Model-Based Gait Enrolment in Real-World Imagery

David K Wagg and Mark S Nixon

Department of Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, UK
dkw02r | msn@ecs.soton.ac.uk

Abstract

We present a model-based approach to gait extraction that is capable of reliable operation on real-world imagery. Hierarchies of shape and motion are employed to yield relatively modest computational demands, avoiding the high-dimensional search spaces associated with complex models. Anatomical data is used to generate shape models consistent with normal human body proportions. Mean gait data is used to create prototype gait motion models, which are adapted to fit individual subjects.

Accuracy is evaluated on subjects filmed from a fronto-parallel view in controlled laboratory conditions, for which some gait parameters are known. We further show that comparable performance is attained in outdoor conditions. As such, we describe a new approach to enrolment for gait recognition technologies, allowing reliable subject gait extraction in real-world imagery.

1. Introduction

Gait may be defined as the individual pattern of movement produced as a person walks. This pattern is sufficiently unique for each individual to be employed as a biometric [Winter91, Nixon99]. Gait analysis is usable from a distance and does not require the subject to be aware of or cooperate with its use, making it particularly valuable in surveillance, or other applications where non-contact operation is required.

This field is currently dominated by face recognition, supported by the role of facial features in the human recognition process. However, gait is more difficult to obscure or disguise, and can be measured from a much wider range of viewpoints. Gait is also more robust with respect to occlusion and variations in illumination, as a gait signature is spatio-temporal rather than a purely spatial measure.

Gait may be best employed in combination with other biometrics, with facial features being an obvious choice. Most approaches to face recognition require a relatively constrained frontal viewpoint, and gait could be employed as a back-up strategy when the subject's face is not visible. Alternatively, multiple cameras could be employed to combine face and gait features, improving overall recognition performance [Shakhnarovich01].

However, enrolment is a more difficult problem for gait, particularly when enrolment conditions cannot be controlled (for example, when enrolling a subject from CCTV footage). Gait enrolment requires the extraction of limb dynamics over a period of time, ideally capturing at least one full gait cycle. In uncontrolled capture conditions, it is likely that other objects will interfere with and occlude the subject; in addition gait is partially self-occluding, as one leg passes in front of the other. To successfully resolve this problem, extraction methodologies must be highly robust to noise and occlusions.

Many existing approaches to gait enrolment are data-driven, typically using the person's silhouette or features derived features from it as a basis for recognition [BenAbdelkader02, Collins02, Huang99, Johnson01, Kale03, Lee02, Phillips02]. This methodology has many advantages, chiefly of speed and simplicity, but has the disadvantage that silhouette dynamics are only indirectly linked to gait dynamics. Noise, occlusions and variations in clothing will all affect silhouette dynamics; it is unclear how a silhouette-based feature set could be normalised for these factors.

Model-based approaches overcome these weaknesses by incorporating knowledge of the shape and dynamics of human gait into the extraction process [Cunado03, Meyer98, Yam02]. The use of a model ensures that only image data corresponding to allowable human shape and motion is extracted, reducing the impact of noise. It also means that gait dynamics are extracted directly by determining joint positions, rather than inferring dynamics from other measures. A model-based approach also has the potential for more general applications, such as animation, user interfaces or model-based coding [Gavrila99].

However, the use of a parametric model introduces its own problems. Success in recognition is dependent on the gait signature being sufficiently complex to incorporate individual variation across the subject population, so that a given subject can be distinguished from all the other subjects under test. As gait is dependent on a large number of parameters (such as joint angles and body segment sizes), this requirement leads to complex models with many free parameters. Finding the best fitting model for the subject thereby necessitates searching a high-dimensional parameter space, with correspondingly high computational requirements.

Most early approaches dealt with this problem by severely limiting model complexity; later solutions have improved on this situation somewhat. [Nash 98] employs a genetic algorithm to cope with the high computational demands, but due to its reliance on stochastic processes this strategy cannot guarantee an optimal model fit. [Lappas02] introduces the dynamic velocity Hough transform, which applies dynamic programming to find an optimal object trajectory using structural evidence and smoothness of motion constraints. However, under this formulation it is difficult to apply parametric motion constraints (such as pendular limb motion).

To reduce the computational requirements of a model-based approach, we employ a model hierarchy composed of shape and motion components.

