Locality constrained representation based classification with spatial pyramid patches

Fumin Shen *,1, Zhenmin Tang, Jingxiong Xu

School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, PR China

A R T I C L E   I N F O

Article history:
Received 26 April 2012
Received in revised form 2 August 2012
Accepted 2 August 2012
Communicated by L. Shao
Available online 23 September 2012

Keywords:
Face recognition
Linear representation
Locality constraint

A B S T R A C T

In this work, we propose a linear representation based face recognition (FR) method incorporating locality information from both spatial features and training samples. Instead of holistic face images, the proposed method is conducted on the spatial pyramid local patches, which are aggregated by a Bayesian based fusion method. The locality constraint on the representation coefficients leads to an approximately sparse representation, which effectively explores the discriminative nature of spatial local features. Different from the sparse representation based classification (SRC) exposing an \( \ell^1 \)-norm constraint on the coefficients, the proposed locality constrained representation based classification (LCRC) is formulated with a computationally efficient \( \ell^2 \)-norm. The proposed method is robust to two crucial problems in face recognition: occlusion and lack of training data. A simple locality based concentration index (LCI) is defined to measure the reliability of each local patch, by which not only the heavily corrupted patches but also the less discriminant ones are rejected. Due to the use of both local patches and the locality constraint, less training data are required by the proposed method. Based on the locality constrained representation, we present three algorithms which outperform the state-of-the-art on the AR and Extended Yale B datasets for both the occlusion and single sample per person (SSPP) problems.

1. Introduction

Linear representation based face recognition methods attract a lot of interests recently due to its efficacy and simplicity. These methods are based on the assumption that a high-dimensional probe face image lies on a low-dimensional subspace spanned by the training samples of the same subject [1]. The decision is made by minimizing the residuals of reconstructing the probe face by a linear combination in terms of training samples with a set of coefficients. In practice, however, these methods do not perform well enough when training samples of each class are not sufficient to model various potential facial variations, e.g., changes of expression, illumination, occlusion, etc. Recently sparse representation based classifier (SRC) [2] has obtained a breakthrough performance. The proposed method is conducted on the spatial pyramid local patches, which are aggregated by a Bayesian based fusion method. The locality constraint on the representation coefficients leads to an approximately sparse representation, which effectively explores the discriminative nature of spatial local features. Different from the sparse representation based classification (SRC) exposing an \( \ell^1 \)-norm constraint on the coefficients, the proposed locality constrained representation based classification (LCRC) is formulated with a computationally efficient \( \ell^2 \)-norm. The proposed method is robust to two crucial problems in face recognition: occlusion and lack of training data. A simple locality based concentration index (LCI) is defined to measure the reliability of each local patch, by which not only the heavily corrupted patches but also the less discriminant ones are rejected. Due to the use of both local patches and the locality constraint, less training data are required by the proposed method. Based on the locality constrained representation, we present three algorithms which outperform the state-of-the-art on the AR and Extended Yale B datasets for both the occlusion and single sample per person (SSPP) problems.

Crown Copyright © 2012 Published by Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail addresses: fumin.shen@gmail.com (F. Shen),
tzm.cs@mail.njust.edu.cn (Z. Tang), xjsxjingsong@gmail.com (J. Xu).

1 His contribution was made when visiting The University of Adelaide.
classification. SPM partitions an image into increasingly fine sub-regions where histograms of local features are computed. Inspired by SPM [8], in this paper we subdivide each image into local patches at different spatial pyramid levels. Then the proposed method is conducted on these patches, by which both the holistic (corresponding to the first level) and local features with increasingly fine resolutions can be taken into classification. A Bayesian based fusion method is then proposed to aggregate the intermediate results with respect to these patches. The Bayesian method is based on the assumption that patches within a face are independent to each other, for simplicity.

In this work, we explore the discriminative nature of locality constrained representation (LCR) of local patches for identifying faces. For local patches, the residual gap between different subjects obtained by the aforementioned L^2-based methods is small when face images suffer from severe distortion, the test image is possibly far from some training samples (even from the same class). The locality constraint encourages the coefficients with respect to nearby samples and simultaneously penalizes the coefficients corresponding to distant ones, which forces the representation discriminant (see examples in Figs. 1 and 3).

Unlike SRC computed by the L^1-minimization, the proposed LCR based classification (LCRC) is formulated with a weighted ridge regression problem.

It is well known that the conventional L^2-minimization usually result in dense solutions. However, we show that, with the locality constraint, the L^2-norm can also lead to a sparse representation. In [9], the authors argue that locality is more essential than sparsity since sparsity dose not necessarily lead to locality but locality always incurs sparsity. Observing that, a classifier based on the sparsity of the coefficients (denoted as LCR-Spr) is presented. The discriminant nature of the locality constraint is validated by the high accuracy of LCR-Spr, which is very close to (sometimes even better) the corresponding residual based LRC.

Taking advantage of the locality constraint, large representation coefficients are concentrated on a small number of entries, which are expected to mainly fall in the same class. Based on that we also represent a class based algorithm C-LCRC, which computes the representation coefficients from one class each time. With a smaller training data matrix, C-LCRC is more efficient.

The method described in this paper effectively addresses two crucial problems in face recognition:

Occlusion. The presence of contiguous occlusion is one of the most challenging problems in the context of robust face recognition. Human may easily recognize a familiar person wearing sunglasses or scarves; however, it is a hard job for a computer to automatically make a correct identification on an obstructed facial image. For linear representation based methods, outliers incurred by occlusion may dramatically bias the regression model and results in a bad representation. The spatial pyramid partition and Bayesian fusion method proposed in this paper can significantly ignore the influence caused by occlusion. In addition, a locality based concentration index (LCI) is defined to measure the reliability of local patches, by which not only non-face patches but also the less discriminant ones (generic to many subject) are rejected.

Lack of training samples. In some real face recognition applications, very few or even only single sample per person (SSPP) is available. The LR based methods (e.g., LRC, SRC) using holistic facial features become unstable in this situation since they do not have enough samples to represent the incoming test image, which make the residual large even for the correct subject. The fact much less inherent facial variations exist in a local patch together with the locality constraint make it possible that much less samples are necessary for our method to cover these variations. Moreover, the proposed Bayesian fusion method can effectively preserve most of the discriminant information. This is verified by our experiments in Section 5.

The remainder of the paper is organized as follows: In Section 2, a brief discussion of related linear representation based methods is given. The proposed method LCRC is described in Section 3 and another two related algorithms are developed in Section 4. In Section 5 the proposed three algorithms and several other methods are evaluated on the AR and Extended Yale B databases. Finally the conclusion and discussion are offered in Section 6.

