A Two-Stage Dimensional Reduction Approach to Low-Dimensional Representation of Facial Images

Jongmoo Choi and Juneho Yi

School of Information & Communication Engineering
Sungkyunkwan University
300 Chunchun-Dong, Jangan-Gu Suwon 440-746, Korea
{jmchoi,jhyi}@ece.skku.ac.kr

Abstract. We present a two-stage dimensional reduction approach to low-dimensional representation. When facial feature data need to be stored in low capacity storing devices, low-dimensional representation of facial images is very important. Our approach is composed of two consecutive mappings of the input data. The first mapping is concerned with best separation of the input data into classes and the second focuses on the mapping that the distance relationship between data points before and after the map is kept as closely as possible. We claim that if data is well-clustered into classes, features extracted from a topology-preserving map of the data are appropriate for recognition when low-dimensional features are to be used. We have presented two novel methods: FLD (Fisher's Linear Discriminant) combined with SOFM (Self-Organizing Feature Map) method and FLD combined with MDS (Multi-Dimensional Scaling) method. Experimental results using Yale, AT&T and FERET facial image databases show that the recognition performance of our methods degrades gracefully when low-dimensional features are used.

1 Introduction

The problem on extremely low-dimensional image representation for face recognition has little been investigated while many researchers study on face recognition robust to illumination, posture and facial expression changes. In practical biometric user authentication systems, low-dimensional feature extraction is one of the most important problem. When facial feature data need to be stored in low capacity storing devices such as bar codes and smart cards, extremely low-dimensional image representation of facial data is very important. It can also be used for data transmission in the internet or mobile environments. Moreover, it is applicable to real-time identification in the case of a large database.

The algorithms like PCA (Principal Components Analysis) [1], FLD (Fisher's Linear Discriminant) [2] and ICA (Independent Components Analysis) [3] can be used for reduction of the dimension of the input data but are not appropriate for low-dimensional representation of high dimensional data because their recognition performance degrade significantly. Although SOFM (Self-Organizing

D. Zhang and A.K. Jain (Eds.): ICBA 2004, LNCS 3072, pp. 131–138, 2004.

Feature Map) [4], PP (Projection Pursuit) [5] and MDS (Multi-Dimensional Scaling) [6] can be employed for low-dimensional data representation, these techniques are suitable for data representation in low-dimensions, usually two or three dimensions. They try to represent the data points in a such way that the distances between points in low-dimensional space correspond to the dissimilarities between points in the original high dimensional space. However, these techniques do not yield high recognition rates mainly because they do not consider class specific information. Our idea is that these methods incorporated with class specific information can provide high recognition rates.

In this research, we present a two-stage dimensional reduction approach to low-dimensional data representation of which the recognition performance degrades gracefully. The proposed approach reduces the dimension of high-dimensional input data as much as possible, while preserving the information necessary for the pattern classification. Our idea is that if data is well-clustered into classes, features extracted from a topology-preserving map of the data are appropriate for recognition when low-dimensional features are to be used. Based on this idea, we apply a mapping to the input data to achieve the most separation of classes, followed by another mapping to preserve the topology of the data that the first map produces. By "topology-preserving map", we mean that vectors in the neighborhood in the input space are projected in the neighborhood in the output space [4].

To experimentally prove our claim, we have presented two novel methods for extremely low-dimensional representation of data with graceful degradation of recognition performance. It is composed of two consecutive mappings of the input data. The first mapping is concerned with best separation of the input data into classes and the second focuses on the mapping in the sense that the distance relationship between data points is kept. Our methods are implemented as the following. The first method employs FLD and SOFM. SOFM preserves the distance relationship before and after the data is transformed. This way, it is possible to represent data in low-dimensions without serious degradation of recognition performance. The second method uses FLD and MDS. The MDS preserves the distance relationship before and after the data is transformed as closely as possible.

