

# Spectral Regression based Local Discriminant Embedding Algorithm for Face Recognition

Bei Huang

School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Email:huangbei\_seu@seu.edu.cn

## Abstract

Local discriminant embedding algorithm (LDE) can get better recognition performance than the conventional dimensionality reduction algorithms based on subspaces techniques, but LDE is weak generalization performance for high dimension small sample and has huge workload to decompose dense matrix. In this paper, the SR-LDE algorithm is proposed. The spectral regression method is introduced into LDE to improve its generalization performance and reduce complexity for dense matrix decomposition. The experiments show that SR-LDE algorithm has better performance on recognition rate and computing speed.

**Keywords:** Face Recognition, Local Discriminant Embedding, Spectral Regression

## 1 Introduction

Face recognition is now widely applied in E-bank, E-commerce, intelligent building and safety monitoring. Although the research has been carried out for several decades, more and better techniques for face recognition are provided. However many unwanted variations result from changes in lighting, facial expression and pose, so the face image space might not be an optimal space for visual perception. In this paper, a spectral regression based Local Discriminant Embedding algorithm (SR-LDE) is proposed to reduce the time of training samples and improve the facial recognition efficiency.

The rest of this paper is organized as follows: Section 2 describes previous work on face recognition. The Spectral Regression based Local Discriminant Embedding algorithm is proposed. Section 4 presents the manifold methods of face analysis in ORL database. Some experimental results are presented. Finally, we provide some concluding remarks.

## 2 Related work

Elastic Bunch Graph Matching algorithm<sup>[1-2]</sup> is

one of early important achievements on face recognition techniques. The algorithm greatly reduced the number of image feature required for face recognition, but didn't reduce the recognition rate. Based on the algorithm, Elastic Graph Dynamic Link Model was proposed to simplify the traditional Dynamic Link Model and integrate it with the Active Contour Model for feature extraction. The method greatly improved the average recognition speed and rate. And Lee expressed a face recognition technique that effectively combines Elastic Graph Matching and Fisherface algorithm<sup>[4]</sup>. In the algorithm, the linear projection per node of an image reduces dimensionality of labeled graph vector. So it could obtain satisfactory results in the recognition rate and speed. In the Eigenfaces method<sup>[5]</sup>, the face images were transformed into a small set of characteristic feature images, which were the principal components of the initial training set of face images, without requiring that they correspond to our intuitive notions of facial parts and features. The method reduced the eigen-dimensionality of the face images, but lower classification performance. Samaria first introduced the Hidden Markov Models into face recognition (HMMs). Through the integration of a priori structural knowledge with statistical information, HMMs can be used successfully to encode face features. But the method is based on grey-level templates, thus lighting changes is sensitive for training images. And it can require a large number of calculations. So Liu<sup>[7]</sup> presented a new face recognition approach based on HMMs and singular value decomposition, which used sample singular values as observation vectors. The method reduced the computing time and increased the recognition rate, but varying illumination conditions still affect the appearance of face images significantly. Du provided a face image pre-processing approach that dealt with the illumination problem to make face recognition robust to illumination variations<sup>[8]</sup>, like as literature 9. In the method, a face image was converted into logarithm domain first, then discrete cosine transform coefficients were used to remove illumination variations.

And Flming et al<sup>[10]</sup> introduced BP neural network into face recognition to improve robustness of face image illumination changes or damaged. However Intrator et al<sup>[11]</sup> used a hybrid (supervised and unsupervised) neural network to carry out the classification of the normalized images. The features the network extracted are more obvious than that BP network extracted. Thus the recognition rate of the network was superior to that of BP. Lin et al<sup>[12]</sup> proposed a face recognition system based on probabilistic decision-based neural networks (PDBNN). The system performs human face detection, eye localization and face recognition in close-to-real-time speed. Because discriminant function obeys probability constraint, PDBNN has more nice properties such as low false acceptance and false rejection rates.

