Neurocomputing 121 (2013) 248–253

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Letters

Two-dimensional relaxed representation

Qiulei Dong *

National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 8 October 2012
Received in revised form
1 March 2013
Accepted 26 April 2013
Communicated by Jinhui Tang
Available online 22 June 2013

Keywords:
Relaxed representation
Sparsity
Image classification

ABSTRACT

In this paper, a novel classification framework called two-dimensional relaxed representation (2DRR) is proposed for image classification. Different from recent popular vector-based representations with/without sparsity which encode a vector signal as a sparse/nonsparse linear combination of elementary vector signals, 2DRR is based on 2D image matrices, where each column of the input matrix signal is represented by a combination of the corresponding columns of the elementary matrices. In order to preserve the global linear coding relationship between the input matrix and these elementary matrices, the proposed 2DRR constrains the coding coefficients corresponding to each column of the input matrix to be locally close. Then two algorithms are derived from the 2DRR framework under the l_2 norm and the l_1 norm respectively. Extensive experimental results show the effectiveness of the proposed algorithms in comparison to three existing algorithms.

1. Introduction

Inspired by the finding in low-level human vision [1,2] that natural images can be generally coded by a sparse combination of structural primitives involving both additive and subtractive interactions, sparse representation has been investigated extensively in pattern recognition and image processing [3–8].

Sparse representation is a kind of representation that accounts for an unknown signal with a linear combination of a small number of elementary signals. Recently, there has been an increasing interest in the sparse representation classification problem. Huang and Aviyente [4] proposed a sparse-representation-based algorithm for signal classification which incorporated reconstruction properties, discrimination power and sparsity. Wright et al. [6] proposed a sparse representation classification (SRC) algorithm for face recognition, where they calculated the sparse representation of an input face image in terms of a set of training images, and performed classification by checking which class yields the least residual. And an extended version of the SRC algorithm was proposed to handle face misalignment in [9]. Yang et al. [10] explored an extended version of the spatial pyramid matching approach for image classification by generalizing vector quantization to sparse coding. Yang and Zhang [7] proposed a Gabor-feature based classification algorithm with a learned Gabor occlusion dictionary. Elhamifar and Vidal [3] formulated the classification problem as a structured sparse recovery problem using two non-convex optimization programs, and proposed convex relaxations for these two optimization programs.

It is worthy to point out that there exist arguments on whether or not the 'sparsity' helps classification recently [11–14]. As opposed to the SRC algorithm [6], Shi et al. [12] considered that the sparsity assumption is not supported by facial data, and proposed an l_2-norm-based classification algorithm (denoted as l_2C in this paper) for the face recognition problem. Furthermore, Zhang et al. [13] analyzed the working mechanism of SRC, and proposed a collaborative representation based classification algorithm with regularized least square (CRC_RLS), which can be considered as a regularization version of the l_2C algorithm [12]. Their experiments shown that the CRC_RLS algorithm has very competitive accuracy for face recognition but with significantly lower complexity compared with the SRC algorithm. Addressing Shi's criticisms [12] on SRC, Wright et al. [14] gave a discussion on the discrepancy between [6,12] within the context of a richer set of experimental results which demonstrate the advantage of the sparse representation.

Here, we do not give a direct evaluation on which is better between SRC and CRC_RLS (or l_2C), as well as whether it is necessary to impose the l_1-norm sparsity constraint on the coding coefficients or not, but propose a classification framework based on image matrices by relaxing the sparsity/nonsparsity constraint with respect to the training sample vectors [6,13,12] to the constraint with respect to the training sample matrices, called two-dimensional relaxed representation (2DRR). Then two algorithms are derived from the 2DRR framework under the l_2 norm and the l_1 norm respectively, which can be considered respectively as a generalized version of the SRC algorithm and a generalized version of the CRC_RLS algorithm.

The remainder of this work is organized as follows: Section 2 introduces related work. Section 3 formulates the 2DRR classification framework, and proposes two algorithms from this framework under the l_2 norm and the l_1 norm respectively. Extensive experimental

* Tel.: +86 10 62535412.
E-mail address: qldong@nlpr.ia.ac.cn

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.04.044

results are reported in Section 4. Finally, some concluding remarks are listed in Section 5.

2. Related work

Here is a brief review of the SRC algorithm [6] which uses the l_1-norm to regularize the coding coefficient vector, as well as the CRC_RLS algorithm [13], which uses the l_2-norm to regularize the coding coefficient vector.

