# Gait Recognition Based on Marker-less 3D Motion Capture

Tomasz Krzeszowski
Rzeszow University of Technology
Powst. Warszawy Av., 35-959 Rzeszow, Poland
tkrzeszo@prz.edu.pl

Bogdan Kwolek
AGH University of Science and Technology
30 Mickiewicza Av., 30-059 Krakow, Poland
bkw@agh.edu.pl

Agnieszka Michalczuk
Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa, Poland

Adam Switonski and Henryk Josinski
Silesian University of Technology
Akademicka 16, Gliwice, 44-101, Poland
henryk.josinski@polsl.pl

## Abstract

*We present an algorithm for view-independent gait-based person identification. The identification is achieved using data obtained by our marker-less 3D motion tracking algorithm. The motion tracking was accomplished by a particle swarm optimization algorithm. The accuracy of the motion tracking algorithm was evaluated using ground-truth data from MoCap. It was determined on 88 sequences with 22 walking performers. We obtained 90% identification accuracy (rank 1) on 230 gait cycles.*

## 1. Introduction

Gait is the product of coordinated, cyclic combination of movements that together lead to individual manner of walking. In recent years, gait-based identification became an active research area due to possible applications in visual surveillance [2]. It is the only feature that can be employed in an identification of the person at a larger distance. Human can effortlessly identify an acquaintance based on walking style, bearing or carriage as one walks. Successful identification of friends can be achieved even if a person is too far to be recognized by his/her face. Although considerable progress in gait biometrics has been made in recent years, reliable gait recognition is still a challenging problem.

In one of the earliest approaches to automatic identification on the basis of walking style, the gait signature was derived from the spatiotemporal pattern of a walking person [11]. Different walkers were distinguished through extracting their spatiotemporal gait patterns obtained from the curve-fitted stick model of a human. The identification was done under assumption that the head and the legs have distinctive signatures in domain of time and translation. On a dataset that consisted of 26 sequences of five different subjects, the classification ratio varied from nearly 60% to 80% depending on weighting factors.

Existing methods for gait recognition can be broadly divided in two main categories, namely model-free and model-based [10]. Model-free methods generally consider the complete motion pattern of human undergoing identification. These approaches strongly depend on the delineated person's silhouette and statistical methods. The majority of the methods belonging to this group can achieve successful identification from a specific viewpoint, usually fronto-parallel (side-view). To accomplish better view-independent gait-based person identification [12] an image-based reconstruction method was proposed [1]. Recently, a gait database [5] was publicly available to evaluate the gait recognition algorithms in the presence of occlusions.

The advantage of the model-based approaches is that they can reliably handle self-occlusion, noise, scale and rotation, as opposed to silhouette-based approaches. The first model-based approach to gait biometrics was developed in [3]. The gait signature was based on the angular motion of the hip and the thigh, where the angular motion of the hip and the thigh was described by Fourier series. A recognition rate of 90% on a database of 10 subjects has been reported.

Model-based gait recognition algorithms are typically built on 2D fronto-parallel body models and represent human body structure explicitly, with support of the biomechanics of human gait [16]. Structural models are widely utilized to describe the topology or the shape of body parts and limbs such as head, trunk, hip, thigh, knee and ankle by measurements like the length and width. In some work coarse human body models were used. For instance, in [8] the ellipses are fitted to different parts of the binary silhouettes and their parameters (e.g. location and orientation) are employed as gait signatures.

The coarse models that are usually employed in 3D approaches to gait recognition are far more resistant to view changes in comparison to 2D ones. Obviously, 3D gait data contains more information in comparison to 2D gait data. It is also more natural representation of human gait. However, most research in gait biometrics has been conducted using 2D approaches. As pointed out in [13] the main reason for that is the lack of a publically available 3D dataset. Another reason is that 3D model based approaches are more complex and the cost of experiments is considerable.

