Gait Extraction and Description by Evidence-Gathering

David Cunado, Jason M. Nash, Mark S. Nixon and John N. Carter
Department of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK
{dc95r|jmn95r|msn|jnc}@ecs.soton.ac.uk

Abstract

Using gait as a biometric is of increasing interest, yet there are few model-based, parametric, approaches to extract and describe moving articulated objects. One new approach can detect moving parametric objects by evidence gathering, hence accruing known performance advantages in terms of performance and occlusion. Here we show how that the new technique can be extended not only to extract a moving person, but also to extract and concurrently provide a gait signature for use as a biometric. We show the natural relationaship between the bases of these approaches, and the results they can provide. As such, these techniques allow for gait extraction and description for recognition purposes, and with known performance advantages of a well-established vision technique.

1 Introduction

Gait is an attractive biometric since, like face recognition, it is non-invasive and can be measured without subject contact or knowledge. An inherent advantage over facial measures is that it is not easily obscured or disguised in many normal situations. Studies in computer vision, medicine and psychology suggest there is a rich potential for gait as a biometric. However, attention in computer vision research has only recently focussed on extracting and describing moving objects, and rather less on moving ambulatory ones.

Medical research has mainly concerned classifying the components of gait to compare pathologically abnormal patients with normal patients. Standard movement patterns of each gait component for pathologically normal people were produced by Murray *et al.* [7, 8]. Allied approaches in psychology have shown how we can recognise people by their gait, even postulating models for human gait [5, 3]

In a heuristic method for automatic recognition, gait was extracted from a sequence of images as a spatio-temporal pattern [11]. This involved finding '*translating blobs in image sequences*'. Although this method achieved a promising recognition rate, the approach was not directly related to the mechanics of walking or body structure. Statistical approaches [6, 4] to gait recognition use eigenspace transformation (EST) based on Principle Component Analysis to create a gait signature. The image sequence is described by its motion content, rather than its relation to a model. Like Niyogi & Adelsen's work [11], these statistical approaches lack an explicit and attributable metric, although they have achieved encouraging results.

There have of course been many vision approaches aimed to model and then to track human movement though an image sequence. Some tracking approaches have used blobs, whereas some of the more sophisticated models use interconnected sticks or cylinders. Again, these lack an explict and unifying theory for moving-feature extraction and description. This is perhaps because human gait requires techniques which can extract and describe a moving articulated object. Ideally these techniques should be able to handle occlusion and noise. These are the attributes of evidence gathering approaches. Recently, the velocity Hough transform (VHT) has been developed [9] to extract moving parametric objects by evidence-gathering. The approach globally integrates local data, so information missed in one scene, yet present in another, still contributes to the final evidence gathered. We show how this approach can be extended first to extract the moving thigh from a sequence of images, ie. to locate a moving articulated object. Then, the motion-based evidence-gathering procedure is used to generate a signature which describes the motion concurrent with its extraction. This process is automatic and is neither heuristic nor application specific: the signature is generated direct from the image sequence.

The VHT is ideally suited to gait analysis since it considers information pertaining to both motion and structure, enabling features with a particular structure exhibiting a certain type of motion, to be identified. Earlier, [10], we have shown how the VHT can be extended for articulated objects. Here, we show how further extension can provide a biometric signature and the relationship between the two techniques and their results. Section 2 shows first how the VHT can be extended to extract an articulated object (the human legs). Then we show how it can be extended to extract and describe human motion (gait), allowing recognition. This avoids problems in the earlier approach [2] which extracted lines in individual frames (using least squares analysis for frames where the lines could not be identified) and then developed the gait signature from the postion and phase of the lines. We also show how a Genetic Algorithm (GA) can be used to handle the large parameter space associated with these new VHT techniques. This leads to practicable techniques, which can identify articulated objects, and extract and describe human motion, as shown in Section 3.

