Marionette Mass-Spring Model for 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

Though interest in gait biometrics continues to increase, there have as yet been few approaches which use model-based algorithms with temporal 3D data. In this paper we describe a new 3D model-based approach using a marionette and mass-spring model to gait biometrics with 3D voxel gait dataset. To model the articulated human body, we use a stick-figure which emulates the marionettes' motion and joint structure. The stick-figure has 11 nodes representing the human joints of head, torso, and lower legs. Each node is linked with at least one other node by a spring. The voxel data points in the next frame have a role as attractor which able to generate forces for each node and then iteratively warp the model into the data. This process is repeated for successive frames for one gait period. The motion kinematics extracted from this tracking process are projected into the sagittal and the frontal plane and used as a gait feature via the discrete Fourier transform. We use 46 subjects where each subject has 4 sample sequences and report encouraging initial gait classification results.

1. Introduction

Gait biometrics enjoys research interest due to its unique benefits compared to other biometric modalities. Some of the benefits are that it is non-invasive, perceivable at a distance, easy to set up in public area and can be hard to conceal [11]. Gait can also be used in fusion with other biometrics to enhance the overall performance [6].

All methods in gait biometrics generally can be classified into model-free or model-based approaches [11]. The model-free approaches heavily depend on the extracted silhouette and statistical methods whilst model-based approaches can use a known gait model as prior information.

Most research in gait biometrics has been conducted using 2D datasets and 2D approaches even tough there are some unique advantages of using 3D. The 3D gait datasets can convey more information and are inherently view invariant as we can synthesize any view. Working in 3D can also bring more consistency regarding the occlusion and multi-interpretation problems.

In the 3D model-free approaches, Seely [13, 12] created the Soton 3D gait dataset and then successfully used it for view-invariant gait recognition using silhouette analysis. He employed simple 2D averaged silhouette methods using view point projection techniques to convert 3D data into 2D view-invariant data. There were three 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 [12].

The 3D model-based approaches can use prior information of the human structure to emulate the real human gait. This model usually is used to tract the object data in order to recover the body pose and obtain the motion trajectories. Another 3D recognition study [1] used a 3D model based on using the match between volumetric cylinders and the voxel data, by correlation filter and dynamic programming.

There are some general approaches for model-based 3D human tracking and pose recovery, though not for biometrics purposes. In the early works, Cheung et al. [2] and Mikic et al. [10] presented systems that can track human body using voxels with acceptable robustneess. Zhang et. al. [15] created a barrel model and use the PSO pose search algorithm for pose estimation and tracking the 3D voxel data. Menier et. al. [8] used a stick figure model to fit it into 3D human point data with E-M algorithm. Some methods also used an physical analogy. Delamarre et. al. [4] proposed a method inspired from a physical phenomena using forces applied on a partially volumetric rigid model using an iterative gradient descent. Luck et. al. [7] also used physical based method to compute the force and align a volumetric model into the subsequent voxel data.

In this paper, we develop a novel 3D model-based approach using marionette and mass-spring model to gait biometrics. Our method combines the use of marionettes'

978-1-4673-0397-2/12/$31.00 ©2012 IEEE

structure and motion principles with the physical mass-spring model for tracking the stick-figure model into the 3D voxel data. For evaluation, we used the gait dataset from the Soton multi-biometric tunnel [13, 12].

2. 3D gait dataset and marionette mass-spring gait model

2.1. 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 [9]. 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.

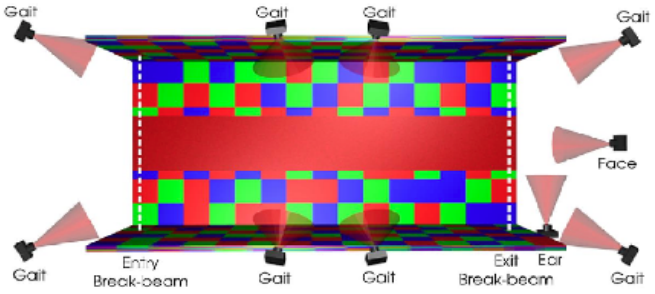


Figure 1. Soton Multi-biometrics Tunnel (from [13])

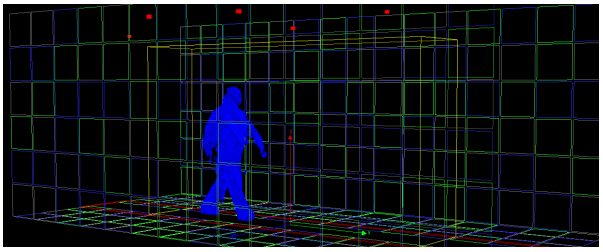


Figure 2. Sample data acquired from Tunnel after reconstruction (from [13])

