Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Towards multi-scale fuzzy sparse discriminant analysis using local third-order tensor model of face images

Xiaoning Song [a,b,*], Zhen-Hua Feng [b], Xibei Yang [c], Xiaojun Wu [a], Jingyu Yang [c]

[a] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China
[b] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK
[c] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form
14 October 2015
Accepted 3 December 2015
Communicated by Ning Wang

Keywords:
Fuzzy sparse discriminant analysis
Multi-scale local analysis
Face recognition

ABSTRACT

Traditional discriminant analysis (DA) methods are usually not amenable to being studied only with a few or even single facial image per subject. The fundamental reason lies in the fact that the traditional DA approaches cannot fully reflect the variations of a query sample with illumination, occlusion and pose variations, especially in the case of small sample size. In this paper, we develop a multi-scale fuzzy sparse discriminant analysis using a local third-order tensor model to perform robust face classification. More specifically, we firstly introduced a local third-order tensor model of face images to exploit a set of multi-scale characteristics of the Ridgelet transform. Secondly, a set of Ridgelet transformed coefficients with respect to each block from a face image are respectively generated. We then merge all these coefficients to form a new representative vector for the image. Lastly, we evaluate the sparse similarity grade between each training sample and class by constructing a sparse similarity metric, and redesign the traditional discriminant criterion that contains considerable fuzzy sparse similarity grades to perform robust classification. Experimental results conducted on a set of well-known face databases demonstrate the merits of the proposed method, especially in the case of insufficient training samples.

1. Introduction

The main purpose of a traditional supervised method is to identify a subset of features that are most predictive or informative in a given dataset. For instance, the discriminant information is a major algebraic cue utilized in a wide variety of applications in image understanding. However, traditional supervised methods are often fail in dealing with many practical scenarios that are intrinsically noisy or affected by various external factors, including illumination, occlusion, pose and expression variations. To tackle this problem, a large number of algorithms have been proposed, and many existing supervised methods perform reasonably well in presence of these variations. Nevertheless, a supervised classifier usually fail in extracting effective features due to the variations in appearance between training samples and gallery ones, especially in the case of insufficient training set.

Recently, locality-oriented feature representation has become a fundamental and efficient learning approach in image analysis by grouping similar objects into the same category, as reported in many previous works, including locally linear embedding (LLE) [1], isomap [2], locality preserving projections (LPP) [3], laplacianface [4], locality pursuit embedding (LPE) [5], neighborhood preserving embedding (NPE) [6–8] and locality preserving canonical correlation analysis (LPCCA) [9]. These methods have the capability to improve the performance over their supervised counterparts, such as the discriminant analysis (DA) [10–12] and maximum margin criterion (MMC) [13]. Despite the success of these locality-oriented learning algorithms in face recognition, some issues have not been properly addressed:

- Many existing learning algorithms are based on the manifold assumption, implying that sufficient training samples are required to characterize the data distribution. However, it is practically hard to obtain a sufficient sampling for intrinsic high dimensional data, such as a face dataset [14–16];
- In general, a supervised method usually convert an image to a long vector by concatenating its row or its column components. However, this strategy does not fully evaluate the adverse effect of spatial correlation among adjacent pixels in an image. In fact, each pixel in a face image is highly correlated with its neighborhoods [17,18].
- Another kind of supervised methods are matrix-based, including the well-known 2D PCA [19] and 2D LDA [20]. Nevertheless,

* Corresponding author at: School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China.
E-mail addresses: xnsong@hotmail.com, xnsong@aliyun.com (X. Song).

http://dx.doi.org/10.1016/j.neucom.2015.12.019
0925-2312/

these matrix-based methods only consider the spatial redundancies in each row or column of an image, which leads to large scatter matrices and highly computational cost in feature extraction [21,22].
- In supervised classification, a sample from the specific class is expected to be assigned a large connecting weight. The traditional locality-oriented learning algorithms, however, often fail to capture piecewise linear relationships between the sample and different classes.

Meanwhile, as an emerging technique in feature representation, sparse coding [23,24] has been used as a strong prior to alleviate the specialization of signal reconstruction [25]. In addition, it has been proven that the coefficients in sparse coding could be discriminative [26]. In fact, for a dataset with linear or piecewise-linear class structure, sparse learning conveys important discriminative information that is automatically encoded in ℓ_1 regularization [27,28]. Recently, inspired by Gabor transform mechanism, Yang et al. [29] proposed a Gabor occlusion dictionary learning algorithm to reduce the computational cost in sparse coding the occluded face images, by which the number of atoms is significantly reduced in the Gabor occlusion dictionary. Meanwhile, Fang et al. [30] proposed a flexible representation algorithm termed multitask adaptive sparse representation (MASR), which not only restricts Gabor features of one test sample to be jointly represented by training atoms from the same class but also promotes the selected atoms for these features to be varied within each class. Though these approaches mentioned above provide promising classification results, some further questions are worth considering: 1) how to find a more natural way to capture the fully spatial redundancies information among the adjacent pixels (or block) in an image; 2) how much of this information can be reorganized to generate sparse similarity with respect to different classes, so as to achieve efficient fuzzy discriminant analysis for robust classification.

