

FACE RECOGNITION USING EXTENDED ISOMAP

Ming-Hsuan Yang
Honda Fundamental Research Labs
800 California Street
Mountain View, CA

ABSTRACT

The Isomap method has demonstrated promising results in finding low dimensional manifolds from samples in the high dimensional input space. While conventional subspace methods compute L_1 or L_2 metrics to represent distances between samples and apply Principal Component Analysis or alike to induce linear manifolds, the Isomap method estimates geodesic distance between samples and then uses multidimensional scaling to induce a low dimensional manifold. 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 classification methods in the literature.

1. MOTIVATION AND APPROACH

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 ability to find low dimensional structures from high dimensional samples [4]. These unsupervised methods are effective in finding compact representations and useful for data interpolation and visualization. On the other hand, Fisher Linear Discriminant (FLD) has shown their success in pattern classification where class labels are available [2] [4]. In contrast 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 [1].

Two dimensionality reduction methods have been recently proposed for learning complex embedding manifolds using local geometric metrics within a single global coordinate system [6] [10]. The Isomap (or isometric feature mapping) method argues that only the geodesic distance reflects the true geometry of the underlying manifold [10].

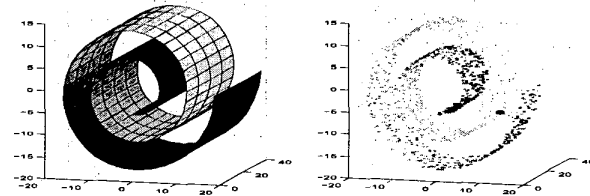


Fig. 1. A complex manifold that shows why Euclidean distances may not be good metrics in pattern recognition.

Figure 1 shows one example where the data points of different classes are displayed in distinct shaded patches on the left hand side and data points sampled from these classes are shown on the right hand side. For any two points on the manifold (circled data points on the right hand side), their Euclidean distance may not accurately reflect their intrinsic similarity and consequently is not suitable for determining true embedding or pattern classification. This problem can be remedied by using geodesic distance (i.e., distance metrics along the manifolds) 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 ϵ in the input space. 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 samples (i.e., the patches of the samples are smooth). The classical multidimensional scaling method is then applied to find a set of low dimensional points with similar pairwise distance. The Locally Linear Embedding (LLE) method captures local geometry of complex embedding manifold by a set of linear coefficients that best approximates each data point from its neighbors in the input space [6]. 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 to find the embedding manifolds that best describe the data points with minimum reconstruction er-

ror, 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.

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 the pairwise geodesic distance as feature vectors. We then apply FLD to find an optimal projection direction to maximize the distances between cluster centers. Experimental results on two 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 literature.

2. 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 appropriate distance metrics $d_X(\mathbf{x}_i, \mathbf{x}_j)$ in the input space X (e.g., Euclidean distance that is often used in face recognition or tangent distance that has been shown to be effective in hand digit recognition [9]). Two methods 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 ϵ -Isomap method includes all the points within some fixed radius ϵ 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 $d_X(\mathbf{x}_i, \mathbf{x}_j) = \infty$ 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 . Since the embedding manifold is unknown, $d_M(\mathbf{x}_i, \mathbf{x}_j)$ is approximated by the shortest path between \mathbf{x}_i and \mathbf{x}_j on G , which is computed by the Floyd-Warshall algorithm [3]:

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

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 our extended Isomap and the original method is that we represent each data point by a feature vector of geodesic distance to any other 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 $m-1$ dimensional vector $\mathbf{f}_i = [D_{ij}]$ where $j \neq i$. The between-class and within-class scatter matrices in Fisher Linear Discriminant are computed by:

$$\begin{aligned} S_B &= \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \\ S_W &= \sum_{i=1}^c \sum_{\mathbf{f}_k \in Z_i} (\mathbf{f}_k - \mu_i)(\mathbf{f}_k - \mu_i)^T \end{aligned}$$

where μ is mean of all samples \mathbf{f}_i , μ_i is 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 determined by the matrix whose orthonormal columns 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 [4]. Finally, each data point \mathbf{x}_i is represented by a low dimensional feature vector computed by $\mathbf{y}_i = W_{FLD} \mathbf{f}_i$.

In practice, there may not be sufficient 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 true structure 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 use 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 problem, we can add a multiple of the identity matrix to the within-scatter matrix, i.e., $S_W + \epsilon I$ (where ϵ is a small number). This also makes the eigenvalue problem numerically more stable. See also [1] for a PCA-based method to overcome the singularity problems.

3. EXPERIMENTS

We test both the original and extended Isomap methods against LLE [6], Eigenface [11] and Fisherface [1] methods using the publicly available AT&T [7] and Yale databases [1]. 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 use the first database as a baseline study and then use the second one to evaluate face recognition methods under varying lighting conditions.

The AT&T (formerly Olivetti) database contains 400 images of 40 subjects. To reduce computational complexity, each face image is downsampled to 23×28 pixels. We

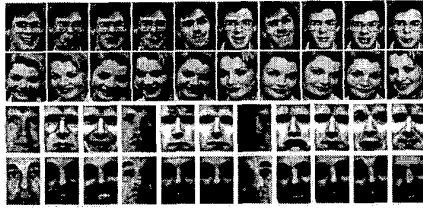


Fig. 2. Face images in the AT&T database (Top) and the Yale database (Bottom).

represent each image by a raster scan vector of the intensity values, and then normalize them to be zero-mean unit-variance vectors. Figure 2 shows images of two subjects. In contrast to images of the Yale database, these images include facial contours, and variation in pose as well as scale. However, the lighting conditions remain relatively constant.