2. Related works

In face recognition community, linear representation based methods have been widely used due to their effectiveness and simplicity. These LR methods are based on the assumption that any probe image lies on a low-dimensional subspace [10,11], and the subspace is spanned by samples from the same subject [11]. The similar idea was previously used in nearest linear combinations (NLC) [11] and nearest feature line (NFL) [12]. Suppose we have a data matrix \( \mathbf{A} \in \mathbb{R}^{m \times n} \) containing the gallery face images from all the C classes with each image in a column vector, then a probe image \( \mathbf{y} \in \mathbb{R}^{m} \) belonging to the ith subject can be approximately represented as a linear combination of samples from the same subject:

\[
\mathbf{y} = \mathbf{A} \mathbf{x}_i, \quad i = 1, 2, \ldots, C, \tag{1}
\]

where \( \mathbf{x}_i \in \mathbb{R}^{n} \) is the coefficient vector and \( N_i \) is the number of training samples of the ith class. After seeking a linear representation of the test image \( \mathbf{y} \) with respect to each class, the nearest subspace (NS) methods [11,13,14] assign \( \mathbf{y} \) as the class with the smallest residual:

\[
\text{identity}(\mathbf{y}) = \arg \min_{1 \leq i \leq C} r_i, \quad i = 1, 2, \ldots, C, \tag{2}
\]

where \( r_i = \| \mathbf{y} - \mathbf{A} \mathbf{x}_i \|_2 \) is the residual with respect to class \( i \) and \( \cdot \|_2 \) denotes L^2-norm. Based on this assumption many variants have been suggested [2,4,7,14].

To obtain the coefficient vector, NLC [11] directly solve the following least squares problem:

\[
\min_{\mathbf{x}_i} \| \mathbf{y} - \mathbf{A} \mathbf{x}_i \|_2, \quad i = 1, 2, \ldots, C. \tag{3}
\]

In NLC [11], the coefficient vector is obtained by a pseudo-inverse matrix. Similarly in [7], the author also formulated face recognition as the above linear regression problem, hence termed as linear regression classification (LRC), which has a closed-form solution \( \hat{\mathbf{x}}_i = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \).

Recently sparse representation classification (SRC) presented in [2] achieves the state-of-the-art performances for face recognition. It incorporates the compressive sensing technique into the LR method. Unlike other popular classification methods in face recognition, all images in the training set (not only one class each time) are used to represent the query image. The SRC problem writes

\[
\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \| \mathbf{x} \|_1 \text{ s.t. } \mathbf{y} = \mathbf{A} \mathbf{x}, \tag{4}
\]

where \( \| \cdot \|_1 \) denotes L^1-norm and \( \mathbf{x} \in \mathbb{R}^{n} \). During the classification phase the residual with respect to subject \( i \) is defined as

\[
r_i = \| \mathbf{y} - \mathbf{A} \hat{\mathbf{x}}_i(\mathbf{x}) \|_2, \tag{5}
\]

where \( \hat{\mathbf{x}}_i(\mathbf{x}) \in \mathbb{R}^{n} \) is a new vector whose only nonzero entries are the entries in \( \mathbf{x} \) with respect to class \( i \). The L^1-norm is utilized to force the representation coefficients sparse, which means only a small number of samples truly participate in the representation. To deal with small dense noise, model (4) is then modified by
solving the following problem:
\[
\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_2 \quad \text{s.t.} \quad \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_2 \leq \epsilon. \tag{6}
\]
where \( \epsilon > 0 \) is the error tolerance. This method is shown to be very robust to random pixel corruption [2]. It is well known that with an appropriate parameter \( \lambda \geq 0 \) we can rewrite (6) to an equivalent unconstrained form:
\[
\min_{\mathbf{a}} \| \mathbf{y} - \mathbf{A} \mathbf{a} \|_2^2 + \lambda \| \mathbf{a} \|_1, \tag{7}
\]
which is known as the lasso problem or basis pursuit denoising problem in signal processing community. Several efficient algorithms for the lasso are available, and [15] provide a comparative study. In the flowing sections, we refer to (7) instead of (6) as the SRC problem.

More recently a similar method with SRC is suggested in [3], where the difference is that the \( \ell^1 \)-regularization is ignored. Due to the use of only \( \ell^2 \)-norm, this method can efficiently deal with high dimensional data (more than 10,000). Another method proposed in [4] called collaborative representation based classification (CRC) simply replace the \( \ell^1 \) constraint in SRC (7) with the \( \ell^2 \) constraint, and obtains comparative performances using eigenfaces [16]. It writes
\[
\min_{\mathbf{a}} \| \mathbf{y} - \mathbf{A} \mathbf{a} \|_2^2 + \lambda \| \mathbf{a} \|_2^2, \tag{8}
\]
which can also be efficiently solved by an analytical solution \( \mathbf{x} = (\mathbf{A}^{\top} \mathbf{A} + \lambda I)^{-1} \mathbf{A}^{\top} \mathbf{y} \), where \( I \in \mathbb{R}^{n \times n} \) is the identity matrix.

Besides the fact that all these three methods utilize an efficient \( \ell^2 \)-norm instead of the \( \ell^1 \)-norm, they achieve comparative (or even better) performances with SRC. Moreover, [3] argues that the sparsity assumption for SRC is not supported by the data and [4] argues that it is the collaborative representation but not the \( \ell^1 \)-norm sparsity constraint that in fact improve the face recognition performance. Yang et al. [17] give a reasonable support for the effectiveness of SRC; however, argue that it is closeness but not sparsity achieved by the \( \ell^1 \)-optimizer that guarantees the effectiveness of SRC. In this work, we will re-examine the role of \( \ell^1 \) and \( \ell^2 \) in classification.

Although these methods mentioned above can effectively handle small noise, they still do not perform well on datasets with severe contiguous occlusions. With images partitioned into several blocks [7,2], both LRC and SRC acquired much better results. However, both the fusion method distance-based evidence fusion (DEF) in [7] and the voting strategy in [2] lose much discriminant information. In this paper, we show that a better fusion method can significantly improve the performance.

Recently several methods consider face recognition with noisy data as robust regression problems, where the squared residuals are replaced with a robust function (such as Huber loss function in [18]). In [19,20] the robust maximum correntropy criterion (a special case of M-estimator), which can effectively deal with non-Gaussian noise, was developed in the regularized sparse representation framework. Both two-stage sparse representation (TSR) [21] and robust sparse coding (RSC) [22] iteratively learn a robust metric to suppress the influence caused by outliers. Good results have been obtained by both these two methods on several datasets. In this paper we focus on the linear method and the robust version can be easily extended. There are several other previous methods related in this category, such as the nearest feature line (NFL) method proposed by Li and Lu [12]. A brief review of these linear representation methods is given in [19].

