Signal Processing: Image Communication 55 (2017) 32–40

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Pair of projections based on sparse consistence with applications to efficient face recognition

Wenjie Zhu*, Yunhui Yan, Yishu Peng

School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Keywords:
Face recognition
Dimension reduction
Pair of projections
Sparse consistence

ABSTRACT

Dimension reduction based feature extraction and classification method show significant performance on the high-dimensional face images. The traditional dimension reduction methods learn a projection based on the Fisher criterion or local structure of the face images. This work aims at learning a pair of projection based on sparse consistence which is measured by sparse constraint and label information for efficient face recognition. The first projection maps the original high-dimensional face images into a low-dimensional space where each face is sparse, and the second one which can also be treated as a classifier guides the sparse low-dimensional face images to the right label. The pair of projections is optimized together using alternative update rules efficiently. Due to the discriminant power of sparse face images and the supervised classifier, the proposed algorithm integrates the supervised and unsupervised information and is more efficient than them for face recognition on both learning and classifying. Experimental results on the challenging Extended Yale B, AR, and LFW face image databases demonstrate the proposed algorithm on both accuracy and efficiency.

1. Introduction

As one of the traditional hot issue in computer vision, face recognition (FR) identifies the face image without label information according to the faces stored with labels automatically. There have been many works [1–5] on FR recently. Normally, the common approaches to FR are confronted with the efficiency problem due to the high-dimension of the face images. Hence, dimension reduction method, as one of the efficient FR, is always used for FR task. Fisher discriminant analysis (FDA) [6,7] and principal component analysis (PCA) [8,9] are the most famous approaches to feature extraction and dimension reduction as the linear transform method for FR. Supervised FDA maps the samples into a low-dimensional space in which the Euclidean distance steered clustering centers of diverse classes are separated possibly while the samples belonging to the same class have a minimal variance. Unsupervised PCA transforms the samples into a low-dimensional space where the samples have a maximal variance. Fisherface and Eigenface corresponding to FDA and PCA have been proposed in [10]. From different aspects, many works extend supervised and unsupervised transform methods for FR, such as kernel version [11–16], samples-character preserving version [17–19], and recently proposed classifier steered version [20,21]. As for robust face recognition in the real world, [22,23] integrate multiple feature into the classification framework and get impressed face recognition results. The basic idea of samples-character preserving version is modeling the samples with a supposed graph and preserving it in the transformed space while learning the projection. The supervised classifier steered discriminant analysis aims at learning a projection for maximizing the residual represented via samples belonging to other classes while minimizing it represented via samples with same labels.

Locality preserving projection (LPP) [18], one of the most famous extended dimension reduction methods, is an alternative to PCA, and it minimizes the Euclidean distance between the sample and the other samples which are close to it in the transformed space. Accordingly, LPP constructs a weight graph among the original samples. Neighborhood preserving embedding (NPE) [17], as a member of the LPP family, the weight matrix is computed between the sample and its neighborhoods which are determined using k-nearest-neighborhood (kNN) algorithm. Compared with LPP, the weight matrix preserves stricter locality. LPP determines the weight of the sample and all other samples before projection learning with Euclidean distance, while NPE determines the weight of the sample and limited samples which have a relatively close Euclidean distance using a least square algorithm. Another graph-preserved projection learning algorithm, sparsity preserving projection (SPP) [19], similar with LPP, also determines the weight of the sample and all the others. SPP constructs the weight graph by using ℓ_1 norm. Furthermore, SPP can handle the noise samples using the extended ℓ_1 sparse solution. Turn to the supervised transform

* Corresponding author.
E-mail addresses: wenjie_zh@126.com (W. Zhu), yanyh@mail.neu.edu.cn (Y. Yan), yishu_peng@126.com (Y. Peng).

http://dx.doi.org/10.1016/j.image.2017.03.012
Received 25 October 2016; Received in revised form 20 March 2017; Accepted 21 March 2017
Available online 23 March 2017
0923-5965/

methods on linear projection learning, the classifier steered discriminant analysis aims at transforming the samples with labels into the space where another supervised classification method works. Sparse representation classifier steered discriminative projection(SRC-DP) [20] learns a projection based on the criteria of SRC which keeps a minimal residual via the training samples belonging to the same class while a maximum one via those belonging to the other classes. Kernel locality-constrained collaborative representation based discriminant analysis (KLCR-DA) [21] is a complex constraint based projection. First, it employs the kernel trick to map the samples into a high-dimension space, and then computes a weight matrix between each sample and another to obtain the locality-constrained collaborative representation. The discriminant constraint, similar with SRC-DP, is adopted to learn the projection in the end. Euclidean distance adopted in FDA devotes to computing the mean value of the samples as the representation of the class. PCA, LPP, NPE, and SPP take into account the assumption that the samples own a typical graph in an unsupervised way. SRC-DP is over dependent on the classifier performance and ignores the character of the discriminant analysis itself for classification.

No matter what criterion used in the dimension reduction, the performance refines based on the following observations: (i) local structure preservation benefits for the feature extraction in the FR task; (ii) the sparse structure is conducive to increase the features' discriminant power.

Although, there still are some issues for the referred dimension reduction-based feature extraction: (i) the weight graph is sensitive to the computation method and may be a time-consuming problem; (ii) the weight graph may be invalid for the small sample problem due to the strict requirement to the training samples.

To this end, this paper will learn a pair of projection based on sparse consistence for efficient FR. The first projection maps the original high-dimensional face images into a low-dimensional space where each face is sparse, and the second one guides the sparse low-dimensional face images to the right label which can also be treated as a classifier. The two projections work together to guarantee the sparse consistence and extract the discriminative feature for FR task.

