Hindawi Publishing Corporation
Discrete Dynamics in Nature and Society
Volume 2009, Article ID 916382, 8 pages
doi:10.1155/2009/916382

Research Article

Locally Linear Discriminate Embedding for Face Recognition

Eimad E. Abusham and E. K. Wong

Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia

Correspondence should be addressed to Eimad E. Abusham, eimad.eldin@mmu.edu.my

Received 21 January 2009; Accepted 12 October 2009

Recommended by B. Sagar

A novel method based on the local nonlinear mapping is presented in this research. The method is called Locally Linear Discriminate Embedding (LLDE). LLDE preserves a local linear structure of a high-dimensional space and obtains a compact data representation as accurately as possible in embedding space (low dimensional) before recognition. For computational simplicity and fast processing, Radial Basis Function (RBF) classifier is integrated with the LLDE. RBF classifier is carried out onto low-dimensional embedding with reference to the variance of the data. To validate the proposed method, CMU-PIE database has been used and experiments conducted in this research revealed the efficiency of the proposed methods in face recognition, as compared to the linear and non-linear approaches.

1. Introduction

Linear subspace analysis has been extensively applied to face recognition. A successful face recognition methodology is largely dependent on the particular choice of features used by the classifier. Linear methods are easy to understand and are very simple to implement, but the linearity assumption does not hold in many real-world scenarios. Face appearance lies in a high-dimensional nonlinear manifold. A disadvantage of the linear techniques is that they fail to capture the characteristics of the nonlinear appearance manifold. This is due to the fact that the linear methods extract features only from the input space without considering the nonlinear information between the components of the input data. However, a globally nonlinear mapping can often be approximated using a linear mapping in a local region. This has motivated the design of the nonlinear mapping methods in this study.

The history of the nonlinear mapping is long; it can be traced back to Sammon's mapping in 1969 [1]. Over time, different techniques have been proposed such as the projection pursuit [2], the projection pursuit regression [3], self-organizing maps or SOM

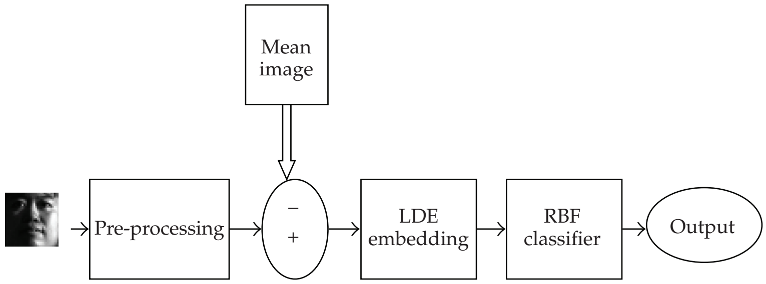

Figure 1: Block diagram of the Locally Linear Discriminate Embedding (LLDE).

[4], principal curve and its extensions [5–8], autoencoder neural networks [9, 10], and generative topographic maps or GTM [11]. A comparison of some of these methods can be found in Mao and Jain [12]. Recently, a new line of nonlinear mapping algorithms was proposed based on the notion of manifold learning. Given a data set that is assumed to be lying approximately on manifold in a high-dimensional space, dimensionality reduction can be achieved by constructing a mapping that respects certain properties of the manifold. Manifold learning has been demonstrated in different applications; these include face pose detection [13, 14], high-dimensional data discriminant analysis [15], face recognition [16–18], analysis of facial expressions [19, 20], human motion data interpretation [21], gait analysis [20, 22], visualization of fibre traces [23], and wood texture analysis [24].

The remainder of this paper is organized as follows. In Section 2, Block diagram of Locally Linear Discriminate Embedding (LLDE) and a Brief review of Locally Linear Discriminate Embedding algorithm are shown. In Section 3, the proposed method is tested on CMU-PIE database and compared to the other methods such as Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA). Finally, a brief conclusion is given.

