

Local-information-based uncorrelated feature extraction

Haitao Zhao

Shanghai Jiaotong University
Institute of Aerospace Science & Technology
Shanghai, China

Shaoyuan Sun

Dong Hua University
Automation Department
Shanghai, China

Zhongliang Jing

Shanghai Jiaotong University
Institute of Aerospace Science & Technology
Shanghai, China

Abstract. In the past few years, the computer vision and pattern recognition community has witnessed the rapid growth of a new kind of feature extraction method, the manifold learning methods, which attempt to project the original data into a lower dimensional feature space by preserving the local neighborhood structure. Among them, locality preserving projection (LPP) is one of the most promising feature extraction techniques. Based on LPP, we propose a novel feature extraction method, called uncorrelated locality preserving projection (ULPP). We show that the extracted features via ULPP are statistically uncorrelated, which is desirable for many pattern analysis applications. We compare the proposed ULPP approach with LPP and principal component analysis (PCA) on the publicly available data sets, FERET and AR. Experimental results suggest that the proposed ULPP approach provides a better representation of the data and achieves much higher recognition accuracies. © 2006 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2163873]

Paper 050791L received Oct. 2, 2005; accepted for publication Dec. 5, 2005; published online Feb. 22, 2006.

1 Introduction

Recently, a new kind of feature extraction method, manifold learning algorithms, based on the local geometry analysis have drawn much attention. Unlike PCA, which aims to preserve the global Euclidean structure of the data space, manifold learning algorithms aim to preserve the inherent manifold structure. Locally linear embedding (LLE), Isomap, Laplacian eigenmap, and locality preserving projection (LPP)¹ are the most important manifold learning algorithms. All of these four algorithms attempt to embed the original data into a submanifold by preserving the local neighborhood structure. They are based on the assumption that the manifold's intrinsic geometry can be fully determined by the local metric and the neighborhood information. Different from LLE, Isomap, and Laplacian eigenmap, LPP is a linear algorithm. It is quite simple and easy to realize. LPP is originally derived by the discretely approximation of the Laplace Beltrami operator on compact Riemannian manifold. He et al.¹ have applied LPP on

face recognition problem. The application has shown the effectiveness of LPP in discovering the the intrinsic geometrical structure of the data.

Based on LPP, we propose a novel algorithm, called uncorrelated locality preserving projection (ULPP). It shares the same locality preserving character as LPP, but at the same time it requires the basis vectors to be statistically uncorrelated. Uncorrelated attributes are highly desirable in pattern analysis, because they contain minimum redundancy.² However, the basis vectors of LPP are statistically correlated, then the extracted features contain redundant information. Moreover, the overlapped information can distort the distribution of the features and even dramatically degrade the performance of LPP. In this paper, we overcome this problem by putting an uncorrelated constraint on the computation of the basis vectors. This makes the proposed ULPP algorithm much more robust.

2 Locality Preserving Projection

LPP¹ can be explained by spectral graph theory. Let us assume that the n -dimensional data set $\{x_1, x_2, \dots, x_N\}$ is distributed on a low-dimensional submanifold. We desire to find a set of d -dimensional ($d < n$) data points $\{y_1, y_2, \dots, y_N\}$ with the same local neighborhood structure as the data $\{x_1, x_2, \dots, x_N\}$. In order to do so, we can construct a weighted graph $G=(V, E, W)$, where V is the set of all points; E is the set of edges connecting the points; and $W=(w_{ij})$ is a similarity matrix with weights characterizing the likelihood of two points. The criterion function of LPP is as follows:

$$\min \sum_{ij} \|y_i - y_j\|^2 w_{ij}. \tag{1}$$

To make the mapping (embedding) not defined only for the original data set, LPP defines a transformation. That is, $y_i = \Phi^T x_i$, where $\Phi = [\phi_1, \phi_2, \dots, \phi_d] \in \mathcal{R}^{n \times d}$. Following some algebraic steps, we see that