For ease of reading, the main notations in the following sections are defined here: Let $Y=[y_1,y_2,\ldots,y_n]\in R^{m\times n}$ denote a test image of size $m\times n$, $y_i(i=1,2,\ldots,n)$ the i-th column of Y, $y=[y_1^T,y_2^T,\ldots,y_n^T]^T$ the column-wise vectorization of Y. Let $D=\{D_1,\ldots,D_i,\ldots,D_k\}$ denote a given training sample set consisting of k subsets, where k is the class number and $D_i(i=1,2,\ldots,k)$ is the subset containing the samples associated with the i-th class.

2.1. Sparse representation for classification

The SRC algorithm [6] employed sparse coding for face recognition based on the assumption that a test face image can be well represented by a small number of the training images. It consists of two main steps: First, the test image is coded sparsely in terms of all the training samples with a few non-zero coefficients since face recognition is a typical small-sample-size problem. Then the test sample is recognized according to the sparse reconstruction residuals corresponding to each class of the training samples.

More concretely, for a test image Y and a given set of training images D, they are all vectorized at first. Then, the sparse coding coefficient vector α is calculated by solving the following l_0-norm minimization problem:

$$\min_{\alpha}\|\alpha\|_0 \quad \text{s.t. } \|y-D\alpha\|_2^2\le\varepsilon \tag{1}$$

where y denotes the column-wise vectorization of Y, '$\|\cdot\|_0$' denotes the l_0-norm. The optimization for (1) is NP-hard, but Donoho [15] shown that this NP-hard problem can be solved by replacing l_0-norm with l_1-norm under some mild condition. Thus, we can obtain an approximate solution by solving the corresponding l_1 optimization problem

$$\min_{\alpha}\|\alpha\|_1 \quad \text{s.t. } \|y-D\alpha\|_2^2\le\varepsilon \tag{2}$$

Based on the Lagrange multiplier method, the Lagrangian formulation of (2) is usually adopted as

$$\hat{\alpha}=\arg\min_{\alpha}\|y-D\alpha\|_2^2+\lambda\|\alpha\|_1 \tag{3}$$

where λ is the tuning parameter. This optimization problem can be efficiently solved by many algorithms [16–18].

Then, y is recognized as the class which gives the minimum reconstruction residual

$$c_y=\arg\min_i\|D\delta_i(\hat{\alpha})-y\|_2 \tag{4}$$

where $\delta_i(\cdot)$ is an indicator function where the elements corresponding to the i-th class are preserved while the rest elements are set to be zero.

2.2. Collaborative representation for classification

Zhang et al. [13] proposed a collaborative representation for classification with regularized least square (CRC_RLS), which encodes an unknown vector sample y with respect to the training set D collaboratively. The coding coefficient vector α is calculated by solving the following regularized least square problem:

$$\hat{\alpha}=\arg\min_{\alpha}\|y-D\alpha\|_2^2+\lambda\|\alpha\|_2^2 \tag{5}$$

where λ is the tuning parameter.

Obviously, the solution to (5) is

$$\hat{\alpha}=(D^TD+\lambda I)^{\dagger}D^Ty \tag{6}$$

where '†' denotes the pseudo-inverse, I denotes the identity matrix.

Then, y is recognized as the class which gives the minimum regularized residual as

$$c_y=\arg\min_i\|D\delta_i(\hat{\alpha})-y\|_2/\|\delta_i(\hat{\alpha})\|_2 \tag{7}$$

It is necessary to point out that the l_2C algorithm [12] is highly similar to the CRC_RLS algorithm in essence. There are only two slight differences between them: (i) the l_2C algorithm only calculate the least squares solution to the first item in (5), but does not introduce any l_2-norm-regularizer, i.e. the designed representation in the l_2C algorithm is a special case of collaborative representation with $\lambda=0$. (ii) The basic reconstruction residual (4) is used as the classification criterion in the l_2C algorithm, while the regularized residual (7) is used in the CRC_RLS algorithm.

3. Two-dimensional relaxed representation

3.1. Motivation and formulation

According to the review of SRC and CRC_RLS in Section 2, it is noted that they are both vector-based algorithms, which represent an input vector with a combination of elementary vectors. And their difference is the norm of the coding coefficient vector: the l_1 norm is used in SRC for constraining the coefficient vector to be sparse, while the l_2 norm is used in CRC for making its solution stable and introducing a certain amount of 'sparsity' [13].

However, in many cases, a datum is in a matrix form, e.g. image data, and it may lose its spatial information to transform it into a vector. In addition, many applications in pattern recognition including face recognition are typical small-sample-size problems. Taking the face recognition as an example, even although these two algorithms have tackled such a small-sample-size problem by encoding the testing sample with respect to the dictionary consisting of all the training samples under the l_1-norm constraint and the l_2-norm constraint respectively as (3) and (5), their reconstruction residuals may be still large so that the consequent classification becomes unstable, especially when the number of the training samples is sufficiently smaller than the sample's dimension, which is a common phenomena in many real applications.