2.2. Marionette spring mass model

We use a marionette mass-spring model as our gait model. Fig. 3 shows our marionette model with 11 nodes and their topology. The model topology are represented with an articulated 3D stick-figure. The human body joints and segments are modelled with nodes (masses) and springs

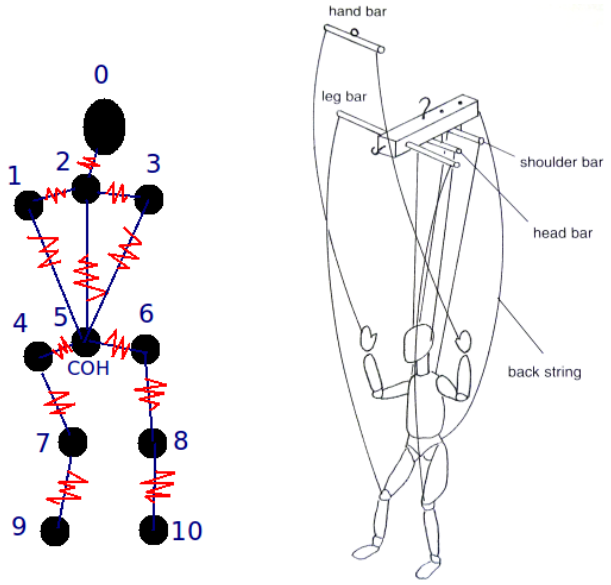


Figure 3. Marionette (from [3]) and marionette mass-spring gait model.

respectively. From the figure, it is clear that we only consider the head, torso, and lower legs but omit the arms because the quality of the voxel data for the arms is poor. We are also not interested in using the arms' kinematics for recognition purposes.

As in a marionette, some body nodes of the model can be moved by application of force. We observe that manipulating the marionette's string controller is similar to applying selected forces to the model's nodes. In our model, the forces arise from voxel points and the springs. The voxel points will behave like attractors to warp the model into the data and the springs will respond so as to maintain the model's topology.

From this model, we can extract the motion kinematics of human gait by tracking the model into the data and then measuring the rotational angle of each body segment. We are also potentially able to measure the shoulder and hip rotation using the two joints position at the arms and hips respectively.

2.3. Notations and model parameter

Let $\mathcal{M} = (P, N)$ be the marionette model with $P = \{p_0, p_1, ..., p_{10}\}$ and $N = \{\{N_0\}, \{N_1\}, ..., \{N_{10}\}\}$ being respectively the set of nodes and the set of neighbour nodes, thereby describing topology. The set of voxel data points for each frame is defined as $A = \{a_1, a_2, ..., a_n\}$.

We use bold symbols to represent vectors such as $\mathbf{x}$. The unit vector oriented from point p to q is denoted $\mathbf{u}_{pq}$ and the Euclidean distance between them is denoted $d(p, q)$.

The marionette model has a parameter k to define the spring elasticity and m for mass. In this paper we use the

same k value for all springs.

2.4. Preprocessing

Preprocessing enables initialization of the marionette model (to set the initial nodes' positions). We also use prior information for the human body anthropometric [14] as commonly used in biomechanics. The preprocessing stage can generate robust estimation of the subject height, heel strikes locations and the centre of hip.

In the preprocessing stage, firstly we seek to process 3D voxel data sequence to obtain 3D bounding boxes. This information will be used to estimate the gait period and the subject's height. We also seek to extract the heel strikes location from subject footprint image. The heel strike information is very important since walking always uses at least one leg to support the body. The heel strikes can also be used as constraints in the tracking process later.

2.4.1 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.

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

2.4.2 Gait period estimation

Using the 3D bounding box data, we can determine the starting and ending frame in one period gait cycle. The heel strike pose corresponds with the widest bounding box or the peak of the filtered width of the 3D bounding box signal. The start frame is that for which a maximum width bounding box occurs and the end frame is the second next maximum width.

