Pattern Recognition Letters 78 (2016) 56–63

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Human gait recognition based on deterministic learning through multiple views fusion☆

Muqing Deng[a,b], Cong Wang[a,b,*], Qingfeng Chen[a,b]

[a] College of Automation, South China University of Technology, Guangzhou 510640, Guangdong, China
[b] Key Laboratory of Biomedical Engineering in Guangdong, Guangzhou 510640, Guangdong, China

ARTICLE INFO

Article history:
Received 11 December 2015
Available online 27 April 2016

Keywords:
Gait recognition
Deterministic learning
Multiple views fusion
Gait variability

ABSTRACT

Gait characteristics extracted from one single camera are limited and not comprehensive enough to develop a robust recognition system. This paper proposes a robust gait recognition method using multiple views fusion and deterministic learning. First, a multiple-views fusion strategy is introduced, in which gaits collected under different views are synthesized as a kind of synthesized silhouette images. Second, the synthesized silhouettes are characterized with four kinds of time-varying gait features, including three width features of the silhouette and one silhouette area feature. Third, gait variability underlying different individuals' time-varying gait features is effectively modeled by using deterministic learning algorithm. This kind of variability reflects the change of synthesized silhouettes while preserving temporal dynamics information of human walking. Gait patterns are represented as the gait variability underlying time-varying gait features and a rapid recognition scheme is presented in published gait databases. Experimental results show that encouraging recognition accuracy can be achieved.

1. Introduction

Gait is an attractive biometric feature for human identification at a distance. Compared with other biometrics, gait has great prominent advantages of being non-contact, non-invasive, unobvious and low resolution requirement [22]. Gait is also difficult to hide or imitate and has great potential applications in video surveillance [9,47].

Despite that much progress has been made for gait recognition [1,16,19,26,36], most of them depend on only one specific view angle and thus have low recognition rate due to the susceptibility to clothes, background, carrying status, shoe type, walking speed, etc. [41]. Gait characteristics extracted with one single camera are relatively limited and not comprehensive enough to develop a robust recognition system [12]. Based on this assumption, one possible method is to get gait sequences from multiple cameras fixed at different view angles and present a recognition scheme based on multi-view gait characteristics fusion. In fact, multiple cameras are now widely used and needed in real-world surveillance environments.

Zhao et al. [48] set up a human 3D model using video sequences captured by multiple cameras. Static and dynamic features were extracted for gait recognition. Lu et al. [32] operated on binary silhouettes and introduced one multiple views fusion recognition scheme on the decision level based on product of sum rule. Kusakunniran et al. [20] used Singular Value Decomposition (SVD) technique to create a new View Transformation Model (VTM), in which the viewing angles of gallery gait data and probe gait data were transformed into the same view angle. Jeong et al. [18] proposed canonical view synthesis based on planar homograpy. Havasi et al. [11] extracted gait characteristics from video-image sequences based on symmetry algorithm for multiple-camera registration. More recently, techniques based on extend SVM [22], correlated motion regression [23], support vector regression (SVR) [21], canonical correlation analysis (CCA) [3] were proposed for multiple views gait recognition.

A robust gait recognition method based on multiple views fusion and deterministic learning is presented in this paper. First, multi-view fusion is employed to synthesize human gait information from two different views images, which is inspired by Huang and Xu [15]. The synthesized silhouettes make full use of two cameras' silhouettes to dispel the influence caused by walking conditions. In contrast to other multi-view methods [48], our method directly fuse human gait information from different views images and extract gait features from the fusion result silhouettes images.

☆ "This paper has been recommended for acceptance by Egon L. van den Broek."
* Corresponding author. Tel.: +86 20 87114615; fax: +86 20 87114612.
E-mail address: wangcong@scut.edu.cn (C. Wang).

http://dx.doi.org/10.1016/j.patrec.2016.04.004
0167-8655/

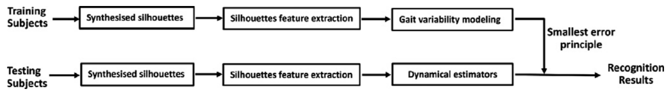

Fig. 1. Flow chart of the proposed method.

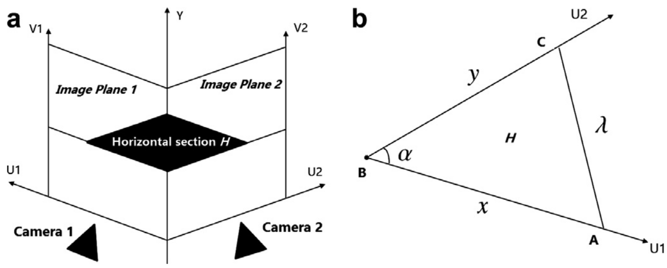

Fig. 2. Illustration of multi-view images fusion: (a)Two cameras configuration; (b)Multi-view images fusion for section projection whose optical axis angle is $\pi - \alpha$.

