

Introduction

Gait recognition refers to the process of identifying individuals from their walking style. Early medical studies [4, 5] indicate that each individual has a distinctive gait pattern if the precise variation of the joint movements during walking is captured and analyzed. However, obtaining this detailed information to identify suspects in surveillance zones such as airports, railway stations, shopping malls, etc., is challenging since these sites are usually overcrowded with uncontrolled movement of people making it difficult to have physical interaction with each and every subject. Other popularly used biometric identification schemes such as finger-print recognition [6, 7, 8, 9] and iris scan [10, 11, 12, 13] will also be ineffective to identify individuals in these situations since these require close interaction with each subject. Even person identification through face recognition [14, 15, 16, 17] is likely to have less application in surveillance sites since the resolution of the face images captured at these sites is usually low due to the long distance of the surveillance cameras from the monitoring zone and also since a person can hide his/her face with a mask or a pair of sunglasses. Gait is the only biometric that is effective in identifying persons from a distance in an unobtrusive manner, and research on Computer Vision-based gait recognition is motivated from the fact that human beings can easily recognize their friends/close relatives from a distance.

Existing research on Computer Vision-based gait recognition utilize silhouette-level information to derive suitable features for gait recognition and do not rely on the availability of precise joint movement information during walking. Most of the early approaches to gait recognition aggregate features from a complete cycle to construct features for recognition. In the later years, several improvements have been made to traditional research on gait recognition such as shifting towards pose-based recognition from cycle aggregation techniques, developing techniques to handle viewpoint variation, co-variate condition changes, and occlusion. However, state-of-the-art research on Computer Vision-based gait recognition suffers from lack of robustness against varying environments such as different walking speeds, co-variate conditions, etc., during training and testing, and are thus not suitable for deployment in real-life situations. In this thesis, we have made attempts to develop new and improved techniques for Computer Vision-based gait recognition with an attempt to extend the state-of-the-art research and make gait recognition more suitable for application in real-life surveillance. We also make a thorough comparative study of our proposed approaches with other recent gait recognition techniques.

1.1 Basics of Gait and Computer Vision-Based Gait Recognition

Gait is a periodic activity in which similar walking stances get repeated over different cycles. A gait cycle may start from any walking pose, covers other poses in a sequential manner, and finally ends with the starting pose. An example of a gait cycle is explained using Figure [1.1](#). As can be seen from the figure, each gait cycle covers two strides: one with right foot forward and the other with left foot forward. Two representative poses in a gait cycle are the double support stance pose and the mid-stance pose. During the double support stance pose, the two feet are farthest from each other and both the feet are grounded whereas during a mid-stance pose the two feet are closest to each other and only one foot is grounded. In any gait cycle there are two double support stance poses and two mid-stance poses and each stride consists of several intermediate walking poses including a double

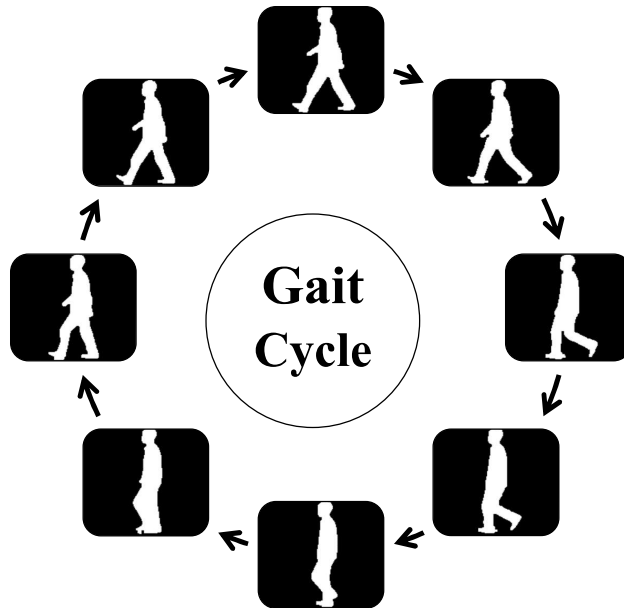


Figure 1.1: An example of a gait cycle with a sequence of binary silhouette images

support stance pose and a mid-stance pose. Since human walking is symmetric in both the halves of a gait cycle, appropriate gait features can be extracted if only a single half cycle of gait is available.

