"Non-locality" Preserving Projection and Its Application to Palmprint Recognition

Jian Yang, David Zhang
Department of Computing,
Hong Kong Polytechnic University, Kowloon,
Hong Kong
{csjyang, csdzhang}@comp.polyu.edu.hk

Jing-yu Yang
Department of Computer Science,
Nanjing University of Science and Technology,
Nanjing 210094, P. R. China
yangjy@mail.njust.edu.cn

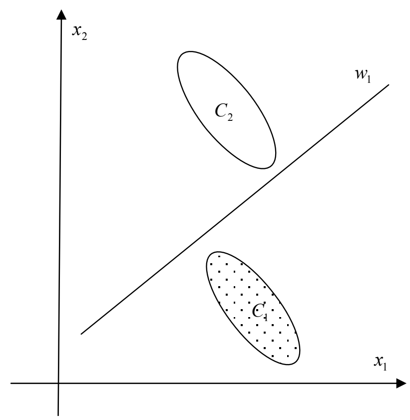
***Abstract*—This paper develops a "Non-locality" Preserving Projection (NLPP) technique for feature extraction. In contrast to the existing Locality Preserving Projection (LPP), a technique based on the characterization of the local scatter, NLPP is a method based on the characterization of the non-local scatter. Intuitively, NLPP should be more effective than LPP when the non-local information plan a dominant role in discrimination. NLPP is tested using the PolyU palmprint database and the experimental results show that NLPP outperforms PCA, LDA and LPP.**

Keywords—feature extraction, manifold learning, subspace learning, biometrics, palmprint recognition

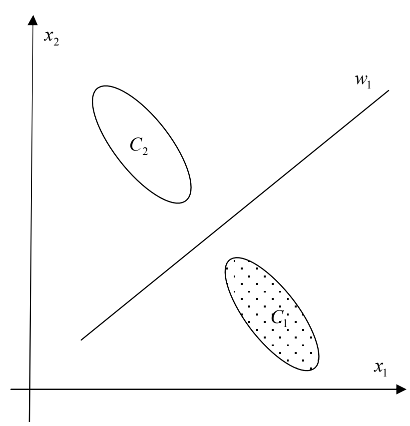
I. INTRODUCTION

Recently, He et al [1, 2] proposed a method called Locality Preserving Projections (LPP) and applied it to face recognition. LPP is a linear subspace method derived from Laplacian Eigenmap [3]. It results in a linear map that optimally preserves local neighborhood information in a certain sense. This map can be viewed as a linear discrete approximation to a continuous map that naturally arises from the geometry of the manifold [1]. Therefore, it can be said that He et al's method built a bridge from manifold learning to subspace learning. In contrast with most manifold learning algorithms, a remarkable advantage of LPP is that it can generate a simple and efficiently-computable linear map, like that of PCA or LDA. This map is also effective, yielding encouraging results on face recognition tasks.

LPP is modeled based on the characterization of "locality". The objective function of LPP is to minimize the local quantity, i.e., the local scatter of the projected data. This criterion, however, cannot guarantee to yield good projections for classification in some cases. Figure 1 (a) and (b) gives two cases where two clusters of samples are uniformly distributed in two ellipses C_1 and C_2. If the locality radius is set as the length of the semi-major axis of the larger ellipse, the direction w_1 is optimal according to the criterion of LPP, since after all samples being projected onto w_1, the local scatter is minimal. In Case (a), w_1 is a suitable projection direction for classification, but in Case (b), w_1 is unsuitable since the projected samples overlap on this direction. It can be seen that in both cases, the non-local quantity, i.e., the inter-cluster scatter, provides important information for discrimination. In this paper, we will explore ways to characterize the "non-locality" and then to utilize the resulting non-local information for classification purpose.



(a)



(b)

Figure 1. Illustration of two clusters of samples in two-dimensional space and the projection directions