In [15], an approach relying on matching 3D motion models to images, and then restoring the motion parameters is proposed. The evaluation was performed on datasets with four people, i.e. 2 women and 2 men walking at 9 different speeds ranging from 3 to 7 km/h by increments of 0.5 km/h. Motion models were constructed using Vicon MoCap system. In [17] a laser range sensor was employed to capture the 3D gait data. The 3D model was fitted into the data to obtain the kinematics information. The gait cycles were synthesized by interpolation of joint positions and their movements from the fitted body models. Large recognition rates on six performers were shown using such high precision data. To overcome the non-frontal pose problem, more recently multi-camera based gait recognition methods have also been developed [4]. In the above mentioned work, joint positions of the full body are used as features.

In this work we present an approach for view-independent gait recognition. The motion parameters are estimated using our algorithm for marker-less 3D human motion capture. The system estimates the 3D motion using video sequences that are acquired by four calibrated and synchronized cameras. The motion tracking is formulated as a dynamic optimization problem. The motion of a walking person is inferred with the help of a 3D human model. The 3D human pose is reconstructed over time through matching the projection of the human body model with the current image observations. The human silhouette is extracted via background subtraction and then the edges are located within the extracted silhouette. The objective function takes into account the normalized distance between the model's projected edges and the closest edges in the image. The optimization of the objective function is achieved by Particle Swarm Optimization. Afterwards, a third order tensor is calculated. Finally, the tensorial gait data are reduced using Multilinear Principal Components Analysis (MPCA) algorithm and then classified.

The performance of the system was determined on a dataset with 22 persons. Every performer completed 4 walks from which 2 were done straight on and 2 diagonally. We show the performance of our motion tracking algorithm using ground-truth data, which were acquired by a commercial motion capture system from Vicon Nexus.

Currently, practical advantages of 3D approaches have not yet been fully explored and investigated. Given their advantages, it is then likely that such approaches will continue to play a significant role in research focusing on using gait as a biometric. The contribution of this work is an algorithm, which on 88 image sequences and 230 gait cycles achieves 90% recognition rate (rank 1).

## 2. Marker-less articulated motion tracking

In this Section we present the 3D human body model and its parameterization. Then we describe the optimization algorithm that is used to achieve 3D motion tracking. Afterwards, the ingredients of the objective function are shown.

### 2.1. 3D human body model

The human body can be represented by a 3D articulated model formed by 11 segments imitating the main parts of the body. The structural model of the human body is in the form of a kinematic chain in which the connections of body parts comprise a parent-child relationship, see Fig. 1. The

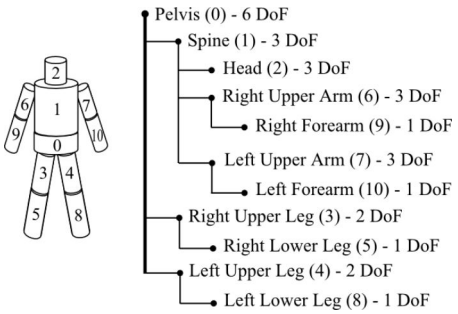


Figure 1. 3D human body model. The main 11 body parts (left), hierarchical structure (right).

pelvis being the root node in the kinematic chain is the parent of the upper legs, which are in turn the parents of the lower legs. Thus, the position of a body limb depends partially on the position of its parent body part and partially on its own pose parameters. This way, the pose parameters of each child body part are described with respect to the local coordinate frame determined by its parent. The 3D geometric model is employed to simulate the human motion and to provide the estimates of the current position, orientation and joint angles. The kinematics of the motion of each body part can be described by a locomotion model.

Although typical human bodies can be represented by such a coarse model, individuals have different body part sizes, limb lengths, and exhibit different ranges of motion. To achieve better precision of tracking the parameters are pre-specified for each individual. For every degree of freedom (DoF) there exists a physical constraint beyond which movement is not allowed. The model is constructed from truncated cones. In human motion simulation it is used to

generate contours that are projected to 2D plane and then matched with the edges on the image. The configuration of the body is parameterized by the position and the orientation of the pelvis in the global coordinate system and the angles between the connected limbs. The number of degrees for each body part of our model can be found on Fig. 1. As we can see, the complete model has 26 degrees of freedom.