The experiments are 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 based on a nearest neighbor classifier. The parameters, such as number of principal components in Eigenface and LLE methods, are empirically determined to achieve the lowest error rate by each method. For Fisherface and extended Isomap methods, we project all the samples onto a subspace spanned by the $c - 1$ largest eigenvectors. The experimental results are shown in Figure 3. Among all the methods, the extended Isomap method using ϵ radius 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 (extended k -Isomap) and the other with ϵ radius (extended ϵ -Isomap) to determine neighboring data points) consistently perform better than their counterparts in the Isomap method by a large margin.

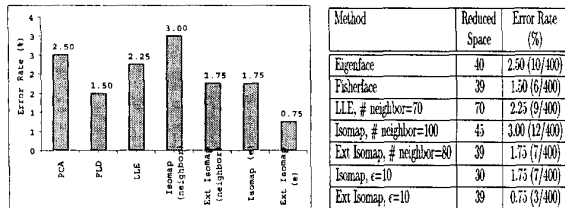


Fig. 3. Results with AT&T database.

The Yale database contains 165 images of 11 subjects that includes variation of facial expression and lighting. For computational efficiency, each image has been downsampled to 29×41 pixels. Similarly, each face image is represented by a centered vector of normalized intensity values.

figure 2 shows 22 closely cropped images of two subjects which include internal facial structures such as the eyebrow, eyes, nose, mouth and chin, but do not contain facial contours.

using the same leave-one-out strategy, we vary 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 4. Both implementations of the extended Isomap method perform better than their counterparts in the Isomap method. Furthermore, the extended ϵ -Isomap method performs almost as well as the Fisherface method (which is one of the best reported methods in the literature) though the original Isomap does not work well.

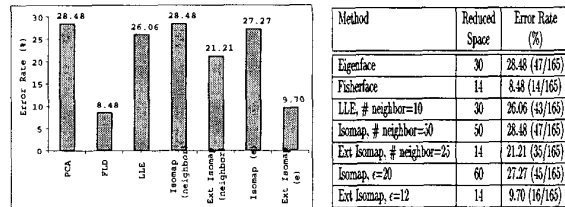


Fig. 4. Results with AT&T database.

Figure 5 shows more performance comparisons between Isomap and extended Isomap methods in both k nearest neighbor as well as ϵ radius implementations. The extended Isomap method consistently outperforms Isomap method with both implementations in all the experiments.

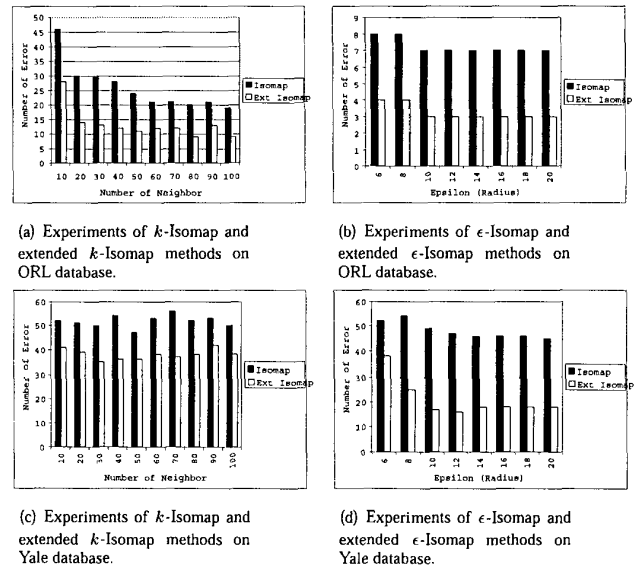


Fig. 5. Performance of Isomap and Extended Isomap methods.

As one example to explain why extended Isomap performs better than Isomap, Figure 6 shows training and test samples of the Yale database projected onto the first two eigenvectors extracted by both methods. The projected samples of different classes are smeared by the Isomap method (Figure 6(a)) whereas the samples projected by the extended Isomap method are separated well (Figure 6(b)).

4. CONCLUDING REMARKS

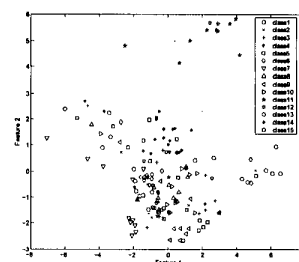
Our experiments suggest the following conclusions:

1. The extended Isomap method performs consistently better than the Isomap method in classification (with both k nearest neighbor and ϵ radius implementations) by a significant margin.
2. Geodesic distance appears to be a better metric than Euclidian 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 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.

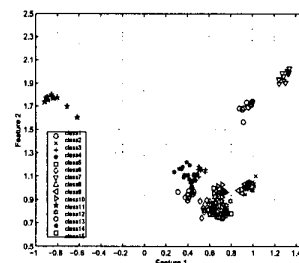
Our future work will focus on efficient methods for estimating geodesic distance, and performance evaluation with large and diverse data sets. We plan to compare the extended Isomap method against other learning algorithms with UCI machine learning data sets, as well as reported face recognition methods using FERET [5] and CMU PIE [8] databases.

5. REFERENCES

- [1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projec-



(a) Isomap method.



(b) Extended Isomap method.

Fig. 6. Samples projected by both methods.

- tion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
- [2] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [3] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989.
- [4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2001.
- [5] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1034, 2000.
- [6] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2000.
- [7] F. Samaria and S. Young. HMM based architecture for face identification. *Image and Vision Computing*, 12(8):537–583, 1994.
- [8] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database of human faces. Technical Report CMU-RI-TR-01-02, Carnegie Mellon University, 2001.
- [9] P. Simard, Y. Le Cun, and J. Denker. Efficient pattern recognition using a new transformation distance. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5, pages 50–58. Morgan Kaufmann, San Mateo, CA, 1993.
- [10] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2000.
- [11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.