J. Vis. Commun. Image R. 39 (2016) 209–217

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Relative distance features for gait recognition with Kinect ☆

Ke Yang *, Yong Dou, Shaohe Lv, Fei Zhang, Qi Lv

National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, China
College of Computer, National University of Defense Technology, Changsha 410073, China

ARTICLE INFO

Article history:
Received 30 June 2015
Revised 24 March 2016
Accepted 30 May 2016
Available online 31 May 2016

Keywords:
Relative distance feature
Gait feature
Anthropometric features
Gait recognition
Biometrics

ABSTRACT

Gait and static body measurement are important biometric technologies for passive human recognition. Many previous works argue that recognition performance based completely on the gait feature is limited. The reason for this limited performance remains unclear. This study focuses on human recognition with gait feature obtained by Kinect and shows that gait feature can effectively distinguish from different human beings through a novel representation – relative distance-based gait features. Experimental results show that the recognition accuracy with relative distance features reaches up to 85%, which is comparable with that of anthropometric features. The combination of relative distance features and anthropometric features can provide an accuracy of more than 95%. Results indicate that the relative distance feature is quite effective and worthy of further study in more general scenarios (e.g., without Kinect).

1. Introduction and related works

Accurate and effective human recognition is a major research area of computer vision, pattern recognition, machine learning, biometrics, and intelligent surveillance. Face [1] and fingerprint recognition [2] technologies have been widely used in commercial and forensic fields. However, such biometrics need to be detected with subjects' cooperation and require high image resolution. A relatively recent trend in biometrics is performing person recognition using full-body characteristics [9]. In contrast to traditional identification technologies, human recognition technology using full-body characteristics can be used from a distance and does not need subjects' cooperation. A large number of surveillance systems built to improve safety and security have been developed recently. Many of these systems include human recognition capabilities, and recognition method using full-body characteristics may be the most suitable one. A surveillance video mainly aims to monitor people; however, most surveillance cameras can only collect low quality videos, and traditional biometrics may be concealed. Low quality videos can still provide sufficient data for the full-body recognition technology, and studies using this biometrics as a forensic tool exist [19]. Full-body characteristics include static body measurement and gait features [9,17].

Static body measurement refers to static body shape [17] or anthropometric feature [9]. Araujo et al. [14] extracted 11 anthropometric features, including skeleton length and height to recognize different people. Gait features model the walking pattern of people. Existing gait recognition methods can mainly be divided into two groups: model-free and model-based technologies [20]. Model-free gait recognition technologies include a variety of silhouette-based methods. Phillips et al. [18] proposed a baseline algorithm by using the correlation of silhouettes. Gait Energy Image (GEI) [3] uses the average image of silhouettes in a gait period to characterize gait features. GEI has become one of the most common methods in recent years because of its simple representation and computation effectiveness. However, GEI's performance could be affected by silhouette quality, namely, segmentation errors. According to Ref. [3], GEI is insensitive to silhouette segmentation errors. But in Ref. [20], the author argue that it is only true for pixel-wise independent random noise where false and miss detection of foreground are with the same probability. When it comes to non-random noise where background and foreground are similar, the segmentation will always be wrong, so GEI is significantly affected by the silhouette quality. An enhanced version of GEI, namely, alpha-GEI is proposed to mitigate such a non-random noise in Ref. [22]. Ref. [23] improves the silhouette quality and recognition accuracy by using standard gait models as prior knowledge. Ref. [24] extends the quality metrics and proposes a novel way of using quality metrics to improve the segmentation pre-processing step. Gait Entropy Image (GEnI) [4] and Chrono-Gait Image (CGI) [5] are also silhouette-based gait representations

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, China.
E-mail address: bityangke@163.com (K. Yang).

http://dx.doi.org/10.1016/j.jvcir.2016.05.020
1047-3203/

proposed based on GEI [3]. GEnI [4] mainly has encoding dynamic gait features and is robust to variable factors, such as clothes and backpacks. CGI [5] uses color to encode gait features to store temporal information. Model-free gait recognition method possesses simple representations, extractable features, and low computational complexities, but addresses viewpoint and occlusion problems poorly [20]. Therefore, there are plenty of works which enhanced the model-free method to handle cross-view [25–28] and occlusion problems [29,30]. Model-based gait recognition methods model the structure of the human body by using body structure parameters [20]. Cunado et al. [6] developed an early model-based method called the pendulum model. The thigh modeled as linked pendulum and gait features are extracted from the frequency component of inclination angle signals. Stick model [7] extracts head, neck, shoulder, chest, pelvic, knee, and ankle positional parameters in the body silhouette according to the knowledge of anatomy and then calculates the kinematic characteristics of various locations to construct the 2D human stick model. All stick models in one gait sequence are linked together to form the gait pattern for recognition.

