

Automatic Gait Recognition

M. S. Nixon*, J. N. Carter*, J. M. Nash*, P. S. Huang*, D. Cunado*, and S. V. Stevenage⁺

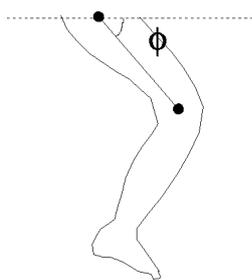
* Department of Electronics and Computer Science | ⁺ Department of Psychology
University of Southampton, Southampton SO17 1BJ, UK

Abstract

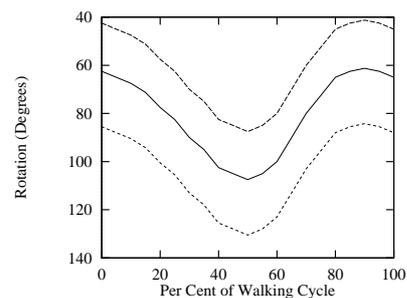
Gait is an emergent biometric aimed essentially to recognise people by the way they walk. Its advantages are that it is non-invasive and that it is less likely to be obscured since it appears to be difficult to camouflage, especially in cases of serious crime. Gait has allied subjects which lend support to the view that gait has clear potential as a biometric. Essentially, we use computer vision to find people and to derive a gait signature from a sequence of images. The majority of current approaches derive motion characteristics, which are then used for recognition. Early results by these studies confirm that there is a rich potential in gait for recognition. Only continued development in technique and in analysis will confirm whether its performance can match that of other biometrics.

1. Introduction

In many applications of person identification many established biometrics are obscured: the face may be hidden or at too coarse a resolution; the ears cannot be seen. However, people need to walk, so their gait is usually apparent. This motivates using gait as a biometric and it has recently attracted interest¹. The Oxford Dictionary definition of gait is: “*Manner of walking, bearing or carriage as one walks*” suggesting that studies can concentrate on different facets of a person’s walk. Apart from perceptibility, another attraction of using gait is that its motion can be hard to disguise.



(a) Angle



(b) Mean Rotation Pattern² (upper and lower lines indicate the standard deviation from the mean)

Fig. 1 Hip Rotation

Clearly there is a view that gait can be used to recognise individuals. That view is not new: Shakespeare used many adjectives to describe gait and in *The Tempest* [Act 4 Scene 1], Ceres observes “High’st Queen of state, Great Juno comes; I know her by her gait”, which suggests that he considered gait (or the cadence of footfalls) to be unique to the individual. While the examination of gait as a biometric is promising, there are limits to the utility of gait information. Most obviously, gait can be obscured by clothing such as skirts, and can be affected both by footwear and by the walking surface. The physical condition of a walker can also alter their gait. For example, pregnancy, fatigue, injury, and drunkenness will all bring about changes to the normal pattern of walking. Another apparent limitation to automated gait recognition is computing power but the continuing advance of processing power will mitigate this. Nevertheless, recent research conducted across allied disciplines including medicine, psychology and computer vision has underlined the wealth of information available from gait, and reaffirms its potential as a cue to identification.

2. Allied Research

2.1 Medical Studies

The aim of medical research has been to classify the components of gait for the treatment of patients with gait problems. Standard movement patterns were produced (2) for normal walkers, by attaching markers to

