Kernel Locality Preserving Low-Rank Representation with Tikhonov Regularization

Yuqi Pan*, Mingyan Jiang‡, Fei Li*.

ABSTRACT

Classification based on Low-Rank Representation (LRR) has been a hot-topic in the field of pattern classification. However, LRR may not be able to fuse the local and global information of data completely and fail to represent nonlinear samples. In this paper, we propose a kernel locality preserving low-rank representation with Tikhonov regularization (KLP-LRR) for face recognition. KLP-LRR is a nonlinear extension of LRR, and it introduces the local manifold structures of data sets into LRR methods. Since the feature information in kernel space has a very high dimensionality, and to fit the proposed KLP-LRR method well, we introduce locality preserving factor and Tikhonov regularization into dimensionality reduction. It can get more discriminant coding information, especially in the aspect of combining local features with global features, where it is capable of improving the recognition rate obviously. Explicit experimental results on AR, the extended Yale B, FERET face databases show KLP-LRR out-performs other comparative methods.

Key words: Low-rank representation, image processing method, face recognition, manifold structures, pattern classification.

1. Introduction

Face recognition is a hot topic in pattern classification and machine learning. In recent years, more and more face recognition applications have been utilized in our daily life. Many of studies have pushed the filed a lot in the last two decades, such as the Nearest Neighbor (NN)[1] and the Nearest Subspace (NS)[2], which are very easy to use and have been widely applied. But neither of them is robust to the noise. Thus many improved methods have been proposed based on NN and NS[3-7].

In the past few years, sparse representation has attracted researchers’ great interest. Sparse representation has emerged widely in the field of sparse reconstruction[8,9], denoising[10], hyperspectral image resolution[11] and so on. Wright et al., proposed a sparse representation-based classification (SRC)[12] method for face recognition. The basic idea of a SRC based classifier is built on the concept that a pixel can be represented as a linear combination of labeled samples via the sparse regularization techniques, for instance, the $\ell_0$-norm regularization and the $\ell_1$-norm regularization. SRC does not have to operate a training process in advance, but it does need labeled data. A test sample to be classified is sparsely represented by the training data, and it will be assigned to the cluster whose labeled samples provide the smallest reconstruction error. The SRC has shown its effectiveness and generality in some experimental results. Researchers began to combine the SRC with many other techniques, such as linear pyramid matching for image classification[13], Gabor feature for SRC to resist the face occlusion[14], locality constrains for image classification [15,16], selecting strategy for testing sample[17] and Independent Component Analysis[18]. M. Yang et al., proposed a robust sparse coding (RSC) for face recognition, which is robust to face occlusion, corruption, lighting and expression changes by seeking for the MLE (maximum likelihood estimation) solution of the sparse coding problem[19]. But there are still several problems in SRC, such as lack of robustness to the occlusion, the high operating time-cost and the rationality of the algorithm itself.

a. School of Information Science and Engineering, Shandong University, No.27 ShanDaNanLu, Jinan, Shandong Province (086)250000, China
1. Yuqi Pan(1990-). Male. Master student. Research interest: pattern classification, intelligent algorithm optimization, data mining. E-mail: chinapanyq@mail.sdu.edu.cn;
2. Corresponding author: Mingyan Jiang(1964-). Male. Professor. Research interest: Artificial neural network, pattern classification, image processing, intelligent algorithm optimization, data mining. E-mail: jiangmingyan@sdu.edu.cn;
3. Fei Li(1986-). Female. Ph.D. student. Research interest: pattern classification, image processing, intelligent algorithm optimization. E-mail: lifeisu@mail.sdu.edu.cn.
Recently, Zhang et al. proposed a classification method called collaborative representation (CRC) for face recognition[20]. It was suggested that it was the collaborative representation (CR) rather than sparse representation that making the essential influence for recognition. Compared to SRC, CRC can achieve the competitive experimental results with much less time-consuming computation. The key difference between SRC and CRC implementations is that the former employs an $\ell_0$ or $\ell_1$ norm regularization while the latter employs an $\ell_2$ norm regularization; thus, the latter can have a closed result, resulting in much lower time-consuming cost. In [21], the kernel version of a collaborative representation based classification (CRC) was proposed and called as KCRC, which employs all the training data from different classes simultaneously (post-partitioning). In [22], a method called kernel locality-constrained collaborative representation based discriminant analysis (KLCR-DA) is proposed. It is a nonlinear version of CRC, moreover, it uses the local structures of data sets, which makes CRC strengthen the ability to represent the query image with local features.

Lately, Liu et al. established Low-Rank Representation (LRR)[23,24] as an efficient method to perform noise correction and subspace segmentation simultaneously. Generally speaking, the purpose of LRR is to find the lowest rank representation among all the observed data that represents all vectors as the linear combination of the atoms in a dictionary with seeking the solution of the nuclear norm minimization problem.

