Multi-view Hand Shape Recognition Using Statistical LLE

Yi-Jay Gu, Pi-Fuei Hsieh, Ming-Hua Yang, and Chung-Hsien Wu
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan 701, Taiwan, R.O.C.
{p7693133, pfhsieh, p7694128}@mail.ncku.edu.tw, and chwu@csie.ncku.edu.tw

ABSTRACT

The image-based object recognition problem becomes complicated when the objects of interest are not posed at a fixed view. This study attempts to recognize the multi-view hand shapes in Taiwanese sign language (TSL) based on a novel statistical locally linear embedding (LLE). The original LLE is an unsupervised nonlinear dimensionality reduction approach that utilizes the local linearity to discover the low dimensional manifold embedded in the high dimensional space. This suggests that LLE may preserve neighborhood configuration in the nonlinear structure of the multi-view hand shape data distribution. For better classification performance, this study proposes a statistical LLE that incorporates the class label information statistically to improve the multiclass classification capability posterior to dimensionality reduction. Experimental results show that the statistical LLE gave a classification performance superior to the original LLE and the linear dimensionality reduction methods such as LDA and PCA in multi-view TSL hand shape recognition problem.

1: INTRODUCTION

While speaking is the most efficient way to communicate with others for people with normal hearing, for the hearing-impaired people, the sign language takes the place of speaking. Sign language is so complicated for normal hearing people that normal hearing people cannot master a sign language without training for a period of time. This embarrassment causes the inconvenience of communication with the hearing impaired people. Thus a sign language recognition system is developed to ease the inconvenience.

Generally, a sign gesture in a sign language can be decomposed to phonemes, which are the smallest contrastive units in the language. Typical phonemes include hand shapes, palm orientations, motion trajectories and facial expressions. In parallel to this analysis, sign language recognition can be realized by identifying phonemes individually and then integrating the results into a conclusion for a sign gesture [1]. Therefore the recognition performance of an individual phoneme affects the overall performance of such sign language recognition. In this paper, we focus on multi-view hand shape recognition, which is associated with hand shape and palm orientation phonemes, when developing a vision-based sign language recognition system.

In this study, an image of size $M \times N$ is regarded as a point located in an MN-dimensional space. In this case, the dimensionality of the data to be processed is relatively large compared with the number of training samples. In order to avoid the curse of the dimensionality [2], the dimensionality reduction process is necessary in high dimensional data classification. The principal component analysis (PCA) and the linear discriminant analysis (LDA) [3][4] are two well-known linear dimensionality reduction methods. PCA and LDA are of limited used when the data distribution is complicated and the scatter matrices cannot sufficiently describe the class separability or when the linear projection inferred by PCA or LDA is not appropriate.

Recently, the locally linear embedding (LLE) [5] algorithm has been proposed to tackle the nonlinear dimensionality reduction problem. LLE is an unsupervised learning algorithm that attempts to discover the embedded manifold by preserving the neighborhood of a data point. The underlying idea of LLE is based on the assumption that data lying on a nonlinear manifold can be viewed as distributing linearly in a local patch if the data are well sampled and lying on a smooth manifold. Although LLE has capability to recover the global nonlinear structure from locally linear fits, the class label information is not taken into account when mapping samples from the high dimensional space to a low dimensional feature space. A supervised LLE (SLLE) algorithm [6][7] has been proposed to utilize the class label information by scaling the Euclidean distance between samples according to their class labels. Although SLLE improves the performance of LLE related to classification, the information provided by the Euclidean distance between samples is not sufficient enough to select proper neighbors for classification. In this paper, we propose a statistical LLE algorithm to incorporate the class label information by estimating the similarity between samples statistically. In experiments, both linear and nonlinear dimensionality reduction approaches were employed to map the segmented hand shape images into a low dimensional space, and the K-NN classifier was used to evaluate the performance.

This paper is organized as follows. Section 2 discusses the hand shape segmentation procedure. Section 3 introduces the original LLE algorithm. Section 4 presents the proposed statistical LLE algorithm. Section 5 shows the classification results in the experiments of multi-view TSL hand shape recognition. Finally, conclusions are inferred and several issues for the future work are mentioned in Section 6.

2: HAND SHAPE SEGMENTATION

Before classifying the hand shape images, the hand shape segmentation procedure was carried out to subtract the background regions in the image. The hand shape segmentation procedure is based on the skin color detection techniques. Two issues that are concerned most in the study of skin color detection are color representations and skin color model. In this paper, we introduce the Gaussian mixture model to model the skin color distribution in RGB color space and compute the probability density to distinguish pixels into hand shape region and the background [8].

Suppose $\mathbf{x}$ is a color vector. We used a Gaussian mixture model to describe the skin color distribution in the color space. The probability density function was computed as

$$\begin{aligned} p(\mathbf{x} \mid \omega_s) &= \sum_{i=1}^{k} w_i \cdot p_i(\mathbf{x} \mid \omega_s) \\ &= \sum_{i=1}^{k} w_i \cdot \frac{1}{(2\pi)^{d/2} \left|\mathbf{C}_i\right|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathrm{T}} \mathbf{C}_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)\}, \end{aligned} \quad (1)$$

where the skin color is denoted by ω_s. $w_i \ldots w_k$ are the proportions of k Gaussian components with mean vectors $\boldsymbol{\mu}_i$ and covariance matrices $\mathbf{C}_i$ in d-dimension and $\sum_{i=1}^{k} w_i = 1,\ w_i > 0$.

In this study, the expectation maximization (EM) algorithm was used to estimate the parameters of a Gaussian mixture model to model the skin color distribution. Utilizing the estimated skin color distribution, a color pixel $\mathbf{x}$ was decided as a skin pixel if

$$p(\mathbf{x} \mid \omega_s) = \sum_{i=1}^{k} w_i \cdot p_i(\mathbf{x} \mid \omega_s) > \tau, \quad (2)$$

where the threshold τ was determined empirically.

3: LOCALLY LINEAR EMBEDDING

3.1: LLE

Essentially, LLE attempts to obtain a low dimensional space in which each point remains their neighboring relationship as in high dimensional space. In other words, the low dimensional space is required to preserve the neighborhood configuration. The LLE algorithm can be generalized to three steps: select neighbors, reconstruct with linear weights and map to embedded coordinates.

The neighborhood of each sample provides prior knowledge for LLE and affects the reconstruction result. The nearest neighbors of each sample can be identified by selecting a fixed number K of nearest neighbors in the Euclidean distance. Another approach is to choose samples within a fixed radius r as neighbors. The process of neighborhood selection can be flexible and various.

The second step of LLE is to reconstruct each sample by the linear combination of its neighbors. Consider the i-th sample $\mathbf{x}_i$ with K neighbors $\mathbf{x}_j$, $i \neq j$. Let w_{ij} be the reconstruction weights, $w_{ij} > 0$ and $\sum_j w_{ij} = 1$. The optimal reconstruction can be derived from minimizing the reconstruction error by properly selecting the reconstruction weights. The reconstruction error is thus formulated as

$$\varepsilon = \sum_i \left\| \mathbf{x}_i - \sum_j w_{ij} \mathbf{x}_j \right\|. \quad (3)$$

The final step of LLE algorithm is to compute the coordinates of the original high dimensional data $\mathbf{x}_i$ in the low dimensional space. The low dimensional embedding is obtained based on the idea that LLE preserves the local linearity from neighbors and the corresponding reconstruction weights. Let $\mathbf{Y}$ denote the collection of $\mathbf{y}_i$. The low dimensional coordinates $\mathbf{y}_i$ in the d-dimensional space can be computed by minimizing the cost function

$$\Phi(\mathbf{Y}) = \sum_i \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_j \right\|, \quad (4)$$

where w_{ij} are the same linear combination weights in the high dimensional space. The steps of LLE algorithm are illustrated in Fig. 1.

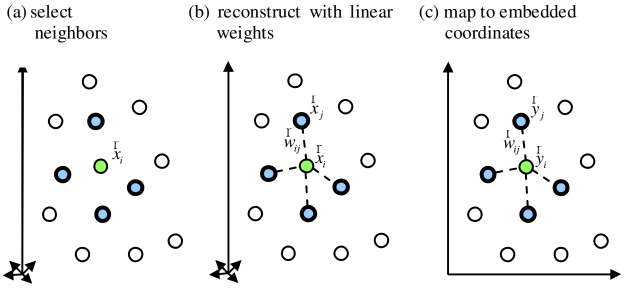
(a) select neighbors
(b) reconstruct with linear weights
(c) map to embedded coordinates
x_i
x_j
w_{ij}
x_i
y_j
w_{ij}
y_i

Fig. 1: Illustration of LLE algorithm.

