

# Extended Isomap for Pattern Classification

**Ming-Hsuan Yang**

Honda Fundamental Research Labs  
Mountain View, CA 94041  
myang@hira.com

## Abstract

The Isomap method has demonstrated promising results in finding low dimensional manifolds from data points in the high dimensional input space. While classical subspace methods use Euclidean or Manhattan metrics to represent distances between data points and apply Principal Component Analysis to induce linear manifolds, the Isomap method estimates geodesic distances between data points and then uses Multi-Dimensional Scaling to induce low dimensional manifolds. Since the Isomap method is developed based on reconstruction principle, it may not be optimal from the classification viewpoint. In this paper, we present an extended Isomap method that utilizes Fisher Linear Discriminant for pattern classification. Numerous experiments on image data sets show that our extension is more effective than the original Isomap method for pattern classification. Furthermore, the extended Isomap method shows promising results compared with best methods in the face recognition literature.

## Introduction

Subspace methods can be classified into two main categories: either based on reconstruction (i.e., retaining maximum sample variance) or classification principle (i.e., maximizing the distances between samples). Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) have been applied to numerous applications and have shown their abilities to find low dimensional structures from high dimensional samples (Duda, Hart, & Stork 2001). These unsupervised methods are effective in finding compact representations and useful for data interpolation and visualization. On the other hand, Fisher Linear Discriminant (FLD) and alike have shown their successes in pattern classification when class labels are available (Bishop 1995) (Duda, Hart, & Stork 2001). Contrasted to PCA which finds a projection direction that retains maximum variance, FLD finds a projection direction that maximizes the distances between cluster centers. Consequently, FLD-based methods have been shown to perform well in classification problems such as face recognition (Belhumeur, Hespanha, & Kriegman 1997).

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

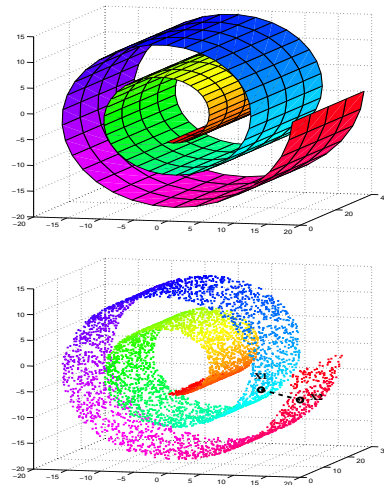


Figure 1: A complex manifold that shows why Euclidean distances may not be good metrics in pattern recognition.

Recently, two dimensionality reduction methods have been proposed for learning complex embedding manifolds using local geometric metrics within a single global coordinate system (Roweis & Saul 2000) (Tenenbaum, de Silva, & Langford 2000). The Isomap (or isometric feature mapping) method argues that only the geodesic distance reflects the intrinsic geometry of the underlying manifold (Tenenbaum, de Silva, & Langford 2000). Figure 1 shows one example where data points of different classes are displayed in distinct shaded patches (top) and data points sampled from these classes are shown (bottom). For a pair of points on the manifold, their Euclidean distance may not accurately reflect their intrinsic similarity and consequently is not suitable for determining intrinsic embedding or pattern classification. The Euclidean distance between circled data points (e.g.,  $x_1$  and  $x_2$  in Figure 1) may be deceptively small in the three-dimensional input space though their geodesic distance on an intrinsic two-dimensional manifold is large. This problem can be remedied by using geodesic distance (i.e., distance metrics along the surface of the manifold) if one is able to compute or estimate such metrics. The Isomap

method first constructs a neighborhood graph that connects each point to all its  $k$ -nearest neighbors, or to all the points within some fixed radius  $\epsilon$  in the input space. For neighboring points, the input space distance usually provides a good approximation to their geodesic distance. For each pair of points, the shortest path connecting them in the neighborhood graph is computed and is used as an estimate of the true geodesic distance. These estimates are good approximations of the true geodesic distances if there are sufficient number of data points (See Figure 1). The classical multidimensional scaling method is then applied to construct a low dimensional subspace that best preserves the manifold's estimated intrinsic geometry.

