Pattern Recognition 47 (2014) 3793–3806

Contents lists available at ScienceDirect

Pattern Recognition

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

Uncooperative gait recognition by learning to rank

Raúl Martín-Félez [a,*], Tao Xiang [b]

[a] Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellón de la Plana, Spain
[b] School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

ARTICLE INFO

Article history:
Received 16 October 2013
Received in revised form
6 April 2014
Accepted 12 June 2014
Available online 27 June 2014

Keywords:
Gait recognition
Covariate conditions
Learning to rank
Transfer learning
Distance learning

ABSTRACT

Gait is a useful biometric because it can operate from a distance and without subject cooperation. However, it is affected by changes in covariate conditions (carrying, clothing, view angle, etc.). Existing methods suffer from lack of training samples, can only cope with changes in a subset of conditions with limited success, and implicitly assume subject cooperation. We propose a novel approach which casts gait recognition as a bipartite ranking problem and leverages training samples from different people and even from different datasets. By exploiting learning to rank, the problem of model over-fitting caused by under-sampled training data is effectively addressed. This makes our approach suitable under a genuine uncooperative setting and robust against changes in any covariate conditions. Extensive experiments demonstrate that our approach drastically outperforms existing methods, achieving up to 14-fold increase in recognition rate under the most difficult uncooperative settings.

1. Introduction

Gait can be used as a behavioral biometric. Compared to physiological biometrics such as fingerprint, iris and face, it has a number of distinctive pros and cons. The key advantage of gait for person identification is that it can operate from a distance and without subject cooperation. This makes gait ideal for situations where direct contact with or cooperation from a subject is not possible, e.g. surveillance in a public space. However, having uncooperative subjects also means that gait is susceptible to changes in various covariate conditions, which are circumstantial and physical conditions that can affect either gait itself or its perception. Examples of these conditions include clothing, surface, load carrying (e.g. carrying a bag), camera view angle, walking speed, and footwear type. This problem is illustrated in Fig. 1, which shows that due to significant changes in covariate conditions, especially view angle and clothing, features of different people (Fig. 1(a), (d)) can be more alike than those of a same subject (Fig. 1(a)–(c)).

As a classification problem (i.e. each person being a different class), gait recognition is challenging. This is not only due to the variable covariate conditions mentioned above, but also because of the lack of training data to cope with the large overlap between classes in the feature space. Specifically, each subject may be captured only in one sequence with a handful of gait cycles for feature extraction, resulting in an extremely under-sampled class distribution. Most existing approaches focus on extracting and selecting the best gait features that are invariable to different conditions [1–3]. However, they are based on human *a priori* knowledge (e.g. the most reliable features are in the most dynamic part of human body, i.e. legs) and select features in the highly overlapped original feature space, which only lead to very limited success. In addition, these methods are designed for addressing specific types of covariate conditions but none of them can cope with large view angle changes. On the other hand, since gait features are particularly sensitive to view angle changes, completely different approaches based on feature transformation [4,5] are developed to deal with the view problem, which in turn do not work on other covariate conditions. Affine moment based features that are invariant to unknown covariant condition changes are proposed in [6]. However, it requires a cooperative setting, relies on clean silhouettes to be extracted from images, and is unable to cope with drastic appearance changes. So far, none of the existing approaches can address all covariate conditions which typically co-exist under an uncooperative setting.

Different from those feature selection and transformation-based methods, some learning-based approaches are also proposed [7–9]. These methods attempt to maximize the inter-class distance whilst minimizing intra-class variations, and can be applied after feature selection/transformation. However, they assume that the same classes/people must be present in both the training and test sets and represented with sufficient samples. Both assumptions are often not valid in practice. More importantly, most existing works use a gallery set composed of gait sequences of people under similar covariate conditions and evaluate their performance on a probe set of possibly different but fixed covariate conditions. They therefore make the implicit

* Corresponding author. Tel.: +34 964 728 358; fax: +34 964 728 435.
E-mail addresses: martinr@uji.es (R. Martín-Félez), t.xiang@qmul.ac.uk, txiang@eecs.qmul.ac.uk (T. Xiang).

http://dx.doi.org/10.1016/j.patcog.2014.06.010
0031-3203/

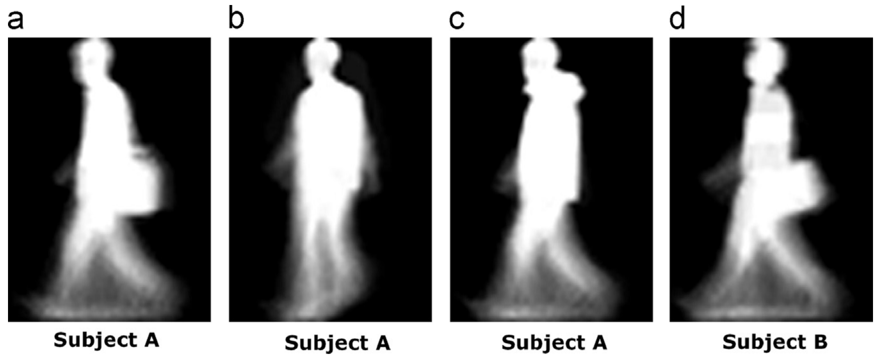

Fig. 1. Comparison of gait representations of Subject A ((a) with a bag, (b) a different viewpoint, and (c) wearing a bulking coat) and Subject B ((d) with a bag). Among (b), (c), and (d), (d) appears to be the best match to (a), because they share the same covariate conditions (view, carrying and clothing), which can easily lead to a wrong match.

