Model-Based 3D Gait Biometrics

Gunawan Ariyanto, Mark S. Nixon
School of Electronics and Computer Science
University of Southampton
Southampton UK, SO17 1BJ
ga08r, msn@ecs.soton.ac.uk

Abstract

There have as yet been few gait biometrics approaches which use temporal 3D data. Clearly, 3D gait data conveys more information than 2D data and it is also the natural representation of human gait perceived by human. In this paper we explore the potential of using model-based methods in a 3D volumetric (voxel) gait dataset. We use a structural model including articulated cylinders with 3D Degrees of Freedom (DoF) at each joint to model the human lower legs. We develop a simple yet effective model-fitting algorithm using this gait model, correlation filter and a dynamic programming approach. Human gait kinematics trajectories are then extracted by fitting the gait model into the gait data. At each frame we generate a correlation energy map between the gait model and the data. Dynamic programming is used to extract the gait kinematics trajectories by selecting the most likely path in the whole sequence. We are successfully able to extract both gait structural and dynamics features. Some of the features extracted here are inherently unique to 3D data. Analysis on a database of 46 subjects each with 4 sample sequences, shows an encouraging correct classification rate and suggests that 3D features can contribute even more.

1. Introduction

All methods in gait biometrics generally can be classified as model-free or model-based approaches [5]. In the model-free approaches, some methods only use the moving shape whilst others combine the shape and the motion. The model-free approaches also heavily depend on the extracted silhouette and statistical methods. Model-based approaches can use prior information of the human structure or a known model (such as gait motion/kinematics model, or a physical pendulum model) to emulate human gait. So far, there have been more model-free techniques than the model-based ones. Even though the model-based approaches are computationally more complex, they can

have advantages such as immunity to noise, slight change of view, and the effects of clothing.

Most research in gait biometrics has been conducted using 2D datasets and/or 2D approaches. The reasons why only few studies concern 3D gait biometrics is perhaps due to complexity and the lack of a publically available 3D dataset. However, imodern computing power has made it possible to study gait biometrics using 3D data with 3D approaches. Clearly, 3D gait data conveys more information than 2D data and it is also the natural representation of human gait. Moreover, 3D data is inherently view invariant because we can synthesize any view. We hypothesise that by using 3D data we can explore more unique factors in human gait.

Yamauchi et al [8] used an active vision sensor to capture the 3D data and then tried to fit the 3D model into the data to obtain the kinematics information. They used a laser range sensor and collected subject poses which represent only the key frames in a gait cycle. In this case, the 3D data are high quality. The complete gait sequence is synthesized by interpolation of joint positions and their movements from the fitted body models. The experimental results show high recognition rates, though currently there are only six subjects used in the experiments.

Seely et al [7, 6] created the Soton 3D gait dataset and then successfully used it for view-invariant gait recognition using model-free analysis. They employed simple 2D average silhouette-based methods using view point projection techniques to convert 3D data into 2D view-invariant data. There were 3 different view-point projections used: side-on, front-on, and top-down projection. The results showed that using 3D gait data can lead to high accuracy (99.6%) and the best performance was achieved by using a combination of projected views[6]. Another study [10] has similar data to that of the Soton 3D dataset, and used it for human tracking rather than recognition.

Given the 3D gait data from the Southampton Biometric Tunnel, we develop the first 3D model-based method to extract gait features from this 3D data. Our model is structural

in that it uses basic gait parameters and articulated cylinders which are fited to the voxel data thereby accruing the advantages of the richer set of features available in 3D and we present introductory results for this new approach.

2. Southampton multi-biometric tunnel

The University of Southampton multi-biometric tunnel provides a constrained environment and is designed for use in high throughput environments such as airports and for the collection of large datasets [4]. The tunnel uses eight synchronised IEEE1394 cameras at 30 fps to capture gait data as shown in Fig. 1. As a subject walks through, the data is acquired automatically. Using a visual hull shape from silhouette reconstruction algorithm, the tunnel is able to produce the 3D voxel gait data. The shape from silhouette reconstruction is simply the calculation of the intersection of projected silhouettes from each camera. Fig. 2 shows a frame of 3D voxel gait data produced by the tunnel.
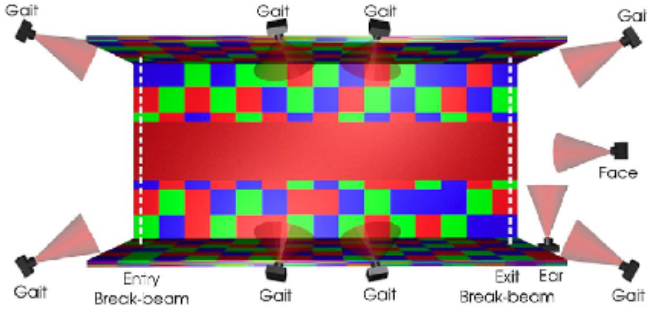