The Locally Linear Embedding (LLE) method captures local geometric properties of complex embedding manifolds by a set of linear coefficients that best approximates each data point from its neighbors in the input space (Roweis & Saul 2000). LLE then finds a set of low dimensional points where each can be linearly approximated by its neighbors with the same set of coefficients that was computed from the high dimensional data points in the input space while minimizing reconstruction cost. Although these two methods have demonstrated excellent results in finding the embedding manifolds that best describe the data points with minimum reconstruction error, they are suboptimal from the classification viewpoint. Furthermore, these two methods assume that the embedding manifold is well sampled which may not be the case in some classification problems such as face recognition since there are typically only a few samples available for each person.

In this paper, we propose a method that extends the Isomap method with Fisher Linear Discriminant for classification. The crux of this method is to estimate geodesic distance, similar to what is done in Isomap, and use pairwise geodesic distances as feature vectors. We then apply FLD to find an optimal projection direction to maximize the distances between cluster centers. Experimental results on three data sets show that the extended Isomap method consistently performs better than the Isomap method, and performs better than or as equally well as some best methods in the face recognition literature.

### Extended Isomap

Consider a set of  $m$  samples  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  and each sample belongs to one of the  $c$  class  $\{Z_1, \dots, Z_c\}$ , the first step in the extended Isomap method is, similar to the Isomap method, to determine the neighbors of each sample  $\mathbf{x}_i$  on the low dimensional manifold  $M$  based on some distance metrics  $d_X(\mathbf{x}_i, \mathbf{x}_j)$  in the input space  $X$ . Such metrics can be Euclidean distance that is often used in face recognition (Turk & Pentland 1991) or tangent distance that has been shown to be effective in hand digit recognition (Simard, Le Cun, & Denker 1993). The assumption is that input space distance provides a good approximation to geodesic distance for neighboring points (See Figure 1). Consequently, input space distance metric can be utilized to determine whether two data points are neighbors or not. The  $k$ -Isomap method uses a  $k$ -nearest neighbor algorithm to determine neighbors while the  $\epsilon$ -Isomap method includes all the points within

some fixed radius  $\epsilon$  as neighbors. These neighborhood relationships are represented in a weighted graph  $G$  in which  $d_G(\mathbf{x}_i, \mathbf{x}_j) = d_X(\mathbf{x}_i, \mathbf{x}_j)$  if  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are neighbors, and  $d_G(\mathbf{x}_i, \mathbf{x}_j) = \infty$  otherwise.

The next step is to estimate geodesic distance  $d_M(\mathbf{x}_i, \mathbf{x}_j)$  between any pair of points on the manifold  $M$ . For a pair of points that are far away, their geodesic distance can be approximated by a sequence of short hops between neighboring data points. In other words,  $d_M(\mathbf{x}_i, \mathbf{x}_j)$  is approximated by the shortest path between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  on  $G$ , which is computed by the Floyd-Warshall algorithm (Cormen, Leiserson, & Rivest 1989):

$$d_G(\mathbf{x}_i, \mathbf{x}_j) = \min\{d_G(\mathbf{x}_i, \mathbf{x}_j), d_G(\mathbf{x}_i, \mathbf{x}_k) + d_G(\mathbf{x}_k, \mathbf{x}_j)\}$$

The shortest paths between any two points are represented in a matrix  $D$  where  $D_{ij} = d_G(\mathbf{x}_i, \mathbf{x}_j)$ .

The main difference between extended Isomap and the original method is that we represent each data point by a feature vector of its geodesic distance to any points, and then apply Fisher Linear Discriminant on the feature vectors to find an optimal projection direction for classification. In other words, the feature vector of  $\mathbf{x}_i$  is an  $m$  dimensional vector  $\mathbf{f}_i = [D_{ij}]$  where  $j = 1, \dots, m$  and  $D_{ii} = 0$ .