II. FUNDAMENTALS

A. Characterization of the Local Scatter

Recall that in PCA, in order to preserve the global geometric structure of data in a transformed low-dimensional space, the global scatter of samples is considered. Instead, if we aim to discover the local structure of data, the local scatter (or *intra-locality scatter*) of samples should be considered. The local scatter can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that are within any local δ -neighborhood ($\delta > 0$). Specifically, two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ are viewed within a local δ -neighborhood provided that $\| \mathbf{x}_i - \mathbf{x}_j \|^2 < \delta$. Let us denote the set $\mathrm{U}^{\delta} = \{(i,j) \big| \| \mathbf{x}_i - \mathbf{x}_j \|^2 < \delta\}$. After the projection of $\mathbf{x}_i$ and $\mathbf{x}_j$ onto a direction $\mathbf{w}$, we get their images y_i and y_j. The local scatter of is then defined by

$$J_L(\mathbf{w}) \triangleq \frac{1}{2}\frac{1}{M_L}\sum_{(i,j)\in \mathrm{U}^{\delta}}(y_i - y_j)^2$$
$$\propto \frac{1}{2}\sum_{(i,j)\in \mathrm{U}^{\delta}}(y_i - y_j)^2 \qquad (1)$$

where M_L is the number of sample pairs satisfying $\| \mathbf{x}_i - \mathbf{x}_j \|^2 < \delta$.

Let us define the adjacency matrix **H**, whose element is given below:

$$H_{ij} = \begin{cases} 1, & \| \mathbf{x}_i - \mathbf{x}_j \|^2 < \delta \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

It is obvious that the adjacency matrix **H** is a symmetric matrix.

By virtue of the adjacency matrix **H**, Eq. (1) can be rewritten by

$$J_L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}(y_i - y_j)^2 \qquad (3)$$

where M is the number of training samples. It follows from Eq. (3) that

$$J_L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}(\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}_j)^2$$
$$= \mathbf{w}^T\left[\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T\right]\mathbf{w}$$
$$= \mathbf{w}^T\mathbf{S}_L\mathbf{w}, \qquad (4)$$

where

$$\mathbf{S}_L = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T. \qquad (5)$$

$\mathbf{S}_L$ is called the local scatter (covariance) matrix.

Due to the symmetry of **H**, it follows that

$$\mathbf{S}_L = \frac{1}{2}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}\mathbf{x}_i\mathbf{x}_i^T + \sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}\mathbf{x}_j\mathbf{x}_j^T - 2\sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}\mathbf{x}_i\mathbf{x}_j^T\right)$$
$$= \left(\sum_{i=1}^{M} D_{ii}\mathbf{x}_i\mathbf{x}_i^T - \sum_{i=1}^{M}\sum_{j=1}^{M} H_{ij}\mathbf{x}_i\mathbf{x}_j^T\right)$$
$$= (\mathbf{X}\mathbf{D}\mathbf{X}^T - \mathbf{X}\mathbf{H}\mathbf{X}^T)$$
$$= \mathbf{X}\mathbf{L}\mathbf{X}^T, \qquad (6)$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M)$, and **D** is a diagonal matrix whose elements on diagonal are column (or row since **H** is a symmetric matrix) sum of **H**, i.e., $D_{ii} = \sum_{j=1}^{M} H_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{H}$ is called Laplacian matrix in [1-3].

It is obvious that **L** and $\mathbf{S}_L$ are both real symmetric matrices. From Eqs. (4) and (6), we know that $\mathbf{w}^T\mathbf{S}_L\mathbf{w} \geq 0$ for any nonzero vector $\mathbf{w}$. So, the local scatter matrix $\mathbf{S}_L$ must be non-negative definite.

In the above discussion, we use δ -neighborhoods to characterize the "locality" and the local scatter. This way is geometrically intuitive but unpopular because it is hard to choose a proper neighborhood radius δ in practice. To void the difficulty, the method of K-nearest neighbors is always used instead in real-world applications. The K-nearest neighbors method can determine the following adjacency matrix **H**, with elements given by:

$$H_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_j \text{ is among K nearest nieghbors of } \mathbf{x}_i \\ & \textit{and } \mathbf{x}_i \text{ is among K nearest nieghbors of } \mathbf{x}_j \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

The local scatter can be characterized similarly by K-nearest neighbor adjacency matrix if Eq. (2) is replaced by Eq. (7).

B. Characterization of the Non-local Scatter

In contrast to the characterization of the local scatter, the non-local scatter (i.e., the inter-locality scatter) can be characterized by the mean square of the Euclidean distance between any pair of the projected sample points that are outside any local δ -neighborhood ($\delta > 0$).

Let us denote the set $\mathrm{U}_N^{\delta} = \{(i,j) \big| \| \mathbf{x}_i - \mathbf{x}_j \|^2 \geq \delta\}$. The non-local scatter is defined by

$$J_N(\mathbf{w}) \triangleq \frac{1}{2}\frac{1}{M_N}\sum_{(i,j)\in \mathrm{U}_N^{\delta}}(y_i - y_j)^2$$
$$\propto \frac{1}{2}\sum_{(i,j)\in \mathrm{U}_N^{\delta}}(y_i - y_j)^2 \qquad (8)$$

where M_N is the number of elements in U_N^{δ}.

