Locality Preserving Projections (LPP)

Xiaofei He

Computer Science Department
The University of Chicago
Chicago, IL 60615

Partha Niyogi

Computer Science Department
The University of Chicago
Chicago, IL 60615

Abstract

Many problems in information processing involve some form of dimensionality reduction. In this paper, we introduce Locality Preserving Projections (LPP). These are linear projective maps that arise by solving a variational problem that optimally preserves the neighborhood structure of the data set. LPP should be seen as an alternative to principal component analysis (PCA) – a classical linear technique that projects the data along the directions of maximal variance. When the high dimensional data lies on a low dimensional manifold embedded in the ambient space, the Locality Preserving Projections are obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. As a result, LPP shares many of the data representation properties of non linear techniques such as Laplacian Eigenmap [4] or Locally Linear Embedding [5]. This is borne out by illustrative examples on a couple of high dimensional image data sets.

1. Introduction

Suppose we have a collection of data points of n-dimensional real vectors drawn from an unknown probability distribution. In increasingly many cases of interest in machine learning and data mining, one is confronted with the situation where n is *very large*. However, there might be reason to suspect that the "intrinsic dimensionality" of the data is much lower. This leads one to consider methods of dimensionality reduction that allow one to represent the data in a lower dimensional space.

A great number of dimensionality reduction techniques exist in the literature. In practical situations, when n is prohibitively large, one is often forced to use linear (or even sublinear) techniques. Consequently, projective maps have been the subject of considerable investigation. Three classical yet popular forms of linear techniques are the methods of principal component analysis (PCA) [1], multidimensional scaling (MDS) [2], and linear discriminant analysis (LDA). Each of these is an eigenvector method designed to model linear variabilities in high-dimensional data.

PCA performs dimensionality reduction by projecting the original n-dimensional data onto the $k \ll n$ dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix. Thus PCA builds a global linear model of the data. Classical MDS finds an embedding that preserves pairwise distances between data points, and it is equivalent to PCA when those distances are Euclidean. Both PCA and MDS are unsupervised learning algorithms. LDA is a supervised learning algorithm. LDA searches for the projective axes on which the data points of different classes are far from each other (maximize between class scatter), while constraining the data points of the same class to be as close to each other as possible (minimizing within class scatter).

Recently, there has been some renewed interest in the problem of developing low dimensional representations when data arises from sampling a probability distribution on a low dimensional manifold embedded in the high dimensional ambient space [4][5][6]. These methods do yield impressive results on some benchmark artificial data sets, as well as on real world data sets. However, their nonlinear property makes them computationally expensive. Moreover, they yield maps that are defined only on the *training* data points and how to evaluate the map on novel *test* points remains unclear.

In this paper, we propose a new *linear* dimensionality reduction algorithm, called **Locality Preserving Projections** (LPP). It builds a graph incorporating neighborhood information of the data set. Using the notion of the Laplacian of the graph, we then compute a transformation matrix which maps the data points to a subspace. This linear transformation optimally preserves local neighborhood information in a certain sense. The representation map generated by the algorithm may be viewed as a linear discrete approximation to a continuous map that naturally arises from the geometry of the manifold [4]. The locality preserving character of the LPP algorithm makes it relatively insensitive to outliers and noise.

The new locality preserving linear maps are interesting from a number of perspectives.

1. The maps are designed to minimize a different objective criterion from the classical linear techniques. Unlike PCA which finds a projection in the directions of maximal variance, the LPP projects to optimally preserve the neighborhood structure of the data. If there is reason to believe that Euclidean distances ($||x_i - x_j||$) are meaningful only if they are small (local), then the LPP finds a projection that respects such a belief. As a result, the LPP makes particular sense when the data lies on a nonlinear manifold embedded in the ambient space. As the experiments in this paper suggest, the LPP yields linear maps whose properties are similar to the non linear maps yielded by the eigenvectors of the true graph Laplacian.

