



# Double linear regressions for single labeled image per person face recognition



Fei Yin <sup>a,\*</sup>, L.C. Jiao <sup>a</sup>, Fanhua Shang <sup>b</sup>, Lin Xiong <sup>a</sup>, Shasha Mao <sup>a</sup>

<sup>a</sup> Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Mailbox 224, No. 2 South TaiBai Road, Xi'an 710071, PR China

<sup>b</sup> Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

## ARTICLE INFO

### Article history:

Received 12 March 2013  
Received in revised form  
10 September 2013  
Accepted 19 September 2013  
Available online 9 October 2013

### Keywords:

Semi-supervised dimensionality reduction  
Label propagation  
Sparse representation  
Linear regressions  
Linear discriminant analysis  
Face recognition

## ABSTRACT

Recently the underlying sparse representation structure in high dimensional data has received considerable attention in pattern recognition and computer vision. In this paper, we propose a novel semi-supervised dimensionality reduction (SDR) method, named Double Linear Regressions (DLR), to tackle the Single Labeled Image per Person (SLIP) face recognition problem. DLR simultaneously seeks the best discriminating subspace and preserves the sparse representation structure. Specifically, a Subspace Assumption based Label Propagation (SALP) method, which is accomplished using Linear Regressions (LR), is first presented to propagate the label information to the unlabeled data. Then, based on the propagated labeled dataset, a sparse representation regularization term is constructed via Linear Regressions (LR). Finally, DLR takes into account both the discriminating efficiency and the sparse representation structure by using the learned sparse representation regularization term as a regularization term of Linear Discriminant Analysis (LDA). The extensive and encouraging experimental results on three publicly available face databases (CMU PIE, Extended Yale B and AR) demonstrate the effectiveness of the proposed method.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

In many fields of scientific research such as face recognition [1], bioinformatics [2], and information retrieval [3], the data are usually presented in a very high dimensional form. This makes the researchers confront with the problem of “the curse of dimensionality” [4], which limits the application of many practical technologies due to the heavy computational cost in high dimensional space, and deteriorates the performance of model estimation when the number of training samples are small compared to the number of features. In practice, dimensionality reduction has been employed as an effective way to deal with “the curse of dimensionality”. In the past years, a variety of dimensionality reduction methods have been proposed [5–10].

According to the geometric structure considered, the existing dimensionality reduction methods can be categorized into three types: global structure based methods, local neighborhood structure based methods, and the recently proposed sparse representation structure [11,12] based methods. Two classical dimensionality reduction methods Principle Component Analysis (PCA) [13] and Linear Discriminant Analysis (LDA) [14] belong to global structure based methods. In the field of face recognition, they are known as “Eigenfaces” [15] and “Fisherfaces” [16]. Two popular local neighborhood structure based

methods are Locality Preserving Projections (LPP) [17] and Neighborhood Preserving Embedding (NPE) [18]. LPP and NPE are named “Laplacianfaces” [19] and “NPEfaces” [18] in face recognition. The representative sparse representation structure based methods include Sparsity Preserving Projections (SPP) [20], Sparsity Preserving Discriminant Analysis (SPDA) [21] and Fast Fisher Sparsity Preserving Projections (FFSPP) [22]. They have also been successfully applied to face recognition. In order to deal with the nonlinear structure in data, most of the above linear dimensionality reduction methods have been extended to their kernelized versions which perform in Reproducing Kernel Hilbert Space (RKHS) [23]. Kernel PCA (KPCA) [24] and Kernel LDA (KLDA) [25] are the nonlinear dimensionality reduction methods corresponding to PCA and LDA. Kernel LPP (KLPP) [17,26] and Kernel NPE (KNPE) [27] are the kernelized versions of LPP and NPE. The nonlinear version of SPDA is Kernel SPDA [21].

One of the major challenges to appearance-based face recognition is recognition from a single training image [28,29]. This problem is called “one sample per person” problem: given a stored database of faces, the goal is to identify a person from the database later in time in any different and unpredictable poses, lighting, etc. from just one image per person [28]. Under many practical scenarios, such as law enforcement, driver license and passport card identification, in which there is usually only one labeled sample per person available, the classical appearance-based methods including Eigenfaces and Fisherfaces suffer big performance drop or tend to fail to work. LDA fails to work since the within-class scatter matrix degenerates to a zero matrix when only one

\* Corresponding author. Tel.: +86 29 88202661; fax: +86 29 88201023.  
E-mail address: [yinfei701@163.com](mailto:yinfei701@163.com) (F. Yin).

sample per person is available. Zhao et al. [30] suggested replacing the within-class scatter matrix with an identity matrix to make LDA work in this setting, although the performance of this Remedied LDA (ReLDA) is still not satisfying. Due to its importance and difficulty, one sample per person problem has aroused lots of interest in face recognition community. To attack this problem, many ad hoc techniques have been developed, including synthesizing virtual samples [31,32], localizing the single training image [33], probabilistic matching [34] and neural network methods [35]. More details on single training image problem can be found in a recent survey [28].