Based on the state-of-art Low Rank Representation theory, we propose a novel classification mechanism called Kernel Locality Preserving Low-Rank Representation with Tikhonov Regularization (KLP-LRR). The motivation why we proposed KLP-LRR are shown below:

1). Since all the testing samples are represented with respect to all the training samples by seeking the low rank matrix representation with solving the nuclear norm regularization problems, little local useful facial information can be used for face recognition, needless to say, the manifold features of images.

2). We want to make full use of all the features in the images, especially the manifold ones. Thus, to extract the desired features, we have to map the training data to a high dimensionality, which makes the problem non-linear. However, the original thought of LRR is to represent all the vectors as the linear combination of the bases in a dictionary. Thus, we need to employ kernel trick technique, and, to keep the integrity of manifold features including the global and the local ones of facial images, the kernel based dimensionality reduction method is also required. In our method, we utilize the Kernel Locality Preserving Projection (KLPP)[25] to reduce the dimension.

3). Besides the high manifold features, we also introduce Tikhonov regularization matrix into our method to enhance the comprehensiveness of using local and global information further.

Because of above reasons, we propose our KLP-LRR method. In KLP-LRR, the input data set is implicitly mapped into a high dimensional kernel feature domain. The purpose of it is not only to convert the non-linear feature information into linear ones, but also to seek the facial manifold features hiding in a high dimension. Nevertheless, the data in kernel feature space is tremendous and unknown, namely, KLP-LRR can not be operated directly. To overcome this difficulty, kernel locality preserving projection (KLPP)[26,27] is introduced. By applying this dimensionality reduction method, we can remap the high dimensional data into a new subspace for KLP-LRR, which contains the local and global manifold features. The proposed KLP-LRR is an innovative combination of locality preserving with LRR. Additionally, our algorithm is robust to luminance variation.

The rest of this paper is organized as follow: Some basic knowledge of LPP, LRR and kernel trick technique is shown in Section 2. We provide KLP-LRR algorithm in detail in Section 3. A variety of experimental results are shown in Section 4. Finally, we make some conclusions in Section 5.

2. Related reviews

2.1 Locality Preserving Projection

Considered the observed data set $\{ x_1, \cdots, x_n \} \in \mathbb{R}^w$, where the original two dimensional image matrix is converted into
one dimensional vector, let \( X = [x_1, \cdots, x_n] \), then \( X \) is an \( m \times n \) size matrix. Let \( S \) be a similarity matrix defined on all pairwise data points. LPP is capable of acquiring the local manifold structure feature information by seeking the optimal solution of the following minimization problem:

\[
W_{opt} = \arg \min_w \sum_y (y_i - y_j)^2 S_{ij} \\
= \arg \min_w \sum_y (w^T x_i - w^T x_j)^2 S_{ij} \\
= \arg \min_w w^T X L X^T w \\
\text{s.t.} \quad w^T X L X^T w = 1
\]

(1)

where \( L = D - S \) is the graph Laplacian matrix, and \( D_{ij} = \sum_j S_{ij} \) is the local density measure around \( x_j \). The symmetry similarity matrix \( S_{ij} \) in LPP is defined as:

\[
S_{ij} = \begin{cases} 
\exp(-\frac{\|x_i - x_j\|^2}{t}), & \frac{\|x_i - x_j\|^2}{t} < \epsilon \\
0, & \text{otherwise}
\end{cases}
\]

(2)

where \( \epsilon \) is the definition of the radius of the local neighborhood. Here \( S_{ij} \) is actually heat kernel weight, the relative knowledge of how to select parameter \( t \) can be found in [28].

The objective function in LPP incurs a heavy penalty if neighboring points \( x_i \) and \( x_j \) are mapped far apart. Therefore, the minimization of the objective function is an attempt to ensure that if \( x_i \) and \( x_j \) are “close”, then \( y_i \) and \( y_j \) are close as well. The optimization will ultimately lead to the following generalized eigenvalue problem:

\[
X L X^T w = \lambda X D X^T w
\]

(3)

Let \( w_0, w_1, \cdots, w_{k-1} \) be the solutions of Eq. (3), ordered by their eigenvalues, \( 0 \leq \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{k-1} \). Then \( W = [w_0, w_1, \cdots, w_{k-1}] \) is the final transformation projection matrix of LPP.

2.2 Low Rank Representation

LRR has been proposed for a kind of method to recover subspace structure by Liu et al.[23,24]. LRR can be applied to segment data into linear subspaces in supervised or unsupervised classification. The purpose of LRR is to seek the lowest rank representation among all the candidates that is able to represent observed data vectors with linear combination of the dictionary atoms, supposed a proper dictionary. The regularized rank minimization problem can be modeled as follows:

\[
\min \text{rank}(Z) \quad \text{s.t.} \quad X = AZ,
\]

(4)

where \( X \) is a collected data matrix, each column of \( X \) is a collection of observed data, \( A \) is a dictionary, and \( Z \) is a lowest rank representation of data \( X \) with respect to a dictionary \( A \). Unfortunately, this formula function is not convex, which means it can not be minimized optimally. In the rank optimal minimization problems, the rank function can be replaced by a new method called the nuclear norm, problem (4) with respect to the following convex minimization problem:
\[ \min_{Z} \|Z\|, \quad s.t. \quad X = AZ \] (5)

For the seek of partitioning the data into the irrespective subspaces, in [23], the collected data matrix \( X \) is regarded as the dictionary. So problem (5) is changed like this:

\[ \min_{Z} \|Z\|, \quad s.t. \quad X = XZ \] (6)

Supposed the noisy data, the optimization formula of LRR is modeled as follows:

\[ \min_{Z} \|Z\| + \lambda \|E\|_{2,1}, \quad s.t. \quad X = XZ + E, \] (7)

where \( XZ \) is low rank and \( E \) is the corresponding representation errors (or noises) of the data matrix \( X \). The \( \|E\|_{2,1} \) is employed to characterize the error term for presenting the sample specific corruptions, \( \|E\| \) is chosen for the small Gaussian noise and for the stochastic occlusion, \( \|E\| \) is an appropriate choice. The coefficient \( \lambda \) (also \( \lambda > 0 \)) is introduced to balance the influence of the two regularization parts.

Under some certain circumstances, the solutions of these problems are quite similar, and Low Rank Representation based Classification ensures existence[29,30]. A large amount of algorithms have been developed to solve the problem about low rank recovery, such as an iterative thresholding method[31], Augmented Lagrange Multiplier method (ALM)[32,33] and Accelerated Proximal Gradient (APG)[34]. Specifically speaking, ALM is a better choice for considering the time-consuming cost and accuracy, which is also employed to solve the face recognition problem in this paper.

2.3 The kernel trick

The kernel trick is a very well-known method and widely used in pattern recognition. Some typical methods employing the kernel trick are KPCA[35], SVM[36-38] and KFDA [39]. With the kernel trick, original non-linear data can be easily mapped from the non-linear input feature space into a high dimensional kernel feature space, where non-linear feature can be reconstructed in a linear way, and then we can use a linear method to solve this problem.

Appropriate selection of a kernel function is capable of presenting the similarity accurately among all samples; however, not all metric distances can be used in kernel methods. In fact, valid kernels are only those satisfying the Mercer’s conditions [40], requiring to be positive semi-definite. For a given non linear mapping function \( \Phi \), the Mercer’s kernel function \( k(\cdot, \cdot) \) can be formulated like this:

\[ k(x, x') = \Phi(x)^T \Phi(x'). \] (8)

Widely applied kernels include linear kernel \( k(x, x') = x^T x' \), the \( t \)-degree polynomial kernel \( k(x, x') = (x^T x' + 1)^t (t \in \mathbb{Z}^+) \), and the Gaussian radial basis function (RBF) kernel \( k(x, x') = \exp(-\gamma \|x - x'\|_2^2) (\gamma > 0) \), and \( \gamma \) is a coefficient of RBF kernel.

With the kernel mentioned above, a linear algorithm can be easily generalized into a non-linear one by employing the kernel trick without mapping the data completely by substituting a chosen kernel.

3. KLP-LRR based Classification

In this section, we will explicitly describe a new image classification method based on LRR.
3.1 why locality preserving projection?

As mentioned in Section 1, LPP is introduced to our method. The reasons why we choose it are shown as follows:

1). We assume that Gaussian core is selected and the lowest rank representation matrix $\mathbf{Z}$ with respect to the observed data $\mathbf{X}$ is already obtained. Actually, Kernel PCA[41] or Kernel LDA (KDA)[42] can be used to extract the features in the high dimensional space. However, Yang et al. has pointed out that the feature extraction algorithms should be consistent with the classification method [43]. Neither PCA nor LDA is designed for KLP-LRR method. So they may not be able to extract the accuracy feature information hidden in the high dimensional domain in KLP-LRR.

2). It is shown that a lot of manifold structures hidden in face image can not be extracted effectively by PCA or LDA. Fortunately, the basic idea behind LPP is that it considers the manifold structure of the data set, and preserves the locality of data in the embedding space. LPP has shown the superiority in the fields of computer vision and pattern recognition. It is the advantage of LPP that makes LRR strengthen the ability of locality representation, and the simplified proof of it is shown as follows:

The Eq. (3) can be rewritten as:

$$\mathbf{X DX^T} \mathbf{w} = \lambda \mathbf{X LX^T} \mathbf{w} = \gamma \mathbf{XLX^T} \mathbf{w}. \tag{9}$$

Let $\mathbf{M}_D = \mathbf{X DX^T}$ and $\mathbf{M}_L = \mathbf{X LX^T}$, then the above equation is equivalent to:

$$\mathbf{M}_D \mathbf{w} = \gamma \mathbf{M}_L \mathbf{w}. \tag{10}$$

Therefore, the original functions of LPP are the eigenvalues with respect to the largest eigenvalues of Eq. (5).

