Pattern Recognition 47 (2014) 2794–2806

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning locality-constrained collaborative representation for robust face recognition

Xi Peng [a], Lei Zhang [a,*], Zhang Yi [a], Kok Kiong Tan [b]

[a] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China
[b] Department of Electrical and Computer Engineering, National University of Singapore, Engineering Drive 3, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 23 October 2012
Received in revised form
6 September 2013
Accepted 18 March 2014
Available online 27 March 2014

Keywords:
Non-sparse representation
Sparse representation
Local consistency
ℓ_2-minimization
Partial occlusions
Additive noise
Non-additive noise
Robustness

ABSTRACT

The models of low-dimensional manifold and sparse representation are two well-known concise models that suggest that each data can be described by a few characteristics. Manifold learning is usually investigated for dimension reduction by preserving some expected local geometric structures from the original space into a low-dimensional one. The structures are generally determined by using pairwise distance, e.g., Euclidean distance. Alternatively, sparse representation denotes a data point as a linear combination of the points from the same subspace. In practical applications, however, the nearby points in terms of pairwise distance may not belong to the same subspace, and vice versa. Consequently, it is interesting and important to explore how to get a better representation by integrating these two models together. To this end, this paper proposes a novel coding algorithm, called Locality-Constrained Collaborative Representation (LCCR), which introduce a kind of local consistency into coding scheme to improve the discrimination of the representation. The locality term derives from a biologic observation that the similar inputs have similar codes. The objective function of LCCR has an analytical solution, and it does not involve local minima. The empirical studies based on several popular facial databases show that LCCR is promising in recognizing human faces with varying pose, expression and illumination, as well as various corruptions and occlusions.

1. Introduction

Sparse representation has become a powerful method to address problems in pattern recognition and computer version, which assumes that each data point $\mathbf{x} \in \mathbb{R}^m$ can be encoded as a linear combination of other points. In mathematically, $\mathbf{x} = \mathbf{Da}$, where $\mathbf{D}$ is a dictionary whose columns consist of some data points, and $\mathbf{a}$ is the representation of $\mathbf{x}$ over $\mathbf{D}$. If most entries of $\mathbf{a}$ are zeros, then $\mathbf{a}$ is called a sparse representation. Generally, it can be achieved by solving

$$(P_0): \ \min\|\mathbf{a}\|_0 \quad \text{s.t. } \mathbf{x} = \mathbf{Da},$$

where $\|\cdot\|_0$ denotes ℓ_0-norm by counting the number of nonzero entries in a vector. P_0 is difficult to solve since it is a NP-hard problem. Recently, compressive sensing theory [1,2] have found that the solution of P_0 is equivalent to that of ℓ_1-minimization problem $(P_{1,1})$ when $\mathbf{a}$ is highly sparse.

$$(P_{1,1}): \ \min\|\mathbf{a}\|_1 \quad \text{s.t. } \mathbf{x} = \mathbf{Da},$$

where ℓ_1-norm $\|\cdot\|_1$ sums the absolute value of all entries in a vector. $P_{1,1}$ is convex and can be solved by a large amount of convex optimization methods, such as basis pursuit (BP) [3], least angle regression (LARS) [4]. In [5], Yang et al. make a comprehensive survey for some popular optimizers.

Benefiting from the emergence of compressed sensing theory, sparse coding has been widely used for various tasks, e.g., subspace learning [6], spectral clustering [7,8] and matrix factorization [9]. In these works, Wright et al. [10] reported a remarkable method that passes sparse representation through a nearest feature subspace classifier, named sparse representation-based classification (SRC). SRC has achieved attractive performance in robust face recognition and has motivated a large amount of works such as [11–13]. The work implies that sparse representation plays an important role in face recognition under the framework of nearest subspace classification [14].

* Corresponding author.
E-mail addresses: pangsaai@gmail.com (X. Peng), leizhang@scu.edu.cn (L. Zhang), zhangyi@scu.edu.cn (Z. Yi), kktan@nus.edu.sg (K.K. Tan).

http://dx.doi.org/10.1016/j.patcog.2014.03.013
0031-3203/

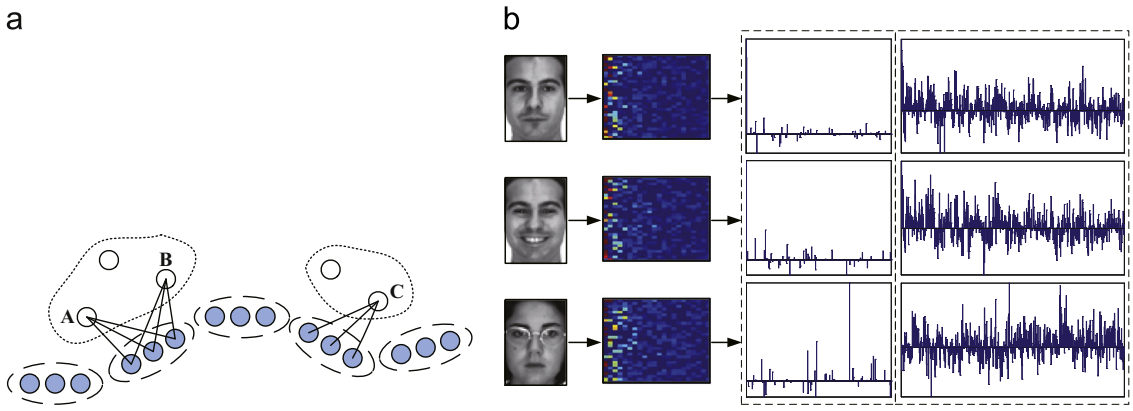


Fig. 1. A key observation. (a) Three face images from two different sub-manifolds are linked to their corresponding neighbors, respectively. (b) The first column includes three images which correspond to the points *A*, *B* and *C* in (a). The second column shows the Eigenface feature [31] matrices for the testing images; The third column includes two parts: the left part is the coefficients of SRC [10], and the right one is of CRC-RLS [18]. From the results, we could see that the representations of nearby points are more similar than that of non-neighboring points, i.e., local consistency could be defined as the similar inputs have similar codes.

However, is ℓ_1-norm-based sparsity really necessary to improve the performance of face recognition? Several recent works directly or indirectly examined this problem. Yang et al. [15] discussed the connections and differences between ℓ_1-optimizer and ℓ_0-optimizer for SRC. They show that the success of SRC should attribute to the mechanism of ℓ_1-optimizer which selects the set of support training samples for the given testing sample by minimizing reconstruction error. Consequently, Yang et al. pointed out that the global similarity derived from ℓ_1-optimizer but sparsity derived from ℓ_0-optimizer is more critical for pattern recognition. Rigamonti et al. [16] compared the discrimination of two different data models. One is the ℓ_1-norm-based sparse representation, and the other model is produced by passing input into a simple convolution filter. Their result showed that two models achieve a similar recognition rate. Therefore, ℓ_1-norm-based sparsity is actually not as essential as it seems in the previous claims. Shi et al. [17] provided a more intuitive approach to investigate this problem by removing the ℓ_1-regularization term from the objective function of SRC. Their experimental results showed that their method achieves a higher recognition rate than SRC if the original data is available. Zhang et al. [18] replaced the ℓ_1-norm by the ℓ_2-norm, and their experimental results again support the views that ℓ_1-norm-based sparsity is not necessary to improve the discrimination of data representation. Moreover, we have noted that Naseem et al. [19,20] proposed Linear Regression Classifier (LRC) which has the same objective function with Shi et al.'s work. The difference is that Shi et al. aimed to explore the role of sparsity while Naseem et al. focused on developing an effective classifier for face recognition.

As another extensively studied concise model, manifold learning (locality preservation model) is usually investigated for dimension reduction by learning and embedding local consistency of original data into a low-dimensional representation [21–23]. Local consistency means that nearby data points share the same properties, which is hardly reflected in linear representation.

Recently, some researchers have explored the possibility of integrating the locality (local consistency) with the sparsity together to produce a better data model. Baraniuk and Wakin [24] successfully bridged the connections between sparse coding and manifold learning, and have founded the theory for random projections of smooth manifold; Majumdar and Ward [25] investigated the effectiveness and robustness of random projection method in classification task. Moreover, Wang et al. [26] proposed a hierarchal images classification method named locality-constrained linear coding (LLC) by introducing dictionary learning into Locally Linear Embedding [27]. Chao et al. [28] presented an approach to unify group sparsity and data locality by introducing the term of ridge regression into LLC; Yang et al. [29] incorporated the prior knowledge into the coding process by iteratively learning a weight matrix of which the entries denotes the similarity between two data points.

In this paper, we proposed and formulated a new kind of local consistency into the linear coding paradigm by enforcing *the similar inputs* (*i.e.*, *neighbors*) *produce similar codes*. The idea is motivated by an observation in biological founds [30] which shows that L2/3 of rat visual cortex activates the same collection of neurons in response to leftward and rightward drifting gratings. Fig. 1 shows an example to illustrate the motivation. There are three face images *A*, *B* and *C* selected from two different individuals, where *A* and *B* came from the same person. This means that *A* and *B* lie on the same subspace and could represent with each other. Fig. 1(b) is a real example corresponding to Fig. 1(a). Either from the Eigenface [31] matrices or the coefficients of the two coding schemes, we can see that the similarity between *A* and *B* is much higher than the similarity between *C* and either of them.

Based on the observation, we proposed a representation learning method for robust face recognition, named Locality-Constrained Collaborative Representation (LCCR). The algorithm obtains a representation for each data point by enforcing the codes of neighboring points are as similar as possible. Furthermore, the objective function of LCCR has an analytic solution, does not involve local minima. Extensive experiments show that LCCR outperforms SRC [10], LRC [17,19], CRC-RLS [18], CESR [13], LPP [32], and linear SVM with Eigenface [31] in face recognition.

Except in some specified cases, lower-case bold letters represent column vectors and upper-case bold ones represent matrices, $\mathbf{A}^T$ denotes the transpose of the matrix $\mathbf{A}$, $\mathbf{A}^{-1}$ represents the pseudo-inverse of $\mathbf{A}$, and $\mathbf{I}$ is reserved for identity matrix.