A velocity filtering algorithm is employed to determine the bulk motion of the person independently of shape parameters. Using this motion information we form a global temporal accumulation describing the person's average shape over the gait sequence. This accumulation is used to robustly estimate the size and shape of the person's body segments, using ellipses for the head and torso and two pairs of lines for each leg, applying anatomical constraints to reduce matching errors. Using this initialisation we can estimate the dominant gait frequency via a measurement of edge strength about the lower leg region over time. Leg motion is estimated by fitting prototype gait curves collected from a clinical gait study, stretched or compressed to fit the subject's gait frequency and hip rotational amplitude.

Our approach currently assumes a single subject moving at a constant speed, fronto-parallel against a cluttered background. However, this approach could be generalised to an arbitrary viewpoint.

We show that this methodology provides a good initial model fit suitable for further adaptation, and is capable of performance in noisy real-world conditions similar to that in controlled laboratory conditions.

2. Gait Signature Extraction

2.1. Bulk Motion Estimation

We may consider the motion of a person in normal gait to be composed of many separate motion components, forming a hierarchy according to the total pixel displacement they are responsible for. At the top of this hierarchy is the person's velocity in the horizontal plane, as a person will move with approximately constant velocity during normal gait (changes in velocity may also distort their gait signature, further justifying this assumption). The second level of the hierarchy is articulated motion; we may consider a third level to be object deformations (for example due to clothing or camera distortion), but this level of detail is considered unnecessary for our current purposes.

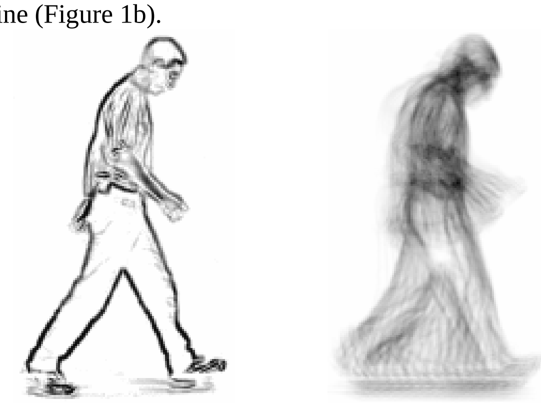
Image data is pre-processed (Figure 1a) using a Gaussian averaging filter for noise suppression, followed by Sobel edge detection and background subtraction (the background is computed by a temporal median of neighbouring frames). This removes all static objects, leaving only edges belonging to moving objects. The extraction process does not require binary edge data, which means that error-prone thresholding can be avoided.

Using a velocity filtering algorithm it is possible to determine object motion independently of shape. This algorithm effectively performs the same global temporal accumulation as the velocity Hough transform [Nash97], but without shape specificity:

$$A_v(i, j) = \sum_{n=0}^{N} I_n\left(i + v\left(\frac{N}{2} - n\right), j\right) \qquad (1)$$

where A_v is the accumulation for velocity v (in pixels per frame), I_n is the image intensity function at frame n, i and j are coordinate indices and N is the number of frames in the gait sequence.

This algorithm sorts objects in the scene according to their velocity and starting position, producing an accumulation for each possible object velocity. Each object's contribution to an accumulation is dependent on its edge strength, the number of frames it is in view of the camera and how close its velocity is to the accumulation velocity. This global averaging process means that objects in each accumulation are relatively unaffected by other objects, greatly reducing the problems associated with objects merging and splitting. At the correct accumulation velocity for an object, edges from each frame will accumulate to a single area, producing an average shape outline (Figure 1b).



(a) Section of pre-processed image data (b) Global temporal accumulation

Figure 1: Motion estimation by temporal accumulation

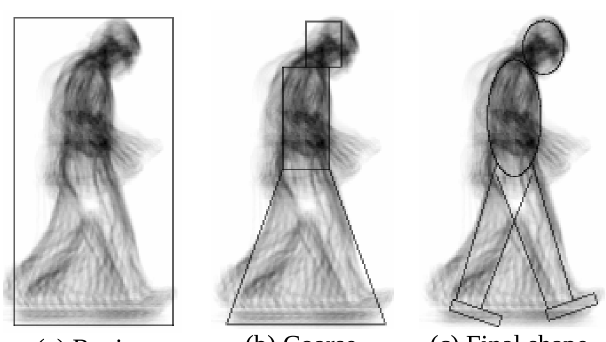
Each moving object in the scene appears as a peak in a plot of maximal accumulation intensity against velocity. Assuming that the person is the most significant moving object in the scene, their velocity can be inferred by selecting the highest peak in this plot (this assumption holds true for most current gait databases). If there are other more significant objects within the scene moving at a similar velocity, we must apply some knowledge of the person's shape to distinguish them from the other objects.