Addressing the above discussions, we propose a representation framework for encoding images more accurately based on image matrices, rather than vectors as in SRC and CRC_RLS, called 2D relaxed representation (2DRR). In the proposed 2DRR, a test image is coded column by column in terms of the corresponding columns of the training images in order to improve the representation accuracy, and simultaneously the coding coefficient vectors associated with neighboring columns of the test image are constrained to be locally close so that the global coding relationship between the test data and the training samples is preserved to a certain extent. More concretely, assuming that the training set is composed of N images of size $m\times n$, n dictionaries $H_i\in R^{m\times N}(i=1,\ldots,n)$ are constructed respectively by the i-th columns of all the training samples. $H_i(i=1,\ldots,n)$ is used to code the i-th column of the test sample, and then these obtained coding coefficient vectors corresponding to neighboring columns are required to be as close as possible. Therefore, the 2DRR framework is formulated as the following optimization problem:

$$\min_{\{\alpha_i\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^n \mathcal{R}_q(\alpha_i) + \eta \sum_{i=1}^n \sum_{j\in Ne(i)} \|\alpha_i-\alpha_j\|_2^2 W_{i,j}$$
$$\text{s.t. } \|Y-[H_1\alpha_1, H_2\alpha_2, \ldots, H_n\alpha_n]\|_2^2 \le \varepsilon \quad (8)$$

where $\mathcal{R}_q(\cdot)(q=1,2)$ is an l_q-norm-regularizer, i.e. $\mathcal{R}_1(\cdot)=\|\cdot\|_1$ and $\mathcal{R}_2(\cdot)=\|\cdot\|_2^2$, $Ne(i)$ denotes the p-neighborhood of i, W is a sparse symmetric weight matrix, where each element $W_{i,j}$ indicates the weight of the penalty item for the coding coefficients corresponding to the neighboring columns ($W_{i,j}$ can be set in the 0/1 manner or in the heat kernel manner).

Based on the Lagrange multiplier method, we can rewrite the formulation (8) of 2DRR as

$$\min_{\{\alpha_i\}_{i=1}^n} \|Y-[H_1\alpha_1, H_2\alpha_2, \ldots, H_n\alpha_n]\|_2^2 + \frac{\lambda}{n}\sum_{i=1}^n \mathcal{R}_q(\alpha_i)$$
$$+\eta \sum_{i=1}^n \sum_{j\in Ne(i)} \|\alpha_i-\alpha_j\|_2^2 W_{i,j} \quad (9)$$

According to the designed 2DRR, two algorithms are derived from the formulation (9) under the l_2 norm and the l_1 norm respectively, which are described in detail in the following two subsections.

3.2. 2DRRl_2 algorithm

When $q=2$, i.e. $\mathcal{R}_q(\alpha_i)=\|\alpha_i\|_2^2$ in (9), the original optimization problem becomes

$$\min_{\{\alpha_i\}_{i=1}^n} \|Y-[H_1\alpha_1, H_2\alpha_2, \ldots, H_n\alpha_n]\|_2^2 + \frac{\lambda}{n}\sum_{i=1}^n \|\alpha_i\|_2^2$$
$$+\eta \sum_{i=1}^n \sum_{j\in Ne(i)} \|\alpha_i-\alpha_j\|_2^2 W_{i,j} \quad (10)$$

Let $\alpha_v=[\alpha_1^T, \alpha_2^T, \ldots, \alpha_n^T]^T$, then (10) is reformulated as

$$\min_{\alpha_v} \|y-H\alpha_v\|_2^2 + \frac{\lambda}{n}\|\alpha_v\|_2^2 + 2\eta\alpha_v^T S\alpha_v \quad (11)$$

where

$$H=\begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_n \end{pmatrix}_{mN\times nN},$$

$S=(V-W)\otimes I_N$, V is a diagonal matrix with $V_{i,i}=\Sigma_j W_{i,j}$, I_N is the N-order identity matrix.

Obviously, (11) is a least-squares problem, whose least-squares solution is

$$\hat{\alpha}_v = \left(H^T H + \frac{\lambda}{n} I_{nN} + 2\eta S\right)^\dagger H^T y \quad (12)$$

where '†' denotes the pseudo-inverse.

Then similar to (7), the test image Y is recognized as the class which gives the minimum regularized residual according to the calculated coefficient vectors $\{\hat{\alpha}_i\}_{i=1}^n$

$$c = \arg\min_i \|[H_1\delta_i(\hat{\alpha}_1), H_2\delta_i(\hat{\alpha}_2), \ldots, H_n\delta_i(\hat{\alpha}_n)]-Y\|_2/\|\delta_i(\hat{\alpha}_v)\|_2 \quad (13)$$

The complete recognition procedure is summarized in Algorithm 1, denoted as 2DRRl_2.