The remainder of paper is organized as follows: Section 2 introduces three related approaches for face recognition based on data representation, i.e., SRC [10], LRC [17,19] and CRC-RLS [18]. Section 3 presents our LCCR algorithm. Section 4 reports the experiments on several facial databases. Finally, Section 5 contains the conclusion.

2. Preliminaries

We consider a set of N facial images collected from L subjects. Each training image is denoted as a vector $\mathbf{d}_i \in \mathbb{R}^M$ corresponding to

the ith column of the dictionary $\mathbf{D} \in \mathbb{R}^{M \times N}$. Without generality, we assume that the columns of $\mathbf{D}$ are sorted according to their labels.

2.1. Sparse representation-based classification

Sparse coding aims at finding the most sparse solution of $P_{1,1}$. However, in many practical problems, the constraint $\mathbf{x} = \mathbf{Da}$ cannot hold exactly since the input $\mathbf{x}$ may include noise. Wright et al. [10] proposed Sparse Representation-based Classifier (SRC) which relaxes the constraint to $\|\mathbf{x}-\mathbf{Da}\|_2 \le \varepsilon$, where $\varepsilon > 0$ is the error tolerance, then, $P_{1,1}$ is rewritten as

$$(P_{1,2}): \ \min \|\mathbf{a}\|_1 \quad \text{s.t. } \|\mathbf{x}-\mathbf{Da}\|_2 \le \varepsilon.$$

Using Lagrangian method, $P_{1,2}$ can be transformed to the following unconstrained optimization problem:

$$(P_{1,3}): \ \underset{\mathbf{a}}{\operatorname{argmin}} \ \|\mathbf{x}-\mathbf{Da}\|_2^2 + \lambda\|\mathbf{a}\|_1,$$

where the scalar $\lambda \ge 0$ balances the importance between the reconstruction error and the sparsity. Given a testing sample $\mathbf{x} \in \mathbb{R}^M$, its sparse representation $\mathbf{a}^* \in \mathbb{R}^N$ can be computed by solving $P_{1,2}$ or $P_{1,3}$.

After getting the sparse representation of $\mathbf{x}$, SRC infers the label of $\mathbf{x}$ by assigning it to the class that has the minimum residual:

$$r_i(\mathbf{x}) = \|\mathbf{x}-\mathbf{D}\delta_i(\mathbf{a}^*)\|_2, \tag{1}$$

$$\text{identity}(\mathbf{x}) = \underset{i}{\operatorname{argmin}}\{r_i(\mathbf{x})\}. \tag{2}$$

where the nonzero entries of $\delta_i(\mathbf{a}^*) \in \mathbb{R}^N$ are the entries in $\mathbf{a}^*$ that are associated with ith class, and identity($\mathbf{x}$) denotes the label for $\mathbf{x}$.

2.2. ℓ_2-minimization-based methods

In [19], Naseem et al. proposed a Linear Regression Classifier (LRC) which achieved comparable accuracy to SRC in the context of robust face recognition. In another independent work [17], Shi et al. used the same objective function with that of LRC to discuss the role of ℓ_1-regularization-based sparsity. The objective function used in [17,19] is

$$\underset{\mathbf{a}}{\operatorname{argmin}} \ \|\mathbf{x}-\mathbf{Da}\|_2^2.$$

In [17], Shi et al. empirically showed that their method (denoted as LRC in this paper for convenience) requires $\mathbf{D}$ to be an over-determined matrix for achieving competitive results, while the dictionary $\mathbf{D}$ of SRC must be under-determined according to compressive sensing theory. Once the optimal code $\mathbf{a}^*$ is calculated for a given input, the classifiers (1) and (2) are used to determine the label for the input $\mathbf{x}$.

As another recently proposed ℓ_2-norm-based model, CRC-RLS [18] estimates the representation of the input $\mathbf{x}$ by relaxing the ℓ_1-norm to the ℓ_2-norm in $P_{1,3}$. The objective function is as follows:

$$\underset{\mathbf{a}}{\operatorname{argmin}} \ \|\mathbf{x}-\mathbf{Da}\|_2^2 + \lambda\|\mathbf{a}\|_2^2,$$

where $\lambda > 0$ is a balance factor.

LRC and CRC-RLS show that ℓ_2-norm-based models can achieve competitive classification accuracy with hundreds of times speed increase, compared with SRC. Under this background, we aim to incorporate the local geometric structures into coding process for achieving better discrimination and robustness to corruptions and occlusions.

3. Locality-constrained collaborative representation

It is challenging to improve the discrimination and the robustness of facial representation because a practical face recognition system requires not only a high recognition rate but also the robustness against various corruptions and occlusions.

3.1. Algorithm description

Locality preservation-based algorithm (LPA) and sparse representation (SR) have been extensively studied and successfully applied to appearance-based face recognition. LPA aims to find a low-dimensional model by learning and preserving some properties shared by nearby points from the original space to another one. Alternatively, SR encodes each testing sample as a linear combination of the training data, which actually depicts a global relationship between testing sample and training ones. In this paper, we aim to propose and formulate a kind of local consistency into coding scheme for modeling facial data. Our objective function has the following form:

$$E(\mathbf{a}) = \|\mathbf{x}-\mathbf{Da}\|_2^2 + \lambda\|\mathbf{a}\|_p + \gamma E_{\mathrm{L}}, \tag{3}$$

where $p = \{1, 2\}$, E_{L} is the locality constraint, $\lambda \ge 0$ and $\gamma \ge 0$ dictate the importance of $\|\cdot\|_p$ and E_{L}, respectively. Then the key is to formulate the shared property of the neighborhood with E_{L}.

Motivated by the biological experiment of Ohki et al. [30] as discussed in Section 1 and the key observation in Fig. 1, we have the following inequality:

$$\|\mathbf{y}_i - \mathbf{Dc}_i^*\|_2^2 \le \|\mathbf{y}_i - \mathbf{Da}^*\|_2^2 \le \|\mathbf{y}_i - \mathbf{D}\overline{\mathbf{a}}^*\|_2^2,$$

where $\mathbf{y}_i$ is a training sample which falls into the local neighborhood of the input $\mathbf{x}$, and $\mathbf{a}^*$ and $\mathbf{c}_i^*$ denote the representation of $\mathbf{y}_i$ and $\mathbf{x}$, respectively. Moreover, $\overline{a}^*$ is the optimal code of a data point that is not close to $\mathbf{y}_i$. This inequality shows that it is better to represent $\mathbf{y}_i$ using the code of nearby point than using that of non-nearby point. In other words, similar inputs should have similar codes.

Thus, E_{L} is defined as follows:

$$E_{\mathrm{L}} = \frac{1}{K} \sum_{\mathbf{y}_i \in \mathbf{Y}(\mathbf{x})} \|\mathbf{y}_i - \mathbf{Da}\|_2^2, \tag{4}$$

which establishes the relationship between $\mathbf{a}$ and $\mathbf{y}_i$. Replacing (4) into (3), it is given by

$$\underset{\mathbf{c}}{\operatorname{argmin}} \ (1-\gamma)\|\mathbf{x}-\mathbf{Da}\|_2^2 + \gamma\frac{1}{K} \sum_{\mathbf{y}_i \in \mathbf{Y}(\mathbf{x})} \|\mathbf{y}_i - \mathbf{Da}\|_2^2 + \lambda\|\mathbf{a}\|_p, \tag{5}$$

where $0 \le \gamma < 1$ balances the importance between the testing image $\mathbf{x}$ and its local neighborhood $\mathbf{Y}(\mathbf{x})$. $\mathbf{Y}(\mathbf{x})$ is searched from the training samples according to prior knowledge or manual labeling. For simplicity, we assume that each data point has K neighbors. The second term measures the contribution of locality, which is helpful to the robustness of the model.

The locality constraint (second term) in (5) is a simplified model to formulate the property of similar inputs having similar codes. From another view, it actually enforces that the reconstruction of $\mathbf{x}$ is still into the local neighborhood of $\mathbf{x}$.

Consider the recent findings [17,18], i.e., ℓ_1-norm-based sparsity might not bring a higher recognition accuracy than ℓ_2-norm-based methods in face recognition, we simplify our objective function (5) as follows:

$$\underset{\mathbf{c}}{\operatorname{argmin}} \ (1-\gamma)\|\mathbf{x}-\mathbf{Da}\|_2^2 + \gamma\frac{1}{K} \sum_{\mathbf{y}_i(\mathbf{x}) \in \mathbf{Y}(\mathbf{x})} \|\mathbf{y}_i(\mathbf{x}) - \mathbf{Da}\|_2^2 + \lambda\|\mathbf{a}\|_2^2. \tag{6}$$

Clearly, (6) achieves the minimum when its derivative with respect to $\mathbf{a}$ is zero. Hence, the optimal solution is

$$\mathbf{a}^* = (\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{D}^T\left[(1-\gamma)\mathbf{x} + \gamma\frac{1}{K} \sum_{\mathbf{y}_i(\mathbf{x}) \in \mathbf{Y}(\mathbf{x})} \mathbf{y}_i(\mathbf{x})\right]. \tag{7}$$

Let $\mathbf{P}=(\mathbf{D}^T\mathbf{D}+\lambda\mathbf{I})^{-1}\mathbf{D}^T$ whose calculation requires re-formulating the pseudo-inverse, it can be calculated in advance and only once as it is only dependent on training data $\mathbf{D}$.

Given a testing image $\mathbf{x}$, the first step is to determine its neighborhood $\mathbf{Y}(\mathbf{x})$ from the training set according to prior knowledge, or manual labeling, etc. In practical applications, there are two widely-used variations for finding the neighborhood:

1. ϵ-ball method: The training sample $\mathbf{d}_i$ is a neighbor of the testing image $\mathbf{x}$ if $\|\mathbf{d}_i-\mathbf{x}\|_2<\epsilon$, where $\epsilon>0$ is a constant.
2. K-nearest neighbors (K-NN) searching: The training sample $\mathbf{d}_i$ is a neighbor of $\mathbf{x}$, if $\mathbf{d}_i$ is among the K-nearest neighbors of $\mathbf{x}$, where $K>0$ can be specified as a constant or determined adaptively.