The between-class and within-class scatter matrices in Fisher Linear Discriminant are computed by:

$$S_B = \sum_{i=1}^c N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$$

$$S_W = \sum_{i=1}^c \sum_{\mathbf{f}_k \in Z_i} (\mathbf{f}_k - \boldsymbol{\mu}_i)(\mathbf{f}_k - \boldsymbol{\mu}_i)^T$$

where  $\boldsymbol{\mu}$  is the mean of all samples  $\mathbf{f}_k$ ,  $\boldsymbol{\mu}_i$  is the mean of class  $Z_i$ ,  $S_{W_i}$  is the covariance of class  $Z_i$ , and  $N_i$  is the number of samples in class  $Z_i$ . The optimal projection  $W_{FLD}$  is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples:

$$W_{FLD} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_m]$$

where  $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$  is the set of generalized eigenvectors of  $S_B$  and  $S_W$ , corresponding to the  $m$  largest generalized eigenvalues  $\{\lambda_i | i = 1, 2, \dots, m\}$ . The rank of  $S_B$  is  $c - 1$  or less because it is the sum of  $c$  matrices of rank one or less. Thus, there are at most  $c - 1$  nonzero eigenvalues (Duda, Hart, & Stork 2001). Finally, each data point  $\mathbf{x}_i$  is represented by a low dimensional feature vector computed by  $\mathbf{y}_i = W_{FLD} \mathbf{f}_i$ . The extended Isomap algorithm is summarized in Figure 2.

The computational complexity and memory requirement of the Isomap and the extended Isomap are dominated by the calculation of all pair shortest paths. The Floyd-Warshall algorithm requires  $O(m^3)$  operations and stores  $O(m^2)$  elements of estimated geodesic distances for straightforward implementations. On the other hand, the MDS procedure in the Isomap method can be time consuming as a result of its iterative operations to detect meaningful underlying dimensions that explain the observed similarities or dissimilarities (distances) between data points.

---

### 1. Constructing neighboring graph

First compute Euclidean distance,  $d_X(\mathbf{x}_i, \mathbf{x}_j)$  between any two points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  in the input space  $X$ . Next connect neighbors of any point  $\mathbf{x}_i$  by finding its  $k$ -nearest neighbors or all the points that are within  $\epsilon$  radius of  $\mathbf{x}_i$ . The procedure results in a weighted graph  $d_G(\mathbf{x}_i, \mathbf{x}_j)$  where

$$d_G(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} d_X(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors} \\ \infty & \text{otherwise.} \end{cases}$$

### 2. Computing shortest path between pairs of points

Compute shortest path between any pair of points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  on  $d_G$  using Floyd-Warshall algorithm, i.e.,

$$d_G(\mathbf{x}_i, \mathbf{x}_j) = \min\{d_G(\mathbf{x}_i, \mathbf{x}_j), d_G(\mathbf{x}_i, \mathbf{x}_k) + d_G(\mathbf{x}_k, \mathbf{x}_j)\}$$

The shortest paths between any two points are represented in a matrix  $D$  where  $D_{ij} = d_G(\mathbf{x}_i, \mathbf{x}_j)$ .

### 3. Determining most discriminant components

Represent each point  $\mathbf{x}_i$  by a feature vector  $\mathbf{f}_i$  where  $f_i = [D_{ij}]$ ,  $j = 1, \dots, m$ . Determine a subspace where the class centers are separated as far as possible by using the Fisher Linear Discriminant method.

---

Figure 2: Extended Isomap Algorithm.

It can be shown that the graph  $d_G(\mathbf{x}_i, \mathbf{x}_j)$  provides increasing better estimates to the intrinsic geodesic distance  $d_M(\mathbf{x}_i, \mathbf{x}_j)$  as the number of data points increases (Tenenbaum, de Silva, & Langford 2000). In practice, there may not be sufficient number samples at one's disposal so that the geodesic distances  $d_G(\mathbf{x}_i, \mathbf{x}_j)$  may not be good approximates. Consequently, the Isomap may not be able to find intrinsic dimensionality from data points and not suitable for classification purpose. In contrast, the extended Isomap method utilizes the distances between the scatter centers (i.e., poor approximates may be averaged out) and thus may perform well for classification problem in such situations. While the Isomap method uses classical MDS to find dimensions of the embedding manifolds, the dimensionality of the subspace is determined by the number of class (i.e.,  $c - 1$ ) in the extended Isomap method.

To deal with the singularity problem of within-scatter matrix  $S_W$  that one often encounters in classification problems, we can add a multiple of the identity matrix to the within-scatter matrix, i.e.,  $S_W + \epsilon I$  (where  $\epsilon$  is a small number). This also makes the eigenvalue problem numerically more stable. See also (Belhumeur, Hespanha, & Kriegman 1997) for a method using PCA to overcome singularity problems in applying FLD to face recognition.