The proposed method for face recognition is mainly based on the Bayesian fusion of spatial pyramid features. In addition to SPM [8], pyramid methods have been widely used in computer vision problems. For example, a pyramid feature descriptor called PHOG was developed in [23] based on SPM and the histogram of gradient orientation (HOG) [24]. Recently another pyramid feature descriptor PCOG was proposed in [25], which consists of a correlogram of orientation gradients over sub-regions at different resolution levels. PCOG was applied in [25,26] for the human motion classification and action segmentation problems.

Our method is largely inspired by the following two works. A coding scheme called locality-constrained linear coding (LLC) was recently proposed in [27] and achieved impressive performances with a linear SVM classifier for image classification. LLC is a fast implementation of the local coordinate coding (LCC) [9] which approximates the high dimensional nonlinear function by a global linear function with respect to a local coordinate coding scheme. The main difference between our method and these two methods is that they are formulated for the image coding or nonlinear function learning problem while our method is to explore the discriminative nature of locality for local patches for face recognition.

3. Locality constrained representation for spatial pyramid patches

3.1. Locality constrained linear representation

Recall that in the SRC formulation (7), the same weight parameter \( \lambda \) is used for all regression coefficients. However, this constraint dose not necessarily hold in practice.\(^2\) Intuitively the coefficients corresponding to less relevant predictors should be penalized, whereas the most relevant predictors should be well kept in the regression model. Penalizing the coefficients by dissimilarity between the probe image and the training samples provides a meaningful way, which writes
\[
\min_{\mathbf{a}} \| \mathbf{y} - \mathbf{A} \mathbf{a} \|_2^2 + \lambda \| \mathbf{W} \mathbf{a} \|_1, \tag{9}
\]
where \( \mathbf{W} \) is a diagonal matrix with its ith diagonal entry \( w_i \) the distance \( d(y,x_i) \):
\[
d(y,x_i) = \frac{\| y - x_i \|_2^2}{\max_i \{ \| y - x_i \|_2^2 \}}, \quad i = 1, \ldots, n. \tag{10}
\]
Here \( x_i \) is the ith column of \( \mathbf{A} \) representing a training sample. The widely used heat kernel function [31,32,27] \( \exp(\| y - x_i \|_2^2/\sigma) \), where \( \sigma > 0 \) is the kernel size, can also be used as the distance. However, one may need to choose the parameters \( \lambda \) and \( \sigma \) carefully such that only the most reconstructive neighbours of the test image can be well kept. We find in practice (10) works well and with that the algorithm is not sensitive to \( \lambda \), which is always set as 1 in our experiments.

It is clear that with the weight parameter \( \mathbf{W} \), coefficients with respect to training samples far away from the test image are penalized heavily and encouraged to be zeros, and those coefficients with respect to samples close to the test image will be well kept. This is reasonable because, intuitively in a subspace spanned by training samples, the query data point is more possibly well and with that the algorithm is not sensitive to \( \lambda \), which is always set as 1 in our experiments.

Another variable selection method adaptive lasso [29] assign the weight parameter \( w_i \) in (9) as \( 1/\| x_i \|_2^2 \), where \( \theta > 0 \) and \( x_i \) is the ith column of the least square (LS) solution \( \hat{x} = (\mathbf{A}^{\top} \mathbf{A})^{-1} \mathbf{A}^{\top} \mathbf{y} \). With an appropriate selection of \( \theta \) and \( \lambda \), the adaptive lasso has the

\(^2\) Actually this constraint may make the lasso estimators largely biased [28] (toward zeros). In general the lasso shrinkage is not consistent [29,30]. More theoretical analysis can be found in [28–30].
especially there exist occlusions or corruptions in face images.

Like SRC (7), however, with the use of \( \ell^1 \)-norm problem (9) is not very efficient in practice for face recognition systems. Several previous works [3,4] argue that \( \ell^1 \)-norm is not necessary for classification, therefore in this paper we relax (9) with \( \ell^2 \)-norm in the constraint as follows:

\[
\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_2 + \lambda \| \mathbf{W} \mathbf{x} \|_2^2.
\] (11)

Formulation (11) is a weighted ridge regression problem which can be efficiently solved with a closed-form solution:

\[
\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{W}^T \mathbf{W})^{-1} \mathbf{A}^T \mathbf{y}.
\] (12)

### 3.1.1. An Bayesian interpretation

Following an idea from [34], we give the above problem a Bayesian interpretation. With a Laplacian prior assumption on the parameter \( \mathbf{x} \), i.e.,

\[
f(\mathbf{x}) = (1/2\gamma) \exp(-\|\mathbf{x}\|_1/\gamma),
\]

where \( \gamma = 1/\lambda \mathbf{w} \), (9) is actually equivalent to a maximum a posterior (MAP) estimation of a linear model with Gaussian distributed noise error. Similarly if we relax the Laplacian prior assumption on the parameter to a Gaussian one, we arrive at the weighted ridge regression problem (11). With a smaller scale parameter \( \gamma \) (corresponding to a larger distance \( \mathbf{w} \), in (10)), the Gaussian density function put more mass near zero, which makes the training samples far away form the probe image more possibly associated with zero or near zero coefficients. Therefore the weighted ridge regression problem (9) is expected to produce an approximately sparse estimation which will be shown in Fig. 1.

### 3.1.2. Comparison with other linear representation methods

Next let us compare the locality constrained representation (LCR) and other two related methods: sparse representation (SR) [2] and collaborative representation (CR) [4] through an example. Fig. 1 shows a face image from the first subject of the AR dataset and its representation coefficients in terms of all training samples (see description in Section 5.3) computed by different methods. It is clearly seen that although both LCR and CR have a weaker sparse representation than SRC, all these three methods correctly concentrate the largest coefficients on the correct subject. Note-worthy is the coefficients obtained by LCR are sparser than that by CR due to the locality regularization. Here ‘sparse’ means most training samples are associated with nearly zero (not exactly zero) coefficients and only a small number of samples (of subject 1 in the example) are with large coefficients (the largest coefficient for CR and LCR is about 0.22 and 0.4, respectively).