Once the neighborhood of the testing image $\mathbf{x}$ is obtained, LCCR just simply projects $\mathbf{x}$ and its neighborhood $\mathbf{Y}(\mathbf{x})$ onto space $\mathbf{P}$ via (7). In addition, the matrix form of LCCR is easily derived, which can used in batch prediction.

$$\mathbf{A}^*=(\mathbf{D}^T\mathbf{D}+\lambda\mathbf{I})^{-1}\mathbf{D}^T\left[(1-\gamma)\mathbf{X}+\gamma\frac{1}{K}\sum_{i=1}^{K}\mathbf{Y}_i(\mathbf{X})\right],$$

where the columns of $\mathbf{X}\in\mathbb{R}^{M\times J}$ are the testing images whose codes are stored in $\mathbf{A}^*\in\mathbb{R}^{N\times J}$, and $\mathbf{Y}_i(\mathbf{X})\in\mathbb{R}^{M\times J}$ denotes the collection of ith-nearest neighbor of $\mathbf{X}$.

The proposed LCCR algorithm is summarized in Algorithm 1, and an overview is illustrated in Fig. 2.

Algorithm 1. Face recognition using locality-constrained collaborative representation (LCCR).

Input: A matrix of training samples $\mathbf{D}=[\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_N]\in\mathbb{R}^{M\times N}$ which are sorted according to the label of $\mathbf{d}_i$, $1\le i\le N$; A testing image $\mathbf{x}\in\mathbb{R}^M$; The balancing factors $\lambda\ge 0$, $0\le\gamma\le 1$, and the size of neighborhood K.

1: Normalize the columns of $\mathbf{D}$ and $\mathbf{x}$ to have a unit ℓ_2-norm, respectively.

2: Calculate the projection matrix $\mathbf{P}=(\mathbf{D}^T\mathbf{D}+\lambda\mathbf{I})^{-1}\mathbf{D}^T$ and store it.

3: For each testing sample $\mathbf{x}$, find the neighborhood $\mathbf{Y}(\mathbf{x})=\{\mathbf{y}_1(\mathbf{x}),\ \mathbf{y}_2(\mathbf{x}),\ldots,\mathbf{y}_K(\mathbf{x})\}$ from the training samples $\mathbf{D}$.

4: Code $\mathbf{x}$ over $\mathbf{D}$ via

$$\mathbf{a}^*=\mathbf{P}\left[(1-\gamma)\mathbf{x}+\gamma\frac{1}{K}\sum_{\mathbf{y}_i(\mathbf{x})\in\mathbf{Y}(\mathbf{x})}\mathbf{y}_i(\mathbf{x})\right].$$

5: Compute the regularized residuals over all classes by

$$r_i(\mathbf{x})=\frac{\|\mathbf{x}-\mathbf{D}\delta_i(\mathbf{a}^*)\|_2}{\|\delta_i(\mathbf{a}^*)\|_2},\qquad(8)$$

where i denotes the index of class.

Output: identity$(\mathbf{x})=\text{argmin}_i\{r_i(\mathbf{x})\}$.

3.2. Discussions

From the algorithm, we can see that the performance of LCCR is positively correlated with that of K-NN searching method. Thus, it is possible to assume that LCCR will be failed if K-NN cannot find the correct neighbors for the testing sample. Here, we give a real example (Fig. 3) to illustrate that LCCR would largely avoid such situations from happening. In the example, the classification accuracy of LCCR is about 94% by using 600 AR images with sunglasses as testing image and 1400 clean ones as training samples.

Fig. 3(a) demonstrates the coefficients and residual of LCCR and CRC-RLS. For LCCR, we adopt Cityblock distance as metric to determine the neighborhood for each querying sample. From the results, we can see that these two methods correctly predicted the identity of the input, whereas K-NN searching could not find the correct neighbors (see Fig. 3(b)). It illustrates that LCCR could work well even though K-NN is failed to get the results. Fig. 3(c) and (d) illustrates another possible situation. That is, CRC-RLS fails to get the correct identity of the input, whereas LCCR successfully obtains the correct identity. Moreover, we could see that the locality constraint of LCCR shrinks the trivial coefficients and redistributes the representation of CRC-RLS. This verifies another claim [26] that locality must lead to sparsity, but not necessary vice versa. The property might play an important role in immunization of LCCR from K-NN failure.

3.3. Computational complexity analysis

The computational complexity of LCCR consists of two parts for offline and online computation, respectively. Suppose the dictionary $\mathbf{D}$ contains n samples with m dimensionality, LCCR takes $O(mn^2+n^3)$ to compute the projection matrix $(\mathbf{D}^T\mathbf{D}+\lambda\mathbf{I})^{-1}\mathbf{D}^T$ and $O(mn)$ to store it.

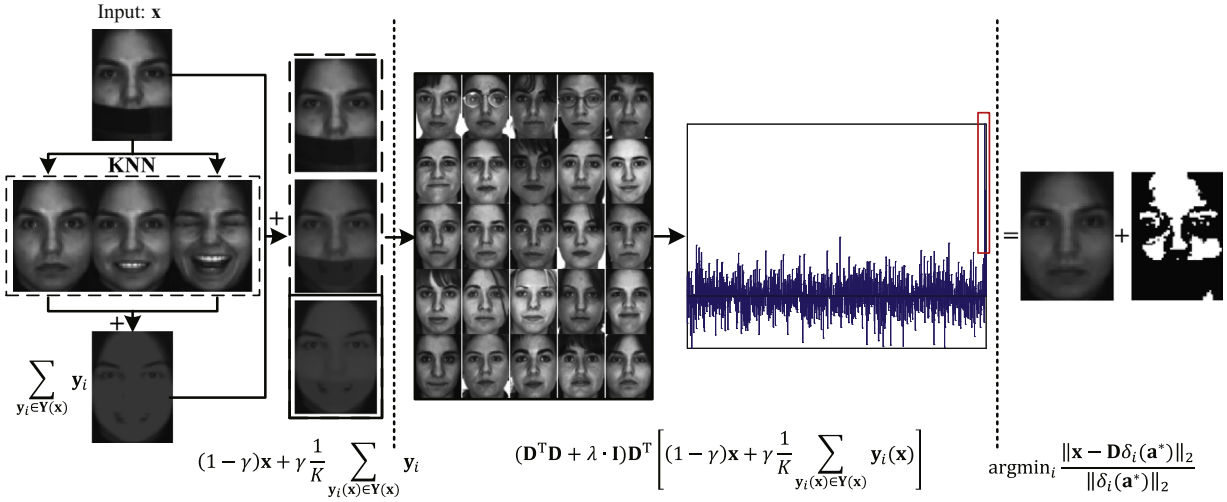


Fig. 2. Overview of the coding process of LCCR, which consists of three steps separated by dotted lines. For a given input $\mathbf{x}$, LCCR find its neighborhood $\mathbf{Y}(\mathbf{x})$ from training data; and then, encodes $\mathbf{x}$ over $\mathbf{D}$ by calculating the optimal representation $\mathbf{a}$ (see bar graph) which produces the minimal reconstruction errors for $\mathbf{x}$ and $\mathbf{Y}(\mathbf{x})$ simultaneously; after that, assigns the input to the class which produces the minimum residual. In the middle part of the figure, we use a red rectangles to indicate the basis vectors which produce the minimum residual. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

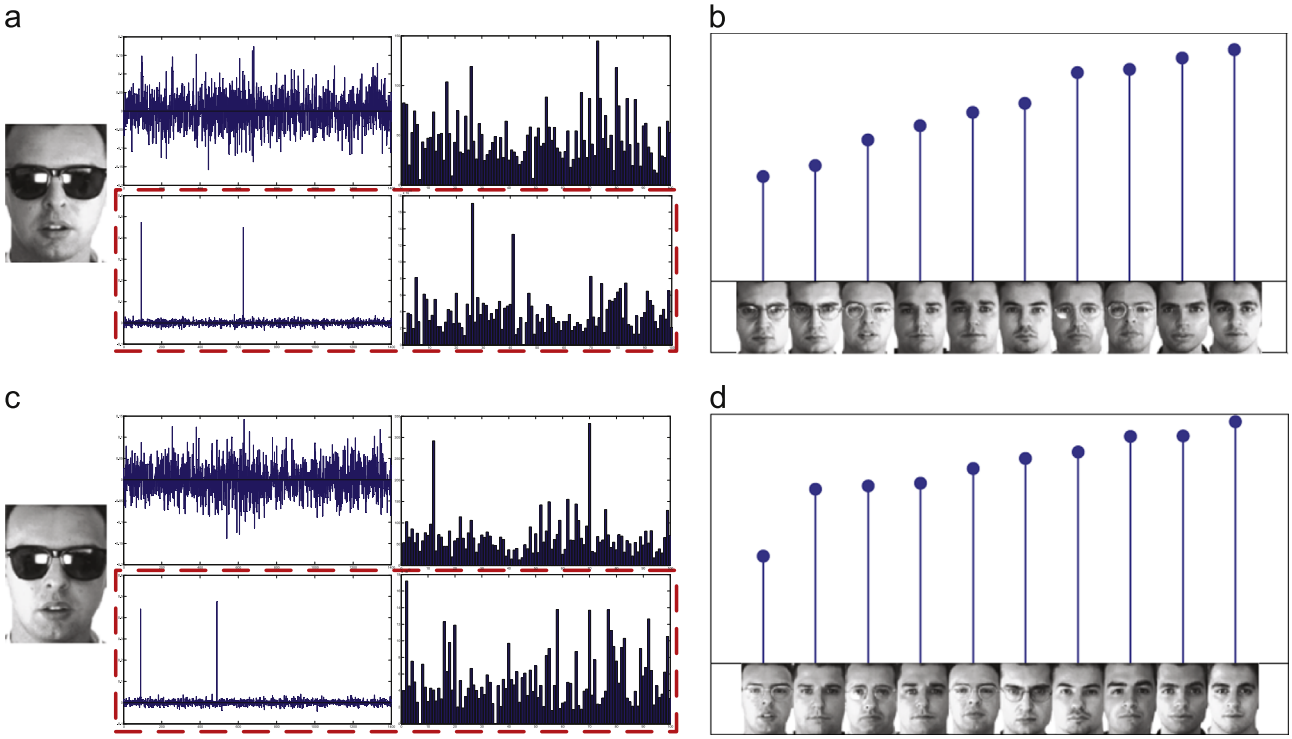


Fig. 3. The effectiveness of the proposed model. (a) A testing face disguised by sunglass comes from the 7th subject of AR database, where the figures (in the red rectangle) in the second row are the coefficients and residual of the input learned by LCCR (λ=0.005, γ=0.9, and k=2), and the figures in the first row are the results of CRC-RLS [18] (λ=0.001 for the best accuracy). (b) The 10 nearest neighbors of the input in terms of Cityblock distance (y-axis). (c) and (d) are the results of another testing sample from the same individual. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