2. Materials and Methods

2.1. Preprocessing and Normalization

Face preprocessing and normalization is significant parts of face recognition systems. Changes in lighting conditions have been found to dramatically decrease the performance of face recognition. Therefore, all images have been preprocessed to obtain a representation of the face which is invariant to illumination, while keeping the information necessary to allow a discriminative recognition of the subjects. Gaussian kernel has been used to estimate the local mean and standard deviation of images to correct nonuniform illumination. The local normalization is computed as follows:

$$g(x,y) = \frac{f(x,y) - m_f(x,y)}{\sigma_f(x,y)}, \tag{2.1}$$

where $f(x,y)$ is the original image, m is an estimation of a local mean of f, and s is an estimation of the local standard deviation.

Figure 1 illustrates a block diagram of the developed method. All face images have to preprocess to obtain a representation of the face which is invariant to illumination by

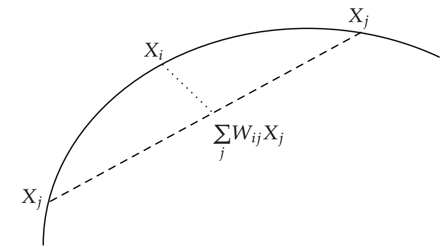

Figure 2: Reconstruction weights of face data and its neighbors.

(2.1). Then we obtain the reconstruction weights by capturing the intrinsic geometry of the neighborhood. The detail of the diagram is illustrated below.

2.2. The LLDE Algorithm

Find reconstruction weights by capturing the intrinsic geometry of the neighborhood. The LLDE creates a locally linear mapping, from the high-dimensional coordinates to the low dimensional embedding, as shown in Figure 2.

Compute the average weight that represent every face data by its neighbors:

$$\varphi(w) = \left\| x_i - \sum_{j=1}^{K} w_{ij} x_{ij} \right\|, \quad x_{ij} \in X \in R^N, \tag{2.2}$$

where x_i means the ith unknown sample, and x_{ij} the corresponding training samples, according to the K values (nearest neighbors).

Computing the low-dimensional embedding D, the following cost function is minimized:

$$\Phi(D) = \sum_{i=1}^{N} \left\| D_i - \sum_{j=1}^{K} W_{ij} D_j \right\|^2, \tag{2.3}$$

where N is the number of training and K is the number of the nearest neighbors.

The RBF classifier is a one hidden layer neural network, with several forms of radial basis activation functions, as follows:

$$f_j(D^*) = \exp \frac{\left\| D^* - \mu_j \right\|^2}{2\sigma_j^2}, \tag{2.4}$$

where σ_j is the width parameter, μ_j is the vector determining the centre of the basis function f, and D^* is the n-dimensional input vector. In an RBF network, a neuron of the hidden layer is activated whenever the input vector is close enough to its central vector. The second layer of the RBF network, that is, the output layer, comprises one neuron to each class. The final classification is given by the output neuron with the greatest output.

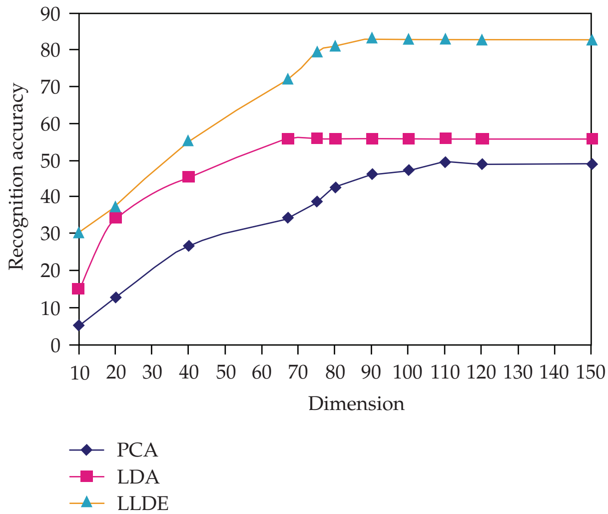

Figure 3: The recognition rates of the PCA, LDA, and LLDE.