The above, which was traditional subspace analysis method, had obtained the better effect in the pattern recognition, but the optimal reconfiguration of all samples was its purpose, and measured the classification error existed between different samples. Thus these methods could only find the global Euclidean structure of all samples, but their local distribution structure was not clear enough. When the data is highly nonlinear, this kind of method could not effectively recover the internal structure of the data, could only describe its statistical distribution. Therefore Seung et al<sup>[13]</sup> studied the manifold learning problems in the view of neurophysiology and hypothesized that a visual memory was stored as a manifold of stable states, and referred to the image data were manifold in high dimension space. The Isomap algorithm<sup>[14]</sup> and Locally Linear Embedding algorithm<sup>[15]</sup>, two manifold learning algorithm found meaningful low-dimension structures hidden in their high dimension observations. The Laplacianface approach<sup>[16-17]</sup> used Locality Preserving Projections (LPP) to map the face images into a face subspace and were the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the face manifold. The method provided a better representation and achieved lower error rates in face recognition. Subsequently, UDP<sup>[18]</sup>、NPE<sup>[19]</sup>、MFA<sup>[20]</sup> 和 LDE<sup>[21]</sup> were proposed. In addition, the classification training way<sup>[22]</sup> can be used to improve the face recognition accuracy.

In LDE algorithm, the neighbor and class relations of data were used to construct the embedding for classification problems. And the algorithm learned the embedding for the submanifold of each class to solve optimization problem. But LDE is weak generalization performance for high dimension small sample and the eigen-decomposition of dense matrices is expensive in both time and memory. The

spectral regression can avoid these problems, thus a revised algorithm SR-LDE was proposed. The main contributions of this paper is that the spectral regression theory is introduced into LDE algorithm, and SR-LDE algorithm is proposed to resolve the issue of feature extraction of the  $n \times n$  dimension, dense matrix. The method can reduce the complication of LDE algorithm.

### 3 Spectral Regression based LDE Algorithm

The objective function of LDE algorithm is as follows:

$$\mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}(\mathbf{D}^p-\mathbf{W}^p)\mathbf{X}^T\mathbf{a} \quad (1)$$

Where,  $\mathbf{X}$  denotes observation set of images in high dimension space,  $\mathbf{D}$  and  $\mathbf{D}^p$  are diagonal matrices,  $\mathbf{W}$  and  $\mathbf{W}^p$  are weight matrices, and

$$D_{ii}=\sum_j W_{ij}, \quad D_{ii}^p=\sum_j W_{ij}^p.$$

**Theorem 1.** Let  $\mathbf{y}$  be the eigenvector of eigen-problem in  $\mathbf{L}\mathbf{y}=\lambda\mathbf{B}\mathbf{y}$  with eigenvalue  $\lambda$ . If  $\mathbf{y}=\mathbf{X}^T\mathbf{a}$ , then  $\mathbf{a}$  is the generalized eigenvector of eigen-problem in  $\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{a}$  with the same eigenvalue  $\lambda$ .

In  $\mathbf{y}=\mathbf{X}^T\mathbf{a}$ , such  $\mathbf{a}$  might not exist. So a possible way is to find  $\mathbf{a}$  which can best fit the equation in the squares sense:

$$\mathbf{a}=\arg\min_a\sum_{i=1}^n(\mathbf{a}^T\mathbf{x}_i-y_i)^2 \quad (2)$$

The optimal solution  $\mathbf{a}$  in Eqn.2 is  $\mathbf{a}=(\mathbf{X}^T\mathbf{X}+\mu\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$ , where  $\mu$  is a parameter to control the amounts of shrinkage. So the appropriate  $\mu$  value can be selected to reduce the deviation and achieve the best effect.

When  $\mu>0$ , this regularized solution  $\mu$  will not satisfy the linear equations system  $\mathbf{y}=\mathbf{X}^T\mathbf{a}$  and  $\mathbf{a}$  will not be the eigenvector of eigen-problem in  $\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{a}$ .

**Theorem 2.** Suppose  $\mathbf{y}$  ( $\mathbf{y}=\mathbf{X}^T\mathbf{a}$ ) is the eigenvector of eigen-problem in  $\mathbf{L}\mathbf{y}=\lambda\mathbf{B}\mathbf{y}$  and  $\mathbf{y}$  is in the space spanned by row vectors of  $\mathbf{X}$ . The corresponding projective function  $\mathbf{a}$  calculated  $\mathbf{a}=(\mathbf{X}^T\mathbf{X}+\mu\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$  will be the eigenvector of eigen-problem in  $\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{a}$  as  $\mu$  decreases to zero.

According to above two theorem, we can get two following corollaries.

