Local Ridge Regression for Face Recognition

Hui Xue[1,2] Yulian Zhu[1] Songcan Chen[*1,2]

[1]*Department of Computer Science & Engineering, Nanjing University of Aeronautics & Astronautics, 210016, Nanjing, P.R. China*

[2]*State Key Lab. for Novel Software Technology, Nanjing University, P.R. China*

Abstract:

Ridge regression (RR) for classification is a regularized least square method to model the linear dependency between covariate variables and labels. By applying appropriate techniques to encode the multivariate labels in face recognition as the vertices of the regular simplex which can separate points with highest degree of symmetry, RR maps the face images into a face subspace where the images from each individual will locate near their individual targets. However, as a holistic method, RR operates directly on a whole face region represented as a vector and thus can not effectively recognize the faces with illumination variations and partial occlusions. In this paper, we present a novel algorithm, termed as Local Ridge Regression (LRR). Different from RR, LRR emphasizes on each local face region matching rather than the whole. As a result, LRR can not only enhance the robustness to the local variations by utilizing the spatial and geometrical information of facial components, but also avoid the dimensionality reduction in the holistic RR as a preprocessing. Furthermore, an efficient cross-validation algorithm is adopted to select the regularization parameters in each local region. Experiments on two standard face databases demonstrate that the proposed algorithm significantly outperforms RR and the two popular linear face recognition techniques (Eigenface and Fisherface). Although we concentrate on ridge regression in this paper, following the proposed line of the research, many current multi-category classifiers can also be applied in face recognition through combining the characteristics of face images and may be obtain better recognition accuracies.

---

[*] Corresponding author: Tel: +86-25-84896481 Ext. 12106; Fax: +86-25-84498069; E-mail: s.chen@nuaa.edu.cn (S. Chen), xuehui@nuaa.edu.cn (H. Xue) and lianyi_1999@nuaa.edu.cn (Y. Zhu)

Keywords: Ridge regression (RR); Local information; Local ridge regression (LRR); Face recognition

1. Introduction

In the past few decades, face recognition has become a hot issue of research in computer vision community. Among various developed techniques in this field, appearance-based method is one of the most widely used techniques which usually represents a face image as a high dimensional vector of pixels[1]. To overcome the difficulty incurred by high dimension, a lot of subspace methods have been proposed where Eigenface[2] and Fisherface[3] are two of the most popular algorithms. Eigenface is an unsupervised method which utilizes the idea of Principle Component Analysis (PCA) to project the original high dimensional data onto a low dimensional subspace that can maximally preserve original image information[4]. Fisherface is a supervised algorithm which combines PCA and Linear Discriminant Analysis (LDA) to extract the most discirminant features that can maximally separate the images of different classes in the resultant face subspace. Although Fisherface further introduces the class information compared to Eigenface, it is affected heavily by the relative positions of the labeled training images due to the weakness of LDA. An et al.[1] have indicated that in the multi-category face recognition problems, while LDA tries to maximize the between-class distances and minimize the within-class distances simultaneously, the pairwise distances can be significantly unbalanced and this may result in bad performance for classes with small pariwise between-class distances in the reduced subspace.

Recently, a new generalized Ridge Regression (RR) method[1] has been proposed to solve the latent problem of Fisherface. Motivated by the fact that the m vertices of a regular m-simplex is the most balanced and symmetric separate points in the (m-1)-dimensional space, the method first encodes the targets for m distinct individuals as the m vertices and then applies the ridge regression to map the training face images into the (m-1)-dimensional subspace so as to the images from each individual will locate near their individual targets[1]. Recognition is performed by mapping the new face images into the subspace and comparing its distance to all the targets. However, although RR has yielded much better recognition performance than Eigenface and Fisherface experimentally, like the two methods, RR is also a holistic technique which operates directly on a whole face region and neglects the local information. As a result, RR is sensitive to the local variations in face images, such as illumination variations and partial occlusions.

