Signal Processing 106 (2015) 245–252

Contents lists available at ScienceDirect

Signal Processing

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

A sparse and discriminative tensor to vector projection for human gait feature representation

Lefei Zhang [a], Liangpei Zhang [b], Dacheng Tao [c], Bo Du [a,*]

[a] Computer School, Wuhan University, Wuhan 430072, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China
[c] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia

ARTICLE INFO

Article history:
Received 4 May 2014
Received in revised form
21 July 2014
Accepted 2 August 2014
Available online 13 August 2014

Keywords:
Tensor representation
Multi-linear principal component analysis
Discriminative locality preserving projection
Feature extraction
Gait recognition

ABSTRACT

In this paper, we introduce an efficient tensor to vector projection algorithm for human gait feature representation and recognition. The proposed approach is based on the multi-dimensional or tensor signal processing technology, which finds a low-dimensional tensor subspace of original input gait sequence tensors while most of the data variation has been well captured. In order to further enhance the class separability and avoid the potential overfitting, we adopt a discriminative locality preserving projection with sparse regularization to transform the refined tensor data to the final vector feature representation for subsequent recognition. Numerous experiments are carried out to evaluate the effectiveness of the proposed sparse and discriminative tensor to vector projection algorithm, and the proposed method achieves good performance for human gait recognition using the sequences from the University of South Florida (USF) HumanID Database.

1. Introduction

Human recognition by analysis of the biometric resources such as face, gait, iris, and palmprint has been thoroughly studied and employed in many applications [1,2]. Among them, gait recognition, i.e., identify a person by the manner of walking patterns extracted from the video, has been reported that could well identify a human subject in low-resolution images that are taken at a distance, has recently become a popular research problem in signal processing community and a number of gait recognition algorithms have been proposed in the recent few years [3–5].

The main challenge of the gait recognition lies in that its performance is affected by many factors, including the environmental factors as well as the physical characteristics of the human subject, such as the camera factors, the observed sequence and preprocessed silhouette quality, the human walking speed, elapsed time, carrying condition, and shoes style. Furthermore, these aforementioned factors are sometimes correlated [2,6,7]. The open source humanID dataset [8] investigated several of these factors and provided the baseline algorithm for human gait recognition, however, how to measure the effect of these factors or to combine these factors is still an open question to date.

Most of the existing approaches for gait recognition could be broadly categorized into two classes: the model based and the silhouette based approaches. The model based methods usually characterise a human subject by some specific models, e.g., the appearance-based models [9,10], the stochastic statistical models [11], the articulated biomechanics models [12,13], and some other parameter-based models such as the method proposed in [14]. In contrast, the silhouette based gait recognition algorithms do not assume certain model of the human subject, but analyze the spatio-temporal shape and motion features of

* Corresponding author.
E-mail address: gunspace@163.com (B. Du).

http://dx.doi.org/10.1016/j.sigpro.2014.08.005
0165-1684/

the silhouette images which are extracted from the sequences [2,15–17]. In this paper, we suppose that the standard silhouette images are available and focus on the silhouette based approaches for human gait feature representation and recognition.

In the computer vision and machine learning research, the effective feature representation of the studied subject is a key issue. Since most of the data dimension reduction algorithms (e.g., the principal components analysis, PCA [18], the linear discriminant analysis, LDA [19], the Sparse PCA [20], and some recently published methods [21,22]) and classification algorithms (e.g., the Support Vector Machines, SVM [23]) in the literature only allow the vector feature as inputs, for the gait recognition task, it is straightforward to resize the silhouette images to the long vectors as the input feature representation. However, the feature dimensionality is usually much larger than the number of gallery sets, which causes an issue that is known as the "small sample-size (SSS) problem" [24,25], and thus results in poor recognition performance [26,27]. The other key shortcoming of the vector feature representation lies in that this scheme has lost many of the original spatial constraints of each pixel in the silhouette images, which hinders the subsequent algorithm to construct the optimal dimension reduction and classification model with only limited training samples.