$$\begin{aligned} \frac{1}{2} \sum_{ij} \|y_i - y_j\|^2 w_{ij} &= \frac{1}{2} \sum_{ij} (\Phi^T x_i - \Phi^T x_j)^T (\Phi^T x_i - \Phi^T x_j) w_{ij} \\ &= \sum_{ki} \phi_k^T x_i D_{ii} x_i^T \phi_k - \sum_{kij} \phi_k^T \text{Diag}(x_i, \dots, x_i) \\ &\quad \times w_{ij} \text{Diag}(x_j^T, \dots, x_j^T) \phi_k \\ &= \text{trace}(\Phi^T X(D - W)X^T \Phi), \end{aligned} \tag{2}$$

where $X = [x_1, x_2, \dots, x_N]$, and $D = \text{diag}(D_{ii})$; its entries are row (or column since W is symmetric) sums of W , $D_{ii} = \sum_j w_{ij}$; $\text{Diag}(x, \dots, x)$ returns a block diagonal matrix with block entry x . $L = D - W$ is the Laplacian matrix.¹

In order to remove the arbitrary scaling factor in the embedding, LPP imposes a constraint as follows:

$$YDY^T = I \Rightarrow \Phi^T XDX^T \Phi = I. \tag{3}$$

This constraint sets the mapping (embedding) scale and makes the vertices with high similarities to be mapped nearer to the origin. Finally, the minimization problem reduces to:

$$\operatorname{argmin}_{\Phi} \operatorname{trace}(\Phi^T X L X^T \Phi). \quad (4)$$

$$\Phi^T X D X^T \Phi = I$$

The transformation matrix Φ that minimizes the objective function can be obtained by solving the generalized eigenvalue problem:

$$X L X^T \phi = \lambda X D X^T \phi. \quad (5)$$

Note that the Laplacian matrix can be viewed as the discretely approximation of the Laplace Beltrami operator on compact Riemannian manifold.¹

Based on the matrix $\Phi = [\phi_1, \dots, \phi_d]$, for any instance $\mathbf{u} = [u_1, u_2, \dots, u_n]^T \in \mathcal{R}^n$, we can define the following linear transform from \mathcal{R}^n to \mathcal{R}^d :

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix} = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_d^T \end{bmatrix} \mathbf{u}. \quad (6)$$

Equation (6) with the constraint (3) forms the LPP transformation. It is easy to obtain the following theorem.

Theorem 1: Any two features v_i and v_j ($j \neq i$) in Eq. (6) are correlated.

Proof: Let the covariance matrix of the data $\{x_i\}_{i=1}^N$ be Ξ . It is easy to obtain the following equation:

$$E[(v_i - E v_i)(v_j - E v_j)] = \phi_i^T \Xi \phi_j. \quad (7)$$

With the constraint (3), we have inequality, in general,

$$\phi_i^T \Xi \phi_j \neq 0 \quad (j \neq i). \quad (8)$$

Therefore, we cannot obtain the following equation in general:

$$E[(v_i - E v_i)(v_j - E v_j)] = 0, \quad (9)$$

thus completing the proof.

3 Uncorrelated Locality Preserving Projection

The algorithmic procedure of uncorrelated locality preserving projection (ULPP) is stated below:

1. Perform PCA projection. We project the data set into a PCA subspace to make the matrix $X D X^T$ become nonsingular. Let the transformation matrix of PCA be Φ_{PCA} .
2. Construct the similarity of any two data points. For each sample x_i , if x_j is among the m -nearest neighbors of x_i , set the similarity $w_{ij} = w_{ji} = x_i^T x_j$; otherwise, set the similarity $w_{ij} = w_{ji} = 0$. The similarity reflects the local structure of the data space.
3. Obtain the diagonal matrix D and the Laplacian matrix L .
4. Compute the uncorrelated locality preserving projections. Let $S_L = X L X^T$, $S_D = X D X^T$ and $I = \operatorname{diag}(1, 1, \dots, 1)$ for concision. Let $\{\phi_1, \phi_2, \dots, \phi_k\}$ be the uncorrelated locality preserving projections, thus we define

$$\Phi^{(k-1)} = [\phi_1, \phi_2, \dots, \phi_{k-1}]. \quad (10)$$

The uncorrelated locality preserving projections $\{\phi_1, \phi_2, \dots, \phi_k\}$ can be iteratively computed as follows:

- (a) Compute ϕ_1 as the eigenvector of $S_D^{-1} S_L$ associated with the smallest eigenvalue.
- (b) Compute ϕ_k as the eigenvector associated with the smallest eigenvalue of the eigenfunction

$$P^{(k)} S_L \phi = \lambda S_D \phi, \quad (11)$$

where

$$P^{(k)} = I - \Xi (\Phi^{(k-1)})^T (\Phi^{(k-1)} \Xi S_D^{-1} \Xi (\Phi^{(k-1)})^T)^{-1} \times \Phi^{(k-1)} \Xi S_D^{-1}. \quad (12)$$

5. Perform the ULPP transformation. Let $\Phi^* = [\phi_1, \phi_2, \dots, \phi_d]$. The embedding is as follow:

$$\mathbf{v} = \Phi_{ULPP}^T \mathbf{u}, \quad (13)$$

where \mathbf{v} is a d -dimensional representation of \mathbf{u} and $\Phi_{ULPP} = \Phi_{PCA} \Phi^*$.

3.1 Justification

In this subsection, we provide the justification of our algorithm.

According to the following theorem, we can compute the k -th uncorrelated direction ϕ_k as the eigenvector of $P^{(k)}$ [Eq. (12)] associated with the the smallest eigenvalue of $P^{(k)}$.

Theorem 2: The k -th uncorrelated direction ϕ_k is the eigenvector corresponding to the smallest eigenvalue of the following eigenfunction:

$$P^{(k)} S_L \phi = \lambda S_D \phi,$$

where

$$P^{(k)} = I - \Xi (\Phi^{(k-1)})^T (\Phi^{(k-1)} \Xi S_D^{-1} \Xi (\Phi^{(k-1)})^T)^{-1} \Phi^{(k-1)} \Xi S_D^{-1}.$$

Proof: Recall the minimization problem of LPP:

$$\operatorname{argmin}_{\Phi} \operatorname{trace}(\Phi^T X L X^T \Phi).$$

$$\Phi^T X D X^T \Phi = I$$

In order to compute the k -th uncorrelated direction ϕ_k , we add the constraints

$$\phi_k \Xi \phi_i = 0 \quad (i = 1, 2, \dots, k-1).$$

We can use the method of Lagrange multipliers to transform the minimization function to include all constraints. Recall $S_L = X L X^T$ and $S_D = X D X^T$, we have

$$L(\phi_k) = \phi_k^T S_L \phi_k - \lambda (\phi_k^T S_D \phi_k - 1) - \sum_{i=1}^{k-1} \gamma_i \phi_k^T \Xi \phi_i. \quad (14)$$

The minimization is performed by setting the partial derivative of $L(\phi_k)$ with respect to ϕ_k equal to zero:

Table 1 Recognition accuracies (%) comparison of FERET data set.

	Eigenface	LPP	ULPP
Rank 1	58.61	72.64	85.90
Rank 2	63.70	78.22	89.49
Rank 3	68.22	81.39	91.07

$$2S_L\phi_k - 2\lambda S_D\phi_k - \sum_{i=1}^{k-1} \gamma_i \Xi \phi_i = 0. \quad (15)$$

Multiplying the left side of Eq. (15) by ϕ_k , we have

$$\lambda = \frac{\phi_k^T S_L \phi_k}{\phi_k^T S_D \phi_k}. \quad (16)$$

Thus, λ represents minimization problem to be optimized.

