Oscillating Models for Perception of Human Motion

Jeffrey E. Boyd

Department of Computer Science
University of Calgary
Calgary AB T2N 1N4
boyd@cpsc.ucalgary.ca

Abstract. There are two fundamentally different approaches to machine analysis of human motion: model-based, that use a high-level kinematic model, and model-free, that use low-level representations of the motion. Although each approach has its advantages, there is currently a perceptual gap between the two. This paper describes a new type of kinematic model that enables us to bridge the perceptual gap. The model is the perceptual equivalent of passive mechanical models that walk without any control mechanism. Thus the model not only describes the linkages in a kinematic chain, it also has an innate resonance that is a gait. If we introduce control to force the model to synchronize with low-level oscillations perceived in a video sequence, the result is a gait model that resonates with an observed gait. We describe a system that demonstrates this new model and the connection between the model-free and model-based representations.

1 Introduction

There are two fundamentally different approaches to machine analysis of human motion: model-based and model-free. In a model-based approach, a system fits observed data to a kinematic model of a human. In contrast, model-free approaches interpret the data without a human kinematic model. Computer vision literature contains successful examples of both approaches, but it is clear that neither method is complete, i.e., neither method can perform all the tasks that may be required in visual analysis of human motion. Model-free methods are simpler, faster, and require little or no camera calibration. Model-based methods are usually slow, often require camera calibration, and assume that the correct kinematic model is known *a priori*, but they can address problems that require knowledge of the kinematic structure.

We focus on two requirements for gait perception [1]. The first is frequency entrainment, i.e., the component motions of the gait must oscillate at the same frequency, or integer multiples of each other, and therefore, the perception of the gait must oscillate at an entrained frequency. The second requirement is phase locking, i.e., the component motions of the gait must maintain relative timing, or phase throughout the gait cycle, and therefore perception of the gait depends on the relative phase of the perceived motion. These requirements span all approaches to gait analysis, both model-based and model free.

Although each approach has its advantages, there is a perceptual gap between the two. In this paper, we propose a novel kinematic model that bridges this gap by allowing frequency and phase information to move from model-free to model-based representations. The new model is the perceptual equivalent of passive dynamic walking models described by McGeer [2, 3], Coleman and Ruina [4], Garcia *et al.* [5], and Collins *et al.* [6]. These mechanical systems walk in the absence of any control system because they naturally resonate in a gait. The model dimensions and mass determine the *natural gait*. The resonance of these models implies that each person has their own innate gait determined by their body size and mass. Given that there is psychological and physiological evidence that suggests a relationship between the perception of an activity and the synthesis of the same activity [1, 7], one can hypothesize that people use their innate gait to perceive the gaits of others. This idea provides the inspiration for a machine vision system that has a *walking model* with an innate gait that the system uses to perceive human motion.

We describe a system that demonstrates the innate-gait walking model and the connection between model-free and model-based representations of gaits. The system employs video phase-locked loops (VPLLs), described by Boyd [8], as a model-free source of frequency and phase data from a gait. A VPLL synchronizes an array of internal oscillators to the oscillations of pixel intensities in a video sequence. In turn, the walking model then synchronizes its oscillations to the frequency and phase of the VPLL oscillators. The concept is similar to Laszlo *et al.* [9]. Whereas Laszlo *et al.* describe a gait model that walks on its own but reacts to variations in terrain in order to create a synthetic gait, we are proposing a model that walks on its own but reacts to timing patterns in a visual stimulus in order to perceive a gait.

2 Background

The following summarizes some recent results in human motion analysis, both model-free and model-based. Where methods are specific to human gait, the method always performs some form of frequency entrainment. In some cases, phase-locking also occurs, although this is not consistent.

2.1 Model-Free Methods

Several model-free methods analyze temporal variations in the shape of moving regions. Little and Boyd [10] use optical flow to identify the moving regions in a gait image sequence, then describe the shape of the region with a set of scalar features that oscillate with the gait. The system extracts the relative phases of the scalar oscillations and forms a phase feature vector that is used to identify individual gaits. Cutler and Davis [11] identify periodicities in a vector of intensities for a tracked object. As a periodic motion goes through its cycles, some frames are similar to others. The periodic behavior that arises from these self-similarities allows the system to distinguish between human, animal and mechanical motion. Baumberg and Hogg describe oscillations in the silhouette of human figure with a vibrating plate model [12].

Other model-free methods analyze temporal variations in pixels or small image regions. Polana and Nelson [13] examine oscillations in the magnitude of the optical flow in a sequence containing periodic motion. They compute a coarse resolution flow magnitude image at eight points in the period of the motion. From this they form a 96-element vector that they use to recognize a broad range of periodic motions. Liu and Picard [14] examine oscillations in pixel intensity for a gait sequence using fast Fourier transforms (FFT). Boyd [8] uses a VPLL to synchronize an array of oscillators with the intensity oscillations observed at pixels in an image sequence. The synchronized oscillators yield a complex image representing the magnitude and phase of the pixel oscillations.

2.2 Model-Based Methods

Motion in a kinematic model occur in the joint angle and limb trajectories. The majority of methods in this area are not specific to gait, i.e., they do not exploit the periodic nature of gaits, but take the more general approach of estimating a series of poses that may or may not be periodic. These methods include work by Rowley and Rehg [15], Wachter and Nagel [16], Wren et al. [17], and Bregler and Malik [18]. Fujiyoshi and Lipton [19] use a simplified kinematic model they call a star skeleton. While they estimate the skeleton on a frame-by-frame basis, the skeleton reveals period limb motion. Bissacco *et al.* [20] extract joint angle trajectories from a motion sequence. They then compute an auto-regressive moving-average (ARMA) model of the joint movement which is in turn used as a feature vector. The method is used to recognize different types of gaits such as running, walking, or walking a stair case. Tanawongsuwan and Bobick [21] use joint angle trajectories derived from a motion capture system. The trajectories are synchronized to a common reference point in the gait and then re-sampled so that all subjects have the same number of samples. Trajectories for the various joints are then combined to form a large feature vector used to recognize individual gaits. Bobick and Johnson [22] describe a system that uses static parameters derived from a gait such as stride and torso length.They demonstrate that the system can recognize individuals.

3 Synchronization of Model to Data

With few exceptions, a common need for frequency entrainment and phase locking unifies gait analysis methods. It is this unification that inspires the system introduced here to connect model-free oscillations to a walking kinematic model. This section describes the VPLLs that synchronize oscillators to a video sequence, a simple kinematic model that has an innate gate, and the process that synchronizes the model with the VPLL oscillators.

3.1 Model-Free Timing from Video Phase-Locked Loops

Figure 1 shows a block diagram of a basic phase-locked loop (PLL). Its basic components are a phase detector, a low-pass loop filter, and an oscillator. The phase detector compares the phases of a sinusoidal input, u_1, and the internal oscillator, u_2, yielding the phase difference, u_d. A low-pass filter smooths u_d to get the loop output, u_f. u_f feeds back to the oscillator to determine the frequency of the oscillation. If u_1 is a steady-state sinusoid, then u_f will be constant and u_2 will oscillate at same frequency as u_1, but with a constant phase difference.

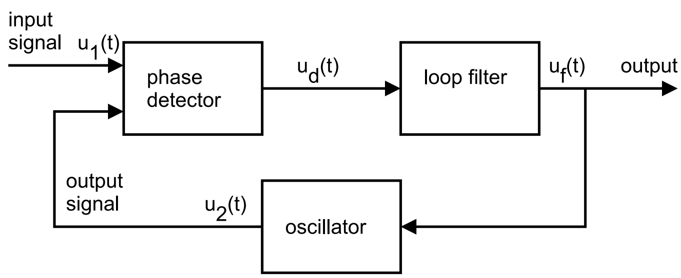

Fig. 1. Block diagram of a basic PLL. A VPLL is an array of PLLs, one per pixel.

VPLLs consists of an array of PLLs, one per pixel. A band pass filter placed at each PLL input selects oscillations of a single frequency. The VPLLs described here use digital-averaging all software PLLs.

Within a video PLL there is an abundance of signals useful in the perception of oscillating motion. The output of the loop, u_f is a phase error, but also gives an instantaneous estimate of the frequency of oscillation. For a gait, all pixels for the walker will lock to the same frequency and so the PLL does frequency entrainment.

Since all oscillators lock on the same frequency, the only difference between oscillators is the phase of their oscillation. The relative phases of the oscillators throughout the image give the relative phases of the intensity oscillations themselves. This performs the task of phase locking.

The use of a digital-averaging phase detector has the added benefit of being able to compute the magnitude of the u_1. This is useful in giving an indication of the amplitude, or strength, of the locked signal. By combining the magnitude with the phase, we get a phasor at each pixel site. The combination of phasors over several pixels forms a pattern that rotates at the locked frequency. Although one can use the pattern to classify and recognize oscillatory motion, in this paper we use selected phasors to set the timing for a walking model.

3.2 A Two-Dimensional Walking Model

A walking model can be derived from different sources such as passive mechanical models [6, 5, 2, 3], or motion capture data. Here we use a model based on observations of gaits generated by Poser [23]. Figure 2 describes this two-dimensional kinematic model. There are two legs, each composed of a shin and a thigh (see Figure 2(a)). The shin and thigh join at two knees and the two legs join at the hip. l_T and l_S are the lengths of the thigh and shin respectively. θ_T is the angle formed by the thigh and vertical reference. θ_S is angle formed by the shin with the thigh. One (θ_T, θ_S) pair is required for each of the two legs so four angles plus l_T and l_S are sufficient to describe the pose of the model at any point in time. Figures 2(b) and (c) describe the trajectories of θ_T and θ_S over one cycle of the gait. ϕ is the phase of the gait cycle and is normalized so that $0 \leq \phi < 1$ for a single cycle. $0 \leq \phi < 1/2$ is called the contact phase (the foot is in contact with the ground, and $1/2 \leq \phi < 1$ is called the swing phase (the foot is swinging forward).

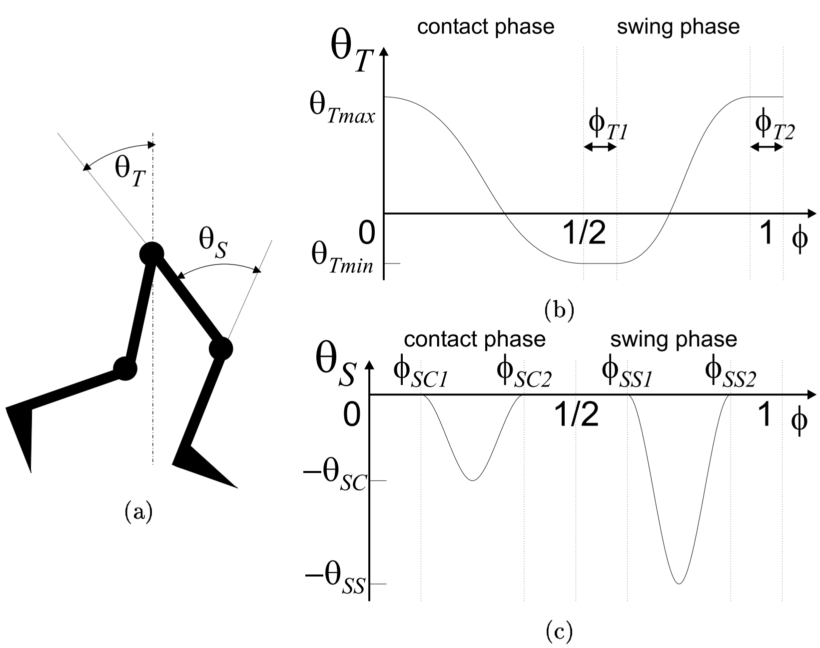

Fig. 2. An oscillating gait model: (a) simple two-dimensional kinematic model of a leg, (b) oscillations in thigh-angle, and (c) oscillations in shin angle.

The model walks by cycling the θ_T and θ_S trajectories through $0 \leq \phi < 1$. The two legs are assumed to be exactly opposite in phase, i.e., the phase difference between them is fixed at 1/2. The ten angular parameters described in Table 1 combine to generate a broad range of subjectively different gaits.

parameter	description
l_T	thigh length
l_S	shin length
θ_{Tmax}	forward extent of thigh motion
θ_{Tmin}	backward extent of thigh motion
ϕ_{T1}	pause after contact phase
ϕ_{T2}	pause after swing phase
θ_{SC}	contact phase knee bend
θ_{SS}	swing phase knee bend
ϕ_{SC1}	start of contact phase knee bend
ϕ_{SC2}	end of contact phase knee bend
ϕ_{SS1}	start of swing phase knee bend
ϕ_{SS2}	end of swing phase knee bend

Table 1. Summary of parameters for gait model of Figure 2.

3.3 Synchronization Process

To synchronize the model with gait data we extract the VPLL oscillator phase for a set of three points (A, B, and C) from images of a tracked gait sequence as shown in Figure 3. The positions of these points have the following significance:

A point of maximum backward thigh motion
B point of maximum forward thigh motion
C point of maximum forward shin/ankle motion

The VPLL phases at these points ($\angle A$, $\angle B$, and $\angle C$) identify key timing/phase relationships in the gait and can therefore be used to synchronize the model. For example, $\angle B - \angle A$ is the delay between forward thigh extension and backward thigh extension, $\angle C - \angle A$ is the delay between thigh extension and knee lock, and $\angle A$ determines when the θ_T trajectory is at its minimum. In normal operation, the VPLL oscillators synchronize to the frequency of pixel intensity oscillations, which for a gait is the step (or footfall) frequency. To synchronize the model we need an oscillator entrained to the fundamental frequency of the gait. For that reason we add a second oscillator to each PLL that is not in the feedback loop and oscillates at half the frequency of the loop oscillator. We synchronize using the phase from the out-of-loop, half-frequency oscillator. $\angle A$, $\angle B$, and $\angle C$ determine the phases in the model parameters by the following equations:

$$\phi = \angle B + \phi_e, \qquad \phi_{T1} = 0, \qquad \phi_{T2} = \angle B - \angle A,$$
$$\phi_{SC1} = 1 + \tfrac{\angle C - \angle A}{2}, \qquad \phi_{SC2} = \phi_{SC1} + 0.2,$$
$$\phi_{SS1} = 0.45, \text{ and} \qquad \phi_{SS2} = \phi_{SC1},$$

where ϕ_e is the constant phase error of the VPLL at the gait frequency. The size of the walking figure in the image determines l_T and l_S. We set θ_{Tmin}, θ_{Tmax}, θ_{SC}, and θ_{SS} to values approximated from observations of several gaits. In the resulting system, the source of timing information in the walking model comes only from the model-free VPLL output.

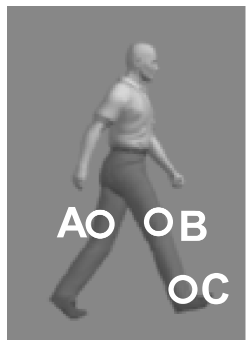

Fig. 3. Synchronization points to connect model-free to model-based representation.

4 Examples

To demonstrate the system we apply it to two gait sequences. The first is a synthetic sequence generated by Poser [23]. Figure 4 shows frames from the sequence with the limbs of the synchronized gait model superimposed. The second sequence is taken from the MoBo data base (Gross and Shi [24]). We use a segmented image sequence from the database. Figure 5 shows frames from the sequence. The segmented figure is dark gray on a black background. Limbs in the synchronized model are plotted in white. In both cases the synchronization results in a machine perception that matches the gait.

Note that we do not use all the parameters in the model. This suggests that future refinements of the model will have fewer degrees of freedom.

In frames 126 and 130 of Figure 4 we see that the shin angle of the model does not match the original image. This is due to the asynchronous gait produced by the Poser software. Since our walking model assumes a synchronous gait, it cannot account for differences between the left and right leg motion. When there are tracking errors it is possible for the synchronization points to drift off the body. This causes model synchronization to be erratic, something that we observed for about 10 frames in the MoBo sequence.

We use the segmented image sequence from the MoBo [24] database. Loosely draped clothing creates intensity variations due to the texture of the folds in the clothing. Although this has not proved to be a problem when viewing VPLL

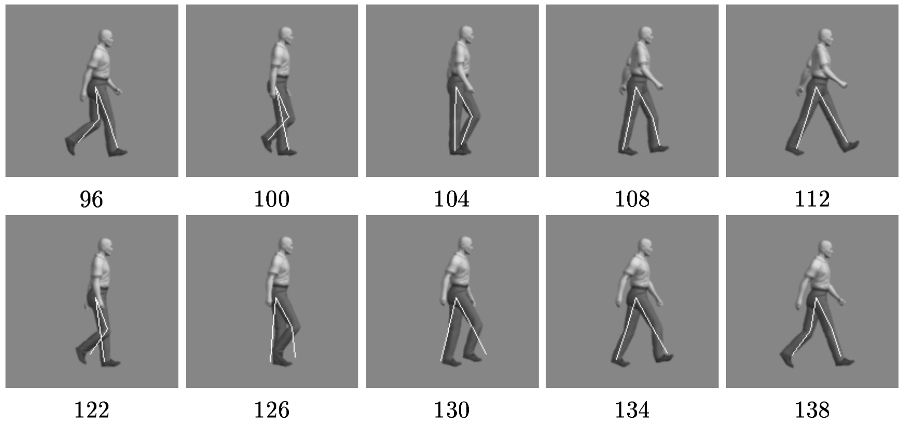

Fig. 4. Results of gait synchronization system applied to synthetic Poser [23] data. Frame number is indicated under each image.

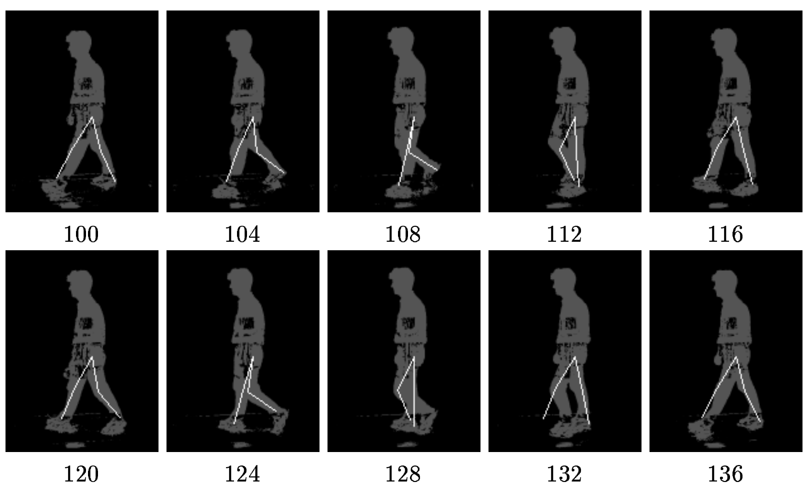

Fig. 5. Results of gait synchronization system applied to sequence from MoBo [24] database. Frame number is indicated under each image.

output as a global timing pattern for gait recognition, the timing at individual pixels is not as reliable. Segmentation avoids the problem by presenting a situation where the person appears as though they were dressed in white walking on a black background. The problem does not occur in the Poser sequence because the images are synthetic. Other approaches that may eliminate problems caused by clothing would include stabilizing (smoothing) the data in space and time.

5 Discussion

The examples of the previous section show that it is possible to synchronize a walking model to data gleaned from a model-free analysis. The linkage to the model is direct though, i.e., the model timing depends only on the differences of VPLL phases and there is no feedback. The VPLL data can vary with lighting, clothing and errors in tracking, and the phase variations pass through to the synchronization process. An improved system would use feedback from the synchronized model to improve the VPLL operation.

The size of the perceptual gap between video data and a complete kinematic model is large. Model-free methods are successful, but they do not span the entire gap. Pose estimation and visual motion capture systems [18, 15–17] do span the gap, but not easily. We suggest that a reliable transition from data to model will require several steps over smaller gaps. At each step, frequency and phase data must move from the data toward the model. The following steps illustrate one possibility.

1. Identify low-level motion features.
2. Describe the shape of the low-level features.
3. Fit a simple model to the shape data.
4. Fit a complete walking kinematic model.