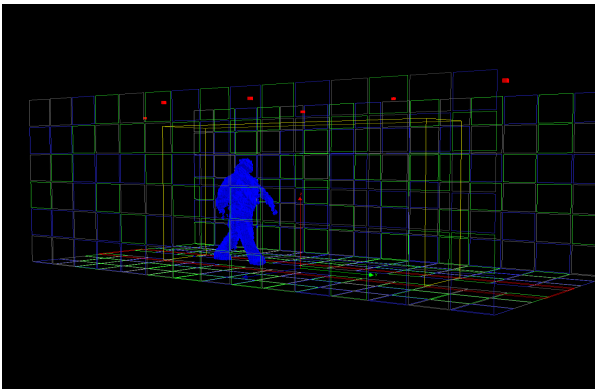Figure 1. Soton Multi-biometrics Tunnel (from [7])

Figure 2. Sample data acquired from Tunnel after reconstruction (from [7])
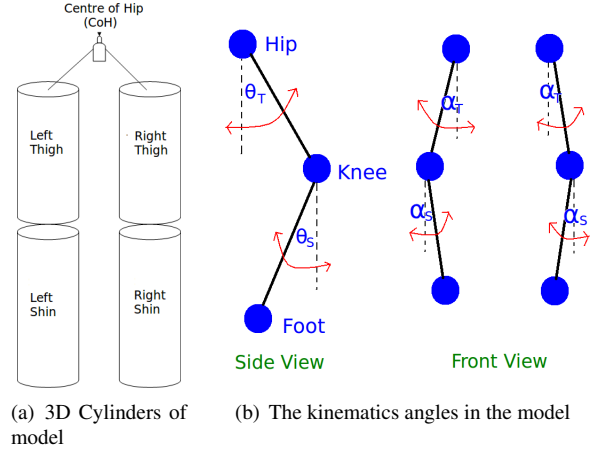

(a) 3D Cylinders of model

(b) The kinematics angles in the model

Figure 3. 3D Human Model

3. Gait feature extraction

3.1. Gait model

Rather than using a simple 2D stick figure model, we use an articulated 3D points cylinders as our gait model. As shown in Fig. 3, there are four cylinders and four joints for modelling the human legs. These cylinders correspond to the thighs and the shins. Two joints with 3D DoF are used at each leg to connect the shin to the thigh and the thigh to the hip. Within each cylinder in the model, we generate a regular cloud of points and a local coordinate frame is defined with the origin at the top of the cylinder. These origin points also correspond to the center of rotation of each body part. The global coordinate system originates at the center of hip, which is currently modelled as a single 3D static point. The length and the initial positions of the cylinders are estimated using subject height and human anthropometric data[9, 3].

The gait kinematics are extracted using the thigh and shin angles of the cylinder gait model and in total we can extract up to eight kinematics angles. The rotation of the cylinder in transversal plane will give no effect, therefore we only consider two rotation angles: θ and α in Fig. 3(b) are rotation angles in the sagittal and the frontal plane, respectively; T, S, R, and L denote the thigh, shin, right, and left (θ_{RT} denotes the sagitall angle of right thigh).

3.2. Preprocessing and structural features extraction

In the preprocessing stage, we seek to process 3D voxel data sequence to obtain a 3D bounding box. This information will be used to estimate the gait cycle period and the subject's height, to initialise the gait model and model-fitting process.

We also seek to extract some structural features of gait using the 3D voxel data. So far we are able to estimate

three structural gait parameters: height, stride length, and footprint pose. The footprint pose can be considered as an novel gait feature because it is inherently unique to the 3D (markerless) gait data.