## Experiments

Two classical pattern classification problems, face recognition and handwritten digit recognition, are considered in order to analyze the performance of the extended and original

Isomap methods. These two problems have several interesting characteristics and are approached quite differently. In the appearance-based methods for face recognition in frontal pose, each face image provides a rich description of one's identity and as a whole (i.e., holistic) is usually treated as a pattern without extracting features explicitly. Instead, subspace methods such as PCA or FLD are applied to implicitly extract meaningful (e.g., PCA) or discriminant (e.g., FLD) features and then project patterns to a lower dimensional subspace for recognition. On the contrary, sophisticated feature extraction techniques are usually applied to handwritten digit images before any decision surface is induced for classification.

We tested both the original and extended Isomap methods against LLE (Roweis & Saul 2000), Eigenface (Turk & Pentland 1991) and Fisherface (Belhumeur, Hespanha, & Kriegman 1997) methods using the publicly available AT&T (Samaria & Young 1994) and Yale databases (Belhumeur, Hespanha, & Kriegman 1997). The face images in these databases have several unique characteristics. While the images in the AT&T database contain facial contours and vary in pose as well as scale, the face images in the Yale database have been cropped and aligned. The face images in the AT&T database were taken under well controlled lighting conditions whereas the images in the Yale database were acquired under varying lighting conditions. We used the first database as a baseline study and then used the second one to evaluate face recognition methods under varying lighting conditions. For handwritten digit recognition problem, we tested both methods using the MNIST database which is the *de facto* benchmark test set.

### Face Recognition: Variation in Pose and Scale

The AT&T (formerly Olivetti) face database contains 400 images of 40 subjects (<http://www.uk.research.att.com/facedatabase.html>). To reduce computational complexity, each face image is downsampled to  $23 \times 28$  pixels for experiments. We represent each image by a raster scan vector of the intensity values, and then normalize them to be zero-mean unit-variance vectors. Figure 3 shows images of a few subjects. In contrast to images of the Yale database shown in Figure 5, these images include facial contours, and variations in pose as well as scale. However, the lighting conditions remain relatively constant.



Figure 3: Face images in the AT&T database.

The experiments were performed using the “leave-one-out” strategy (i.e.,  $m$  fold cross validation): To classify an image of a person, that image is removed from the training set of  $(m - 1)$  images and the projection matrix is computed. All the  $m$  images in the training set are projected to a reduced space and recognition is performed using a nearest neighbor classifier. The parameters, such as number of principal components in Eigenface and LLE methods, were empirically determined to achieve the lowest error rate by each method. For Fisherface and extended Isomap methods, we project all samples onto a subspace spanned by the  $c - 1$  largest eigenvectors. The experimental results are shown in Figure 4. Among all the methods, the extended Isomap method with  $\epsilon$  radius implementation achieves the lowest error rate and outperforms the Fisherface method by a significant margin. Notice also that two implementations of the extended Isomap (one with  $k$ -nearest neighbor, i.e., extended  $k$ -Isomap, and the other with  $\epsilon$  radius, i.e., extended  $\epsilon$ -Isomap) to determine neighboring data points) consistently perform better than their counterparts in the Isomap method by a significant margin.

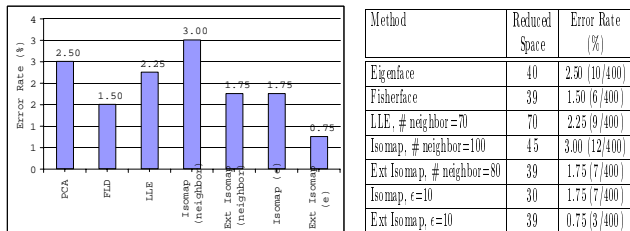


Figure 4: Results with the AT&T database.

### Face Recognition: Variation in Lighting and Expression

The Yale database contains 165 images of 11 subjects with facial expression and lighting variations (available at <http://cvc.yale.edu/>). For computational efficiency, each image has been downsampled to  $29 \times 41$  pixels. Similarly, each face image is represented by a centered vector of normalized intensity values. Figure 5 shows closely cropped images of a few subjects which include internal facial structures such as the eyebrow, eyes, nose, mouth and chin, but do not contain facial contours.