### 2.2. Articulated motion tracking

3D motion tracking can be perceived as dynamic optimization problem. In such an approach, in each frame, the object state is determined using an objective function expressing the matching between the projected 3D model and the image features. Recently, particle swarm optimization (PSO) [6] has been successfully applied in full body motion tracking [18, 7]. In such an approach each particle represents a hypothesis about the current object state. The particles follow simple position and velocity update rules and they are evaluated on the basis of a fitness function. The fitness function is utilized in the evaluation of the degree of similarity between the real and the hypothesized pose. Motion tracking can be achieved by a sequence of static PSO-based optimizations, followed by re-diversification of the particles to cover the possible poses in the next time step. Typically, the particles are propagated according to weak transition model with parameters allowing them to cover possible state changes between consecutive images. Estimating the 3D pose in each frame is non-linear, high-dimensional optimization problem.

In this work the tracking of human motion is achieved using the Annealed Particle Swarm Optimization (APSO) [7]. In the APSO algorithm the objective function is annealed and then it undergoes a quantization. Thanks to such an operation the similar function values are clustered. That means that in every iteration the algorithm determines the set of the best particles. The swarm selects the global best location from a set of candidate best locations.

Since the motion tracking can be cast as dynamic optimization problem, the tracking can be obtained by a sequence of static PSO-based optimizations, followed by re-diversification of the particles to cover the potential poses that can arise in the next time step. The re-diversification of the particle $i$ can be achieved on the basis of normal distribution, which is concentrated around the state estimate $\hat{x}_{t-1}$ in time $t-1$: $x_t^i \leftarrow \mathcal{N}(\hat{x}_{t-1}, \Sigma)$, where $\Sigma$ denotes the covariance matrix of the Gaussian distribution, whose diagonal elements were determined experimentally.

### 2.3. Fitness function

The fitness function expresses the degree of similarity between the real and the estimated human pose. The 3D model in a candidate pose is projected into 2D image plane. Given the projected model, for each camera two fitness values are calculated. The first value depends on the overlap with the current silhouette, whereas the second one reflects a match between the model edges and the closest edges in the image, see Fig. 2.

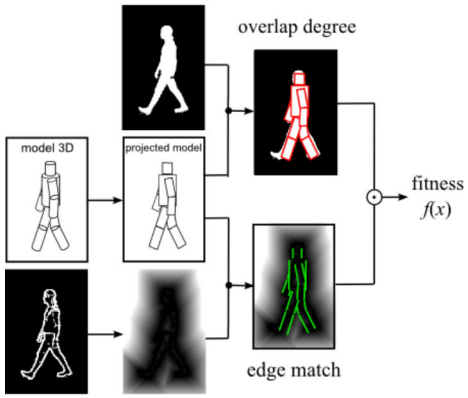


Figure 2. Determining the fitness function.

Figure 3 depicts calculation of the edge distance map, which is used in matching the edges. The person is segmented using a background subtraction algorithm. The binary mask representing the extracted person is morphologically dilated and then used to suppress edges not belonging to person. The edges are extracted using the Sobel operator. The masked gradient image, see image after the block *and* on Fig. 3, is used in extraction of the edge distance map.

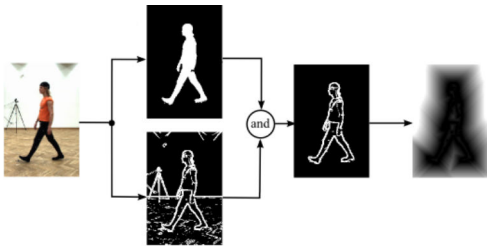


Figure 3. Determining the edge distance map.