For using as much local information hidden in face images as possible to relax the influence of local variation for recognition, recently various local region matching techniques have been developed[4-10]. The general idea of local region matching techniques is to first locate several facial features (components), and then classify the faces by comparing and combining the corresponding local statistics[11]. Heisele et al.[12] further indicated that comparing the component (local) system and the global systems, the former outperforms the latter in recognition rate larger than 60%[11]. Consequently, in this paper, we also apply local region matching techniques into RR and present a novel algorithm, termed as Local Ridge Regression (LRR). LRR first partitions an originally whole image into L equally sized local regions in non-overlapping or overlapping ways, and then collects all those local regions sharing the same original feature components respectively from the training set to compose L corresponding local region training sets. Ridge regression is performed on each of such L local region sets to directly train different L classifiers by the different face features. The new unlabeled face image is identified by also partitioning into L local regions in the same way as the training phase and classifying each local region by the corresponding classifier. The final recognition result is obtained by assembling the total L results from the L classifiers and voting. In this way, not only is the spatial and geometrical information in a face image preserved in each local region, but also the influence of local variations is restrained in several local regions by the classifiers' voting so as to greatly improve the recognition accuracy.

There are three major contributions of the proposed LRR. First, LRR is more robust to the local variations than the holistic RR. Second, LRR can simultaneously train a set of classifiers corresponding to different local regions and thus it is quite suitable for parallel computation to greatly improve the computational efficiency of the holistic RR, especially in the large-sized image cases. Third, the partition of local regions in LRR is independent on the dimension of the face image. Consequently, it can avoid the dimensionality reduction in the holistic RR as a preprocessing. Furthermore, it can also escape the latent dimensional curse when the dimension of the images is quite large.

The rest of the paper is organized as follows. In section 2, we briefly review the holistic RR method. Section 3 presents the proposed LRR. Section 4 provides experimental results on two face databases to illustrate the superiority of LRR. Some conclusions are drawn in Section 5.

2. Ridge Regression (RR)

Suppose there are *m* individuals for recognition. RR algorithm usually has three parts: labeling training images, learning classifier and recognition.

Firstly, RR chooses the regular simplex vertices as the individual targets and uses these targets as the multivariate labels of the training images[1]. Let $\mathbf{T}_i \in \mathbf{R}^{m-1}$ ($i = 1, 2, \cdots, m$) are the vertices of one regular *m*-simplex and $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \cdots, \mathbf{T}_m]$. RR constructs $\mathbf{T}$ as follows[1]:

1. Let $\mathbf{T}_1 = [1, 0, \cdots, 0]^T$ and $\mathbf{T}_{1,i} = -1/(m-1)$, for $i = 2, \cdots, m$.
2. For $1 \le k \le m-2$,

$$T_{k+1,k+1} = \sqrt{1 - \sum_{i=1}^{k} T_{i,k}^2}$$

$$T_{k+1,j} = -\frac{T_{k+1,k+1}}{m-k-1}, \quad j = k+2, \cdots, m \tag{1}$$

$$T_{i,k+1} = 0, \quad k+1 < i \le m-1$$

Then, RR treats the face recognition as a ridge regression problem to locate the images from each individual as near their individual targets as possible. As a result, the task of learning classifier in RR is to find a matrix $\mathbf{W}$ that can model the linear dependency between the image $\mathbf{x}_i$ and the label $\mathbf{Y}_i$, where $\mathbf{Y}_i = \mathbf{T}_j$ if $\mathbf{x}_i$ belongs to the *j*th individual, $i = 1, \cdots, n$, $j = 1, \cdots, m$. Meanwhile, RR also penalizes the norm of $\mathbf{W}$ to reduce the variance of the estimate as the regularization term. Therefore, the objective function (2) of RR is minimizing

$$J(\mathbf{W}) = \sum_{i=1}^{n} \left\| \mathbf{Y}_i - \mathbf{W}^T \mathbf{x}_i \right\|^2 + \lambda \left\| \mathbf{W} \right\|^2 \tag{2}$$

where λ is the regularization parameter to balance the bias and variance of the estimate.

Taking derivative of (2) with respect to the W and equaling it to zero, we have the matrix $\mathbf{W}$ as

$$\mathbf{W} = (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I})^{-1} \mathbf{X}\mathbf{Y}^T \tag{3}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ and $\mathbf{Y} = [\mathbf{Y}_1, \cdots, \mathbf{Y}_n]$.

Finally, let $\mathbf{x}$ be a new image. RR compares the distances between $\mathbf{W}^T \mathbf{x}$ and the individual target $\mathbf{T}_i$ and identifies $\mathbf{x}$ as that with minimal distance.

