Class-label Locally Linear Embedding in Face Recognition

Pang Ying Han and Andrew Teoh Beng Jin, Fazly Salleh Abas, Ooi Shih Yin

***Abstract*—Locally Linear Embedding (LLE) is an unsupervised non-linear manifold learning method, which has spurred increased interest in face recognition research recently. However, it is commonly known that a supervised method that considering the class-specific information always outperforms the unsupervised one, especially in biometric recognition task. In this paper, we propose a supervised LLE technique, known as class-label Locally Linear Embedding (cLLE). cLLE aims to discover the nonlinearity of high-dimensional data by minimizing the global reconstruction error of the set of all local neighbours in the data set. cLLE method is using user class-specific information in neighbourhoods selection and thus preserves the local neighbourhoods. Since the locality preservation is correlated to the class discrimination, the proposed cLLE is expected superior to LLE in face recognition. Experimental results on three face databases demonstrate the success of the proposed technique.**

***Index Terms*—face recognition, Locally Linear Embedding, FisherSpace, class-specific information**

I. Introduction

The most popular technique in computer based face recognition is based on the subspace-based approach. These approaches are mostly originated from Turk and Pentland's work, known as Eigenfaces [1]. The key idea of Eigenfaces is the optimal linear transformation which aims to preserve the maximum variance of data vector. Soon after, Eigenfaces is extended by Belhumeur et al. by introducing Fisher's Linear Discriminant Analysis (FLD) to improve the discrimination capability, denoted as Fisherfaces [2]. Fisherfaces seeks a projection direction that maximizes the between-class scatter, but minimizes the within-class scatter. There are other subspace-based techniques, such as Bayesian algorithm using probabilistic subspace that extracting global interpretations [3], local feature analysis that extracting local topographic representation [4], independent component analysis that revealing statistically independent components [5], Non-negative factorization that obtaining a part-based linear facial representation [6].

Real face images are always with varying illumination, pose and expression. These intra-class variations create highly non-linear sub-manifold in the high-dimensional image space [7]. The linear dimensionality reduction methods are sub-optimal to face recognition as they are not able to reveal the underlying non-linear structure of the face data. Recently, a powerful non-linear manifold learning technique, based on neighbourhood preservation, has been proposed by Saul and Roweis, namely Locally Linear Embedding (LLE) [8][9]. The mapping to a single low-dimensional coordinate system is derived from the symmetries of locally linear reconstructions.

Hadid et al. assessed the performance of LLE on one candidate's real face images in different poses and compared the results with PCA and Self-Organizing Map (SOM) [10]. The LLE embedding recovered perfectly the different poses and the face distribution was well distributed according to the poses. But, the PCA embedding tended to be scattered and the representations of some face images with different poses were jumbled up. Zhang et al. and Wang et al. reported the application of LLE in multi-pose face synthesis [11][12]. These papers illustrated the LLE projection of a face image sequence when the head rotated from left to right. The projected points by LLE scattered along a smooth curve and in the arrangement associated with the head rotation trend. In [7][13], LLE was demonstrated to be superior to PCA in few face databases.

The main assumption behind LLE is that the data set is sampled from a (possibly non-linear) manifold, embedded in the high dimensional space. LLE embeds the data to a low dimensional space, in which the local geometry of the high-dimensional data is preserved, via neighbourhood preservation, in the embedded space. However, LLE is an unsupervised technique in which the class-specific information of data is lacked of. It is believed that the recognition ability will be improved if they are considered. Ridder et al. proposed a supervised version of LLE, namely Supervised LLE (SLLE) [14]. The idea is to find a projection that separates within-class structure from between-class structure, by "adding" the distances between samples belonging to different classes, while leaving them unchanged if they are from the same class. Inspired by Ridder's work, we propose a new variant of supervised-based LLE, namely class-label LLE (cLLE). We assume that the local neighbourhoods of each data sample should be composed of samples belonging to the same class only. Here, we are not only maximizing between-class structure, but also minimizing within-class structure. Therefore, neighbours of a sample will have higher possibility to be picked from the same class. Thus, it can obtain better local neighbourhood preservation than LLE, leading to a good discrimination.