### 3.2. Spatial pyramid local patches and Bayesian based fusion

Face recognition becomes far more challenging in the presence of occlusions, which may dramatically bias the estimations through traditional techniques, such as least squares (LS). The inverse effect of occlusion can be significantly alleviated by utilizing the spatial information of face images. In this work, the proposed method in the previous section is conducted on the spatial pyramid local patches. SPM [8] partitions an image into increasingly fine sub-regions where histograms of local features are computed. Similarly we subdivide each image into \( 2^l \times 2^l \) non-overlapping blocks at different levels, \( l = 0, 1, \ldots, L \). Then totally \( T = \sum_{l=0}^{L} 4^l \) patches are generated from each image. We refer to the corresponding training data and test sample of the \( j \)th patch from level \( l \) as \( \mathbf{y}^{(l,j)} \) and \( \mathbf{y}^{(l)} \). Fig. 2 illustrates a three-level pyramid for partitioning a face image. With the precomputed locality weight matrix \( \mathbf{W}^{(l)} \), each patch is processed independently,

\[
\hat{\mathbf{x}}^{(l)} = \min_{\mathbf{x}} \| \mathbf{y}^{(l)} - \mathbf{A}^{(l)} \mathbf{x} \|_2 + \lambda \| \mathbf{W}^{(l)} \mathbf{x} \|_2.
\] (13)

For matching problems SPM aggregates all levels by weighting each level \( l \) with \( 1/2^{L-l} \), which is inversely proportional to its level width. Through that a smaller weight is associated with a larger sub-region which involves increasing dissimilar features [8]. For face recognition, we aggregate these levels in another way. All the patches are first downsampled into the same size and then aggregated with their corresponding reconstruction errors (residuals). This is more straightforward because it is residuals that

---

**Fig. 1.** Representation of a downsampled 20 \( \times \) 20 image from subject 1 in the AR dataset by LCR, CR, and SR. (a) Coefficients and (b) residuals with respect to training samples in different subjects. The ratio of the two smallest residuals is shown on the top of each chart of (b).
we have \( T \) from class follows:

maximum a posteriori (MAP) estimation of the class label

which is produced by the procedure of the proposed method in Algorithm 1.

This is similar to (14) and actually in practice we find these two situations showing non-Gaussian noise and kernel function to form a correntropy-based sparse model, which is identity should be well kept. A heuristic method to measure the score of patches is obtained: 1 from level 0, 4 from level 1, and 16 from level 2. Each illustration of image partition in a three-level spatial pyramid. Totally 21 blocks are obtained: 1 block: 1 − 5 block: 6 − 21

For simplicity, we assume that the patches within a face are concentrated on only a small fraction of entries and all other coefficients are zeros or nearly zeros. In Section 3.4 we describe an effective way to automatically reject the invalid patches due to the locality information.

3.3. Sparsity induced by locality for local patches and its discriminant nature

Let us first see an example demonstrating the effectiveness of LCR for local patches. Fig. 3 shows a comparison of LCR, CR, and SR for both a clean patch and a corrupted one from pyramid level 2 of a 20 \( \times \) 20 dimensional image as in Fig. 2 (see settings in Section 5.2). For the clean patch, Fig. 3a shows that both LCR and SR have a sparse\(^3\) representation and concentrate the large coefficients on a few entries. In contrast, CR has a dense representation where

\[ \hat{c} = \max_i \prod_{t=1}^T P(M_t | c_i) \]

Here \( P(M_t | c_i) \) represents the likelihood of the patch \( M_t \) which is from class \( i \). It is natural that we can model the likelihood by the corresponding residual as in (14), \( P(M_t | c_i) = \exp(-\beta r_i^T) \), so we obtain

\[ \hat{c} = \max_i \prod_{t=1}^T \exp(-\beta r_i^T) \]

which is

\[ \hat{c} = \max_i \exp \left( -\beta \sum_{t=1}^T r_i^T \right) . \]

This is similar to (14) and actually in practice we find these two fusion methods achieve very similar performances. We describe the procedure of the proposed method in Algorithm 1.

\begin{algorithm}
\caption{Locality constrained representation based classification (LCRC).}
1: \textbf{Input}: Partition each image into \( T \) local patches at different spatial pyramid levels as described in Section 3, and we get the training data matrices \( A^{(i)} \in \mathbb{R}^{m \times n}, t = 1, \ldots, T \) and test patch vectors \( y^{(i)} \in \mathbb{R}^n, t = 1, \ldots, T \). Set the regularization parameter \( \lambda > 0 \) and the scale parameter \( \beta > 0 \).
2: Normalize \( y \) and columns in \( A^{(i)} \) to be \( l^2 \)-norm unit vectors.
3: For each patch, compute the diagonal locality matrix \( W^{(i)} \) with its entries: \( w^{(i)}_{ij} = \| y^{(i)} - x^{(i)} \|_2 / Q, i = 1, \ldots, n \). Here \( Q \) normalizes \( W^{(i)} \) to have maximum entry value 1.
4: For each patch, compute the representation coefficients: \( z^{(i)} = (A^{(i)} A^{(i)} + \lambda W^{(i)} W^{(i)})^{-1} A^{(i)} y^{(i)} \), and the residuals corresponding to different classes \( r_i \) by Eq. (15). (For datasets with heavy occlusions, first discard those unreliable patches using the validation method described in Section 3.4.)
5: \textbf{Output}: Identify \( y \) by (16) or (21).
\end{algorithm}

It is not surprising that by partitioning images into patches at different pyramid levels one can obtain a more robust estimation, since occlusion existing in one patch will not affect estimations on other patches. For an incoming test image with occlusion (see Fig. 2 for example), we get several patches (14 patches in the example) without or very small occlusion. With these ‘clean’ patches, one can obtain a more accurate estimation than that just using the whole image. In addition, by this partition scheme and fusion method both holistic information (from level 0) and increasingly local information (from sub-regions at level 1 to L) are taken into account for classification.

To eliminate the impact of sunglasses and scarves occlusion, the modular approach is used in [27], which simply partitions the image into blocks and then aggregate results of these individual blocks by majority voting or the competing method distance-based evidence fusion (DEF). However, both these two strategies only use information from part of these blocks for classification. For the voting method [2], all blocks which lead to dissimilar class labels with the majority one in the classification phase are discarded. If the image is heavily corrupted, it is likely that the clean patches are discarded, which makes voting unstable. Moreover this method gets the final decision based on the intermediate decisions (class labels) instead of the more informative residuals. For the DEF method [7], which actually use only one block with the smallest residual and useful information from all other blocks is lost.