3.2: Generalization

LLE provides a manifold learning approach to reducing the dimensionality of the data in a nonlinear way. However, LLE lacks a way to project the new arrival testing data to the low dimensional embedding like the linear dimensionality reduction approaches [9]. For this reason, a generalization procedure is needed to map new arrival data on the manifold without repeating the expensive eigenvalue computation when solving the cost function in (4).

In order to map a new data on the low dimensional space, the locally linear property and reconstruction weights are used. For a new arrival data $\mathbf{x}_{N+1}$, the nearest K neighbors are selected first. The second step is to compute the linear weights w_{N+1} that optimize the reconstruction $\mathbf{x}_{N+1}$ from its neighbors with the sum-to-one constraint, $\sum_j w_{N+1j} = 1$. Next, the new coordinate $\mathbf{y}_{N+1}$ in the low dimensional space can be determined as $\mathbf{y}_{N+1} = \sum_j w_{N+1j} \mathbf{y}_j$, which is the linear combination of the same neighbors in the high dimensional space with the optimal reconstruction weights.

4: STATISTICAL LLE

The data of each class in the high dimensional space are not necessarily normally distributed. In order to delineate the distribution more accurately, each class is modeled by a set of Gaussian clusters. In other words, several Gaussian subclasses are defined to describe a hand shape class. Various clustering algorithms have been developed for a long time, such as *k*-means clustering, hierarchical clustering, and fuzzy clustering. In this paper, the *k*-means algorithm was employed for its efficiency and easy implementation.

Assume that there are *K* clusters in each of *L* classes. After clustering, all the training samples are labeled into one of the *KL* clusters. Given a *D* dimensional sample **x**, the probability density of the cluster ω_{ik} is determined as

$$p(\mathbf{x}\,|\,\omega_{ik}) = \frac{1}{(2\pi)^{D/2}\left|\mathbf{S}'_{ik}\right|^{1/2}}\exp\{-\frac{1}{2}d^2(x)\}, \tag{5}$$

$$d^2(x) = (\mathbf{x}-\mathbf{m}_{ik})^T(\mathbf{S}_{ik}')^{-1}(\mathbf{x}-\mathbf{m}_{ik}), \tag{6}$$

where $\mathbf{S}'_{ik} = \mathbf{S}_{ik} + r\cdot\mathbf{I}$ is the sample covariance matrix of the *k*-th cluster in class ω_i plus a generalization term to prevent singularity and make the estimation more robust. In practice, the value of *r* is determined empirically.

Since the subclasses are mutually exclusive and statistically exhaustive, the likelihood of each class for the sample **x** can be determined by the sum of the subclass likelihood as

$$p(\mathbf{x}\,|\,\omega_i) = \sum_{k=1}^{K} p(\mathbf{x}\,|\,\omega_{ik})\,. \tag{7}$$

The objective of statistical LLE is to derive a relationship between samples and each class based on the statistics. Without losing the generality, assume that samples are independent. Under this assumption, the likelihood of any pairs of sample **x** and sample $\mathbf{x}'$ belonging to the same class ω_i is equal to the product of their class memberships associated with ω_i. Therefore, the relationship between any pairs of samples for a class ω_i can be evaluated by the minus logarithm of the maximum likelihood that the two samples belong to the same class. A relationship *d* is defined by

$$d(\mathbf{x},\mathbf{x}') = \max_i\left\{-\log p(\mathbf{x}\,|\,\omega_i)p(\mathbf{x}'\,|\,\omega_i)\right\}. \tag{8}$$

The measurement *d* incorporates the class information provided by the statistical cluster model for each class. In statistical LLE, *d* takes the place of original Euclidean distance. Later, the *K* nearest neighbors of each samples are determined according to *d*. The subsequent steps follow the same procedure of LLE.

5: EXPERIMENTS

5.1: Dataset

A total of 25 frequently used TSL hand shapes, as shown in Fig. 2, were selected for the experiments. The hand shape images of size 240×352 pixels were captured in the lab under a controlled light source and a stationary background. For each hand shape, 2 rotations, yaw and pitch, according to different orientations were performed to generate hand shape images with 20 different views in each orientation.