Multimedia University, 75450 Melaka, Malaysia, yhpang@mmu.edu.my
Yonsei University, Seoul, South Korea, bjteoh@yonsei.ac.kr
Multimedia University, 75450 Melaka, Malaysia, fazly.salleh.abas@mmu.edu.my
Multimedia University, 75450 Melaka, Malaysia, syooi@mmu.edu.my

II. Locally Linear Embedding

Locally Linear Embedding (LLE) is a non-linear sub-manifold learning algorithm that embeds a high dimensional data into a lower dimensional space while preserving local topological structure via neighbourhood preservation [8][9]. Let $\{\tilde{\Gamma}_i \in \Re^d \mid i=1,\ldots, M\}$ represents the input data as d-dimensional points in an Euclidean vector space and the output embedding vectors are $\{\boldsymbol{Y}_i \in \Re^t \mid i=1,\ldots, M\}$, where $t \ll d$.

Step 1: Neighbourhood Computation

Calculate the Euclidean distances between all the pre-processed points, $\tilde{\Gamma}_i$, and then choose the k nearest neighbours for each data point.

Step 2: Linear Weights Reconstruction

Calculate a weight matrix, $\boldsymbol{W}$ which contains the best linear approximation of each data point $\tilde{\Gamma}_i$ from its neighbours. We minimize the following cost function:

$$\Phi(\boldsymbol{W}) = \sum_{i=1}^{M} \left| \tilde{\Gamma}_i - \sum_{j=1}^{M} \boldsymbol{W}_{ij} \tilde{\Gamma}_j \right|^2 \tag{1}$$

Note that each of the data point $\tilde{\Gamma}_i$ is reconstructed only from its neighbours, i.e. $W_{ij} = 0$ if $\tilde{\Gamma}_j$ is not a neighbour of $\tilde{\Gamma}_i$. This means that every weight vector W_i contains only at most k non-zero elements and the matrix $\boldsymbol{W}$ is sparse. Without loss of generality, the weights sum up to one for each point, this simplifies the optimization problem. Consider a specific $\tilde{\Gamma}$ with k nearest neighbours $\boldsymbol{\eta}_j$ and reconstruction weights W_j, the equation (1) can be rewritten in the following form:

$$\begin{aligned} \left| \tilde{\Gamma} - \sum_{j=1}^{k} W_j \boldsymbol{\eta}_j \right|^2 &= \left| \sum_{j=1}^{k} W_j \left(\tilde{\Gamma} - \boldsymbol{\eta}_j \right) \right|^2 \\ &= \sum_{j=1}^{k} \sum_{i=1}^{k} W_j W_i C_{ji} \end{aligned} \tag{2}$$

with local covariance matrix $C_{ji} = (\tilde{\Gamma} - \boldsymbol{\eta}_j).(\tilde{\Gamma} - \boldsymbol{\eta}_i)$. To minimize this error under the constraint that the rows of W sum to one, a Lagrange multiplier can be used [15].

$$L(W, \lambda) = \sum_{j=1}^{k} \sum_{i=1}^{k} W_j W_i C_{ji} + \lambda \left(\sum_{j=1}^{k} W_j - 1 \right) \tag{3}$$

where $\lambda \neq 0$. Saul and Roweis recommended solving the linear equation $\sum_j C_{ji} W_i = 1$ and rescaling the weights so that they sum to one, which yields the same result. However, the matrix $\boldsymbol{C}$ might be singular, especially when neighbours k more than the input dimensions d. It also may appear in the cases with $k \ll d$. To overcome this problem, a small multiple can be added to the $\boldsymbol{C}$ matrix:

$$\boldsymbol{C} := \boldsymbol{C} + r\boldsymbol{I} \tag{4}$$

r is a small regularization parameter that will have only a negligible effect on the results.

Step 3: Embedding Coordinates Computation

In this step, a low-dimensional vector $\boldsymbol{Y}_i$ is constructed. This can be done by minimizing the following cost function:

$$\Phi(\boldsymbol{Y}) = \sum_{i=1}^{M} \left| \boldsymbol{Y}_i - \sum_{j=1}^{M} W_{ij} \boldsymbol{Y}_j \right|^2 \tag{5}$$