For each querying sample $\mathbf{x}$, LCCR needs $O(mn)$ to search the K-nearest neighbors of $\mathbf{x}$ from $\mathbf{D}$. After that, the algorithm projects $\mathbf{x}$ into another space via (7) in $O(m^2n)$. Thus, the computational complexity of LCCR is $O(m^2n)$ for each unknown sample. Note that, the computational complexity of LCCR is same with that of LRC [17,19] and CRC-RLS [18]. Moreover, it is more competitive than SRC [10] even though the fastest ℓ_1-solver is used to obtain the sparse representation. For example, SRC takes $O(t_1m^2n+t_1mn^2)$ to encode each sample over $\mathbf{D}$ when Homotopy optimizer [33] is adopted to get the sparsest solution, where Homotopy optimizer is one of the fastest ℓ_1-minimization algorithm according to [5] and t denotes the number of iterations of Homotopy algorithm. From the above analysis, it is easy to find that a medium-sized data set may bring up the scalability issues with the models. To address this problem, a potential choice is to perform dimension reduction and sampling techniques to reduce the size of problem in practical application as suggested in [34].

4. Experimental verification and analysis

In this section, we report the performance of LCCR over five publicly accessed facial databases, i.e., Labeled Faces in the Wild-a database (LFW) [35], AR [36], ORL [37], the Extended Yale database B [38], and Multi-PIE [39]. We examine the recognition results of the proposed algorithm with respect to (1) discrimination, (2) robustness to corruptions, (3) and robustness to continue occlusions.

4.1. Experimental configuration

We compared the classification results of LCCR with four linear coding models (SRC [10], CESR [13], LRC [17,19] and CRC-RLS [18]) and a subspace learning algorithm (LPP [32]) with the nearest neighbors classifier (1NN). Moreover, we also reported the results of linear SVM [40] over the original inputs. Note that SRC, CESR, LRC, CRC-RLS and LCCR directly code each testing sample over training data without usability of dictionary learning method. SRC and CESR get classification results by finding which subject produces the minimum reconstruction error, whereas LRC, CRC-RLS and LCCR (these algorithms without ℓ_1-regularization) get the result by finding which subject has the minimum regularized residual. Moreover, in these models, only LCCR incorporates locality constraint into coding scheme. For a comprehensive comparison, we report the performance of LCCR with five basic distance metrics, i.e., Euclidean distance (ℓ_2-distance), Seuclidean distance (standardized Euclidean distance), Cosine distance (the cosine of the angle between two points), Cityblock distance (ℓ_1-distance), and Spearman distance.

For computational efficiency, as did in [10,18], we performed Eigenface [41] to reduce the dimensionality of data set throughout the experiments. Moreover, SRC requires the dictionary $\mathbf{D}$ to be an under-determined matrix, and Shi et al. [17] claimed that their model (named as LRC in [19]) will achieve competitive results when $\mathbf{D}$ is over-determined. For an extensive comparison, we investigate the performance of the tested methods except SRC over two cases.

We solved the ℓ_1-minimization problem in SRC by using the CVX [42], a package for solving convex optimization problems, and got the results of LRC, CRC-RLS and CESR by using the source codes from the homepages of the authors. All experiments are carried out using MATLAB on a 2.5 GHz machine with 2.00 GB RAM. Moreover, the MATLAB code of LCCR and the used databases can be downloaded at http://www.machineilab.org/users/pengxi/.

4.2. Recognition results with varying parameters

Parameter determination is a big challenge in pattern recognition and computer vision. To examine the influence of parameters over the recognition result, we carried out some experiments

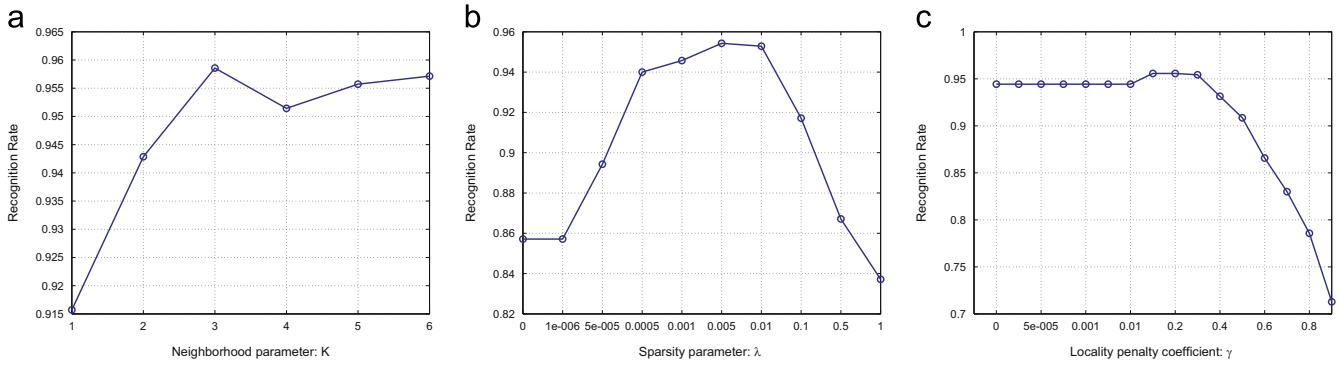


Fig. 4. Recognition accuracy of LCCR using Cityblock distance on a subset of AR database with dimensionality 2580. (a) The recognition rates versus the variation of the neighborhood parameter K, where $\lambda=0.005$ and $\gamma=0.2$. (b) The recognition rates versus the variation of the sparsity parameter λ, where $K=5$ and $\gamma=0.2$. (c) The recognition rates versus the variation of the locality constrained coefficient γ, where $K=3$ and $\lambda=0.005$.

using AR database. In each test, we varied the value of one parameter and fixed the value of the others. Fig. 4 illustrates the recognition rates of LCCR with varying K, λ, and γ. From the results, we could see that LCCR performs stable when λ is in the range of [0.0005 0.01] and γ is assigned to a small value. γ and K are used to incorporate locality, whose optimal value are dependent on the prior of data distribution. If pairwise distance is better than linear coefficient in finding neighbors, a bigger γ and K is helpful to improve the discrimination of LCCR. Otherwise, a smaller one is preferable. Moreover, if the data set is well-sampled, the optimal value of K should be the dimensionality of subspace for integrating enough discrimination into the representation.

In the following experiments, as did in [6,7], we report the best classification results of all the tested methods under different parameter configurations. The value range used to find the best values for LCCR has been shown into Fig. 4, and these possible values of λ also are tested for SRC and CRC-RLS. In all tests, we randomly split each data set into two parts for training and testing and compare the performance of the algorithms using the same partition to avoid the difference in data sets.

4.3. Recognition with various features

In this section, we evaluate the performance of LCCRs using Labeled Faces in the Wild-a database (LFW) which contains 13,123 images captured from uncontrolled environment with variations of pose, illumination, expression, misalignment and occlusion, etc. By following [29], we use a subset of aligned LFW [35] which consists of 143 subjects with no less than 11 samples per subject, and the first 10 samples are used for training and the remaining samples are used for testing. Moreover, we extract four features, e.g., intensity value (Gray value), low-frequency Fourier transform feature (FFT) [43], Gabor magnitude [44] and Local Binary Pattern (LBP) [45], for all the tested algorithms. For each feature, "divide and conquer" strategy is adopted. In details, each image is partitioned into 2×2 blocks; and then LDA [31] is performed in each block to get the discrimination-enhanced feature; after that, all blocks' features are concatenated to form the final feature vector.

From Table 1, we can conclude

- LCCRs generally outperform the other methods in the all tests. For example, LCCR with Cosine distance has more than 2.63%, 1.93%, 3.5%, and 4.7% higher rate than the second best algorithm over four different features.
- Cosine and Spearman distance are more helpful than the other three metrics to improve the recognition rate of LCCR. Moreover, the combination of LCCR with Spearman distance achieves the highest classification rate (73.62%) for the data set.

Table 1
The recognition accuracy of competing algorithms on the LFW database with four features.

Algorithms	Gray (%)	FFT [43] (%)	Gabor [44] (%)	LBP [45] (%)
SVM [40]	3.10	5.76	42.35	18.48
LPP [32]	5.72	6.86	54.88	53.46
SRC [10]	7.29	33.56	68.66	61.63
CESR [13]	9.62	37.57	65.74	63.05
LRC [17,19]	7.11	13.99	25.40	26.28
CRC-RLS [18]	7.25	14.03	25.40	26.31
LCCR + Cityblock	8.93	21.96	64.14	64.10
LCCR + Seuclidean	8.78	20.77	63.81	62.03
LCCR + Euclidean	9.59	22.19	64.58	66.55
LCCR + Cosine	**12.25**	39.50	72.16	**67.75**
LCCR + Spearman	11.48	**39.54**	**73.62**	66.51

- CESR and SRC obtain the second and third best result in most cases, especially, when Gabor and FFT features are used. In fact, we have found that LRC, CRC and LCCRs will achieve better results if the classification is performed by finding which subject has the minimal residual but the regularized minimal residual, i.e., the classifier 1 but 8 is used. However, this case is validate only for LFW data set. In the following experiments, the classifier 8 is a better choice for LRC, CRC and LCCRs.

4.4. Recognition on clean images

In this section, we examine the performance of seven competing methods over 4 clean facial data sets. Here, clean image means an image without occlusions or corruptions, just with variations in illumination, expression, etc.

(1) *ORL database* [37] consists of 400 different images of 40 individuals. For each person, there are 10 images with the variation in lighting, facial expression and facial details (with or without glasses). For computational efficiency, we cropped all ORL images from 112×96 to 56×48, and randomly selected 5 images from each subject for training and used the remaining 5 images for testing.