3.2.1 3D bounding box

Let the voxels representing a person at frame j be $V^j(x, y, z)$ and the location of each voxel is $v_i = (x_i, y_i, z_i)$. The bounding box is the smallest which encloses the connected voxel data. It can be extracted from values of x (frontal/length), y (sagittal/width), and z (transverse/height) axis for a filled voxel as described in Eqn. 1. Fig. 4 shows typical bounding box BB extracted here for x, y, and z axis respectively.

$$BB_a = \max_i a_i - \min_i a_i \quad \text{where} \quad a \in x, y, z \quad (1)$$
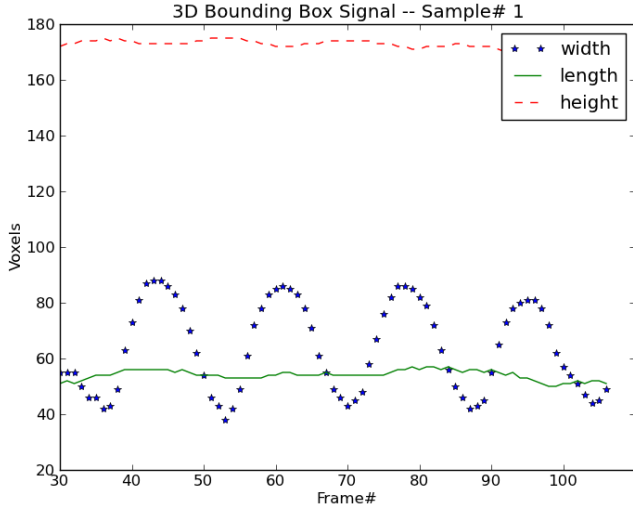

Figure 4. 3D Bounding Box

3.2.2 Gait cycle estimation

Using the 3D bounding box data, we can determine the starting and ending frame in a gait cycle. The heel strike pose corresponds with the peak of the filtered width of the 3D bounding box signal, as shown in Fig. 4. The start frame is that for which a maximum width bounding box occurs (frame# 44) and the end frame is the second next maximum width (frame# 78).

3.2.3 Subject height

The object height h is derived from the average of the transverse plane bounding box in one gait cycle (N frames) as

shown in Eqn. 2.

$$h = \frac{\sum_{j=1}^{N} BB_z^j}{N} \quad (2)$$

3.2.4 Footprints orientation

We extract the footprint image using a cross-sectional plane of the voxel data. During the walking cycle, each heel is in a static position on the ground for around half gait cycle. Therefore, we project one or more planes around the ankle into the ground plane G^j for each frame and then accumulate it to derive the footprints image F as described in Eqn. 3. We use Otsu thresholding and morphology operation to obtain clean footprint and locate the heel strike positions. In order to obtain the orientation angle for each foot print, we employ a moment analysis as described in [1]. The footprints feature consists of four orientation angles $[\beta_1, \beta_2, \beta_3, \beta_4]$.

$$
\begin{aligned}
G^j(x, y) &= V^j(x, y, z = 0) \\
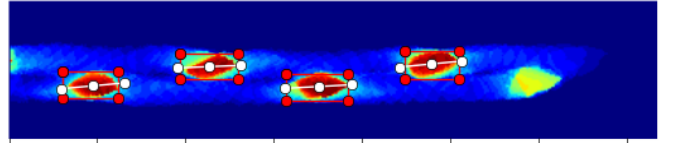F(x, y) &= \sum_{j=1}^{S_j} G^j(x, y) \quad (3)
\end{aligned}
$$

Figure 5. Footprints and its orientation image

3.2.5 Stride length

Using the footprint information above, we are able to extract the stride length by calculating the distance between the adjacent footprints of the same foot.

3.3. Dynamic features extraction

We have developed a model-fitting framework to collect some dynamics features of gait. So far we have collected two kind of 3D dynamic features, i.e. centre of hip (CoH) and gait kinematics. The model-fitting framework use a correlation filter to generate a correlation energy map between the gait model and the voxel data. Dynamic programming is used to extract the gait kinematics trajectories from the energy maps by selecting the most efficient path in the whole sequence.
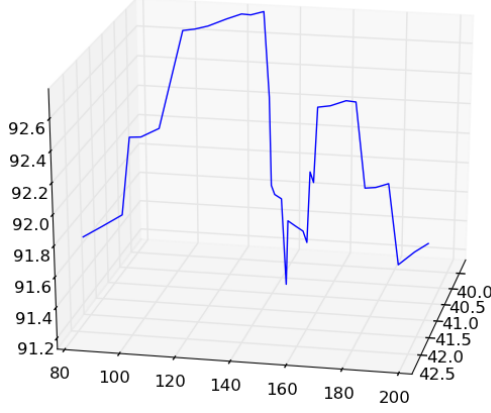
Figure 6. Center of hip in one sequence

3.3.1 Model-fitting

Our shape model will be fitted into the 3D data using a correlation algorithm. The cylinders in the shape models behave like a 3D "filter" correlated with the 3D data. In this process, two objects are moved over each other to generate a distance measurement describing where they best match (are closest to each other). In our case we rotate each cylinder of the shape model.