In our previous work [43,45], we successfully applied deterministic learning algorithm to human gait recognition. Deterministic learning theory is capable of capturing the dynamics information underlying the temporal gait features, which contains more in-depth information and is robust to different walking status or occlusion problem compared with the static features of gait signals [39]. Following this idea, we attempt to capture the gait variability via deterministic learning from multi-view silhouettes. This kind of variability reflects the change of body poses or silhouettes between consecutive frame sequences while preserving temporal dynamics information of human walking. Gait patterns are represented as the gait variability underlying time-varying gait features. Based on the modeling results, a set of dynamical estimators are constructed to represent the training gait patterns. By comparing the set of estimators with a test gait pattern, a set of synchronization(recognition) errors are generated. A test gait pattern can be rapidly recognized according to the smallest error principle. Fig. 1 shows the flow-chart of our method.

2. Feature extraction and representation

Before training and recognition, gait sequences from two different views are synthesized and converted into a sequence of gait feature signals at this preprocessing stage.

2.1. Silhouette extraction

The first step to the proposed method is silhouette extraction. Using the method proposed in [34], silhouettes in a walking sequence are extracted. Mathematical morphology method is used to fill in holes and remove noises. Edge images are produced by applying the Canny operator with hysteresis thresholding. Finally, the body silhouette is determined followed by dilation and erosion. A bounding box is then placed around the part of the motion image and the silhouettes are resized to the same height.

2.2. Multiple views fusion

Two cameras at two different view directions are used in this paper, as seen in Fig. 2a. Image planes (*Image plane* 1, 2) of the two cameras are both parallel to the vertical Y axis. V axis in Image plane is parallel to Y axis as well. For an arbitrary horizontal section H of the human body, segments AB, BC are the projections of H onto *Image plane* 1 and *Image plane* 2, respectively. Angle between the two image planes is denoted by α. Therefore, angle between U1 axis and U2 axis is α and optical axis angle between the two cameras is $\pi - \alpha$. Fig. 2b illustrates the horizontal section projection whose optical axis angle is $\pi - \alpha$, when α is a sharp angle. Lengths of segment AB, BC are denoted by x and y. A new synthesized silhouettes width variable, denoted as λ, can be introduced as follows:

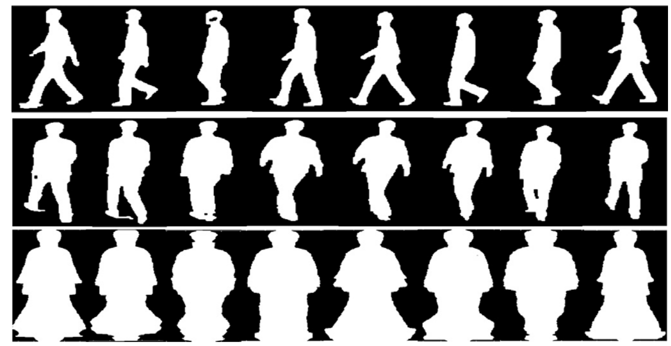

Fig. 3. Synthesizedsilhouette from two cameras(optical axis angle is 60°).

$$\lambda = \sqrt{x^2 + y^2 - 2xy\cos\alpha}, \tag{1}$$

Physically, projections from arbitrary horizontal section onto two different Image planes are synthesized as the variable λ. In other words, human silhouette widths at two different cameras are synthesized and represented as a new variable λ.

When $\alpha > \pi/2$, angle between U1 axis and U2 axis is an obtuse angle. For the reason that only direction angle, regardless of its sign, is considered in projection, the angle between the two image planes α is modified. The modified angle, denoted as β, is recalculated as follow:

$$\beta = \begin{cases} \alpha, & \text{if } \alpha < \pi/2 \\ \pi - \alpha, & \text{if } \alpha > \pi/2 \end{cases} \tag{2}$$

Synthesized silhouettes width variable λ is modified as follow:

$$\lambda = \sqrt{x^2 + y^2 - 2xy\cos\beta}, \tag{3}$$

As shown in Fig. 3, we calculate synthesized silhouette width variable λ for each horizontal section and form the so-called synthesized silhouette image. The first two rows are walking silhouette sequence for one person recorded in two different cameras. The last row is a sequence of the synthesized silhouette, which reflects quasi-periodic characteristics of human gait. As gait characteristics with two degrees of freedom (vertical-axis and horizontal section) are involved, synthesized silhouette contains more information than original silhouette.

2.3. Gait feature extraction and representation

2.3.1. Extraction of width features from synthesized silhouette

Width of the silhouette can be defined as the distance between left and right extremities of the silhouette, which has been proved as a good method for representing two-dimensional silhouette [19,26,45]. Width feature contains structural as well as dynamical information of gait.

Left and right boundaries are traced from the synthesized silhouette. The gait synthesized silhouette is divided into four equal regions from top to bottom, namely region 1, region 2, region 3 and region 4, as shown in Fig. 4. X axis denotes the row index and Y axis denotes the width associated with that row. H is the height of the silhouette. Along a given row, the width is the difference between leftmost and rightmost boundary pixels in that row. (X, Y) denotes an arbitrary pixel in the silhouette image. Y_X^L and Y_X^R represent the Y-coordinates of the leftmost and rightmost boundary