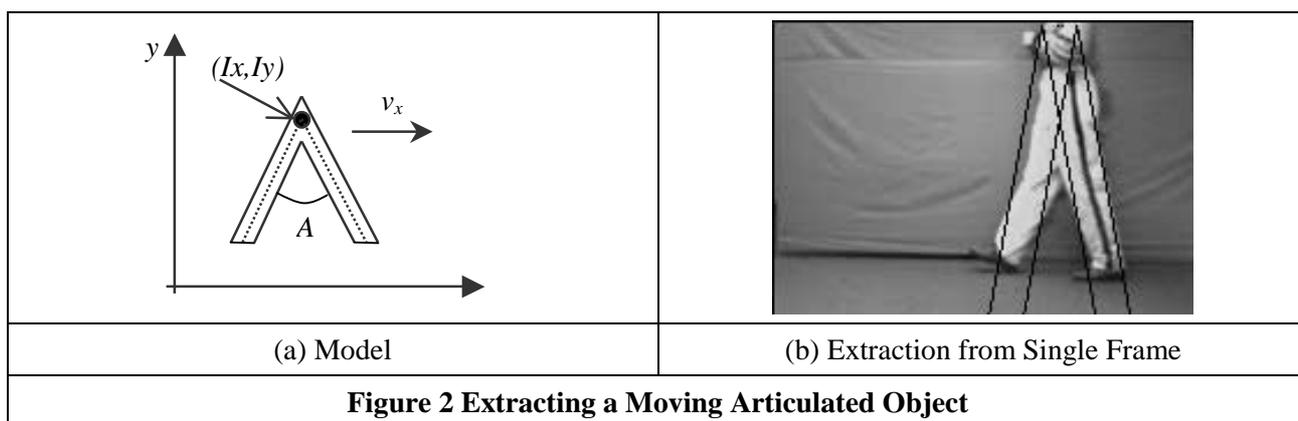
the joints of each individual. These patterns then served as the baseline for examination of the walking pattern of patients with gait abnormalities. In all, there appear to be twenty distinct gait components, some of which can be measured only from an overhead view of the subject. Among the most informative of these were the pelvic and thorax rotations (see Figure 1), ankle rotations, and pelvic tipping which, together with measures of spatial displacement revealed stability across repeated trials and differentiation across different walkers (2). From this work, Murray considered that gait could usefully be viewed as a “total walking cycle”, or as a periodic signal. Normal individuals showed a regular and systematic deviation from this periodic signal, while abnormal walking patterns showed a more extreme deviation from the standard pattern. The underlying principle, though, is that when the total gait cycle is considered, individual differences do emerge and thus gait can be considered ‘unique’. While the purpose of medical studies has been in the analysis of gait across populations, the study of gait by psychologists has focused much more clearly on gait as a biometric, and it is to this literature that we now turn.

2.2 Psychology of Gait

In the earliest studies of gait perception³ participants were presented with images produced from points of light attached to body joints. This technique generated what have become called ‘point-light displays’ which allowed researchers to investigate the perception of biological motion without the influence of biological form. When the points were viewed in static images, they were perceived to form a random arrangement such as star constellations or lights on a Christmas tree. However, when the point light displays were animated, they were immediately perceived as representing a human in motion. Later work confirmed that a human figure can be perceived very easily from such moving displays and more recently⁴ it has been shown that point light displays are sufficient for the discrimination of different types of motion such as jumping and dancing. As such, the human visual system appears adept at perceiving motion and further studies have shown how perceivers can even recognise gender and identity from gait. One recent study⁵, using video as well as point light displays, has shown that humans can indeed recognise people by their gait, and can learn gait patterns for the purposes of recognition. The study confirmed that, even under adverse viewing conditions, individual differences in gait could still be perceived. Consequently, psychology, too, presents us with a strong case for considering gait as a synchronous, symmetric pattern of movement from which individual difference can be perceived.

2.3 Modelling the Human Body and its Motion

Many studies have considered human motion extraction and tracking, though not for biometric purposes. The human body is an extremely complex object, being highly articulated and capable of a variety of motions. Stick figure models and volumetric models are commonly used for three-dimensional tracking, and the ribbon and blob models are less popular. There is an extensive literature of models and tracking¹.