3. Local Ridge Regression (LRR)

In this section, we propose a novel algorithm to solve the sensitivity of local variations in RR. Relatively to the holistic RR, here we abuse the terminology, i.e. local, to name the proposed algorithm as Local Ridge Regression (LRR). Following the line of the research in the local region matching methods, LRR also involves three steps: local region partition, classifier training and classification. It is noteworthy that LRR avoids the general dimension reduction preprocessing in the classifier training phase due to the partition, and thus it is much simpler.

3.1. Local Region Partition

Generally, there are two different techniques to implement the partition, that is, local components and local regions. Local components are areas occupied by the facial components, such as eyes, noses and mouths, and centered independently at the component centers; Local regions are local windows centered at designated coordinates of a common coordinate system[11]. Zou et al.[11] have verified that comparison of corresponding local regions is better than comparing corresponding facial components. So, in this paper, we adopt the simplest rectangular regions to partition images, which not only are conveniently used but also can better preserve the spatial and geometrical information in the original images[4-6].

Suppose that there are n $W_1 \times W_2$ images belonging to m individuals in the training set, and these individuals possess $n_1, n_2, \cdots, n_m$ face images respectively. Each image is first divided into L equally sized local regions in a non-overlapping way which are further concatenated into corresponding column vectors with dimensionality of $W_1 \times W_2 / L$. Then we collect these vectors at the same position of all face images to form a specific local region training set, in this way, L separate local region sets are formed[5]. This process is illustrated in Fig. 1.

Non-overlapping partition sometimes may divide the relation between each sub-image and lead to totally neglect the relation between local regions. Consequently, we also attempt the overlapping partition way which can connect the adjacent local regions and combine the different information in each regions. The process is illustrated in Fig. 2. In Section 4, we will verify the conjecture that comparison of overlapping regions is better than comparing non-overlapping regions.

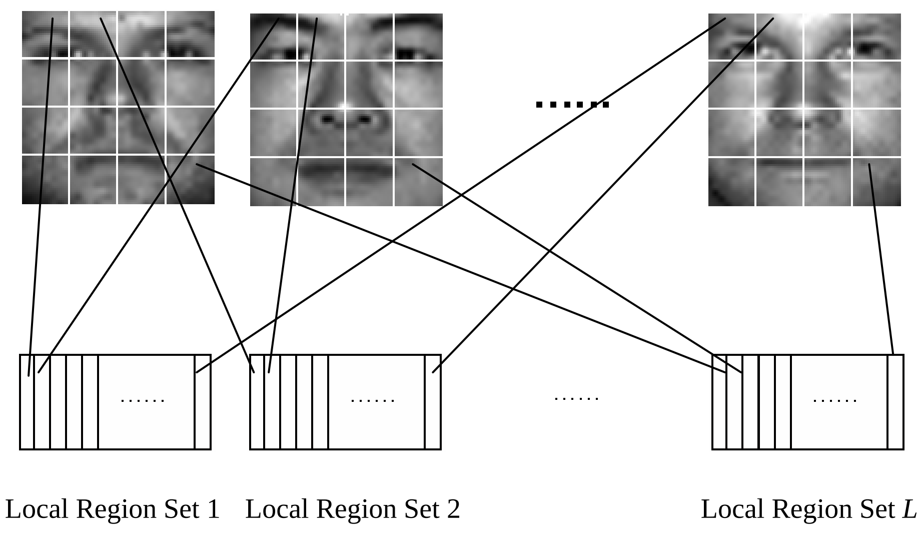

Fig.1. The construction of local region face image sets (Images are from the Extended Yale face database B[13])

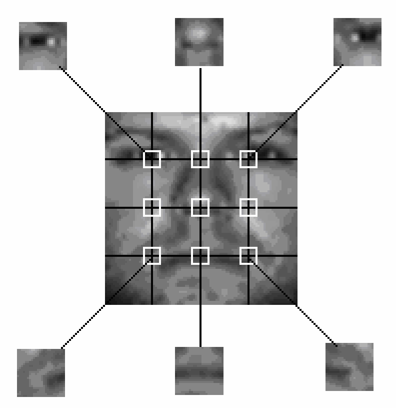

Fig.2. The construction of overlapping local region face image set (The image is also from [13])

3.2. Classifier Training

After partitioning L local regions, we can apply the ridge regression in each local region set to find the corresponding matrix $\mathbf{W}^d$, $d=1,\cdots,L$.