assumption that the data are collected in a cooperative manner so that the covariate conditions are known *a priori*. This essentially deprives gait of its most useful characteristic as an uncooperative and non-intrusive biometric.

In this paper, a novel approach is proposed which casts gait recognition as a learning to rank problem – a completely different perspective from previous approaches. More specifically, given a training and a test datasets consisting of gait features of different people who may even be captured from a completely different scene, we learn a bipartite ranking model. The model aims to learn a ranking function in a higher dimensional space where true matches and wrong matches become more separable than in the original space. The output of the model is a ranking function which gives a higher score if a pair of gait feature vectors belong to the same person than to different people. This new formulation has three distinctive advantages over the previous ones: (1) This model is data-driven and can address all covariate conditions including view, i.e. one model for all. (2) Critically, unlike most previous approaches, it does not make any assumption about the gallery and probe sets having the same covariate conditions, either within each set or across the two sets. This makes it particularly suitable for uncooperative person identification, where gait should be used. (3) It does not suffer from the class under-sampling problem. Specifically, since it is based on bipartite ranking, there are only two classes during training: true matches and wrong matches; this also means that gait features from different people captured in different scenes/datasets can be used for training. In essence, it performs cross-class and cross-dataset transfer learning and is able to learn from an auxiliary dataset where plenty of data are available. We assume those data contain the covariate conditions we are modeling, but we do not assume that we know which particular gait sequence contains which covariate (uncooperative setting).

Extensive experiments have been conducted on three benchmark large gait datasets, covering both indoor and outdoor environments. They assess effects of changes in a number of covariate conditions (view angle, surface, carrying conditions and clothing changes) either alone or in combination under both uncooperative and cooperative settings. Results probe that our approach significantly outperforms other contemporary methods, especially under the most demanding uncooperative gait recognition tasks, where an up to 14-fold increase in recognition rate is observed. In addition, our framework is shown to be effective regardless which gait representation is chosen.

2. Related work

2.1. Gait representations

Most existing gait recognition techniques extract gait information from silhouettes obtained from video sequences. One of the simplest yet effective representations is Gait Energy Image (GEI) [7] (see Fig. 1), which is obtained by averaging silhouettes across a gait cycle. However, it has been shown to be sensitive to various covariate conditions [7,10]. To overcome this problem, a number of variants of GEI have been proposed. Yang et al. [3] propose to enhance the dynamic regions of GEI, which are located by a variance analysis. Bashir et al. [1] present a method to distinguish the dynamic and static areas of GEI by using Shannon entropy at each GEI pixel, resulting in a new gait representation called GEnI. Shing and Biswas [2] improve the construction of GEI by using sway alignment instead of upper body alignment, which favors the perception of dynamic information. The basic idea of these methods is to select GEI features from the most dynamic areas of human body, i.e. legs and arms, which are less affected by changes in carrying conditions, clothing, and surface. Various other silhouette-based gait representations have been also developed, including Average Energy Image (AEI) [11], Frame Difference History Image (FDHI) [12] and an optical flow based representation [13]. Most of the recently proposed gait representations are designed to be insensitive against certain covariate condition changes. However, none of them is capable of coping with all covariate conditions since there are so many of them and each one has effects on different aspects of gait [13]. The framework proposed in this paper can improve the recognition performance of any gait representation regardless whether they are designed to be invariant to different covariate condition changes or not, as demonstrated by our experiments (Section 4.3.1).

2.2. Gait feature selection and transformation

Given a gait representation, recognition can be performed by template matching, i.e. using the one-nearest-neighbor (1NN) classifier based on a certain distance metric. However, to alleviate the effects of various covariate conditions, existing approaches have exploited feature selection and transformation. Feature selection methods such as [1–3] select those features from a gait representation that are more invariant to a given covariate condition. Nevertheless, selecting features in the highly overlapped original feature space typically relies on human a priori knowledge (e.g. the most reliable features are in the most dynamic parts of the human body) which only leads to limited success. Others propose to transform the features. On the one hand, some methods perform transformation to represent unknown gait conditions to recreate known covariate conditions. This is usually the preferred method to deal with camera view angle changes [4,5]. Gait features from one view are mapped to another by a learned View Transformation Model (VTM). Recognition is then performed after different views are transformed to the same. A different method is proposed by Bashir et al. [14] which does not reconstruct gait features in different views, but models their