These techniques appear not to have benefited from evidence gathering. A new velocity Hough Transform (HT) can find moving objects by evidence gathering. This accrues the HT’s known advantages of good performance in noise and/or occlusion. The velocity HT has been extended to locate moving articulated objects⁶, and the modelling strategy is illustrated in Figure 2(a). The position of the hip (I_x, I_y) can be considered as having a negligible vertical motion and a constant horizontal velocity v_{Ix} with oscillatory influence due to the rotation of the hip about the vertical axis. Accordingly, a model of the horizontal velocity of the hip is

$$v_x = v_{lx} + B_{lx} \cos(\omega_{lx} \pi t + \phi_{lx}) \quad (1)$$

The two legs can be modelled as a pair of articulated lines whose internal angle A varies as

$$A = B_A \sin(\omega_A \pi t + \phi_A) \quad (2)$$

The result of extracting the human from an image sequence is illustrated in Figure 2(b). This can be used to good effect to find a moving pair of articulated lines in an image sequence, with the HT's inbuilt advantages in terms of performance in noise and in occlusion (i.e., when walking behind a lamppost). In fact, it affords an appropriate initialisation for statistical approaches, since it can isolate precisely the region of interest.

3. Automatic Gait Recognition

3.1 Current Approaches

In what was perhaps the earliest approach to automatic recognition by gait, the gait signature was derived from the spatio-temporal pattern of a walking person⁷. Here, in XT dimensions (translation and time), the motions of the head and of the legs have different patterns. These patterns were processed to determine the body motion's bounding contours and then a five stick model was fitted. The signature was derived by normalising the fitted model for velocity and then by using linear interpolation to derive normalised gait vectors. This was then applied to a database of 26 sequences of five different subjects, taken at different times during the day. Depending on the values used for the weighting factors in a Euclidean distance metric, the classification rate varied from nearly 60% to just over 80%, equivalent to human performance and a promising start indeed.

Later, optical flow was used to derive a gait signature⁸, again automatically by computer vision, aiming not to use a model of a human walking but more to analyse motion content for recognition purposes. The motion features were defined as the variation of the first- and second-order moments of the dense optical flow, the 'shape of motion'. Then, the periodic structure of the features was described by the relative phase of the spectral components, referenced to one of the co-ordinates of the centroid. Recognition rates exceeding 90% were reported on a database of seven subjects, the rate depending on the number and type of motion feature used. Again, an encouraging result.

Another approach was aimed more at generic object-motion characterisation⁹, using gait as an exemplar of their approach. The approach was similar in function to spatio-temporal image correlation, but used the parametric eigenspace approach to reduce computational requirement and to increase robustness. The approach first derived body silhouettes by subtracting adjacent images, with further processing to reduce noise. Then, the images were projected into eigenspace, a well-established approach in automatic face recognition. Eigenvalue decomposition was then performed on the sequence of silhouettes where the order of the eigenvectors corresponds to frequency content. Recognition from a database of 10 sequences of seven subjects showed classification rates of 100% for 16 eigenvectors and 88% for eight, compared with 100% for the (computationally much more demanding) spatio-temporal image correlation approach, and with robustness to noise. We now consider two other recent approaches to automatic gait recognition, in detail.

3.2 Automatic Gait Recognition by Statistical Measurement

Eigenspace transformation (EST) based on Principal Component Analysis (PCA) or Karhunen-Loeve Transform is well known to offer a potent metric in automatic face recognition and in gait analysis, but without using data analysis to increase classification capability. PCA based on the global covariance matrix of the full set of image data is not sensitive to class structure in the data. Linear Discriminant Analysis (LDA), also called Canonical Analysis (CA), can be used to optimise the class separability of different face classes and improve the classification performance. The features are obtained by maximising between-class variation whilst minimising within-class variation. Unfortunately, this approach has a high computational cost. Moreover, the within-class covariance matrix obtained via CA alone may be singular. A new approach combines canonical space transformation based on CA with eigenspace transformation, for gait analysis^{10,11}. Combining EST with canonical space transformation (CST) reduces the data dimensionality and optimises the class separability of different gait sequences, simultaneously. This gives a 'statistical' approach to automatic gait recognition wherein the image sequence is described as a whole, neither by a model- nor by a motion-based approach, but by one which describes the motion content.