2. The locality preserving quality of LPP is likely to be of particular use in information retrieval applications. If one wishes to retrieve audio, video, text documents under a vector space model, then one will ultimately need to do a nearest neighbor search in the low dimensional space. For example, the popular latent semantic indexing (LSI) method projects the high dimensional data onto a low dimensional space obtained from a singular value decomposition (SVD; closely related to PCA) on the original data matrix. Since LPP is designed for preserving local structure, it is likely that a nearest neighbor search in the low dimensional space will yield similar results to that in the high dimensional space. This makes for an indexing scheme that would allow quick retrieval.

3. LPP is linear. This makes it fast and suitable for practical application. While a number of non linear techniques have properties (1) and (2) above, we know of no other linear projective technique that has such a property.

4. LPP is defined every where. Recall that nonlinear dimensionality reduction techniques like ISOMAP, LLE, Laplacian eigenmaps are defined only on the training data points and it is unclear how to evaluate the map for new test points. In contrast, the Locality Preserving Projection may be simply applied to any new data point to locate it in the reduced representation space.

As a result of all these features, we expect the LPP based techniques to be a natural alternative to PCA based techniques in exploratory data analysis, information retrieval, and pattern classification applications.

The rest of this paper is organized as follows: Section 2 describes the proposed Locality Preserving Projection algorithm. Section 3 presents a justification of this algorithm. Some computational issues in LPP are discussed in Section 4. Preliminary experimental results are shown in Section 5. Finally, we give concluding remarks and future work in Section 6.

2. Locality Preserving Projections

2.1. The linear dimensionality reduction problem

The generic problem of linear dimensionality reduction is the following. Given a set $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ in $\mathbf{R}^l$, find a transformation matrix A that maps these k points to a set of points $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k$ in $\mathbf{R}^m$ $(m \gg l)$, such that $\mathbf{y}_i$

”represents” $\mathbf{x}_i$, where $\mathbf{y}_i = A\mathbf{x}_i$.

Our method is of particular applicability in the special case where $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k \in \mathcal{M}$ and $\mathcal{M}$ is a manifold embedded in $\mathbf{R}^l$

2.2. The algorithm

Locality Preserving Projection (LPP) is a linear approximation of the nonlinear laplacian eigenmap [4]. The algorithmic procedure is formally stated below:

1. **Constructing the adjacency graph**: Let G denote a graph with k nodes. We put an edge between nodes i and j if $\mathbf{x}_i$ and $\mathbf{x}_j$ are ”close”. There are two variations:

   (a) ϵ-neighborhoods. [parameter $\epsilon \in \mathbf{R}$] Nodes i and j are connected by an edge if $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \epsilon$ where the norm is the usual Euclidean norm in $\mathbf{R}^l$.

   (b) n nearest neighbors. [parameter $n \in \mathbf{N}$] Nodes i and j are connected by an edge if i is among n nearest neighbors of j or j is among n nearest neighbors of i.

   Note: The method of constructing an adjacency graph outlined above is correct if the data actually lie on a low dimensional manifold. In general, however, one might take a more utilitarian perspective and construct an adjacency graph based on any principle (for example, perceptual similarity for natural signals, hyperlink structures for web documents, etc.). Once such an adjacency graph is obtained, LPP will try to optimally preserve it in choosing projections.

2. **Choosing the weights**: Here, as well, we have two variations for weighting the edges:

   (a) Heat kernel. [parameter $t \in \mathbf{R}$]. If nodes i and j are connected, put

   $$W_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}}$$

   The justification for this choice of weights can be traced back to [1].

   (b) Simple-minded. [No parameter]. $W_{ij} = 1$ if and only if vertices i and j are connected by an edge.

3. **Eigenmaps**: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem:

   $$XLX^T\mathbf{a} = \lambda XDX^T\mathbf{a}$$

   where D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W, $D_{ii} = \Sigma_j W_{ji}$. $L = D - W$ is the Laplacian matrix. The i^{th} column of matrix X is $\mathbf{x}_i$.