It is necessary to transfer frequency and phase information at each step from one level of representation to the next. The key will lie in choosing the representations that best facilitate the transfer.

6 Conclusions

We described a system that connects a model-free and a model-based representation of a gait. While the connection is direct, it is successful in demonstrating that one representation can be used to synchronize another. Thus it shows the way toward connecting a series of representations to smoothly bridge the perceptual gap between data and model.

References

1. B. I. Bertenthal and J. Pinto. Complementary processes in the perception and production of human movements. In L. B. Smith and E. Thelen, editors, *A Dynamic Systems Approach to Development: Applications*, pages 209–239. MIT Press, Cambridge, MA, 1993.
2. T. McGeer. Passive dynamic walking. *The International Journal of Robotics Research*, 9(2):62–82, April 1990.
3. T. McGeer. Passive walking with knees. In *IEEE International conference on Robotics and Automation*, pages 1640–1645, 1990.
4. M. J. Coleman and A. Ruina. An uncontrolled toy that can walk but cannot stand still. *Physcial Review Letters*, 80(16):3658–3661, April 1998.

5. M. Garcia, A. Chatterjee, A. Ruina, and M. Coleman. The simplest walking model: stability, complexity, and scaling. *ASME Journal of Biomechanical Engineering*, 120(2):281–288, April 1998.
6. S. H. Collins, M. Wisse, and A. Ruina. A 3-d passive-dynamic walking robot with two legs and knees. *International Journal of Robotics Research*, April 2001.
7. F. Fazio, D. Perani, N. A. Borghese, S. Ferrari, F. Lo Monaco, V. Blasi, and M. C. Gilardi. How can virtual reality be? a pet study of motor actions observation. In *Human Brain Mapping Conference*, 1999.
8. J. E. Boyd. Video phase-locked loops in gait recognition. In *International Conference on Computer Vision*, pages 696–703, Vancouver, BC, 2001.
9. J. Laszlo, M. van de Panne, and E. Fiume. Limit cycle control and its application to the animation of balancing and walking. In *SIGGRAPH 96*, pages 155–162, 1996.
10. J. J. Little and J. E. Boyd. Recognizing people by their gait: the shape of motion. *Videre*, 1(2):1–32, 1998.
11. R. Cutler and L. Davis. Robust periodic motion and motion symmetry detection. In *Computer Vision and Pattern Recognition 2000*, 2000.
12. A. M. Baumberg and D. C. Hogg. Learning spatiotemporal models from training examples. Technical Report 95.9, University of Leeds School of Computer Studies, March 1995.
13. R. Polana and R. Nelson. Detection and recognition of periodic, nonrigid motion. *International Journal of Computer Vision*, 23(3):261–282, June 1997.
14. F. Liu and R. W. Picard. Finding periodicity in space and time. In *International Conference on Computer Vision*, 1998.
15. H. A. Rowley and J. M. Rehg. Analyzing articulated motion using expectation-maximization. In *Computer Vision and Pattern Recognition 97*, Peurto Rico, June 1997.
16. S. Wachter and H.-H. Nagel. Tracking of persons in monocular image sequences. In *Nonrigid and Articulated Motion Workshop*, Peurto Rico, June 1997.
17. C.R. Wren, A Azarbayenjani, T. Darrell, and A. P. Pentland. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.
18. C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Computer Vision and Pattern Recognition 1998*, pages 8–15, Santa Barbara, June 1998.
19. H. Fujiyoshi and A. J. Lipton. Real-time human motion analysis by image skeletonization. In *DARPA Image Understanding Workshop*, Monterey, California, November 1998.
20. A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *Computer Vision and Pattern Recognition 2001*, volume II, pages 52–57, Kauai, HI, 2001.
21. R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Computer Vision and Pattern Recognition 2001*, volume II, pages 726–731, Kauai, HI, December 2001.
22. A. Bobick and A. Johnson. Gait recognitin using static activity-specific parameters. In *Computer Vision and Pattern Recognition 2001*, volume I, pages 423–430, Kauai, HI, December 2001.
23. MetaCreations Corporation. *Poser 4 user guide*, 1999.
24. R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.