2 Gait Evidence Gathering

2.1 Articulated Objects

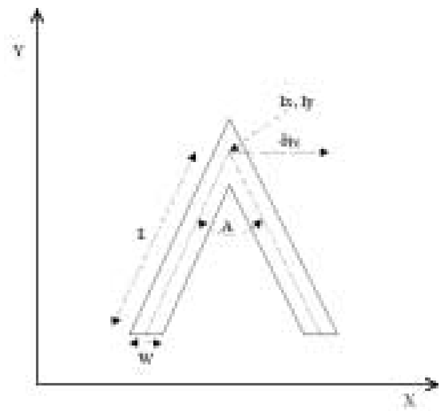

Figure 1: Articulated Lines

A simple model of the legs, a pair of thick lines, which are free to articulate about a single point, is shown in Fig. 1. The model is defined by: Ix and Iy, the x and y co-ordinates, respectively, of the point of articulation of a subject's legs i.e. the hip; A, the angle between the legs (measured relative to the centre of each leg); L, the length of each leg; and W, the width. Ambulation, can then be simulated through the variation of the appropriate model parameters with time.

The dynamics of the model are derived from medical studies, which indicate that human gait is periodic, with the rotation pattern of each thigh during a gait cycle being approximately sinusoidal in nature, [7, 8]. To parallel this behaviour, the model parameter A, the angle between the legs, is varied in a sinusoidal manner.

$$A = B_A \sin(\omega_A \pi t + \phi_A) \tag{1}$$

where B_A corresponds to the amplitude of the parameter variation with time t and ω_a and ϕ_A are the associated frequency and phase of the variation.

Similary, the lateral motion exhibited by a subject's hip during ambulation can be modelled as a constant forward velocity coupled with a sinusoidal term [7] as

$$Ix = V_{Ix} + B_{Ix} \sin(\omega_{Ix} \pi t + \phi_{Ix}) \tag{2}$$

To determine the absolute position of the model's point of articulation, or hip, for a given time t, Eqn. (2) is integrated with respect to time, to give Eqn. (3).

$$\begin{aligned} Ix = Ix_0 + V_{Ix} t + \frac{B_{Ix}}{\omega_{Ix}} \sin(\omega_{Ix} \pi t + \phi_{Ix}) \\ - \frac{B_{Ix}}{\omega_{Ix}} \sin(\phi_{Ix}) \end{aligned} \tag{3}$$

The initial x co-ordinate of the articulation point is Ix_0 and V_{Ix} is a constant velocity term. The amplitude of the sinusoidal variation is given by B_{Ix}, and ω_{Ix} and ϕ_{Ix} are the associated frequency and phase. The final term of the expression is a constant of integration which ensures that there is zero displacement from Ix_0 at time $t = 0$.

Clearly, the gait model as described is only intended as an approximation, with consideration not being given to vertical motion of the hip or independent motion of the lower leg, ie. there are no knees. However, the structure and dynamics of the model are sufficient to estimate the relative thigh motion and the corresponding hip rotation pattern, which is a primary cue for gait recognition.

Having proposed a parameteric description of the feature of interest, namely a walking pair of legs, the principles of the VHT can be employed to identify instances of the feature from a temporal sequence of images. As with the conventional HT, the VHT uses an inverse image-to-model space mapping in conjunction with edge point data, to gather evidence for the structural and motion parameters of the feature of interest. Using the edge point data, it is possible to estimate initial values, i.e. values in the first frame of the sequence, for the model's point of articulation, the subject's hip, given a range of the remaining model parameters to search over and that each frame of the sequence is labelled with a time reference value, relative to the intial frame.

Accordingly, extraction of an instance of the proposed gait model requires a 11-D accumulator space to store votes for the respective model parameters. In each frame, pairs of edge pixels are chosen at random and possible values for the point of articulation are calculated for a predefined range of angles, A; given that the point of articulation is the intersection point of two lines bisecting each leg (see Fig. 1) and that each edge point is assumed to lie on a seperate leg. Possible values for the initial point of articulation, i.e. its position in the first frame of the sequence, are then identified using Eqns. (1) and (3), and the time index of the current frame. The accumulator cells corresponding to the identified initial points of articulation, and the associated variation of parameters A and I_x, generated by the current pair of edge pixels, are then incremented. Once all edge pixels in the current frame have been processed, the process repeats for the next frame of the sequence, voting in the same accumulator space, until all frames have been analysed. The result of the process is a peak in the accumulator space correspsonding to the best estimate of the gait model's parameters. If multiple walkers are present in the image sequence, then multiple peaks will occur in the accumulator space, with each peak corresponding to the gait model parameters of a particular individual.

