RHYTHMIC GAIT SIGNATURES FROM VIDEO WITHOUT MOTION CAPTURE

Jeffrey E. Boyd

Department of Computer Science
University of Calgary
boyd@cpsc.ucalgary.ca

Akil Sadikali

Department of Computer Science
University of Calgary
akil.sadikali@gmail.com

ABSTRACT

The goal of gait biometrics is usually to identify individual people from a distance, often without their knowledge. As such, gait biometrics provide a source of data that ties a visible pattern of motion to an individual. We describe our work to convert one particular biometric gait signature into a rhythmic sound pattern that is unique for different individuals. We begin with a camera viewing a person walking on a treadmill, then extract a phase configuration that describes the timing pattern of motions in the gait. The timing pattern is then converted to a rhythmic percussion pattern that allows one to hear differences and similarities across a population of gaits. We can also hear phase patterns in a gait independent of the actual frequency of the gait. Our approach avoids the inconvenience and cost of traditional motion capture methods. We demonstrate our system with the sonification of 25 gaits from the CMU Motion of Body database.

1. INTRODUCTION

Gait is ubiquitous: it is important for personal mobility, and we frequently observe the gaits of the people around us. Our observations of gait are usually visual but are occasionally audio. We often feel that we can identify a friend from afar by viewing their gait. Familiar colleagues produce sounds through their footsteps in the corridor that we recognize even when we cannot see them. This paper presents some of our work aimed at finding connections between gait and sound. While our motivation is largely based on curiosity, the conversion of human motion to sound has potential applications in athletics and therapy.

An obvious approach to gait sonification is to start with a motion capture system to acquire temporal signals corresponding to joint trajectories of a person as they move. Motion capture is a well-developed technology, offering accurate joint trajectories in real time. However, motion capture has some disadvantages. Motion capture is expensive (at least with respect to the apparatus we propose in this paper). Video- and marker-based systems require that all motion be performed within the field of view of a set of cameras. Motion that covers large distances requires many cameras resulting in increased costs. It takes time to attach the markers to a subject. An alternative to video and markers is to attach sensors to the body (even more time-consuming than markers), but this can interfere with the motion of a subject and even be dangerous for some athletic activities.

Past interest in gait biometrics suggests methods of analyzing gait without conventional motion capture [1]. This is because the use of markers or sensors on a subject's body would not be practical for biometrics. Furthermore, biometrics by necessity find variations in gait that can identify individuals. Therefore, if a biometric gait recognition can produce data unique to an individual gait, it should also be possible to sonify that data to produce a sound that uniquely corresponds to that gait.

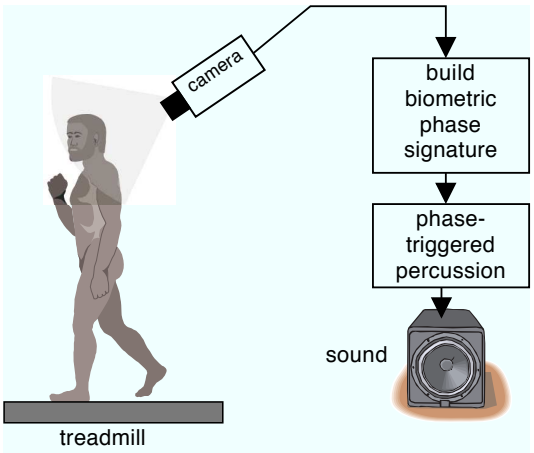

Figure 1: Schematic of phase-locked gait sonification system. The system captures video images of a subject walking on a treadmill and then builds a biometric signature based on the phase configuration of the gait. The system then sonifies the gait signature to allow a clinician and the subject to hear the phase relationships in the gait.

Figure 1 shows the gait sonification system that we propose. A camera views a person walking on a treadmill, but there are no markers or sensors placed on the body. The system builds a biometric signature from which it extracts phase data that describe the relative timing of motions within the gait [2]. As phase signals pass thresholds, the system triggers percussion events to produce a rhythmic portrait off the gait. We demonstrate the system with gaits found in a database intended for testing biometric systems. It is possible to hear the differences and similarities among gaits.

2. BACKGROUND

2.1. Sound and Motion

Many have investigated relationships between human motion and sound. Effenberg [3] and Effenberg and Melzer [4] describe methods for sonification of human motion. They measure the motion of subjects using a variety of methods including a motion capture and pressure-sensitive plates. They display properties of the gait such

as force, velocity, and acceleration of body parts by coding data values to pitch. Higher pitches indicate faster velocities or higher accelerations. Effenberg concludes that augmenting the visual display of the motion with sonified data allowed observers to better estimate some motion parameters. Schaffert et al. [5] examine the use of sonification in training elite athletes. Vogt et al. [6] describe a sound feedback system in which a subject "triggers and controls sound parameters" with movement. Some studies describe the use of musical rhythms to influence gait [7, 8]. In these studies, subjects try to match external rhythms to their gait pattern while an observer records their success. The observations focused only on the heal strike and ignored the remainder of the gait. The analysis and observations relied on manual interpretation of the data.

2.2. Gait Biometrics

Biometric systems extract features from people such as finger prints, patterns in the iris and voice properties, to form a numerical *signature* that is unique to an individual. The signature can therefore be used to recognize individuals or verify their identity. While the primary goal of biometric systems is recognition and verification, the extraction of unique physical features has applications in other areas. Since our goal is to produce sounds that relate to how individual people walk, *gait biometrics* offer methods that can extract from a gait precisely the information we need.

Recent interest in biometrics that can be collected covertly resulted in the publication of a plethora of methods for gait biometrics [1]. This desire for covert acquisition means that gait biometrics do not use markers or sensors placed on the body, as is normally required by a motion capture system. The absence of markers and sensors frees a gait analysis system to be used with more versatility, and at lower cost.

A critical property for the perception of gait, and indeed for producing a gait, is phase locking [9]. This means that the various body parts that are moving periodically in the gait are moving at the same frequency and with a fixed phase difference. For example, the left and right legs operate in opposing phases, the right arm swings in phase with the left leg, and the full extension of the shin (knee lock) normally happens slightly after the forward extension of the thigh. Subtle variations in these phase relationships can provide clues to identity, a fact that is exploited by Boyd [2]. Boyd uses an array of phase locked loops to determine the phase of pixel-intensity oscillations in a sequence of video images of a gait. Given that the pixels are alternately covered and uncovered by body parts as they move through the gait cycle, the phase of pixel intensities is directly related to the phase of motion of the corresponding body parts. The *phase configuration* of a gait acts as a biometric signature for recognition, and can also recognize variations in gait across individuals such as walking on an incline versus on a level surface, and walking fast versus walking slow.

2.3. Gait Databases

The biometrics community has provided several publicly distributed gait databases suitable for testing a variety of gait analysis methods. Among the best known gait databases are the University of California, San Diego [10], Carnegie Mellon University *motion of body* (MoBo) [11], University of Southampton [12], and University of South Florida [13] databases.

We demonstrate our system with the MoBo database. It contains samples for 25 subjects. The subjects walk on a treadmill and are recorded by multiple cameras from different viewing angles. Samples for each subject show walking slowly, walking fast, walking on an incline, and walking while carrying a ball. Each sequence contains images covering 10 seconds of time, sampled at 30 frames per second.

3. GAIT BIOMETRIC SONIFICATION SYSTEM

Our gait sonification system (Figure 1) consists of three parts: video gait capture, computation of the biometric signature, and the sonification of that signature. This section describes these components.

3.1. Video Gait Capture

Our testing was based on the subset of the Mobo database that shows the *fast walk* from the side. The side view best reveals the leg, arm, and body motion in the gait. We restricted ourselves to the *fast walk* samples only so that we would have a consistent way to compare the gaits of the 25 individuals.

While it is possible to compute the biometric signatures from figures against an arbitrary background, an initial figure-to-background segmentation forces the amplitudes of pixel oscillations to be uniform. The MoBo database provides all sequences with segmented figures. If one wishes to move beyond the MoBo samples, chroma-keying or any of a number of background subtraction methods published in the computer vision literature can perform this task.

3.2. Biometric Signature

The biometric signature is the *phase configuration* proposed by Boyd [2], and summarized in the following.

3.2.1. Video Phase Locked Loops

A phase-locked loop (PLL), shown in Figure 2(a), is a control system that synchronizes the oscillations in a *voltage-controlled oscillator* (VCO), u_2, to an incoming oscillating signal u_1. The VCO has a center frequency, i.e. the frequency at which it oscillates when the input is zero. A change in the input to the oscillator changes the frequency of its oscillations. Note that the term *voltage* reveals the PLL's origins in electrical engineering. For our purposes, the oscillator is controlled by a numerical input value.

To understand the role of the feedback in the PLL, suppose that u_1 is a sinusoid at the center frequency of the VCO. The PLL reaches a steady state where u_1 and u_2 have identical frequency and phase. The phase difference, u_d, computed by the *phase detector*, is zero. Ignoring the *loop filter* for the moment, the zero phase difference feeds back to the VCO which continues to oscillate at the center frequency and stays in phase with u_1. Now suppose that u_1 increases in frequency. This will cause the phase difference to increase, and the frequency of VCO to increase until it matches the frequency of u_1. In the new steady state, u_2 has the same frequency as u_1, and the phase difference, u_d is constant. Thus, for a sinusoidal input, the PLL will reach a steady state where the VCO matches the frequency of, and is *phase-locked* with the input. The role of the *loop* filter is to remove high-frequency output from the phase detector that is not related to the phase difference. For our purposes, the PLL is a mechanism to measure and track oscillations in images.

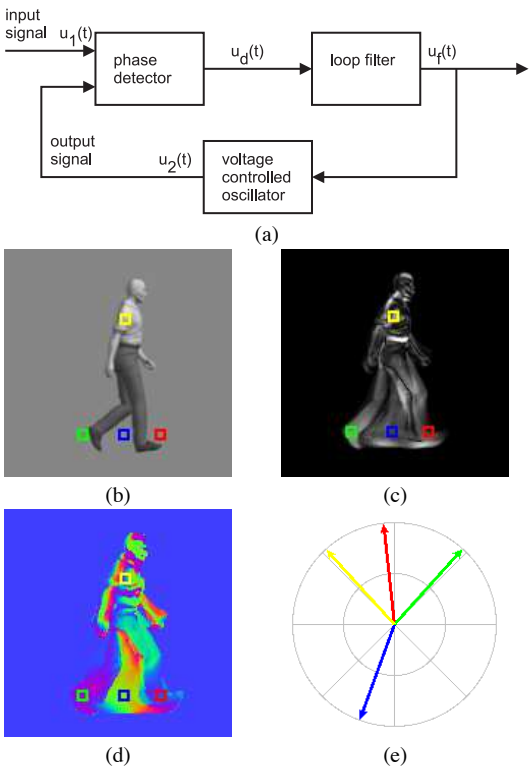

Figure 2: Video Phase Lock Loops: (a) the feedback loop used to lock on to the oscillations in a single pixel, (b) sample raw input image with four points selected, (c) the magnitude and (d) phase of the corresponding oscillations in the video sequence, and (e) the *phasors* (phase vectors) corresponding to the four selected points. As the gait proceeds in time, the phase vectors rotate counter clockwise.

A video phase-locked loop (VPLL) is simply an array of independent PLLs, one for each pixel in a video sequence. Each component PLL locks onto the the oscillations at its position in the image. Since the gaits are themselves phase locked, the component motions of the gait oscillate with the same frequency. Therefore the PLLs in a VPLL all lock to the same frequency, i.e., the fundamental frequency of the gait, and the relative phases of the PLL oscillators are the relative phases of oscillations in the gait image. The array of phase measurements for a video sequence is a *phase configuration* that can be used as a biometric signature.

Figure 2(b)-(e) illustrates the VPLL in operation. Figure 2(b) shows a single frame from a video sequence of a person walking. A VPLL locks onto the oscillations in each pixel to produce two images: a magnitude image (showing the magnitude of the oscillations, Figure 2(c)), and a phase image (showing the relative phases of the oscillations, Figure 2(d)). We can use these images as a whole, or examine the phases at select positions. Figure 2(e) shows *phasors* (phase vectors) for the points delineated in Figure 2(b)-(d), plotted on a unit circle. The phasors rotate with the gait making one rotation per stride in the gait. It is the relative phases that are useful as a biometric, and that we want to hear in our sonification of the gait.

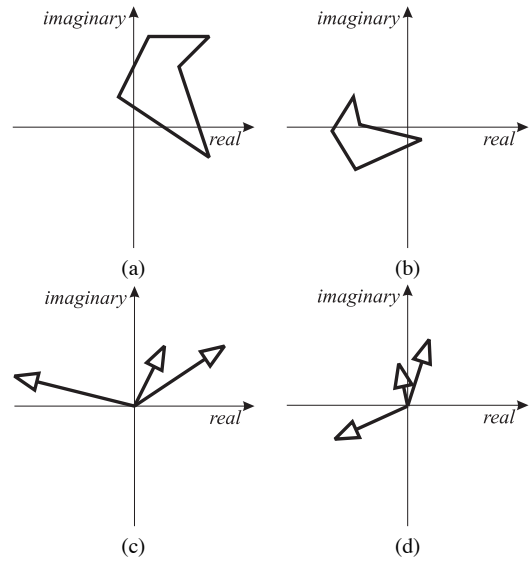

Figure 3: Procrustes analysis applied to shape and phase configurations. In the conventional application, a shape is represented by a vector of complex vertices. The shape in (a) is the same as the shape (b) because one is a translated, scaled, and rotated version of the other. A phasor configuration is also a vector of complex numbers. The configuration in (c) is the same as that in (d) because one is a rotated and scaled version of the other. Rotation is always about the origin so translation is omitted.

3.2.2. Directional Statistics

The VPLL operates an array of PLLs independently. In practice, variations in the position of the walker on the treadmill over time, sporadic errors in background subtractions, and spurious changes in the gait (e.g., as the subject raises an arm) all affect the phase measurements at some pixels. The PLL tracks these faithfully, but when transformed into a sound, we hear the anomalies more than we hear the gait. To prevent this, we *stabalize* the phase configurations by averaging over time. Given that the phases are directions that vary over time, we use *directional statistics*.

Procrustes shape analysis is a method in directional statistics [14] that can summarize (by finding means) and compare (using distance measures) shapes. We can represent a phase or timing pattern in a gait as a set of directions, which is mathematically equivalent to a shape, making Procrustes analysis a useful tool for analyzing the phasor patterns that emerge from gaits.

The following is a summary based on Mardia and Jupp [14]. Describe a shape in two dimensions using a vector of k complex numbers, $\mathbf{z} = [z_1, z_2, \ldots, z_k]^T$, called a configuration. Two configurations, $\mathbf{z}_1$ and $\mathbf{z}_2$, represent the same shape if by a combination of translation, scaling, and rotation, their configurations are equal, i.e.,

$$\begin{aligned} \mathbf{z}_1 &= \alpha \mathbf{1}_k + \beta \mathbf{z}_2, \quad \alpha, \beta \in \mathcal{C} \\ \beta &= |\beta| e^{i \angle \beta}, \end{aligned}$$

as shown in Figure 3(a) and (b). That is, $\alpha \mathbf{1}_k$ translates $\mathbf{z}_2$, and $|\beta|$ and $\angle\beta$ scale and rotate $\mathbf{z}_2$. It is convenient to center shapes by defining the centered configuration $\mathbf{u} = [u_1, u_2, \ldots, u_k]^T$, $u_i = z_i - \bar{z}$, and $\bar{z} = \sum_{i=1}^{k} z_i/k$. We can find the mean of a set of n shapes by finding the μ that minimizes the objective function

$$\min_{\alpha_j, \beta_j} \sum_{j=1}^{n} \|\mu - \alpha_j \mathbf{1}_k - \beta_j \mathbf{u}_j\|^2. \quad (1)$$

To find μ we compute the matrix

$$\mathbf{S}_u = \sum_{j=1}^{n} (\mathbf{u}_j \mathbf{u}_j^*)/(\mathbf{u}_j^* \mathbf{u}_j). \quad (2)$$

The *Procrustes mean shape*, $\hat{\mu}$, is the dominant eigenvector of $\mathbf{S}_u$, i.e., the eigenvector that corresponds to the greatest eigenvalue of $\mathbf{S}_u$.

Although Procrustes shape analysis is intended for treating two-dimensional shapes, it is easily adapted to handling vectors of phasors [2]. A vector of complex phasors, or a phasor configuration, is equivalent to a shape configuration, as illustrated in Figure 3(c) and (d). Shapes are invariant through translation, scaling, and rotation. Translational invariance is achieved by using the centered configuration $\mathbf{u}$. When using phasors the issue of translation becomes irrelevant. All phasors rotate about the origin, $0 + 0i$, at the entrained frequency, and the configurations, $\mathbf{z}$, are already centered, i.e., $\mathbf{z} = \mathbf{u}$.

3.2.3. Mean Configuration as Biometric Signature

In the examples reported later in this paper, we produce a biometric signature that is a mean phase configuration for each subject by the following steps.

1. Allow the VPLL time to lock for the first 100 frames of a sequence.
2. Align the VPLL output for the next 40 frames ($1.3s$ at 30*fps*) so that the oscillating figures have a stationary center.
3. Crop and resample the oscillating region to 21 by 21 pixels. The lower resolution makes the computation of the eigenvectors of $\mathbf{S}_u$ tractable.
4. Compute the Procrustes mean over the 40 21-by-21 configurations by computing the eigenvalues and eigenvectors of $\mathbf{S}_u$.

The end result is a biometric signature that is 441-element complex vector representing the relative phases of pixel oscillations in the observed gait.

3.3. Sound Generation

Figure 4 describes the process by which we convert the gait biometric into percussive sound. First, we expand the biometric signature in time to form a periodic sequence using

$$\theta_i(t) = \left(\phi_i + 2\pi \frac{t}{20}\right) \bmod 2\pi, \quad (3)$$

where ϕ_i is the phase of the ith element of the biometric signature, and $\theta_i(t)$ is the value of the corresponding element expanded at time $t = 0, 1, \ldots, 19$.

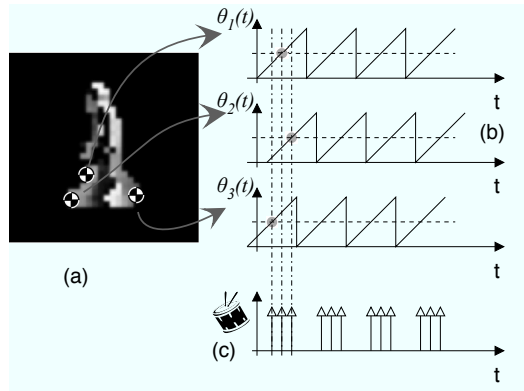

Figure 4: Methods for sonification of the gait signature. (a) The procrustes mean configuration is expanded in time and sampled at selected points. Pixel values correspond to a phase in the range $[0 \ldots 2\pi)$. (b) The temporal signal at the selected points is a phase ramp in time. (c) As each phase signal crosses a phase threhsold, the system triggers a percussion event. The resulting sound is a rhythmic pattern synchronized to the gait and correlated with the individual gait.

When sonifying the biometric signature, we play $\theta(t)$ at 30*fps*. Note that Equation (3) forces the expanded sequence to have a period of 20 samples. Viewed at 30 *fps*, this corresponds to a stride period of $0.67s$, a value within the range typical of human gaits. Consequently, all sonified sequences have the same gait frequency and the similarities or differences we hear are due only to the phase information, not the frequency of the individual gaits. Figure 4(a) shows a single frame from an expanded sequence.

Then next step is to extract the phase at selected positions in the gait. We have no rules about which points to select, but in order to compare gaits, we must be consistent across our set of subjects. We opted for three positions within the biometric signature:

1. the forward extent of the knee motion ($\theta_1(t)$),
2. the forward extent of the foot motion ($\theta_2(t)$), and
3. the rearmost extent of the foot motion ($\theta_3(t)$).

Figure 4(a) shows these positions. Equation (3) forces the temporal signal at the sample points to form a ramp with a period of 20 frames, and relative phases determined by the biometric signature, ϕ, as shown in Figure 4(b).

The last step is to trigger percussion sound events as the $\theta_1(t)$, $\theta_2(t)$, and $\theta_3(t)$ cross a reference phase, Figure 4(c). In this case, the reference phase is π (or 0.5 normalized to the circumference of the unit circle).

4. IMPLEMENTATION AND TESTING WITH MOBO

We implemented the biometric signature computations and the temporal expansion (Equation (3) in Octave (*http://www.gnu.org/software/octave/*). The expanded sequences were then stored as movie clips, each 20 frames in duration, showing a single cycle of the expanded biometric signature.

A Pure Data (PD, *http://puredata.info/*) *patch* does the final steps of the sonification. *Gem* extensions to PD read the 20-frame

clips, playing them cyclically. Custom extensions written in both *C* and *Python* sample the video and implement the phase triggers. The PD patch produces a synthetic drum sound in response to the phase triggers.

We manually selected three positions as described above for each of the 25 sequences from the MoBo database corresponding to a side view of the fast walk. The system allows us to switch subjects (recalling the manually selected positions), experiment with different sample position and sounds, and to record resulting sounds. All of this can be run on a current generation laptop computer with suitable headphones or speakers.

5. DISCUSSION

While experimenting with our system, we identified roughly seven groups of distinct rhythmic audio pattern. The grouping is subjective and we do not know whether or not other observers would make the same groupings. It is, however, reasonable to say that there are distinctive patterns, but that the patterns aren't specific enough to easily resolve all 25 subjects. This is consistent with Boyd's observation that as a biometric, the phase information adds modest but measurable improvements to the recognition of individuals [2].

If we adjust the playback rate of the sequences so that they match the frequency of the original gait, we get a much different impression of the variations in gait. As suggested by Kuo [15], body mass and dimensions affect gait frequency. Therefore, when we extract phase only and ignore the frequency information, we remove an important part of what disambiguates gaits. It is perhaps a strength of our approach that we can separate these aspects of the gait.

Computing a gait signature off-line, then playing back for sonification has limited practical use. For real value, we must compute and sonify in real-time to give immediate feedback to the person walking or perhaps to a clinical observer. While we have experimented with this, it is inherently more difficult to do for the following reasons.

1. A person's position on a treadmill tends to change gradually as they drift forward and backward over time. For real-time sonification, this tracking has to be done reliably in real-time.
2. A mechanism is necessary to find the selected sample positions reliably and continuously in light of the tracking problem mentioned above. Furthermore, variety in body and gait dimensions also confounds automatic selection of sample positions.
3. The stability of the sonified rhythm is improved with Procrustes averaging. As implemented, this is slow and we have reduced the spatial resolution of our data to compensate. Methods for efficient on-line computation of the eigenvalues and eigenvectors would ameliorate this.

Traditional motion capture methods may solve some of these problems, but they do so with high-cost equipment and a loss of convenience.

6. REFERENCES

[1] J. E. Boyd and J. J. Little, "Silhouette-based gait recognition," in *Encyclopedia of Biometrics*, S. Z. Li, Ed. Springer, 2009, pp. 646–652.

[2] J. E. Boyd, "Synchronization of oscillations for machine perception of gaits," *Computer Vision and Image Understanding*, vol. 96, pp. 35–59, 2004.

[3] A. O. Effenberg, "Movement sonification: Effects on perception and action," *IEEE Multimedia*, vol. 12, no. 2, pp. 53–59, 2005.

[4] A. Effenberg and J. Melzer, "Motionlab sonify: a framework for the sonification of human motion data," in *Ninth International Conference on Information Visualisation*, 2005, pp. 17–23.

[5] N. Schaffert, K. Mattes, and A. O. Effenberg, "A sound design for the purposes of movement optimization in elite sport (using the example of rowing," in *International Conference on Auditory Display*, Copenhagen, Denmark, May 2009.

[6] k. Vogt, D. Pirró, I. Kobenz, R. Höldrich, and G. Eckel, "Physiosonic-movement sonification as auditory feedback," in *International Conference on Auditory Display*, Copenhagen, Denmark, May 2009.

[7] J. Hamburg and A. A. Clair, "The effects of a movement with music program on measures of balance and gait speed in healthy older adults," *Journal of Music Therapy*, pp. 212–226, 2003.

[8] M. J. Staum, "Music and rhythmic stimuli in the rehabilitation of gait disorders," *Journal of music therapy*, vol. 20, no. 2, pp. 69–87, 1983.

[9] B. I. Bertenthal and J. Pinto, "Complementary processes in the perception and production of human movements," in *A Dynamic Systems Approach to Development: Applications*, L. B. Smith and E. Thelen, Eds. Cambridge, MA: MIT Press, 1993, pp. 209–239.

[10] J. J. Little and J. E. Boyd, "Recognizing people by their gait: the shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.

[11] R. Gross and J. Shi, "The cmu motion of body (mobo) database," Robotics Institute, Carnegie Mellon University, Tech. Rep. CMU-RI-TR-01-18, June 2001.

[12] J. Shutler, M. Grant, M. Nixon, and J. Carter, "On a large sequence-based human gait database," in *Fourth Int. Conf. Recent Advances in Soft Computing*, Nottingham, UK, 2002, pp. 66–71.

[13] S. Sarkar, J. Phillips, Z. Liu, I. Robledo, P. Prother, and K. W. Bowyer, "The humanid gait challenge problem: data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, February 2005.

[14] K. V. Mardia and P. E. Jupp, *Directional statistics*. Chichester: Wiley, 2000.

[15] A. D. Kuo, "A simple model of bipedal walking predicts the preferred speed-step length relationship," *Journal of Biomechanical Engineering*, vol. 123, pp. 264–269, June 2001.