$$\mathbf{W}^d = [\mathbf{X}^d(\mathbf{X}^d)^T + \lambda\mathbf{I}]^{-1}\mathbf{X}^d\mathbf{Y}^T \tag{4}$$

where $\mathbf{X}^d = [\mathbf{x}_1^d, \mathbf{x}_2^d, \cdots, \mathbf{x}_n^d]$ is the pixel matrix of the dth local region set of the training images, and $\mathbf{Y} = [\mathbf{Y}_1, \cdots, \mathbf{Y}_n]$ is the label matrix as in the original RR algorithm. λ is the regularization parameter. Due to the partition independent on the dimension of the images, the LRR algorithm avoids the dimension reduction preprocessing in the original RR and directly learns the local classifiers.

Especially, the regularization parameter λ is a crucial hyper-parameter in the LRR which controls the good generalization performance of the trained classifier. Hence, here we will discuss the option of λ in detail. A popular way to estimate λ is cross-validation. In k-fold cross-validation, the training dataset is randomly split into k disjoint subsets. A classifier is trained for k times on stochastic k -1 subsets and a subset is left out as the validation set to be used for estimating the generalization error at the same time[14]. Finally, the classifier corresponding to the parameter with the lowest average estimated risk is chosen. However, the original implementation of k-fold cross-validation trains a predictor for each split of the data and thus has much expensively computational complexity if k is large[1]. An et al.[1] further developed an efficient technique for general k-fold cross-validation of the generalized RR with multivariate labels. So, here we also adopt this technique to estimate λ in each local region set.

Generally, for a new image $\mathbf{x}$, we first partition it into L local regions as in the training images. Then the corresponding predicted label in each local region is

$$\begin{aligned}\mathbf{Y}^d(\mathbf{x}) &= (\mathbf{W}^d)^T \mathbf{x}^d \\ &= \mathbf{Y}(\mathbf{X}^d)^T[\mathbf{X}^d(\mathbf{X}^d)^T + \lambda\mathbf{I}]^{-1}\mathbf{x}^d \\ &= \mathbf{Y}[(\mathbf{X}^d)^T\mathbf{X}^d + \lambda\mathbf{I}]^{-1}(\mathbf{X}^d)^T\mathbf{x}^d \\ &\triangleq (\mathbf{A}^d)^T(\mathbf{X}^d)^T\mathbf{x}^d \end{aligned} \tag{5}$$

where $\mathbf{x}^d$ denotes the pixel vector of the dth local region in the new image, and

$$\mathbf{A}^d = [(\mathbf{X}^d)^T\mathbf{X}^d + \lambda\mathbf{I}]^{-1}\mathbf{Y}^T \tag{6}$$

In k-fold cross-validation, we split the data into k approximately equally sized subsets $\{\mathbf{x}_{l,i}\}_{i=1}^{n_l}$, $l = 1,\cdots,k$. As in [1], we also split the label matrix $\mathbf{Y}$ and $\mathbf{A}^d$ into k sub-matrices as follows:

$$\mathbf{Y}^T = \left[\mathbf{Y}_{(1)}, \mathbf{Y}_{(2)}, \cdots, \mathbf{Y}_{(k)}\right]^T, \quad \mathbf{A}^d = \left[\mathbf{A}^d{}_{(1)}, \mathbf{A}^d{}_{(2)}, \cdots, \mathbf{A}^d{}_{(k)}\right]^T \tag{7}$$

where $\mathbf{Y}_{(l)} = \left[\mathbf{Y}_{l,1}^T, \mathbf{Y}_{l,2}^T, \cdots, \mathbf{Y}_{l,n_l}^T\right]^T$.

Without loss of generality, we leave the lth subset aside as the validation set. Then the corresponding predicted labels $\mathbf{Y}_{cv}^{(l)}$ can be directly computed as follows:

$$\mathbf{Y}_{cv}^{(l)} = \mathbf{Y}_{(l)} - \mathbf{B}_{ll}^{-1}\mathbf{A}_{(l)}^d, \quad l = 1,\cdots,k \tag{8}$$

where

$$\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1k} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{1k}^T & \mathbf{B}_{2k}^T & \cdots & \mathbf{B}_{kk} \end{bmatrix} \triangleq [(\mathbf{X}^d)^T\mathbf{X}^d + \lambda\mathbf{I}]^{-1} \tag{9}$$

and $\mathbf{B}_{ij} \in \mathbf{R}^{n_i \times n_j}$, for $i, j = 1,\cdots,k$. For more details, the readers can refer to [1].