The minimum value here is invariant under the rotations and translations of the image points. So, all coordinates are required to be centred on the origin, i.e. $\sum_{i=1}^{M} \boldsymbol{Y}_i = \vec{0}$ and are constrained to have unit covariance, i.e. $1/M \sum_{i=1}^{M} \boldsymbol{Y}_i \boldsymbol{Y}_i^T = \boldsymbol{I}$. We can rewrite the equation (5) to be:

$$\Phi(\boldsymbol{Y}) = \boldsymbol{Y}^T \boldsymbol{M} \boldsymbol{Y} \tag{6}$$

where $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$. Thus, equation (6) can be minimized by solving the eigen-decomposition of the matrix $\boldsymbol{M}$. The eigenvectors corresponding to the 2nd to $(M+1)^{st}$ smallest eigenvalues form the embedding $\boldsymbol{Y}$. The corresponding eigenvector with zero eigenvalue is discarded since it is representing a free translation mode.

Step 4: Mapping a Test Data Point onto the Embedding Vectors

k nearest neighbours of the test data $\tilde{\Gamma}_{test}$ are selected among the training data points. Then, the linear weights W_j that best reconstruct $\tilde{\Gamma}_{test}$ from its j^{th} neighbours are computed. These linear weights W_j are constrained to the sum-to-one, $\sum_j W_j = 1$. Finally, the embedded test data, Υ_{test}, is computed via $\Upsilon_{test} = \sum_j W_j \boldsymbol{Y}_j$, where the sum is over the outputs corresponding to the all neighbours of $\tilde{\Gamma}_{test}$ [13].

III. Class-Label Locally Linear Embedding

LLE is an unsupervised technique. The performance of LLE can be improved by utilizing class-specific information during neighbourhood selection in the training phase. The idea is to find a mapping that separates the between-class structures/ manifolds. For fully disjoint/ separated manifolds, the local neighbourhood of a sample from a class should be composed of samples belonging to the same class only. In class-label locally linear embedding, denoted as cLLE, the neighbourhood size, k, is redefined as the number of training images per class.

cLLE algorithm maximizes the between-class scatter, but minimizes the within-class scatter via the projection of the pre-processed points, $\tilde{\Gamma}_i$ onto FisherSpace (also known as LDA subspace). Then, based on the distances between the projected points $\breve{\Gamma}_i$, it selects minimum-distance k points as the neighbour set of each data point, $\tilde{\Gamma}_i$. With this, neighbours of a sample will have higher possibility to be picked from the same class. The FisherSpace is a set of vectors $\wp = [\wp_1, \wp_2, \ldots, \wp_Q]$ that can be obtained by satisfying [2]

$$\wp = \arg\max \left| \wp^T S_b \wp / \wp^T S_w \wp \right| \quad (7)$$

in which S_b is the between-class scatter matrix and S_w is the within-class scatter matrix. S_b and S_w are defined as:

$$S_b = \sum_{j=1}^{Q} \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T ;$$

$$S_w = \sum_{j=1}^{Q} \sum_{i=1}^{Q'_j} \left(\tilde{\Gamma}_i^{\,j} - \mu_j\right)\left(\tilde{\Gamma}_i^{\,j} - \mu_j\right)^T \quad (8)$$

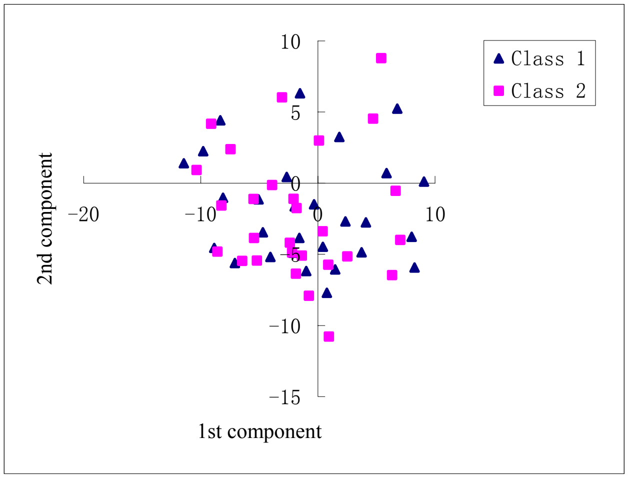
where $\tilde{\Gamma}_i^{\,j}$ is the i^{th} sample of class j, μ_j is the mean of class j, Q is the number of class, Q'_j is the number of samples in class j and μ is the mean of all the classes.