Model-fitting initialisation and CoH extraction

In order to fit the gait model into the voxel data, we need to find the position of CoH in the data. This CoH will be used as a stationary point of the gait model in the fitting process. We estimate the CoH (c_x, c_y, c_z) by using the mean of the voxel data and the subject heigh as described in Eq. 4. Where $I(x, y, z)$ is the value/intensity of the voxel i.e. $\{0, 1\}$ and h is the subject height. A sample of the extracted CoH is shown in Fig. 6.

$$
\begin{aligned}
c_x &= \frac{1}{s_x s_y s_z} \sum_{x=1}^{s_x} \sum_{y=1}^{s_y} \sum_{z=1}^{s_z} x I(x, y, z) \\
c_y &= \frac{1}{s_x s_y s_z} \sum_{x=1}^{s_x} \sum_{y=1}^{s_y} \sum_{z=1}^{s_z} y I(x, y, z) \\
c_z &= 0.530 \times h
\end{aligned} \tag{4}
$$

Correlation filter and energy map

$M = \{m_i\}$ is the 3D thigh cylinder points of the model behaving as "thigh-filter" and $D = \{d_i\}$ is the 3D data points derived from the 3D voxel data. The rotation operation of the thigh has 2 DoF using θ and α angles. We use a transformation matrix $\mathbf{T}(\alpha, \theta)$ as in eq. 5 to implement this rotation

operation. This allows us to parameterise the correlation process with α and θ (as in Fig. Figure:model2).

$$
\mathbf{T}(\alpha, \theta) =
\begin{pmatrix}
\cos(\alpha) & 0 & \sin(\alpha) \\
0 & 1 & 0 \\
-\sin(\alpha) & 0 & \cos(\alpha)
\end{pmatrix}
\times
\begin{pmatrix}
1 & 0 & 0 \\
0 & \cos(\theta) & -\sin(\theta) \\
0 & \sin(\theta) & \cos(\theta)
\end{pmatrix}
\tag{5}
$$

We use the least squares distance to measure the correlation energy between the model and the data. The correlation energy map $\mathbf{E}$ is defined by calculating the correlation for each possible parameter value (α and θ) over s_m points as described in Eqn. 6.

$$
\begin{aligned}
E[\alpha, \theta] &= \sum_{i=1}^{s_m} dist(m_i \times \mathbf{T}(\alpha, \theta), D)^2 \\
dist(m, D) &= \min_i dist(m, d_i)
\end{aligned} \tag{6}
$$

The range of θ and α are ± 45 and ± 7.5 degree respectively. We choose a resolution of 6 and 1 degree for θ and α. Fig. 7 shows the correlation energy map of sample 1 at frame 65. The darker the color means the lower energy value and it has better fit.
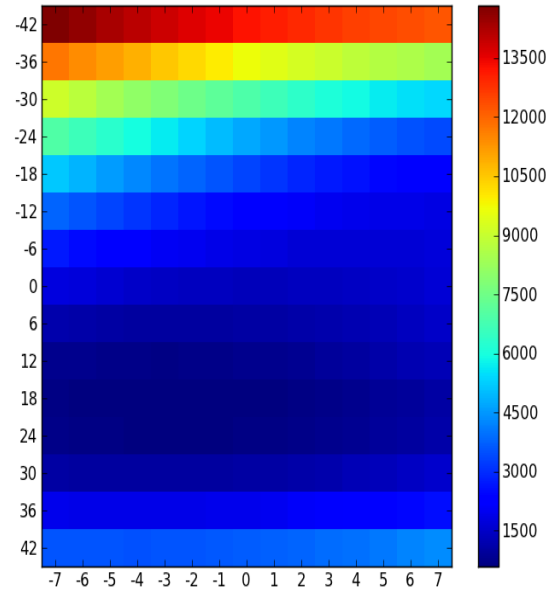

Figure 7. Correlation Energy Map

Dynamic programming

Since gait is a spatio-temporal activity, we process the gait sequence as a whole rather than frame by frame thus making it possible to filter noise. We can phrase the problem of

extracting the best gait kinematics trajectory as a shortest path problem.

