

Orthogonal Neighborhood Preserving Embedding for Face Recognition

Xiaoming Liu¹, Jianwei Yin^{1*}, Zhilin Feng², Jinxiang Dong¹, Lu Wang³

¹Department of Computer Science and Technology, Zhejiang University, China

²Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, China

³Department of Automation, Tsinghua University, Beijing, 100084, P.R.China

Email: {liuxiaoming, zjuyjw}@cs.zju.edu.cn

*Corresponding author: Jianwei Yin, Email: zjuyjw@cs.zju.edu.cn

ABSTRACT

In this paper, we propose a new algorithm called Orthogonal Neighborhood Preserving Embedding (ONPE) for face recognition. ONPE can preserve local geometry information and is based on the local linearity assumption that each data point and its k nearest neighbors lie on a linear manifold locally embedded in the image space. ONPE is based on Neighborhood Preserving Embedding (NPE), but overcomes the metric distortion problem of NPE, while metric distortion usually leads to performance degradation. Besides, we propose a classification method (ONPC) based on the ONPE, which use local label propagation method in the reduced space for face recognition. ONPC is based on the natural assumption that the local neighborhood information is also preserved in reduced space, and the label of a data point can be obtained in the reduced space by the labels of its neighbors. Experimental results on two face databases demonstrate the effectiveness of our proposed method.

Index Terms—Face Recognition, Manifold Learning, Eigenface, Neighborhood preserving embedding

1. INTRODUCTION

Face recognition is one of the most challenging problems in computer vision and pattern recognition. Numerous methods have been proposed for face recognition over the past few decades. Among these methods, Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] are the most popular techniques, which assume that the samples lie on a linearly embedded manifold. However, a lot of research has shown that facial images possibly lie on a nonlinear submanifold [3, 4]. When using PCA and LDA for dimensionality reduction, they will fail to discover the intrinsic dimension of the image space. Recently, a number of manifold learning methods are proposed to discover the nonlinear structure of the manifold by investigating the local geometry of the samples, such as LLE [3], Isomap [4], LTSA [5, 6].

Neighborhood Preserving Embedding (NPE) [7] is a

recently proposed linear dimensionality reduction algorithm. NPE aims to preserving the local manifold structure, unlike PCA, which aims to preserving the global Euclidean structure. However, different from PCA, NPE is non-orthogonal. Therefore, it can not preserve the metric structure of the high-dimensional space and suffers from the problem of dimensionality estimation [10].

In this paper, we propose a new algorithm called Orthogonal Neighborhood Preserving Embedding. Orthogonal NPE is fundamentally based on NPE. It shares the same neighborhood preserving character as NPE, but at the same time it requires the basis functions to be orthogonal. Orthogonal basis functions preserve the metric structure of the original high dimensional space. Moreover, with the ONPE projection, we proposed to utilize a LNP (Local Neighborhood labels Propagation) method [8] for classification (ONPC) on the reduced space, naturally extends the assumption of LNP. ONPC is based on the reasonable assumption that local neighborhood information is preserved in the reduced space and the label information of a point can be obtained from neighbor points. Note that original LNP is applied in original high-dimensional space, and is inefficient for large data set. The original NPE utilize KNN (K-nearest neighbors) for classification, and is not optimal due to its ignoring of the local geometry information.

The rest of the paper is organized as follows: the Orthogonal Neighborhood Preserving Embedding (ONPE) algorithm and the classification extension (ONPC) are described in Section 2. Experimental results on face datasets are shown in Section 3. Finally, conclusions are summarized in Section 4.

2. ORTHOGONAL NEIGHBORHOOD PRESERVING EMBEDDING

2.1. ONPE algorithm

First of all, we would like to point out the non-orthogonality property of NPE. Recall that in NPE [7], the basis vectors of NPE is the first k eigenvectors associated with the smallest eigenvalues of the eigen-problem

$$XX^T \mathbf{b} = \lambda XX^T \mathbf{b} \quad (1)$$

The basis vectors satisfy the following constraints:

$$\mathbf{b}_i^T XX^T \mathbf{b}_j = 0 \quad (i \neq j) \quad (2)$$

The transformation of NPE is non-orthogonal. Actually, it is XX^T orthogonal.

The algorithm procedure of ONPE is stated below:

In face recognition, usually the number of feature dimensions (D) is much larger than the number of samples (n). Then, the $D \times D$ matrix XX^T is singular. To overcome this problem, we can apply PCA to project the x_i into a subspace without losing information and the matrix XX^T becomes non-singular.

Preprocess-PCA projection (Optional): Project the data point x_i into the PCA subspace by throwing away the components corresponding to zero eigenvalues. Denote the transform matrix of PCA by A_{PCA} . For the simple of explanation, we will use x_i to denote the data after PCA projection also. If the preprocess is not applied, let A_{PCA} denotes the identity matrix.