Given c training classes to be learnt, where each class represents a walking sequence of a single subject, $\mathbf{x}'_{i,j}$ is the j -th image (of n pixels) in class i and N_i is the number of images in i -th class. The total number of training images is

$$N_T = N_1 + N_2 + \dots + N_C \quad (3)$$

and the training set is represented by $[\mathbf{x}'_{1,1} \dots \mathbf{x}'_{1,N_1}, \mathbf{x}'_{2,1}, \dots, \mathbf{x}'_{c,N_c}]$. First, the brightness of each sample image is normalised and the mean \mathbf{m}_x is used to form an $n \times N_T$ matrix \mathbf{X}

$$\mathbf{X} = [\mathbf{x}_{1,1} - \mathbf{m}_x, \dots, \mathbf{x}_{1,N_1} - \mathbf{m}_x, \dots, \mathbf{x}_{c,N_c} - \mathbf{m}_x] \quad (4)$$

EST uses the eigenvalues and eigenvectors, generated by the data covariance matrix derived from the product $\mathbf{X}\mathbf{X}^T$, to rotate the original data co-ordinates along the direction of maximum variance. Calculating the eigenvalues and eigenvectors of the $n \times n$ matrix $\mathbf{X}\mathbf{X}^T$ is computationally intractable for typical image sizes. Based on singular value decomposition, we can compute the eigenvalues of $\mathbf{X}^T\mathbf{X}$ where the matrix size is $N_T \times N_T$ and is much smaller than the $n \times n$ matrix. The eigenvectors of $\mathbf{X}^T\mathbf{X}$ are used as an orthogonal basis to span a new vector space. Each image can be projected to a single point in this space. According to PCA, the image data can be approximated by taking only the largest eigenvalues and their associated eigenvectors. This partial set of k eigenvectors spans an eigenspace where the points $\mathbf{y}_{i,j}$ are the projections of the original images $\mathbf{x}_{i,j}$ by the eigenspace transformation matrix, $[\mathbf{e}_1, \dots, \mathbf{e}_k]$, as

$$\mathbf{y}_{i,j} = [\mathbf{e}_1, \dots, \mathbf{e}_k]^T \mathbf{x}_{i,j} \quad (5)$$

After this transformation, each original image can be approximated by the linear combination of these eigenvectors. In CST, the classes of the transformed vectors resulting from eigenspace calculation are used to calculate a scatter matrix \mathbf{S}_b , a within-class matrix \mathbf{S}_w and a between class matrix \mathbf{S}_b which reflect the dispersion, the variance and the variance of the difference, respectively. The objective of CST is to minimise \mathbf{S}_w and to maximise \mathbf{S}_b , simultaneously. This is achieved by minimising the generalised Fisher linear discriminant function \mathbf{J} , where

$$\mathbf{J}(\mathbf{W}) = \mathbf{W}^T \mathbf{S}_b \mathbf{W} / \mathbf{W}^T \mathbf{S}_w \mathbf{W} \quad (6)$$



Figure 3 Statistical Gait Recognition: Template and Processed Data

The ratio of the variances is maximised by the selection of the feature \mathbf{W} that minimises its rate of change. Supposing \mathbf{W}^* to be the optimal solution and that \mathbf{w}_i^* be its column vector which is a generalised eigenvector and corresponds to the i -th largest eigenvalue λ_i , then

$$\mathbf{S}_b \mathbf{w}_i^* = \lambda_i \mathbf{S}_w \mathbf{w}_i^* \quad (7)$$

After the generalised eigenvalue equation is solved, we obtain a set of eigenvalues and eigenvectors that span the canonical space where the classes are much better separated and the clusters are much smaller.