Multiplying the left-hand side of Eq. (15), respectively, by $\phi_j^T \Xi S_D^{-1}$ ($j=1, 2, \dots, k-1$), we have

$$2 \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_{k-1}^T \end{bmatrix} \Xi S_D^{-1} S_L \phi_k - \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_{k-1}^T \end{bmatrix} \Xi S_D^{-1} \Xi \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{k-1} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{k-1} \end{bmatrix} = 0. \quad (17)$$

Let $\Gamma^{(k)} = [\gamma_1, \gamma_2, \dots, \gamma_{k-1}]^T$, thus we can obtain

$$\Gamma^{(k)} = 2(\Phi^{(k)} \Xi S_D^{-1} \Xi (\Phi^{(k)})^T)^{-1} \Phi^{(k)} \Xi S_D^{-1} S_L \phi_k. \quad (18)$$

Since Eq. (15) can be written as

$$2S_L\phi_k - 2\lambda S_D\phi_k - \Xi (\Phi^{(k-1)})^T \Gamma^{(k)} = 0, \quad (19)$$

substituting Eq. (18) into Eq. (19), we have

$$S_L\phi_k - \lambda S_D\phi_k - \Xi (\Phi^{(k-1)})^T (\Phi^{(k)} \Xi S_D^{-1} \Xi (\Phi^{(k)})^T)^{-1} \Phi^{(k)} \Xi S_D^{-1} S_L \phi_k = 0, \quad (20)$$

or in another form

$$[I - \Xi (\Phi^{(k-1)})^T (\Phi^{(k-1)} \Xi S_D^{-1} \Xi (\Phi^{(k-1)})^T)^{-1} \Phi^{(k-1)} \Xi S_D^{-1}] S_L \phi_k = \lambda S_D \phi_k, \quad (21)$$

thus completing the proof.

4 Experiment Results

4.1 Experiments on FERET Data Set

FERET³ is a very popular database. Face image variations in the FERET database include pose, illumination, facial expression, and aging. In our experiment, we use a subset of the FERET data set. The data set we selected contains 72 people and 6 images for each individual. All images are aligned by the centers of the eyes and mouth and then scaled with resolution 112×92 . We randomly select 3 images from each person for training, while the rest of the images of each individual are selected for testing. The ex-

Table 2 Recognition accuracies (%) comparison on AR data set.

	Eigenface	LPP	ULPP
Rank 1	70.58	72.42	80.32
Rank 2	79.07	77.96	84.44
Rank 3	83.40	80.43	86.51

periments are repeated 20 times and the average accuracies of rank 1 to rank 3 are recorded and shown in Table 1. The results show ULPP gives the best performance. The recognition accuracy of ULPP increases from 85.90% with rank 1 to 91.07% with rank 3, which is about 10% higher than the recognition accuracy of LPP.

4.2 Experiments on AR Data Set

AR⁴ is also a popular database. Face image variations in the AR database include illumination, facial expression, and occlusion. In our experiment, 952 images (119 individuals, 8 images per person), which involve large variations in facial expression, are selected. We manually cropped the face portion of the images. All the cropped images are aligned at the centers of eyes and mouth and then normalized with resolution 112×92 . We randomly select 4 images from each person for training, while the rest of the images of each individual are selected for testing. The experiments are repeated 30 times and the average accuracies of rank 1 to rank 3 are recorded and shown in Table 2. From Table 2, it can be seen that ULPP outperforms the other two methods. The recognition accuracy of ULPP increases from 80.32% with rank 1 to 86.51% with rank 3, which is about 5% higher than the recognition accuracy of LPP.

5 Conclusion

This paper presents a novel feature extraction method, called uncorrelated locality preserving projection (ULPP). ULPP can extract uncorrelated features which contain minimum redundancy and are highly desirable for many applications. In this paper, we present a theoretical study on ULPP and provide an implementation of it. The experiments on face image data show the superiority of ULPP over the other both PCA and LPP.

Acknowledgment

This project is partially supported by National Natural Science Foundation of China (60502042) and 2010 Shanghai EXPO Special Project (2004BA908B07, 04DZ05807).

References

1. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005).
2. J. Ye, R. Janardan, Q. Li, and H. Park, "Feature extraction via generalized uncorrelated linear discriminant analysis," *Proc. 21st Int. Conf. Mach. Learning*, Banff, Alberta, Canada, 2004.
3. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000).
4. A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report #24* (June 1998).