To address the aforementioned difficulties, some researchers suggest to use matrix and tensor representation instead of vector representation [28,29], as reported in the works of two-dimensional PCA [30,31], two-dimensional LDA [15], tensor subspace analysis [32], multi-linear PCA [16,33], and so on. Especially in the framework of multi-linear PCA, the authors introduce a multi-linear principal component projection that captures most of the original tensorial input variation. However, the standard multi-linear PCA algorithm ignores the discriminative information provided by the gallery set, and its output data is still in a tensor format, which could not be directly processed by a conventional classifier. In this paper, we introduce an efficient tensor to vector projection algorithm for human gait feature representation and recognition. The main contributions of this paper are summarized as follows:

1. We suggest to represent the observed gait sequences as 3-order-tensors by which the data structure of both the spatial and temporal domains is well preserved. Based on the tensor representation, we further introduce the multi-linear PCA algorithm, which performs feature extraction by finding a low-dimensional tensor subspace that captures most of the data variation of original input gait sequence tensors.
2. Followed by the multi-linear PCA algorithm, we adopt a discriminative locality preserving projection with sparse regularization to transform the refined tensor data to the final vector feature representation, by which the class separability is improved and the potential model overfitting is simultaneously avoided.

The output feature representation by our proposed sparse and discriminative tensor to vector projection (SDTTV) could be simply employed for subsequent classification, and numerous experiments indicate that the proposed algorithm achieves good performance for human gait recognition. The remainder of this paper is organized as follows. In the following section, we give a brief description of related tensor algebra, and then present the proposed SDTTV algorithm in detail. After that, the experiments are reported in Section 3, followed by the conclusion.

2. The proposed sparse and discriminative tensor to vector projection algorithm

In this section, we first review some multi-dimensional or tensor signal processing rules that are related to the proposed SDTTV algorithm, then we introduce the multi-linear PCA approach, which aims to find a low-dimensional tensor subspace of original input tensors while most of the data variation could be well preserved. Finally, we provide the optimization of SDTTV algorithm by combining the discriminative locality preserving projection and a sparse regularization to transform the refined tensor data to the final vector feature representation for subsequent gait recognition.

2.1. Related tensor algebra

The notationsused in this paper are followed by convention in the tensor papers, e.g., vectors are denoted by lowercase boldface and italic letters, such as $\boldsymbol{x}$, matrices by uppercase boldface and italic, such as $\boldsymbol{W}$, and tensors by calligraphic letters, such as $\mathcal{A}$. For a M-order-tensor $\mathcal{A} \in \mathbb{R}^{K_1 \times K_2 \times \cdots \times K_M}$, in which K_i suggests the size of this tensor in each mode, and the elements of $\mathcal{A}$ are denoted with indices in lowercase letters, i.e., $\mathcal{A}_{k_1,k_2,\ldots,k_M}$, in which each k_i addresses the i-mode of $\mathcal{A}$, and $1 \le k_i \le K_i$, $i \in (1,2,\ldots,M)$. Unfolding tensor $\mathcal{A}$ along the i-mode is defined by keeping the index k_i fixed and varying the other indices, the result of which is denoted as a matrix $\boldsymbol{A}_{(i)} \in \mathbb{R}^{K_i \times \prod_{j \neq i} K_j}$. The i-mode product of a tensor $\mathcal{A}$ by a matrix $\boldsymbol{W} \in \mathbb{R}^{J_i \times K_i}$ is a tensor with entries $(\mathcal{A} \times_i \boldsymbol{W})_{k_1,\ldots,k_{i-1},j_i,k_{i+1},\ldots,k_M} = \sum_{k_i} \mathcal{A}_{k_1,\ldots,k_M} \boldsymbol{W}_{j_i,k_i}$. The Frobenius norm of a tensor $\mathcal{A}$ is given by $\|\mathcal{A}\| = \sqrt{\sum_{k_1} \cdots \sum_{k_M} \mathcal{A}^2_{k_1,k_2,\ldots,k_M}}$, and the Euclidean distance between two tensors $\mathcal{A}$ and $\mathcal{B}$ could be measured by $\|\mathcal{A} - \mathcal{B}\|$. For more detailed information, refer to [34,35]. As a summary, Table 1 lists the fundamental symbols defined in tensor algebra related to this paper.

2.2. Multi-linear principal components analysis

Suppose we have N gait tensors in the gallery set, i.e., $\{\mathcal{A}_n \in \mathbb{R}^{K_1 \times K_2 \times K_3}, n = 1,2,\ldots,N\}$, in which K_1 and K_2 are the height and width of each grey-level gait frame image from a subject sequence, respectively, and K_3 is the number of frames in a sequence. In order to achieve the objective of finding a low-dimensional tensor subspace of original input tensors in which most of the data variation could be well preserved, the multi-linear PCA algorithm suggests to adopt a series of