The following section gives a brief overview of the feature extraction and dimensional reduction methods that have preciously been used for object recognition. In section 3, we describe the proposed methods: FLD combined with SOFM method and the FLD combined with MDS method. Let us call them 'FLD+SOFM' and 'FLD+MDS' methods, respectively. We report the experimental results on the recognition performance of FLD+SOFM and FLD+MDS methods in section 4.

2 Dimensional Reduction and Topology-Preserving Map

There have been reported many algorithms for dimensional reduction and feature extraction. One group of dimensional reduction methods can be referred to as *topology-preserving mapping* and another group as *well-clustered mapping*. Among the former group are SOFM, MDS and GTM (Generative Topographic Mapping) [7] and these methods are used mainly for data visualization or data compression. FLD, Kernel FLD [8] and multi-layer neural networks are examples of the latter group and are mostly used for pattern classification [9].

We can achieve very low-dimensional data representation with graceful degradation of performance by using a *topology-preserving map* when the data is well clustered into classes. However, the typical facial image data in real environments do not have well-clustered distribution and it is not guaranteed to achieve high classification performance by a *topology-preserving map* although we can get a low-dimensional data set. Accordingly, we have to focus more on the discriminant power rather than dimensional reduction in the case.

3 Two-Stage Dimensional Reduction

We present a two-stage dimensional reduction approach to low-dimensional data representation by applying two different maps in a row. The first stage is only concerned with best separation of classes. Once the data is rendered well-separated into classes by the first stage map, the second stage map only focuses on preservation of topological continuity before and after the map of the data. As previously described, the idea is based on the fact that if data is well-clustered into classes, features extracted from a topology-preserving map of the data are appropriate for recognition when extremely low-dimensional features are to be used.

3.1 Method I: FLD+SOFM

Let us $\mathbf{x}_k \in \mathbb{R}^N, k = 1, \cdots, M$ be a set of training data. FLD produces a linear discriminant function $\mathbf{f}(\mathbf{x}) = \mathbf{W^T x}$ which maps the input data onto the classification space. FLD finds a matrix $\mathbf{W}$ that maximizes

$$J(\mathbf{W}) = \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|} \tag{1}$$

where $\mathbf{S}_b$ and $\mathbf{S}_w$ are between- and within-class scatter matrices, respectively. $\mathbf{W}$ is computed by maximizing $J(\mathbf{W})$. That is, we find a subspace where, for the data projected onto the subspace, between-class variance is maximized while minimizing within-class variance. As a result of the first map, we obtain $\mathbf{z} = \mathbf{W}^T\mathbf{x}$.

After the stage of FLD, the next stage maps $\mathbf{z}$ onto a low-dimensional feature space $\mathbf{f} = \mathbf{G}(\mathbf{z})$ by SOFM. SOFM is a kind of competitive network. SOFM first determines the winning neuron using a competitive layer. Next, weight vectors for all neurons within a certain neighborhood of the winning neuron are updated using the Kohonen rule [4]. When a vector is presented, the weights of the winning neuron and its neighbors move toward the input pattern. After learning, the neurons of the output layer will be a feature map revealing a distance relationship within input patterns.

3.2 Method II: FLD+MDS

Let us $\mathbf{x}_k \in \mathbb{R}^N, k = 1, \cdots, M$ be a set of observations and $\mathbf{D}$ be a dissimilarity matrix. Classical MDS is an algebraic method to find a set of points in low-dimensional space so that the dissimilarity are well-approximated by the interpoint distances.