3. Results and Discussion

3.1. CMU-PIE Database

One of the largest datasets developed to investigate the affect of Pose, Illumination, and Expression. It contains images of 68 people, each under 13 different poses, 43 different illumination conditions, and 4 different expressions [25]. In the experiments conducted in this study, 6 out of 13 poses for each person were selected. Out of 43-illumination configurations, 21 were selected to typically span the set of variations; these covered the left to the right profile.

The Locally Linear Discriminate Embedding (LLDE) is a manifold learning technique, in which the local geometric properties within each class are preserved, based on the local neighbour structure, and the local structure is generally important for discriminate analysis. Each face image can linearly be approximated by its neighbours, with the same set of coefficients, computed from the high-dimensional data in the input space while minimizing reconstruction cost. For classification, the resulting embedding spaces are fed to Gaussian Radial Basis Function to produce feature vectors. A Gaussian Radial Basis Function could be a good choice for the hidden layers; it is widely used and researched tool for (nonlinear) function approximation, which is a central theme in pattern analysis and recognition. The transformation from the input space to the hidden-unit space is nonlinear. On the other hand, the transformation from the hidden space to the output space is linear.

Suppose that each hidden layer node is a Gaussian Radial Basis Function equation (2.4) and μ_j is the centre of the j the class. The closer D_i to the μ_i is, the higher the value of the Gaussian function will be produced. The outputs of the hidden layer can be viewed as a set of discriminate features, extracted from the input space. Figure 3 shows the plots of the recognition rate versus the dimensionality of the PCA, LDA, and LLDE. The dimensions used are ranging between 10 and 150; based on the figure, the LLDE was shown to significantly outperform the PCA, and LDA. The novelty of the proposed method is to extract discriminate nonlinear features and to solve the problem of using the linear methods to extract features

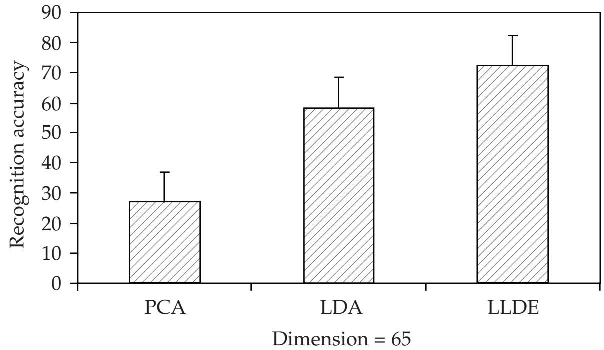

Figure 4: The average recognition rates of the PCA, LDA, and LLDE, across 10 tests (dimension 65).

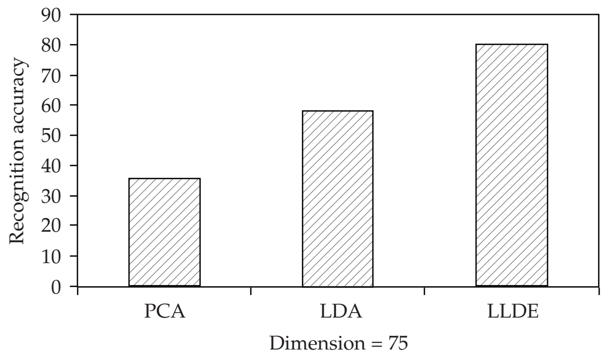

Figure 5: The average recognition rates of the PCA, LDA, and LLDE, across 10 tests (dimension 75).

Table 1: The average error rates (%) of the PCA, LDA, and LLDE, across ten tests and four dimensions.

PCA	LDA	LLDE
60.75	48.15	16.98

from nonlinear manifold; the global nonlinear structure of nonlinear data will be destroyed by applying linear methods so that the recognition rate is dropped down. The proposed LLDE is invariant to translations and rescaling and takes full advantages of the property manifold learning methods that are superior to linear feature extraction methods.

As shown in the figure, the recognition rates of 49.66%, 55.9%, and 83.1% were achieved by the PCA, LDA, and LLDE, with the reduced feature dimensions of 110, 67, and 95, respectively. For performance assessment and consistent, experiments are conducted on reduced selected dimension (65, 75, 90, 110). The average recognition rates are presented in Figures 4, 5, 6, and 7. The upper bound of the dimensionality of LDA is $c-1$, where c is the number of individuals.