2.4.3 Subject height estimation

The object height h is derived from the average of the transverse plane bounding box in one gait period (N frames) as shown in Eqn. 2.

$$h = \frac{\sum_{j=1}^{N} BB_z^j}{N} \tag{2}$$

2.4.4 Center of hip estimation

We estimate the center of hip COH = (c_x, c_y, c_z) by using the mean of the voxel data and the subject height as described in Eq. 3. Where $I(x, y, z)$ is the value/intensity of the voxel i.e. $\{0, 1\}$ and h is the subject's height.

$$\begin{aligned} c_x &= \frac{1}{s_x s_y s_z} \sum_{x=1}^{s_x} \sum_{y=1}^{s_y} \sum_{z=1}^{s_z} x I(x, y, z) \\ c_y &= \frac{1}{s_x s_y s_z} \sum_{x=1}^{s_x} \sum_{y=1}^{s_y} \sum_{z=1}^{s_z} y I(x, y, z) \\ c_z &= 0.530 \times h \end{aligned} \tag{3}$$

2.4.5 Footprints extraction and heel strikes detection

We extract the footprint image $F(x, y)$ 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 period. Therefore, we project one or more planes near the ankle height into the ground plane G^j for each frame and then accumulate a footprint image F as described in Eqn. 4. We use optimal thresholding and morphology to obtain clean footprint and locate the heel strike positions.

$$\begin{aligned} G^j(x, y) &= V^j(x, y, z = 0) \\ F(x, y) &= \sum_{j=1}^{S_j} G^j(x, y) \end{aligned} \tag{4}$$

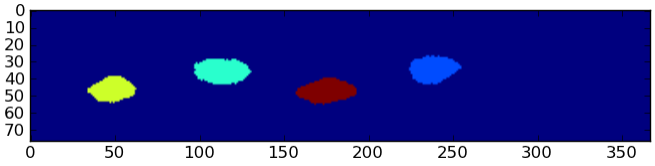


Figure 4. Extracted footprints

3. Tracking system

In order to obtain the gait kinematics from the subjects, we use a tracking approach using the marionette mass-spring gait model. The tracking module will warp the model nodes' positions each time with new voxel data in the next frame. The voxel data here works as an attractor that is able to pull the nodes in the direction of motion of the data. The pulling mechanism can be seen as an analogy of the string controller in the marionette system.

3.1. Model initialization

With prior knowledge of the human anthropometric, preprocessing results and the first frame pose, which is a heel strike pose, we can initialize the model straight away. The ankle nodes positions are placed at the footprint locations. While the knee nodes are interpolated between the hip and ankle nodes. Fig 5 shows the initialization model nodes imposed into the voxel data points.

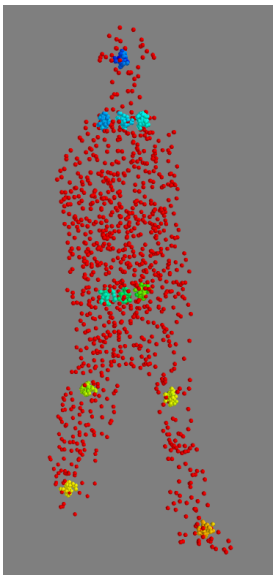


Figure 5. Initialization model nodes imposed into the voxel points

3.2. Attractor force

The voxel data points generated from the voxel binary data with a uniform sampling parameter $s < 1$. The lower value of s will generate smaller number of voxel points, hence will increase the computational speed. Based on our experiment, $0.01 < s < 0.1$ gives a good balance between resolution and speed.

The voxel data points will generate an attraction force on each node of the marionette model. Each node in the model is only affected by a set attractor nearby. To select the attractor sets for each node we employs a hierarchical clustering approach. The first phase is to cluster body-parts globally using all the voxels. The body parts consist of head, torso, and the two legs. The head is represented by a point and the others are represented by body segments (lines). Each voxel point in A will be assigned to one of the defined body part clusters AS based on the Euclidean distance from the point to the body part. Therefore, $A = \{AS_{Head}, AS_{Torso}, AS_{R.Leg}, AS_{L.Leg}\}$.

The second phase is a local nodes clustering to refine the results of previous clusters. For each body parts cluster AS, the set of voxel points will be divided into nodes cluster AP_i. Optionally, we can add a dummy node in the middle of each body segment to increase the accuracy of correct corresponds between the model node and voxel points attractors. According to Fig. 3, we have $AS_{R.Leg} = \{AP_4, AP_{dummy}, AP_7, AP_{dummy}, AP_9\}$ where AP_{dummy} is a cluster for dummy node.

Finally, the attractor force **FA** applied at a node p_i is defined by:

$$\mathbf{FA}_p = \|\mu_i - p_i\| * \mathbf{u}_{p\mu} \tag{5}$$

where μ_i is the mean of points in node cluster AP_i.