The proposed gait model however, is a relatively simple one, in terms of the stucture and dynamics of the human leg, and is designed only to estimate the relative motion of a subject's thighs, and in turn the hip rotation pattern. To obtain more comprehensive measurements pertaining to a subject's gait, which are required if gait is to be used as a biometric, demands a higher level of confidence in the extracted thigh data and for consideration to be given to the motion of the lower leg. Extension of the proposed gait model to meet these requirements is possible, yet the dynamics of the lower leg are considerably more complex than those exibited by the thigh and are not easily modelled. Hence, a Fourier based model is now described, which faciliates a more precise description of the dynamics of the thigh and is more readily extendable to include the motion of the lower leg.

2.2 Extracting and Describing Human Gait

Earlier [2], an approach to using gait as a biometric modelled the lower limbs as two inter-connected pendula (Fig. 2), and gait was considered as the motion of these pendula. This did not use the new precepts of the VHT which can handle situations where the legs are obscured. Any periodic signal, with period T, can be represented by a Fourier Series (FS). The motion of the thigh can be better represented by a FS as it does not generalise the motion to a specific funtion. Gathering evidence for a given feature using VHT techniques requires a structural model whose spatial positioning at a given time is governed by a motion model. As such, the VHT can lead to a new strategy for automatic gait signature extraction from an image sequence.

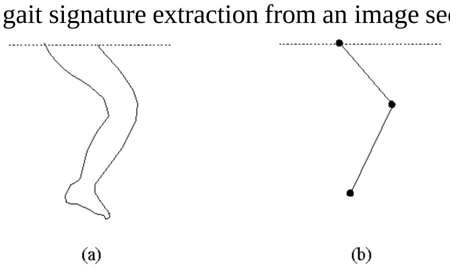

Figure 2: (a) Leg outline. (b) Pendulum model of the leg.

Fig. 3 shows the model of the thigh used for temporal evidence gathering. The thigh is described by a point $\bar{c}$ that represents the hip joint, or pelvis, and a line passing through $\bar{c}$ at an angle ϕ. The pelvis has horizontal and vertical velocities v_x and v_y, respectively. The angle ϕ of the line is governed by the hip rotation model. An arbitrary point $\bar{r}$ on the line is given by

$$\bar{r} = \bar{c}(t) + \lambda \bar{u}(t), \tag{4}$$

where $\bar{u}(t)$ is the unit vector of the direction of the line, $\bar{c}(t)$ is the position vector of the pelvis and λ can take any real value. The values of $\bar{u}(t)$ and $\bar{c}(t)$ are given by

$$\bar{u}(t) = (-\sin\phi(t), \cos\phi(t))$$
$$\text{and} \quad \bar{c}(t) = (c_{xo} + c_x(t), c_{yo} + c_y(t)),$$

where c_{xo} and c_{yo} are the initial co-ordinate values of the pelvis. Decomposing Eqn. 4 into x and y parts gives equations that describe the co-ordinates of points on the moving line;

$$r_x = c_{xo} + c_x(t) - \lambda \sin\phi(t) \tag{5}$$
$$r_y = c_{yo} + c_y(t) + \lambda \cos\phi(t). \tag{6}$$

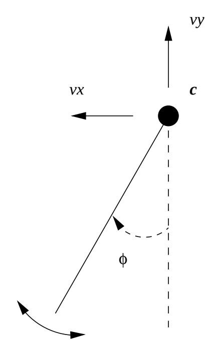

Figure 3: Model of thigh for temporal evidence gathering.