Remark. When the number N of the training samples is too large, the sizes of the matrices H and S in (12) would be large accordingly and it would be of high computational cost to calculate the matrix inverse in (12). Therefore, in this case, the alternated least squares (ALSs) strategy is used to calculate the solution to (10), i.e. at each alternation step only α_i is updated by fixing the rest variables $\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n$.

Algorithm 1. 2DRRl_2.

Input: A training image set D for k classes, a test image Y
Output: Y's label c
1 Normalize the training images to have unit Frobenius norm;
2 Construct $\{H_i\}_{i=1}^n$ according to D;
3 Code Y over $\{H_i\}_{i=1}^n$ according to (12);
4 Compute the regularized residuals:
5 $e_i(Y)=\|[H_1\delta_i(\hat{\alpha}_1), H_2\delta_i(\hat{\alpha}_2), \ldots, H_n\delta_i(\hat{\alpha}_n)]-Y\|_2/\|\delta_i(\hat{\alpha}_v)\|_2$
$i=1,2,\ldots,n$
6 Identity of Y: $c=\arg\min_i(e_i(Y))$;

3.3. 2DRRl_1 algorithm

When $q=1$, i.e. $\mathcal{R}_q(\alpha_i)=\|\alpha_i\|_1$ in (9), the original optimization problem becomes

$$\min_{\{\alpha_i\}_{i=1}^n} \|Y-[H_1\alpha_1, H_2\alpha_2, \ldots, H_n\alpha_n]\|_2^2 + \frac{\lambda}{n}\sum_{i=1}^n \|\alpha_i\|_1$$
$$+\eta \sum_{i=1}^n \sum_{j\in Ne(i)} \|\alpha_i-\alpha_j\|_2^2 W_{i,j} \quad (14)$$

Similar to (10), (14) is reformulated as

$$\min_{\alpha_v} \|y-H\alpha_v\|_2^2 + \frac{\lambda}{n}\|\alpha_v\|_1 + 2\eta\alpha_v^T S\alpha_v \quad (15)$$

The second-order cone programs (SOCPs) [19] can be used to solve this l_1-regularized problem (15) effectively.

Then similar to (4), the test image Y is recognized as the class which gives the minimum reconstruction residual according to the calculated coefficient vectors $\{\hat{\alpha}_i\}_{i=1}^n$

$$c = \arg\min_i \|[H_1\delta_i(\hat{\alpha}_1), H_2\delta_i(\hat{\alpha}_2), \ldots, H_n\delta_i(\hat{\alpha}_n)]-Y\|_2 \quad (16)$$

The complete recognition procedure is summarized in Algorithm 2, denoted as 2DRRl_1.

Algorithm 2. 2DRRl_1.

Input: A training image set D for k classes, a test image Y
Output: Y's label c
1 Normalize the training images to have unit Frobenius norm;
2 Construct $\{H_i\}_{i=1}^n$ according to D;
3 Code Y over $\{H_i\}_{i=1}^n$ by solving the minimization problem (15);
4 Compute the residuals:
5 $e_i(Y)=\|[H_1\delta_i(\hat{\alpha}_1), H_2\delta_i(\hat{\alpha}_2), \ldots, H_n\delta_i(\hat{\alpha}_n)]-Y\|_2$
$i=1,2,\ldots,n$
6 Identity of Y: $c=\arg\min_i(e_i(Y))$;

3.4. Algorithmic analysis

3.4.1. Parameter selection

As seen from (9) that the introduced parameter η is used to tune the tradeoff between the reconstruction residual and the distribution of the coding coefficient vectors. When $\eta\to 0$, the constraint on the pairwise coding coefficient vectors corresponding to neighboring columns of the test image are weakened, and then the reconstruction residual becomes lower, but the coding vectors may be distributed more dispersedly due to the loss of the global coding relationship between the test image and the training images, consequently more columns of the test image may be represented mainly by a combination of the corresponding columns of interclass training images. When $\eta\to\infty$, the constraint on the pairwise coding coefficient vectors corresponding to neighboring columns of the test image are reinforced, but the reconstruction residual is prone to be larger, consequently the classification becomes unstable, especially when the number of the training samples is sufficiently smaller than the sample's dimension. In the extreme case when $\eta=\infty$, all the coding vectors are required to be equivalent, then the 2DRR under the l_1/l_2 norm is in essence equivalent to SRC/CRC_RLS.