For the simplicity of the following proof, we introduce some notions here. Define $n \times m$ matrices:

$$\mathbf{F}_D = \mathbf{D}^{\frac{1}{2}} \mathbf{X}^T, \quad \mathbf{F}_L = \mathbf{L}^{\frac{1}{2}} \mathbf{X}^T. \tag{11}$$

It is obvious that with above notations, we have

$$\mathbf{M}_D = \mathbf{F}_D \mathbf{F}_D^T, \quad \mathbf{M}_L = \mathbf{F}_L \mathbf{F}_L^T. \tag{12}$$

Note that both the two matrices $\mathbf{M}_D$ and $\mathbf{M}_L$ contain locality information of observed image data since the Laplacian matrix $\mathbf{L}$ and the density measure matrix $\mathbf{D}$ include manifold structures and its density features, respectively, which helps LRR to reconstruct the testing image matrix.

Thus, Laplacian Score (LS) [44] proposed by He et al. is introduced into our method. The objective function of Laplacian Score is defined as follows:

$$\mathbf{L}^L_{r} = \frac{\sum_y (f_{ri} - f_{pj})^2 S_{ij}}{\text{Var}(f_{ri})}, \tag{13}$$

where $\mathbf{L}^L_{r}$ is the Laplacian score of the $r$-th feature, $f_{ri}$ and $f_{pj}$ denote the $r$-th feature of the samples $x_i$, respectively. $\text{Var}(f_{ri})$ is the estimated variance of the $r$-th feature. $\mathbf{S}$ is the unsupervised weight matrix of the $k$ nearest neighbor graph, which is defined as Eq. (2).

3.2 Classification based on KLP-LRR

Suppose that samples are mapped from the original feature space $\mathbb{R}^m$ into a high dimensional feature space $\mathbb{F}$ by a nonlinear mapping $\phi$, $\phi: \mathbb{R}^m \rightarrow \mathbb{F}$. Denote $\phi_x = [\phi(a_1), \phi(a_2), \ldots, \phi(a_n)]$ and $\phi_x = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]$ represent the
matrix consisting of all the training samples and all the test samples after the non-linear mapping $\phi$, where $\phi_j \in \mathbb{R}^{D \times p}$ with $D \gg m$, whose $D$ is the dimension of the feature space $F$. Then, it is easy for us to represent the image of all the testing samples with respect to the image of all training samples in this kernel feature space $F$ as follows:

$$\Phi_X = \Phi_A Z.$$  \hspace{1cm} (14)

Hence, the problem of Low-Rank Representation in $F$ can be formulated like this:

$$\min_Z \|Z\|_F \quad \text{s.t.} \quad \Phi_X = \Phi_A Z,$$  \hspace{1cm} (15)

where $Z$ is the lowest-rank representation matrix of the testing data in term of training data in the high dimensional feature space $F$. Due to unpredictability of $\Phi_X$ and $\Phi_A$, Eq. (15) can not be solved directly. We have to introduce dimensionality reduction method in $F$; since the whole images need to be projected from $F$ into a low-dimensional subspace. Let $P \in \mathbb{R}^{m \times d}$ is the projection matrix in $F$, then, we can rewrite the Eq. (14) as follows:

$$P^T \Phi_X = P^T \Phi_A Z.$$  \hspace{1cm} (16)

Obviously, it can be seen that $P$ need to be connected with all the images data in order that dot products of images can be replaced by a kernel. With the reconstruction of the projection matrix in a kernel-based dimensionality reduction method in KLPP, the projection vector is a linear combination of images in $F$. We can obtain

$$P_j = \sum_{i=1}^n v_j, \phi(a_i) = \Phi_A v_j,$$  \hspace{1cm} (17)

where $P_j$ is the $j$-th projection vector of $P=[P_1, P_2, \ldots, P_d] \in \mathbb{R}^{d \times m}$ and $v_j=[v_{j,1}, v_{j,2}, \ldots, v_{j,n}]^T$ is called the pseudo-projection vector corresponding to the $j$-th projection vector.

To keep the locality features of the images, we introduce Tikhonov regularization. Eq. (16) can be rewritten as follows:

$$\hat{P}_j = \Phi_A (\Gamma_{\phi(x)} v_j),$$  \hspace{1cm} (18)

where the new locality biasing Tikhonov matrix $\Gamma_{\phi(x)}$ has the form of:

$$\Gamma_{\phi(x)} = \begin{bmatrix} \|\Phi(x) - \Phi(a_1)\|_F & 0 & \cdots & 0 \\ \|\Phi(x) - \Phi(a_n)\|_F & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ 0 & \cdots & \ddots & \|\Phi(x) - \Phi(a_n)\|_F \end{bmatrix}.$$  \hspace{1cm} (19)

Here, the crucial motivation for the introduction of Tikhonov regularization is to compensate the shortage of discriminative representation feature information, especially in some very low feature subspace dimensionalities. The effectiveness of Tikhonov regularization will be discussed in Section 4.4.