**Corollary 1.** If  $\mathbf{y}$  is the eigenvector of eigen-problem in  $(\mathbf{D}-\mathbf{W})\mathbf{y}=\lambda(\mathbf{D}^p-\mathbf{W}^p)\mathbf{y}$  and  $\mathbf{y}=\mathbf{X}^T\mathbf{a}$ , then  $\mathbf{a}$  is the eigenvector of eigen-problem in  $\mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}(\mathbf{D}^p-\mathbf{W}^p)\mathbf{X}^T\mathbf{a}$ .

**Proof**

$$\begin{aligned}
X(D-W)X^T a &= X(D-W)y \\
&= X\lambda(D^P - W^P)y \\
&= \lambda X(D^P - W^P)y \\
&= \lambda X(D^P - W^P)X^T a
\end{aligned}$$

Where  $y$  is low dimension embedding of the high dimension data set  $X$ .

The  $y$  is equal to  $X^T a$ , therefore the optimal solution  $a$  can be calculated by ridge regression as follows:

$$a = (X^T X + \mu I)^{-1} X^T y \quad (3)$$

When  $\mu > 0$ , this regularized solution will not satisfy the linear equations system  $y = X^T a$ .

**Corollary 2.** Suppose  $y(y = X^T a)$  is the eigenvector of eigen-problem in  $X(D-W)X^T a = \lambda X(D^P - W^P)X^T a$ , if  $y$  is in the space spanned by row vectors of  $X$ , The corresponding projective function  $a$  will be the eigenvector of eigen-problem in  $X(D-W)X^T a = \lambda X(D^P - W^P)X^T a$  as  $\mu$  decreases to zero.

**Proof** Suppose rank  $(X) = r$ , the Singular Value Decomposition of  $X$  is

$$X = U\Lambda V^T \quad (4)$$

Where  $\Lambda = \text{diag}(\sigma_1, \dots, \sigma_r)$ ,  $U \in R^{n \times r}$ ,  $V \in R^{m \times r}$ , we have  $U^T U = V^T V = I$ . The  $y$  is in the space spanned by row vectors of  $X$ , therefore  $y$  is in the space spanned by column vectors of  $V$ . Thus  $y$  can be represented as the linear combination of the column vectors of  $V$ . Moreover, due to the column vectors of  $V$  are linear independent, the combination is unique.

Suppose the combination coefficients are  $b_1, \dots, b_r$ ,  $b = [b_1, \dots, b_r]^T$ , we have:

$$\begin{aligned}
Vb &= y \Rightarrow V^T Vb = V^T y \Rightarrow \\
b &= V^T y \Rightarrow VV^T y = y
\end{aligned} \quad (5)$$

To continue our proof, we introduce the concept of pseudo inverse of a matrix. So the pseudo inverse of the matrix  $X$  can be computed as  $X^+ = V\Lambda^{-1}U^T$  or

$$X^+ = \lim_{\mu \rightarrow 0} (X^T X + \mu I)^{-1} X^T$$

If  $X^T X$  is singular matrix, then  $(X^T X)^{-1}$  does not exist. Therefore the regularized least squares solution in Eqn.(3)

$$\begin{aligned}
a &= (X^T X + \mu I)^{-1} X^T y \\
&\stackrel{\mu \rightarrow 0}{=} (X^T)^+ y \\
&= U\Lambda^{-1}V^T y
\end{aligned} \quad (6)$$

Combine with the equation in Eqn.(5), we have:

$$\begin{aligned}
X^T a &= V\Lambda U^T a \\
&= V\Lambda U^T U\Lambda^{-1}V^T y \\
&= VV^T y = y
\end{aligned} \quad (7)$$

So,  $a$  is the eigenvector of eigen-problem

in  $X(D-W)X^T a = \lambda X(D^P - W^P)X^T a$ .

In traditional manifold algorithm, first compute the projective vector  $a$ , then compute the test data set  $X$  by  $a$ . However, in our proposed method, first compute the eigenvector  $y$  of the training samples, next compute the projective vector  $a$  by spectral regression, finally get test data set  $X$ . Thus the SR-LDE algorithm does not eigen-decompose dense matrix with  $n \times n$  dimension. The dimensionality of the eigen-decomposition matrix is  $m \times m$ . In face images, the dimension  $n$  of face feature is usually much large than the dimension  $m$  of face sample, that is  $m \ll n$ . Thus our method can greatly reduce the algorithm complication.