As the fast development of the digital photography industry, it is possible to have a large set of unlabeled images. In this background, a more natural and promising way to attack one labeled sample per person problem is semi-supervised dimensionality reduction (SDR). Semi-supervised Discriminant Analysis (SDA) [29] is one SDR method which has been successfully applied to single labeled image per person face recognition. SDA first learns the local neighborhood structure using the unlabeled data and then uses the learned local neighborhood structure to regularize LDA to obtain a discriminant function which is as smooth as possible on the data manifold. Laplacian LDA (LapLDA) [36], Semi-supervised LDA (SSLDA) [37], and Semi-supervised Maximum Margin Criterion (SSMMC) [37] are all reported semi-supervised dimensionality reduction methods which can improve the performance of their supervised counterparts like LDA and Maximum Margin Criterion (MMC) [38]. These methods consider the local neighborhood structure and can be unified under the graph embedding framework [37,39]. Despite the success of these SDR methods, there are still some disadvantages: (1) these SDR methods are based on the manifold assumption which requires sufficiently many samples to characterize the data manifold [40]; (2) the adjacency graphs constructed in these methods are artificially defined, which brings the difficulty of parameter selection of neighborhood size and edge weights. To resolve these issues, Sparsity Preserving Discriminant Analysis (SPDA) [21] was presented. SPDA first learns the sparse representation structure through solving  $n$  (number of training samples)  $\ell^1$  norm optimization problems, and then uses the learned sparse representation structure to regularize LDA. SPDA has achieved a good performance on single labeled image per person face recognition, but it still has some shortages: (1) it is computationally expensive since  $n \ell^1$  norm optimization problems need to be solved in learning the sparse representation structure and (2) the label information is not taken advantage of in learning the sparse representation structure.

To tackle the above problems, we propose a novel SDR method, named Double Linear Regressions (DLR), which simultaneously seeks the best discriminating subspace and preserves the sparse representation structure. More specifically, a Subspace Assumption Based Label Propagation (SALP) method, which is accomplished using Linear Regressions (LR), is first presented to propagate the label information to the unlabeled data. Then, based on the propagated labeled dataset, a sparse representation regularization term is constructed via Linear Regressions (LR). Finally, DLR takes into account both the discriminating efficiency and the sparse representation structure by using the learned sparse representation regularization term as a regularization term of linear discriminant analysis. It is worthwhile to highlight some aspects of DLR as follows:

- (1) DLR is a novel semi-supervised dimensionality reduction method aiming at simultaneously seeking the best discriminating subspace and preserving the sparse representation structure.
- (2) DLR can obtain the sparse representation structure via  $n$  small class specific linear regressions. Thus, it is more time efficient than SPDA.

- (3) In DLR, label information is first propagated to all the training set. Then it is used in learning a more discriminative sparse representation structure.
- (4) Unlike SDA, there are no graph construction parameters in DLR. The difficulty of selecting these parameters is avoided.
- (5) Our proposed label propagation method SALP is quite general. It can be combined with other graph-based SDR methods to construct a more discriminative graph.

The rest of the paper is organized as follows. Section 2 gives a brief review of LDA and RDA. DLR is proposed in Section 3. DLR is compared with some related works in Section 4. The experimental results and discussions are presented in Section 5. Finally, Section 6 gives some concluding remarks and future work.

## 2. A brief review of LDA and RDA

Before we go into the details of our proposed DLR algorithm, a brief review of LDA and RDA is given in the following.

### 2.1. LDA

Given a set of samples  $\{\mathbf{x}_i\}_{i=1}^n$ , where  $\mathbf{x}_i \in \mathbb{R}^m$ , let  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$  be the data matrix consisting of all samples. Suppose samples are from  $K$  classes. LDA aims at simultaneously maximizing the between-class scatter and minimizing the within-class scatter. The objective function of LDA is defined as follows:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}, \quad (1)$$

$$\mathbf{S}_B = \sum_{k=1}^K N_k (\mathbf{m} - \mathbf{m}_k)(\mathbf{m} - \mathbf{m}_k)^T, \quad (2)$$

$$\mathbf{S}_W = \sum_{k=1}^K \sum_{i \in C_k} (\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T, \quad (3)$$

where  $\mathbf{m}_k = 1/N_k \sum_{i \in C_k} \mathbf{x}_i$ ,  $\mathbf{m} = 1/n \sum_{i=1}^n \mathbf{x}_i$ ,  $C_k$  is the index set of samples from class  $k$ , and  $N_k$  is the number of samples in class  $k$ .  $\mathbf{S}_B$  is called between-class scatter matrix and  $\mathbf{S}_W$  is called within-class scatter matrix. The optimal  $\mathbf{w}$  can be computed as the eigenvector of  $\mathbf{S}_W^{-1} \mathbf{S}_B$  that corresponds to the largest eigenvalue [14].

When there is only one labeled sample per class, LDA fails to work because  $\mathbf{S}_W$  is a zero matrix. The Remedied LDA (ReLDA) for one labeled sample per class scenario was proposed by Zhao et al. [30] in which  $\mathbf{S}_W$  is replaced by an identity matrix.

### 2.2. RDA

Despite its simplicity and effectiveness for classification, LDA suffers from the Small Sample Size (SSS) problem [41]. Among the methods designed to attack this problem, Regularized Discriminant Analysis (RDA) [42,43] is a simple and effective one, whose objective function is defined as follows:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w} + \lambda_1 \mathbf{w}^T \mathbf{w}} \quad (4)$$

where  $\lambda_1$  is the tradeoff parameter. The optimal  $\mathbf{w}$  can be computed as the eigenvector of  $(\mathbf{S}_W + \lambda_1 \mathbf{I})^{-1} \mathbf{S}_B$  that corresponds to the largest eigenvalue.

Download English Version:

<https://daneshyari.com/en/article/532058>

Download Persian Version:

<https://daneshyari.com/article/532058>

[Daneshyari.com](https://daneshyari.com)