Then, weight matrix, $\boldsymbol{W}$, and embedding coordinates, $\boldsymbol{Y}$, are computed based on the local neighbourhood, using equation (1) and (1). During recognition stage, the test data $\tilde{\boldsymbol{\Gamma}}_{test}$ are projected onto the FisherSpace, $\wp$ constructed during the training phase to maximize the between-class scatter and minimize the within-class scatter. Then, *k* nearest neighbours of the test data $\tilde{\boldsymbol{\Gamma}}_{test}$ are selected among the training data points based on the distance between the projected data points, $\breve{\boldsymbol{\Gamma}}_{test}$. Next, the linear weights W_j that best reconstruct $\tilde{\boldsymbol{\Gamma}}_{test}$ from its j^{th} neighbours are computed. These linear weights W_j are constrained to the sum-to-one, $\sum_j W_j = 1$ .Finally, the embedded test data, Υ_{test}, is computed via $\Upsilon_{test} = \sum_j W_j \boldsymbol{Y}_j$.

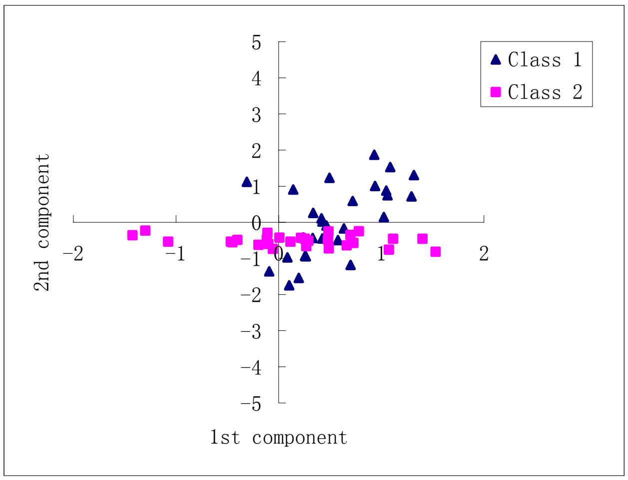
IV. CLLE ON FACE MANIFOLD

We illustrate the embedding robustness of cLLE and compare it with the embeddings of PCA and LLE on *Yale* database. From Fig. 1, the two sets of data point after PCA embedding tends to be scattered and are jumbled up; In the LLE embedding, the face distribution is more controlled. On the other hand, cLLE performs much better than LLE in face manifold discovery, in which the face distribution is well-controlled and the between-class structures are separated. Therefore, cLLE approach works well to discover the nonlinearity of the face data compared to PCA and LLE.

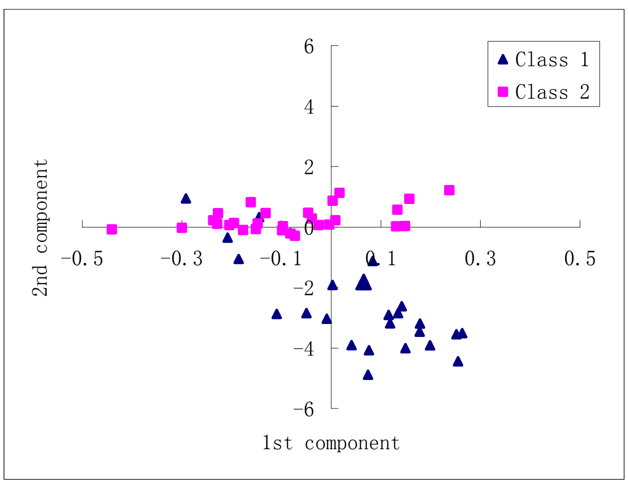
We also explore the effect of the number of training samples per class to the performance of cLLE embedding. Fig. 2 illustrates the cLLE embeddings using 2 (small number of training samples per class) and 12 (large number of training samples per class) training images per class on *Yale* database. Compare these two face distributions, we can observe that the cLLE embedding using 12 training images per class is better-distributed. In addition, there is a clear separation between the class structures. The reason is by using larger training samples per class cLLE is able to learn how to best discriminate between faces of different classes and obtain more accurate class label information for neighbourhood selection, leading to a better local neighbourhood preserving power. To support our claim, we study this effect on real face data using the images of three different face databases in section 5.