Let $\tilde{v} = \Gamma_{\phi(x)} v_j$, then we have:

$$\hat{P}_j = \Phi_A \tilde{v}_j.$$  \hspace{1cm} (20)

Let $V=[\tilde{v}_1, \ldots, \tilde{v}_d]$ be the new projection matrix, and Eq(20) can be expressed as follows:

$$\hat{P} = \Phi_A V.$$  \hspace{1cm} (21)

then substituting Eq. (21) into Eq. (16), we get:

$$V^T \Phi_A^T \Phi_A V = V^T \Phi_A^T \Phi_A Z.$$  \hspace{1cm} (22)
where \( \Phi^T \Phi \in \mathbb{R}^{n \times p} \) and \( \Phi^T \Phi \in \mathbb{R}^{m \times n} \) are the kernel Gram matrix, which can be calculated from the kernel function.

That is to say, for any samples \( a \) and \( x \), we have \( \phi(a)^T \phi(x) = k(a,x) \) and \( \phi(a)^T \phi(a) = k(a,a) \), where \( k(\cdot,\cdot) \) is a kernel function. Therefore, we have:

\[
\Phi^T \Phi = \begin{bmatrix} \phi(a_1)^T \phi(x_1) & \cdots & \phi(a_p)^T \phi(x_p) \\ k(a_1,x_1) & \cdots & k(a_p,x_p) \\ \vdots & \ddots & \vdots \\ k(a_1,x_1) & \cdots & k(a_p,x_p) \end{bmatrix},
\]

\[
\Phi^T \Phi = \begin{bmatrix} \phi(a_1)^T \phi(a_1) & \cdots & \phi(a_p)^T \phi(a_p) \\ k(a_1,a_1) & \cdots & k(a_p,a_p) \\ \vdots & \ddots & \vdots \\ k(a_1,a_1) & \cdots & k(a_p,a_p) \end{bmatrix}.
\]

Thus, we have a feasible optimization minimization problem:

\[
\min_{\mathbf{Z}} \|\mathbf{Z}\|, \quad \text{s.t.} \quad V^T \mathbf{X} = V^T A \mathbf{Z},
\]

where \( A = \Phi^T \Phi \) and \( \mathbf{X} = \Phi^T \Phi \).

Under the noisy circumstance, the optimization problem can be formulated as follows:

\[
\min_{\mathbf{Z}} \|\mathbf{Z}\| + \lambda \|E\|_{2,1}, \quad \text{s.t.} \quad V^T \mathbf{X} = V^T A \mathbf{Z} + E,
\]

here \( E \) is the corresponding reconstruction error in the high dimensional feature space, the coefficient \( \lambda > 0 \) is employed to balance the effects of the two parts. To get the optimal solution of problem (26), the pseudo-transformation matrix \( V \) is required. In this work, we calculate the pseudo-matrix \( V \) in the following way:

We apply the dimensionality reduction method in KLPP [26, 27] with the transformation matrix. The pseudo-transformation vector \( v_j \in \mathbb{R}^n \) can be obtained by solving the eigenvalue problem:

\[
n\lambda_v = \Phi^T \Phi v,
\]

where \( n \) is the number of training samples, and \( v_j \) is normalized eigen-vector so that \( \lambda_v v = 1 \). The first \( d \) eigen-vector \( v_j \) is taken, which is corresponding to the first \( d \) largest eigen-value \( \lambda_j, \ j = 1, \ldots, d \) and \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \).

Then, the pseudo-transformation matrix \( V = [v_1, v_2, \cdots, v_d] \in \mathbb{R}^{n \times d} \).

With the method mentioned above, we can obtain the pseudo-transformation matrix \( V \). Then we can solve the problem (26) by ALM mentioned in Section 2.2. With the solution coefficient matrix \( Z \), we make the classification of the \( l \) th testing samples \( x_l \) with respect to \( Z \) as follows. In every class \( i \), we define \( \zeta(z_i) \in \mathbb{R}^n \) is the discriminant function which gets the non-zero coefficients in \( z_i \) in terms of the \( i \) th class only. \( z_i \in \mathbb{R}^n \) as the \( l \) th column of \( Z, (l = 1, 2, \cdots, p) \).

The \( i \) th representation for the given testing sample \( x_l \) in the reduced subspace can be formulated as \( V^T A \zeta(z_i) \). Then, we classify \( x_l \) based on minimizes the residual between \( V^T x_l \) and its representations:
\[
\min_i r(x_i) = \left\| P^T X_i - V^T A \xi_j (z_j) \right\|_2 / \left\| \xi_j (z_j) \right\|_2,
\]

where \( X_i \) is the \( i \)th column of \( X \). The complete classification procedure of KLP-LRR is shown in Algorithm 1.