Figure 5: Face images in the Yale database.

Using the same leave-one-out strategy, we varied the number of principal components to achieve the lowest error rates for Eigenface and LLE methods. For Fisherface and extended Isomap methods, we project all samples onto a subspace spanned by the  $c - 1$  largest eigenvectors. The experimental results are shown in Figure 6. Both implementations of the extended Isomap method perform better than their counterparts in the Isomap method. Furthermore, the extended  $\epsilon$ -Isomap method performs almost as well as the Fisherface method (which is one of the best methods in the face recognition literature) though the original Isomap does not work well on the Yale data set.

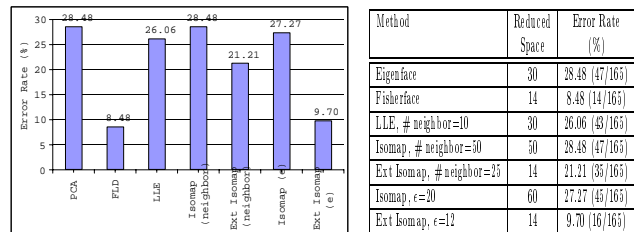


Figure 6: Results with the Yale database.

Figure 7 shows more performance comparisons between Isomap and extended Isomap methods in both  $k$ -nearest neighbor as well as  $\epsilon$  radius implementations. The extended Isomap method consistently outperforms the Isomap method with both implementations in all the experiments.

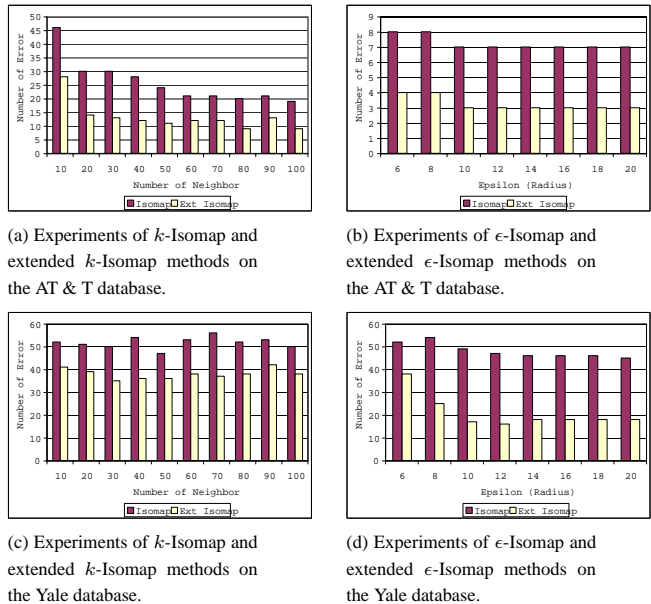
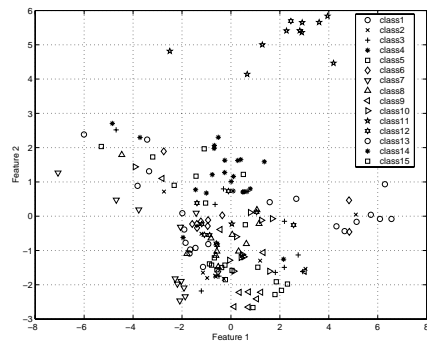
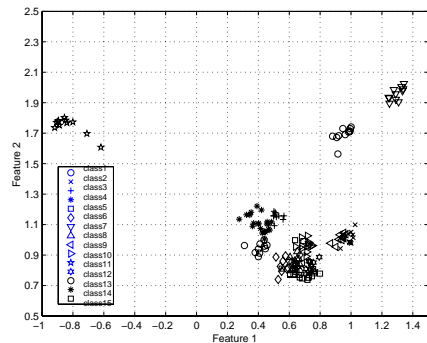


Figure 7: Performance of the Isomap and the extended Isomap methods.

As one example to explain why extended Isomap performs better than Isomap, Figure 8 shows training and test samples of the Yale database projected onto the first two



(a) Isomap method.



(b) Extended Isomap method.

Figure 8: Samples projected by the Isomap and the extended Isomap methods.

eigenvectors extracted by both methods. The projected samples of different classes are smeared by the Isomap method (Figure 8(a)) whereas the samples projected by the extended Isomap method are separated well (Figure 8(b)).

### Handwritten Digit Recognition

The MNIST database of handwritten digits comprises a training set of 60,000 examples, and a test set of 10,000 examples (publicly available at <http://www.research.att.com/~yann/exdb/mnist/index.html>). The images are normalized while preserving their aspect ratio, and each one is centered in a  $28 \times 28$  window of gray scales by computing the center of mass of the pixels, and translating the image so as to position this point at the center. Some handwritten digit images of the MNIST database are shown in Figure 9.

Due to computational and memory constraints, we randomly selected a training set of 1,500 MNIST images and a non-overlapping test set of 250 images for experiments. We repeated the same experiment five times and varied the parameters to achieve the lowest error rates in each run. As a baseline study, each image is represented by a raster scan vector of intensity values without applying any feature extraction algorithms. Figure 10 shows the averaged results by  $k$ -Isomap,  $\epsilon$ -Isomap and our extended methods. The extended Isomap methods consistently outperform the original Isomap methods in our experiments with the MNIST data sets.

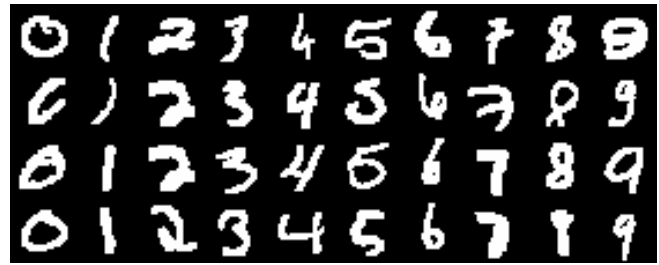


Figure 9: MNIST digit images.

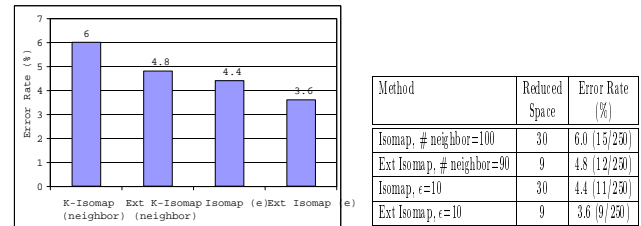


Figure 10: Results with the MNIST database.

We note that the results achieved by the extended Isomap, shown in Figure 10, are not as competitive as the best methods in the literature (Burges & Schölkopf 1997) (LeCun *et al.* 1998) (Belongie, Malik, & Puzicha 2001), the reason being that no feature extraction algorithm is applied to the raw images which is in direct contrast to the best results reported in the literature (such as support vector machines, convolutional neural networks and shape contexts that use complex feature extraction techniques with nonlinear decision surfaces). Our MNIST digit recognition experiments serve as a baseline study to investigate the improvement of the extended Isomap over the original one in visual pattern recognition problems. The recognition rates can be improved by applying sophisticated feature extraction techniques or with  $k$ -nearest neighbor classifier in the lower dimensional subspace.

### Discussion

The experimental results of two classical pattern recognition problems, face recognition and handwritten digit recognition, show that the extended Isomap method performs better than the Isomap method in the pattern classification tasks. On the other hand, we note that the Isomap method is an effective algorithm to determine the intrinsic dimensionality of the embedding manifolds and thus suitable for interpolation between data points (See (Tenenbaum, de Silva, & Langford 2000) for examples).

The feature extraction issues that we alluded to in the handwritten digit recognition experiments can be addressed by combining kernel tricks to extract meaningful and nonlinear features of digit patterns. One advantage of kernel methods is that it provides a computationally efficient method that integrates feature extraction with large margin classifier. We plan to extend the Isomap method with kernel Fisher

Linear Discriminant (Mika *et al.* 2000) (Roth & Steinhage 2000) (Baudat & Anouar 2000) for pattern classification.