In summary, the inner product matrix of raw data $\mathbf{B} = \mathbf{X^T X}$ can be computed by $\mathbf{B} = -\frac{1}{2}\mathbf{HDH}$, where $\mathbf{X}$ is the data matrix $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_M] \in \mathbb{R}^{N\times M}$ and $\mathbf{H}$ is a centering matrix $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}}{\mathbf{M}}\mathbf{1^T 1}$. $\mathbf{B}$ is real, symmetric and positive semi-definite. Let the eigendecomposition of $\mathbf{B}$ be $\mathbf{B} = \mathbf{V\Lambda V^T}$, where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{V}$ is a matrix whose columns are the eigenvectors of $\mathbf{B}$. The matrix $\hat{\mathbf{X}}$ for low-dimensional feature vectors can be obtained as $\hat{\mathbf{X}} = \mathbf{\Lambda_k^{1/2} V_k^T}$ where $\mathbf{\Lambda}_k^{1/2}$ is a diagonal matrix of k largest eigenvalues and $\mathbf{V}_k$ is its corresponding eigenvectors matrix. Thus, we can compute a set of feature vectors, $\hat{\mathbf{X}}$, for a low-dimensional representation. See [10] for a detailed description.

We could not map new input vectors to features by using the classical MDS because the map is not explicitly defined in the method [11]. We used a method that achieves mapping onto an MDS subspace via PCA based on the relationship between MDS and PCA. Let $\mathbf{Y_{MDS}}$ be a set of feature vectors in an MDS subspace and $\mathbf{Y_{PCA}}$ be a set of feature vectors in a PCA subspace. Let $\mathbf{\Lambda_{MDS}}$ denotes the diagonal matrix of eigenvalues of inner product matrix $\mathbf{B}$. Then, the relationship between PCA and MDS is

$$\mathbf{Y_{PCA}} = \mathbf{\Lambda_{MDS}}^{1/2}\mathbf{Y_{MDS}}. \tag{2}$$

For the purpose of low-dimensional feature extraction, we need to compute projections onto FLD and MDS subspaces. Let $\mathbf{p}$ be an input pattern, then the feature vector in FLD+MDS space becomes

$$\mathbf{f_{FLD+MDS}} = (\mathbf{\Lambda_{PCA}^{-1/2}})\mathbf{W_{PCA}^T W_{FLD}^T}\ \mathbf{p}. \tag{3}$$

See [12] for a detailed description.

4 Experimental Results

We have evaluated the recognition performance of the proposed FLD+SOFM and FLD+MDS methods as follows.

4.1 Experiment I: FLD+SOFM with Yale and AT&T Databases

We have used Yale [13] and AT&T [14] databases. Yale database contains 165 images of 15 persons and AT&T database contains 400 images of 40 persons. We tightly cropped and normalized all the facial images.

In the SOFM stage, the entire training patterns are represented by the indices of neurons corresponding to two-dimensional map. In testing, only the node that is the most similar to the given input pattern is activated. As a result,

Table 1. Correct Recognition Rates (%) (C: number of class)

Dimension	Methods	Yale (C=15)	AT&T (C=40)
2	PCA	16.4	11.9
	FLD	41.8	11.9
	SOFM	64.3	71.3
	FLD+SOFM	96.4	86.2
	FLD+MDS	65.5	42.5
C-1	PCA	87.3	94.0
	FLD	98.2	94.8
	FLD+MDS	100.0	91.8

input patterns are classified into classes of the activated nodes. In the proposed method, the number of input neurons in SOFM is the same as the dimension of feature vectors obtained from the FLD stage. The output layer represents a two dimensional square map.

We have applied cross validation because the performance of the SOFM algorithm varies depending on the initial parameters. First, we change the number of grids. After learning using multiple SOFMs, we evaluate the performance using the validation set. We have decided the number of neurons as the number of grids that have the highest average recognition performance. Secondly, after the number of neurons is settled, multiple SOFMs with various initial parameters are learned by the learning set. Then we select the SOFM that has high performance corresponding to the upper 10% in the validation set.

As shown in Table 1, FLD+SOFM method performs better than the others in the case of very low-dimensional representation. The recognition rate of FLD is high (98.2%) when a sufficient number, C-1, of features are used. However, the recognition rate degraded significantly to 41.8% when only two dimensional representation of the data is used. The recognition rate of SOFM is higher than that of FLD when two dimensional representation is employed.