Indeed it is necessary to reject the heavily corrupted patches and in the meanwhile to effectively fuse information from the remaining ones. In Section 3.4 we describe an effective way to automatically reject the invalid patches due to the locality information.

\[ \hat{c} = \max_i \prod_{t=1}^T P(M_t | c_i) \]

Here \( P(M_t | c_i) \) represents the likelihood of the patch \( M_t \) which is from class \( i \). It is natural that we can model the likelihood by the corresponding residual as in (14), \( P(M_t | c_i) = \exp(-\beta r_i^T) \), so we obtain

\[ \hat{c} = \max_i \prod_{t=1}^T \exp(-\beta r_i^T) \]

which is

\[ \hat{c} = \max_i \exp \left( -\beta \sum_{t=1}^T r_i^T \right) . \]

This is similar to (14) and actually in practice we find these two fusion methods achieve very similar performances. We describe the procedure of the proposed method in Algorithm 1.
variation of the coefficients with respect to different subjects is small, and the largest coefficient is only 0.027 which is far smaller than that of SR (0.34) and LCR (0.44). This observation demonstrates that besides the \(\ell^1\)-norm constraint, the \(\ell^2\)-norm constrained by locality can also lead to sparsity. And locality results in sparse representations in a more natural way: test images are more likely to be represented by their neighbours.

Furthermore, localization of the representation coefficients is also helpful in classification. From Fig. 3b we can see that LCR obtains a larger ratio of the two smallest residuals (5.34) than that by SR (2.72) and CR (1.06). Consistent with the example shown in Fig. 1, this result further shows the discriminative nature of locality. Although CR correctly associates its smallest residual with the test subject (subject 1), the gap between residuals corresponding to different subjects is very small. Using only the \(\ell^2\)-norm regularization without locality constraint CR does not perform as well as LCR.

On the heavily corrupted patch, all these methods fail. The dense coefficients (Fig. 3c) provide little information for classification, which is validated by the corresponding residuals shown in Fig. 3d. Apparently the heavily corrupted patch is not reliable for classification because it may be relatively closer to an unrelated subject (see the residual for subject 38 in Fig. 3d). We next describe an effective way to measure the reliability of a patch through the locality information.

### 3.4. Validation based on locality concentration index

In [2], Wright et al. present a sparse representation based validation method, which rejects an invalid image if the proposed sparsity concentration index (SCI) of its coefficient vector \(\mathbf{a}\) is below a threshold. This method is based on the argument that a valid test image should have a sparse representation whose nonzero entries concentrate mostly on one subject [2]. Similarly, we assume that a valid patch should be close to some samples belonging to the same class and far away from those in other classes. A local patch far away from (or not close to) patches of any subject is expected to be less helpful for classification and should be discarded. This usually happens when a local patch is heavily corrupted. With the precomputed locality vector \(\mathbf{w} = \text{diag}(\mathbf{W})\) of a test image, the following locality concentration index (LCI) is defined to measure the reliability of an image (patch):

**Definition 1 (locality concentration index (LCI)).** The LCI of a locality vector \(\mathbf{w} \in \mathbb{R}^d\) is defined as

\[
\text{LCI}(\mathbf{w}) = 1 - \frac{C \cdot \min_{i \in [N]} |d_i(\mathbf{w})|_1}{\|\mathbf{w}\|_1} \in [0, 1].
\]

If LCI(\(\mathbf{w}\)) = 0, the test image is evenly far away from (or close to) all classes, and if LCI(\(\mathbf{w}\)) is nearly 1, the test image will be very close to images from at least one subject.\(^4\) An image or patch is valid if its LCI is greater than a threshold.\(^4\)

---

\(^4\) Note that in the second situation, the local patch is possibly very close to more than one subject and in that case this patch seems not discriminant between these subjects. However, this patch is expected to be effectively represented by the training patches from these nearby subjects and this patch is also taken into classification.
regarded invalid and rejected if
\[ \text{LCI}(\mathbf{w}) < = \tau, \]
where \( \tau > 0 \) is the threshold.

Recall the example in Fig. 2, all the training data and the face image with sunglasses disguise are first resized to a resolution of \( 20 \times 20 \) and then subdivided into local patches in a 3-level spatial pyramid. The locality vector \( \mathbf{w} \) of each patch against the corresponding training patches of all classes is computed. Fig. 4 shows the LCI values of these totally 21 locality vectors. We can clearly see that the patches with heavily corruptions are all associated with low LCI values. Before classification rejecting the unreliable patches with low LCI values will be helpful to robust face recognition. In Section 5.2 improved accuracies are obtained using this method on the AR dataset with sunglasses and scarf occlusions. Note that LCI values for some patches without corruptions (patches 15 and 16 in the example) are also possibly low, and that is because the locality variation of this patch between some subjects is relatively small, which means these patches are less discriminant even they are ‘clean’.

4. Classification based on locality constrained representation

Besides the algorithm shown in Section 3.2, in this section we propose another two algorithms based on the locality constrained representation.

4.1. Sparsity decision rule based classification

As shown in previous sections, localization of the representation coefficients provides useful information for classification. In this section we design a classifier directly based on the locality induced sparse coefficients. Ideally the coefficient vector \( \mathbf{a} \) of a probe image \( \mathbf{y} \) should concentrate a small number of its largest entries on the training samples from the same subject. Based on that, we assign \( \mathbf{y} \) to the class with the largest coefficient vector in terms of \( \ell^2 \)-norm:
\[ \text{identity}(\mathbf{y}) = \max_i |\delta_i(\mathbf{a})|_2, \quad i = 1, \ldots, C. \]

For the patch based algorithm, we can easily modify Algorithm 1 by substituting steps 5 and 6 with the output.
\[ \text{identity}(\mathbf{y}) = \max_i \sum_{t=1}^T |\delta_i(\mathbf{a}^{(t)})|_2, \quad i = 1, \ldots, C. \]

We refer to this algorithm as LRC-Spr in the following sections. In Section 5, we will show that the classifier perform close to (and in some cases even slightly better than) the residual based one. As the example illustrated in Section 3, this again shows the discriminative nature of locality in face recognition.

4.2. Classification using data from homo-class

Most of the methods we discussed in the former sections are based on collaborative representation, i.e., taking training samples from all subjects to reconstruct the probe image. This method is helpful especially when training data size of each class is small, which takes advantage of the fact that face images from different subject share similarities [4]. A comparative study about the relationships of collaborative (sparse) representation based classification with the class based nearest neighbour (NN) and nearest subspace (NS) is given in the supplementary material of [2].