The objective function for camera $j$ is calculated on the basis of following expression: $f^{(j)}(x) = 1 - (f_1(x)^{w_1} \cdot f_2(x)^{w_2})$, where $x$ stands for the state (pose), whereas $w$ denotes weighting coefficients that were determined experimentally. The function $f_1(x)$ reflects the degree of overlap between the extracted body and the projected 3D model into 2D image. The function $f_2(x)$ reflects the edge distance map-based fitness. The objective function for all four cameras is determined according to the following expression: $f(x) = \frac{1}{4}\sum_{j=1}^{4} f^{(j)}(x)$.

The calculation of the objective function is the most consuming operation. Moreover, in multiview tracking the 3D model is projected and then rendered in each camera's view.

Thus, in our approach the objective function is calculated by OpenMP threads, which communicate via the shared memory. Each core calculates the fitness score for single camera and every PSO thread has access to the shared memory with the objective function values.

## 3. Gait characterization and recognition

The marker-less motion tracking was achieved using color images of size $960 \times 540$, which were acquired at 25 fps by four synchronized and calibrated cameras. Each pair of the cameras is approximately perpendicular to the other camera pair. A commercial motion capture system from Vicon Nexus was employed to provide the ground truth data. The system uses reflective markers and sixteen cameras to recover the 3D location of such markers. The data are delivered with rate of 100 Hz and the synchronization between the MoCap and multi-camera system is achieved using hardware from Vicon Giganet Lab.

A set of $M = 39$ markers was attached to main body parts. From the above set of markers, 4 markers were placed on the head, 7 markers on each arm, 12 on the legs, 5 on the torso and 4 markers were attached to the pelvis. Given such a placement of the markers on the human body and the estimated human pose, the corresponding positions of virtual markers on the body model were determined. Figure 4 illustrates the distances between ankles, which were determined by our marker-less motion tracking algorithm and the MoCap system. High overlap between both curves formulates a rationale for the usage of the marker-less motion tracking to achieve view-independent gait recognition. In particular, as we can observe, the gait cycle and the stride length can be calculated with sufficient precision.

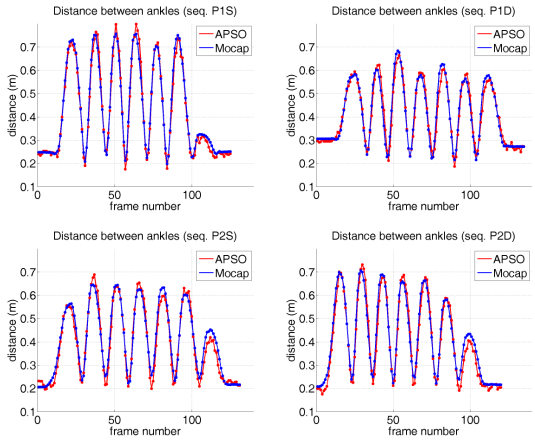


Figure 4. Distance between ankles during walking in sequences $P1$ (straight and diagonal) and $P2$ (straight and diagonal).

Figure 5 illustrates components of a typical model-based system for gait recognition. Given a gallery database, consisting of gait patterns from a set of known subjects, the objective of the gait recognition system is to determine the identity of the probe samples. In this work we treat each gait cycle as a data sample, which is represented by a third order tensor. A tensor is a multidimensional object, whose elements are addressed by indices. The number of indices determines the order of the tensor, whereas each index defines one of the tensor modes [9].

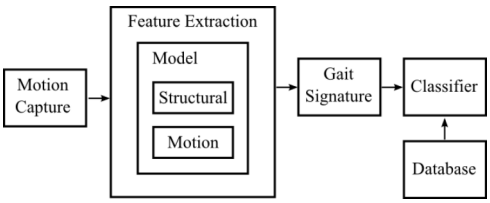


Figure 5. A typical model-based gait recognition system.