We choose dynamic programming to select the optimal trajectory based on the correlation energy maps. A correlation energy map $E = [e(j,k)]$ can be considered as a matrix with row index j and column index k. The accumulation of this map in one gait cycle will form a 3D array. The first step is to transverse the entire map and compute the cumulative minimum energy M for all possible connected links for each entry. Eq. 7 shows an example of using 9-connected links to connect between two energy maps in adjacent frames.

$$
\begin{aligned}
M(i,j,k) \;=\; & e(i,j,k) + min(M(i-1,j-1,k-1), \\
& M(i-1,j-1,k), M(i-1,j-1,k+1), \\
& M(i-1,j,k-1), M(i-1,j,k), \\
& M(i-1,j,k+1), M(i-1,j+1,k-1), \\
& M(i-1,j+1,k), M(i-1,j+1,k+1))
\end{aligned}
\tag{7}
$$

At the end of this process, the minimum value of the last matrix in M will indicate the end of the minimal connected trajectories in one gait cycle. Hence, in the second step we backtrack from this minimum entry in M to find the path of the optimal trajectory.

Kinematics features extraction

We use a hierarchical approach of model-fitting to extract the kinematics features. Firstly, we fit the thigh cylinder into the data and then extracting the thigh kinematics angle. Secondly, we use this thigh kinematics angle to update the knee joint and shin position. Finally, we fit the shin cylinder to get the shin. The hierarchical approach has an advantage of reducing the optimal pose search space and the complexity of computation.

4. Evaluation and Analysis

The database contains 46 different subjects; four recordings of each subject were taken, giving a total of 184 samples.

### 4.1. Model-fitting result

Fig 8 shows the model-fitting result. The cylinder gait model is able to fit into the voxel data even though the voxel data quality is imperfect as shown in Fig. 8(a) and 8(b). Fig. 8(c) shows limitations associated with the hierarchy and the data in that the thigh is extracted well, but the shin has less fidelity to the 3D data.

Fig 9 shows all θ_{RT} and θ_{RS} from all samples in the database and shows that human walking has a large variance of kinematics features.
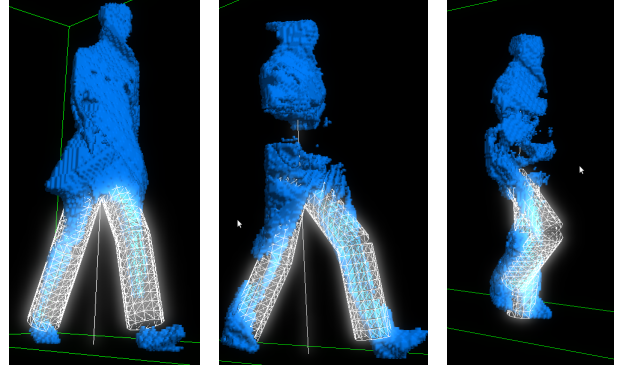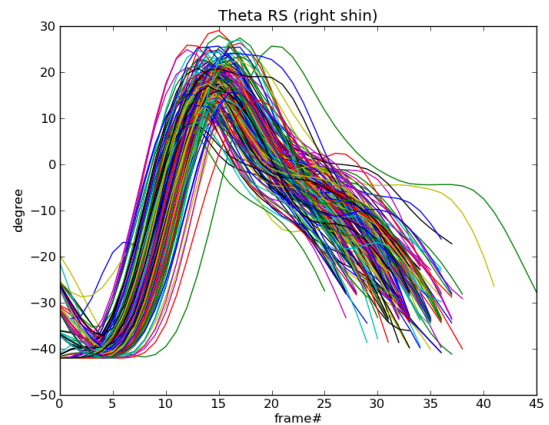