3.3. Spring force

During the tracking and warping process, the spring will maintain the model topology by inserting spring forces to the nodes when the states are changing. Given two nodes p and q linked by a spring with elasticity constant k and resting length l, the spring force applied from p to q $\mathbf{E}_{pq}$ will follow Hooke's law as:

$$\mathbf{E}_{pq} = k(d(p,q) - l_{pq}).\mathbf{u}_{pq} \tag{6}$$

and the total spring force for a model node p with a set of its neighbour N_p is defined as follow:

$$\mathbf{FS}_p = \sum_{q \in N_p} \mathbf{E}_{pq} \tag{7}$$

3.4. Updating the model

During the warping process, the model will be iteratively updated until reaching the stopping criteria. There are three states in the model that need updating. Those states are:

1. Acceleration **a**
2. Speed **v**
3. Position **x**

The total force affecting a model node p is a summation of its spring force and attractor force.

$$\mathbf{F}_p = \mathbf{FS}_p + \mathbf{FA}_p \tag{8}$$

The acceleration a_p of node p is defined as below:

$$\mathbf{a}_p = \frac{\mathbf{F}_p}{m} \tag{9}$$

where m is the mass of the node. The speed is updated with:

$$\mathbf{v}^{t+1} = f\mathbf{v}^t + dT.\mathbf{a}^{t+1} \tag{10}$$

where dT is a constant time step, and f is the friction coefficient. Finally, the position of the node is updated with:

$$\mathbf{x}^{t+1} = \mathbf{x}^t + dT.\mathbf{v}^{t+1} \tag{11}$$

We also need to update the model rest length l state for each new frame in order to adapt the model to the new fitted pose. We consider both the original and the new stable distance length for updating the rest length state. For two nodes p and q with the rest length l_{pq} and new distance $d(p,q)$, the updated rest length is as follow:

$$l_{pq} = \alpha.l_{pq} + (1-\alpha).d(p,q) \tag{12}$$

where α value is between 0 and 1. Empirically, $0.5 \leq \alpha \leq 1$ has given reasonably good results.

3.5. Stability and stopping criteria

The model updating process will be terminated when their states are considered stable. We define a minimum threshold τ for **v** and **a** to stop the iterative process.

3.6. Heel strikes and COH constraints

Based on prior knowledge, the centre of hip (COH) and heel strikes position are static during the iterative warping process. Therefore, the nodes corresponding to the COH and ankle at the heel strike will be initialised for each frame by the data from the preprocessing result. Given the accuracy of the preprocessing stage these nodes are not updated during the warping process.

4. Gait features and classification

4.1. Gait features

Gait features are created by selecting and combining the kinematics and structural information.

We only use the lower legs motion for the kinematic feature. We project the legs motion in sagittal and frontal plane then extract the motion angle of thigh and shin. The kinematics feature is represented by a set of angles $\{\theta_{RT}, \alpha_{RT}, \theta_{LT}, \alpha_{LT}, \theta_{RS}, \alpha_{RS}, theta_{LS}, \alpha_{LS}\}$ where θ, α, T, and S are sagittal angle, frontal angle, thigh, and shin respectively.

We extract the gait kinematics features using The Discrete Fourier Transform (DFT). After applying the DFT to the filtered kinematics data, we will obtain new information about the frequency components of the subject's gait. These frequency components will be used for classification. We use a Fast Fourier Transform (FFT) with 64 points. We are also able to extract the gait structural features such as height, stride length and footprint direction (which is the orientation between one footprint and the next).

4.2. Classification

We classify the subject based on two data types:

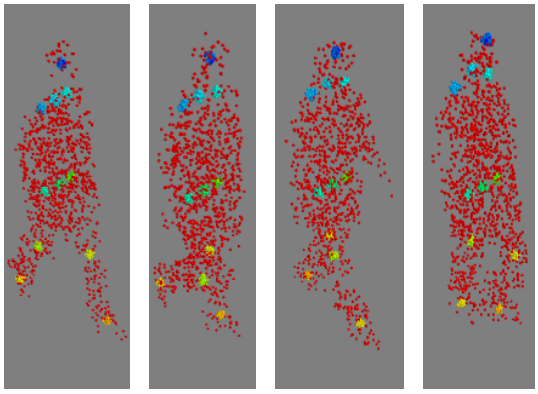
1. structural features: height, stride length and footprint direction.

2. kinematics features: the frequency components of the kinematics angles.

k-NN was used as a classifier with leave one out cross validation (LOOCV). Three different values of k were used: 1, 3, and 5. For the dynamic feature we evaluated three different distances functions of the FFT components, i.e Magnitude, Magnitude weighted phase and Euclidean.