Different from these methods and the method described in Section 3, we propose another algorithm based on homo-class classification. As mentioned above, LCRC need much less training samples due to the use of local patches. Moreover, the locality constraint effectively concentrate the large representation coefficients of a valid test image (patch) on its neighbours, which are expected to mainly fall in the same class (see the example shown in Fig. 3). We will see that directly reconstructing the test image by samples from only one class each time does not significantly affect the performance of LCRC in most cases. Given training data of each class \( \mathbf{A} \), and a test image \( \mathbf{y} \), the class based locality constrained representation based classification (C-LCRC) writes
\[ \hat{\mathbf{z}}_i = \arg \min_{\mathbf{x}} |\mathbf{y} - \mathbf{A}_i \mathbf{x}|_2^2 + \lambda |\mathbf{W}_i \mathbf{x}|_2^2, \quad i = 1, \ldots, n, \]
where \( \mathbf{W}_i \) is the locality matrix based on class \( i \). Then we get the residual with respect to class \( i \),
\[ r_i(\mathbf{y}) = |\mathbf{y} - \mathbf{A}_i \hat{\mathbf{z}}_i|_2, \quad i = 1, \ldots, n. \]

Similar as (14), the patch based algorithm can be easily obtained by modifying (27) as
\[ s_i(\mathbf{y}) = \sum_{t=1}^T \exp(-\beta |\mathbf{y}^{(t)} - \mathbf{A}_i^{(t)} \hat{\mathbf{z}}^{(t)}_i|_2), \quad i = 1, \ldots, n, \]
where \( s_i(\mathbf{y}) \) is the aggregated score of \( \mathbf{y} \) with respect to class \( i \). Here \( \mathbf{A}_i^{(t)} \) and \( \mathbf{y}^{(t)} \) are the partitioned training and test data of patch \( t \), and \( \hat{\mathbf{z}}^{(t)}_i \) is the corresponding estimated coefficients. After getting scores corresponding to all classes, \( \mathbf{y} \) is then assigned to the class with the highest score.

In Section 5, we will show that C-LCRC also achieves high accuracies on databases in various conditions. In addition, by solving a set of small-size problems instead of a large problem, the proposed C-LCRC becomes even more efficient than LCRC.
Fig. 5 shows a comparison of the running time various methods need to recognize a face from 38 subjects enrolled in the Extended Yale B database with increasing dimensions: 25, 50, 100, ..., 6400. It is clearly seen that the class based algorithm is much faster especially on high dimensional data. Note that SRC is implemented using the efficient Homotopy method [37], denoted as SRC-Hom.

5. Experimental results

In this section, we evaluate the proposed three algorithms: LCRC, LCRC-Spr, C-LCRC on public benchmark databases for face recognition. We will first demonstrate the robustness of our methods to contiguous occlusion: both artificial and natural. And then we will show the efficacy of the proposed method with insufficient training data or SSPS. Several state-of-the-art methods are also performed for comparison. As well as the methods discussed in the former section, we will also compare our methods with extended SRC [38], which is proposed most recently for FR problem with insufficient training samples. Apart from the original training samples, the method constructs an extra intra-class variant dictionary which also participates in the sparse representation. This method obtains superior results than the original SRC on several datasets (see details in [38]).

The code for TSR is obtained from the authors [21]. We implement SRC and extended SRC using Homotopy$^5$ and to its accuracy and efficiency [15] for the lasso problem (7). We set $\lambda = 0.001$ for these two algorithms. Due to the settings in [2,7], the modular methods for SRC and LRC are carried out with images partitioned into $4 \times 2$ blocks, and the voting and DEF algorithms are used to combined results for these blocks, respectively. For CRC, we set $\lambda = 0.001 \times n/700$ according to the authors [4]. We set $\lambda = 1$ for our methods in all experiments unless otherwise specified. We set $\beta$ as 1 for the proposed LCRC and LCRC-Spr and 10 for the class based C-LCRC. In all our experiments, we set three spatial pyramid levels, i.e., $L = 2$.

5.1. Face recognition with random block occlusion

In this section, we evaluate our methods on face recognition problem with artificial contiguous occlusion. Following [2], a square monkey face is placed on each test image at a random location which is unknown to the algorithms. Two occluded example samples are shown in Fig. 6. Specially we use the Extended Yale B database [40], which consists of 2414 frontal face images from 38 subjects under various lighting conditions. The images are cropped and normalized to 192 x 168 pixels [14]. Half of the images were randomly selected for training (i.e., about 32 images per subject), and the remaining half are for testing. In our experiment, the simple downsampled images are used for features, with resolution 40 x 40. In order to evaluate the performance of various methods on this data each was run on five sets of images with randomly placed occlusions from 20% to 50%. Experiments are also carried out on original data without occlusion for baseline. Recognition rates of different methods are reported in Table 1.

We can clearly see that the proposed three methods obtain the best results in all situations. It is noteworthy that with occlusions below 10% LCRC gets 100% recognition rate. When occlusion increases to 50% accuracy rate of LCRC is still above 88%, while the best result of all other methods is 73.1% obtained by SRC-voting. Both the two non-modular methods CRC and TSR do not perform well with large occlusions. LCRC-Spr and C-LCRC perform very close to LCRC except that when occlusion increases to 50%, the former two classifiers obtains around 87% accuracies which is lower than that of the third one by about 1%. The close performances of LCRC and LCRC-Spr show that a good representation is more important than the decision rule.

For fair comparison, we also carry out CRC, TSR and SRC with spatial pyramid features (CRC-SP, TSR-SP and SRC-SP in Table 1). We can see that with the spatial pyramid features the accuracies of these three methods are significantly improved. TSR-SP and SRC-SP also outperform the other two modular methods LRC-DEF and SRC-voting by large gaps especially when with large occlusions. However, they are still inferior to LCRC and its two extensions, which is mainly because the locality information indeed boosts the FR performance as mentioned above.

We also evaluate the performance of LCRC with $\ell^1$-norm regularization as in (9). We set $\lambda = 0.1$ for LCRC-$\ell^1$. As can be seen, LCRC-$\ell^1$ obtains a very close performance on this dataset and the $\ell^1$-norm regularization does not necessarily improve the accuracy for LCRC as stated before. This is consistent with the arguments in [3,4].

5.2. Face recognition with disguise

The AR database [41] consists of over 4000 facial images from 126 subjects (70 men and 56 women). For each subject 26 facial images were taken in two separate sessions. The images exhibit a

---

$^5$ The code is obtained via http://www.eecs.berkeley.edu/~software/l1bench mark/.

$^6$ In the classification phase, (16) and (20) yield a very similar result for LCRC and the reported results are based on [21].
decision fusion method is not robust enough for heavily corrupted cases. SRC-voting and LRC-DEF do not perform as well as our methods, which show the voting and DEF decision fusion method is not robust enough for heavily corrupted data as described in Section 3.2. With our pyramid features, performance of the robust method TSR is even improved, with very high accuracies 99% and 97% for the sunglasses and scarves cases, respectively.