We identify each local region $\mathbf{x}_{cv}^d$ of the validation images by comparing the distances from the predicted labels $\mathbf{Y}_{cv}^{(l)}$ to the individual targets $\mathbf{T}_i$, $i = 1,\cdots,m$. Then we sum up all recognition errors in the k-fold cross-validation in the corresponding local region set and choose the optimal parameter λ with the minimal error for each local classifier.

In summary, the procedure of classifier training in each local region set can be formally stated as follows:

Input: The local region set $\mathbf{X}^d = [\mathbf{x}_1^d, \mathbf{x}_2^d, \cdots, \mathbf{x}_n^d]$, $d = 1, \cdots, L$

Output: λ and $\mathbf{W}^d$.

1. Label the multivariate label $\mathbf{Y}_i$ of $\mathbf{x}_i^d$ as the regular simplex vertices $\mathbf{T}_j$, $j = 1, \cdots, m$;
2. Choose the regularization parameter λ:
   - 2.1. Compute $[(\mathbf{X}^d)^T \mathbf{X}^d + \lambda \mathbf{I}]^{-1}$;
   - 2.2. Compute $\mathbf{A}^d$ and $\mathbf{B}_{ll}$ from (6) and (9) respectively;
   - 2.3. Compute the predicted label $\mathbf{Y}_{cv}^{(l)}$ from (8), $l = 1, \cdots, k$;
   - 2.4. Identify the validation images $\mathbf{x}_{cv}^d$ by comparing the distances from the predicted labels $\mathbf{Y}_{cv}^{(l)}$ to the individual targets $\mathbf{T}_j$, $j = 1, \cdots, m$;
   - 2.5. Sum up all recognition errors in the *k*-fold cross-validation and choose the optimal parameter λ with the minimal error;
3. Compute $\mathbf{W}^d$ from (4).

3.3. Classification

For an unknown face image $\mathbf{x}$, we will classify it by classifiers' voting. In this way, the influence of local variations, such as illumination variations and partial occlusions, will be restricted in the several local regions so as to greatly improve the recognition robustness to the variations. As described in Section 3.2, we first partition $\mathbf{x}$ into L local regions. Then in each local region, the image's identity is determined by comparing the distances from the predicted label produced by the corresponding local classifier to the individual targets $\mathbf{T}_i$ s. Since one classification result for the unknown image is generated independently in each local region, there will be total L results from L local regions. Let the probability of the image $\mathbf{x}$ belonging to the *c*th class is[6]

$$P_c = \frac{1}{L}\sum_{i=1}^{L} q_c^i \tag{10}$$

where $q_c^i = \begin{cases} 1, & if \ \ the \ \ ith \ \ sub-image \ \ belongs \ \ to \ \ the \ \ cth \ \ class \\ 0, & otherwise \end{cases}$

Then the final classification result is

$$Identity(\mathbf{x}) = \arg\max_{1 \le c \le m}(P_c) \tag{11}$$

4. Experiments

4.1. Face Image Databases

We carry out our experiments on two face image databases: the AR face database[15] and the Extended Yale face database B[13].

The AR database contains 100 individuals with different facial expressions, illumination conditions and occlusions. Each individual has 26 face images taken in two sessions. The first session has 13 face images named from 01 to 13, including neutral expression (01), different facial expression (02-04), different lighting (05-07), occlusions with sunglasses (08-10) and a scarf (11-13) under different lighting. The second session exactly duplicates the first session two weeks later[11]. For psychophysical experiments have indicated that eye is most important for recognition[16, 17], here we omit the pictures occluded by sunglasses. We use "01-07" pictures in the first session from each individual as gallery. And our experiments are conducted on four probe sets: AR11-13 ("11", "12" and "13" pictures, occlusions with a scarf in Session 1), AR15-17 ("15", "16" and "17" pictures, different expressions in Session 2), AR18-20 ("18", "19" and "20" pictures, different lighting conditions in Session 2), AR24-26 ("24", "25" and "26" pictures, occlusions with a scarf in Session 2). The 2000 images are all cropped into the same size of 66×48 pixels.