The data extracted by the motion tracking algorithm were stored in ASF/AMC data format. For a single gait cycle consisting of two strides a third order tensor $30 \times 11 \times 3$ was calculated. Before feeding the gait samples to dimensionality reduction, the tensorial inputs need to be normalized to the same dimension in each mode. The number of frames in each gait sample has some variation and therefore the number of frames in each gait sample was normalized. The normalized time mode dimension was chosen to be 30, which was roughly the average number of frames in each gait sample. The second dimension of the tensor is equal to the number of bones (excluding pelvis), i.e. 10 plus one element in which the distance between ankles and person height are stored. The third dimension accounts for three angles, except the eleventh vector that contains distance between ankles and person's height. Such a gait signature was then reduced using Multilinear Principal Components Analysis (MPCA) algorithm [9].

MPCA is a multilinear extension of Principal Component Analysis (PCA) algorithm, which is commonly utilized for dimension reduction in analysis of high-dimensional data. In attempting to preserve the data structure, MPCA seeks for low-dimensional projections and, thereby, decreases dimensionality more efficiently than PCA. Thus, the MPCA transformation aims to capture as high a variance as possible, accounting for as much of the variability in the data as possible, subject to the constraint of mode-wise orthogonality. There is one orthogonal transformation for each dimension (mode). MPCA determines a tensor-to-tensor projection that captures most of the signal variation present in the original tensorial representation.

## 4. Experimental results

The marker-less motion tracking system was evaluated on video sequences with 22 walking individuals. In each

image sequence the same actor performed two walks, consisting in following a virtual line joining two opposite cameras and following a virtual line joining two nonconsecutive laboratory corners. The first subsequence is referred to as 'straight', whereas the second one is called 'diagonal'. Given the determined pose estimate, the model was overlaid on the images. Figure 6 depicts some results which were obtained for person 1 in a diagonal walk. The degree of overlap of the projected 3D body model with the performer's silhouette reflects the tracking accuracy.

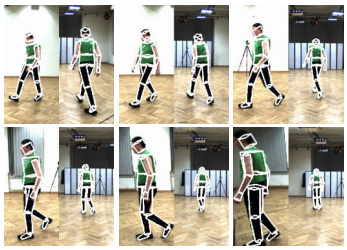

Figure 6. Articulated 3D human body tracking in sequence $P1D$. Shown are results in frames #20, 40, 60, 80, 100, 120 (from left to right and from top to bottom). The left sub-images are seen from view 1, whereas the right ones are seen from view 2.

In Fig. 7 are shown some motion tracking results that were obtained in the same image sequence, but with the performer following a virtual line connecting two opposite cameras. The body model is overlaid on the images from the right profile view and the frontal view. The discussed results were obtained in 20 iterations per frame using the APSO algorithm consisting of 300 particles.

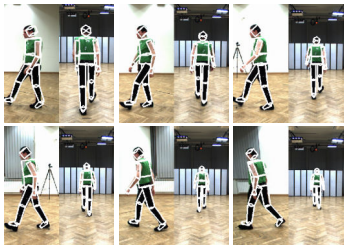

Figure 7. Articulated 3D human body tracking in sequence $P1S$. Shown are results in frames #20, 40, 60, 80, 100, 120.

The plots shown in Fig. 8 illustrate the accuracy of motion estimation for some joints. As we can observe, the average tracking error of both legs is about 50 mm and the maximal error does not exceed 100 mm. As we can see on Fig. 8 the error for the right hand is somewhat larger in comparison to errors of the remaining body parts. Such errors arise because of the similar color distribution of the wooden parquet with the skin color distribution. The errors can be reduced through more sophisticated image processing.

In Tab. 1 are presented some quantitative results that were obtained in the sequence $P1$ ($P1S$ and $P1D$). The

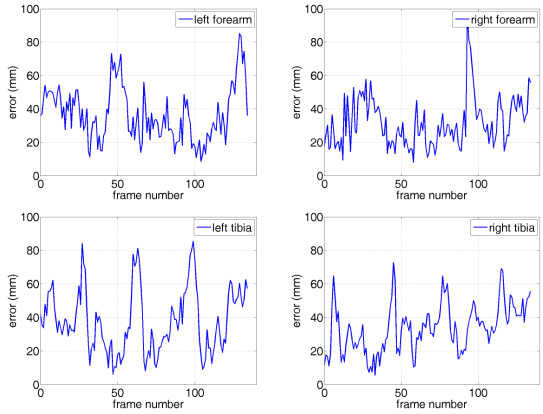


Figure 8. Tracking errors [mm] vs. frame number in seq. $P1D$.

errors were calculated on the basis of 39 markers. For each frame they were computed as average Euclidean distance between individual markers and the recovered 3D joint locations [14].