## Conclusion

In this paper, we present an extended Isomap method for pattern classification when the labels of data points are available. The Isomap method is developed based on reconstruction principle, and thus it may not be optimal from the classification viewpoint. Our extension is based on Fisher Linear Discriminant which aims to find an optimal project direction such that the data points in the subspace are separated as far away as possible.

Our experiments on face and handwritten digit recognition suggest a number of conclusions:

1. The extended Isomap method performs consistently better than the Isomap method in classification (with both  $k$  nearest neighbor and  $\epsilon$  radius implementations) by a significant margin.
2. Geodesic distance appears to be a better metric than Euclidean distance for face recognition in all the experiments.
3. The extended Isomap method performs better than one of the best methods in the literature on the AT & T database. When there exist sufficient number of samples so that the shortest paths between any pair of data points are good approximates of geodesic distances, the extended Isomap method performs well in classification.
4. The extended Isomap method still performs well while the Isomap method does not in the experiments with the Yale database. One explanation is that insufficient samples result in poor approximates of geodesic distances. However, poor approximates may be averaged out by the extended Isomap method and thus it performs better.
5. Though the Isomap and LLE methods have demonstrated excellent results in finding the embedding manifolds that best describe the data points with minimum reconstruction error, they are suboptimal from the classification viewpoint. Furthermore, these two methods assume that the embedding manifold is well sampled which may not be the case in face recognition since there are typically only a few samples available for each person.

Our future work will focus on efficient methods for estimating geodesic distance, and performance evaluation with large and diverse databases. We plan to extend our method by applying the kernel tricks, such as kernel Fisher Linear Discriminant, to provide a richer feature representation for pattern classification. We also plan to compare the extended Isomap method against other learning algorithms with UCI machine learning data sets, as well as reported face recognition methods using FERET (Phillips *et al.* 2000) and CMU PIE (Sim, Baker, & Bsat 2001) databases.

## References

Baudat, G., and Anouar, F. 2000. Generalized discriminant analysis using a kernel approach. *Neural Computation* 12:2385–2404.

Belhumeur, P.; Hespanha, J.; and Kriegman, D. 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):711–720.

Belongie, S.; Malik, J.; and Puzicha, J. 2001. Matching shapes. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 1, 454–461.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press.

Burges, C. J., and Schölkopf, B. 1997. Improving the accuracy and speed of support vector machines. In Mozer, M. C.; Jordan, M. I.; and Petsche, T., eds., *Advances in Neural Information Processing Systems*, volume 9, 375. The MIT Press.

Cormen, T. H.; Leiserson, C. E.; and Rivest, R. L. 1989. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company.

Duda, R. O.; Hart, P. E.; and Stork, D. G. 2001. *Pattern Classification*. New York: Wiley-Interscience.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.

Mika, S.; Rätsch, G.; Weston, J.; Schölkopf, B.; Smola, A.; and Müller, K.-R. 2000. Invariant feature extraction and classification in kernel spaces. In Solla, S.; Leen, T.; ; K.-R.; and Müller., eds., *Advances in Neural Information Processing Systems 12*, 526–532. MIT Press.

Phillips, P. J.; Moon, H.; Rizvi, S. A.; and Rauss, P. J. 2000. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(10):1090–1034.

Roth, V., and Steinhage, V. 2000. Nonlinear discriminant analysis using kernel functions. In Solla, S.; Leen, T.; ; K.-R.; and Müller., eds., *Advances in Neural Information Processing Systems 12*, 568–574. MIT Press.

Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500).

Samaria, F., and Young, S. 1994. HMM based architecture for face identification. *Image and Vision Computing* 12(8):537–583.

Sim, T.; Baker, S.; and Bsat, M. 2001. The CMU pose, illumination, and expression (PIE) database of human faces. Technical Report CMU-RI-TR-01-02, Carnegie Mellon University.

Simard, P.; Le Cun, Y.; and Denker, J. 1993. Efficient pattern recognition using a new transformation distance. In Hanson, S. J.; Cowan, J. D.; and Giles, C. L., eds., *Advances in Neural Information Processing Systems*, volume 5, 50–58. Morgan Kaufmann, San Mateo, CA.

Tenenbaum, J. B.; de Silva, V.; and Langford, J. C. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500).

Turk, M., and Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1):71–86.