In addition, as seen from (9), the parameter λ tunes the tradeoff between the reconstruction residual and the sparsity of the solution to (9). (The l_2-norm introduces a certain amount of sparsity to the solution although this sparsity is weaker than that by the l_1-norm [13].) With the increase of λ, the sparsity of the solution is reinforced, and the reconstruction residual increases so that the representation accuracy decreases. Therefore, in order to represent an input image effectively, λ is generally set to a relatively small/moderate value. When $\lambda=0$ and $\eta=0$, the 2DRR with only the first item in (9) is in essence equivalent to l_2C.

Summarizing the above discussions, learning these two tuning parameters is necessary for obtaining a good classification performance. Here, the leave-one-out cross-validation is used to determine these two parameters.

3.4.2. Relationship between 2DRRl_1/2DRRl_2 and SRC/CRC_RLS/l_2C

From the above discussions, it is noted that 2DRR is constructed by relaxing the constraint on the coding coefficient vector with respect to the training vector samples [6,13,12] to the constraint on a set of coding coefficient vectors in terms of the columns of the training matrix samples, which can be considered as a generalization of the sparse representation (SR) [6], the collaborative representation (CR) [13], and the l_2-norm-based representation [12], i.e. all these three representations are special cases of 2DRR under different constraints. More concretely, when $\eta=\infty$, 2DRRl_1/2DRRl_2 is in essence equivalent to SRC/CRC_RLS. When $\lambda=0$ and $\eta=\infty$, the 2DRRl_2 is equivalent to l_2C. It can be concluded further that, when 2DRRl_1(2DRRl_2) and SRC(CRC_RLS/l_2C) are used for classification with a given set of training samples, the lowest value of the misclassification rates computed by 2DRRl_1(2DRRl_2) with different combinations of λ and η should be no larger than the lowest value of the misclassification rates computed by SRC (CRC_RLS) with different choices of λ. Extensive experimental results in Table 4 confirm this conclusion.

3.4.3. Reconstruction residual analysis

Reconstruction residual can reflect the representation accuracy of an algorithm effectively. Here, the reconstruction residuals with respect to all the training samples by our algorithms, SRC, CRC_RLS, and l_2C are analyzed. According to the first item in (3) and (5), the reconstruction residual by the vector-based algorithms (SRC, CRC_RLS, and l_2C) is defined as

$$RV(\alpha)=\|y-D\alpha\|_2 \tag{17}$$

Similarly, according to the first item in (9), the reconstruction residual by our matrix-based algorithms is defined as:

$$RM(\{\alpha_i\}_{i=1}^n)=\|Y-[H_1\alpha_1,H_2\alpha_2,\ldots,H_n\alpha_n]\|_2 \tag{18}$$

Obviously, both $RV(\alpha)$ and $RM(\{\alpha_i\}_{i=1}^n)$ are used for measuring the difference between the input image and the corresponding reconstructed images under the l_2 norm. We introduce the following proposition on the reconstruction residuals by 2DRRl_2 and l_2C, and give a brief proof.

Proposition. *Let $\hat{\alpha}$ be the optimal solution to the minimization problem (5) with $\lambda=0$, and let $\{\hat{\alpha}_i\}_{i=1}^n$ be the optimal solution to the minimization problem (10) with $\lambda=0$. For an arbitrarily selected parameter η, $RM(\{\hat{\alpha}_i\}_{i=1}^n)\le RV(\hat{\alpha})$.*

Proof. Since $\{\hat{\alpha}_i\}_{i=1}^n$ is the optimal solution to the minimization problem (10) with $\lambda=0$, then

$$\begin{aligned}
&\|Y-[H_1\hat{\alpha}_1,H_2\hat{\alpha}_2,\ldots,H_n\hat{\alpha}_n]\|_2^2+\eta\sum_{i=1}^{n}\sum_{j\in Ne(i)}\|\hat{\alpha}_i-\hat{\alpha}_j\|_2^2\\
&\quad\le\|Y-[H_1\hat{\alpha},H_2\hat{\alpha},\ldots,H_n\hat{\alpha}]\|_2^2+\eta\sum_{i=1}^{n}\sum_{j\in Ne(i)}\|\hat{\alpha}-\hat{\alpha}\|_2^2\\
&\quad=\|Y-[H_1\hat{\alpha},H_2\hat{\alpha},\ldots,H_n\hat{\alpha}]\|_2^2=\|y-D\hat{\alpha}\|_2^2
\end{aligned} \tag{19}$$

Therefore, according to the definitions (17) and (18), we have

$$\begin{aligned}
&RM(\{\hat{\alpha}_i\}_{i=1}^n)^2\le\|Y-[H_1\hat{\alpha}_1,H_2\hat{\alpha}_2,\ldots,H_n\hat{\alpha}_n]\|_2^2+\eta\sum_{i=1}^{n}\sum_{j\in Ne(i)}\|\hat{\alpha}_i-\hat{\alpha}_j\|_2^2\\
&\le\|y-D\hat{\alpha}\|_2^2=RV(\hat{\alpha})^2\\
&\Longrightarrow RM(\{\hat{\alpha}_i\}_{i=1}^n)\le RV(\hat{\alpha})\quad\square
\end{aligned}$$