Table 2 reports the classification accuracy of the tested algorithms over various dimensionality. Note that the Eigenface with 200D retains 100% energy of the cropped data, which makes the investigated methods achieve the same rates over 2688D. From the results, LCCRs outperform the other algorithms, and the best results are achieved when Cityblock distance is used to search the nearest neighbors. Moreover, we can find that all the algorithms achieve a higher recognition rate in the original space except LRC. One possible reason is that the cropped operation degrades the

performance of LRC, another reason may attribute to the used classifier. We have found that if another nearest subspace classifier [10] is adopted with linear regression-based representation, the accuracy of LRC is slightly decreased from 89% to 88.00% over the original data and from 91% to 90% with 120D.

(2) *AR database* [36] includes over 4000 face images of 126 people (70 male and 56 female) which vary in expression, illumination and disguise (wearing sunglasses or scarves). Each subject has 26 images consisting of 14 clean images, 6 images with sunglasses and 6 images with scarves. As did in [10,18], we used a subset that contains 1400 clean faces randomly selected from 50 male subjects and 50 female subjects. For each subject, we randomly permute the 14 images and take the first half for training and the rest for testing. Limited by the computational capabilities, as did in [18], we resized all images from original 165×120 to 60×43 (2580D) and convert it to gray scale.

(3) *Extended Yale B database* [38] contains 2414 frontal-face images with size 192×168 over 38 subjects, as suggested in [10,18], we carried out the experiments using the cropped and normalized images of size 54×48. For each subject (about 64 images per subject), we randomly split the images into two parts with equal size, one for training, and the other for testing. Similar to the above experimental configuration, we calculated the recognition rates over dimensionality 54, 120 and 300 using Eigenface, and 2592D in the original data space. Table 4 shows that LCCRs again outperform its counterparts across various spaces, especially when the Spearman distance is used to determine the neighborhood of testing samples.

(4) *Multi-PIE database* (MPIE) [39] contains the images of 337 subjects captured in 4 sessions with simultaneous variations in pose, expression and illumination. As did in [18], we used 14 frontal images with 14 illuminations[1] and neutral expression from Session 1 for training, and used 10 frontal images[2] from Sessions 2 to Session 4 for testing. Note that, we used all the 249 subjects from Session 1 and discarded the images that not belong to the subjects of Session 1 from Session 2 to 4. Moreover, all images are downsized from 100×82 to 50×41.

From Tables 2 to 5, we draw the following conclusions:

1. LCCRs generally outperforms SVM (original input), SRC (sparse representation), CESR (robust sparse representation), LRC (linear regression-based model) and CRC-RLS (collaborative representation) over the tested cases.
2. LCCRs perform better in a low-dimensional space than a high-dimensional ones. For example, on the Extended Yale B, the difference in accuracy between LCCR and CRC-RLS (the second best method) changed from 4.06% (54D) to 2.49% (120D) and to 1.17% (300D). It again corroborates our claim that local consistency is helpful to improving the discrimination of data representation, since the low-dimensional data contain few information than higher one.
3. CESR is more competitive in the original space at the cost of computing cost. For example, it outperforms the other models over MPIE-S4 in classification accuracy where its time cost about 11003.51 s, compared with 3104.82 s of SRC, 54.59 s of LRC, 54.79 s of CRC-RLS and 59.82 s of LCCR.
4. SRC, LRC and CRC-RLS achieve the similar performance, and SRC is more competitive in the low-dimensional feature spaces. The results are consistent with the reports in [18]. For example, in the experiments of Zhang over MPIE-S2 with 300D, the accuracy scores of SRC and CRC-RLS are about 93.9% and 94.1%, respectively, comparing with 93.13% and 94.88% in our experiments. Moreover, CRC-RLS and LRC achieve similar recognition rates with the difference less than 1% across various feature spaces.

[1] Illuminations 0,1,3,4,6,7,8,11,13,14,16,17,18,19.
[2] Illuminations 0,2,4,6,8,10,12,14,16,18.

Table 2
The maximal recognition accuracy of competing algorithms on the ORL database.

Dim	**54** (%)	**120** (%)	**200** (%)	**2688(%)**
SVM [40]	90.00	92.50	93.50	93.00
LPP [32]	86.00	86.50	86.50	86.50
SRC [10]	92.00	96.50	86.00	–
CESR [13]	89.50	88.50	89.00	97.50
LRC [17,19]	92.50	91.00	89.00	89.00
CRC-RLS [18]	94.50	94.00	94.50	95.00
LCCR + Cityblock	**97.50**	**97.50**	**98.00**	**98.00**
LCCR + Seuclidean	96.00	96.50	96.00	96.50
LCCR + Euclidean	96.00	96.00	96.50	96.50
LCCR + Cosine	96.00	96.50	96.50	96.50
LCCR + Spearman	96.00	96.00	96.00	96.00

Table 3
The maximal recognition accuracy of competing algorithms on the AR database.

Dim	**54** (%)	**120** (%)	**300** (%)	**2580(%)**
SVM [40]	73.43	81.00	82.00	83.14
LPP [32]	39.29	43.57	53.86	53.86
SRC [10]	81.71	88.71	90.29	–
CESR [13]	74.00	81.43	84.57	84.57
LRC [17,19]	80.57	90.14	93.57	82.29
CRC-RLS [18]	80.57	90.43	94.00	94.43
LCCR + Cityblock	**86.14**	**92.71**	**95.14**	**95.86**
LCCR + Seuclidean	85.00	91.86	94.43	95.43
LCCR + Euclidean	84.00	91.29	94.14	94.86
LCCR + Cosine	83.43	90.86	94.00	94.57
LCCR + Spearman	84.71	90.71	94.14	94.43

Table 4
The maximal recognition accuracy of competing algorithms on the extended Yale B database.

Dim	**54** (%)	**120** (%)	**300** (%)	**2592(%)**
SVM [40]	84.52	92.72	95.28	95.45
LPP [32]	35.93	54.55	70.78	75.66
SRC [10]	93.71	95.12	96.44	–
CESR [13]	92.30	94.95	95.53	96.11
LRC [17,19]	92.88	95.61	97.85	90.48
CRC-RLS [18]	92.96	95.69	97.90	98.26
LCCR + Cityblock	93.21	96.03	97.93	98.34
LCCR + Seuclidean	93.21	95.70	97.93	98.34
LCCR + Euclidean	93.21	95.70	97.93	98.34
LCCR + Cosine	93.46	95.78	97.93	98.59
LCCR + Spearman	**97.02**	**98.18**	**99.10**	**99.59**

4.5. Recognition on partial facial features

The ability to work on partial face features is very interesting since not all facial features play an equal role in recognition. Therefore, this ability has become an important metric in the face recognition researches [46]. We examine the performance of the investigated methods using three partial facial features, i.e., right eye, nose, as well as mouth and chin, sheared from the clean AR faces with 2580D (as shown in Fig. 5(a)). For each partial face feature, we generate a data set by randomly selecting 7 images per subject for training and the remaining 700 for testing. It should be

noted that [10] conducted the similar experiment on Extended Yale B which includes less subjects, smaller irrelevant white background, and more training samples per subject than our case.

Fig. 5(b) shows that LCCRs achieve better recognition rates than SVM, SRC, LRC and CRC-RLS for right eye as well as mouth and chin, and the second best rates for the nose. Some works found that the most important feature is the eye, followed by the mouth, and then the nose [47]. We can see that the results for SVM, CRC-RLS and LCCR are consistent with the conclusions even though the dominance of the mouth and chin over the nose is not very distinct.

4.6. Face recognition with block occlusions

To examine the robustness to block occlusion, similar to [10,17,18], we get 700 testing images by replacing a random block of each clean AR image with an irrelevant image (baboon) and use 700 clean images for training. The occlusion ratio increases from 10% to 50%, as shown in Fig. 6(a). We investigate the classification accuracy of the methods across Eigenface space with 300D (Fig. 6(b)) and cropped data space with 2580D (Fig. 6(c)).

Fig. 6(b)–(d) shows that LCCRs generally outperform the other models with considerable performance margins. Especially, with the increase of the occlusion ratio, the difference in recognition rates of LCCRs and the other methods becomes larger. For example, when the occlusion ratio is 50%, in 300 dimensional space, the accuracy of LCCR with Cityblock distance is about 19.7% higher than SVM, about 44.1% higher than LPP, about 26.0% higher than SRC (CVX), about 35.1% higher than CESR, about 19.6% higher than LRC, and about 15.1% higher than CRC-RLS. Note that, different ℓ_1-solvers will lead to different results for SRC. For Example, if SRC adopts Homotopy algorithm [33] to get the sparest solution, the recognition rate will increase from 25.43% (with CVX) to 36.14% such that the performance dominance decreases from 26% to 15.3%. Moreover, CESR achieves the best results at the cost of computational cost when the original data is available and the occluded ratio ranges from 20% to 40%.

On the other hand, it is easy to find that LRC, CRC-RLS and LCCRs are more robust than SRC and SVM, which implies that the ℓ_1-regularization term cannot yield better robustness than the ℓ_2-regularization term, at least for the Eigenface space. Moreover, the models achieve better results in higher dimensional space, even though the difference of classification accuracy between higher dimensional space and lower ones is less than 1% except CESR has an obvious improvement.

4.7. Face recognition with real occlusions

In this sub-section, we examine the robustness to real possible occlusions of the investigated approaches over the AR data set. We use 1400 clean images for training, 600 faces wearing by sunglasses (occluded ratio is about 20%) and 600 face wearing by scarves (occluded ratio is about 40%) for testing, separately. In [10], Wright et al. only used a third of disguised images for this test, i.e., 200 images for each kind of disguises. In addition, we also investigate the role of K-NN searching in LCCR.

We examine two widely used feature schemes, namely, the holistic feature with 300D and 2580D, as well as the partitioned feature based on the cropped data. The partitioned feature scheme firstly partitions an image into multiple blocks (8 blocks as did in

Table 5
The maximal recognition accuracy of competing algorithms on the multi PIE database.