By virtue of the adjacency matrix **H** in Eq. (2) or (7), the non-local scatter can be rewritten by

$$J_N(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-H_{ij})(y_i - y_j)^2 \quad (9)$$

It follows from Eq. (9) that

$$J_N(\mathbf{w}) = \mathbf{w}^T\left[\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-H_{ij})(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T\right]\mathbf{w}$$

$$= \mathbf{w}^T\mathbf{S}_N\mathbf{w}, \quad (10)$$

where

$$\mathbf{S}_N = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}(1-H_{ij})(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T \quad (11)$$

$\mathbf{S}_N$ is called the non-local scatter (covariance) matrix. It is easy to show $\mathbf{S}_N$ is also a non-negative definite matrix.

Let us define the matrix $\mathbf{H}_N = (1-H_{ij})_{M\times M}$. Similar to the derivation of Eq. (6), we have

$$\mathbf{S}_N = \mathbf{X}\mathbf{L}_N\mathbf{X}^T \quad (12)$$

where $\mathbf{L}_N = \mathbf{D}_N - \mathbf{H}_N$, $\mathbf{D}_N$ is a diagonal matrix whose elements on diagonal are column (or row) sum of $\mathbf{H}_N$, i.e., $(\mathbf{D}_N)_{ii} = \sum_{j=1}^{M}(\mathbf{H}_N)_{ij}$.

III. Methods

A. Locality Preserving Projection (LPP)

Let us define the matrix $\mathbf{S}_D = \mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}$. LPP seeks to *minimize* the local scatter under the condition that the projection axes are $\mathbf{S}_D$-orthogonal. The optimization model of LPP is given by [1, 2]

$$\arg\min_{\mathbf{w}^{\mathrm{T}}\mathbf{S}_D\mathbf{w}=1} J_L(\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{S}_L\mathbf{w}, \quad (13)$$

which is equivalent to

$$\arg\min J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_L\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_D\mathbf{w}} \quad (14)$$

If the local scatter matrix $\mathbf{S}_D$ is non-singular, the criterion in Eq. (14) can be minimized directly by calculating the generalized eigenvectors of the following generalized eigen-equation:

$$\mathbf{S}_L\mathbf{w} = \lambda\mathbf{S}_D\mathbf{w} \quad (15)$$

The projection axes of LPP can be selected as the generalized eigenvectors $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_d$ of $\mathbf{S}_L\mathbf{w} = \lambda\mathbf{S}_D\mathbf{w}$ corresponding to *d smallest* positive eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_d$.

Otherwise, we can perform PCA first and make $\mathbf{S}_D$ non-singular in the PCA-transformed space. Then, LPP can be performed based on PCA-transformed data.

B. "Non-locality" Preserving Projection (NLPP)

Let us redefine the matrix $\mathbf{S}_D = \mathbf{X}\mathbf{D}_N\mathbf{X}^{\mathrm{T}}$. NLPP seeks to *maximize* the non-local scatter under the condition that the projection axes are $\mathbf{S}_D$-orthogonal. The optimization model of NLPP is

$$\arg\max_{\mathbf{w}^{\mathrm{T}}\mathbf{S}_D\mathbf{w}=1} J_N(\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{S}_N\mathbf{w}, \quad (16)$$

which is equivalent to

$$\arg\max J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_N\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_D\mathbf{w}} \quad (17)$$

If the local scatter matrix $\mathbf{S}_D$ is non-singular, the criterion in Eq. (14) can be maximized directly by calculating the generalized eigenvectors of the following generalized eigen-equation:

$$\mathbf{S}_N\mathbf{w} = \lambda\mathbf{S}_D\mathbf{w} \quad (18)$$

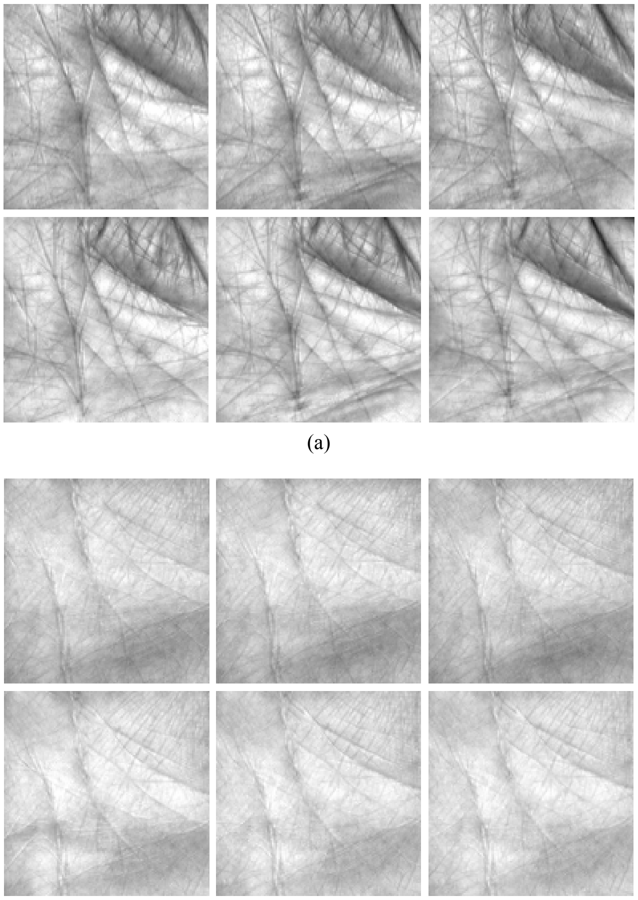
The projection axes of NLPP can be selected as the generalized eigenvectors $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_d$ of $\mathbf{S}_N\mathbf{w} = \lambda\mathbf{S}_D\mathbf{w}$ corresponding to *d largest* positive eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d$.

Otherwise, similar to the operation in LPP, PCA is first used for dimension reduction and then NLPP is performed in the PCA-transformed space.

IV. Experiments

The proposed method is tested using the PolyU palmprint database which contains 600 grayscale images of 100 different palms with six samples for each palm [4]. Six samples from each of these palms were collected in two sessions, where the first three were captured in the first session and the other three in the second session. The average interval between the first and the second sessions is two months. In our experiments, the central part of each original image was automatically cropped. The cropped images was resized to 128×128 pixels and pre-processed by histogram equalization. Figure 2 shows some sample images of two palms.

According to the protocol of this database, the images captured in the first session are used for training and the images captured in the second session for test. Thus, for each palm class, there are three training samples and three testing samples. PCA [5], LDA [6], LPP and NLPP are, respectively, used for palm feature extraction. Note that LDA, LPP and NLPP all involve a PCA phase due to the singularity issues. In this phase, the number of principal components, m, is set as 120. The K-nearest neighborhood parameter K in LPP and NLPP is chosen as K = 2. After feature extraction, a nearest neighbour classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimension are listed in Table 1. This table shows that NLPP outperforms PCA, LDA and LPP. The recognition rate of NLPP is up to 99.7%, i.e., there is only one sample missed.



(a)

(b)

Figure 2. Samples of the cropped images in PolyU Palmprint database

TABLE I. THE MAXIMAL RECOGNITION RATES (%) OF PCA, LDA, LPP AND NLPP ON POLYU PALMPRINT DATABASE AND THE CORRESPONDING DIMENSIONS

Method	PCA	LDA	LPP	NLPP
Recognition rate	86.0	97.7	98.7	99.7
Dimension	90	95	95	100

V. CONCLUSION

In this paper, we propose a "non-locality" preserving projection technique (NLPP) and demonstrate its effectiveness for feature extraction using the PolyU palmprint database. Actually, the "locality" and "non-locality" can be characterized under a unified framework, e.g. maximizing the non-local scatter and minimizing the local scatter at the same time. We will perform this work and publish the results in another paper.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation of China under Grants No. 60503026, No. 60332010, No. 60472060, and No. 60473039, and the CERG fund from the HKSAR Government and the central fund from the Hong Kong Polytechnic University.

REFERENCES

[1] X. He, P. Niyogi, "Locality Preserving Projections", Neural Information Processing Systems 16 (NIPS'2003).

[2] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, "Face Recognition Using Laplacianfaces", IEEE Trans. Pattern Anal. Machine Intell., 2005, 27(3), pp. 328- 340.

[3] M. Belkin, P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation", Neural Computation, 2003, 15 (6):1373-1396.

[4] D. Zhang, "PolyU palmprint database", http://www4.comp.polyu.edu.hk/~biometrics/

[5] M. Turk and A. Pentland, "Eigenfaces for recognition", J. Cognitive Neuroscience, 1991, 3(1), pp.71-86.

[6] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval", IEEE Trans. Pattern Anal. Machine Intell., 1996,18(8), pp. 831-836.