In application, this analysis is applied to human silhouettes derived by subtracting the background from the image and then thresholding the result. Fig. 3(a) shows a human template from the original sequence and Fig. 3(b) shows one of the eigenvectors extracted from the sequence of moving templates. The trajectories in eigenspace overlap and their centroids are very close together. After CST the trajectories are much better separated and with lower individual variance as in Fig. 3(c) (though for three dimensions only). Recognition from the canonical space is accomplished using the distance between the accumulated centre to each

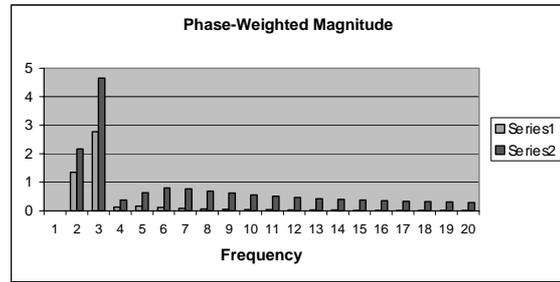
centroid. On five sequences of five people from the database of the Visual Computing Group, University of California, San Diego, an 85 % classification rate was achieved by CST alone whereas 100% was achieved with combined EST and CST (as evidenced by the cluster size and separation).

3.3 Automatic Gait Recognition by Feature-Based Measurement

An alternative approach to collecting the motion information in an image sequence is to find feature(s) and describe their motion. Gait signatures have been derived from frequency components of the variation in inclination of the thigh, as extracted by computer vision techniques¹². A bi-pendular model is used as the leg motion is periodic and each part of the leg (upper and lower) appears to have pendulum-like motion. Fourier theory allows periodic signals to be represented by spectra - the gait motion of the lower limbs can be described in such a way. The model of legs for gait motion allows these rotation patterns to be treated as periodic signals and so Fourier Transform techniques can be used to obtain a spectrum. The spectra of different subjects can then be compared for distinctive, or unique, characteristics. Earlier, the spectrum of the variation of the thigh was derived by edge detection, followed by line extraction, both derived from separate frames¹. Then, missing data were interpolated prior to derivation of the biometric signature.



(a) Frame of Sequence with Extracted Result



(b) Spectral Signatures for Two Subjects

Figure 4 Example of Model-Based Recognition (a) Original and (b) Signature

The approach has been superseded by one which generates the signature direct from edge detected versions of the image sequence¹². This again uses the Velocity HT to collect data over the entire sequence of images, but extracts the signature rather than just locating an articulated subject. The horizontal motion of the pelvis is modelled similar to Eqn. 1 whilst the inclination of the thighs is expressed by a Fourier series as

$$\phi(t) = a_0 + 2 \sum_{k=1}^N [b_k \cos(k\omega_0 t) - c_k \sin(k\omega_0 t)] \quad (8)$$

This describes the variation in the angle shown in Fig. 1(a). To avoid the high dimensional accumulator space consistent with the large number of parameters in Eqns. 1 and 8, the velocity HT is applied by way of a genetic algorithm. A single frame of a sequence with the result superimposed (i.e. the line calculated for that frame by the extracted parameters) is shown in Fig. 4(a). The resulting descriptors are used to calculate the transform data. The magnitude spectrum drops to near zero above the fifth harmonic, again agreeing with earlier (medical) work. The phase spectra are much more different than the magnitude spectra but some phase components carry little information since their respective magnitude component is very small. As such, the phase data is weighted by the magnitude data to reduce contributions where the magnitude is small. This gives the signature, as shown in Fig. 4(b) for two subjects.

The k -nearest neighbour rule was then used to classify the transform data using the 'leave one out' rule, for $k = 3$ and for $k = 1$. Four video sequences were acquired for each of ten subjects. The correct classification rates (CCR) are summarised in Table 1 which gives analysis for classification by magnitude spectra alone, and for multiplying the magnitude spectra by the phase, both for two values of k . Note that the magnitude component of the FT is time-shift invariant; it will retain its spectral envelope regardless of where in time the FT is performed. The phase component does not share this characteristic, and a time shift in the signal will change the shape of the phase envelope. Accordingly, the rotation patterns were aligned to start at the same point, to allow phase comparison. This is because the magnitude plots do not confer discriminatory ability whereas the phase plots do. The multiplication appears reasonable, since gait is not characterised by extent of flexion alone, but is controlled by musculature which in turn controls the way the limbs move. Accordingly, there is physical constraint on the way we move our limbs. However, we cannot

use phase alone, since some of the phase components occur at frequencies for which the magnitude component is too low to be of consequence. By multiplication of the spectra, we retain the phase for significant magnitude components. Clearly, in this analysis, using phase-weighted magnitude spectra gives a much better classification rate (100%) than use of magnitude spectra alone (80%), for both values of k .