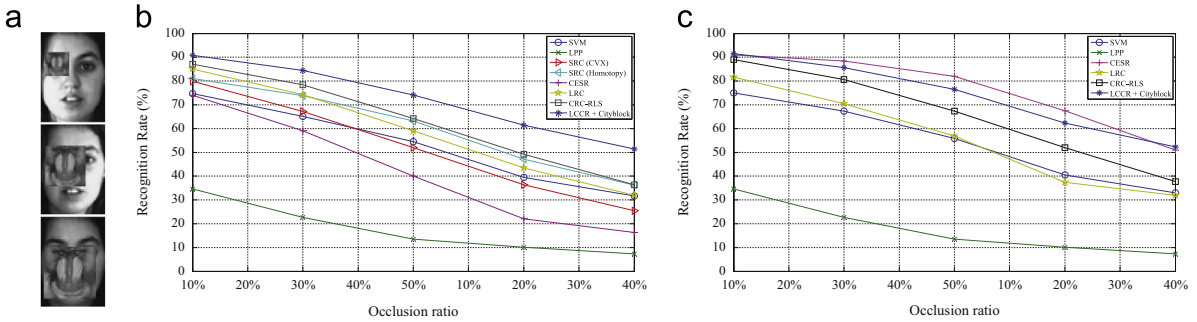
Dim	**300**			**2050**		
Dataset	MPIE-S2 (%)	MPIE-S3 (%)	MPIE-S4 (%)	MPIE-S2 (%)	MPIE-S3 (%)	MPIE-S4 (%)
SVM [40]	91.33	85.13	89.20	91.45	85.75	89.43
LPP [32]	40.12	27.44	31.20	31.49	31.00	31.49
SRC [10]	93.13	90.60	94.10	–	–	–
CESR [13]	92.41	87.38	91.94	94.46	92.06	**96.17**
LRC [17,19]	94.64	89.88	93.37	83.19	70.25	75.03
CRC-RLS [18]	94.88	89.88	93.60	95.30	90.56	94.46
LCCR + Cityblock	**95.36**	91.25	**95.54**	**96.08**	91.94	95.89
LCCR + Seuclidean	95.06	91.25	94.51	95.84	91.56	95.20
LCCR + Euclidean	95.06	91.31	94.57	95.84	91.63	95.14
LCCR + Cosine	95.12	90.88	94.69	95.78	91.38	95.14
LCCR + Spearman	95.12	**91.75**	94.40	95.78	**92.19**	95.20

a

b

Features	**Right Eye**	**Mouth and Chin**	**Nose**
Dim	**308**	**798**	**224**
SVM [40]	70.71%	41.29%	37.14%
LPP [32]	58.57%	14.43%	53.71%
SRC [10]	84.00%	70.71%	**78.00%**
CESR [13]	81.57%	56.00%	70.43%
LRC [17, 19]	72.86%	32.86%	70.57%
CRC-RLS [18]	83.14%	73.86%	73.57%
LCCR + Cityblock	**86.86%**	**76.29%**	75.86%
LCCR + Seuclidean	84.86%	75.57%	75.29%
LCCR + Euclidean	85.29%	75.00%	76.00%
LCCR + Cosine	84.43%	74.57%	75.29%
LCCR + Spearman	84.57%	75.86%	75.14%

Fig. 5. Recognition accuracy with partial face features. (a) An example of the three features, right eye, mouth and chin, and nose from left to right. (b) The recognition rates of competing methods on the partial face features of the AR database.



d

Dim	300 (Eigenface)					2580				
Occlusionratio	10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
LCCR + Cityblock	90.71%	84.43%	74.14%	61.43%	51.43%	91.57%	85.57%	76.43%	62.29%	52.29%
LCCR + Seuclidean	90.00%	80.14%	67.86%	53.71%	39.57%	91.00%	82.14%	70.57%	56.29%	41.71%
LCCR + Euclidean	88.86%	80.86%	68.14%	55.00%	43.29%	89.71%	81.86%	71.00%	57.00%	43.86%
LCCR + Cosine	88.57%	80.29%	69.29%	56.57%	44.00%	89.00%	82.43%	71.00%	57.29%	45.57%
LCCR + Spearman	88.29%	81.00%	70.00%	57.14%	43.29%	89.57%	83.00%	72.57%	59.43%	44.14%

Fig. 6. Experiments on AR database with varying percent block occlusion. (a) From top to bottom, the occlusion percents for test images are, 10%, 30%, and 50%, respectively. (b) and (c) are the recognition rates under different levels of block occlusion on AR database with 300D (Eigenface) and 2580D, respectively. (d) The recognition rates of LCCRs with 300D and 2580D.

[10,18,29], see Fig. 7(a) and (c), then conducts classification on each block independently, and after that, aggregates the results by voting.

Fig. 7(e) reports the recognition rates of all the tested methods. For the images occluded by sunglasses, LCCR with Cityblock distance and CESR achieve remarkable results with the holistic feature scheme, their recognition accuracy are nearly double that of the other methods. This considerable performance margin contributes to the accuracy of *K*-NN searching based on Cityblock distance (see Fig. 7(b)).

For the images occluded by scarves, LCCR achieves the highest recognition rate over the full dimensional space, and the second highest rates using Eigenface. However, the difference in rates between LCCR and other non-iterative algorithms (LRC, CRC-RLS) is very small due to the poor accuracy of *K*-NN searching as shown in Fig. 7(d). Furthermore, the partitioned feature scheme produces higher recognition rates than the holistic one for all competing methods, which is consistent with previous report [10].

From the above experiments, it is easy to conclude that the preservation of locality is helpful to coding scheme, especially when the real structures of data cannot be found by traditional coding scheme. Moreover, the performance ranking of LCCR with five distance metrics is same with that of *K*-NN searching with the used metrics.

4.8. Face recognition with corruption

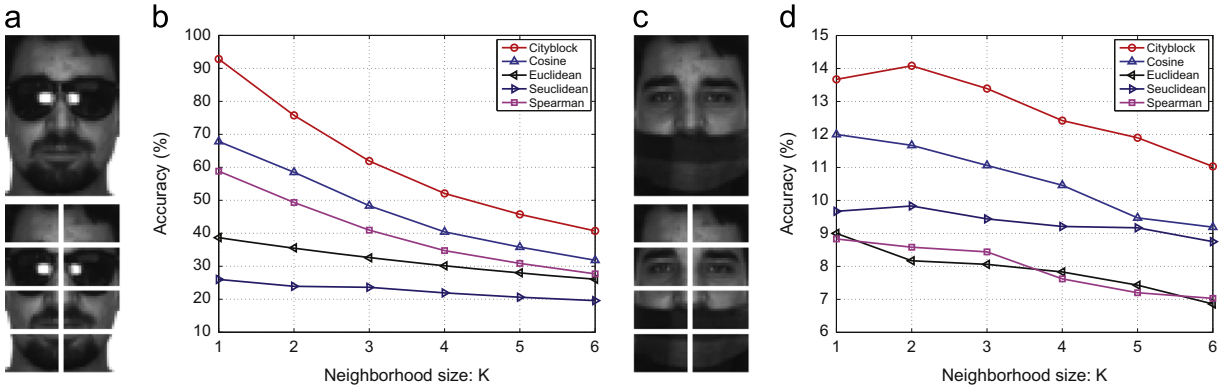
We test the robustness of LCCR against two kinds of corruption using the AR data set containing 2600 images of 100 individuals. For each subject, we use 13 images for training (7 clean images, 3 images with sunglasses, and 3 images with scarves), and the remaining 13 images for testing. Different from [10] which tested the robustness to corruption using the Extended Yale B database, our case is more challenging for the following reasons. First, AR images contain real possible occlusions, i.e., sunglasses and scarves, while Extended Yale B is a set of clean images without disguises. Second, AR includes more facial variations (13 versus 9), more subjects (100 versus 38), and a smaller samples for each subject (26 images per subject versus 64 images per subject). Third, we investigated two kinds of corruption, white noise (additive noise) and random pixel corruption (non-additive noise) which are two commonly assumed in face recognition problem [10,17,19]. For the white noise case (the top row of Fig. 8), we add random noise from normal distribution to each testing image $\mathbf{x}$, that is, $\tilde{\mathbf{x}} = \mathbf{x} + \alpha\mathbf{n}$, and restrict $\tilde{\mathbf{x}} \in [0\ 255]$, where α is the corruption ratio from 10% to 90% with an interval of 20%, and $\mathbf{n}$ is the noise following a standard normal distribution. For the random pixel corruption case (the bottom row in Fig. 8), we replace the value of a percentage of pixels randomly chosen from each test image with the values following a uniform distribution over $[0\ p_{max}]$, where p_{max} is the largest pixel value of current image.

To improve the anti-noise ability of SRC [10], Wright et al. generate a new dictionary $[\mathbf{D}\ \mathbf{I}]$ by concatenating an identity matrix $\mathbf{I}$ with the original dictionary $\mathbf{D}$, where the dimensionality of $\mathbf{I}$ equals to that of data. The use of $\mathbf{I}$ has been verified to be effective in improving the robustness of ℓ_1-norm-based models [48,10] at the cost of time-consuming. Therefore, it is a tradeoff between robustness and efficiency for the algorithms. Will the strategy still work for ℓ_2-minimization-based models? In this sub-section, we fill this gap by comparing the results by coding over these two dictionary.

Tables 6–9 are the recognition rates of the tested methods across feature space (Eigenface with 300D) and full dimensional space (2580D). We did not reported the results of SVM and LPP with the strategy of expanding dictionary since the methods are not belong to the facility of linear coding scheme. Moreover, SRC requires the dictionary is an over-completed matrix to obtain the sparse solution via ℓ_1-optimazation program. Based on the results, we have the following conclusions:

First, the proposed LCCRs are much superior to SVM, LPP, SRC, CESR, LRC and CRC-RLS. For example, in the worst case (the white Gaussian noise corruption ratio is 90%, the best result of LCCR is about 90.54% (Table 7), compared to 82.92% of SVM (Table 7), 2.62% of LPP (Table 6), 84.08% of SRC (Table 8), 73% of CESR (Table 6), 87.15% of LRC (Table 9), and 88.39% of CRC-RLS (Table 7). In the case of random pixel corruption, one can see when the corruption ratio reaches 70%, all methods fail to perform recognition except LCCR in the two data spaces and CESR in the full dimensional space.