1. Constructing an adjacency graph: Let G denote a graph with n nodes. The i -th node corresponds to the data point x_i . There are two ways to construct the adjacency graph: k nearest neighbors or ε neighborhood [3, 7]. In the following explanation, we assume the k nearest neighbors construction method is applied.

2. Computing the weights: The basic assumption is that each data point along with its k nearest neighbors (approximately) lies on a locally linear manifold, and each data point x_i is reconstructed by a linear combination of its k nearest neighbors. Let $W \in R^{n \times n}$ denote the weight matrix with W_{ij} having the weight of the edge from node i to node j , and 0 if there is no such edge. The weights on the edges are computed by minimizing the following reconstruction error function

$$\varepsilon(W) = \sum_i \left\| x_i - \sum_j W_{ij} x_j \right\|_2^2 \quad (3)$$

with constraints

$$\sum_j W_{ij} = 1, j = 1, 2, \dots, n$$

Obviously, the more similar x_j to x_i , the larger W_{ij} will be. To resolve it for a point x_i , we can define the local Gram matrix

$$G_{pl} = (x_i - x_p)^T (x_i - x_l) \in R^{k \times k}$$

the local Gram matrix containing the pairwise inner products among the neighbors of x_i , given that the neighbors are centered with respect to x_i . It can be easily proved that the constrained least squares problem has the following closed-form solution [9]

$$w_i = \frac{\sum_p G_{ip}^{-1}}{\sum_{pl} G_{pl}^{-1}} \quad (4)$$

where w_i represents the i -th column of W .

3. Computing the Orthogonal Neighborhood Preserving Embeddings: Define Let $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ be the orthogonal neighborhood preserving embeddings, we define:

$$A^{(k-1)} = [\mathbf{a}_1, \dots, \mathbf{a}_{k-1}] \quad (5)$$

$$S^{(k-1)} = [A^{(k-1)}]^T (XX^T)^{-1} A^{(k-1)} \quad (6)$$

The orthogonal neighborhood preserving vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ can be iteratively computed as follows:

1) Compute \mathbf{a}_1 as the eigenvector of $(XX^T)^{-1} XX^T$ associated with the smallest eigenvalues, where $M = (I - W)^T (I - W)$.

2) Compute \mathbf{a}_k as the eigenvector of

$$J^{(k)} = \{I - (XX^T)^{-1} A^{(k-1)} [S^{(k-1)}]^{-1} [A^{(k-1)}]^T\} \cdot (XX^T)^{-1} XX^T \quad (7)$$

associated with the smallest eigenvalues of $J^{(k)}$.

4. ONPE projection: Let $A_{ONPE} = [\mathbf{a}_1, \dots, \mathbf{a}_d]$, the embedding is as follows:

$$x \rightarrow y = A^T x \quad (8)$$

$$A = A_{PCA} A_{ONPE} \quad (9)$$

where \mathbf{y} is a d -dimensional representation of \mathbf{x} . A is the transform matrix.

Note, due to space limits, the theory justification is omitted here, which is similar to [10].

2.2. Orthogonal Neighborhood Preserving Classification

In [11], a local manifold matching method was proposed for face recognition. In [8], Want et al. proposed a semi-supervised classification method based on Linear Neighborhood label Propagation (LNP), and utilizes the local geometry information, but it operate in the original high-dimensional space. So as noted by the author, it is inefficient for very high-dimensional data set.

We follow the way of LNP, but with a further natural assumption that the local neighborhood relationship in low-dimensional space is preserved as in original high-dimensional space. Note that this is the aim of NPE projection algorithm, and we operate in the reduced space by apply ONPE first. That is, we classify data points on the reduced space after ONPE projection $Y = A^T X$, $y_i = A^T x_i, i = 1, \dots, n$. Suppose there are c classes and the label set becomes $L = \{1, 2, \dots, c\}$. Let M be a set of $n \times c$ matrices with non-negative real-value entries. Any matrix $F = [F_1^T, F_2^T, \dots, F_n^T]^T \in M$ corresponds to a specific classification on Y which labels y_i as $z_i = \arg \max_{j \leq c} F_{ij}$. Initially, we set $F_0 = Z$, where $Z_{ij} = 1$ ($1 \leq j \leq c$) if y_i is labeled as j , and $Z_{ij} = 0$ otherwise, and for unlabeled points, $Z_{ij} = 0$. The main classification procedure is list in table 1 (For details please refer to [8]).



Fig. 1. Sample face images from the ORL database. There are 10 face images for each subject with different facial expression.