Noting that Equation 1 simply shifts and accumulates each frame, we can improve computational efficiency by first run-length encoding the input data. This representation is shift-invariant, and as runs of zero magnitude edge strength can simply be discarded, this reduces the order of the algorithm to $O(V{\cdot}E{\cdot}N)$, where V is the number of possible velocities, E is the mean number of edge points in each frame and N is the number of frames in the gait sequence. Further performance improvement can be accrued by downsampling input frames and applying a coarse-to-fine velocity search strategy.

2.2. Shape Estimation

The temporal accumulation computed during the bulk motion estimation stage forms an average global view of the person's shape. Parameters that do not change over the course of the gait cycle can therefore be determined from the temporal accumulation; as it is robust with respect to noise and occlusion, static parameters can be estimated with confidence.

The size and proportions of the person are estimated in a hierarchical fashion using anatomical constraints, derived from data published in [Winter90]. A region-growing algorithm is first applied to find all edges belonging to the person. This algorithm is initialised at the peak point in the accumulation, and an aspect-ratio constrained rectangular region is expanded about the point until all significant edges have been encompassed (Figure 2a).



(a) Region expansion (b) Coarse segmentation (c) Final shape estimate

Figure 2: Shape extraction hierarchy

Using this initialisation the approximate height of the person is estimated, using a fixed body segmentation based on mean anatomical proportions (Figure 2b). The final shape model (Figure 2c) consists of two ellipses for the head and torso, two rectangles for the feet and two pairs of lines for each leg. The parameters describing the head and torso are determined by template matching within the locality of the initial segmentation, constrained by mean anatomical proportions. The leg and foot shape parameters are computed as a fixed proportion of the subject's height and torso width, again based on mean anatomical data.

Note that although all shape dynamics are lost in the accumulation process, it is still possible to estimate the amplitude of hip rotation, which may be used to aid articulated motion estimation.

2.3. Articulated Motion Estimation

The motion of the leg during normal gait is periodic, and may be approximately modelled by a single sinusoid [Cunado03]. Applying this assumption, we can estimate a person's gait cycle frequency by measuring edge strength within the outer region of their legs, throughout the gait sequence (Figure 3).

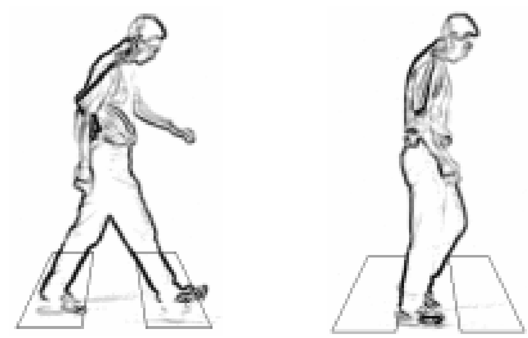
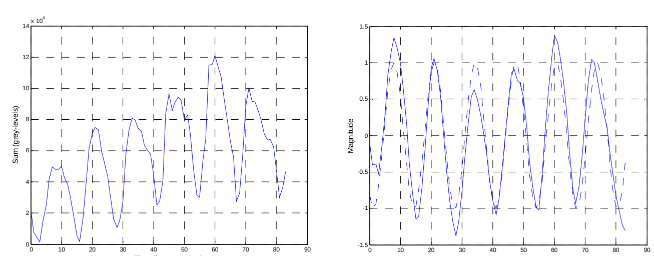


Figure 3: Gait cycle frequency estimation using within-region edge strength measurements

These measurements form a signal with approximately sinusoidal shape, distorted and contaminated by noise due to varying illumination, occlusion and motion estimation errors. Figure 4a depicts this signal for an example outdoor sequence:



(a) Measured signal (b) Normalised data and sinusoidal fit

Figure 4: Gait cycle frequency estimation

Signal distortions are corrected by using low-order polynomials to model variation in the mean level of the sinusoid (numerator of Equation 2), and local variations in sinusoid magnitude (denominator of Equation 2):

$$S_n = \frac{S - p(S)}{p(\ |S - p(S)|\ )} \quad (2)$$

where S_n is the normalised signal, S is the original signal and $p(x)$ denotes the best 2nd-order polynomial fit to signal x, computed by least-squares regression.