The structural model presents the edge point values of the model at any time instance. The motion of the model is described by the functions $c_x(t)$ and $c_y(t)$, which control the movement of the pelvis, and $\phi(t)$ which controls the hip rotation. The horizontal motion of the pelvis is mainly influenced by its transverse rotation and the average ground speed. The horizontal and vertical position $c_x(t)$ and $c_y(t)$ of the pelvis can be modelled as

$$c_x(t) = -\frac{A}{\omega_0}\sin\psi + \left(V_x t + \frac{A}{\omega_0}\sin(\omega_0 t + \psi)\right) \tag{7}$$
$$\text{and} \quad c_y(t) = V_y t, \tag{8}$$

where V_y reflects the slope of the walking surface and V_x is the average velocity, A is the amplitude and ψ is the phase of transverse rotation, and ω_0 is the angular velocity of the gait cycle.

For a constant walking speed, the hip rotation $\phi(t)$ is a periodic function with period T. A FS can represent any periodic signal with fundamental frequency $\omega_0 = \frac{2\pi}{T}$. The function $\phi(t)$ in Eqn. 5 and 6 can be modelled by using the rectangular form of the FS coefficients $a_k = b_k + jc_k$ to give an expression for $\phi(t)$;

$$\phi(t) = a_0 + 2\sum_{k=1}^{N}\left[b_k \cos k\omega_0 t - c_k \sin k\omega_0 t\right], \qquad (9)$$

where N is the number of harmonics. This equation reflects the periodic nature of the hip rotation.

Eqns. 5 and 6 are used to map edge points in the image space into a parameter space whose axes are the parameters of interest. A given edge point (r_x, r_y, t) votes for the possible values of c_{xo} and c_{yo} over an exhaustive range of the remaining parameters. All edge points in the image sequence are used, related edge points cast votes for the same parameter values, the ones relating them together. Peaks in the parameter space represent maximum likelihood estimates of the model parameters.

Cunado *et al.* [2] constructed a gait signature from the magnitude and phase components of the hip rotation. As this is a periodic signal, its Fourier transform (FT) is a train of impulses at the harmonics of the fundamental frequency. The amplitude of these impulses is given directly from the coefficients of the FS representation. As the model parameters include the FS coefficients b_k and c_k, the magnitude and phase components of the hip rotation can be found directly from the evidence gathering process;

$$|X(j\omega)| = 2\pi \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_0)\sqrt{b_k^2 + c_k^2} \qquad (10)$$

$$\arg(X(j\omega)) = \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_0)\tan^{-1}\left(\frac{c_k}{b_k}\right). \qquad (11)$$

Work done by Angeloni *et al.* [1] shows that the magnitude spectrum of the hip rotation drops to near-zero at around 5 Hz. Accordingly, only the low order components of the FS are required. The signature is now derived from the whole image sequence using a model based approach, avoiding the problems inherent in the earlier approach.

2.3 Implementation Considerations

As with all Hough based approaches, the VHT gait evidence gathering technique is subject to the problems of high computational complexity and large storage requirements. Even for the articulated gait model, the high dimensionality of the parameter space can render the technique impractical so a genetic algorithm is used to perform a heuristic search of the parameter space, rather than the exhaustive search performed by conventional Hough based techniques. This gives significant savings in the amount of memory and computational time required.

A GA based VHT (GaVHT) employs a population of chromosomes, each of which represents an instance of the feature of interest, namely a model of the legs. Individual chromosomes consist of a string of genes, each of which is a binary encoding of a particular model parameter, be it one of the 11 defining parameters of the simple gait model or one of the series of coefficients used in the Fourier representation.

The search for the optimal gait model parameters begins with a randomly initialised population of chromsomes. An iterative procedure is then used to produce a new generation of chromosomes, where chromosomes are selected according to their fitness and combined, using the probabilistic operators *cross-over* and *mutation*.

The fitness of an individual chromosome is determined by the degree of correlation between the instance of the gait model it represents and the image data. As the gait models described in this paper are dynamic, the degree of correlation is found by matching a multi-frame template, corresponding to an instance of the model, with the image sequence as a whole. The number of coincident pixels between the template and the image data thus determines the fitness of the particular chromosome.