4.2 Experiment II: FLD+MDS with FERET Database

We have compared the recognition performance of FLD [2] and the proposed FLD+MDS method using a part of FERET database [15]. The whole set of images, U, consists of three subsets named ‘ba’, ‘bj’ and ‘bk’. Basically, the whole set U contains images of 200 persons and each person in the U has three different images within the ‘ba’, ‘bj’ and ‘bk’ sets. The ‘ba’ set is a subset of ‘fa’ which has images with normal frontal facial expression. The ‘bj’ set is a subset of ‘fb’. The images of ‘fb’ have some other frontal facial expressions. The ‘ba’ and ‘bj’ set contain 200 images of 200 persons, respectively. The ‘bk’ set is equal to the ‘fc’ of which images were taken with different cameras and under different lighting conditions. The ‘bk’ set contains 194 images of 194 persons.

For the experiment, we have divided the whole set U into training set (T), gallery set (G) and probe set (P). No one within the training set (T) is included in

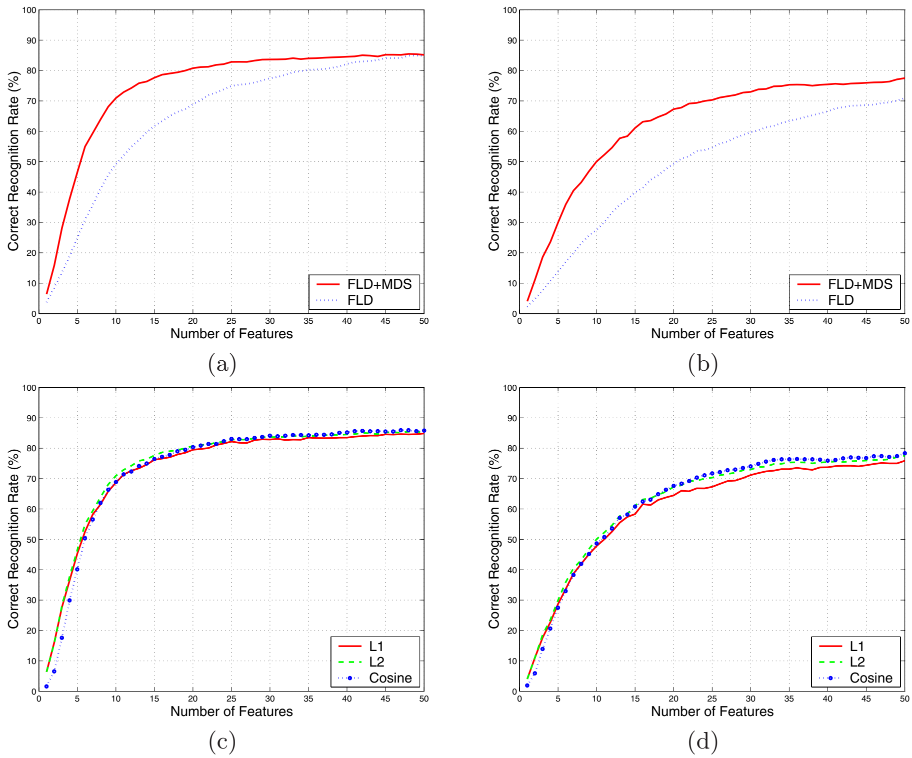

Fig. 1. Comparison of recognition rates: (a) and (b) represent recognition rates for 'ba'-'bj' set and 'ba'-'bk' set, respectively. (c) and (d) represent recognition rates for various distance measures in the case of 'ba'-'bj' set and 'ba'-'bk' set, respectively

the gallery and the probe sets. The experiment consists of two sub-experiments; The first experiment is concerned with evaluation regarding normal facial expression changes. We use the 'ba' set as the gallery and the 'bj' set as the probe. The second experiment is to evaluate the performance under illumination changes. We have assigned the 'ba' set to the gallery and the 'bk' set to the probe. In addition, we randomly selected 50% of the whole set in each sub-experiment in order to reduce the influence of a particular training set because a facial recognition algorithm based on statistical learning depends on the selection of training images. Thus, a training set contains 100 persons in each sub-experiment.