To fairly compare these methods, SRC [2], CRC [4] and LCRC are also evaluated using both whole images and our spatial pyramid features. Individual results of SRC and CRC on local patches are aggregated by voting as in [2,4]. The comparative results are shown in Table 3. From Table 3, we can see that LCRC achieves the best recognition rates for both cases. Specifically, LCRC obtains a 66% accuracy which outperforms CRC and SRC by 8%. With the spatial pyramid features, accuracies of all methods are largely improved. However, accuracies of LCRC are still higher (by 2.5% and 3.5%) than the best results of the other two methods: SRC and CRC. The superior results of LCRC indeed show the discriminant ability of locality information for face recognition, especially on the occluded data.

Take the sunglasses case for example, Fig. 8 shows the impact of validation using LCI on local patches for LCRC (left) and LCRC-Spr (right) where each image is resized to 400 pixels. When the validation threshold \( t \) is set as 0.01 for LCRC, accuracies of all methods are largely improved. However, accuracies of LCRC are still higher (by 2.5% and 3.5%) than the best results of the other two methods: SRC and CRC. The superior results of LCRC indeed show the discriminant ability of locality information for face recognition, especially on the occluded data.

Table 1
Classification accuracy (%) on the Extended Yale B database with using 40 × 40 downsampled pixels. Monkey faces with various sizes are randomly placed on the test faces. Every result set: accuracy rate (standard deviation) is calculated based on 5 runs.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Occlusion rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>CRC</td>
<td>98.3 ± 0.1</td>
</tr>
<tr>
<td>TSR</td>
<td>98.5 ± 0.2</td>
</tr>
<tr>
<td>SRC-voting</td>
<td>99.6 ± 0.2</td>
</tr>
<tr>
<td>LRC-DEF</td>
<td>99.1 ± 0.2</td>
</tr>
<tr>
<td>CRC-SP</td>
<td>99.6 ± 0.1</td>
</tr>
<tr>
<td>TSR-SP</td>
<td>99.8 ± 0.1</td>
</tr>
<tr>
<td>SRC-SP</td>
<td>99.8 ± 0.1</td>
</tr>
<tr>
<td>LCRC</td>
<td>99.9 ± 0.0</td>
</tr>
<tr>
<td>LCRC-Spr</td>
<td>100 ± 0.0</td>
</tr>
<tr>
<td>C-LCRC</td>
<td>99.9 ± 0.0</td>
</tr>
</tbody>
</table>

Table 2
Comparison of different methods conducted on both whole images and our spatial pyramid local patches on the AR database using 40 × 40 downsampled pixels.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Using whole images</th>
<th>Using pyramid patches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SRC</td>
<td>CRC</td>
</tr>
<tr>
<td>Sunglasses</td>
<td>50</td>
<td>58</td>
</tr>
<tr>
<td>Scarf</td>
<td>54</td>
<td>82</td>
</tr>
</tbody>
</table>

number of variations including various facial expressions (neutral, smile, angry, and scream), illuminations (left light on, right light on and all side lights on) and occlusion by sunglasses and scarves. Of the 126 subjects available 100 have been randomly selected for testing (50 males and 50 females) and the images are cropped to 165 × 120 pixels. Eight images of each subject with various facial expressions but without occlusions were selected for training. Testing was carried out on two images of each subject wearing sunglasses and two wearing scarves. All the images are resized to a resolution of 40 × 40, and we simply use the raw pixels for input features.

Two test example images of subjects wearing sunglasses and two wearing scarves from the AR dataset are shown in Fig. 7. For LCRC and LCRC-Spr, we reject patches via (23) and \( t \) is chosen as 0.5. Recognition rates of various methods are summarized in Table 2. C-LCRC obtains 99.5% and 98% recognition rates for datasets with sunglasses and scarf disguises, respectively, which beats all the other state-of-the-art methods compared in this experiment. Both LCRC and LCRC-Spr achieve almost the same results as C-LCRC. Note that no extra spatial prior knowledge is known to the proposed approach. SRC-voting and LRC-DEF do not perform as well as our methods, which show the voting and DEF decision fusion method is not robust enough for heavily corrupted data as described in Section 3.2. With our pyramid features, performance of the robust method TSR is even improved, with very high accuracies 99% and 97% for the sunglasses and scarves cases, respectively.

Table 3
Classification accuracy (%) on the AR database with using 40 × 40 downsampled pixels. Monkey faces with various sizes are randomly placed on the test faces. Every result set: accuracy rate (standard deviation) is calculated based on 5 runs.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Occlusion rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>CRC</td>
<td>98.3 ± 0.1</td>
</tr>
<tr>
<td>TSR</td>
<td>98.5 ± 0.2</td>
</tr>
<tr>
<td>SRC-voting</td>
<td>99.6 ± 0.2</td>
</tr>
<tr>
<td>LRC-DEF</td>
<td>99.1 ± 0.2</td>
</tr>
<tr>
<td>CRC-SP</td>
<td>99.6 ± 0.1</td>
</tr>
<tr>
<td>TSR-SP</td>
<td>99.8 ± 0.1</td>
</tr>
<tr>
<td>SRC-SP</td>
<td>99.8 ± 0.1</td>
</tr>
<tr>
<td>LCRC</td>
<td>99.9 ± 0.0</td>
</tr>
<tr>
<td>LCRC-Spr</td>
<td>100 ± 0.0</td>
</tr>
<tr>
<td>C-LCRC</td>
<td>99.9 ± 0.0</td>
</tr>
</tbody>
</table>

Fig. 7. Images from two subjects in the AR database wearing sunglasses (left) and two wearing scarves (right).
5.3. Recognition from insufficient training samples

In this section we test our methods with insufficient training data using the AR database. We only use the data without occlusions (14 samples of each subject) in this test. For each subjects, the first seven images from session 1 are used for training and the first seven images from session 2 are for testing. All the images are resized to a resolution of $20 \times 20$, and the raw pixels are used for input features. Similar as the setting in [38], we reduce the number of training samples for each subject from 7 to 2 one by one. In this test all images are with only illumination and expression changes, therefore we take all patches into classification without patch validation. Extended SRC is also carried out for comparison, which constructs the intra-class variant dictionary by subtracting the centroid of each class form all images from the same class [38].