(a) Sample# 78   (b) Sample# 112   (c) Sample# 109

Figure 8. Model fitting result
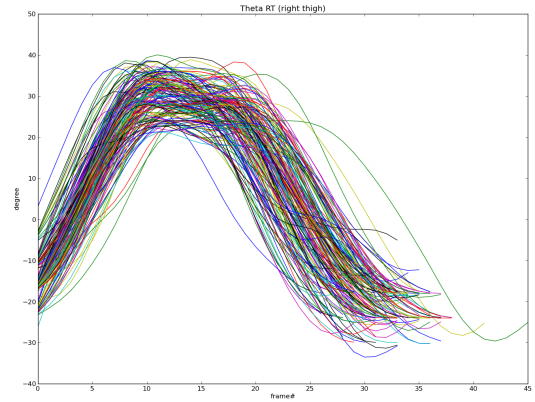
(a) Theta right thigh (θ_{RT}) angles

(b) Theta right shin (θ_{RS}) angles

Figure 9. The extracted kinematics of theta (θ)

Feature	Gait Signature	Type Data	Distance
Height	Height	Scalar	Euclidean
Stride	Stride	Scalar	Euclidean
Footprints	Footprints	Vector	Euclidean
Gait Kinematics	DFT	Vector	Vary
Centre of Hip	DTW	Vector	Euclidean

Table 1. Gait signature.

4.2. Gait Signatures

Gait signatures are created by selecting and combining the statics parameters and the dynamic features. Table 1 shows all gait signatures used in this experiment.

For gait kinematics we extract the gait signature using The Discrete Fourier Transform (DFT). After applying the DFT to the gait kinematic data, we will obtain new information about the frequency components of the subject's gait. These frequency components will be used as gait features.

The DFT is defined by Equation 8. A series x of N complex numbers is transformed into another series X of N complex numbers. The number X_i represents the ith multiple of the fundamental frequency. We use the Fast Fourier Transform (FFT) with 64 points.

$$X_i = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi i}{N}ik} \qquad i = 0, \dots, N-1 \qquad (8)$$

The CoH features behave like a time series datasets with 3D values. We use Dynamic Time Warping (DTW) algorithm to measure the similarity of the COH features. DTW involves the creation of a matrix in which the distance between every possible combination of time instances is stored. This distance is calculated in terms of the feature values of the points using euclidean distance function.

4.3. Classification

We classified the subject based on two data types:

1. Structural features: height, stride length and footprint pose.

2. Dynamic features: frequency components of the kinematic angles

k-NN was used as a classifier with leave one out cross validation. The evaluation was basic in that more advanced approaches such as fusion or classifier could putatively improve performance.

For the kinematics gait signature we used four distances functions of the magnitude of the FFT components, and of the product of magnitude and phase. Three different values of k were used: 1, 3, and 5. We also tried five different

methods of data normalization. k=1 gave the best performance and the data normalization methods were not able to significantly improve the performance.

Table 2 shows the classification rate using structural parameters showing that height is the most discriminatory feature. The best performance of 56.0% was achieved for k=1 with height and footprint. It is interesting to note that the footprint feature can increase recognition capability.

Chosen Features	k=1	k=3	k=5
Height	**41.3**	37.5	37.5
Stride	7.1	4.4	3.8
Footprint	17.4	7.1	12.5
Height & Stride	42.9	28.0	31.0
Height & Footprint	**56.0**	44.6	38.0

Table 2. k-NN classification results (%) for structural features

Table 3 shows the detail of our classification results using the kinematic features. The comparison of distance methods shows that the Euclidean distance to be best. The best recognition performance is achieved by using the α_{RT} angle which is the movement of right thigh in the frontal plane, confirming an earlier study of the potency of the front view for gait recognition [2].

Chosen Features	Chosen Distance Methods			
	Mag	Mag-Phase	Eucl	DTW
θ_{RT}	30.4	19.0	28.8	19.0
α_{RT}	26.6	33.7	**54.9**	30.9
θ_{LT}	38.5	27.1	33.1	21.7
α_{LT}	22.2	30.9	40.7	28.8
θ_{RS}	25.5	15.2	23.9	8.1
α_{RS}	13.0	19.0	28.8	18.4
θ_{LS}	20.1	16.8	21.7	5.4
α_{LS}	18.4	27.7	40.7	22.2

Table 3. Correct classification rate (%) for kinematics feature with k=1

In order to determine the best kinematics features, we evaluated all possible feature subsets. In this case, with 8 kinematics angle features 255 combinations of features were used. Again all classification experiments used leave one out cross validation. Table 4 shows the result of four best kinematics feature subsets. From this table, we can see that the best kinematics features subset is (α_{RT}, α_{LT}, and α_{LS}) with 76.6% correct classification. After combining this best kinematic feature subset with CoH feature we obtained 79.4% correct classification.