   Let $\mathbf{a}_0, \mathbf{a}_1, \cdots, \mathbf{a}_{k-1}$ be the solutions of equation (1), ordered according to their eigenvalues, $\lambda_0 < \lambda_1 < \cdots < \lambda_{k-1}$. Thus, the embedding is as follows:

   $$\mathbf{x}_i \to \mathbf{y}_i = A^T\mathbf{x}_i$$

   $$A = \begin{pmatrix} \mathbf{a}_0, & \mathbf{a}_1, & \cdots, & \mathbf{a}_m \end{pmatrix}$$

   where $\mathbf{y}_i$ is a m-dimensional vector, and A is a $l \times m$ matrix.

3. Justification

Locality Preserving Projection is fundamentally based on laplacian eigenmap [4]. Following the discussion in [4], in this section, we provide a justification of this linear dimensionality reduction algorithm.

3.1. Optimal Embeddings of Laplacian Eigenmap

The following section is based on the standard spectral graph theory. See [3] for a comprehensive reference and [4] for applications to data representation.

Recall that given a data set we construct a weighted graph $G = (V, E)$ with edges connecting nearby points to each other. Consider the problem of mapping the weighted graph G to a line so that connected points stay as close together as possible. Let $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T$ be such a map. A reasonable criterion for choosing a "good" map is to minimize the following objective function [4]

$$\sum_{ij} (y_i - y_j)^2 W_{ij}$$

under appropriate constraints. The objective function with our choice of weights W_{ij} incurs a heavy penalty if neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ are mapped far apart. Therefore, minimizing it is an attempt to ensure that if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "close" then y_i and y_j are close as well.

It is possible to show that this minimization problem reduces to finding

$$\arg\min_{\substack{\mathbf{y} \\ \mathbf{y}^T D\mathbf{y} = 1}} \mathbf{y}^T L \mathbf{y}$$

where $L = D - W$ is the Laplacian matrix. D is a diagonal weight matrix; its entries are column (or row, since W is symmetric) sums of W, $D_{ii} = \sum_j W_{ji}$. The constraint $\mathbf{y}^T D\mathbf{y} = 1$ removes an arbitrary scaling factor in the embedding. Matrix D provides a natural measure on the vertices of the graph. The bigger the value D_{ii} (corresponding to the i^{th} node) is, the more "important" is that node. Laplacian matrix is symmetric, positive semidefinite matrix which can be thought of as an operator on functions defined on vertices of G. The close relationship between the graph Laplacian and the Laplace Beltrami operator on the manifold is discussed in [4]. The vector $\mathbf{y}$ that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem,

$$L\mathbf{y} = \lambda D\mathbf{y} \tag{1}$$

3.2. Optimal Linear Embedding

One disadvantage of laplacian eigenmap is that it cannot produce a transformation function. That is, if a data point is not in the data set, then there is no way to map it into the dimensionality-reduced space. Moreover, the computational complexity of laplacian eigenmap is determined by the size of the data set. If the size is very large, then the nonlinear algorithm is computationally extensive. To overcome these problems, we propose Locality Preserving Projections, which yield linear approximations of laplacian eigenmaps.

Suppose $\mathbf{a}$ is a transformation vector of linear laplacian eigenmap, that is, $\mathbf{y}^T = \mathbf{a}^T X$, where the i^{th} column vector of X is $\mathbf{x}_i$. The minimization problem is as follows

$$\arg\min_{\substack{\mathbf{a} \\ (\mathbf{a}^T X) D (\mathbf{a}^T X)^T = 1}} (\mathbf{a}^T X) L (\mathbf{a}^T X)^T$$

We will now switch to a Lagrangian formulation of the problem. The Lagrangian is as follows

$$\mathcal{L} = (\mathbf{a}^T X) L (\mathbf{a}^T X)^T - \lambda (\mathbf{a}^T X) D (\mathbf{a}^T X)^T$$

Requiring that the gradient of $\mathcal{L}$ vanish gives the following eigenfunction

$$L'\mathbf{a} = \lambda D'\mathbf{a} \tag{2}$$

where $L' = XLX^T$, $D' = XDX^T$. It is easy to show that the matrices D' and L' are symmetric and positive semi-definite. The vectors $\mathbf{a}_i (i = 1, 2, \cdots, k)$ that minimize the objective function are given by the minimum eigenvalue solutions to the generalized eigenvalue problem.

