J. Vis. Commun. Image R. 25 (2014) 53–63

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Pose Depth Volume extraction from RGB-D streams for frontal gait recognition

Pratik Chattopadhyay *, Aditi Roy, Shamik Sural, Jayanta Mukhopadhyay

School of Information Technology, Indian Institute of Technology, Kharagpur, West Bengal 721302, India

ARTICLE INFO

Article history:
Available online 4 March 2013

Keywords:
Frontal gait recognition
Microsoft Kinect
Depth registered silhouette
Pose Depth Volume
RGB-D stream
Silhouette
Voxel volume
Key pose

ABSTRACT

We explore the applicability of Kinect RGB-D streams in recognizing gait patterns of individuals. Gait energy volume (GEV) is a recently proposed feature that performs gait recognition in frontal view using only depth image frames from Kinect. Since depth frames from Kinect are inherently noisy, corresponding silhouette shapes are inaccurate, often merging with the background. We register the depth and RGB frames from Kinect to obtain smooth silhouette shape along with depth information. A partial volume reconstruction of the frontal surface of each silhouette is done and a novel feature termed as Pose Depth Volume (PDV) is derived from this volumetric model. Recognition performance of the proposed approach has been tested on a data set captured using Microsoft Kinect in an indoor environment. Experimental results clearly demonstrate the effectiveness of the approach in comparison with other existing methods.

1. Introduction

Like several other biometric based identification methods, gait has been studied extensively as a biometric feature in recent years. An advantage of gait recognition is that, unlike other existing biometric methods like finger print detection, iris scan and face detection, gait of a subject can be recognized from a distance without active participation of the subject. This is because detailed textured information is not required in gait recognition. Capturing position variation of human limbs during walking is the main aim of gait recognition and this can be done using binary silhouettes extracted from images which may not be of very high quality. Over the years, while *model based* and *appearance based* gait recognition has captured significant attention from researchers, *appearance based* gait recognition entirely from the frontal view was not given much focus. Also, use of depth cameras in gait recognition is quite rare.

In this paper, we concentrate on gait recognition from frontal view (*frontal gait recognition*) only. An advantage of frontal gait recognition is that walking videos captured from this viewpoint do not suffer from self-occlusion due to hand swings which prevails in fronto-parallel view. Also, since the camera is positioned right in front of a walking person, videos can be captured in a narrow corridor like situation as well. However, a disadvantage of frontal gait recognition is that binary silhouettes extracted from RGB video frames cannot represent which limb (left/right) of a walking person is nearer to the camera and which one is behind. Thus, pose ambiguity cannot be adequately resolved, leading to incorrect gait recognition. This information deficiency is not present in depth images, where depth values indicate whether the right limb is forward and the left limb is backward or the other way round. Variation of depth in limb positions together with variation of shape is an important element of frontal gait recognition.

Recently developed depth cameras like Kinect [1,2] can efficiently capture the depth variation in different human body parts while walking. But the depth video frames so obtained are quite noisy, as a result of which extracted object silhouettes are not often clean. In contrast, silhouettes extracted from the RGB video frames are much cleaner but shape variation over a gait cycle is not enough for the extraction of useful gait features. In order to capture both color and shape information in a single frame, we combine information from both the RGB and the depth video streams from Kinect to derive a new gait feature. Each silhouette from the depth frame of Kinect is projected into the RGB frame coordinates using a standard registration procedure, thereby forming a silhouette in transformed space which we term as *depth registered silhouette*. Previously, registration of Kinect depth and RGB frames has been used for 3D reconstruction using depth videos captured from multiple views of an object [4]. However, to the best of our knowledge, no gait recognition method exists which fuses both RGB and depth information for deriving gait features. It may be noted that, there is no publicly available frontal gait database with both color and depth video frames of walking persons recorded simultaneously.

---

* Corresponding author.
E-mail addresses: pratikc@sit.iitkgp.ernet.in (P. Chattopadhyay), aditi.roy@sit.iitkgp.ernet.in (A. Roy), shamik@cse.iitkgp.ernet.in (S. Sural), jay@cse.iitkgp.ernet.in (J. Mukhopadhyay).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.02.010

So, we have built a new database using Microsoft Kinect by capturing walking sequences of 30 individuals.

The proposed gait feature is termed as Pose Depth Volume (PDV). It is derived from a partial volumetric reconstruction of each depth registered silhouette. First, a certain number of *depth key poses* are estimated from a set of training samples and each frame of an entire walking sequence of a subject is classified into an appropriate depth key pose. A PDV is constructed corresponding to each such pose by averaging voxel volumes corresponding to all the frames which belong to that pose. Thus, the number of PDVs of each subject is same as the number of depth key poses. Each voxel in a PDV indicates the number of times an object voxel occurs in that position for that particular depth key pose within a complete gait cycle. A classifier is trained with gait cycles of subjects in the training data set and a different gait cycle is used for testing the accuracy of recognition.