TABLE 1 ONPC ALGORITHM

Input: $X = \{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_n\} \in \mathbb{R}^D$, $\{x_i\}_{i=1}^l$ are labeled, $\{x_u\}_{u=l+1}^n$ are unlabeled. The initial label matrix Z . The number of nearest neighbors k . The reduced dimension d . The constant α , defaults to 0.99.

Output: The labels of all the data points.

1. Dimension reduction by ONPE: $Y = A^T X$, note that here we also obtain the weight matrix W during ONPE.
2. Construct the propagation matrix $P = W$, and iterate $F_{t+1} = \alpha P F_t + (1 - \alpha) Z$ until convergence.
3. Let F^* be the limit of the sequence F_t . Output the labels of each data point x_i (y_i) by $z_i = \arg \max_{j \leq c} F_{ij}^*$.

3. EXPERIMENT RESULTS

In this Section, we investigate the use of ONPE on face analysis and recognition, and compare ONPE with Eigenface (PCA based), Fisherface (LDA based) and NPE. Eigenface and Fisherface are two of the most popular linear techniques for appearance-based face recognition.

The proposed method has been tested on ATT ORL database and Yale database. In ORL database, there are ten different grey images for each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). The size of each image is 92×112 pixels with 256 grey levels per pixel. 10 sample images of one individual are displayed in Fig. 1. The Yale database was constructed at the Yale Center for Computational Vision and Control. It contains 165 gray scale images of 15 individuals. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink).

The first experiment for recognition is tested on the ORL database. For comparison with NPE, the same preprocessing as in [7] is applied, and we use the preprocessed data which is available on the web

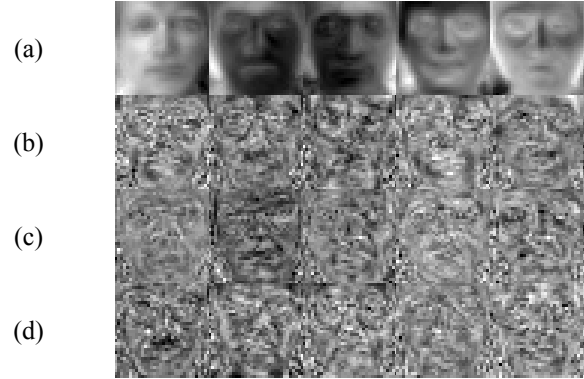


Fig. 2. The first 5 basic vectors on Yale face database. (a) Eigenfaces, (b) Fisherfaces, (c) NPEfaces, (d) ONPEfaces.

(<http://www.ews.uiuc.edu/dengcai2/Data/data.html>). The size of each preprocessed image in the experiments is 32×32 pixels, with 256 grey levels per pixel. Thus, each image can be represented as a 1024-dimensional vector in image space. As in [7], the nearest-neighbor method using Euclidean metric was employed for recognition on Eigenface, Fisherface and NPE, while for ONPE, the label propagation method (ONPC) is applied.

We have discussed how to learn an orthogonal neighborhood preserving subspace. The images of faces in the training set are used to learn such a face subspace. The subspace is spanned by the basis vectors. Therefore, any image in the face subspace can be represented as a linear combination of the basis vectors. We can display the basis vectors as a sort of feature images. Using the Yale face database as the training set, we present the first 5 basis vectors in Fig. 2, together with Eigenfaces and Fisherfaces, NPEfaces and ONPEfaces presentation. It is very interesting to see that the NPE and ONPE faces are similar to Fisherfaces.

For each individual, $l(=3,4,5)$ images are randomly selected for training and the rest are used for testing. For each given l , the results are averaged over 10 random splits. In general, the recognition rates vary with the dimension of the face subspace. Fig 3 shows the plot of recognition accuracy versus dimensionality d reduction for Eigenface, Fisherface, NPE and ONPE on ORL database (for $l=3,4$). The best result obtained in the optimal subspace and the corresponding dimensionalities for each method on ORL database are shown in Table 2. Note that, the upper bound of the dimensionality of reduced space is $n-1$, $c-1$, n and n for PCA, LDA, NPE and ONPE, respectively.

TABLE 2 PERFORMANCE COMPARISONS ON ORL

Method	3 Train	4 Train	5 Train
Eigenfaces	81.3%(106)	81.3%(118)	86.2%(58)
Fisherfaces	86.7%(38)	92.4%(39)	92.5%(39)
NPEfaces	87.2%(62)	93.1%(76)	94.3%(58)
ONPEfaces	91.7%(42)	95.2%(34)	97.8%(44)

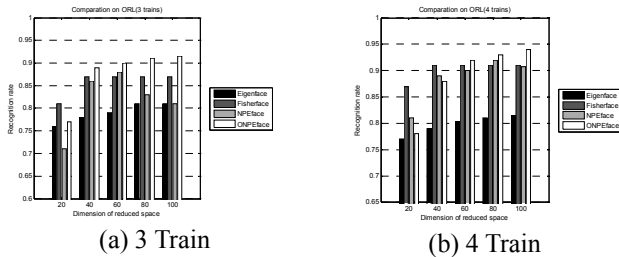


Fig. 3. Recognition rate vs dimension of reduced space on ORL.