PCA embedding

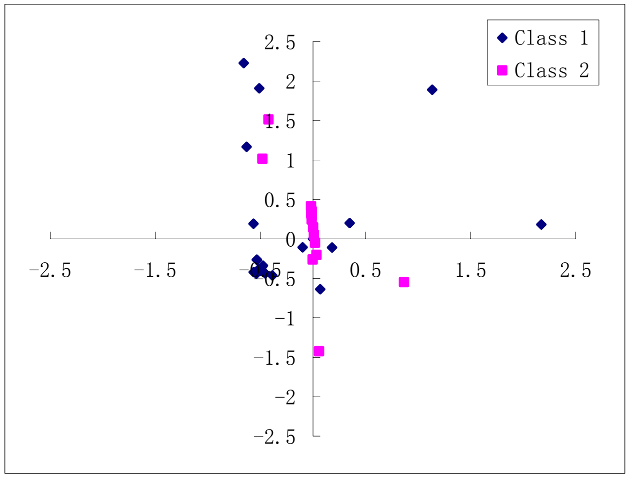


LLE embedding

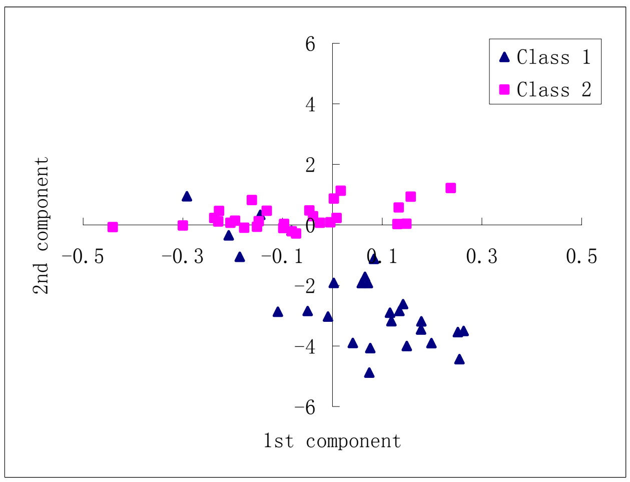


cLLE embedding

Fig. 1. PCA, LLE and cLLE embeddings on face data of the *Yale* databases based on two-class problem



using 2 training images per class



using 12 training images per class

Fig. 2. cLLE embeddings using 2 and 12 training images per class on the *Yale* database.

V. Experimental results and Discussions

In this section, we investigate the performance of our proposed class-label Locally Linear Embedding (cLLE) method in face verification. The system performance is compared with PCA (as baseline, also known as Eigenfaces), FLD (or Fisherfaces) and LLE. Three databases: Olivetti Research Laboratory Database (*ORL*), Aleix Martinez and Robert Benavente Database (*AR*), and Yale Face Database (*Yale*) are used for testing. Each database is partitioned into training and testing sets. Both training and testing sets contain all persons, but there is no overlap in the sample image between the training and testing sets. We conduct experiment with 2-fold cross-validation strategy. In the first-fold test, the first half samples per person are used for training and the remaining samples are used for testing. In the second-fold test, the training set is formed by the last half samples per person and the first half samples are used for testing.

A. Results on the ORL database

The *ORL* database contains 400 images of 40 persons (10 images per person). Some images have different variations including facial expression (open or close eyes, smiling or non-smiling) and facial details (glasses or without glasses). Therefore, there are 5x40=200 images in training set and 5x40=200 images in testing set at each experiment run. The recognition error is the average error rate (AER) obtained in both tests. We show the best results and the optimal dimensionality obtained (in bracket) obtained by PCA (baseline), FLD, LLE and cLLE methods in Table I. For FLD, we project all the samples onto a subspace spanned by the Q-1 largest eigenvectors, where Q is the number of class. We can see that our algorithm performs the best. Fig. 3 shows the recognition error in term of Average Error Rate (AER) on the testing set against the number of features in Euclidean distance metric. Besides, we also explore the effect of the number of training image of the algorithms to the system verification performance. Table II shows the best results and the optimal dimensionality (in bracket) obtained by PCA (baseline), FLD, LLE and cLLE methods for 2 and 8 training images per person settings. It is worthwhile to note that in the cases when only two training images per person are available, our method and FLD performs poorly compared to PCA and LLE. The result is consistent with the observation in [16] that LDA method performs poorer when the training set is small. However, when using 8 training images per class, cLLE method obtains the best result.