Existing literature shows that majority of research on Computer Vision-based gait recognition focus on deriving appearance-based features that either encode the shape information from the binary silhouettes extracted from the input RGB frames, or exploit the dynamic information available in the sequence to generate features at the granularity of key poses. Ease of computation and high accuracy have made appearance-based gait recognition very popular among the research community. Also, these approaches are generally robust to changes in the color-appearance of an individual since the corresponding features are extracted from a set of cropped and normalized binary silhouettes and not from the input RGB frames. Obtaining the desired binary silhouette frames from the input RGB frames

is done through a set of pre-processing steps and this is usually associated with any appearance-based gait feature extraction algorithm. The steps involved in the pre-processing operation are explained in further detail next.

The extraction of silhouette images from the corresponding RGB images in the presence of a fixed background can be done using any standard background subtraction procedure, e.g., [18]. The process generally includes three basic steps: (a) background subtraction to obtain a difference image, (b) binarization of the difference image with an experimentally determined threshold, (c) cropping the best-fit bounding box around the silhouette and normalizing the box to a fixed width and height. The threshold is chosen suitably so as to get clean silhouette information to the extent possible by avoiding inclusion of background noise and shadows. The third step enables computation of gait features with similar dimensions from each frame of a binary silhouette sequence. The first two rows of Figure 1.2 show the RGB frames extracted from the videos of a walking person, whereas the next two rows show the corresponding silhouette images after background subtraction using a threshold of 0.57 on the difference images obtained by subtracting the background image from the corresponding RGB frames. Since the thesis aims at developing effective pose-based gait recognition algorithms and not object detection, we have relied on simple threshold-based background subtraction which provides fairly good segmentation results for the purpose of gait recognition. Also, most of the datasets used in the thesis for evaluation provide binary silhouettes extracted from the RGB frames and these have been used in our experiments directly. However, more sophisticated recent algorithms such as the one used in [19] could be used to automate the background subtraction process. The final two rows in the above figure correspond to the cropped and normalized background subtracted silhouettes.

1.2 Categorization of Approaches Based on Viewpoint

Based on the camera viewpoint, gait recognition techniques can be broadly classified as fronto-parallel (side) view or frontal (front/back) view gait recognition. For

1.2 Categorization of Approaches Based on Viewpoint



Figure 1.2: RGB to silhouette image, first two rows consist of RGB images, next two rows consist of the corresponding silhouette images before dimension normalization, and the last two rows correspond to the cropped and normalized silhouettes.

front-parallel gait recognition, the line-of-sight of the camera is perpendicular to the direction of motion of the walking persons, whereas for frontal gait recognition the camera line-of-sight is parallel to the direction of motion of walking persons. The first two rows of Figure [1.2](#) show a fronto-parallel view gait sequence, whereas the four rows of Figure [1.3](#) show a sample sequence captured from the frontal view. From the figures, it can be seen that although the swing of the limbs is clearly noticeable for the fronto-parallel sequence, the same is not so prominent for the frontal gait sequence. This can also be verified by comparing the sequences from the two views as shown in Figs. [1.4](#) and [1.5](#). While the periodic pattern of gait is observable from the curves corresponding to the fronto-parallel view (i.e., Fig.