The extended Yale face database B contains 38 individuals and around 64 near frontal images under different illuminations per individual. All image data are manually aligned, cropped and then resized to 32×32 pixels just as in [1]. A random subset with *l* (*l*=5, 10, 20, 30) images per individual is taken with labels to form the training set, and the rest of the database is the testing set[1]. For each given l, the experiments is repeated over 50 random splits by using the matlab data files in [13], and the average results are reported.

4.2. Evaluation of Classification Performance

We compare the proposed LRR with the most popular face recognition methods: Eigenface, Fisherface and the originally holistic RR on the two face databases. In LRR, we attempt the different partition sizes according to the different sizes of the images in the databases by cross-validation and the selected experimental results are reported. In the AR database, the sizes of local regions are 11×8, 22×16 and 33×24; And in the extended Yale database B, the sizes are designated to 4×4, 8×8 and 16×16 respectively. We also attempt the non-overlapping and overlapping partition ways. In the overlapping way, the adjacent local regions overlap each other almost 50%. The corresponding classification results are listed in Table 1 and 2 respectively, where Eigenface, Fisherface and the holistic RR all involve the dimension reduction and the optimal results are reported.

As shown in Table 1, LRR is significantly superior to the other three holistic methods in all the probe sets in the AR database, basically within all combinations of the size of local regions and non-overlapping or overlapping way. Especially in the AR11-13, AR18-20 and AR24-26 corresponding to different light conditions and occlusions with a scarf, LRR shows the surprisingly high robustness to these local variations and the optimal recognition error rates are less than 50% of those in the other algorithms.

The similar conclusion can also be drawn in the extended Yale database B in Table 2. The face images in the database mostly have pose and illumination variations. LRR also shows the best classification performance corresponding to all the different training sets. Especially, when the numbers of the training images are smaller, such as in 5Train and 10Train, LRR achieves much better recognition accuracies.

Furthermore, it is noteworthy that, the options of the appropriate size of local regions and overlapping way are still open problems. However, here we can capture some empirical observations about the options in LRR. Obviously, from Tables 1 and 2, in the two databases, the optimal classification performances of LRR are accomplished both in the middle partition size and in overlapping way, which even exceed the other combinations in LRR over 50%. These exactly accords with our conjecture in Section 3.1, that is, the middle-size and overlapping partition way can connect more spatial and geometrical information in adjacent local regions.

Table 1. Classification performance (error rate %) comparison on the AR face database

	AR11-13	AR15-17	AR18-20	AR24-26
Eigenface	89.00	22.00	23.33	94.67
Fisherface	64.00	19.00	16.33	83.33
RR	35.33	8.33	4.67	66.00
LRR 11×8, non-overlapping	23.00	18.00	3.67	50.67
LRR 11×8, overlapping	17.67	15.33	4.33	44.67
LRR 22×16, non-overlapping	24.33	12.67	2.33	54.33
LRR 22×16, overlapping	***11.00***	***8.00***	***0.67***	***33.33***
LRR 33×24, non-overlapping	30.33	14.33	4.67	50.33
LRR 33×24, overlapping	23.00	8.67	1.00	41.67

Table 2. Classification performance (error rate %) comparison on the extended Yale face database B

	5 Train	10 Train	20 Train	30 Train
Eigenface	63.60	46.40	30.40	22.50
Fisherface	24.50	12.40	8.70	13.30
RR	23.80	12.00	4.77	2.28
LRR 4×4, non-overlapping	52.09	43.46	37.12	35.23
LRR 4×4, overlapping	26.33	16.39	11.12	7.67
LRR 8×8, non-overlapping	11.96	4.85	2.63	1.78
LRR 8×8, overlapping	***10.24***	***3.85***	***2.03***	***1.42***
LRR 16×16, non-overlapping	19.42	8.63	4.26	2.29
LRR 16×16, overlapping	23.24	11.13	5.26	3.02

5. Conclusions

In this paper, we have proposed a new face recognition technique LRR based on the insights of the originally holistic RR. To overcome the sensitivity of RR to local variations, LRR adopts the popular local region matching techniques. As a result, LRR not only enhances the robustness to the variations, but also effectively avoids the latent dimensional curse when the dimension of the images is very large. Experimental results demonstrate the surprisingly good classification performance of LRR. It is worth to note that, although we concentrate on the improvement of RR in the whole paper, the proposed line of the research about LRR is general. Through combining the spatial and geometrical information of facial components in local way, many current multi-category classifiers can also be applied in face recognition and may obtain better recognition performance, which deserves our future researches.