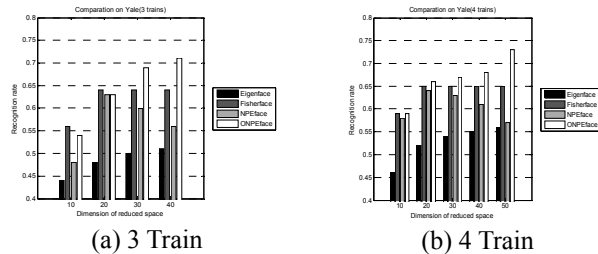


Fig. 4. Recognition rate vs dimension of reduced space on Yale.

As can be seen, our ONPE algorithm almost outperforms the other methods across all the values of dimension. Although LDA perform well when the value of d is small, their performance impairs as the dimension d increases. From Table 2, we can see that the dimension d of ONPE corresponding to the top recognition rate is low and the performance of ONPE improves significantly as the number of l of training samples per individual increases. The Fisherface method performed comparatively to ONPE as the size of the training set increases. Moreover, the optimal dimensionality obtained by NPE and ONPE and Fisherface is much lower than that obtained by Eigenface. It is also observed that ONPE can achieve similar recognition accuracy with much smaller dimension compared with NPE. Note that in NPE recognition, the KNN method is used, while in our ONPE, the classification method is based on our proposed method which takes into account the local neighborhood information.

Similar experiments are also applied to the Yale face database. The similar preprocessing is applied on the Yale database. Similar to the experiment on ORL database, a random subset with $l(=3,4)$ images per individual was taken with labels to form the training set. The rest of the database was considered to be the testing set. For each given l , we average the results over 10 random splits. The experimental protocol is the same as before. The recognition results are shown in Fig 4. Our ONPE method almost outperformed all the other methods.

One shortcoming of our method is that it needs more time to calculate the basic images for its iterative way. It needs further improvement for real time applications.

4. CONCLUSION

We have proposed a new algorithm for manifold learning called Orthogonal Neighborhood Preserving Embedding (ONPE). The new algorithm combines the orthogonal advantage of PCA transform and Neighborhood Preserving Embedding. Orthogonal NPE can have the neighborhood preserving power as NPE, while it does not suffer from the problem of metric distortion. Besides, we proposed a classification method (ONPC) for ONPE based on local label propagation with a natural assumption which utilizes the local geometry information. Experiments on face recognition demonstrated the effectiveness of the proposed method.

Acknowledgement: The work has been supported by the National High-Tech. R&D Program for 863, China (No. 2006AA01Z170, No.2006AA1Z171) and the highlight R&D Program of Zhejiang Province(No. 2006C11206).

REFERENCES

- [1] M.A. Turk, and A.P. Pentland, "Face recognition using eigenfaces". in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.
- [2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection". *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711-720, July 1997.
- [3] S.T. Roweis, and L.K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding". *Science*, 2000. 290: pp. 2323-2326.
- [4] J.B. Tenenbaum, V. Silva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction". *Science*, 2000. 290(12): pp. 2319-2323.
- [5] X.M. Liu, J.W. Yin, Z.L. Feng, and J.X. Dong, "Incremental Manifold Learning Via Tangent Space Alignment". in *Proceeding of Artificial Neural Networks in Pattern Recognition. LNAI*, 2006. pp.107-121.
- [6] Z. Zhang, and H. Zha, "Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment". *SIAM Journal on Scientific Computing*, 2006. 26(1): pp. 313-338.
- [7] X. He, D. Cai, S. Yan, and H.J. Zhang, "Neighborhood Preserving Embedding". in *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 2005: p. 1208-1213.
- [8] F. Wang, and C. Zhang. "Label propagation through linear neighborhoods". in *Proceedings of the 23rd international conference on Machine learning*. 2006. pp.985-992.
- [9] L.K. Saul, and S.T. Roweis, "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds". *Journal of Machine Learning Research*, 2004. 4(2): pp. 119-155.
- [10] D. Cai, X. He, and J. Han, "Orthogonal Laplacianfaces for Face Recognition". *IEEE transactions on image processing*, 2006. 15(11): pp. 3609-3614.
- [11] W. Liu, W. Fan, Y. Wang, and T. Tan. "Local Manifold Matching for Face Recognition". in *IEEE International Conference on Image Processing*. 2005, (2):pp.926-929