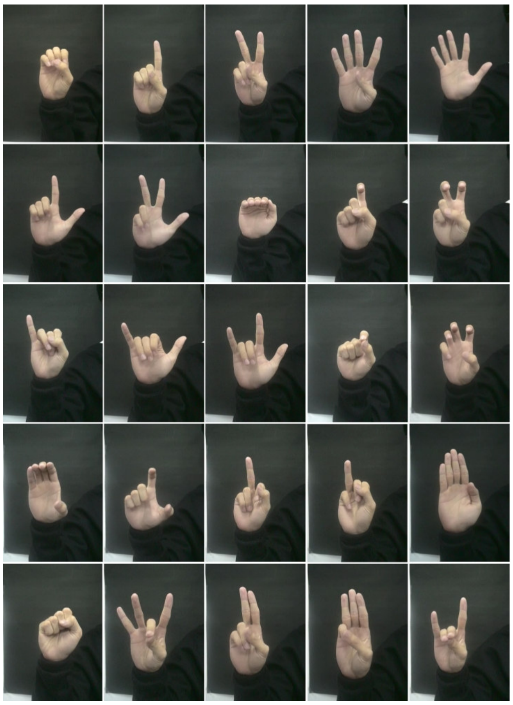

Fig. 2: TSL hand shapes used in experiments.

5.2: Multi-view TSL Hand Shape Recognition: Subject-dependent Experiments

In this experiment, each TSL hand shape was regarded as a class. The TSL hand shape segmented images were divided into training and testing dataset. For each hand shape, 1120 hand shape images with multi-view were randomly selected as training dataset and the other 840 images were used for testing.

The traditional linear approaches, such as PCA and LDA, were employed to reduce the dimensionality of the multi-view TSL hand shape data. In order to estimate the hand shape data distribution more precisely, *K*-means clustering were performed to define six subclasses from each original class before applying LDA. In nonlinear approaches, LLE, SLLE and the proposed statistical LLE were applied respectively with 25 neighbors for reconstruction weights in the neighborhood selection step. The parameter that controls the amount of the class label information to be incorporated into SLLE was set to 2.0 in this experiment. The proposed statistical LLE was applied to the data set in which each class had been divided into six clusters. In the classification stage, *K*-NN classifier (*K* = 5) was used to classify the reduced low dimensional data. The classification results are shown in Fig. 3 for comparison.

In Fig. 3, the original LLE algorithm gave the lowest recognition rats, which matches the results in previous literatures [5]. This indicates that LLE lacks the capability to handle the multiclass classification problem. PCA and LDA had comparative performance. SLLE improved the capability of LLE in classification due to the use of class information. The proposed statistical LLE achieved the best recognition rate in this experiment. As mentioned in Sec. 3.1, the neighborhood selection step provides an opportunity to incorporate the prior knowledge into analysis. Compared to SLLE that scales the Euclidean distance based on the class labels alone, the statistical LLE has integrated more information about the high-dimensional distribution. When there were a sufficiently large number of training samples to obtain reliable statistics, the statistical LLE gave a promising performance.

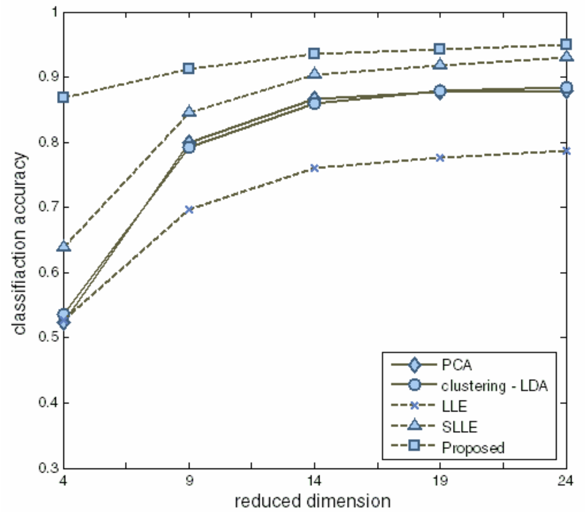

Fig. 3: Classification results of the subject-dependent tests.

5.3: Subject-independent Experiments

Subject-independent experiments were carried out to test the classification performance in the case in which a new subject had never given any training samples. A set of TSL hand shape images that belong to a new subject is all used as testing samples. The strategy of leave-one-subject-out tests was performed to recognize the hand shape images of each subject. The classification performance for each approach was evaluated by averaging the recognition accuracy of seven subjects. The classification results are showed in Fig. 4.