Furthermore, tensor subspace models are one of the modern research directions in the face recognition. Many researches have showed that representing the images as tensors of arbitrary order can further improve the performance of algorithms in most cases[18-24]. Consequently, how to generalize LRR to tensor learning is another interesting topic for future study.

Acknowledgment

This work was supported by National Natural Science Foundations of China (60773060 and 60773061) and Natural Science Foundations of Jiangsu Province (BK2006187 and BK2008xxx). We also thank Prof. Xiaoyang Tan for beneficial discussion.

References

[1] S. An, W. Liu, and S. Venkatesh. Face recognition using kernel ridge regression. CVPR, 2007.

[2] M. Turk and A. Pentland. Eigenfaces for recognition. J. Cognitive Neurosci,3(1): 71-86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfances vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence,19(7): 711-720, 1997.

[4] S. Chen and Y. Zhu. Subpattern-based principle component analysis. Patter Recognition,37(5): 1081-1083, 2004.

[5] K. Tan and S. Chen. Adaptively weighted sub-pattern PCA for face recognition. Neurocomputing,64: 505-511, 2005.

[6] Q. Hong, S. Chen, and X. Ni. Sub-pattern canonical correlation analysis with application in face recognition. ACTA AUTOMATICA SINICA,34(1): 21-30, 2008 (In Chinese).

[7] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. CVPR: 84-91, 1994.

[8] A.M. Martinez. Recognizing imprecisely localized, partically occluded and expression variant faces

from a single sample per class. IEEE Trans. Pattern Analysis and Machine Intelligence,24(6): 748-763, 2002.
[9] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. Image and Vision Computing, 2008.
[10] X. Geng and Z. Zhou. Image region selection and ensemble for face recognition. JCST, 2006.
[11] J. Zou, Q. Ji, and G. Nagy. A comparative study of local matching approach for face recognition. IEEE Trans. on Image Processing,16(10): 2617-2628, 2007.
[12] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. Comput. Vis. Image Understand.,91(1): 6-12, 2003.
[13] The Extended Yale Face Database B. http://ews.uiuc.edu/dengcai2/Data/data.html.
[14] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. Wiley, 2001.
[15] A.M. Martinez and R. Benavente. The AR Face Database. CVC Technical Report #24, June, 1998.
[16] D.I. Perrett, P.A.J. Smith, D.D. Potter, A.J. Mistlin, A.S. Head, A.D. Milner, and M.A. Jeeves. Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc. of the Royal Society of London. Series B: Biological Sciences,223: 293-317, 1985.
[17] E. Bart and S. Ullman. Class-based feature matching across unrestricted transformations. IEEE Trans. on Pattern Analysis and Machine Intelligence,30(9): 1618-1631, 2008.
[18] H. Zhou, Y. Yuan, and A.H. Sadka. Application of semantic features in face recognition. Patter Recognition,41(10): 3251-3256, 2008.
[19] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang. Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval. IEEE Trans. on Image Processing,16(11): 2811-2821, 2007.
[20] D. Tao, X. Li, X. Wu, and S.J. Maybank. General tensor discriminant analysis and Gabor features for gait recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence,29(10): 1700-1715, 2007.
[21] D. Tao, X. Li, X. Wu, W. Hu, and S.J. Maybank. Supervised tensor learning. Knowledge and Information Systems,13(1): 1-42, 2007.
[22] T. Zhang, D. Tao, and J. Yang. Discriminative locality alignment. The 10th European Conference on Computer Vision (ECCV), 2008.
[23] D. Tao, J. Sun, J. Shen, X. Wu, X. Li, S.J. Maybank, and C. Faloutsos. Bayesian tensor analysis. IEEE International Joint Conference on Neural Networks (IJCNN), 2008.
[24] T. Zhang, K. Huang, X. Li, J. Yang, and D. Tao. Discriminative orthogonal neighborhood preserving projections for classification. IEEE Trans. on Systems, Man, and Cybernetics, Part: B, 2008.