---

**Algorithm 1. KLP-LRR**

1. Input: a training matrix \( A \in \mathbb{R}^{m \times k} \) for \( k \) classes, a testing matrix \( X \in \mathbb{R}^{m \times p} \), parameter \( \lambda > 0 \).
2. Choose a Laplacian core kernel \( k(\cdot, \cdot) \) and its parameters.
3. Calculate Gram matrix \( A \) and \( X \).
4. Calculate matrix \( V \) by selecting the KLPP projection method and computing Tikhonov regularization matrix.
5. Normalize the columns of \( V^T A \) and \( V^T X \) to have unit nuclear norm.
6. Solve the problem (26) and obtain matrix \( Z \in \mathbb{R}^{m \times p} \).
7. Compute the residual
   for \( l = 1, 2, \ldots, p; i = 1, 2, \ldots, k \)
   \[
   r_l (x_i) = \left\| X_i - A \delta_l (z_l) \right\|_2 / \left\| \delta_l (z_l) \right\|_2.
   \]
8. Output: \( \text{class}(x_i) = \arg \min_i r_l (x_i) \).

---

### 4. Experimental Results

In this section, we verify the proposed algorithm in face recognition and compare with SRC, CRC, kernel SRC (KSRC), and LRR. To prove the effectiveness of our algorithm, we perform experiments on three public available face databases for face recognition: the AR database, the Extended Yale B database, the FERET database. In SRC and KSRC, the \( l_1 \)-norm regularization is applied to solve the minimization problem with the linprog function in the optimization toolbox of MATLAB. For CRC, we do appreciate the authors for providing their relative codes [20]. The dimension of the data is cropped at 180 for PCA features extraction in LRRC. The parameters of the methods here are set according to the following description. The parameter \( \epsilon \) in SRC is set to \( 10^{-4} \). In CRC, let parameter \( \lambda = 0.01 \). In KSRC, the polynomial kernel is employed with the parameters \( d = 3 \) and \( \lambda = 10^{-4} \). In LRRC, let \( \lambda = 1 \). In each database, we choose several amount of training data and testing data. We first find the optimal parameters of KLP-LPP, then implement our algorithm on the testing data and compare its performance with other methods.

Our experimental environment is shown as follows: software version, MATLAB 2010b; operating system, Windows 7; computer hardware, Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz and 8GB (1333Mhz) RAM.

#### 4.1 AR database

AR database is made up of more than 4000 face images of 126 persons under the front view of faces with different expressions, illuminations and occlusions. Every person includes 26 images which are taken into two sessions (separated in two weeks) and each session contains 13 images. We chose a subset of the database consisting of 50 male and 50 female samples. In every sample, 14 images with different illumination, expression changes and two different kinds of occlusions, the images are cropped to size 40\(!\times\)40. Fig. 1 shows some sample images used in our experiments.
In this experiment, we choose the first seven images in the first session for training set while choosing 7 images taken in the first and second session with no intersection for testing set. Hence, there are 700 training samples and 700 testing samples.

The average recognition rates of the testing data are shown in Table 1. It is obviously seen that KLP-LPP works better than any other methods. We further show the recognition rates with the feature dimension of 20, 40, 60, 80, 100, 120, 140, 160, 180, 200 and compare with SRC, CRC, KSRC, LRRC (shown in Table 2). In Table 2, we can see that KLP-LPP outperforms the other algorithms when the dimension is higher than 80, and when the dimension is lower than 100, KSRC get a highest recognition rate at most cases.

### Table 1

<table>
<thead>
<tr>
<th>Methods</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>87.24</td>
</tr>
<tr>
<td>CRC</td>
<td>84.91</td>
</tr>
<tr>
<td>KSRC</td>
<td>93.29</td>
</tr>
<tr>
<td>LRRC</td>
<td>83.61</td>
</tr>
<tr>
<td>KLP–LRR</td>
<td><strong>94.89</strong></td>
</tr>
</tbody>
</table>

### 4.2 Extended Yale B database

The Extended Yale B database consists of 2414 front-view face images of 38 individuals taken under 64 different controlled illuminations. All the images are cropped to the actual facial area. The original size of images stored in database is $192 \times 168$ pixels. In this experiment, we crop the size of the image to $64 \times 64$. **Fig.2** shows an example of subjects from the Extended Yale B database.

At first, we randomly chose half of the image of each class for training data (32 images per subject). The rest ones are made up to the testing data. The experimental results are shown in Table 3 and 4, in which the former one is the average performance comparison, the latter one is the explicit recognition rate from subspace feature dimension 20 to 200.

### Table 3

<table>
<thead>
<tr>
<th>Methods</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>89.301</td>
</tr>
<tr>
<td>CRC</td>
<td>87.183</td>
</tr>
</tbody>
</table>
We can easily get a conclusion that the results achieved by KLP-LRR are the best of all the competitive algorithms. Table 4 shows KSRC has a slight advantage when the dimension is less than 100. KLP-LRR shows its advantage and begins to get the top recognition rate when the dimensionality is higher than 100.

### 4.3 FERET database

The FERET database is often used as a standard database evaluating face classification algorithms. It consists of 14051 images of total 1199 subjects. In this experiment, we performed the proposed method on a subset of 1400 images of 200 persons; each person has seven images with the variations in facial poses, expressions and illuminations. The images are re-sized to $40 \times 40$. Some examples are shown in Fig. 3.