Main contributions of this paper are as follows:

(i) We explicitly introduce unsupervised and supervised information into the pair of projection learning procedure by employing sparse constraint and label information. By doing so, the pair of projections can extract the discriminant feature of the samples from a low level to a high level;.

(ii) We optimize the pair of projections together, design the second projection to keep the consistence of the first one guided sparse samples, and analyze the time complexity and convergence of the proposed algorithm to guarantee the efficiency of the algorithm;.

(iii)Comparison experiments on the three challenging face databases well validate the classification accuracy and the comparison of computation time on training phase testifies the efficiency of the proposed algorithm.

The remainder of this paper is organized as follows: Section 2 introduces the related work on dimension reduction for FR and Section 3 displays the proposed algorithm; Section 4 presents the experimental results on famous face databases and Section 5 concludes this paper.

2. Representative dimension reduction methods on projection learning

For ease of presentation, we introduce the notations used in the whole work. The bold uppercase letter denotes the matrix and the bold lowercase letter denotes the vector. Given a real $m \times n$ matrix $\mathbf{A} = (a_{ij})_{m\times n}$, $\mathbf{a}^i \in \mathbb{R}^n (i = 1, 2, \ldots, m)$ and $\mathbf{a}_j \in \mathbb{R}^m (j = 1, 2, \ldots, n)$ are respectively the i-th row and j-th column vectors of $\mathbf{A}$. The j-th element of the vector $\mathbf{a}$ is denoted as $\mathbf{a}[j]$. The ℓ_1-norm and ℓ_2-norm of $\mathbf{a}$ are defined as $\| \mathbf{a} \|_1$ and $\| \mathbf{a} \|_2$ respectively. The Frobenius norm of $\mathbf{A}$ is denoted as $\| \mathbf{A} \|_F$.

In the numerous unsupervised dimension reduction methods on the FR, PCA, LPP and SPP are the three representative ones. FDA and SRC-DP are represented as the supervised ones due to the performance on FR.

Given a set of data points $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ from C classes, where $\mathbf{x}_j \in \mathbb{R}^D$:

(i) PCA aims to seek a low-dimensional space where the points have a maximum variance. For each data point x_j, PCA obtains the low-dimensional point $\mathbf{y}_j = \mathbf{P}^T\mathbf{x}_j$ by solving the optimization problem:

$$\mathbf{P}^* = \arg \max_{\|\mathbf{P}\|_2=1} \sum_{j=1}^{n} \|\mathbf{y}_j - \tilde{\mathbf{y}}\|_2^2 \tag{1}$$

where $\tilde{\mathbf{y}}$ is the mean of $\mathbf{y}_j (j = 1, 2, \ldots, n)$.

(ii) LPP computes the projection to enforce that the points which are close to each other in the original space should keep the relation in the low-dimensional space as well. The relation of the data points so called local structure preservation involved into LPP makes it keep more discriminative information than PCA. LPP aims to solve the following optimization problem:

$$\mathbf{P}^* = \arg \max_{\mathbf{P}} \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \mathbf{W}_{ij} \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{n\times n}$ is the weight graph defined as:

$$\mathbf{W}_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{t}}, & \mathbf{x}_i \in kNN(\mathbf{x}_j), \mathbf{x}_j \in kNN(\mathbf{x}_i) \\ 0, & otherwise \end{cases} \tag{3}$$

and $kNN(\mathbf{x}_i)$ indicates the k-nearest-neighbor set of $\mathbf{x}_i$, t is the kernel size.

(iii) Diverse from LPP, SPP constructs the weight graph via computing the global sparse relationship of the original data. According to the weight graph which achieved by minimizing a regularization-related objective function, SPP gets the low-dimensional point $\mathbf{y}_j = \mathbf{P}^T\mathbf{x}_j$ by:

$$\mathbf{P}^* = \arg \max_{\mathbf{P}} \sum_{j}^{n} \| \mathbf{P}^T\mathbf{x}_j - \mathbf{P}^T\mathbf{X}\mathbf{w}_j \|_2 \tag{4}$$

where $\mathbf{W}$ is derived from the following modified ℓ_1 minimization problem:

$$\arg \min_{\mathbf{w}_j} \| \mathbf{w}_j \|_1 \quad s.t. \quad \mathbf{x}_j = \mathbf{X}\mathbf{w}_j, \quad \sum_j \mathbf{w}_j = 1, \quad \mathbf{w}_{jj} = 0 \tag{5}$$

(iv) FDA aims at obtaining a set of discriminant vectors which maximize the Fisher criterion in the low-dimensional space. The projection P is derived from the following optimization:

$$\mathbf{P}^* = \arg \max_{\mathbf{P}} \frac{\mathbf{P}^T S_B \mathbf{P}}{\mathbf{P}^T S_W \mathbf{P}} \tag{6}$$

where S_W and S_B are the within-class and between-class scatter matrices respectively and defined as follows [24]:

$$S_W = \frac{1}{n} \sum_{i=1}^{C} \sum_{j=1}^{n} (\mathbf{x}_{ij} - \mathbf{m}_i)(\mathbf{x}_{ij} - \mathbf{m}_i)^T, \; S_B = \frac{1}{n} \sum_{i=1}^{C} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{7}$$

where $\mathbf{m}_i$ is the mean of the i-th class and m is the mean of the set of data across all the classes. Intuitively, FDA employs the Fisher criterion to maximize the between-class distance while minimizing the within-class distance in the dimension reduction process.

(v) SRC-DP uses sparse representation-based classifier to steer the design of a feature extraction method. Diverse from FDA, SRC-DP maximizes the between-class reconstruction residual while minimizing the within-class reconstruction residual in the dimension reduction process. The within-class and between-class scatter matrices are defined as follows: