Proc. of SPIE Conference on Biometric Technology for Human Identification VII, (Orlando, USA), April 2010

Gait Curves for Human Recognition, Backpack Detection and Silhouette Correction in a Nighttime Environment

Brian DeCann and Arun Ross

West Virginia University, Morgantown, West Virginia, USA

ABSTRACT

The need for an automated surveillance system is pronounced at night when the capability of the human eye to detect anomalies is reduced. While there have been significant efforts in the classification of individuals using human metrology and gait, the majority of research assumes a day-time environment. The aim of this study is to move beyond traditional image acquisition modalities and explore the issues of object detection and human identification at night. To address these issues, a spatiotemporal gait curve that captures the shape dynamics of a moving human silhouette is employed. Initially proposed by Wang et al.,[1] this representation of the gait is expanded to incorporate modules for individual classification, backpack detection, and silhouette restoration. Evaluation of these algorithms is conducted on the CASIA Night Gait Database, which includes 10 video sequences for each of 153 unique subjects. The video sequences were captured using a low resolution thermal camera. Matching performance of the proposed algorithms is evaluated using a nearest neighbor classifier. The outcome of this work is an efficient algorithm for backpack detection and human identification, and a basis for further study in silhouette enhancement.

Keywords: gait recognition, object detection, silhouette extraction, gait curves, nighttime biometrics

1. INTRODUCTION

The use of biometric traits to recognize individuals is gaining traction as a legitimate method for establishing identity in a variety of applications.[2] While physical biometric cues such as face, fingerprint and iris are typically used in such applications, recent research has demonstrated the possibility of using ancillary information such as gender, height, weight, age and ethnicity to improve the recognition accuracy of biometric systems.[3] Further, the feasibility of using behavioral cues such as gait, signature and keystroke dynamics has also been explored. Particularly, gait recognition - based on the process of human locomotion - is considered an attractive biometric for surveillance applications primarily because it is non-invasive and can be perceived in low resolutions.[4]

Psychological studies have shown that humans are able to recognize individuals at a distance by their gait.[5,6] As a result, there has been substantial interest in mathematically quantifying an individual's gait. Nearly all methods begin with the extraction of a binary silhouette of the human object from the background scene of a video sequence. The simplest method to accomplish this is background subtraction,[7] though advanced methods include use of gaussian mixture models[8] and fuzzy logic.[9] Should segmentation prove erroneous or difficult, population based hidden Markov models have shown to be capable of silhouette enhancement.[10] Following the extraction of a silhouette, most methods can be categorized as being either model-free or model-based.

Model-based approaches use information collected from known structure of individuals or through models of the human body. Biped models are the most common, but vary on level of complexity and type of information extracted. Features that have been extracted through models include spectra of thigh inclination,[11] thigh rotation,[12] stride and elevation parameters[13,14] and cadence.[15] The primary reason for classifying gait in this manner is that these models allow for robust feature extraction. Since features are collected from a known structure, errors arising from silhouette shape changes are not prevalent. Thus, these approaches are less likely to be affected by errors due to changes in clothing or differences in camera view. However, the increased model complexity and processing requirements may limit the application of these models in real-time environments.

Brian DeCann: E-mail: bdecann@mix.wvu.edu (Contact author)
Arun Ross: E-mail: arun.ross@mail.wvu.edu

Model-free approaches generally aim to extract features based on analyzing a moving shape. Early examples of model-free approaches include spatiotemporal XYT gait signature derivation,[16] principal component analysis[17] and linear discriminant analysis.[18] Additional features such as contour width,[19] procrustes shape analysis,[1] amplitude spectra of key poses,[20] and the direct use of Hidden Markov Models (HMMs)[21] have been proposed. The work by Liu et al.[22] demonstrated that the discriminatory information carried in individual stances can be used to opimize silhouette shape distortion caused by varying walking surface and presence of objects. In addition, their release of the HumanID Gait Challenge established a benchmark for gait related algorithms, experiments and datasets.[23] Recent developments include the Gait Energy Image (GEI)[24] and Head Torso Image (HTI)[25] algorithms which perform template matching of the entire silhouette or specific regions of the silhouette. Pseudoshape representations using a normalized height and width have also been explored.[26] The previous three algorithms are unique in that they were the first to to be used for the problem of gait recognition at night. In general, the primary advantage of a model-free methodology is simplicity, as features are entirely derived from silhouette shape dynamics. However, a commonly cited concern of these methods is their inability to adapt to silhouette variations that may arise as a result of clothing changes or a different camera viewpoint.

Often studied in relation to gait recognition are the problems of human tracking and object detection from video. There have been a number of methods developed for this purpose. W^4 is perhaps one of the most well known trackers of both humans and objects from silhouettes.[27] This work in tracking was significant in that it provided a benchmark for detecting foreground objects, distinguishing people from other moving objects and tracking multiple targets simultaneously. This algorithm was later expanded to include features related to periodicity.[28]

The motivation of this paper is to present a silhouette extraction and matching scheme that can be used for gait recognition in a nighttime environment. Nighttime environments are chosen in this study as they represent practical surveillance scenarios. In addition, the infrared spectrum is a desirable modality for image capture because it is invisible to the human eye, does not induce shadows, and resolved images are for the most partunaffected by additional sources of background radiation. The proposed method will be further advanced to demonstrate how the performance of gait recognition can be increased by incorporating an estimation scheme that detects a backpack on the human subject. Rectification of human silhouettes to account for backpacks is explored because many gait recognition algorithms have shown to degrade in matching performance in the presence of such objects.[14, 24–26] Thus, if it were possible to detect the presence of an object on the human subject, the possibility of estimating the true silhouette shape and improving recognition in a nighttime environment is explored.

In this work the use of gait curves and procrustes distance analysis is investigated for deployment in a nighttime environment. The method is inspired by previous work by Wang et al.[1] Experiments have been conducted on the CASIA Infrared Night Gait Database consisting of 1530 video sequences corresponding to 153 subjects. Video sequences depict subjects walking at a normal pace (without a backpack), slow pace, fast pace, and normal pace (while carrying a backpack).

2. METHODOLOGY

2.1 Overview

The proposed algorithm for human identification and backpack detection contains several stages. Within a video sequence of N frames, consider a specific frame I_k. The first step is preprocessing, wherein image I_k is prepared for segmentation. In the next step a binary silhouette image is produced. With a binary silhouette, static features are derived leading to the extraction of the gait curve. A backpack detection module is then applied, followed by a silhouette correction should a backpack be detected. The end result is a series of spatiotemporal gait curves, which are converted into a single shape for classification.

2.2 Preprocessing and Silhouette Extraction

Thermal imagery differs significantly from that of the visible spectrum. Images are often of low contrast and resolution, and there is little pixel variation between the subject and background. Thus, prior to attempting silhouette extraction, the intensity values of image I_k are adjusted such that 1% of the pixel values are saturated

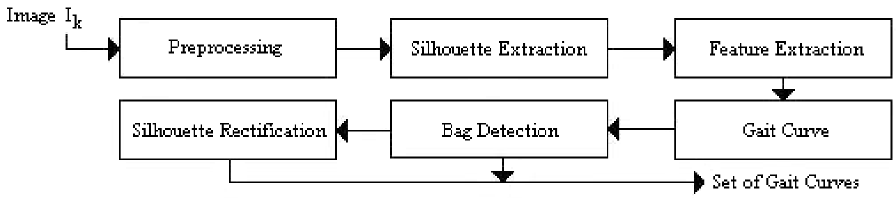

Figure 1. The key stages of the method described in this paper.

at the extrema of 0 and 1, resulting in image I_{adj}. Background subtraction is then used to create a difference image in time. Background noise and additional anomalies are removed through a series of threshold filters and morphological operations. The resulting binary image is denoted by $S = \{s_{i,j}\}$ where $i = 1, 2, \ldots n_h$, $j = 1, 2, \ldots, n_w$, n_h is the height of the image and n_w is the width of the image.

$$I_{diff} = \text{abs}(I_{adj} - I_0) \tag{1}$$

$$S = \begin{cases} 1, & I_{diff} > 0 \\ 0, & I_{diff} \leq 0 \end{cases} \tag{2}$$

Figure 2 illustrates the process of resolving a complete silhouette.

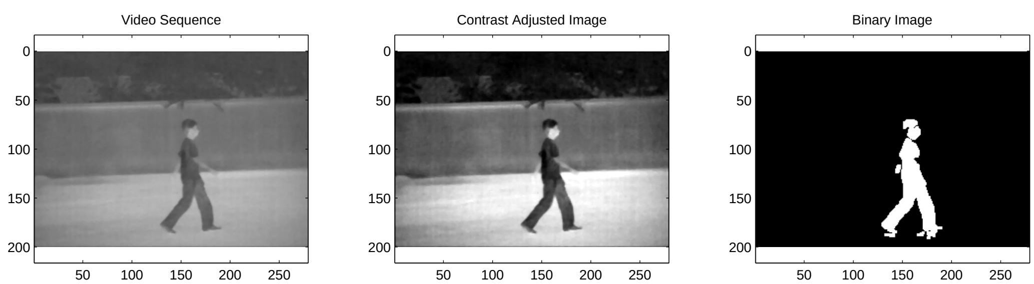

Figure 2. Left: Unprocessed video segment. Center: Filtered and contrast-adjusted image. Right: Resolved silhouette.

2.3 Static Feature Extraction

As a silhouette traverses across the viewing plane, static parameters can be collected at each frame. Raw silhouette height for example, can be extracted by calculating the maximum difference in vertical silhouette coordinates. Let $v_{min} = \{i_{vmin}, j_{vmin}\}$ and $v_{max} = \{i_{vmax}, j_{vmax}\}$ denote the pixels corresponding to the minimum and maximum vertical coordinates, respectively, for which $S(i,j) = 1$. Then, the height h can be simply computed as $h = i_{vmax} - i_{vmin}$.

Since this measure is relative to the distance of a subject from the camera, it cannot be used as a unique feature without knowledge of local markers or subject depth.[29] Despite this shortcoming, raw silhouette height can be used in the design of additional features. The first such use is in isolating the coronal plane of the silhouette. In a matrix coordinate system, v_{min} represents the top of the head. This is fairly consistent across any gait sequence, and can be treated as the peak of the coronal plane. v_{max} however, will shift to either the left or right foot, depending on variations in stance. The terminus of the coronal plane is then located at $v_{ter} = \{i_{vmax}, j_{vmin}\}$. An alternative method to determine the coordinates of the coronal plane would involve computing the centroid, which is the center of mass of the silhouette. However, presence of carried objects, arm sway, or segmentation errors such as holes can greatly distort the horizontal position of the centroid. Thus, identification of the coronal plane using v_{min} and v_{ter} is favored as it is less susceptible to these effects. Refer to Figure 3 for a fully labeled silhouette.

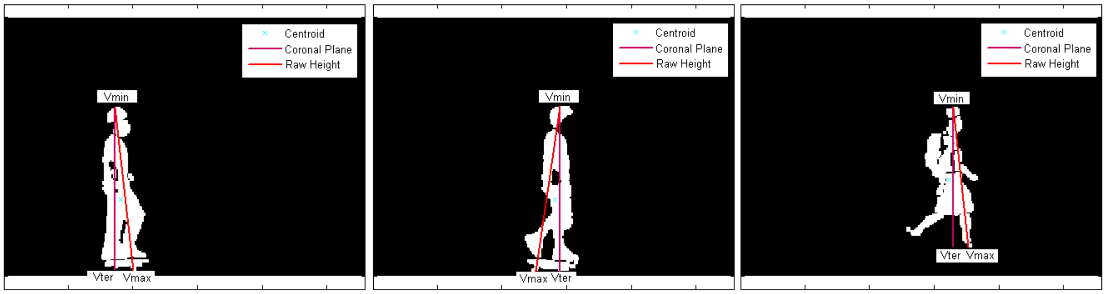

Figure 3. Labeled silhouettes. Note the difference in marking the coronal plane and centroid.

2.4 Spatiotemporal Feature Extraction

Following calculation of the coronal plane coordinates, the left and rightmost pixel locations of the outermost contour are obtained for each row in the silhouette. That is, for the pth row where $p \in [i_{vmin}, i_{vmax}]$ these pixels are denoted as g_p^{left} and g_p^{right}, respectively. Subtraction of the horizontal position of the coronal plane from these point sets yields a space normalized contour, denoted as the gait curve, G_k, for the k^{th} frame in the sequence.

Thus, the evolution of the gait curve across several frames can be regarded as a spatiotemporal feature for shape based analysis. In other words, the output of an arbitrary function $F(G_1, G_2, \ldots, G_N)$ is a single gait curve encompassing the shape dynamics of a particular video sequence. For example, the output of function F could be the mean of the G_k's.

$$F(G_1, G_2, \ldots, G_n) = \frac{1}{N}\sum_{k=1}^{N} G_k. \tag{3}$$

While Equation 3 represents one potential method for representing a set of gait curves, alternative solutions exist as well. The Procrustes meanshape[30, 31] is a mathematically elegant measure of representing and evaluating shape sets. This measure is particularly attractive because subsequent distance comparisons are invariant to translation, rotation and scale. Computation of Equation 3 cannot guarantee these properties. In order to use this measure, the point set in all G_k's must be normalized to contain exactly the same number of points. The set is also vectorized by conversion from spatial coordinates to the complex plane. Subtraction of the vector mean aligns each gait curve at the origin. These operations are summarized in the equations below.

$$\mathbf{z_k} = Re(G_k) + jIm(G_k); \tag{4}$$

$$\bar{\mathbf{z}} = \sum_{i=1}^{k} \frac{\mathbf{z_i}}{k}; \tag{5}$$

$$\mathbf{u_k} = \mathbf{z_k} - \bar{\mathbf{z}}; \tag{6}$$

$$\mathbf{u} = [\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_N}]; \tag{7}$$

$$S_u = \sum_{j=1}^{N} \frac{(\mathbf{u_j}\mathbf{u_j^T})}{(\mathbf{u_j^T}\mathbf{u_j})}. \tag{8}$$

Following the series of N frames, a series of N vectorized gait curves are created. Extracting the first eigenvector of scatter matrix S_u results in the averaged gait curve, denoted as $\bar{\mathbf{G}}$. Here, it should be noted that at least 1 full gait cycle should be completed in order to produce a reliable $\bar{\mathbf{G}}$. For the CASIA Night Gait dataset, the minimum number of frames that produces this result is approximately 10. This constraint is necessary to ensure that enough information has been captured to create a distinguishable $\bar{\mathbf{G}}$.

An agglomeration of N gait curves, as well as the respective average for three different subjects is illustrated in Figure 4. The procrustes distance between any two shape representations, $(\bar{\mathbf{G_1}}, \bar{\mathbf{G_2}})$, is then

$$\mathbf{d}(\bar{\mathbf{G_1}}, \bar{\mathbf{G_2}}) = 1 - \frac{\left|\bar{\mathbf{G_1}}^{\mathbf{T}}\bar{\mathbf{G_2}}\right|^2}{\left\|\bar{\mathbf{G_1}}\right\|^2 \left\|\bar{\mathbf{G_2}}\right\|^2} \tag{9}$$

where, the smaller the resulting value, the similar the shapes.

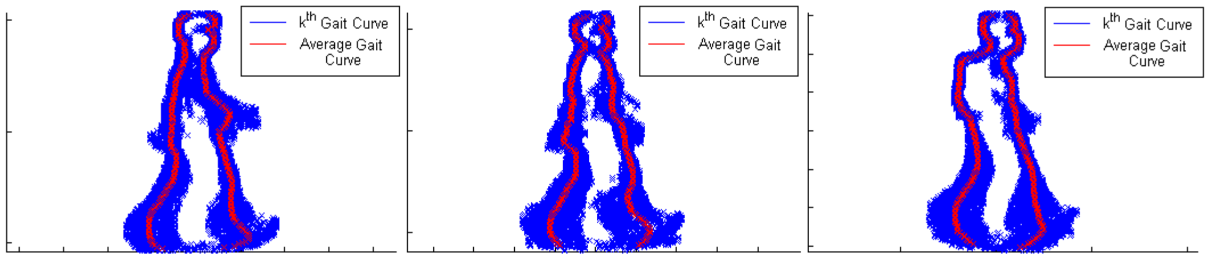

Figure 4. Gait curves for three different subjects. The mean gait curve is included.

2.5 Backpack Detection

The gait curve G can be further exploited by considering it as a one dimensional signal. This is accomplished simply by evaluating the distance between the jth component of each point from the horizontal position of the coronal plane. The end result is a signal of length M that indicates the horizontal distance between each gait curve point to the coronal plane. For the purpose of backpack detection, the back region is of particular interest. If a person is carrying a backpack, it would be expected that the distance of the curve would be greater in the back region than if a backpack were not present. Intuitively this is likely since the presence of a backpack should outwardly distort the silhouette shape. Thus, the area under the curve in the back region should be greater given the presence of a bag. Since the signal has also been previously interpolated to exactly M points, the position of any region of interest in this signal can be considered to be consistent across frames. This allows for estimation of a window where the back region likely exists. However, this signal is also a function of capture depth. To account for this, a normalization factor γ is included to scale the curve according to the waist region. The waist is chosen for its consistency in the gait cycle (i.e., it does not perturb much when the motion is perpendicular in the field of view). Refer to Figure 5 for a labeled example of this signal.

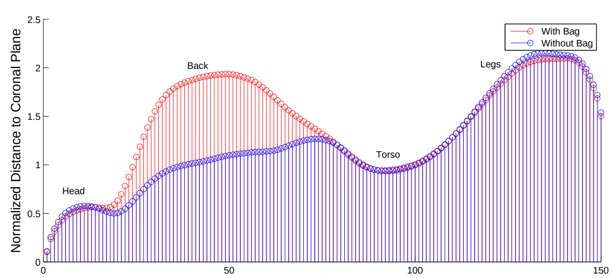

Figure 5. Back half of the 1-D gait curve. Note the back region between values 25 to 75.

In deriving features for backpack detection, let $y[m], m = 1, 2, \ldots, M/2$, denote a signal representation of a gait curve, as shown in Figure 5. Note that this representation is strictly the back half of a particular gait curve. In observing the statistics of Figure 5, the intuitive notion about the silhouette shape for subjects with and without bags is verified. This information also provides the necessary window in which to target backpack related features. In this case, a loosely defined back window, w_{back} is the interval [25,75].