Fig. 9 shows the comparative recognition results of various methods. The proposed approach LCRC and LCRC-Spr achieve very similar performances, both of which outperform all other methods in all situations. In particular, when all the seven training samples per subject are available, LCRC achieves 98% accuracy which is higher than that of SRC-voting, LRC-DEF by 7.9% and 31.3%, respectively. With only two training samples per class, LCRC still achieves more than 80% recognition rate while accuracies of all other methods are below 64%. Not surprisingly, the class-based algorithm C-LCRC does not perform as well as the other two LCRC algorithms on this dataset, and the accuracy gap becomes larger as the training data size decreases. However, we can see that C-LCRC still outperforms other methods compared in this experiment. On this dataset, extended SRC [38] does not perform as well as the modular SRC, however much better than LRC-DEF. We also compare LRC-DEF with LRC and CRC which use data without partition, and the DEF algorithm performs even worse than the original LRC (in some cases) and CRC on this dataset. It is not surprising because there are no occlusion on this dataset and all blocks should participate in the final classification, while the DEF selects only one block (corresponding to the smallest residual).

5.4. Recognition from single sample per person

We next test the robustness of our method on the SSPP problem using the AR dataset. For each subject, the first image with natural expression and illumination from session 1 is used for training, and the rest 12 images with expression and illumination changes and sunglasses and scarves disguises from session 1 are for testing. We resize all the images to three different resolutions $20 \times 20$, $32 \times 32$, $40 \times 40$. For LCRC and LCRC-Spr, we set the same validation parameter $\tau$ as in Section 5.2. For C-LCRC, to eliminate the inverse effect of occlusion in the dataset, a large $\beta$ (40) is chosen. For extended SRC, the first 20 subjects in session 2 (260 images) are used to construct the intra-class variant dictionary by two ways [38]: (1) subtracting the centroid of each class form all images from the same class, and (2) subtracting the natural image form other images from the same class.

Table 4 shows the recognition rates of various methods with different feature dimensions. According to the table, LCRC-Spr and LCRC perform the best in all dimensional feature spaces. 

---

**Table 4**

<table>
<thead>
<tr>
<th>Approach</th>
<th>CRC</th>
<th>ExSRC1</th>
<th>ExSRC2</th>
<th>SRC-voting</th>
<th>LRC-DEF</th>
<th>LRC</th>
<th>LRC-DEF</th>
<th>C-LCRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 $\times$ 20</td>
<td>47.6</td>
<td>84.5</td>
<td>83.5</td>
<td>67.8</td>
<td>65.2</td>
<td>89</td>
<td><strong>89.6</strong></td>
<td>83.6</td>
</tr>
<tr>
<td>32 $\times$ 32</td>
<td>55.3</td>
<td>87.5</td>
<td>88</td>
<td>73.8</td>
<td>66.9</td>
<td><strong>92.3</strong></td>
<td>91.8</td>
<td>86.3</td>
</tr>
<tr>
<td>40 $\times$ 40</td>
<td>54.6</td>
<td>88.4</td>
<td>87.7</td>
<td>75</td>
<td>66.5</td>
<td>91.7</td>
<td><strong>92.1</strong></td>
<td>85.1</td>
</tr>
</tbody>
</table>

* A similar setting as in [38] is used for extended SRC except $20 \times 20$ instead of $27 \times 20$ downsamples images are used. The results reported by the authors [38] are: about 93% accuracy with seven images per class available and 78% accuracy with two images per class.
Specifically, LCRC-Spr achieves 89.6% recognition rate with 400 dimensional raw pixel features which outperforms SRC-voting, LRC-DEF by than 21.8% and 24.4%, respectively. This shows that LCRC copes well with the single training sample FR problems and also confirms that LCRC needs much less training samples than the other methods. Due to the use of intra-class variant dictionary, extended SRC performs much better than CRC and even the modular approaches SRC-voting and LRC-DEF on this single training sample problem, however still worse than the proposed LCRC and LCRC-Spr. Both these two LCRC algorithms obtain above 91% accuracy in higher dimensional feature spaces, while the best accuracy 88.4% of all other methods is obtained by extended SRC with feature dimension 1600. This result is impressive, since only one training sample is available for each subject and the test data incorporates both expression, illumination changes and severe facial disguises. Note that compared to other methods, extended SRC requires extra training samples to construct the bases dictionary on the one training sample problem, which possibly cannot be satisfied in real-world applications.

Note that for the classed based algorithms LRC-DEF and C-LCRC, without collaboration of data from other classes the single training sample problem becomes even more challenging, since the probe image is actually represented by only one image each time. However, C-LCRC still achieves much higher accuracies than other collaborative representation based methods (CRC and the SRC modular approach), which is due to the effectiveness of the spatial pyramid partition and fusion method.

6. Conclusions and future work

In this work we propose a new face recognition method incorporating locality on both representation samples and spatial features. The locality constraint enforces the representation sparse, which effectively concentrates the large representation coefficients on a small number of training samples, while other ones are nearly zeros. The spatial pyramid local patches instead of holistic features are used to significantly boost the classification performances. Due to both, the proposed method is very robust for two critical problems in face recognition: occlusion and lack of training data.

Based on the locality constrained representation, we proposed three algorithms. The first two: LCRC and LCRC-Spr take training samples from all classes into representing the probe image, while the third one C-LCRC is homo-class based. All these three algorithms outperform the state-of-the-art on the public Yale B and the third one C-LCRC is homo-class based. All these three algorithms take training samples from all classes into representing the probe image, while other ones are nearly zeros. The spatial pyramid local patches instead of holistic features are used to significantly boost the classification performances. Due to both, the proposed method is very robust for two critical problems in face recognition: occlusion and lack of training data.

In our method, each test patch is represented by its corresponding training patches at the same location and patches in a face are considered independently. This may not work well on datasets without well aligned, for example, with large pose variations. How to extend the proposed methods by modelling the dependency between patches appears to be interesting in the future work.

References


Fumin Shen received his bachelor degree in Applied Mathematics from Shandong University, China. Currently he is a Ph.D. student in School of Computer Science, Nanjing University of Science and Technology, China. His major research interests include computer vision and machine learning, including face recognition, image analysis, hashing methods, and robust statistics with its applications in computer vision.

Zhenmin Tang received his Ph.D. degree from Nanjing University of Science and Technology, Nanjing, China. He now is a professor and also the head of School of Computer Science, Nanjing University of Science and Technology. His major research areas include intelligent system, pattern recognition, image processing, embedded system. He has published over 80 papers. He is also the leader of several key programs of National Nature Science Foundation of China.

Jingsong Xu is a Ph.D. student at Pattern Recognition and Intelligent System, Nanjing University of Science and Technology. He received the B.Sc. degree from the same university in 2007. Currently, he is visiting University of Technology, Sydney. His research interests include computer vision and machine learning.