Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

Discriminative sparse projections for activity-based person recognition

Haibin Yan

School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 23 September 2015
Received in revised form
20 November 2015
Accepted 20 November 2015

Keywords:
Person recognition
Gait recognition
Metric learning
Sparse coding
Activity analysis

ABSTRACT

In this paper, we propose an activity-based person recognition approach based on discriminative sparse projections (DSPs) and ensemble metric learning. Unlike gait recognition where only the walking activity is utilized for human identification, we aim to recognize people from more types of activities such as eating, drinking, running, and so on. Our motivation is inspired by the fact that people do not always walk in person identification systems and gait recognition could fail to work in this scenario. Given each video clip, we first extract the binary human body mask in each frame by using background substraction. Then, we propose a DSP method to project these body masks into a low-dimensional subspace and cluster them into a number of clusters simultaneously. Subsequently, each video clip is pooled as a histogram feature for activity representation. Lastly, we propose an orthogonal ensemble metric learning (OEML) method to learn a distance metric to exploit more discriminative information for recognition. Experimental results on five benchmark activity databases are presented to show the efficacy of our proposed approach. Moreover, we apply our approach to gait recognition and also achieve competitive results with the state-of-the-art methods on the widely used CASIA-B gait dataset.

1. Introduction

Over the pase two decades, person recognition has been intensively studied in computer vision and pattern recognition [1–8] due to its wide applications such as access control, human–computer interaction, visual surveillance, and information security. Most existing person identification approaches such as face recognition [3], fingerprint recognition [2], palmprint identification [6] and iris recognition [9], usually require person cooperation and are generally not suitable for person identification at a distance. Unlike these biometric recognition techniques, gait-based person identification has attracted much attention in recent years [4,10–12] because gait provides a noninvasive way to recognize people at a distance.

While a number of gait recognition methods have been proposed in the literature [4,10–16], one key shortcoming of gait recognition is that only human walking activity is considered and utilized for person identification and these gait recognition systems are likely to fail to work when people perform other activities such as eating, drinking, and running rather than walking in these systems. In many real-world applications, people may not always walk in the scene and it is very likely that they are performing other activities besides walking in the scene. Since gait can provide enough discriminative information for human identification, a natural question arises: is it possible to identify people from different types of activities rather than gait since gait can be considered as a special case of general human activities, as shown in Fig. 1? If so, how to effectively explore discriminative features of these activities to achieve this goal? In this paper, we provide a positive answer to these two questions.

Intuitively, the manner in which humans perform different activities can provide some distinctive information for person identification because human body information is generally distinct for different persons. Moreover, different dynamic information observed in other activities are also discriminative. Similar to gait recognition, people may perform the same activity in different manners. While video-based face and gait recognition has been extensively studied over the past decade [13,14,17–22], there has been extremely few attempts on using other activities rather than gait for person identification in the literature. In this paper, we propose a new approach to activity-based person identification. Given each activity video clip, we first extract the binary body mask in each frame using background substraction. Then, we propose a discriminative sparse projections (DSPs) method to project these body masks into a low-dimensional subspace and cluster them into a number of clusters, and represent each video clip with a histogram feature. Lastly, we propose an orthogonal ensemble metric learning (OEML) method to learn a distance metric to exploit more discriminative information for recognition. Experimental results are presented to show the efficacy of our

E-mail address: eyanhaibin@bupt.edu.cn

http://dx.doi.org/10.1016/j.neucom.2015.11.111
0925-2312/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: H. Yan, Discriminative sparse projections for activity-based person recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.111

Fig. 1. Two activity examples of three persons. The left three images in each row are three frames from the eating activity and the right three images in each row are three frames from the drinking activity, respectively. Image frames in each row are from the same person. The objective of this work is to identify people at a distance from his/her different activities.

proposed approach. Moreover, we also apply our approach to gait recognition and achieve competitive results with the state-of-the-arts, which further demonstrates the effectiveness of our approach.

The rest of the paper is organized as follows. Section 2 briefly reviews related works. Section 3 details the proposed methods. Section 4 presents the experimental results, and Section 5 concludes the paper.

2. Related work

In this section, we briefly review three related topics: (1) gait recognition, (2) human activity recognition, and (3) metric learning.

Gait recognition: A number of gait recognition methods have been proposed in the literature [4,10–16], and these methods can be mainly classified into two categories: model-based and motion-based. Model-based methods explicitly model human body and perform matching in each frame of a gait sequence, and motion-based methods characterize human gait pattern by using a compact representation. Since model-based methods usually require human motion modeling and parameter estimation from gait sequences, which are still challenging for current imperfect vision techniques, motion-based methods are more popular in gait recognition. As we mentioned before, only the walking activity is exploited in gait recognition, and this study aims to identify people from more other general activities besides human gaits.

Human activity recognition: In computer vision, a large number of activity recognition methods have been proposed in recent years [23–32]. Unlike activity recognition which aims to recognize the type of human activity from videos, activity-based human identification is a relatively new research topic, and there has been only a few seminal studies in recent years [33–36]. To our best knowledge, Gkalelis et al. [33] was the first attempt to formally address the problem of activity-based human identification by using fuzzy c-means (FCM) and linear discriminant analysis (LDA). Their method was further evaluated on more activity datasets and encouraging results were achieved to show the feasibility of human identification using activities [34]. More recently, Lu et al. [35] presented a sparse coding method for activity-based human identification. Since the quantization error is reduced, their method achieved better performance than [33]. However, both FCM and sparse coding are not discriminative enough since they are generative methods. Moreover, these methods performed feature quantization in the original feature space, which may not be effective enough because some irrelevant and redundancy information are contained in this space. To address these shortcomings, we propose a discriminative sparse projections (DSPs) method to learn a low-dimensional subspace for feature quantization, so that the irrelevant information of human body masks is discarded in the learned subspace and a discriminative codebook can be obtained for feature encoding.

Metric learning: There have been a number of metric learning algorithms proposed in the literature [37–47]. Existing metric learning methods can be mainly divided into two categories: unsupervised and supervised. Unsupervised methods aim to learn a low-dimensional manifold where the geometrical information of the samples is preserved, and supervised methods aim to seek an appropriate distance metric for classification tasks. While existing metric learning methods have achieved reasonably good performance in many visual analysis applications, these methods usually learn a global discriminative metric for the whole training data at once, which usually suffers from high-dimensional feature representations. To address this, PCA is usually applied to reduce the feature dimensionality before metric learning. However, such a preprocessing may lose some discriminative infirmation. In this paper, we propose a new EOML method to learn multiple projections from randomly sampled subsets of training samples, and orthogonalize these projections and combine them into a distance metric. Hence, no PCA preprocessing is required in our method and our method can perform dimensionality reduction and metric learning simultaneously. Moreover, the basic vectors of our learned distance metric are orthogonal to each other such that they are more compact than those of most existing metric learning methods.

Please cite this article as: H. Yan, Discriminative sparse projections for activity-based person recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.111