As model pixels are mapped to image pixels to determine the fitness of a given chromosome, the GaVHT actually performs template matching. Nevertheless, the results are identical to the conventional VHT, [12], and the forward model-to-image space mapping is considerably more straightforward to implement.

Cross-over and *mutation* are probabilistic operations which directly affect the binary bit strings of the individual chromosomes. *Cross-over* uses a pair of parent chromosomes, cuts them at a random position in their bit strings and joins the opposing sections to form two new, child, chromosomes. The *mutation* operator randomly mutates the bit strings of the new chromosomes, complementing the affected bit. The probability of mutation is usually very low, of the order of $\frac{1}{1000}$, and is designed to ensure that some genetic variation propagates throughout subsequent generations.

As the parent chromosomes are selected in proportion to their fitness, only the fittest chromosomes, or those whose gait parameters best match the image data, are used to create the next generation. Accordingly, after a number of generations the whole population is concentrated in the high performance section of the parameter space and eventually converges on the optimal solution, namely the best estimate of gait parameters for the subject present in the image sequence.

3 Results

To demonstrate the efficacy of evidence-gathering as a means of automated gait description, a series of tests were performed. Both the articulated gait model and the Fourier based approach were used to determine the location of a subject's thigh in an image sequence. The sequences feature a subject walking parallel to the plane of the camera, against a plain background. Each frame of the sequence was digitised and then edge detected, using the Canny operator, assigned a time index value, relative to the first frame of the sequence, and used as the input data for the GaVHT. An example frame of a gait sequence is shown in Fig. 4, together with its corresponding edge map. This figure clearly demonstrates the problem of missing information, with the one leg almost completely occluding the other. Only by considering motion in the evidence gathering process, combined with the information present in the preceding and subsequent frames, is it possible to extract the position of the occluded leg.

(a) Gait image

(b) Edge map

Figure 4: Sample gait data

Fig. 5 shows the results of using the GaVHT, in conjunction with the articulated gait model of section 2.1, to extract the relative motion of a subject's thighs. Clearly, the GaVHT has successfully identified the initial position of the subject and tracked their position throughout the image sequence. Moreover, the nature of the angular variation associated with the thighs has been well determined, which is demonstrated by their extracted position in each frame of the sequence. Although, in some instances the influence of the lower leg, combined with the lack of knees in the articulated model, has led to an error in positional estimates of the thigh. Nevertheless, the GaVHT has determined the relative dynamics of the thighs with sufficient accuracy to enable the gait period of the subject to be identified. However, the slight errors in the estimated thigh position illustrate that the articulated model is not suited to use.

The result of the GaVHT implementing the Fourier model is shown Fig. 6. The extracted thigh model matches well the expected position of the moving thigh. The estimated position of the hip is marked by the line normal to the long line representing the thigh. Although the extracted model does not correlate exactly with the stripe on the subject's leg, the GaVHT has extracted the best fit model in-

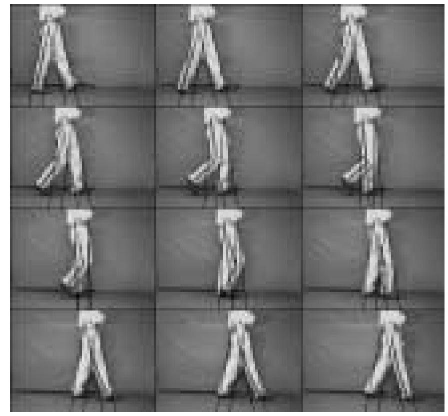

Figure 5: Simple Gait model results. Sequences run from top left to bottom right.

Figure 6: Fourier model results. Sequences run from top left to bottom right.

stance to the input data. The true location and motion of the thigh cannot be found due to the nature of the feature; being obscurred by the subject's clothing. Nervertheless, the Fourier model represents a good estimate of the structure and motion of the feature, and the results demonstrate a high fidelity to the input data. Moreover, the model is readily extensible to enable consideration to be given to independent motion of the lower leg, thus creating a more comprehensive gait signature. Classification on the original database of five subjects led to a 100% classification rate and we are currently working on a new extended database to quantify the nature of the results this new technique can deliver.