Table 1. Average errors in sequence $P1$ ($P1S$ and $P1D$). The images from seq. $P1S$ are shown on Fig. 7, whereas the images from seq. $P1D$ are depicted on Fig. 6.

| | | | Seq. $P1S$ | Seq. $P1D$ |
|---|---|---|---|---|
| | #particles | it. | error [mm] | error [mm] |
| PSO | 100 | 10 | 64.4±35.6 | 56.3±30.2 |
| | 100 | 20 | 55.0±29.4 | 56.1±32.7 |
| | 300 | 10 | 57.2±30.8 | 55.7±34.7 |
| | 300 | 20 | 54.6±30.3 | 50.0±27.6 |
| APSO | 100 | 10 | 53.2±28.9 | 49.1±24.4 |
| | 100 | 20 | 49.7±28.1 | 45.0±23.6 |
| | 300 | 10 | 49.3±26.3 | 44.9±22.0 |
| | 300 | 20 | 45.6±24.6 | 41.6±21.1 |

Table 2 shows the average errors, which were obtained on images from another sequence, called $P2$ ($P2S$ and $P2D$). For each evaluated sequence $P$xx the depicted errors are the result of the averaging over ten runs of the PSO/APSO with unlike initializations.

Table 2. Average errors in a sequence $P2$ ($P2S$ and $P2D$).

| | | | Seq. $P2S$ | Seq. $P2D$ |
|---|---|---|---|---|
| | #particles | it. | error [mm] | error [mm] |
| APSO | 100 | 10 | 56.1±30.7 | 47.4±25.0 |
| | 100 | 20 | 53.5±30.2 | 43.6±22.1 |
| | 300 | 10 | 54.0±29.2 | 43.2±22.0 |
| | 300 | 20 | 52.9±29.5 | 40.8±19.9 |

Table 3 illustrates the recognition accuracy that was obtained on 88 image sequences, each containing both straight

and diagonal walks, which were performed by a sole actor. Each sequence consisted of 2 or 3 full gait cycles. Given such a collection of sequences, we obtained a database with 230 gait cycles. 10-fold cross-validation was used to evaluate the performance of the proposed algorithm. The data obtained by motion capture were smoothed by a filter with temporal window of size 9.

Table 3. Identification accuracy [%].

| | | Naïve Bayes | | | MLP | | |
|---|---|---|---|---|---|---|---|
| $Q$ | $A$ | Rank 1 | Rank 2 | Rank 3 | Rank 1 | Rank 2 | Rank 3 |
| 0.60 | 1 | 48.7 | 68.7 | 81.7 | 38.7 | 59.6 | 77.8 |
| 0.70 | 2 | 60.4 | 80.4 | 89.1 | 53.5 | 75.2 | 86.1 |
| 0.80 | 4 | 60.0 | 80.9 | 87.8 | 55.2 | 80.0 | 86.5 |
| 0.90 | 12 | 76.5 | 86.5 | 93.0 | 69.1 | 80.0 | 87.4 |
| 0.95 | 32 | 77.4 | 89.6 | 94.3 | 73.5 | 87.0 | 92.6 |
| 0.97 | 40 | 78.7 | 89.6 | 93.9 | 73.5 | 86.5 | 90.0 |
| 0.98 | 60 | 78.7 | 86.1 | 92.2 | 75.6 | 85.6 | 91.7 |
| 0.99 | 120 | 84.3 | 90.9 | 93.9 | 89.6 | 94.3 | 97.0 |

In the person identification we employed Naïve Bayes (NB) and multilayer perceptron (MLP) classifiers. The parameter $Q$ denotes the ratio of variations in the MPCA, which should be kept in each mode, whereas $A$ denotes the number of the attributes corresponding to given $Q$. In practical recognition tasks, $Q$ is commonly set to a large value to capture most of the variation. As one can observe, for $A = 120$ attributes the MLP classifier achieves 90% recognition accuracy (rank 1), whereas the NB gives almost 85% recognition accuracy. The results show that using 3D gait data from our marker-less motion capture can lead to high identification accuracy.