Second, all investigated algorithms perform worse with increased corruption ratio and achieve better results in white noise corruption (additive noise) than random pixel corruption



e

Disguise	sunglasses			scarves		
Feature	Holistic		Partitioned	Holistic		Partitioned
Dim	300	2580	2580	300	2580	2580
SVM [40]	47.83%	48.67%	40.17%	13.50%	13.83%	41.67%
LPP [32]	14.83%	18.50%	88.83%	20.33%	24.17%	82.00%
SRC [10]	57.00%	-	93.00%	**69.83%**	-	91.83%
CESR [13]	21.17%	**95.50%**	**97.50%**	31.50%	16.67%	91.33%
LRC [17, 19]	52.83%	49.17%	88.17%	68.50%	57.50%	91.83%
CRC-RLS [18]	53.00%	71.50%	88.33%	68.50%	89.17%	92.17%
LCCR + Cityblock	**93.00%**	93.50%	91.17%	68.67%	89.17%	92.50%
LCCR + Seuclidean	57.83%	74.50%	91.50%	68.50%	89.17%	92.17%
LCCR + Euclidean	66.00%	79.00%	90.83%	68.67%	89.17%	92.50%
LCCR + Cosine	76.83%	85.33%	93.83%	68.83%	**89.50%**	**93.83%**
LCCR + Spearman	70.67%	81.17%	95.83%	68.67%	89.33%	**93.83%**

Fig. 7. Recognition on AR faces with real possible occlusions. (a) The top row is a facial image occluded by sunglass, whose partitioned blocks are shown as below. (b) The accuracy of *K*-NN searching using Cityblock distance, Cosine distance, Euclidean distance, Seuclidean distance and Spearman distance on the AR images with sunglasses (2580D). (c) Similar to (a), the top row is a face occluded by scarf, and its partitions below. (d) The precision of *K*-NN searching using Cityblock, Cosine, Euclidean, Seuclidean, Spearman as distance metrics on the AR images with scarves (2580D). (e) The recognition rates of competing methods across different experimental configurations.

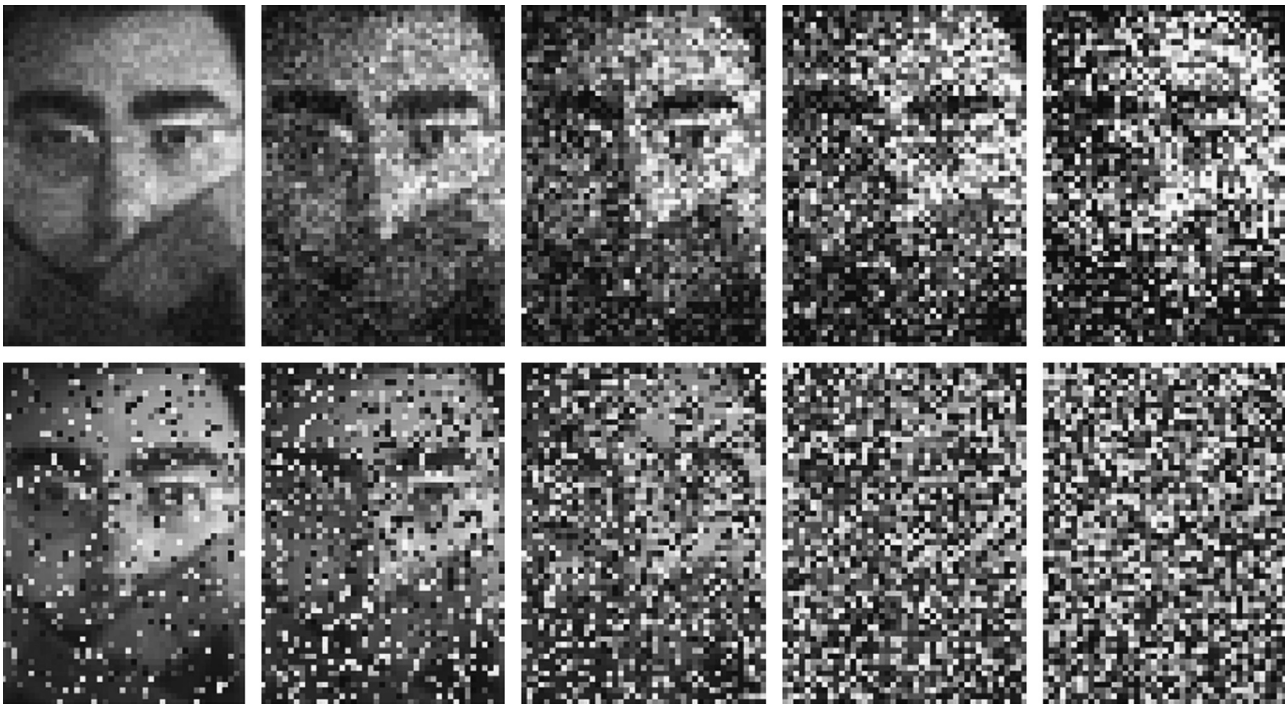


Fig. 8. Testing images from AR database with additive noise and non-additive noise. Top row: 10%, 30%, 50%, 70%, 90% white noises are added into test image; Bottom row: the case of random pixel corruption with 10%–90% percentages, respectively.

Table 6
The robustness of different methods over AR database with 300D (coding over **D**).

Corruptions	White Gaussian noise					Random pixel corruption				
Corrupted ratio	**10%**	**30%**	**50%**	**70%**	**90%**	**10%**	**30%**	**50%**	**70%**	**90%**
SVM [40]	91.77	91.38	90.23	88.62	82.69	91.54	81.92	45.46	8.23	2.00
LPP+1NN [32]	29.31	8.46	4.08	2.38	2.62	5.69	2.62	2.00	1.46	1.17
SRC [10]	92.62	91.23	86.54	78.31	62.62	89.62	72.31	38.85	8.23	2.00
CESR [13]	89.69	87.85	85.38	80.85	73.00	87.38	76.31	43.23	12.38	1.46
LRC [17,19]	93.39	92.39	88.85	81.85	67.62	91.77	77.00	45.77	13.62	2.54
CRC-RLS [18]	94.77	94.39	92.85	90.92	87.31	84.08	88.69	65.46	20.77	2.92
LCCR + Cityblock	**97.00**	**96.00**	94.54	92.31	89.08	**96.54**	92.31	79.69	37.08	5.23
LCCR + Seuclidean	96.31	95.85	94.46	92.39	88.54	95.69	90.08	65.85	20.77	3.00
LCCR + Euclidean	95.77	95.23	94.23	92.31	88.31	95.39	90.39	67.23	20.92	3.23
LCCR + Cosine	95.62	95.31	93.92	92.15	88.69	94.85	89.46	65.62	20.85	3.69
LCCR + Spearman	96.15	95.39	**94.69**	**93.08**	**89.77**	95.54	**92.54**	**83.00**	**59.31**	**13.69**

Table 7
The robustness of different methods over AR database with 2580D (Coding over **D**).

Corruptions	White Gaussian noise					Random pixel corruption				
Corrupted ratio	**10%**	**30%**	**50%**	**70%**	**90%**	**10%**	**30%**	**50%**	**70%**	**90%**
SVM [40]	91.92	91.31	89.85	88.62	82.92	91.69	81.46	44.77	7.69	1.92
LPP+1NN [32]	37.69	10.31	4.54	3.00	2.54	6.92	2.54	2.15	1.46	1.47
CESR [10]	90.85	86.69	84.38	78.85	70.08	91.00	90.77	**90.54**	**66.08**	13.08
LRC [17,19]	78.69	33.77	4.62	4.62	2.69	21.77	3.77	2.39	0.92	1.15
CRC-RLS [18]	94.85	94.77	93.23	90.85	88.39	94.15	89.08	67.08	22.69	2.62
LCCR + Cityblock	**97.54**	96.08	95.08	93.15	**90.54**	**96.85**	**93.23**	78.77	29.77	4.54
LCCR + Seuclidean	96.92	**96.23**	**95.39**	92.92	89.00	96.00	90.77	67.31	22.69	3.00
LCCR + Euclidean	96.08	95.62	94.85	92.23	88.92	95.62	91.15	68.31	22.92	3.00
LCCR + Cosine	96.08	95.46	94.39	92.54	89.46	95.31	90.77	67.15	23.23	3.62
LCCR + Spearman	96.54	95.23	94.69	**93.31**	90.39	95.85	92.92	83.31	60.69	**13.85**

Table 8
The robustness of different methods over AR database with 300D (Coding over [**D E**]).

Corruptions	White Gaussian noise					Random pixel corruption				
Corrupted ratio	**10%**	**30%**	**50%**	**70%**	**90%**	**10%**	**30%**	**50%**	**70%**	**90%**
SRC [10]	92.62	91.08	90.46	87.92	84.08	91.62	83.38	56.31	14.92	2.08
CESR [13]	85.69	82.46	80.31	73.69	63.15	83.38	69.00	34.69	10.62	2.92
LRC [17,19]	92.62	92.15	91.00	88.85	86.31	91.15	84.46	49.92	8.31	1.85
CRC-RLS [18]	92.62	92.15	91.00	88.85	86.31	91.15	84.46	49.92	8.31	1.85
LCCR + Cityblock	**93.08**	92.39	91.69	89.62	86.62	92.62	87.54	69.46	34.00	**4.92**
LCCR + Seuclidean	92.85	92.39	91.69	89.62	86.46	91.85	85.23	50.00	8.31	2.46
LCCR + Euclidean	92.85	92.39	91.69	89.92	86.62	91.92	85.15	51.23	10.23	2.54
LCCR + Cosine	92.69	92.69	91.92	89.92	87.00	92.15	84.92	49.92	8.38	2.62
LCCR + Spearman	92.85	**92.85**	**92.15**	**90.15**	**87.54**	**92.69**	**87.85**	**75.92**	**56.54**	1.38

Table 9
The robustness of different methods over AR database with 2580D (coding over [**D E**]).