## 4 Results

In this section, in order to investigate the performance of our proposed SR-LDE for face recognition, the ORL face database is used in our experiments. Let  $\mu = 0.01$ , compare the performance of SR-LDE with that of PCA, LDA, LPP and LDE.

### 4.1 Experiment in ORL Database

The ORL face database is set up by AT&T laboratory of Cambridge University, contains 10 different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/ no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The size of each image is  $92 \times 112$  pixels, with 256 grey levels per pixel. In our experiments, the size of each image is normalized to  $64 \times 64$  pixels.

In our experiments,  $G$  ( $G=2, 3, 4, 5$ ) images per person were randomly selected for training and remaining  $10-G$  images for testing. In order to eliminate the influence of random factors, the random selection process is repeated 10 times for the selected  $G$ . We average the results over 10 splits. In different number of training samples and testing samples, the test results of different algorithms are reported on Tab.1 and Fig.1 shows the performance of recognition rate as a function of the eigenvector dimensionality for G2/P8. The test results of G3/G7, G4/G6 and G5/P5 is similar to G2/G8.

Tab.1: Recognition rates of different algorithms on ORL

Num. Algorithm	G2/P8	G3/P7	G4/P6	G5/P5
PCA	70.4% (79)	82.7% (65)	85.2% (82)	88.1% (80)
LDA	77.5% (37)	86.6% (39)	88.5% (39)	93.2% (39)
LPP	79.8% (39)	87.2% (39)	88.9% (38)	94.4% (39)
LDE	80.7% (28)	89.1% (37)	91.5% (42)	95.4% (28)
SR-LDE	83.2% (35)	91.2% (30)	92.6% (48)	96.2% (31)

Where  $G_n/P_n$  means the number of training samples and testing samples respectively, the value in ( ) means eigen-dimensionality when get the optimal recognition rate.

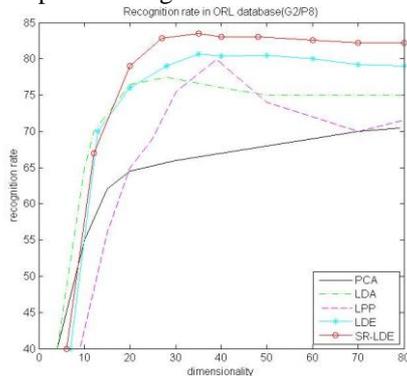


Fig.1: Recognition rate and the dimensionality of each algorithm on ORL

#### 4.2 Analysis

According to Fig.1, the main observations from the performance comparisons include:

(1) Whether the traditional subspace dimensionality reduction algorithm, or manifold learning algorithm, with the increasing number of test samples, the recognition rate of each algorithm is improved, but the manifold learning algorithms have better recognition effect than the traditional dimensionality reduction methods. It is indisputable that the performance of SR-LDE algorithm is best, because the algorithm uses spectral regression to calculate the optimal solution of eigenvector. So it can solve the problems of the small samples and avoid eigen-decomposing dense matrices.

(2) In dealing with large scale and high dimension data set, because of high computational complexity, the traditional graph embedding algorithm is not directly applied to face recognition. However the spectral regression algorithm can avoid decomposing dense matrix, effectively solve the generalized eigenvalue of large scale and high dimension data set, so as to shorten the time of calculation. In ORL, different number of training samples requires training time as shown in Tab. 2.

Tab. 2: different number of training samples and

training time (unit:s)				
algorithm samples	PCA	LDA	LDE	SR-LDE
10×68	3.658	5.230	7.431	1.365
20×68	7.793	8.477	9.278	1.522
30×68	7.949	8.739	9.414	1.756

#### 5 Conclusion

In face recognition algorithms, the manifold learning algorithms have better recognition effect than the traditional dimensionality reduction methods, therefore LDE algorithm is selected as the research object and the spectral regression algorithm is introduced in this paper. The proposed SR-LDE algorithm can solve the problems of the high dimension small sample and the eigen-decomposition of dense matrices. The experiments show that SR-LDE algorithm has better performance on recognition rate and computing speed.