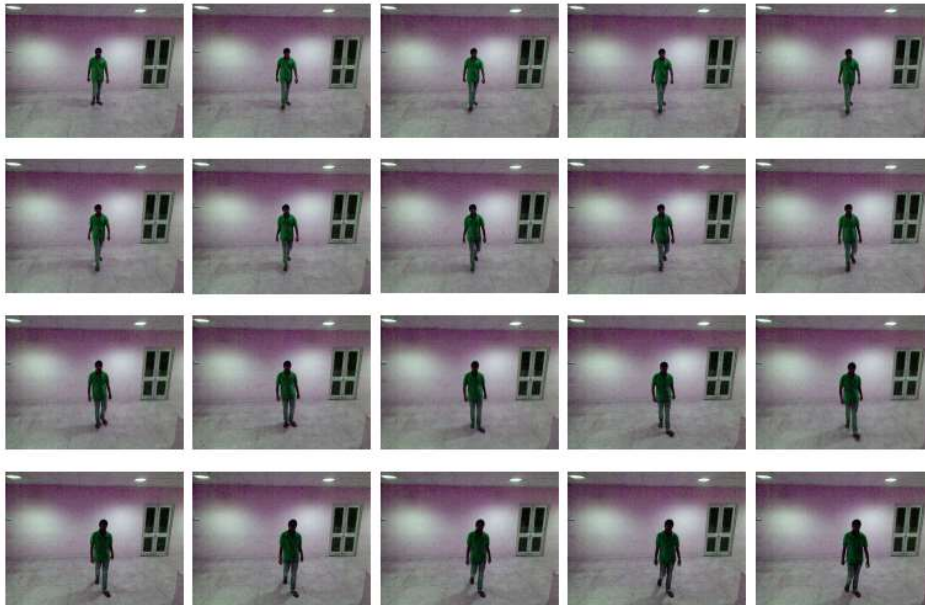


Figure 1.3: A sample RGB gait sequence from the frontal view

1.4), the same is quite difficult to notice from the frontal view (i.e., Fig. 1.5). It has also been mentioned in a few previous studies that walking video from the fronto-parallel view provides maximum information about the gait of a subject.

1.3 Challenges in Computer Vision-Based Gait Recognition

Existing Computer Vision-based gait recognition techniques usually employ binary silhouettes for feature extraction and do not use the RGB frames directly. This is since gait recognition is a long-term process where the clothing conditions of a subject usually alter and distinctive features cannot be derived from RGB information. Thus, accurate extraction of binary silhouettes from the captured image frames through background subtraction is very important for effective gait recognition. Next, we list out a few challenges in research on gait recognition.

1.3 Challenges in Computer Vision-Based Gait Recognition



Figure 1.4: Binary silhouette frames extracted from a fronto-parallel view walking sequence

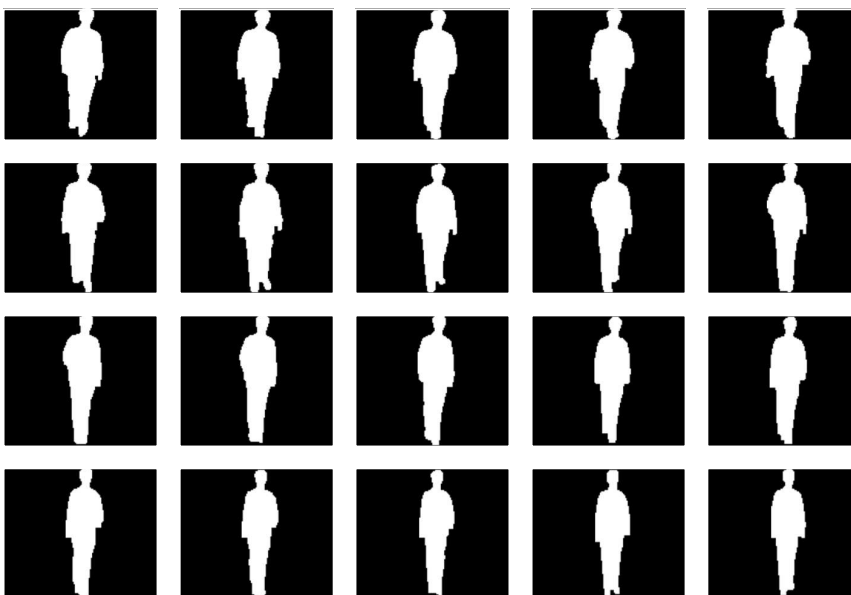


Figure 1.5: Binary silhouette frames extracted from a frontal view walking sequence