Corruptions	White Gaussian noise					Random pixel corruption				
Corrupted ratio	**10%**	**30%**	**50%**	**70%**	**90%**	**10%**	**30%**	**50%**	**70%**	**90%**
CESR [13]	91.85	87.15	82.08	71.85	57.62	91.38	**91.15**	**90.23**	**76.85**	8.08
LRC [17,19]	93.00	92.46	91.69	90.15	87.15	92.15	84.92	51.15	8.54	1.77
CRC-RLS [18]	93.00	92.46	91.69	89.23	87.15	92.15	84.92	51.15	8.54	1.77
LCCR + Cityblock	**93.39**	93.00	92.15	90.15	87.39	93.08	88.23	70.15	34.08	5.08
LCCR + Seuclidean	93.31	93.00	92.15	90.15	87.31	92.69	85.85	51.39	8.62	2.46
LCCR + Euclidean	93.31	92.77	91.92	90.31	87.62	92.62	86.00	52.46	10.69	2.54
LCCR + Cosine	93.15	**93.23**	92.23	90.39	87.77	92.69	85.62	51.59	8.62	2.77
LCCR + Spearman	**93.39**	93.15	**92.62**	**90.54**	**88.31**	**93.31**	88.54	77.00	56.54	**13.62**

(non-additive noise). Moreover, the improvement of CESR is obvious when the original data is used to test. As discussed in the above, the improvement is at the cost of computational efficiency. For the other methods, they perform slightly better (less than 1%) in the full-dimensional space except LRC.

Third, the results show that coding over [**D I**] is helpful in improving the robustness of SRC and LRC, but it has negative impact on the recognition accuracy of CESR, CRC-RLS and LCCR. For example, when white noise ratio rises to 90% for the Eigenface (Table 6), expanding **D** leads to the variation of the recognition rate from 62.62% to 84.08% for SRC, from 73.00% to 63.15% for CESR, from 67.62% to 86.31% for LRC, from 87.31% to 86.31% for CRC-RLS, and from 89.77% to 87.54% for LCCR with Spearman distance. The conclusion has not been reported in the previous works.

5. Conclusions and discussions

It is interesting and important to improve the discrimination and robustness of data representation. The traditional coding algorithm gets the representation by encoding each datum as a linear combination of a set of training samples, which mainly depicts the global structure of data. However, it will be failed when the data are grossly corrupted. Locality (Local consistency) preservation, which keeps the geometric structure of manifold for dimension reduction, has shown the effectiveness in revealing the real structure of data. In this paper, we proposed a novel objective function to get an effective and robust representation by enforcing the similar inputs produce similar codes, and the function possesses analytic solution.

The experimental studies showed that the introduction of locality makes LCCR more accurate and robust to various occlusions and corruptions. We investigated the performance of LCCR with five basic distance metrics (for locality). The results imply that if better K-NN searching methods or more sophisticated distance metrics are adopted, LCCR might achieve a higher recognition rate. Moreover, the performance comparisons over two different dictionaries show that it is unnecessary to expand the dictionary **D** with **I** for ℓ_2-norm-based coding algorithms.

Each approach has its own advantages and disadvantages. Parameter determination maybe is the biggest problem of LCCR which requires three user-specified parameters. In the future works, it is possible to explore the relationship between locality parameter k and the intrinsic dimensionality of sub-manifold. Moreover, the work has focused on the representation learning, however, dictionary learning is also important and interesting in this area. Therefore, an possible way to extend this work is exploring how to reflect local consistency in the formation process of dictionary.

Conflict of interest statement

None declared.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of this paper. The authors also thanks Dr. Meng Yang to provide some data sets for experiments. This work was supported by National Basic Research Program of China 973 Program under Grant No. 2011CB302201, Program for New Century Excellent Talents in University under grant NCET-11-0355, National Natural Science Foundation of China under grant 61332002, and National Natural Science Foundation of China under grant 61322203.

References

[1] E.J. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215.
[2] D.L. Donoho, For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution, Commun. Pure Appl. Math. 59 (6) (2006) 797–829.
[3] S.S.B. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[4] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–451.
[5] A. Yang, A. Ganesh, S. Sastry, Y. Ma, Fast ℓ^1-minimization algorithms and an application in robust face recognition: a review, in: Proceedings of International Conference on Image Processing, 2010, pp. 1849–1852.
[6] X. Peng, Z. Lei, Z. Yi, Constructing l2-graph for subspace learning and segmentation, ArXiv e-prints arXiv:1209.0841.
[7] B. Cheng, J. Yang, S. Yan, Y. Fu, T. Huang, Learning with ℓ^1-graph for image analysis, IEEE Trans. Image Process. 19 (4) (2010) 858–866.
[8] E. Elhamifar, R. Vidal, Sparse Subspace Clustering: Algorithm, Theory, and Applications, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35 (2013) 2765–2781.
[9] C. Wang, X. He, J. Bu, Z. Chen, C. Chen, Z. Guan, Image representation using Laplacian regularized nonnegative tensor factorization, Pattern Recognit. 44 (10) (2011) 2516–2526.
[10] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[11] H. Zhang, N.M. Nasrabadi, Y. Zhang, T.S. Huang, Joint dynamic sparse representation for multi-view face recognition, Pattern Recognit. 45 (4) (2012) 1290–1298, http://dx.doi.org/10.1016/j.patcog.2011.09.009.
[12] H. Zhang, Y. Zhang, T.S. Huang, Simultaneous discriminative projection and dictionary learning for sparse representation based classification, Pattern Recognit. 46 (1) (2013) 346–354.
[13] R. He, W.-S. Zheng, B.-G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1561–1576.
[14] S.Z. Li, J. Lu, Face recognition using the nearest feature line method, IEEE Trans. Neural Netw. 10 (2) (1999) 439–443.
[15] J. Yang, L. Zhang, Y. Xu, J. Yang, Beyond sparsity: the role of l1-optimizer in pattern classification, Pattern Recognit. 45 (3) (2012) 1104–1118.
[16] R. Rigamonti, M.A. Brown, V. Lepetit, Are sparse representations really relevant for image classification? in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2011, pp. 1545–1552.
[17] Q. Shi, A. Eriksson, A. van den Hengel, C. Shen, Is face recognition really a compressive sensing problem? in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 553–560.
[18] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 471–478.
[19] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
[20] I. Naseem, R. Togneri, M. Bennamoun, Robust regression for face recognition, Pattern Recognit. 45 (1) (2012) 104–118.
[21] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: Proceedings of IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1208–1213.
[22] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, J. Mach. Learn. Res. 7 (2006) 2399–2434.
[23] S.C. Yan, D. Xu, B.Y. Zhang, H.J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.
[24] R.G. Baraniuk, M.B. Wakin, Random projections of smooth manifolds, Found. Comput. Math. 9 (1) (2009) 51–77.
[25] A. Majumdar, R.K. Ward, Robust classifiers for data reduced via random projections, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 40 (5) (2010) 1359–1371, http://dx.doi.org/10.1109/TSMCB.2009.2038493.
[26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 3360–3367.
[27] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[28] Y. Chao, Y. Yeh, Y. Chen, Y. Lee, Y. Wang, Locality-constrained group sparse representation for robust face recognition, in: Proceedings of IEEE International Conference on Image Processing, 2011, pp. 761–764.
[29] M. Yang, L. Zhang, D. Zhang, S. Wang, Relaxed collaborative representation for pattern classification, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2224–2231.
[30] K. Ohki, S. Chung, Y.H. Ch'ng, P. Kara, R.C. Reid, Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex, Nature 433 (7026) (2005) 597–603.

[31] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[32] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[33] M.R. Osborne, B. Presnell, B.A. Turlach, A new approach to variable selection in least squares problems, IMA J. Numer. Anal. 20 (3) (2000) 389–403.
[34] X. Peng, L. Zhang, Z. Yi, Scalable sparse subspace clustering, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013.
[35] Y. Taigman, L. Wolf, T. Hassner, Multiple one-shots for utilizing class label information, in: BMVC, 2009, pp. 1–12.
[36] A. Martinez, R. Benavente, The AR Face Database, 1998.
[37] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV), 1994, pp. 138–142.
[38] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[39] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.
[40] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, Liblinear: a library for large linear classification, J. Mach. Learn. Res. 9 (2008) 1871–1874.
[41] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[42] M. Grant, S. Boyd, Graph Implementations for Nonsmooth Convex Programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, Springer, London, 2008, pp. 95–110.
[43] Y. Su, S. Shan, X. Chen, W. Gao, Hierarchical ensemble of global and local classifiers for face recognition, IEEE Trans. Image Process. 18 (8) (2009) 1885–1896.
[44] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Trans. Image Process. 11 (4) (2002) 467–476.
[45] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European Conference on Computer Vision, Springer, 2004, pp. 469–481.
[46] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, B. Vijayakumar, Partial holistic face recognition on FRGC-II data using support vector machine, in: Proceedings of Computer Vision and Pattern Recognition Workshop, 2006, pp. 48–53.
[47] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face recognition by humans: nineteen results all computer vision researchers should know about, Proc. IEEE 94 (11) (2006) 1948–1962.
[48] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognit. 43 (1) (2010) 331–341.

Xi Peng is currently a Ph.D. candidate in the machine intelligence laboratory, Sichuan University. His research interests included representation learning and object recognition in supervised and unsupervised way.

Lei Zhang received the B.S. and Masters degrees in mathematics and the Ph.D. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2002, 2005, and 2008, respectively. She was a Post-Doctoral Research Fellow in the Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, Hong Kong, from 2008 to 2009. Currently, she is a Professor at Sichuan University, Chengdu. Her current research interests include theory and applications of neural networks based on neocortex computing.

Zhang Yi received the Ph.D. degree in mathematics from the Institute of Mathematics, Chinese Academy of Science, Beijing, China, in 1994. He is currently a Professor at Sichuan University, Chengdu, China, and acts as the Dean of the College of Computer Science, Sichuan University. He is the Founder of the Machine Intelligence Laboratory, College of Computer Science, Sichuan University. He has published more than 130 scientific articles in neural network, machine intelligence and data mining. Currently, he is an associate editor of IEEE Transactions on Neural Networks, the editorial board member of several famous journals, e.g. Cognitive Neurodynamics. His current research interests include machine intelligence, neural networks, data mining, machine learning, and neocortex computing toward intelligent machines.

Kok Kiong Tan received the B.Eng. degree in electrical engineering with honors in 1992 and the Ph.D.degree in 1995, all from National University of Singapore. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, National University of Singapore. His current research interests are in the areas of advanced control and auto-tuning, precision instrumentation and control, and general industrial automation.