Pattern Recognition 56 (2016) 40–49

Contents lists available at ScienceDirect

Pattern Recognition

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

Quadratic projection based feature extraction with its application to biometric recognition

Yan Yan [a], Hanzi Wang [a,*], Si Chen [b], Xiaochun Cao [c], David Zhang [d]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Fujian 361005, China
[b] School of Computer and Information Engineering, Xiamen University of Technology, Fujian 361024, China
[c] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[d] Biometrics Research Centre, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form
7 December 2015
Accepted 16 February 2016
Available online 24 February 2016

Keywords:
Biometric recognition
Feature extraction
Quadratic projection
Semidefinite programming
Lagrange duality

ABSTRACT

This paper presents a novel quadratic projection based feature extraction framework, where a set of quadratic matrices is learned to distinguish each class from all other classes. We formulate quadratic matrix learning (QML) as a standard semidefinite programming (SDP) problem. However, the conventional interior-point SDP solvers do not scale well to the problem of QML for high-dimensional data. To solve the scalability of QML, we develop an efficient algorithm, termed DualQML, based on the Lagrange duality theory, to extract nonlinear features. To evaluate the feasibility and effectiveness of the proposed framework, we conduct extensive experiments on biometric recognition. Experimental results on three representative biometric recognition tasks, including face, palmprint, and ear recognition, demonstrate the superiority of the DualQML-based feature extraction algorithm compared to the current state-of-the-art algorithms.

1. Introduction

A typical statistical pattern recognition system usually consists of four modules: a sensor module, a preprocessing module, a feature extraction module, and a classification module [1]. Among these four modules, the feature extraction module plays a critical role in the success of the system. The objective of feature extraction is to find a specific representation which encodes relevant information from input data, so that not only is the computational complexity of subsequent classifiers reduced but also the useful features can be used to perform the desired tasks [2].

Usually, the real-world data can be represented as a high-dimensional vector [3]. For instance, an image of size 80×80 can be viewed as a point in a 6400 dimensional feature space. However, the high dimensionality of data prevents from direct usage of learning techniques in a high-dimensional space. A common way to deal with this problem is to make use of feature extraction techniques, or more specifically, use dimensionality reduction techniques [2,4-6] to project the original high-dimensional data onto a low-dimensional space.

Recently, biometric recognition, which refers to the task of automatic identification of individuals based on their physiological and/or behavioral characteristics, has received much attention due to its wide range of applications, such as law enforcement, access control, and video surveillance [1,3,7]. A number of biometrics have been proposed in recent years (e.g., [8–12]). Two kinds of biometric characteristics are usually used, i.e., physiological characteristics (such as face, palmprint, and ear) and behavioral characteristics (such as gait, signature). Despite decade-long efforts, building an automatic and robust biometric recognition system remains a challenging problem due to variations in illumination, pose, occlusion, etc.

During the fast few decades, numerous feature extraction methods have been put forward to deal with the biometric recognition problems. For example, Qian et al. [13] proposed the discriminative histograms of local dominant orientation (D-HLDO) method for biometric image feature extraction. Shekhar et al. [14] developed a joint sparse representation for robust multimodal biometrics recognition. Beside feature extraction, feature selection is also extensively investigated to discover the knowledge related to biometric data. Different from feature extraction, which generates new features from functions of the original features, feature selection returns a subset of the features from a large feature pool. Boosting [15,16] and Lasso [17] have been successfully used to perform feature selection in face detection and recognition. Sun et al. [18] proposed an optimization formulation for ordinal feature

* Corresponding author. Tel.: +86 592 2580063; fax.: +86 592 2580063.
E-mail addresses: yanyan@xmu.edu.cn (Y. Yan), hanzi.wang@xmu.edu.cn (H. Wang), chensi@xmut.edu.cn (S. Chen), caoxiaochun@iie.ac.cn (X. Cao), csdzhang@comp.polyu.edu.hk (D. Zhang).

http://dx.doi.org/10.1016/j.patcog.2016.02.010
0031-3203/

selection for iris and palmprint recognition. Guo et al. [19] presented the feature band selection for the online multispectral palmprint recognition. Ghoualmi et al. [20] proposed a feature selection method based on the genetic algorithm for ear authentication. Kumar et al. [21] suggested to use feature selection and combination to improve the performance of bimodal biometric system.

Until now, a large number of feature extraction methods have been developed. However, many methods mainly consider the first order statistics of data, which are indeed non-linearly distributed. Even though nonlinear feature extraction methods are introduced to handle the non-linearly distributions, the computational cost of these methods is high. On the other hand, high order statistics which capture the complex statistical relationship of the data can be beneficial for feature extraction and feature selection, potentially leading to superior performance.