1.3.1 Presence of Wearable Objects

The training and test sequences of the same subject may not have similar appearances due to the presence of wearable objects. A person may not carry any co-variate objects during capturing of the training video, whereas he/she may wear a hat, or a coat, or carry a bag during capturing of the test sequence. The presence of co-variate objects distorts the appearances of the binary silhouettes and the gait features thus extracted do not have appearance similarity with that of the gallery features extracted from silhouettes without co-variate objects.

1.3.2 Changes in Camera Viewpoint

The appearances of human silhouettes vary from different viewpoints. Hence, for accurate recognition a pair of gallery and test subjects must be compared using features corresponding to the same viewpoint. If the training and test sequences are captured from different views, a suitable view transformation model needs to be employed to predict the silhouettes from the same viewpoint. Recently, Generative Adversarial Networks are being extensively used to develop these view transformation models. However, there are scopes for improving the existing approaches so as to obtain improved gait recognition accuracy.

1.3.3 Changes in Walking Speed or Camera Frame Rate

Existing pose-based gait features predetermine a set of key poses and next map the frames of an input sequence to the appropriate key poses. These methods are not effective if the walking speed of a person is very fast or the camera frame rate is very low causing the number of frames captured to be less than the number of key poses.

1.3.4 Presence of Occlusion

A gait recognition algorithm works effectively if at least a complete gait cycle is available. However, in real-life, it may not be always possible to capture the frames belonging to a complete cycle due to the presence of occlusion by static and dynamic objects. A few approaches to occlusion handling in gait recognition either

1.4 Contributions of the Dissertation

predict/reconstruct the missing frames from the available unoccluded frames, or match the two sequences based on the common set of key poses present in both the sequences. However, these approaches are not robust against high degrees of occlusion.

1.4 Contributions of the Dissertation

We have made three major contributions in the domain of pose-based gait recognition in this thesis as described in detail in Chapters [3](#), [4](#), and [5](#). Here, we provide a brief summary of our main contributions.

1.4.1 Exploiting Contour Information for Gait Recognition

Existing pose-based gait recognition approaches utilize complete silhouette information to extract the relevant gait features. These binary silhouettes mostly contain shape/structural information pertaining to an individual and do not put too much emphasis on the dynamic information. Rather, features derived from the silhouette contours appear to be more effective in preserving these dynamic features. Through this contribution, we improve upon the existing gait recognition approaches that derive features from the contour information. Popular gait features in this category include Active Energy Image [\[20\]](#) and Frame Difference Energy Image [\[21\]](#), each of which computes the difference between adjacent silhouettes in a sequence to compute a difference image and next aggregates these difference images over a gait cycle. Gait recognition using these features is not expected to be accurate if the differences between the adjacent silhouettes are not substantially high which can happen if the frame rate is very fast or if the person walks at a very slow pace. As an improvement, we propose to extract the boundary pixels of each silhouette instead of differencing adjacent silhouettes so as to retain the contour information at a high resolution. We further compute a pose-based gait feature termed as the *Pose-based Boundary Energy Image* by computing features at the granularity of a pre-estimated key pose set. The proposed feature is time-efficient and has been seen to provide improved accuracy compared to state-

of-the-art techniques on different data sets. The steps of the proposed approach including the key results and a thorough comparative study have been provided in Chapter 3. A research paper out of the initial work has been published in the IEEE International Conference on IOT, Electronics and Mechatronics, and the complete work has been communicated for possible publication in the Springer Sadhana journal.

1.4.2 Using a Dictionary of Key Poses For Improved Pose-Based Feature Extraction

Each of the existing pose-based gait recognition approaches as well as the method used in our first contribution estimate a key pose set with a fixed number of key poses before computing the features for gait recognition with the assumption that the number of frames in a gait sequence is less than the number of elements in the key pose set. However, such an assumption may not hold true always due to varying frame rate or varying speed of walking persons. In our second contribution, we have made an attempt to counter this problem by considering a dictionary of key pose sets with varying number of key poses in each set instead of a single key pose set. As gait features here also we rely on pose-based contour features, but instead of computing a single set of features, we derive multiple sets of features for each of the different key pose sets, which we term as the Generalized Active Energy Image (*GAEI*) features. Further, comparison between the test subject and each of the different gallery subjects is done based on the number of key poses from the different sets where the *GAEI* features closely match. We have done extensive experiments on different data sets to evaluate the gait recognition accuracy using these features and observe that the proposed pose-based gait analysis approach through a dictionary of key pose sets is more accurate compared to the existing pose-based techniques that use a single key pose set and it can be conveniently integrated with any suitable gait template. The related approach along with experimental results have been reported in Chapter 4. A research article out of this work has been published in the Springer Multimedia Tools and Applications journal.

1.5 Organization of the Thesis

1.4.3 Dynamic Gait Energy Image for Gait Recognition

The use of multiple key pose sets, as in the previous contribution, require extraction of multiple features corresponding to each of these key pose sets. It can effectively handle minor changes in walking speed or minor differences in the frame rate of video capture between the training and testing video sequences. However, it also makes the gait feature extraction and recognition process highly time-intensive. Also, determination of the key poses for each set through constrained K-Means clustering following the algorithm in [1] is erroneous if the gallery gait sequences are not properly aligned. Usually this alignment (i.e., choosing the same starting pose for each sequence) is done through standard image processing algorithms and its effectiveness is significantly dependent on the accuracy of silhouette extraction. Manual alignment from a large number of sequences also requires huge time and effort. To overcome these limitations, we propose to extract the maximal unique walking poses from the gait cycles with the maximum number of frames in the gallery set for gait recognition. We also relax the constraint proposed in [1] for mapping of frames to the appropriate key poses by allowing two consecutive frames to get mapped to even non-consecutive key poses. This helps in exploiting the information contained in a single key pose set more effectively and eliminates the need to determine a dictionary of key pose sets. The resulting gait feature derived at the granularity of key poses is termed as the *Dynamic Gait Energy Image (DGEI)*. In the same work, we also use a Generative Adversarial Network (GAN) model to synthetically generate the *DGEI* features without co-variate objects from those constructed using gait sequences with co-variate objects. Extensive experimental studies using the publicly available CASIA B, TUM-GAID, and OU-ISIR TreadMill B data sets verify the effectiveness of our approach and emphasize its superiority over the state-of-the-art techniques. The proposed method along with experimental results are explained in detail in Chapter 5. A research article out of this work has been published in the Elsevier Neurocomputing journal.

1.5 Organization of the Thesis

The rest of the thesis is structured as follows:

Chapter 2 presents a thorough literature survey on traditional and modern gait recognition approaches with descriptions on popular data sets that are used for research on person identification through gait recognition. This chapter also introduces the tools and frameworks used to implement the various gait recognition algorithms in the thesis.

Chapter 3 describes the construction of the Boundary Energy Image (BEI) feature and its extension to Pose-based Boundary Energy Image along with experimental results and comparative study with existing approaches.

Chapter 4 describes an approach for overcoming the shortcomings of existing pose-based gait recognition approaches that pre-determine a set of key poses before computing the gait features, and also construction of new and effective gait features termed *Generalized Active Energy Image (GAEI)* along with experimental results and comparative study.

Chapter 5 describes another improvement over the traditional pose-based features that is more time-efficient compared to the one discussed in Chapter 4 and performs gait recognition satisfactorily even in the presence of co-variate conditions. In this chapter, we also present a comparative study among all the proposed gait recognition approaches in the thesis.

Finally, **Chapter 6** concludes the thesis and points out the future scopes of research in the area of gait recognition.