In this experiment, we used the first six images of each subject for training, and randomly-chosen ones are used for testing data. Thus, there are 1200 training samples and 1000 testing ones.

![Some sample images in FERET database](image)

**Fig. 3** Some sample images in FERET database

The average experimental result is shown in Table 5. KLP-LRR gets the highest recognition rate again.

We further test recognition accuracy of KLP-LRR with the feature subspace dimensions of 20, 40, 60, 80, 100, 20, 140, 160, 180, and 200. Table 6 shows the recognition rates versus feature dimension by SRC, CRC, KSRC, LRRC. It can be clearly seen that SRC and CRC have a similar performance and worse than LRRC. When the dimensionality is 20 and 40, KSRC outperforms the others. While the dimensionality is higher than 40, KLP-LPP presents a outstanding performance.

Based on the experimental results on the three different face databases, we can get the conclusion that KSRC has a better performance under a low dimensionality feature subspace. When the dimensionality of feature subspace becomes high, KLP-LRR gets the top recognition rate, due to its ability of combining the locality manifold structures of face image and non-linear representation. Experimental results indicate that the KLP-LRR is a promising classification algorithm.

### Table 5

Average recognition rates(percent) on FERET database

<table>
<thead>
<tr>
<th>Methods</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>70.2</td>
</tr>
<tr>
<td>CRC</td>
<td>73.1</td>
</tr>
<tr>
<td>KSRC</td>
<td>80</td>
</tr>
<tr>
<td>LRRC</td>
<td>78.1</td>
</tr>
<tr>
<td>KLP–LRR</td>
<td>86.6</td>
</tr>
</tbody>
</table>

### 4.4 The Effectiveness of Tikhonov Regularization

To demonstrate the effectiveness of the introduced Tikhonov regularization, we did a comparison experiment in AR, Extended Yale B and FERET face database respectively. The training and testing samples are as same as the previous experiments. We compare the KLP-LRR with the one without Tikhonov regularization from dimensionality of feature
subspace 20 to 200. Fig. 4 illustrates the explicit experimental results. Specific experimental data is shown in Table. 2, 4 and 6, marked by KLP-LPP*.

![Fig. 4 The effectiveness of Tikhonov Regularization](image)

From Fig. 4, we can find the obvious fact that Tikhonov regularization can obviously enhance the recognition rate, especially, when the dimensionality of feature subspace is extremely low. With exhaustive analysis of experimental results, we have found that recognition rates of the same method without Tikhonov regularization is 3.3% on average lower than the one with Tikhonov regularization. Specifically speaking, the disparities are actually between 1.88% and 6.5% when the dimensionality of feature subspace is below 100, which are worse than those above 100 feature subspace dimensionality (between 1.27% and 4%). As we can see from Fig.4, as the increase of the dimensionality of feature subspaces, the gap between KLP-LRR with Tikhonov regularization and the one without Tikhonov slightly decreases. The premier reason is that with the increase of the dimensionality, there is more discriminative feature information to employ for improving the recognition accuracy. With the best knowledge of authors, we draw a conclusion that there is lack of adequate useful feature information when the dimensionality of feature subspace is pretty low while Tikhonov regularization can definitely compensate this shortage of discriminative feature information and reduce the residual to some extent.

5. Conclusions

In this paper, we have proposed a classification method called Kernel Locality Preserving Low Rank Representation (KLP-LRR). The main contributions of KLP-LRR are presented as follows: (1) KLP-LRR incorporates the idea of local preserving projection into low rank representation, which means the manifold features of face images can be used in the reconstruction of the representation. (2) KLP-LRR embeds the Laplacian score in low rank representation, which extracts the manifold features hidden in high dimensionality effectively, improves the ability of locality reconstruction of LRR and combines the local and global features of images. (3) Tikhonov regularization is introduced into low rank representation at the first time, which can further improve the ability of locality reconstruction of LRR and make full use of the local and global features of images. With large amounts of experimental results, we have shown that the proposed method is competitive with the state-of-art other classification methods (SRC, CRC, KSRC, and LRRC). The experimental results on different face database, including, AR database, Extended Yale B database, and FERET database, demonstrate that KLP-LRR greatly improves the classification performance over the compared algorithms.

Our future work will be focused on the robustness of our method including severe variety of expressions, poses, illuminations and occlusions. We plan to introduce a kind of deep learning scheme to seek the optimal parameters of KLP-LPP in an automatic way.
Acknowledgment

This work is supported by the Natural Science Foundation of Shandong Province (ZR2014FM039).

Table 2. Recognition rates on AR database with different feature subspaces.