This proposition shows that the proposed 2DRRl_2 is able to represent an input image more accurately compared with l_2C. Furthermore, when λ is set to a small/moderate value but not zero, we speculate from the proof of the introduced proposition that the similar conclusion on the reconstruction residuals by 2DRRl_2(2DRRl_1) and CRC_RLS (SRC) is tenable, i.e. the computed reconstruction residual by 2DRRl_2 (2DRRl_1) is smaller than CRC_RLS(SRC) under a same small/moderate λ. Currently, we are not able to give a theoretical proof for the speculation yet. However, extensive experimental results in Section 4 show that the speculation holds true.

4. Experiments

The proposed 2DRRl_1 and 2DRRl_2 are all implemented on a Core 2 Duo 2.53 GHz PC. To further evaluate the proposed algorithms, they are compared with three state-of-the-art algorithms, SRC [6], l_2C [12], and CRC_RLS [13].

4.1. Databases

All these algorithms are tested on the following three face databases where all the images are normalized to have unit Frobenius norm:

1. Yale database [20]: The Yale database consists of 165 gray images of 15 persons. The images contain variations in lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised and wink), and with/without glasses. The original face images in Yale database are aligned by fixing the locations of two eyes first, and then they are cropped and resized to 32×32 pixels.
2. ORL database [21]: The ORL database contains 400 images of 40 individuals. The images are captured at different times and with different variations including expression (open or closed eyes, smiling or non-smiling) and facial details (glasses or no-glasses). Each image is manually cropped and normalized to the size of 32×32 pixels.
3. AR database: A subset which contains 50 male subjects and 50 female subjects, is chosen from the AR database [22] in this experiments. For each subject, 14 images with only illumination change and expressions are selected, and all the images are cropped to 60×43 pixels.

4.2. Numerical results

At first, the images in the Yale database are used to test the influence of the parameter p which is used to describe the size of the neighborhood $Ne(\cdot)$ in (8), and the weight matrix W is defined in the 0/1 manner. A random subset with $m(=3, 4, 5)$ images for each person in the Yale database is selected for training and the rest for testing. For each given m, both 2DRRl_1 with $p=2, 4, 6, 8, 10$ and 2DRRl_2 with $p=2, 4, 6, 8, 10$ are performed independently 10 times, and the corresponding mean value of the error rates (MVERs) and the standard deviation of the error rates (SDERs) are shown respectively in Tables 1 and 2. It can be noted from these two tables that the computed MVER by 2DRRl_1/2DRRl_2 varies slightly with p, probably because p controls the size of the neighborhood of each coding coefficient vector and indirectly influences the closeness between the coding coefficient vectors associated with neighboring columns of a test image to some extent. In the above experiments, 2DRRl_1 with $p=4$ achieves comparably good performance in most cases, while 2DRRl_2 with $p=6$ achieves comparably good performance in most cases. Therefore, in the following experiments, the parameter p for 2DRRl_1 is set to 4, and the parameter p for 2DRRl_2 is set to 6.

To further test the performances of all these algorithms, in each of the referred databases in Section 4.1, a random subset with $m(=3, 4, 5)$ images for each person is selected for training and the rest for testing. For each given m, all the algorithms are performed independently 10 times, and the MVER and the SDER in these databases are shown respectively in Tables 3–5.

In addition, to compare the reconstruction accuracy of all these algorithms, for each given $m(=3, 4, 5)$, Table 6 lists respectively the mean value of the reconstruction residuals (MVRRs) and the standard deviation of the reconstruction residuals (SDRRs) according to (17) and (18) in the Yale database by all these algorithms, and Fig. 1 shows a set of the corresponding reconstructed images.

Here are some points revealed in Tables 3–6, and Fig. 1:

1. As seen from Tables 3–5, in all the referred databases, 2DRRl_2 performs better than CRC_RLS and l_2C, while 2DRRl_1 performs better than SRC, which confirms the conclusion in Section 3.4.2.
2. Compared with 2DRRl_1, the error rates computed by 2DRRl_2 are lower in the Yale database and the AR database, but higher in the ORL database. Compared with SRC, the error rates computed by CRC_RLS are higher in the Yale database, but quite close in the ORL database and the AR database. The above observations show that the sparsity assumption cannot necessarily guarantee improving the classification accuracy for the face recognition problem which is a typical small-sample-size problem. In addition, it is noted that both CRC_RLS and SRC outperform l_2C, probably due to the fact that the introduced regularizers on the coding coefficient vectors in both CRC_RLS and SRC make their solutions more stable.
3. As seen from Table 6 and Fig. 1, the reconstruction residuals computed by 2DRRl_2 and 2DRRl_1 are lower than those by the other referred algorithms, and the corresponding recovered images by 2DRRl_2 and 2DRRl_1 have better reconstruction quality than those by these referred algorithms, probably because the proposed 2DRR is based on 2D image matrices and codes an image column by column in terms of the corresponding columns of the training images, which further confirms our speculation in Section 3.4.3.