Table 5 shows the classification results using the dynamics features. The CoH cannot achieve high recognition when used alone, but can be combined with the other dynamic features to good effect. On the other hand, the kine-

Features Subset	Correct Class. Rate (%)
$\alpha_{RT}, \alpha_{LT}, \alpha_{LS}$	76.6
$\alpha_{RT}, \alpha_{LT}, \alpha_{RS}, \alpha_{LS}$	74
$\theta_{RT}, \alpha_{RT}, \theta_{LT}, \alpha_{LT}, \alpha_{RS}, \alpha_{LS}$	72.3
$\theta_{RT}, \alpha_{RT}, \theta_{LT}, \alpha_{LT}, \alpha_{LS}$	71.2

Table 4. Four best kinematics features subsets

matics features, especially the frontal angles, have much potential to be used as main gait features.

Chosen Features	k=1	k=3	k=5
CoH	**20.1**	11.4	12.0
Kinematics	**76.6**	57.6	54.0
CoH & Kinematics	**79.4**	65.8	63.4

Table 5. k-NN classification results (%) for dynamic features

This results show that our kinematics features have greater discriminatory capability than the structural parameters. Although our best correct classification rate does not yet equal that achieved by a model-free approach [6], our system is directly related to the subject's gait and not their clothing or appearance and allows for deeper analysis of the gait mechanisms and can be related to biomechanical models of gait.

5. Conclusions and further work

We have developed the first model-driven approach to gait biometrics for use with 3D data. The model uses structural gait parameters such as height, stride and footprint poses (the latter being unique to 3D data and is not available in data wherein the subject walks in a plane normal to that of the camera). We use articulated cylinders to fit the 3D volumetric subject data, and this provides the inclinations of the shin and thigh in 3D. By Fourier analysis of the trajectories of these angles, recognition capability can be increased. These results are largely introductory and show that discriminatory capability can be achieved. Clearly results could be improved by a more sophisticated classifier and the hierarchical structure used herein, and by deploying feature subset selection but the basic results here confirm that discriminatory capability can be achieved by this new approach. The approach also reveals a selection of measures which are new to gait analysis. Naturally, the model-based approach could be used to complement the discriminatory potential of the model-free approach. We also intend to further study the relationship of this model to biomechanics, and the advantage of its relation to human gait in that it does not depend as much on appearance and clothing as do the model-free approaches.

References

[1] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, (Q2), 1998. 3

[2] M. Goffredo, J. N. Carter, and M. S. Nixon. Front-view gait recognition. In *IEEE Second International Conference on Biometrics: Theory, Applications and Systems (BTAS 08)*, October 2008. 6

[3] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *IEEE International Conference on Automatic Face and Gesture Recognition 2008*, September 2008. 2

[4] L. Middleton, D. K. Wagg, A. I. Bazin, J. N. Carter, and M. S. Nixon. A smart environment for biometric capture. In *IEEE Conference on Automation Science and Engineering*, 2006. 2

[5] M. Nixon, T. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, 2006. 1

[6] R. Seely, M. Goffredo, J. N. Carter, and M. S. Nixon. View invariant gait recognition. *Handbook of Remote Biometrics: for Surveillance and Security*, April 2009. 1, 7

[7] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter, and M. S. Nixon. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *Biometrics: Theory, Applications and Systems*. IEEE, September 2008. 1, 2

[8] K. Yamauchi, B. Bhanu, and H. Saito. Recognition of walking humans in 3d: Initial results. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 45–52, June 2009. 1

[9] J.-H. Yoo, M. S. Nixon, and C. J. Harris. Extracting human gait signatures by body segment properties. In *Proceedings of the Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 35–, Washington, DC, USA, 2002. IEEE Computer Society. 2

[10] Z. Zhang, H. S. Seah, C. K. Quah, A. Ong, and K. Jabbar. A multiple camera system with real-time volume reconstruction for articulated skeleton pose tracking. In *Proceedings of the 17th international conference on Advances in multimedia modeling - Volume Part I*, MMM'11, pages 182–192, Berlin, Heidelberg, 2011. Springer-Verlag. 1