The thigh angle variation (*A* in figure 1) and variation of the hip position for the subject in Fig. 5 are shown in Figs. 7 and 8 respectively. In both instances, the articulated gait model and the Fourier model have been used to extract the required information. The thigh angle variation pat-

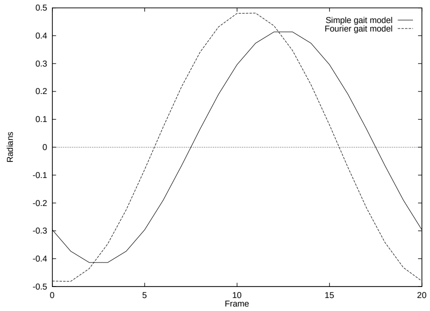

Figure 7: Extracted thigh angle variation pattern for subject in figure 5.

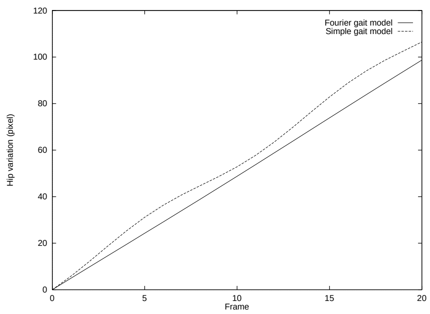

Figure 8: Extracted hip position variation for subject in figure 5.

terns determined by each gait model are clearly very similar, suggesting a high level of confidence in the extracted information. There are slight differences in the amplitude and phase of thigh angle variation of the two models, although this can be attributed to the improved accuracy of the Fourier based approach over the simple gait model. The results of Fig. 8, depicting the variation of the subject's hip position along the x axis throughout the sequence, parallel those of the thigh angle variation. As before, there is a slight difference in the extracted parameters of the two gait models, but there is an overiding simularity in the overall trend.

4 Conclusions

Model-based approaches to gait feature extraction provide a unified analytic procedure where results can be verified by matching with image data. Evidence gathering is well-known for its ability to handle noise and occlusion, as increasingly occurrent in image sequences. Evidence gathering techniques have been developed for a (moving) articulated model, representing human thighs, and to provide a gait signature automatically from the motion of the thighs. Independent results confirm that both techniques can extract features which match well with image data, and which match well with each other. Accordingly, there appears a rich potential for motion based evidence gathering techniques for extracting gait as a biometric.

References

[1] C. Angeloni, P. O. Riley, and D. E. Krebs. Frequency content of whole body gait kinematic data. *IEEE Transactions on Rehabilitation Engineering*, 2(1):40–46, 1994.

[2] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. *Lecture Notes in Computer Science (Proceedings of AVBPA'97)*, pages 95–102, March 1997.

[3] D. D. Hoffman and B. E. Flinchbaugh. The interpretation of biological motion. *Biological Cybernetics*, 42:195–204, 1982.

[4] P. S. Huang, C. J. Harris, and M. S. Nixon. Canonical space representation for recognising humans by gait or face. In *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 180–185, Tucson, Arizona, April 1998.

[5] G. Johansson. Visual perception of biological motion and a model for its anlaysis. *Percept. Psychophys.*, 14:210–211, 1973.

[6] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[7] M. P. Murray. Gait as a total pattern of movement. *American Journal of Physical Medicine*, 46(1):290–332, 1967.

[8] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone Joint Surgery*, 46-A(2):335–360, 1964.

[9] J. M. Nash, J. N. Carter, and M. S. Nixon. Dynamic feature extraction via the velocity Hough transform. *Pattern Recognition Letters*, 18:1035–1047, 1997.

[10] J. M. Nash, J. N. Carter, and M. S. Nixon. Extraction of moving articualted-objects by evidence gathering. In *Proceedings of the British Machine Vision Conference*, pages ?–?, 1998.

[11] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[12] J. Sklansky. On the Hough transform for curve detection. *IEEE Transactions on Computers*, 27:923–926, 1978.