Frequency estimation is performed by fitting a fixed-amplitude sinusoid to the data, selecting the frequency and phase that minimises squared error (Figure 4b).

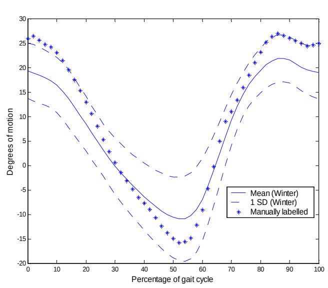
This frequency information can be applied directly using sinusoidal joint rotation models [Cunado03, Yam02]. A single sinusoid is adequate to approximately model the rotation of the hip and knee joints:

$$\theta_h(t) = A_h \sin\left(wt + \varphi_h\right) + \psi_h \quad (3)$$

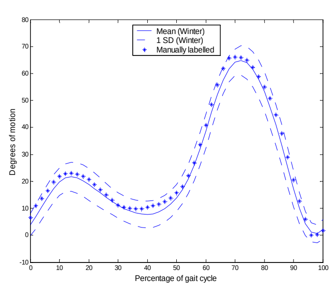
$$\theta_k(t) = A_k \sin\left(wt + \varphi_h + \varphi_k\right) + \psi_k \quad (4)$$

where $\theta_h(t)$ and $\theta_k(t)$ are the respective hip and knee joint rotations (measured relative to the vertical axis) at time t, A_h and A_k are the joint rotational amplitudes, w is the gait cycle frequency (in radians per frame), ϕ_h is the starting hip joint phase, ϕ_k is a constant phase offset, ψ_h and ψ_k are constant amplitude offsets.

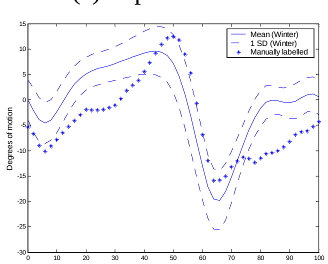
However, accuracy can be improved by more closely modelling human gait. Clinical gait studies have quantitatively measured the pattern of movement produced as people walk, by attaching markers to each joint. Mean gait patterns from [Winter91] were used to produce prototypical rotation models for the hip, knee and ankle joints. Figure 5 shows these models, together with joint angles manually extracted from a sequence in the Southampton HiD database [Shutler02]. Note that by clinical convention rotations are measured in degrees of motion, as opposed to rotation relative to the vertical axis.



(a) Hip rotation



(b) Knee rotation



(c) Ankle Rotation

Figure 5: Mean joint rotation patterns

This comparison suggests that the mean rotation models for the hip and knee match well to a typical subject. Ankle rotation is not such a good match, as the subjects in the clinical study were barefoot, as opposed to a typical subject who will be wearing shoes. However, the mean ankle rotation model still provides a better basis than a simple sinusoidal model would. The motion of the pelvis is not modelled at this point; the positions of the hip joints are assumed to coincide, remaining at the same level throughout the gait sequence.

The discrete Fourier transform (DFT) of each model is computed, creating continuous representations of the shape of the models. To match the subject's gait, the DFT models are scaled to match the subject's estimated gait cycle frequency and hip amplitude. Cycle phase is estimated by temporally matching leg templates to edge strength over the whole sequence, selecting the phase that maximises template correlation. Matching globally in this fashion increases resistance to noise, and can be performed quickly when only one search parameter is required.

Finally, the vertical oscillation of the subject's upper body is modelled by a single sinusoid with parameters proportional to the subject's height and gait motion:

$$Y(t) = A_y \sin 2\left(wt + \varphi_h + \pi/8\right) + \psi_y \quad (5)$$

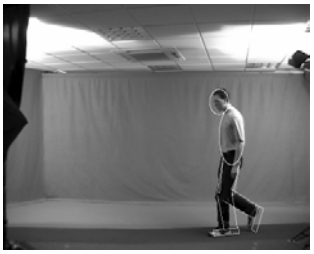
where $Y(t)$ is the y-coordinate of the torso at time t, A_y is the amplitude of oscillation, w is the gait cycle frequency, ϕ_h is the starting hip joint phase and ψ_y is the centre of oscillation.

