Image and Vision Computing 32 (2014) 1117–1126

Contents lists available at ScienceDirect

Image and Vision Computing

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

Attribute-based learning for gait recognition using spatio-temporal interest points ☆

Worapan Kusakunniran

Faculty of Information and Communication Technology, Mahidol University, Thailand

ARTICLE INFO

Article history:
Received 22 February 2014
Received in revised form 28 September 2014
Accepted 14 October 2014
Available online 1 November 2014

Keywords:
Gait recognition
Human identification
Spatio-temporal
Attribute-based learning

ABSTRACT

This paper proposes a new method to extract a gait feature from a raw gait video directly. The Space–Time Interest Points (STIPs) are detected where there are significant movements of human body along both spatial and temporal directions in local spatio-temporal volumes of a raw gait video. Then, a histogram of STIP descriptors (HSD) is constructed as a gait feature. In the classification stage, the support vector machine (SVM) is applied to recognize gaits based on HSDs. In this study, the standard multi-class (i.e. multiple subjects) classification can often be computationally infeasible at test phase, when gait recognition is performed by using every possible classifiers (i.e. SVM models) trained for all individual subjects. In this paper, the attribute-based classification is applied to reduce the number of SVM models needed for recognizing each probe gait. This process will significantly reduce the test-time computational complexity and also retain or even improve the recognition accuracy. When compared with other existing methods in the literature, the proposed method is shown to have the promising performance for the case of normal walking, and the outstanding performance for the cases of walking with variations such as walking with carrying a bag and walking with varying a type of clothes.

1. Introduction

Gait recognition has been studied broadly in a surveillance perspective because it is capable of identifying humans at a distance by inspecting their walking manners. A large number of gait recognition methods have been developed and published, which can be roughly divided into two categories: 1) model-based methods [1–10]; 2) model-free methods [11–22]. These methods require a pre-processing to extract gait appearances such as shape contours, silhouettes, skeletons, and body joints for further gait analysis. This pre-processing will consume additional computing time. It may also extract incomplete gait appearances which will negatively influence performances of gait analyses in later stages. Also, these gait appearances are sensitive to variations and partial occlusions caused by several factors such as carrying a bag and varying a type of clothes [23].

1.1. Related works

The model-based methods generally aim to model kinematics of human joints in order to measure physical gait parameters such as trajectories, limb lengths, and angular speeds [24]. For example, Cunado et al. [3] modeled legs as an interlinked pendulum. Then, a phase-weighted Fourier magnitude spectrum was used to recognize gait features which were derived from frequency components of changes in an inclination of human thighs. Bobick and Johnson [4,10] used activity-specific static body parameters for gait recognition without directly analyzing dynamics of gait patterns. Kovac and Peer [2] proposed a skeleton model-based gait recognition method by modeling gait dynamics and eliminating influences of subjects' appearances on recognition. Tafazzoli and Safabakhsh [8] applied active contour models and the Hough transform to model movements of articulated parts of the human body. Bouchrika and Nixon [5] used elliptic Fourier descriptors to extract crucial features from human joints. However, the methods in this category have to deal with localizations of the human joints, which are not robust on a markerless motion [7]. It is also difficult to extract underlying models from gait sequences [15].

The model-free methods typically analyze gait sequences without explicit modeling of the human body structure [24]. The different methods in this category have been developed from different perspectives. For example, BenAbdelkader et al. [21] proposed an EigenGait method using image self-similarity plots. Chai et al. [2] introduced a Perceptual Shape Descriptor technique for recognizing gaits. Tan et al. [25] used eight kinds of projective features to describe human gait and PCA was applied for gait feature dimension reduction. Han and Bhanu [11] proposed a concept of Gait Energy Image (GEI), and combined real and synthetic templates to improve the accuracy of gait recognition. Liu et al. [12] employed a population Hidden Markov Model (pHMM) to model human gaits and generated dynamics-normalized stance-frames to recognize individuals. Recently, Wang et al. [13] developed a temporal template, named Chrono-Gait Image (CGI), by encoding gait contours using a multi-channel mapping function. Roy et al. [14]

☆ This research project is supported by Mahidol University.
E-mail address: worapan.kus@mahidol.ac.th.

http://dx.doi.org/10.1016/j.imavis.2014.10.004
0262-8856/

modeled a gait cycle using a chain of key poses which were then averaged to generate the gait feature, called Pose Energy Image (PEI).

1.2. The proposed framework

Without any pre-processing on a raw gait video, the proposed method is to extract and recognize gait features directly from the video on a spatio-temporal feature domain. In this paper, the pre-processing is referred only to a foreground–background segmentation and/or a shape contour/boundary/skeleton extraction. Fig. 1 shows the framework of the proposed method for gait recognition (the detailed frameworks of its key processes will be shown in Figs. 2, 3, and 4). In these figures, rectangles represent inputs/outputs, while ellipses represent processing steps.

In Fig. 1, gait recognition is to find the best matched identity of a probe gait against other gaits in a gallery dataset. First, spatio-temporal interest points (STIPs) are detected from each gait video individually. STIPs are local structures in a spatio-temporal domain where image values have significant local variations in both space and time. These variations are linked to significant movements of human body in a gait video. Therefore, STIP is an interest point of a dominant walking pattern, which can be used to represent characteristics of each individual gait.