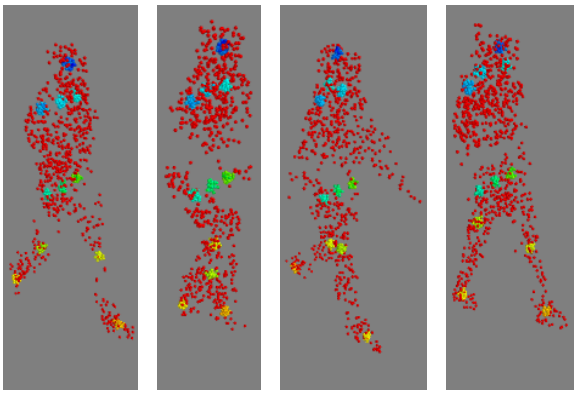
5. Results and analysis

We used Soton multi-biometric tunnel gait dataset containing 46 different subjects; four recordings of each subject were taken, giving a total of 184 samples.



(a) frame# 1 (b) frame# 7 (c) frame# 17 (d) frame# 28

Figure 6. The tracking result from good voxel data



(a) frame# 1 (b) frame# 7 (c) frame# 17 (d) frame# 28

Figure 7. The tracking result from imperfect voxel data

5.1. Tracking system results

The constant and parameter values used for our test are $k = 20, m = 1, dT = 0.1, s = 0.02, \tau_a = 0.0001$ and $\tau_v = 0.0001$. Fig 6 shows the tracking result for sample frames where the model nodes are imposed into the voxel points. In this sequence, one gait period consists of 36 frames.

The marionette model can handle occluded and imperfect (partly missing) voxel data as swhon in Fig. 7. In this sequence, it has 29 frames in one gait period.

5.2. Gait recognition analysis

Table 1 shows the classification rate using structural features showing that height is the most discriminatory feature. The best performance of 58.2% was achieved for k=1 with height and footprint. It is interesting to note that the footprint feature can increase recognition capability quite better than the stride.

Table 2 shows the detail of our classification results using the kinematic features. The best recognition performance is achieved by using the α_{RT} angle which is the movement of

Chosen Features	k=1	k=3	k=5
Height	**41.3**	37.5	37.5
Stride	7.1	4.4	3.8
Footprint	21.2	10.3	15.2
Height & Stride	42.9	28.0	31.0
Height & Footprint	**58.2**	47.8	43.5

Table 1. k-NN classification results (%) for structural features

right thigh in the frontal plane, confirming an earlier study of the potency of the front view for gait recognition [5]. The comparison of distance methods shows that the Euclidean distance is the best.

Chosen Features	Chosen Distance Methods		
	Mag	Mag-Phase	Eucl
θ_{RT}	29.3	19.6	28.3
α_{RT}	25.5	35.3	**52.2**
θ_{LT}	37.5	28.3	32.6
α_{LT}	25.5	31.5	41.8
θ_{RS}	24.5	16.8	24.5
α_{RS}	15.2	19.0	32.1
θ_{LS}	21.2	17.9	21.7
α_{LS}	19.0	30.4	40.8

Table 2. Correct classification rate (%) for kinematic features with k=1

Although our best correct classification rate does not yet equal that achieved by a model-free approach [12], 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.

6. Conclusions and Further Work

We have described a new model-based approach to 3D gait recognition. This is inherently viewpoint in variant, a factor that restricts appearance based gait biometrics approaches. Our new model is warped to fit voxel data of a 3D walking subject. The model is based on a marionette and mass spring system which represents the arrangement and interconnection of human vertices. Our initial results show good fit tracking though a sequence of frames and that there is varying perspicacity in the measures derived by this model. The model can handle some occluded and imperfect voxel data and the results encourage further development along these lines. There is also a possibility of fusing with appearance-based metrics, since the 3D data can be projected so as to be observed from any viewpoint.

References

[1] G. Ariyanto and M. S. Nixon. Model-based 3d gait biometrics. In *IJCB*, 2011.

[2] G. K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:77, 2003.

[3] D. Currell. *Making and Manipulating Marionettes*. The Crowood Press, 2004.

[4] Q. Delamarre and O. Faugeras. 3d articulated models and multiview tracking with physical forces. *Comput. Vis. Image Underst.*, 81:328–357, March 2001.

[5] 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] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image Vision Comput.*, 25(6):817–832, 2007.

[7] J. P. Luck and D. E. Small. Realtime markerless motion tracking using linked kinematic chains. In *JCIS*, pages 849–854, 2002.

[8] C. Menier, E. Boyer, and B. Raffin. 3d skeleton-based body pose recovery. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, 3DPVT '06, pages 389–396, Washington, DC, USA, 2006. IEEE Computer Society.

[9] 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.

[10] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *Int. J. Comput. Vision*, 53:199–223, July 2003.

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

[12] 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.

[13] 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.

[14] D. Winter. *The Biomechanics and Motor Control of Human Movement*. John Wiley and Sons, 2 edition, 1990.

[15] 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.