Table 1
MVER(SDER) comparison with different values of p by 2DRRl_1 in the Yale database.

m	MVER%(SDER)				
	$p=2$	$p=4$	$p=6$	$p=8$	$p=10$
3-Train	27.5(2.36)	**27.1(2.62)**	28.1(3.10)	28.8(2.28)	29.9(1.78)
4-Train	22.9(2.56)	22.6(1.91)	**22.4(3.52)**	23.3(2.96)	24.3(2.96)
5-Train	19.7(3.67)	**18.7(2.07)**	19.4(2.39)	21.1(3.01)	21.7(2.79)

Table 2
MVER(SDER) comparison with different values of p by 2DRRl_2 in the Yale database.

m	MVER%(SDER)				
	$p=2$	$p=4$	$p=6$	$p=8$	$p=10$
3-Train	27.5(2.79)	27.3(2.84)	**26.7(2.36)**	27.2(2.56)	27.4(2.95)
4-Train	22.2(2.60)	21.7(2.44)	**21.4(3.34)**	**21.4(2.74)**	21.6(1.56)
5-Train	18.9(2.36)	18.6(3.06)	**18.1(3.19)**	18.6(3.56)	18.3(3.45)

Table 3
MVER(SDER) comparison on the Yale database.

Methods	MVER%(SDER)		
	3-Train	4-Train	5-Train
SRC	30.7(0.63)	27.1(5.41)	23.8(3.90)
2DRRl_1	27.1(2.62)	22.6(1.91)	18.7(2.07)
l_2C	36.4(2.82)	32.1(4.84)	26.9(6.20)
CRC_RLS	36.3(2.67)	31.6(4.40)	26.9(6.11)
2DRRl_2	26.7(2.36)	21.4(3.34)	18.1(3.19)

Table 4
MVER(SDER) comparison on the ORL database.

Methods	MVER%(SDER)		
	3-Train	4-Train	5-Train
SRC	14.2(0.93)	10.0(2.04)	6.6(1.88)
2DRRl_1	11.3(1.58)	7.7(2.27)	4.2(0.84)
l_2C	16.2(0.90)	13.8(3.08)	8.9(2.07)
CRC_RLS	14.6(1.21)	11.1(2.22)	6.8(0.27)
2DRRl_2	13.4(1.46)	9.9(2.36)	5.5(1.22)

Table 5
MVER(SDER) comparison on the AR database.

Methods	MVER%(SDER)		
	3-Train	4-Train	5-Train
SRC	18.1(5.64)	11.8(3.96)	6.6(1.94)
2DRRl_1	16.0(4.10)	9.5(2.74)	6.2(0.62)
l_2C	26.1(7.82)	19.4(6.16)	15.0(3.77)
CRC_RLS	18.3(6.16)	11.0(4.28)	6.1(2.17)
2DRRl_2	13.1(4.95)	7.6(3.78)	5.2(0.81)

Table 6
Reconstruction residual comparison in the Yale database.

Methods	MVRR(SDRR)		
	3-Train	4-Train	5-Train
SRC	0.2491(0.0563)	0.2230(0.0503)	0.2105(0.0514)
2DRRl_1	0.2166(0.0456)	0.2080(0.0372)	0.1850(0.0336)
l_2C	0.2486(0.0564)	0.2223(0.0505)	0.2094(0.0516)
CRC_RLS	0.2494(0.0566)	0.2229(0.0505)	0.2103(0.0517)
2DRRl_2	0.2104(0.0410)	0.1998(0.0390)	0.1826(0.0381)

5. Conclusions

In this paper, we propose a novel classification framework based on 2D image matrices, called two-dimensional relaxed representation (2DRR). In the 2DRR framework, the input matrix is coded