As expected, the subject-independent experiments produced lower classification accuracy than the subject-dependent experiment because hand shapes posed by different people (subjects) often have a larger variance than same people. In the subject-dependent experiment, hand shape images that belong to the same subject were possibly included in training sample set. Therefore these hand shapes of the same subject made a large proportion of nearest neighbors when the Euclidean distance was used to determine the nearest neighbors in LLE and SLLE. Note that the classification result of SLLE deteriorated more seriously than others. That is primarily because the variance due to different subjects made the Euclidean distance no longer reliable as in the subject-dependent experiment when the samples belonging to the same subject were not included.

Relying on the statistics estimated from training samples, the proposed statistical LLE integrated more training samples than SLLE to determine the nearest neighbors. Statistics based on a set of training samples provides more reliable information than the distance based on one sample.

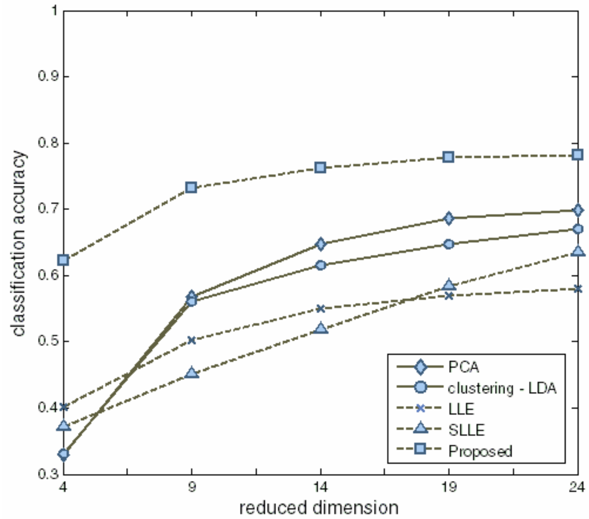

Fig. 4: Classification results of the subject-independent tests.

6: CONCLUSIONS

In this paper we present a procedure for recognizing the multi-view TSL hand shapes. The procedure is composed of two stages: feature extraction and classification. In feature extraction stage, a Gaussian mixture model is used to describe the skin color distribution and to obtain the segmented hand shape images. Then dimensionality reduction approaches are applied to the segmented hand shape images. In classification stage, the *K*-nearest neighbor classifier is used to classify the reduced dimensional hand shape data.

In dimensionality reduction, we propose a statistical LLE to enhance the capability of the original LLE algorithm related with multiclass classification. Unlike the original LLE, which is an unsupervised approach, the proposed statistical LLE is a supervised approach that utilizes the class label information in a statistical way. In the experiments of the multi-view TSL hand shape classification, the statistical LLE yielded better recognition performance than linear approaches, such as PCA and LDA, and nonlinear approaches, such as LLE and SLLE.

For more general and flexible usage of statistical LLE, it would be interesting to investigate how to automatically adjust the parameters, including the number of clusters for delineating a class and the number of neighbors for defining the neighborhood of a sample. The parameters generally vary with different data distributions.

ACKNOWLEDGMENTS

The authors would like to thank the participators in hand shape recording for their devotion to setting up the TSL multi-view hand shape database. Also, the authors are grateful to the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC94-2614-E-006-087.

REFERENCES

[1] S. C. W. Ong and S. Ranganath, "Automatic sign language analysis: a survey and the future beyond lexical meaning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873-891, June 2005.

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* second ed., New York: Academic Press, 1990.

[3] A. K. Jain, R. P.W. Duin and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, Jan. 2000.

[4] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification,* second ed., New York: John Wiley and Sons, Inc., 2001.

[5] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding." *Science*, vol. 290, pp. 2323-2326, Dec. 2000.

[6] D. Ridder, O. Kouropteva, O. Okun, M. Pietikäinen and R. P. W. Duin, "Supervised locally linear embedding." *Proc. of Joint Int'l Conf. on ICANN/ICONIP*, 2003.

[7] D. Lian, J. Yang, Z. Zheng and Y. Chang, "A facial expression recognition system based on supervised locally linear embedding," *Pattern Recognition Letters*, vol. 26, pp. 2373-2389, 2005.

[8] M. H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its applications in image and video databases," *Conf. on Storage and Retrieval for Image and Video Database*, vol. 3656, 1999.

[9] O. Kouropteva, O. Okun and M. Pietikäinen, "Incremental locally linear embedding," *Pattern Recognition*, vol. 38, pp. 1764-1767, 2005.