#### ACKNOWLEDGMENT

This research is supported by the National High Technology Research and Development Program (863) of China Grant No. 2013AA014001.

#### References

- [1] L. Wiskott, J. Fellous, N. Kruger et al. Face recognition by elastic bunch graph matching[J]. IEEE Trans. Pattern Anal.Mach.Intell, 1997,19(7):775-779.
- [2] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. EWurtz, W. Konen, Distortion Invariant Object Recognition in the Dynamic Link Architecture[J], IEEE Transactions On Computers, 1993,42(3):300-311.
- [3] R. Lee, J. Liu, J. You. Face recognition:elastic relation encoding and structural matching [C].Proc. of IEEE Intl.Cmf.Sust.Man Cybern, 1999:172-177.
- [4] H. J. Lee, W. S. Lee, J. N. Chung. Face Recognition Using Fisherface Algorithm And Elastic Graph Matching[C].Proc.Int'l Conf.Image Process, 2001.vol.1:998-1001
- [5] M.A.Turk, A.P.Pentland. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.
- [6] F.Samaria. Face Recognition Using Hidden Markov Models[D]. PhD thesis.Univ.of Cambridge.1994
- [7] X.J.Liu, D.F.Wang, L.F.Zhang, Y.G.Shi and M.Y.Zou. An Approach for Face Recognition Based on Singular Value Decomposition and

- Hidden Markov Model[J]. Chinese Journal of Computers, Chinese, 2003, 26(3): 85-89
- [8] S.Du, M.Shehata, W.Badawy. A novel algorithm for illumination invariant DCT-based face recognition[C]. 25<sup>th</sup> IEEE Canadian Conference on Electrical and Computer Engineering:Vision for a Greener Future, 2012.
- [9] D.M.Maria, N.Michele, R.Daniel and W.Harry. Robust Face Recognition for Uncontrolled Pose and Illumination Changes[J]. IEEE Transactions on Systems,Man,and Cybernetics:Systems, 2013,43(1):149-163
- [10] M.Fliming, G.Cottrell. Categorization of faces using unsupervised feature extraction[C]. Proceeding of Conference Neural Networks. 1990:65-70
- [11] N.Intrator, D.reisfeld, Y.Yeshurun. Face recognition using a hybrid supervised/unsupervised neural network[J]. Pattern Recognition Letters,1996, 17(1):67-76
- [12] S. H. Lin,S. Y. Kung and L. J. Lin. Face recognition / detection by probabilistic decision-based neural network[J]. IEEE Transactions on Neural Networks , 1997,8(1): 114-132.
- [13] [13] H.S. Seung, D.D. Lee. The manifold ways of perception[J]. Science, 2000, 290(5500): 2268-2269.
- [14] J.B.Tenenbaum,V.de Silva and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction[J].Science, 2000,290(5500):2319-2323
- [15] S.T.Roweis,L.K.Saul. Nonlinear dimensionality reduction by locally linear embedding[J]. Science,2000,290(5500):2323-2326
- [16] [16] X.F. He, S.C. Yan, Y.X. Hu, N.Partha and H.J.Zhang. Face recognition using laplacianfaces[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence,2005,27(3):328-340
- [17] M.Belkin, P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]. Advances in Neural Information Processing Systems 14, Vancouver, British Columbia, Canada, 2001, Dec.3-8:585-591
- [18] J.Yang, D.Zhang, J.Y.Yang, B.Niu. Globally Maximizing, Locally,Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007,29(4):650-664
- [19] X.F.He, D.Cai, S.C.Yan, H.J.Zhang. Neighborhood preserving embedding[C]. IEEE International conference on computer vision (ICCV), 2005:1208-1213
- [20] S.C.Yan , D.Xu , B.Y.Zhang. H.J.Zhang, Q.Yang and S.Lin. Graph Embedding and Extensions : A General Framework for Dimensionality Reduction[J] . IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007, 29(1): 40—51
- [21] H. T. Chen , H. W. Chang , and T. L.Liu. Local Discriminant Embedding and Its Variants[C] . Proc . IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 846—853.
- [22] F.K.Brenda, J.B.Mark, C.K.Joshua, W.V.B.Richard and K.J.Anil. Face Recognition Performance: Role of Demographic Information[J].IEEE Transactions on Information Forensics and Security, 2012,7(6):1789-1801