As shown in Figure 1, FLD+MDS method performs better than the others in the case of low-dimensional representation. The experimental results show that low-dimensional data representation with graceful degradation of recognition performance can be achieved by using an inter-distance preserving map after the input data is rendered well clustered into classes. The recognition rate for a given number of features in these figures was obtained by averaging thirty experiments.

We can see that there is no significant performance difference between the three distance measures (L1, L2 and cosine).

5 Conclusion

This research features a novel approach to low dimensional reduction of facial data that do not give significant degradation of the recognition rate. We have proposed two methods. The FLD+SOFM method achieves very accurate recognition rates although only two dimensional features are used for recognition. The FLD+MDS method also outperforms FLD method when represented in a low-dimensional space. These results experimentally prove that if data is tightly clustered and well separated into classes, a few features extracted from a topology-preserving map of the data are appropriate low dimensional features for recognition without significant degradation of recognition performance.

Our methods are practically useful for face recognition, especially when facial feature data need to be stored in low capacity storing devices such as bar codes and smart cards. It is also readily applicable to real-time face recognition in the case of a large database.

Acknowledgements

This work was supported in part by grant No. R01-1999-000-00339-0 from the Basic Research Program KOSEF and BK21.

References

[1] Turk, M., Pentland, A.: "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. 131
[2] Belhumeur, P., Hespanha, J., Kriegman, D.: "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711–720, 1997. 131, 135
[3] Bartlett, M. S., Martin, H., Sejnowski, T. J.: "Independent Component Representations for Face Recognition," *Proceedings of the SPIE*, Vol. 3299, pp. 528–539, 1998. 131
[4] Kohonen, T.: *Self-Organizing Maps*, Springer-Verlag, 1995. 132, 133
[5] Friedman, J. K., Tukey, J. W.: "A Projection Pursuit Algorithm for Exploratoty Data Analysis," *IEEE Trans on computers*, vol. 23, pp. 881–889, 1974. 132
[6] Duda, R. O., Hart, P. E., Stork, D. G.: *Pattern Classification*, John Wiley & Sons, Inc., 2001. 132
[7] Bishop, C. M., Svensén, M.: "GTM: The Generative Topographic Mapping," *Neural Computation*, Vol. 10. No. 1, pp. 215–234, 1998. 133
[8] Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Müller, K. R.: "Fisher Discriminant Analysis with Kernels," *IEEE Neural Networks for Signal Processing IX*, pp. 41–48, 1999. 133

[9] Carreira-Perpoñán, M.: “A Reivew of Dimension Reduction Techniques,” *Technical Report CS-96-09*, Dept. of Computer Science University of Sheffield, 1997. 133
[10] Pcekalska, E., Paclík, P., Duin, R. P. W.: “A Generalized Kernel Approach to Dissimilarity-based Classification,” *Journal of Machine Learning Research*, vol. 2, pp. 175–211, 2001. 134
[11] Chandrasiri, N. P., Park, M. C., Naemura, T., Harashima, H.: ”Personal Facial Expression Space based on Multidimensional Scaling for the Recognition Improvement”, Proc. IEEE ISSPA’99, pp. 943–946, 1999. 134
[12] Choi, J. M., Yi, J. H.: “Low-Dimensional Image Representation for Face Recognition,” *Workshop on Multimodal User Authentication*, 2003. 134
[13] http://cvc.yale.edu/projects/yalefaces/yalefaces.html. 134
[14] http://www.uk.research.att.com/facedatabase.html. 134
[15] Phillips, P. J., Moon, H. J., Rizvi, S. A., Rauss,. P. J.: “The FERET Evaluation Methodology for Face-Recognition Algorithms,” *IEEE Trans. on PAMI*, vol. 22, No. 10, pp. 1090–1104, 2000. 135