Note that L' and D' are two $l \times l$ matrices, where l is the dimensionality of the original space. While in the nonlinear case, L and D are two $k \times k$ matrices, where k is the number of data points in the data set. In most real world applications, $k \gg l$. This property makes LPP much more computationally tractable than the nonlinear laplacian eigenmap.

4. Computational Issues of LPP

As desceibed in Section 3, we need to solve a generalized eigenfunction, $L'\mathbf{a} = \lambda D'\mathbf{a}$. If the matrix D' is inversable, then the generalized eigenfunction can be transformed to a ordinary eigenfunction, i.e. $D'^{-1}L'\mathbf{a} = \lambda\mathbf{a}$. However, the eigensystem computation of the ordinary eigenfunction could be unstable since the matrix $D'^{-1}L'$ need not be symmetric. To avoid this problem, the following method diagonalizes the symmetric matrix, and produces a stable eigensystem computation procedure.

By applying eigenvector decomposition, D' can be decomposed into the product of three matrices as follows:

$$D' = USU^T$$

where $S = diag\{\lambda_1, \lambda_2, \cdots, \lambda_l\}$ are the eigenvalues and $U = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_l]$ are the corresponding eigenvectors of D'. U is othonormal in that $U^TU = I$ and $UU^T = I$. Now, we can rewrite equation (2) as follows:

$$\begin{aligned} & L'\mathbf{a} = \lambda D'\mathbf{a} \\ \Longrightarrow \quad & L'\mathbf{a} = \lambda USU^T\mathbf{a} \\ \Longrightarrow \quad & U^TL'U(U^T\mathbf{a}) = \lambda S(U^T\mathbf{a}) \end{aligned}$$

If D' is non-singular, then each diagonal entry of S is greater than zero (D' is positive semi-definite, hence its eigenvalues are real numbers, and no less than zero). In this case, $S^{-1/2}$ exists and can be easily obtained from S. If D' is singular, we define $S^{-1/2}$ as follows:

$$(S^{-1/2})_{ii} = \begin{cases} S_{ii}^{-1/2}, & \text{if } S_{ii} > 0; \\ 0, & \text{otherwise.} \end{cases}$$

In fact, $S^{-1/2}$ is a generalized inverse of $S^{1/2}$. Now, we have the following equation:

$$S^{-1/2}U^TL'US^{-1/2}(S^{1/2}U^T\mathbf{a}) = \lambda S^{1/2}U^T\mathbf{a}$$

Let $W = S^{-1/2}U^TL'US^{-1/2}$ and $\mathbf{b} = S^{1/2}U^T\mathbf{a}$. Finally, equation (2) can be simplified to solve the following eigenvector problem:

$$W\mathbf{b} = \lambda\mathbf{b} \tag{3}$$

where W is a symmetric matrix. So we can first solve a symmetric eigenfunction (3) to obtain λ and $\mathbf{b}$, then compute $\mathbf{a}$ as $US^{-1/2}\mathbf{b}$. In this way, we have transformed the generalized eigenvector problem to a ordinary eigenvector problem of symmetric matrix, which could be computed more stably and efficiently.

5 Experimental Results

In this section, we will discuss several applications of the LPP algorithm. We begin with a 2-D data visualization example. Then, we apply the algorithm to face analysis.