In this paper, we propose a novel nonlinear feature extraction framework, which takes advantage of the quadratic projection technique. Compared with the traditional linear projection technique, the quadratic projection technique exploits the second order statistics of data. It is well-known that the quadratic classifiers are optimal for the data under Gaussian distributions. Even when the data is not Gaussian-distributed, we can still expect quadratic projection to perform better than linear projection under general conditions since more high-order information is taken into consideration in quadratic projection.

More specifically, we propose a novel nonlinear feature extraction framework based on the quadratic projection technique. Different from the traditional linear projection technique (which obtains a feature vector based on a linear form), the quadratic projection technique uses a quadratic form to extract a feature vector, where each feature is extracted by using the homogeneous polynomial of degree two in a number of original features. In the proposed framework, a set of quadratic matrices is learned to distinguish each class from all other classes. Mathematically, we formulate quadratic matrix learning (QML) as a standard semidefinite programming (SDP) problem. To solve the scalability of QML, we further develop an efficient algorithm which significantly reduces the computational complexity of the conventional interior-point SDP solvers [6].

In this paper, we will motivate and study this new framework within the context of biometric recognition. We use biometrics data as a case study to illustrate the effectiveness of the proposed framework. Experimental results on three representative biometric recognition tasks (including face, palmprint, and ear recognition) show that the proposed algorithm achieves better performance than the linear projection based and kernel/tensor based feature extraction algorithms.

In summary, the main contributions of our work are summarized as follows:

1. A novel feature extraction framework based on the quadratic projection technique is proposed to extract discriminative features, where a set of quadratic matrices is learned. Experimental results on biometric recognition tasks show the effectiveness of the proposed framework.
2. We develop an efficient algorithm for quadratic matrix learning (QML) via the Lagrange duality theory. Our proposed algorithm is much more scalable than the traditional SDP solvers. The importance of this improvement is that it thereby allows us to apply QML to high-dimensional data.

The rest of this paper is organized as follows. Section 2 describes related work. Section 3 presents the details of the proposed quadratic projection based feature extraction framework, where a novel algorithm is developed for efficient QML. Experimental results on three biometric recognition tasks are given in Section 4. Finally, Section 5 provides the concluding remarks.

2. Related work

Feature extraction can be performed in a linear or nonlinear way. The linear feature extraction based algorithms usually perform a linear mapping of input data onto a low-dimensional feature space. Typical algorithms include principal component analysis (PCA) [22], linear discriminant analysis (LDA) [23,9], locality preserving projections (LPP) [24], margin fisher analysis (MFA) [4], class-dependence feature analysis (CFA) [25,26], local discriminative gaussians (LDG) [27], and low rank matrix factorization [28]. Recently, a large number of distance metric learning algorithms [29–32] have been proposed to perform linear feature extraction. These algorithms are computationally efficient. However, their performance can degrade in cases with non-linearly distributed data existing in many real-world applications.

Nonlinear feature extraction algorithms are based on the intuition that input data lies on a nonlinear manifold in a high-dimensional space. A direct and natural way to extend the linear feature extraction algorithms to nonlinear cases is to take advantage of the kernel technique [33,34], which does not have to explicitly compute the nonlinear mapping between the input space and the feature space. The kernel-based nonlinear algorithms find nonlinear projections by nonlinearly mapping data onto a higher-dimensional feature space, but it still performs linear projections in the new feature space.

Other types of nonlinear algorithms include manifold learning techniques, such as ISOMAP [35], locally linear embedding (LLE) [36], and local tangent space alignment (LTSA) [37]. Nevertheless, many manifold learning algorithms suffer from the so-called out-of-sample problem [38], i.e., these algorithms provide mapping only for training data but not for unseen test data.

The multilinear subspace learning (MSL) techniques [39,40] have also been developed for finding a low-dimensional representation of high-dimensional tensor data through direct mapping. There are three types of multilinear projections according to the forms of input and output of a projection [39], i.e., vector-to-vector projection (VVP), tensor-to-tensor projection (TTP), and tensor-to-vector projection (TVP). Although the MSL algorithms preserve the structure in original data by operating on natural tensor representations, most of these algorithms are based on iterative schemes and usually converge to local solutions.

Generally speaking, the distributions of real-world data (such as biometric data) show highly non-linear and non-convex. Therefore, the non-linear feature extraction is beneficial for the subsequent classification. However, the kernel extension is computationally expensive, while the multi-linear based algorithms often offer local optimal solutions. In this paper, we develop a novel nonlinear feature extraction framework, which leverages the quadratic projection technique to encode high order statistics of the biometric data. Note that both the proposed algorithm and the MSL algorithms (using 2D matrix data) [39,40] aim to optimize a matrix. However, in the proposed framework, the optimized matrix is a quadratic matrix required to satisfy the positive semidefinite constraint (usually not required in the MSL algorithms). Furthermore, compared with the MSL algorithms which attempt to obtain one matrix to distinguish all classes, the proposed framework obtains multiple quadratic matrices, where each quadratic matrix is trained to separate one class from the other classes.