TABLE I
PERFORMANCE COMPARISONS IN TERM OF AER ON THE *ORL* DATABASE

Method	AER (%)
PCA (baseline)	10.8423 (50)
LDA (FLD)	10.1128 (39)
LLE	11.9500 (50)
cLLE	**4.3096** (50)

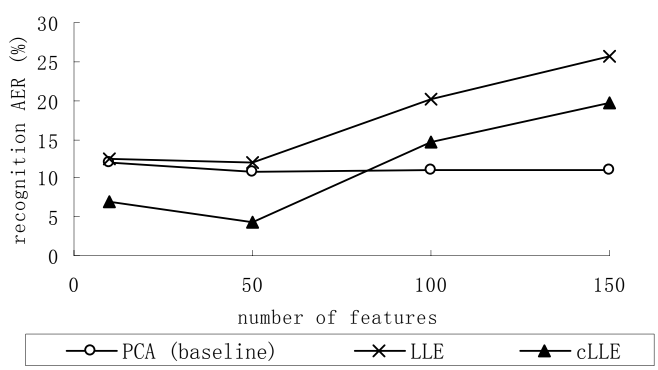


Fig. 3. Recognition error rate against the number of features on *ORL* database.

TABLE II
PERFORMANCE COMPARISONS FOR 2 AND 8 TRAINING IMAGES PER PERSON ON THE *ORL* DATABASE

Method	2 train	8 train
PCA	**10.8393** (80)	10.7612 (150)
LDA (FLD)	38.2512 (39)	8.9863 (39)
LLE	15.9816 (10)	9.3489 (50)
cLLE	39.9880 (50)	**5.6971** (50)

B. Results on the AR database

In our experiment, 102 subjects with 14 non-occluded images per class are selected from the *AR* database. Table III shows the best results and the optimal dimensionality obtained (in bracket) obtained by PCA (baseline), FLD, LLE and cLLE methods. Our algorithm and FLD perform the best with about 12% error rate. Fig. 4 shows the recognition error rate on the testing set against the number of features in Euclidean distance metric. Table IV shows the best results and the optimal dimensionality (in bracket) obtained by PCA (baseline), LLE and cLLE methods for 2 and 7 training images per person. cLLE and FLD methods obtains the poor performance when the number of samples per class is too small for training purpose. However, when using 7 training images per class, the both methods obtain the best result.

TABLE III
PERFORMANCE COMPARISONS IN TERM OF AVERAGE ERROR RATE (AER) ON THE *AR* DATABASE

Method	AER (%)
PCA	30.0194 (150)
LDA (FLD)	**12.3914** (101)
LLE	26.2318 (100)
cLLE	**12.6026** (50)

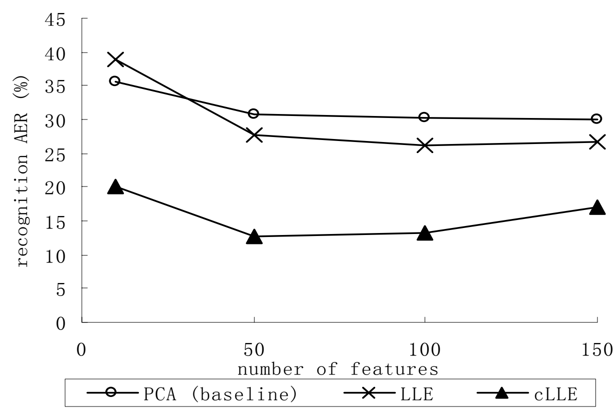


Fig. 4. Recognition error rate against the number of features on the *AR* database.

TABLE IV
PERFORMANCE COMPARISONS FOR 2 AND 7 TRAINING IMAGES PER PERSON ON THE *AR* DATABASE

Method	2 train	7 train
PCA	**17.0913** (150)	30.0194 (150)
LDA (FLD)	39.0739 (101)	**12.3914** (101)
LLE	19.2872 (10)	26.2318 (100)
cLLE	31.1205 (10)	**12.6026** (50)

C. Results on the Yale database

Yale images are induced with significant "artificial" illumination variations, using spotlight to create the illumination changes. We randomly selected a subset of *Yale* containing 24 images from 38 persons. These images possess significant illumination variations with the setting of the light source direction with respect to the camera axis at ± 35 degrees azimuth and ± 40 degrees elevation. Note that a positive azimuth implies that the light source was to the right of the subject while negative means it was to the left. Positive elevation implies above the horizon, while negative implies below the horizon. There are 12x38=456 images in training set and 12x38=456 images in testing set at each experiment run. The best results and the optimal dimensionality obtained (in bracket) obtained by PCA (baseline), FLD, LLE and cLLE methods are displayed in Table V. Our algorithm obtains the lowest error rate of 6.9%. Fig. 5 shows the recognition true error in term of Average Error Rate (AER) on the testing set against the number of features in Euclidean distance metric. Table VI shows the best results and the optimal dimensionality (in bracket) obtained by PCA (baseline), FLD, LLE and cLLE approaches for 2, 10 and 15 training images per person. Again, FLD and cLLE methods obtain poor performance when the number of samples per class is too small for training purpose. However, when using 10 training images per class, the performances of LLE and cLLE methods are comparative. When more images per class are added in the training set, let say 15 training images per class, cLLE method obtains the best result.

TABLE V
PERFORMANCE COMPARISONS IN TERM OF AVERAGE ERROR RATE (AER) ON THE *YALE* DATABASE

Method	AER (%)
PCA (baseline)	23.0722 (50)
LDA (FLD)	7.7397 (37)
LLE	16.4172 (50)
cLLE	**6.90496** (50)

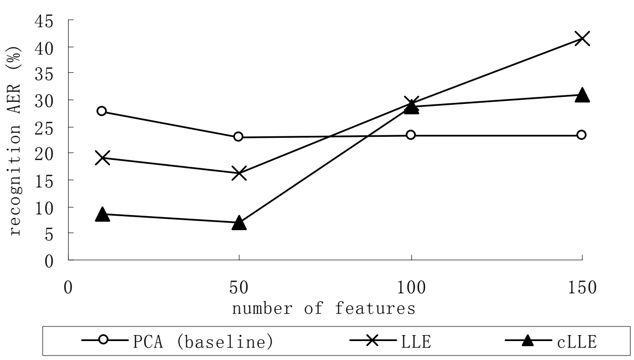


Fig. 5. Recognition error rate against the number of features on the *Yale* database.

TABLE VI
PERFORMANCE COMPARISONS FOR 2, 10 AND 15 TRAINING IMAGES PER PERSON ON THE *YALE* DATABASE

Method	2 train	10 train	15 train
PCA	**15.1760** (50)	19.3812 (50)	34.2211 (50)
LDA (FLD)	37.9979 (37)	12.2891 (37)	14.7581 (37)
LLE	**15.8549** (10)	**9.1218** (50)	10.6920 (50)
cLLE	26.0313 (10)	**9.5727** (50)	**8.7550** (50)

D. Discussions

We summarize the experimental results as below:

1. With sufficient training dataset, our proposed cLLE approach consistently outperforms PCA method in the all experiments on three different face databases: *ORL*, *AR*, and *Yale* databases. The reason is cLLE approach is able to perfectly disclosure the underlying nonlinear structure of the face manifold, in which PCA method is not able to do so.

2. With sufficient training dataset, cLLE method consistently outperforms LLE method on the three different databases. This demonstrates that in neighbourhood selection, if the distance between each data point and other data points are computed and ranked by considering the class-specific information, a better local neighbourhood preservation can be obtained. Since the locality preservation is correlated to the class discrimination, the proposed cLLE approach is expected to be superior to LLE in face recognition. This has been illustrated through the observation to the performance comparison between cLLE and LLE methods in face verification.

3. cLLE method significantly obtains superior recognition performance to PCA and LLE approaches for small number of features. This demonstrates that the proposed cLLE method is able to extract more discriminative features than the others.

4. However, the proposed cLLE method suffers from recognition performance degradation when small number of training samples per class is used. The reason is that using too small number of training images per class is not sufficient for cLLE method to learn the underlying class structure and obtain accurate class-specific information for neighbourhood selection. The result is consistent with the observation in [16] that LDA method performs poorer when the training set is small. Nevertheless, when the training data set is large, the performance of cLLE approach is guaranteed.

VI. CONCLUSIONS

We have proposed a new algorithm for face representation and verification. This algorithm is a supervised version of Locally Linear Embedding (LLE). LLE has been used as an unsupervised technique that not considering the class-specific information. It is believed that the recognition ability will be improved, especially in biometric recognition applications, if class-specific information is considered. Thus, our proposed method, called as class-label LLE (cLLE), uses class-specific information during training for neighbourhood selection. This proposed technique holds an assumption that the local neighbourhood of each sample should be composed of image samples belonging to the same class only. Therefore, there is a higher possibility to select the neighbours of each data point from the class that itself belongs to. Consequently, the proposed technique can have better local neighbourhood preservation, leading to higher class discrimination. It is expected that the proposed cLLE approach is able to achieve good performance in face verification. We have evaluated the effectiveness of cLLE method on three different face databases, namely *ORL*, *AR*, and *Yale* databases. The experimental results show that the proposed technique is able to produce an encouraging recognition performance compared to PCA and LLE approaches. The proposed technique reduced the errors with about 5% on *ORL*, 14% on *AR*, and 10% on *Yale* databases compared to the original feature extractor – LLE. However, our proposed technique also suffers from performance degradation when smaller training images per class are used. With smaller training set, cLLE method has insufficient information to learn the underlying class structure and thus it is unable to obtain the accurate class-specific information for neighbourhood selection. Therefore, our future work is to deal with the issue of how to reduce the number of training samples needed for good recognition result.

REFERENCES

[1] M. Turk, A. Pentland, "Eigenfaces for recognition,: *J. Cognitive Neuroscience*, 3(1), pp. 71-86, 1991.

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, pp. 711-720, 1997.

[3] B. Moghaddam, W. Wahid, A. Pentland, "Beyond eigenfaces: probabilistic matching for face recognition," in *Proceeding of Third Intl. Conf. Automatic Face and Gesture Recognition*, pp. 30-35, 1997.

[4] P. Penev, J. Atick, "Local feature analysis: A general statistical theory for object representation," *Neural Systems*, 7(3), pp. 477-500, 1996.

[5] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, "Face recognition by Independent Component Analysis," *IEEE Transaction on Neural Networks*, 13, pp. 1450-1464, 2002.

[6] D.D Lee, H.S. Seung, "Learning the parts of objects by Non-Negative Matrix Factorization," *Nature*, 401, pp. 788-791, 1999.

[7] J.P. Zhang, H.X. Shen, Z.H. Zhou, "Unified Locally Linear Embedding and Linear Discriminant Analysis Algorithm (ULLELDA) for face recognition," in *Proceeding of 5th Chinese Conference on Biometric Recognition*, pp. 296-304, 2004.

[8] S. Lawrence, S. Roweis, "An introduction to Locally Linear Embedding," URL: <http://www.cs.toronto.edu/~roweis/lle/publications.html>.

[9] S. Roweis, S. Lawrence, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290(5500), pp. 2323-2326, 2000.

[10] A. Hadid, O. Kouropteva, M. Pietikainen, "Unsupervised learning using Locally Linear Embedding: experiments with face pose analysis," in *Proceeding of the 16th International Conference on Pattern Recognition*, pp. 111-114, 2002.

[11] J. Wang, C.S. Zhang, Z.B. Kou, "An analytical mapping for LLE and its application in multi-pose face synthesis," in *Proceeding of 14th British Machine Vision Conference*, 2003.

[12] C.S. Zhang, J. Wang, N.Y. Zhao, D. Zhang, "Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction," *Pattern Recognition Journal*, vol. 37, pp. 325-336, 2004.

[13] D. Liang, J. Yang, Z.L. Zheng, Y.C. Chang, « A facial expression recognition system based on supervised locally linear embedding," *Pattern Recognition Letters*, vol. 26, pp. 2374-2389, 2005.

[14] D.D. Ridder, O. Kouropteva, O. Okun, M. Pietikainen, R.P.W. Duin, "Supervised Locally Linear Embedding," in *Proceeding of Joint Int. Conf. ICANN/ICONIP*, pp. 333-341, 2003.

[15] A. Mizrahi, M. Sullivan, "Calculus and Analytic Geometry," Wadsworth Publishing Company, 3rd Edition, 1990.

[16] M.M. Aleix, C.K. Avinash, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23(2), pp. 228-233, 2001.