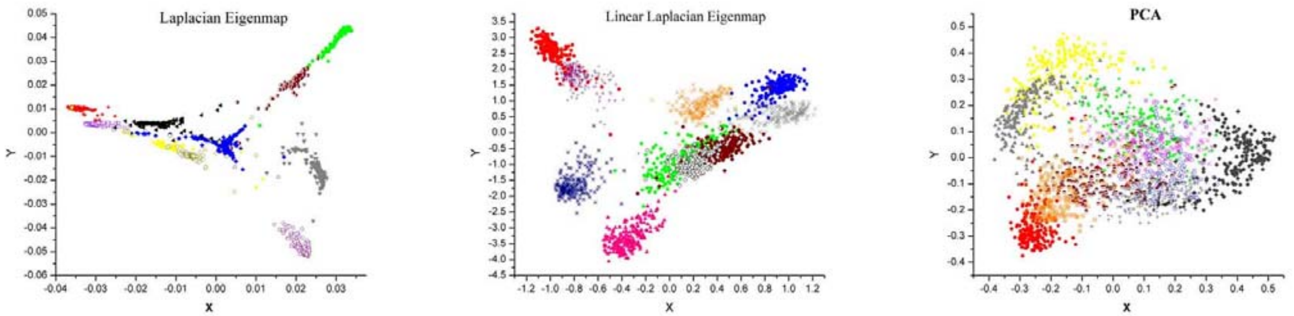

Figure 1: The handwritten digits ('0'-'9') are mapped into a 2-dimensional space. The left panel is a representation of the set of all images of digits using the Laplacian eigenmap. The middle panel shows the results of Locality Preserving Projection using the first two directions to represent the data. The right panel shows the results of principal component analysis. Each color corresponds to a digit.

5.1. 2-D Data Visulization

An experiment was conducted with the handwritten digit dataset [8]. The dataset contains 10 classes ('0'-'9'), and each class has 200 images. Each image is represented by a 649-dimensional vector. The data points are mapped to a 2-dimensional space using different dimensionality reduction algorithms. The experiment results are shown in Fig. 1. As can be seen, LPP performs much better than PCA. It has much more discriminating power than principal component analysis. The Locality Preserving Projection has almost the same performance as nonlinear Laplacian eigenmap, but it is computationally much more tractable.

5.2. Manifold of Face Images

Much research [5][6][7] has suggested that the human face images reside on a manifold embedded in the image space. In this subsection, we applied LPP to images of faces. The same face image dataset used in [5] is used for this experiment. Fig. 2 shows the mapping results. The images of faces are mapped into the 2-dimensional plane described by the first two coordinates of the Locality Preserving Projection. One note should be pointed out that the mapping from image space to low-dimensional space obtained by our method is linear, rather than nonlinear as in most previous work. The linear algorithm does detect the nonlinear manifold structure of images of faces to some extent. Some representative faces are shown next to the data points in different parts of the space. As can be seen, the images of faces are clearly divided into two parts. The left part are the faces with closed mouth, and the right part are the faces with open mouth. Moreover, the smooth changes in facial expression and viewing points of faces can be clearly seen from left to right, from top to bottom.

6 Conclusions

In this paper, we propose a new linear dimensionality reduction algorithm called Locality Preserving Projection. It is based on the same variational principle that gives rise to the Laplacian eigenmap [4]. As a result it has similar locality preserving properties.

Our approach also has a major advantage over recent nonparametric techniques for global nonlinear dimensionality reduction such as [4][5][6];its eigenproblem has a size which scales as the dimensionality of the data points rather than the number of data points. For massive datasets, this saving in run time and memory can be enormous.

Performance improvement of this method over Principal Component Analysis is demonstrated through several experiments. Though our method is still a linear algorithm, it is somehow capable of discovering the nonlinear structure of the data manifold.

References

[1] I. T. Jolliffe, *Principal Component Analysis,* Springer-Verlag, New York, 1989.

[2] T. Cox and M. Cox, *Multidimensional Scalling,* Chapman & Hall, London, 1994.

[3] Fan R. K. Chung, *Spectral Graph Theory,* Regional Conference Series in Mathematics, number 92, 1997.

[4] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering ," *Advances in Neural Information Processing Systems 15*, Vancouver, British Columbia, Canada, 2001.

[5] Sam T. Roweis, and Lawrence K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol 290, 22 December 2000.

[6] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol 290, 22 December 2000.

[7] H. Sebastian Seung and Daniel D. Lee, "The Manifold Ways of Perception," *Science*, vol 290, 22 December 2000.

[8] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases", http://www.ics.uci.edu/ mlearn/MLRepository.html. Irvine, CA, University of California, Department of Information and Computer Science, 1998.

