}^{165 \times 120}$  to  $\mathbb{R}^{1000}$ . For the SRC and CRC methods, we still used all training samples as a dictionary (2000 atoms). For the DKSVD method, we used 1600 atoms for the dictionary learning. For the DCRC method, we used a dictionary with 100 atoms, and with the parameters  $\alpha = 1$ ,  $\lambda = 0.001$ ,  $\lambda_1 = 0.01$ ,  $\lambda_2 = 0.0001$ . The results are given in Table 3.

The results show that the recognition rates of the four methods decreased in different degrees compared with the

**Table 2.** Face recognition rate and time for classifying one test sample of different methods using near-infrared database.

Method	Recognition rate (%)	Time (ms)
SRC	98.8	0.9139
CRC	99.0	0.3928
D-KSVD	98.0	0.04858
DCRC	99.2	0.01918

SRC: sparse representation-based classification; CRC: collaborative representation-based classification; D-KSVD: discriminative-KSVD; DCRC: discriminative collaborative representation-based classification.



**Figure 5.** Samples from the AR database.

**Table 3.** Face recognition rate and time for classifying one test sample of different methods using AR visible database.

Method	Recognition rate (%)	Time (ms)
SRC	77.17	205.9
CRC	83.66	115.1
D-KSVD	98.0	1.300
DCRC	99.1	0.4965

SRC: sparse representation-based classification; CRC: collaborative representation-based classification; D-KSVD: discriminative-KSVD; DCRC: discriminative collaborative representation-based classification.

near-infrared results. The recognition rate of DCRC performs better than that of SRC, CRC, and D-KSVD. The occurrence of such results may be the interfering with noise. The D-KSVD and DCRC methods have the discriminative ability, while the SRC and CRC methods focus on representation of raw signals. Our method is robust for face recognition. The proposed method is the fastest one of the three methods.

## Conclusions

A fast discriminative CR image classification method DCRC was proposed in this article. It incorporated the

within-class scatter and the linear classification error terms into the objective function, so the method had the discriminative ability for classification. In order to decrease the computing time, we added CR mechanism to the objective function. The experimental results showed that DCRC had better recognition performance than the other three methods, and it was suitable for multimodal image (infrared or visible) classification. The calculating speed of our method was improved a lot.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China (grant no 61403398).

## References

1. Elad M and Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans Image Process* 2006; 15(12): 3736–3745.
2. Protter M and Elad M. Image sequence denoising via sparse and redundant representation. *IEEE Trans Image Process* 2009; 18(1): 27–35.
3. Mairal J, Bach F, Ponce J, et al. Discriminative learned dictionaries for local image analysis. In: *Proceedings of 2008 IEEE conference on computer vision and pattern recognition (CVPR)*, Anchorage, AK, USA, June 2008, pp. 1–8. IEEE.
4. Yang J, Wright J, Huang T, et al. Image super-resolution via sparse representation. *IEEE Trans Image Process* 2010; 19(11): 2861–2873.
5. Zhang Q and Li B. Discriminative K-SVD for dictionary learning in face recognition. In: *Proceedings of 2010 IEEE conference on computer vision and pattern recognition (CVPR)*, San Francisco, CA, USA, June 2010, pp. 2691–2698. IEEE.
6. Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 2009; 31(2): 210–227.
7. Kang C, Liao S, Xiang S, et al. Kernel sparse representation with local patterns for face recognition. In: *18th IEEE international conference on image processing*, Brussels, Belgium, September 2011, pp. 3009–3012. IEEE.
8. Yang M, Zhang L, Feng X, et al. Sparse representation based fisher discrimination dictionary learning for image classification. *Int J Comput Vis* 2014; 109(3): 209–232.
9. Zhao L, Sun Y, Hu Y, et al. Robust face recognition based  $l_{21}$ -norm sparse representation. In: *5th international conference on digital home*, Guangzhou, China, November 2014, pp. 25–29. IEEE.

10. Lai J and Jiang X. Class-wise sparse and collaborative patch representation for face recognition. *IEEE Trans Image Process* 2016; 25(7): 3261–3272.
11. Xu L, Wu X, Chen K, et al. Supervised within-class-similar discriminative dictionary learning for face recognition. *J Vis Commun Image Represent* 2016; 38: 561–572.
12. Liu H, Guo D, and Sun F. Object recognition using tactile measurements: kernel sparse coding methods. *IEEE Trans Instrum Meas* 2016; 65(3): 656–665.
13. Liu H, Liu Y, and Sun F. Traffic sign recognition using group sparse coding. *Inform Sci* 2014; 266: 75–89.
14. Aharon M, Elad M, and Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 2006; 54(11): 4311–4322.
15. Liu H, Liu Y, and Sun F. Robust exemplar extraction using structured sparse coding. *IEEE Trans Networks Learn Syst* 2015; 26(8): 1816–1821.
16. Pham DS and Venkatesh S. Joint learning and dictionary construction for pattern recognition. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Anchorage, AK, USA, June 2008, pp. 1–8. IEEE.
17. Jiang Z, Lin Z, and Davis LS. Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans Pattern Anal Mach Intell* 2013; 35(11): 2651–2664.
18. Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation. In: *IEEE international conference on computer vision (ICCV)*, Barcelona, Spain, November 2011, pp. 543–550. IEEE.
19. Zhang L, Yang M, and Feng X. Sparse representation or collaborative representation: which helps face recognition? In: *IEEE international conference on computer vision (ICCV)*, Barcelona, Spain, November 2011, pp. 471–478. IEEE.
20. Liu H, Sun F, Fang B, et al. Robotic room-level localization using multiple sets of sonar measurements. *IEEE Trans Instrum Meas* 2017; 66(1): 2–13.
21. Li F, Liu H, Xu X, et al. Multi-modal local receptive field extreme learning machine for object recognition. In: *International joint conference on neural networks (IJCNN)*, Vancouver, BC, Canada, July 2016, pp. 1696–1701. IEEE.
22. Zhang Z, Xu Y, Yang J, et al. A survey of sparse representation: algorithms and applications. *IEEE Access* 2015; 3: 490–530.
23. Turk M and Pentland A. Eigenfaces for recognition. *J Cognitive Neurosci* 1991; 3(1): 71–86.
24. Yong X, Qi Z, and David Z. Bimodal biometrics based on a representation and recognition approach. *Opt Eng* 2011; 50(3): 037202–037202.
25. Martinez A and Benavente R. *The AR face database*. CVC Tech. Report no. 24, 1998.
26. Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding. *J Mach Learn Res* 2010; 11(1): 19–60.