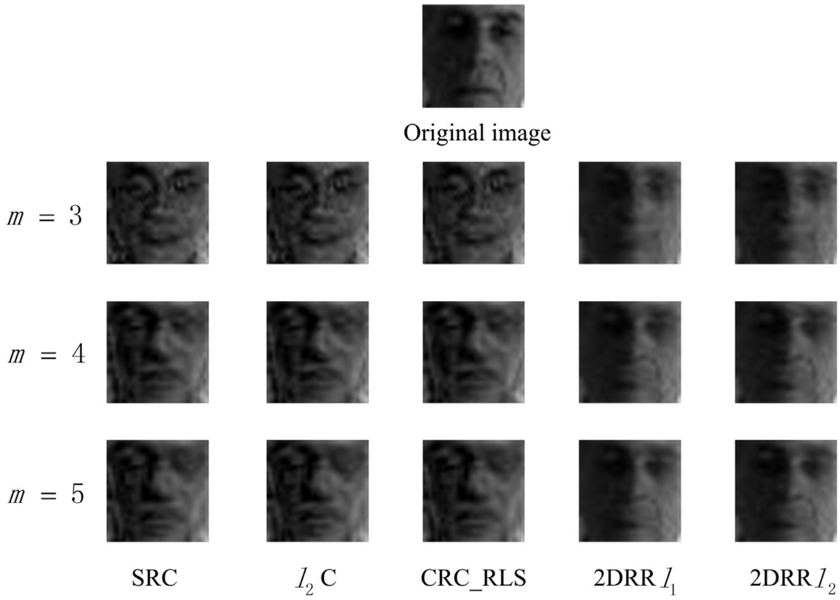

Fig. 1. Reconstructed image examples in the Yale database.

column by column with respect to the corresponding columns of the elementary matrices, and simultaneously the coding coefficients corresponding to each column of the input matrix are constrained to be locally close so that the global linear relationship between the input matrix and these elementary matrices is preserved to some extent. Then two algorithms are derived from the 2DRR framework under the l_2 norm and the l_1 norm respectively.

The advantages of the proposed 2DRR are: (1) It can represent a matrix signal more accurately compared with these vector-representations including sparse representation and collaborative representation. (2) 2DRRl_2 and 2DRRl_1 derived from the 2DRR framework have better performances for face recognition than these vector-representation-based classification algorithms.

Acknowledgment

This work was supported by National Natural Science Foundation of China under Grant nos. (60905020, 61173098).

References

[1] B. Olshausen, D. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.
[2] B.A. Olshausen, D.J. Fieldt, Sparse coding with an overcomplete basis set: a strategy employed by v1, Vision Res. 37 (1997) 3311–3325.
[3] E. Elhamifar, R. Vidal, Robust classification using structured sparse representation, in: CVPR, 2011, pp. 1873–1879.
[4] K. Huang, S. Aviyente, Sparse representation for signal classification, in: NIPS, 2006, pp. 609–616.
[5] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration (2007) 53–69.
[6] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[7] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with gabor occlusion dictionary, in: ECCV (6), 2010, pp. 448–461.
[8] H. Zhang, J. Yang, Y. Zhang, N.M. Nasrabadi, T.S. Huang, Close the loop: joint blind image restoration and recognition with sparse representation prior, in: ICCV, 2011, pp. 770–777.
[9] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, Y. Ma, Towards a practical face recognition system: robust registration and illumination by sparse representation, in: CVPR, 2009, pp. 597–604.
[10] J. Yang, K. Yu, Y. Gong, T.S. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: CVPR, 2009, pp. 1794–1801.
[11] R. Rigamonti, M.A. Brown, V. Lepetit, Are sparse representations really relevant for image classification? in: CVPR, 2011, pp. 1545–1552.
[12] Q. Shi, A. Eriksson, A. van den Hengel, C. Shen, Is face recognition really a compressive sensing problem? in: CVPR, 2011, pp. 553–560.
[13] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: ICCV, 2011, pp. 471–478.
[14] J. Wright, A. Ganesh, A. Yang, Z. Zhou, Y. Ma, Sparsity and robustness in face recognition, arXiv:1111.1014.
[15] D.L. Donoho, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution 59 (2004) 797–829.
[16] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[17] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, A method for large-scale $l1$-regularized least squares, IEEE J. Sel. Top. Signal Process. 1 (4) (2007) 606–617.
[18] J. Yang, Y. Zhang, Alternating direction algorithms for l1-problems in compressive sensing, SIAM J. Sci. Comput. 33 (1–2) (2011) 250–278.
[19] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[20] Yale university face database, ⟨http://cvc.yale.edu/projects/yalefaces/yalefaces.html⟩, 2002.
[21] Olivetti-Research-Laboratory, The olivetti & oracle research laboratory face database, ⟨http://www.cam-orl.co.uk/facedatabase.html⟩, 1994.
[22] A. Martinez, R. Benavente, The AR face database, CVC Technical Report No. 24.

Qiulei Dong received the B.S. degree in automation from Northeastern University, Shenyang, China, in 2003, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2008.

He is currently an associate professor with the Institute of Automation, Chinese Academy of Sciences. His research interest covers pattern classification and 3D computer vision.