The rest of the paper is organized as follows. Section 2 introduces the Kinect RGB-D camera and basic functionality of its different parts. A brief background study on gait is also included in this section. Section 3 illustrates the sequence of steps followed in deriving our proposed gait feature. Positioning of the Kinect camera and construction of the data set together with experimental results are presented in Section 4. Finally, Section 5 concludes the paper and points out future scope of work.

2. Background

2.1. Basics of RGB-D Kinect camera

RGB-D cameras [2,3] are useful for providing depth and color information of an object simultaneously. Kinect, developed by Microsoft, is one such type of camera [1]. It captures depth information through its infrared projector and sensor. The infrared laser emitted by Kinect draws a structured pattern on the object surface. The infrared camera senses the depth from this pattern using a technology which is based on the structured light principle. Apart from the infrared projector and sensor, Kinect also has a color video camera and an array of microphones. The color camera returns RGB frames while the microphone array helps in audio capture. A detailed description of Kinect functionality is given in [1]. For our application, we use RGB and depth video streams from Kinect for deriving gait features.

2.2. Gait recognition literature review

Gait recognition approaches are broadly categorized as *model based* [8,17], *appearance based* [10–12,14–16] and *spatiotemporal based* [13]. *Model based* methods try to capture the kinematics of human motion. Although these approaches are invariant to viewpoint and scaling, the requirements of high computational overhead and very good quality silhouettes have limited their use for practical purposes. In contrast, *appearance based* methods directly extract useful features from binary silhouettes without building any model and the complexity of implementation of *appearance based* approaches is also much less compared to *model based* techniques. Very high quality binary silhouettes is not required in *appearance based* methods and computational burden is also less, making them suitable for practical uses. *Spatiotemporal based* gait recognition considers spatial as well as temporal domain information.

Temporal template based gait feature is a type of *appearance based* gait recognition that is in use more in recent times due to its robust nature against random noise and simplicity in implementation. It started with the development of two features namely, motion energy image (MEI) and motion history image (MHI) [7]. Later gait energy image (GEI), another *appearance based* method, was proposed by Han and Bhanu [10] which compresses one entire gait cycle into a single image by averaging all the silhouettes over the gait cycle. Each pixel in GEI has a gray level value indicating the number of times an object pixel occurs in that position over an entire gait cycle. But, as a result of averaging, two pixels in a GEI may have the same gray levels even if the object pixels in these positions do not occur simultaneously. Thus, GEI captures intrinsic dynamic characteristics of walking with less resolution. Enhanced gait energy image (EGEI) [14], frame difference energy Image (FDEI) [15] and active energy image (AEI) [16] were proposed to address this problem of loss of kinematic information and also to make the feature more robust against different clothing conditions, object carrying conditions, etc.

Dynamic gait motion features can be detected with greater accuracy with pose energy image (PEI) [20]. PEI has been shown to achieve better performance at the cost of greater execution time than GEI and its variants. In PEI, rather than averaging silhouettes over an entire gait cycle, a fixed number of key poses are extracted from the gait cycle and silhouette averaging is done over all silhouettes that belong to each key pose. Basically, these key poses are representatives of the entire gait cycle and are derived in a manner such that significant variations are present between any two key poses. Each pixel gray level in PEI indicates the number of times an object pixel occurs in that position for that particular pose.

Most of the previous approaches used either side view of walking or stereo view from multiple angles. But more emphasis was given on fronto-parallel view rather than frontal view since fronto-parallel walking sequence carries most information about the gait of an individual [6]. Few attempts have been made towards gait recognition solely from the frontal view [18,19]. An approach for *model based* frontal gait recognition using spherical space model is described in [19]. But the only *appearance based* gait recognition from frontal view developed so far is gait energy volume (GEV) [18]. GEV first projects pixels on each depth image frame of an entire gait cycle into world coordinates to create a 3D voxel volume and then computes average of all the voxel volumes obtained over a gait cycle. Since averaging is done over all the frames in a gait cycle, recognizing dynamic variation in walking is affected to a great extent in GEV. Again, as silhouette shapes obtained from depth frames are noisy, directly projecting the pixels of depth frames into 3D loses substantial information about the actual shape of the object.

3. Gait recognition using Pose Depth Volume

In this section, we describe a new feature called Pose Depth Volume (PDV). The applicability of the feature is tested on videos captured by Microsoft Kinect. Instead of using depth videos directly as done in case of GEV, we combine RGB and depth information from Kinect to obtain better silhouettes along with depth information. To capture intrinsic dynamics of gait better than GEV, we divide an entire gait cycle into a number of depth key poses. Averaging of voxel volumes is done over all the frames belonging to a particular depth key pose. The steps followed in deriving the proposed feature are described in detail in the following subsections.

3.1. Combining information from Kinect color and depth video frames

During the recording phase, subjects walk in front of a static background. RGB video frames and corresponding depth video frames are recorded at a fixed rate. The tilt angle of the Kinect and the height at which it is placed are suitably adjusted so that at least one full gait cycle of a walking individual is always captured by both the depth and the RGB cameras of Kinect. An example RGB frame and its corresponding depth frame are shown in Fig. 1(a) and (b), respectively.