Table 1 shows the average recognition error rates, the comparison of the proposed method, and KPCA Plus LDA, and GDA, across ten tests and four dimensions (65, 75, 90, and 110). From the table, it is obvious that the performance of the proposed method is better and it achieves the lowest error rate as compared to the standard linear methods of PCA and LDA.

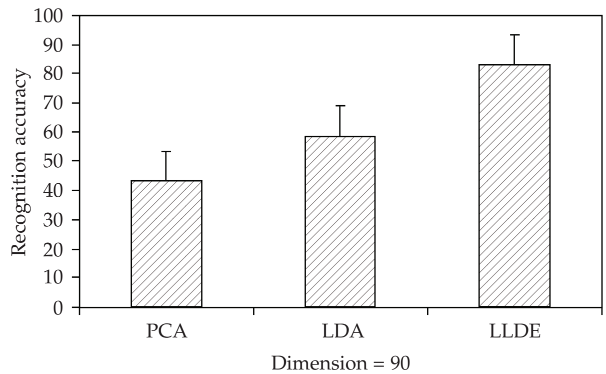

Figure 6: The average recognition rates of the PCA, LDA, and LLDE across 10 tests (dimension 90).

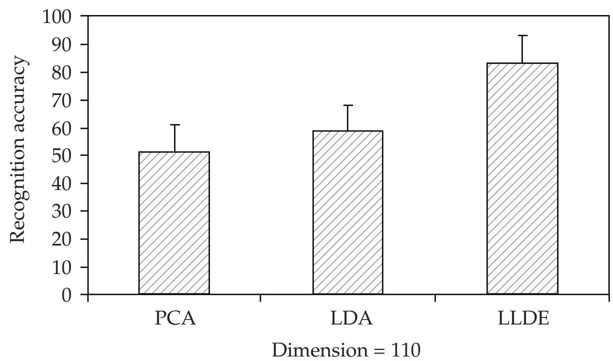

Figure 7: The average recognition rates of the PCA, LDA, and LLDE, across 10 tests (dimension 110).

The comparison of the proposed methods and KPCA Plus LDA [26] and GDA [27] is shown in Figure 8. From the figure, it is clear that the performance of the proposed methods is consistently better as compared to other nonlinear discriminant methods. The method was shown to achieve a maximum accuracy of 83.1%, as compared to only 77.22% and 79.92% by the KPCA and GDA, respectively. This is attributed to the number of the features obtained by the LLDE, which are not limited to $c - 1$, and where c is the number of subjects.

4. Conclusion

Literature surveys and previous studies stated that if linear classifier does not work well, then there are at least two potential reasons for this: (1) regularization was not done well or no robust estimators were used; (2) intrinsically nonlinear: since our dataset is high-dimensional data and the nature of face images is nonlinear, then it is recommended to use an appropriate nonlinear feature space. The proposed method performs an implicit reduction over the whole set of features and effectively extracts the most discriminate features, as shown by the results from the experiments. We think that this is significant when the runtime speed is as important as the actual classification rate: if only a subset of the features is used. In addition to that the proposed method does not suffer from the Small Size (SSS) problem. Our experiments did show clearly that our method is superior to state-of-the art methods.

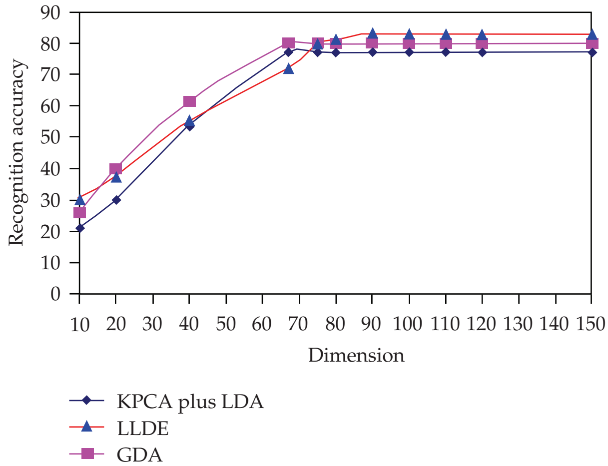

Figure 8: The recognition rates of the KPCA plus LDA, GDA, and LLDE.