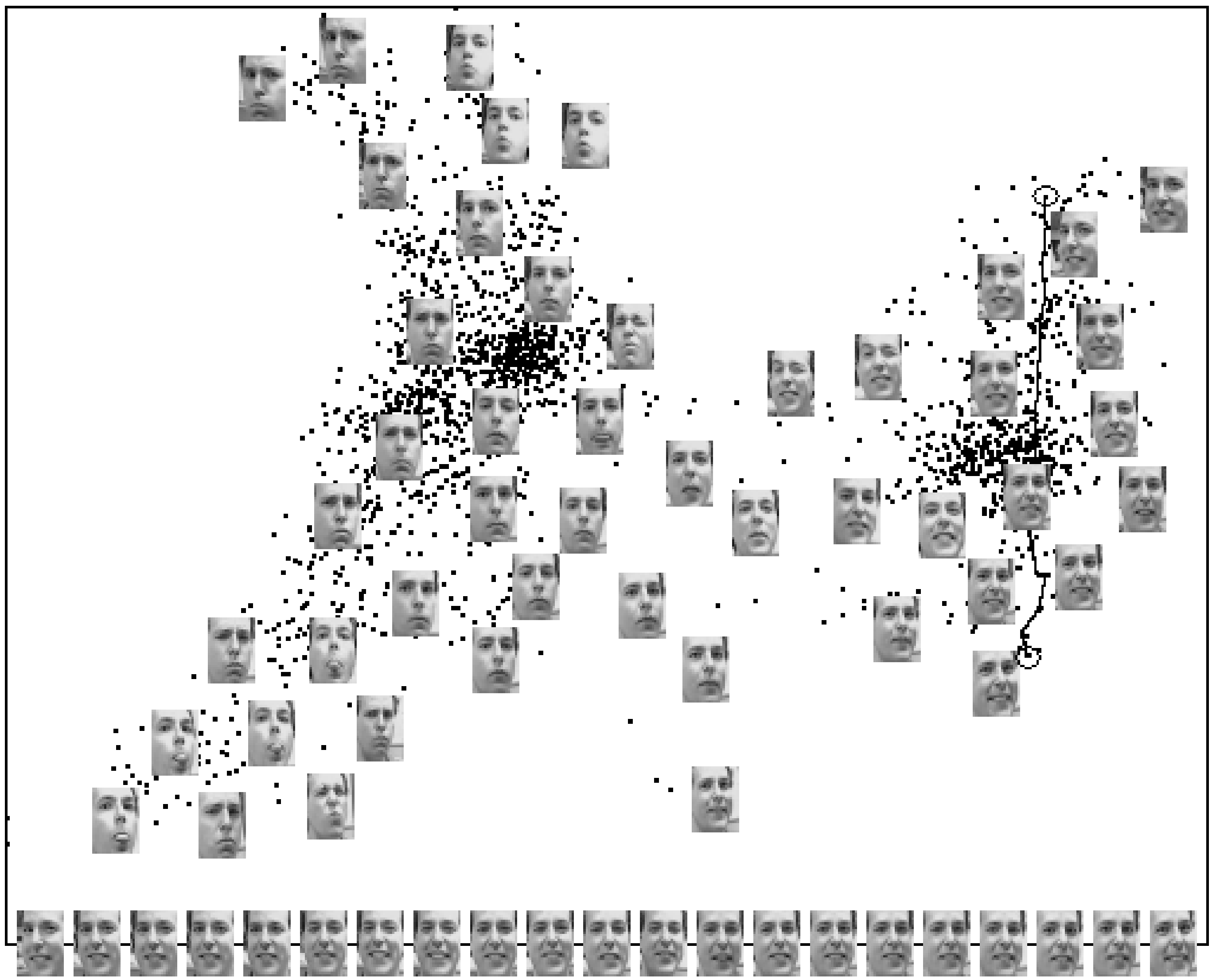

Figure 2: A two-dimensional representation of the set of all images of faces using the Locality Preserving Projection. Representative faces are shown next to the data points in different parts of the space. As can be seen, the facial expression and the viewing point of faces change smoothly.