No. of Nearest Neighbours	Magnitude CCR	Phase x Magnitude CCR
$k = 1$	80 %	100 %
$k = 3$	80 %	100 %

Table 1. Model-Based Classification Performance.

4. Conclusions

Throughout this article, we have argued for the use of gait as a biometric. Its use is attractive because it is non-invasive and less likely to be disguised than other biometrics. Allied studies in physiology and psychology suggest that unique gait patterns can be modelled and recognised, and in computer vision, a number of approaches have modelled the body and tracked it through image sequences. Of more direct importance to the field of biometrics are the recent developments suggesting that computer vision can be used to recognise people by their gait. The majority of current approaches are motion-based, combining the image sequence by its motion or by statistical analysis. Only one technique is feature-based and its results can clearly be identified with the data from which they were derived.

The research has focused more on development of new techniques for the task, but results suggest that further development is certainly warranted, both in application and development. Clearly, gait would benefit from an established database on which to assess new developments both in recognition procedure and in technique. As such, future work will strengthen and develop methodology and will establish more precisely the results that can be achieved by this new biometric.

Acknowledgements

We gratefully acknowledge financial and research support from the Home Office.

References

- 1 M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, S. V. Stevenage, Automatic Gait Recognition, *In: A. K. Jain, R. Bolle and S. Pankanti Eds.: Biometrics: Personal Identification in Networked Society*, pp. 231-250, 1999
- 2 M. P. Murray, Gait as a total pattern of movement, *American Journal of Physical Medicine*, **46**(1), pp. 290-332, 1967
- 3 G. Johansson, Visual perception of biological motion and a model for its analysis, *Perception and Psychophysics*, **14**, pp. 201-211, 1973
- 4 G. P. Binham et al., Dynamics and the Orientation of Kinematic Forms in Visual Event Recognition, *J. Experimental Psychology: Human Perception and Performance*, **21**(6) pp. 1473-1493, 1995
- 5 S. V. Stevenage, M. S. Nixon and K. Vince, Visual analysis of gait as a cue to identity, To appear *Applied Cognitive Psychology*, 1999
- 6 J. M. Nash, J. N. Carter and M. S. Nixon, Extracting moving articulated objects by evidence gathering, *Proc. BMVC 98*, **2**, pp. 609-618, Southampton, Sept. 1998
- 7 S. A. Niyogi, E. H. Adelson, Analyzing and recognizing walking figures in XYT, *Proc. CVPR.*, pp. 469-474, 1994.
- 8 J. J. Little and J. E. Boyd, Recognizing people by their gait: the shape of motion, *Videre*, **1**(2), 1998, at <http://mitpress.mit.edu/e-journals/VIDE/001/v12.html>
- 9 H. Murase and R. Sakai, Moving object recognition in eigenspace representation: gait analysis and lip reading, *Pattern Recognition Letters*, **17**, pp 155-162, 1996
- 10 P. S. Huang, C. J. Harris and M. S. Nixon, Recognizing humans by gait via parametric canonical space, To appear: *Artificial Intelligence in Engineering*, 1999
- 11 P. S. Huang, C. J. Harris and M. S. Nixon, Human gait recognition in canonical space using temporal templates, To appear: *IEE Proc. Image, Vision and Signal Processing* 1999

12 D. Cunado, J. M. Nash, M. S. Nixon and J. N. Carter, Gait Extraction and Description by Evidence-Gathering, *Proc. 2nd Int. Conf. on Audio- and Video-Based Biometric Person Authentication AVBPA99*, Washington D.C., pp 43-48, March 1999