In this paper, we present a multi-scale fuzzy sparse discriminant analysis using a local third-order tensor image model to perform robust face classification. The proposed algorithm has three main contributions:

- Firsly, we introduce a third-order tensor model in image representation, by which an image is divided into m_1 local blocks with size $m_2 \times m_3$. Then a set of Ridgelet transformed coefficients with respect to each block are calculated. The final representation vector of the image is formed by concatenating the coefficients across all the blocks of the image.
- Secondly, we calculate the sparse similarity grade between each training sample and class in the training set by constructing a sparse similarity metric, that is, the fuzzy sparse statistical properties of class "*i*" for the *j*th sample can be evaluated. Thereby, the contributions of all training samples from a specific class becomes critical to build feature extractor with good discriminant capability.
- Lastly, we redesign a new discriminant analysis criterion that contains considerable fuzzy sparse similarity grades, by this means, each test sample can be classified into multi-classes under different similarity grades and the label of test sample $\mathbf{y}$ can be finally determined by sparse subspace spanned by the optimal discriminant vectors.

The remainder of this paper is organized as follows. Section 2 provides concise outlines of Ridgelet transform, traditional discriminant analysis and sparse representation, which can serve as prerequisites to the proposed algorithm detailed in Section 3. Section 4 reports comprehensive experimental results on several commonly used face image databases, including ORL, XM2VTS, FERET and AR. A more detailed discussion and analysis of the proposed algorithm is presented in Section 5. At last, Section 6 concludes the whole paper.

2. Backgrounds

In this section, we briefly introduce the fundamental conceptions of the prerequisite knowledge, including Ridgelet transform [31,32], classical discriminant criterion [10–12] and sparse representation [27,28,33–36].

2.1. Ridgelet transform

To overcome the weakness of wavelets in higher dimensional space, Candes and Donoho [31] have pioneered Ridgelets that effectively deals with line singularities in 2-D, by mapping a line singularity into a point singularity with Radon transform [37]. Then, the wavelet transform is used to handle the point singularity in Radon domain. Thus, Ridgelet transform represents lines and other singularities along lines in a more efficient way than wavelets. Given an integrable bivariate function $f(\mathbf{x})$, the continuous Ridgelet transform (CRT) is defined by [32]

$$R(a,b,\theta) = \int \psi_{a,b,\theta}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \tag{1}$$

where $\int \psi_{a,b,\theta}$ is the Ridgelets given by

$$\psi_{a,b,\theta}(\mathbf{x}) = a^{-1/2}\psi((\mathbf{x}_1 \cos\theta + \mathbf{x}_2 \sin\theta - b)/a) \tag{2}$$

Here, ψ is the smooth univariate function with sufficient decay and satisfying admissibility condition

$$\int |\psi(\xi)|^2/|\xi|^2 d\xi < \infty \tag{3}$$

which holds only if ψ has a vanishing mean. The exact reconstruction formula is

$$f(\mathbf{x}) = \int_0^{2\pi}\int_{-\infty}^{\infty}\int_0^{\infty} R(a,b,\theta)\psi_{a,b,\theta}(\mathbf{x})\frac{da}{a^3}db\frac{d\theta}{4\pi} \tag{4}$$

Note that CRT is similar to 2-D continuous wavelet transform (CWT) except that the point parameters are replaced by the line parameters.

2.2. Discriminant analysis

The traditional DA criterion aims to maximize the ratio of between-class scatter matrix to within-class scatter matrix. Given a set of n_i points belonging to class C_i, we can define the mean of each class as

$$\boldsymbol{\mu}_i = \frac{1}{n_i}\sum_{k \in C_i} \mathbf{x}_k \tag{5}$$

where $i = 1, \ldots, C$, C is the number of classes. The within-class scatter matrix is defined as

$$\mathbf{S}_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{\mathbf{x}_k \in C_i}(\mathbf{x}_k - \boldsymbol{\mu}_i)(\mathbf{x}_k - \boldsymbol{\mu}_i)^{\mathrm{T}} \tag{6}$$

where N is the total number of image samples $N = \sum_{i=1}^{C} n_i$. The between-class scatter matrix is defined as

$$\mathbf{S}_b = \frac{1}{C}\sum_{i=1}^{C}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}} \tag{7}$$

Please cite this article as: X. Song, et al., Towards multi-scale fuzzy sparse discriminant analysis using local third-order tensor model of face images, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.12.019