The joint positions extracted by this process only approximate the true joint positions (the estimation process effectively assumes average gait motion, or no individuality). However, these positions form a strong basis for further model adaptation, which would make recognition possible.

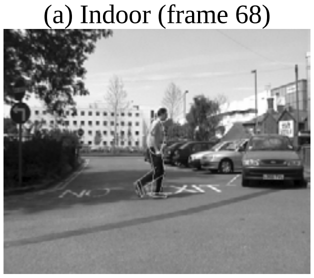
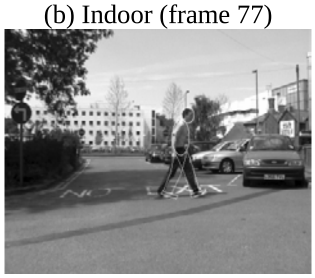
3. Results

The performance of the gait extraction process was evaluated on sequences of two subjects from the Southampton HiD database [Shutler02]. Each subject was filmed from a fronto-parallel viewpoint, in controlled laboratory conditions and in noisy outdoor conditions, allowing the noise-resistance to be tested in isolation from other variables. The database is encoded in Digital Video (DV) format at a resolution of 720x576 pixels, recorded at a rate of 25 frames per second. Each sequence typically consists of 80-100 frames, or around 3 full gait cycles.

The extraction process is fast, with approximately 75% of the total processing time taken up by pre-processing. A 2.4GHz Pentium 4-based PC was used for all testing, requiring approximately 30 seconds processing time for each sequence. Figures 6 and 7 give some examples of the extraction process, showing good overall performance, especially on the outdoor data. Note that there is some error evident in shape estimation, and also some error caused by the assumption that the left and right hip joints coincide.

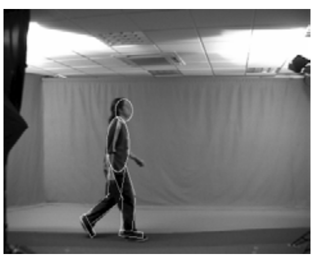



(a) Indoor (frame 68) (b) Indoor (frame 77)

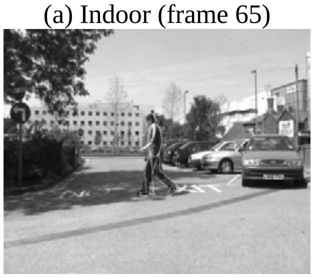
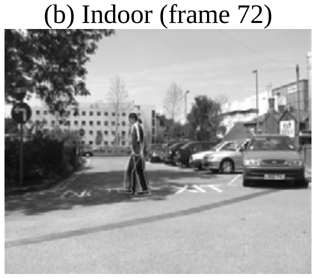



(c) Outdoor (frame 39) (d) Outdoor (frame 47)

Figure 6: Sample extraction results for subject 013




(a) Indoor (frame 65) (b) Indoor (frame 72)

(c) Outdoor (frame 49) (d) Outdoor (frame 56)

Figure 7: Sample extraction results for subject 014

The set-up of the indoor data allows an approximation to ground truth to be made by chroma-key extraction of the subject's silhouette [Shutler02]. From this silhouette data the frame numbers at which the subject's heel strikes the floor are recorded, so that a comparison can be made with the automatically extracted result. The heel-strike frames were estimated from the automatic extraction by finding the knee rotation minima over the sequence.

Although this does not yield an exact measure of the extraction performance, this evaluation can be performed automatically on a large number of sequences. Table 1 shows the results of this evaluation for 56 indoor test sequences split equally over four subjects:

Table 1: Extraction performance under controlled conditions – RMS error in predicted heel-strike frames

Subject	Mean	Standard Deviation
013 (M)	0.933	0.236
014 (M)	0.954	0.458
033 (F)	0.741	0.209
037 (F)	0.979	0.363

The mean error in estimating the point of heel-strikes is around ±1 frame for all subjects, comparable to typical human labelling error. This is an encouraging result, demonstrating that we can successfully track the motion of the subject's legs in relatively clean indoor conditions. To demonstrate robustness, the extraction process was repeated on outdoor data, totalling 64 sequences of the same four subjects. As no ground truth data is available for the outdoor dataset, extraction performance is estimated by comparing the gait cycle period extracted from the outdoor data to that of the indoor data (Figures 8 and 9):

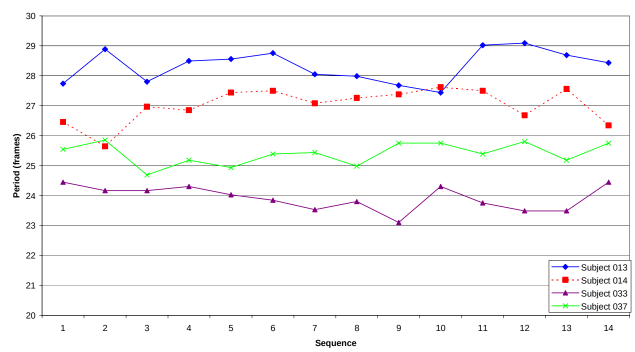


Figure 8: Period extraction for indoor data

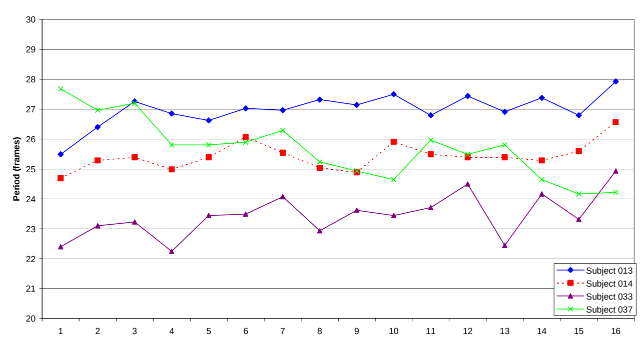


Figure 9: Period extraction for outdoor data

The extracted period is generally consistent between different gait sequences for each subject. Some of the subjects exhibit a reduced gait period on the outdoor dataset, indicating increased cadence. This may be due to the walking surface, or possibly because the subjects do not have a limited walking track in the outdoor dataset. However, even with only one gait parameter most of the subjects can be distinguished from one another.

For a more detailed view of performance, one indoor and one outdoor sequence was manually labelled for each test subject. The positions of the hip, knee and ankle joints were recorded, for comparison against the automatically extracted joint positions. The error is measured by a Euclidean distance metric, normalised to a percentage of the height of the subject. This error is given for a mean gait cycle, averaged over the sequence.

Figure 10 shows the errors measured at each joint position for subject 013 from the Southampton HiD database. Note that some error is expected of the human labelling, estimated at around 1% of subject height (the height of a subject is typically around 300 pixels on the indoor data or 200 pixels on the outdoor data).

This comparison shows that the additional increase in error when moving from controlled laboratory conditions to outdoor conditions is relatively small. It also shows that the additional complexity imposed by the use of mean gait rotation models is justified, resulting in a significant reduction in error over the sinusoidal models (Equations 3 and 4). The motion produced by these models is noticeably more natural in appearance to the human observer, suggesting that further improvement in performance is possible.

4. Conclusions

We have presented a new model-based gait enrolment technique to allow the use of gait analysis on real-world imagery. A model hierarchy of shape and motion keep the computational requirements of this approach to a minimum, while retaining the well-known robustness of a model-based approach.

Anatomical data and mean gait data is applied to produce shape and motion models adhering to known human proportions and gait dynamics, minimising the modelling error in this approach.

We have shown that we can reliably locate joint positions for the purposes of gait analysis in real-world imagery, with only a small loss in accuracy compared to controlled laboratory conditions. Future work will extend this approach by adapting the mean gait models to match each individual, so that recognition may be performed on the gait parameterisation thus obtained.

Acknowledgements

We gratefully acknowledge partial support by the European Research Office of the US Army under Contract No. N68171-01-C-9002.

References

[BenAbdelkader02] C BenAbdelkader, R Cutler and L Davis. "Stride and Cadence as a Biometric in Automatic Person Identification and Verification." *Proc. FGR*, pp. 372-377, 2002.

[Collins02] R T Collins, R Gross and J Shi. "Silhouette-based Human Identification from Body Shape and Gait." *Proc. FGR*, pp. 351-356, 2002.

[Cunado03] D Cunado, M S Nixon and J N Carter. "Automatic Extraction and Description of Human Gait Models for Recognition Purposes." *CVIU*, **90** (1), pp. 1-41, 2003.

[Gavrila99] D M Gavrila. "The Visual Analysis of Human Movement: A Survey." *CVIU*, **73** (1), pp. 82-98, 1999.

[Huang99] P. S. Huang, C. J. Harris and M. S. Nixon. "Recognizing Humans by Gait via Parametric Canonical Space." *Artificial Intelligence in Engineering*, **13** (4), pp. 359-366, 1999.

[Johnson01] A Y Johnson and A F Bobick. "A Multi-View Method for Gait Recognition Using Static Body Parameters." *Proc. AVBPA*, pp. 301-311, 2001.

[Kale03] A Kale, N Cuntoor, B Yegnanarayana, A N Rajagopolan and R Chellappa. "Gait Analysis for Human Identification." *Proc. AVBPA*, 2003.

[Lee02] l Lee and W E L Grimson. "Gait Analysis for Recognition and Classification." *Proc. FGR*, pp. 155-162, 2002.

[Meyer98] D Meyer, J Posl and H Niemann. "Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture Densities." *Proc. BMVC*, pp. 459-468, 1998.

[Nash97] J M Nash, J N Carter and M S Nixon. "Dynamic Feature Extraction via the Velocity Hough Transform." *Pattern Recognition Letters*, **18**, pp. 1035–1047, 1997.

[Nash98] J M Nash, J N Carter and M S Nixon. "Extraction of Moving Articulated-Objects by Evidence Gathering." *Proc. BMVC*, pp. 609-618, 1998.

[Phillips02] P J Phillips, S Sarkar, I Robledo, P Grother and K Bowyer. "The Gait Identification Challenge Problem: Data Sets and Baseline Algorithm." *Proc. FGR*, pp. 137-142, 2002.

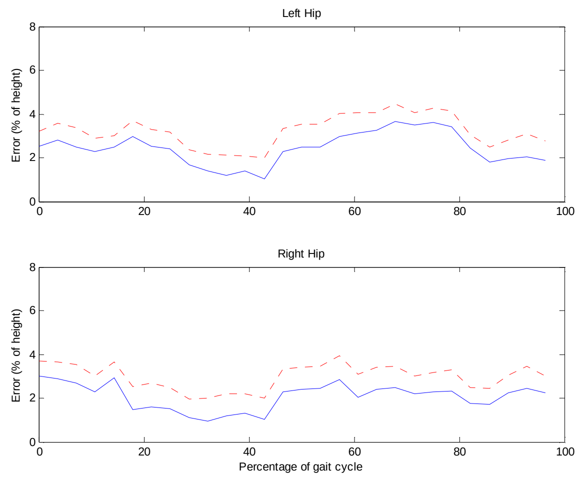
[Shakhnarovich01] G Shakhnarovich, L Lee and T Darrell. "Integrated Face and Gait Recognition from Multiple Views". *Proc. CVPR*, pp. 439-446, 2001.

[Shutler02] J D Shutler, M G Grant, M S Nixon and J N Carter. "On a Large Sequence-based Human Gait Database." *Proc. RASC*, pp. 66-71, 2002.

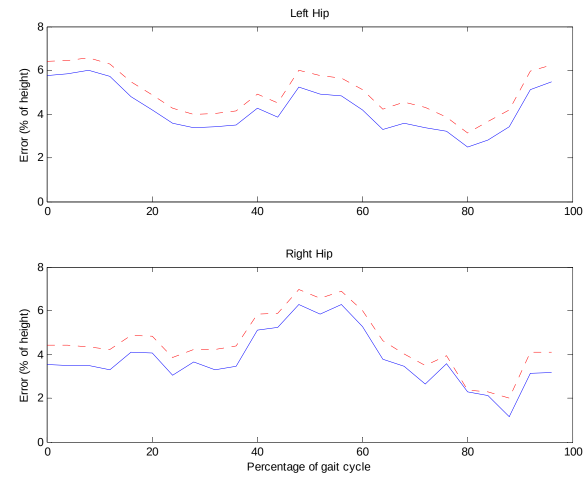
[Winter90] D A Winter. "Biomechanics and Motor Control of Human Movement (2nd Edition)." *John Wiley and Sons*, 1990.

[Winter91] D A Winter. "The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological." *University of Waterloo press, Ontario*. 1991.

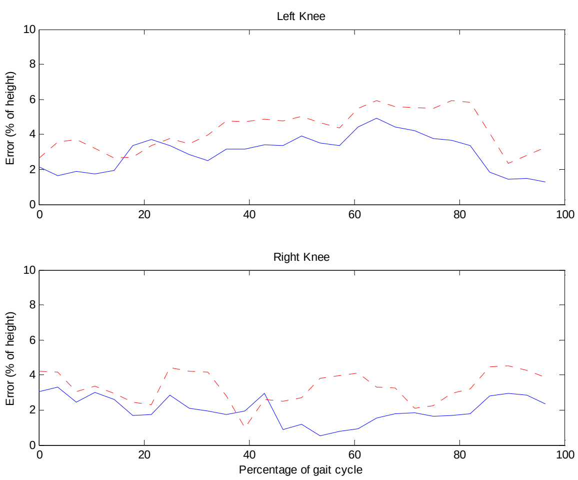
[Yam02] C Yam, M S Nixon and J N Carter. "On the Relationship of Human Walking and Running: Automatic Person Identification by Gait." *Proc. ICPR*, pp. 287-290, 2002.



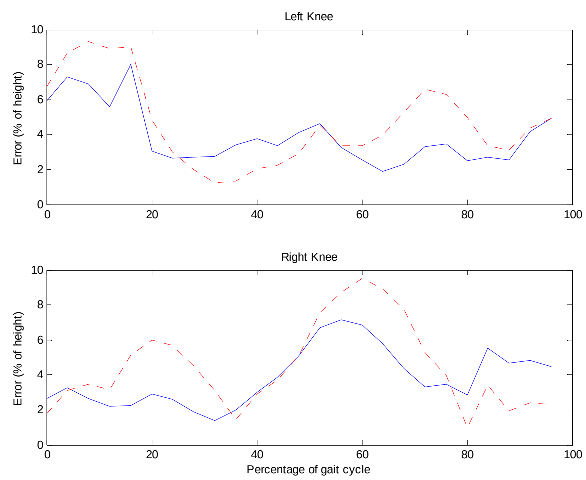
(a) Hip position (laboratory conditions)



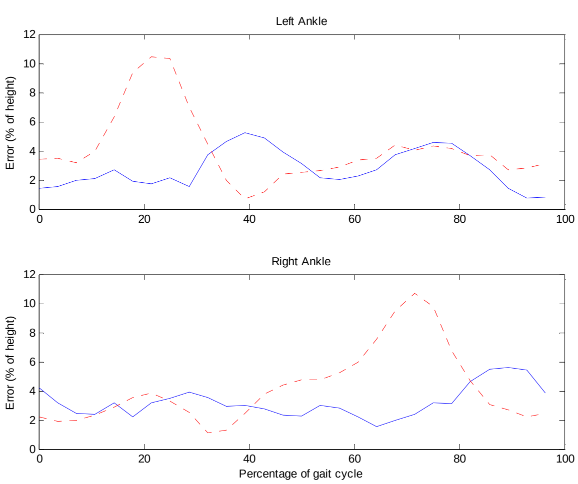
(b) Hip position (outdoor conditions)



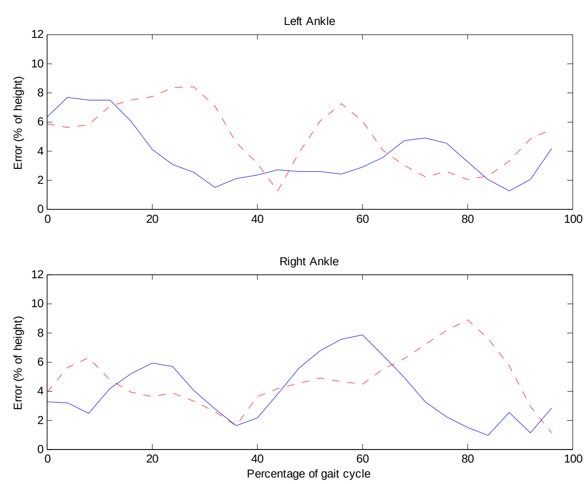
(c) Knee position (laboratory conditions)



(d) Knee position (outdoor conditions)



(e) Ankle position (laboratory conditions)



(f) Ankle position (outdoor conditions)

Figure 10: Error in automatically extracted joint positions from manually labelled positions (subject 013).
Solid line – mean gait models Dotted line – sinusoid model