Second, a histogram of STIP descriptors (HSD) is constructed as a gait feature. Each STIP descriptor is computed by applying a histogram of oriented gradients (HOG) and a histogram of optical flow (HOF) on a 3D video patch (i.e. width × height × time) in a neighborhood of each detected STIP. It well describes walking patterns around the interest point in space and time.

Third, a support vector machine (SVM) [26] is used as a classification method. SVM has been employed to classify gaits efficiently [17,27–30]. However, SVM has been applied based on the standard one-vs.-all strategy [31] which can be computationally infeasible at a test phase in a large scale multi-class classification (i.e. a large number of subjects). This is because its test-time complexity grows linearly with a number of subjects.

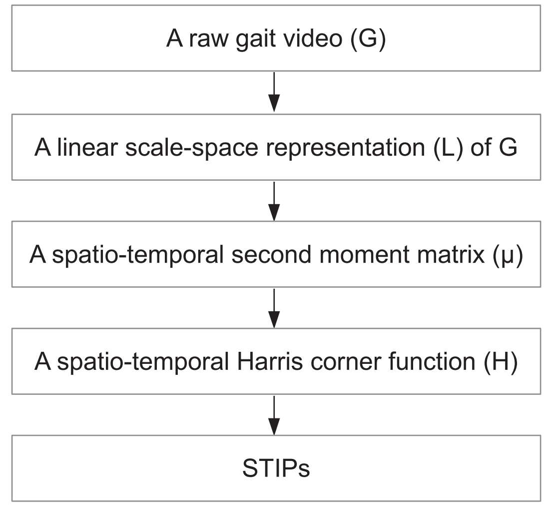

Fig. 2. The proposed framework of the key point detection in a gait video.

Recently, several tree-based methods [32–37] have been proposed to address sublinear testing cost for large multi-class tasks. However, they are limited to several practical constraints [38] including: 1) the tree-based methods cannot be further speeded up by using parallel computing which is a very important and useful methodology to solve a computational problem in real applications; 2) the tree-based methods consume an expensive memory for storing and loading all classifiers at all nodes of the tree in the test phase.

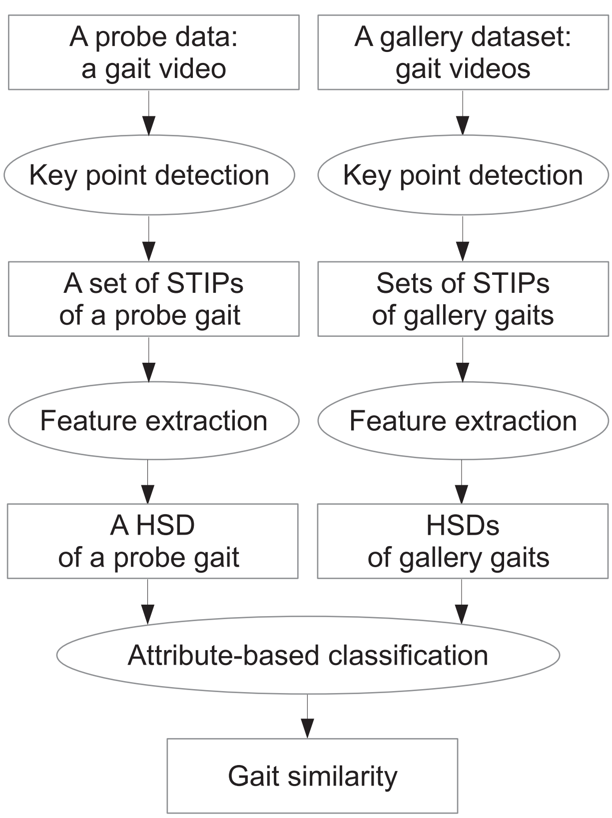

Fig. 1. The proposed framework of gait recognition.

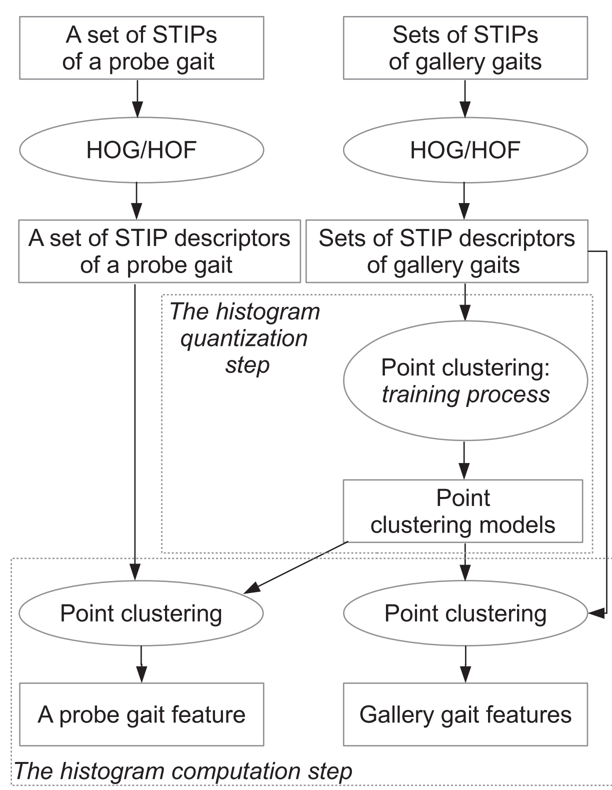

Fig. 3. The proposed framework of the gait feature extraction.