References

[1] J. W. Sammon, "A non-linear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. 18, no. 5, pp. 401–409, 1996.

[2] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. 23, pp. 881–890, 1974.

[3] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 817–823, 1981.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer, New York, NY, USA, 2001.

[5] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[6] B. Kégl, A. Krzyzak, T. Linder, and K. Zeger, "Learning and design of principal curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 281–297, 2000.

[7] A. J. Smola, S. Mika, B. Schölkopf, and R. C. Williamson, "Regularized principal manifolds," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 179–209, 2001.

[8] R. Tibshirani, "Principal curves revisited," *Statistics and Computing*, vol. 2, no. 4, pp. 183–190, 1992.

[9] P. Baldi and K. Hornik, "Neural networks and principal component analysis: learning from examples without local minima," *Neural Networks*, vol. 2, no. 1, pp. 53–58, 1989.

[10] D. DeMers and G. Cottrell, "Non-linear dimensionality reduction," in *Advances in Neural Information Processing Systems*, vol. 5, pp. 580–587, MIT Press, Cambridge, Mass, USA, 1993.

[11] C. M. Bishop, M. Svensén, and C. K. I. Williams, "GTM: the generative topographic mapping," *Neural Computation*, vol. 10, no. 1, pp. 215–234, 1998.

[12] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 296–317, 1995.

[13] A. Hadid, O. Kouropteva, and M. Pietikainen, "Unsupervised learning using locally linear embedding: experiments in face pose analysis," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, pp. 111–114, 2002.

[14] S. Z. Li, X. Lv, and H. Zhang, "View-subspace analysis of multi-view face patterns," in *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS '01)*, pp. 125–132, IEEE Computer Society, Washington, DC, USA, 2001.

[15] C. Bouveyron, S. Girard, and C. Schmid, "High-dimensional discriminant analysis," *Communications in Statistics: Theory and Methods*, vol. 36, no. 13–16, pp. 2607–2623, 2007.

[16] M.-H. Yang, "Face recognition using extended isomap," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, pp. 117–120, 2002.

[17] J. Zhang, S. Z. Li, and J. Wang, "Nearest manifold approach for face recognition," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 223–228, Seoul, Korea, 2004.

[18] E. E. Abusham, D. Ngo, and A. Teoh, "Fusion of locally linear embedding and principal component analysis for face recognition (FLLEPCA)," in *Proceedings of the 3rd International Conference on Advances in Patten Recognition (ICAPR '05)*, vol. 3687 of *Lecture Notes in Computer Science*, pp. 326–333, 2005.

[19] Y. Chang, C. Hu, and M. Turk, "Probabilistic expression analysis on manifolds," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 520–527, 2004.

[20] A. Elgammal and C.-S. Lee, "Inferring 3D body pose from silhouettes using activity manifold learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 681–688, 2004.

[21] O. C. Jenkins and M. J. Matarić, "A spatio-temporal extension to isomap nonlinear dimension reduction," in *Proceedings of the 21st International Conference on Machine Learning (ICML '04)*, pp. 441–448, 2004.

[22] A. Elgammal and C.-S. Lee, "Separating style and content on a nonlinear manifold," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 478–485, 2004.

[23] A. Brun, H. J. Park, H. Knutsson, and C. F. Westin, "Colouring of DT-MRI fiber traces using Laplacian eigenmaps," in *Proceedings of the 9th International Conference on Computer Aided Systems Theory*, vol. 9, pp. 48–51, 2003.

[24] M. Niskanen and O. Silvén, "Comparison of dimensionality reduction methods for wood surface inspection," in *Proceedings of the 6th International Conference on Quality Control by Artificial Vision*, pp. 178–188, 2003.

[25] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pp. 53–58, Washington, DC, USA, 2002.

[26] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[27] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.