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>63.27</td>
<td>85.31</td>
<td>87.27</td>
<td>88.19</td>
<td>90.57</td>
<td>90.63</td>
<td>91.14</td>
<td>92.05</td>
<td>92.27</td>
<td>91.74</td>
</tr>
<tr>
<td>CRC</td>
<td>66.52</td>
<td>80.6</td>
<td>85.06</td>
<td>85.24</td>
<td>86.71</td>
<td>87.71</td>
<td>89.14</td>
<td>89.5</td>
<td>89.43</td>
<td>89.22</td>
</tr>
<tr>
<td>KSRC</td>
<td>83.32</td>
<td>92.72</td>
<td>92.36</td>
<td>93.82</td>
<td>94.9</td>
<td>95.16</td>
<td>95.27</td>
<td>95.07</td>
<td>95.3</td>
<td>94.93</td>
</tr>
<tr>
<td>LRR</td>
<td>62.29</td>
<td>74.43</td>
<td>76.22</td>
<td>78.46</td>
<td>86.57</td>
<td>87.71</td>
<td>92.14</td>
<td>92.34</td>
<td>92.61</td>
<td>93.29</td>
</tr>
<tr>
<td>KLP-LRR</td>
<td>80.07</td>
<td>87.25</td>
<td>90.54</td>
<td>92.74</td>
<td>94.69</td>
<td>95.21</td>
<td>95.87</td>
<td>96.03</td>
<td>95.61</td>
<td>95.06</td>
</tr>
<tr>
<td>KLP-LRR*</td>
<td>85.52</td>
<td>91.57</td>
<td>93.54</td>
<td>95.01</td>
<td>96.57</td>
<td>97.18</td>
<td>97.14</td>
<td>97.75</td>
<td>97.69</td>
<td>96.97</td>
</tr>
</tbody>
</table>

Table 4. Recognition rates on Extended Yale B database with different feature subspaces.

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>75.57</td>
<td>80.37</td>
<td>88.37</td>
<td>89.18</td>
<td>91.67</td>
<td>92.21</td>
<td>93.2</td>
<td>93.2</td>
<td>94.18</td>
<td>95.06</td>
</tr>
<tr>
<td>CRC</td>
<td>51.42</td>
<td>85.63</td>
<td>85.85</td>
<td>91.99</td>
<td>89.14</td>
<td>91.1</td>
<td>91.2</td>
<td>95.72</td>
<td>95.26</td>
<td>94.62</td>
</tr>
<tr>
<td>KSRC</td>
<td>76.42</td>
<td>88.37</td>
<td>89.18</td>
<td>91.67</td>
<td>92.21</td>
<td>93.2</td>
<td>92.21</td>
<td>92.76</td>
<td>91.99</td>
<td>92.21</td>
</tr>
<tr>
<td>LRR</td>
<td>52.52</td>
<td>85.85</td>
<td>89.47</td>
<td>91</td>
<td>91.2</td>
<td>93.2</td>
<td>94.62</td>
<td>94.51</td>
<td>95.73</td>
<td>96.38</td>
</tr>
<tr>
<td>KLP-LRR</td>
<td>70.07</td>
<td>84.37</td>
<td>87.84</td>
<td>89.23</td>
<td>90.45</td>
<td>93.72</td>
<td>93.69</td>
<td>94.85</td>
<td>95.62</td>
<td>96.11</td>
</tr>
<tr>
<td>KLP-LRR*</td>
<td>75.24</td>
<td>88.16</td>
<td>91.2</td>
<td>93.91</td>
<td>94.84</td>
<td>96.75</td>
<td>97.38</td>
<td>97.89</td>
<td>97.68</td>
<td>98.57</td>
</tr>
</tbody>
</table>

Table 6. Recognition rates on FERET database with different feature subspaces.

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>120</th>
<th>140</th>
<th>160</th>
<th>180</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>27</td>
<td>53.5</td>
<td>61</td>
<td>70</td>
<td>77</td>
<td>77.5</td>
<td>82</td>
<td>83.5</td>
<td>87</td>
<td>83.5</td>
</tr>
<tr>
<td>CRC</td>
<td>33</td>
<td>47.5</td>
<td>70</td>
<td>77</td>
<td>79.5</td>
<td>81</td>
<td>83.5</td>
<td>86.5</td>
<td>86</td>
<td>87</td>
</tr>
<tr>
<td>KSRC</td>
<td>53.5</td>
<td>77.5</td>
<td>77</td>
<td>81</td>
<td>83.5</td>
<td>81</td>
<td>86</td>
<td>86.5</td>
<td>87.5</td>
<td>86.5</td>
</tr>
<tr>
<td>LRR</td>
<td>42</td>
<td>63</td>
<td>70</td>
<td>82.5</td>
<td>84</td>
<td>88</td>
<td>85.5</td>
<td>89</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>KLP-LRR</td>
<td>51</td>
<td>72.5</td>
<td>78.5</td>
<td>82</td>
<td>85</td>
<td>87.5</td>
<td>89</td>
<td>90.5</td>
<td>91</td>
<td>90.5</td>
</tr>
<tr>
<td>KLP-LRR*</td>
<td>57.5</td>
<td>77</td>
<td>82.5</td>
<td>87.5</td>
<td>89</td>
<td>91.5</td>
<td>92.5</td>
<td>93</td>
<td>92.5</td>
<td>92.5</td>
</tr>
</tbody>
</table>

References