The complete human motion capture system was written in C/C++. The evaluation of the identification accuracy was performed using WEKA software.

## 5. Conclusions

We have presented a view-independent algorithm for gait recognition. The recognition is done using 3D gait data obtained by our marker-less human motion tracking algorithm. We demonstrated the tracking performance of the algorithm using the MoCap data as ground-truth. High-dimensional tensor data were reduced by the MPCA algorithm and subsequently classified. Experiments on multiview video sequences demonstrated that the algorithm achieves high recognition accuracy. The identification accuracy (rank 1) on 230 gait cycles performed by 22 persons is equal to 90%.

## Acknowledgement

This work was supported by the National Science Centre (NCN) within the research project N N516 483240, the National Centre for Research and Development (NCBiR) within the project OR00002111 and U-2051 DS/M.

## References

[1] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image Vision Comput.*, 27(8):1194–1206, 2009.

[2] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis. Gait recognition: a challenging signal processing technology for biometric identification. *Signal Proc. Mag.*, 22:78–90, 2005.

[3] D. Cunado, M. S. Nixon, and J. N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Int. Conf. AVBPA*, pages 95–102, 1997.

[4] J. Gu, X. Ding, S. Wang, and Y. Wu. Action and gait recognition from recovered 3-D human joints. *IEEE Trans. Sys. Man Cyber. Part B*, 40(4):1021–1033, 2010.

[5] M. Hofmann, S. Sural, and G. Rigoll. Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *Int. Conf. WSCG*, pages 99–104, 2011.

[6] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Int. Conf. on Neural Networks*, pages 1942–1948, 1995.

[7] B. Kwolek, T. Krzeszowski, and K. Wojciechowski. Swarm intelligence based searching schemes for articulated 3D body motion tracking. In *Int. Conf. ACIVS*, pages 115–126, 2011.

[8] L. Lee, G. Dalley, and K. Tieu. Learning pedestrian models for silhouette refinement. In *ICCV*, pages II:663–670, 2003.

[9] H. Lu, K. Plataniotis, and A. Venetsanopoulos. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Trans. on Neural Networks*, 19(1):18–39, 2008.

[10] M. S. Nixon and J. Carter. Automatic recognition by gait. *Proc. of the IEEE*, 94(11):2013–2024, 2006.

[11] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in XYT. In *CVPR*, pages 469–474, 1994.

[12] R. Seely, M. Goffredo, J. Carter, and M. Nixon. View invariant gait recognition. In *Handbook of Remote Biometrics: for Surveillance and Security*, pages 61–82. Springer, 2009.

[13] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon. University of Southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *Proc. of the 2nd IEEE Int. Conf. on Biometrics*, pages 1 –6, 2008.

[14] L. Sigal, A. Balan, and M. Black. HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *Int. Journal of Computer Vision*, 87:4–27, 2010.

[15] R. Urtasun and P. Fua. 3D tracking for gait characterization and recognition. In *Int. Conf. FG*, pages 17–22, 2004.

[16] C. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Rec.*, 37(5):1057 – 1072, 2004.

[17] K. Yamauchi, B. Bhanu, and H. Saito. Recognition of walking humans in 3d: Initial results. In *CVPR Workshops*, pages 45 –52, june 2009.

[18] X. Zhang, W. Hu, X. Wang, Y. Kong, N. Xie, H. Wang, H. Ling, and S. Maybank. A swarm intelligence based searching strategy for articulated 3D human body tracking. In *IEEE Workshop in conj. with CVPR*, pages 45–50, 2010.