Many previous studies argue that recognition performance based completely on the gait feature is limited [9,16,17]. Lombardi et al. [17] proposed a two-point gait that encodes a complete human dynamic gait feature and found that recognition ability using a two-point gait separately is limited, but when combined with the body shape features, the recognition accuracy has a significant improvement. The experimental results of Veeraraghavan et al. [16] showed that body shape is more important than gait features in a video-based gait recognition task. Andersson and Araujo [9] also suggested that using static anthropometric features performs far better than using gait features in skeleton-based human recognition. The present study aims to determine whether the distinguishing ability of the gait itself is limited, or existing features do not adequately capture the right features. This study investigates this problem in Kinect-based human recognition.

Model-based gait recognition methods are robust to occlusion and viewpoint but require complex calculations to extract the human body model. Thus, most of the works are based on model-free methods. The emergence of Kinect has changed this phenomenon [8]. Kinect is the motion controller of XBOX-360 released in 2010 by Microsoft. Conventional color image, depth image of the scene, and human skeleton data stream can be simultaneously extracted from Kinect. The skeletal tracking function of Kinect can provide real-time 3D coordinates of 20 human skeleton points, eliminating the need for complex extraction procedures of human model and providing a great convenience for robust gait recognition technology to light, viewpoint. Similarly, full-body characteristics used in Kinect skeleton-based human recognition also include static body measurement and gait features [9].

The most common division between gait features, according to the human gait theory, is temporal parameters, spatial parameters and kinematic parameters [21]. Spatial and temporal parameters are the intuitive gait features including step length, speed, gait cycle, average stride length, and so on. Kinematic parameters are usually characterized by the joint angles between body segments and their relationships to the events of the gait cycle [21]. Preis et al. [8] extracted spatiotemporal parameters – step length and speed to perform the human recognition task. Since both of these features are changing constantly during walking, using these features is easy to confuse different people. Andersson and Araujo [9] extracted the temporal parameters, spatial parameters and kinematic parameters such as step length, cycle time, speed and features extracted from lower limb angles as gait features. Angle-based kinematic features have poor performance because lower limb angles are disturbed by noise. When people walk in parallel to Kinect, the joints of the distant body parts may be occluded and inferred; when the joints used are inferred, extracting kinematic features from lower limb angles as gait features is inaccurate. And the references such as [9] use the angle-based features from both sides. Ball et al. [10] also extracted the lower limb angle-based kinematic features as gait features.

In several works such as [11,12], distance-based gait features are proposed to model the human gait. The statistical characteristics of the distance of one sequence such as mean values are extracted as the gait features, and the distance-based gait features have distinguishing ability to some extent. And the distance-based gait features can be divided into two categories: relative distance-based and absolute distance-based gait features. These two classes of features can also name them as relative distance features and absolute distance features. In the Kinect depth sensor coordinate system [Fig. 1(a)], absolute gait features are the statistical characteristics extracted from the relative coordinate values of specific skeleton point pairs.

Vertical Distance Features (VDF) was developed by Ahmed et al. [11]. VDF describes the statistical characteristics of the absolute coordinate value changes of some joints when walking. In the most common scenarios that people walk parallel to the Kinect depth sensor's coordinate system's x-axis, the curve of skeleton points' absolute coordinate values (in the coordinate system in Fig. 1(a), it represents the x-axis direction) is not periodic, and the curve can be approximately modeled by an n-order trigonometric-polynomial interpolant function according to Ref. [21]. The curves lose the periodic characteristic of the gait and the parameters of the function need a relative long sequence to fit. The curve of the relative coordinate values of specific skeleton point pairs is still periodic, keeping the periodic characteristic of gait. Thus, the relative distance features are more suitable to model the gait from this perspective.

Gianaria et al. [12] extracted spatiotemporal parameters such as step length, speed, relative distance features (such as average distances between two elbows, two hands, and two ankles), and also the absolute gait features extracted from some joints' absolute coordinate changes as gait features. Chattopadhyay et al. [13] used the relative distance features (such as the mean values of relative distance from front elbow, wrist, hand, ankle, and foot to hip center) and hip center's velocity as gait features. Using statistical characteristics of the relative motion of unconnected joints as representation of gait is robust; however, only using the mean value is insufficient. The standard deviation of the relative distances models the amplitude of movement of people when walking. Since the amplitudes of different people' skeleton joints' movements are different, the standard deviation has the discrimination ability to some extent. It could be a feasible way to combine the mean value and the standard deviation to achieve better performance.

In this study, we extract relative distance features (i.e., the mean and standard deviation of relative coordinate values between the joints) as gait features. The recognition accuracy of the proposed robust relative distance features is up to 85%. This accuracy result is comparable with the anthropometric features and does not require calculation of the gait cycle. Moreover, relative distance features and anthropometric features complement each other well. Recognition accuracy of more than 95% can be achieved when these features are combined. The experimental results have verified that gait features own sufficient recognition capability; relative distance features are an effective representation of gait in the Kinect-based scenario, and worthy of further study in general scenarios.

The rest of the paper is organized as follows. The proposed method is presented in Section 2. The experiments and results are described in Section 3. Conclusion is presented in Section 4.