View Independent Human Gait Recognition using Markerless 3D Human Motion Capture

Tomasz Krzeszowski[2,1], Bogdan Kwolek[2,1], Agnieszka Michalczuk[1], Adam Switonski[1,3], and Henryk Josinski[1,3]

[1] Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa, Poland
bytom@pjwstk.edu.pl
[2] Rzeszów University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
{tkrzeszo,bkwolek}@prz.edu.pl
[3] Silesian University of Technology
Akademicka 16, Gliwice, 44-101 Poland
{Adam.Switonski,Henryk.Josinski}@polsl.pl

Abstract. We present an algorithm for view-independent human gait recognition. The human gait recognition is achieved using data obtained by our markerless 3D motion tracking algorithm. The tensorial gait data were reduced by multilinear principal component analysis and subsequently classified. The performance of the motion tracking algorithm was evaluated using ground-truth data from MoCap. The classification accuracy was determined using video sequences with walking performers. Experiments on multiview video sequences show the promising effectiveness of the proposed algorithm.

1 Introduction

People are able to identify acquaintances on the basis of their walking style, bearing or carriage as one walks. Successful identification can be done even when acquaintance is too far to be recognized by his/her face. Gait arises from coordinated, cyclic combination of movements that lead to individual style or manner of walking. It is the only feature suitable for human identification when the subject is far from the camera. Recently, vision-based gait recognition has attracted increased attention due to possible applications in intelligent biometrics and visual surveillance systems [2].

The seminal work in the area of gait recognition was done in the psychology field [8]. It is commonly believed upon that human visual system is very sensitive to motion stimuli, although the exact mechanism of perception is not fully clear. Johansson's seminal psychophysical experiments [6] demonstrated that humans can recognize biological motion, such as gait on the basis of a small set of points attached to the human body, which are called Moving Light Displays (MLDs). These findings inspired computer vision community to start research on extracting characteristic gait signatures from image sequences for human identification.

In one of the earliest approaches to automatic identification by gait, the gait signature was derived from the spatio-temporal pattern of a walking person [13]. It was assumed that the head and the legs have distinctive signatures in XT (translation and time) dimensions. They first find the bounding contours of the walker, and then fit a five elements stick model on them. On a database of 26 sequences of five different subjects, taken at different times during the day the classification rate varied from nearly 60% to 80% depending on weighting factors.

There are two major approaches to gait recognition, namely appearance based (model free) and model based ones [12]. The majority of the approaches proposed for gait-based identification rely on analysis of image sequences acquired by a single camera. The main drawback of such approaches is that they can perform recognition from a specific viewpoint, usually fronto-parallel. To achieve view-independent motion classification [14], recently, an image-based reconstruction method was proposed [1]. However, in tracking real human movements the 2D models are often unsuitable due to their inability to take into account the self-intersection constraints as well as to encode joint angle limits. Recently, a gait database [5] has been developed to stimulate research on gait recognition in the presence of occlusion.

The major advantage of the model-based approaches is that they can reliably handle self-occlusion, noise, scale and rotation, as opposed to silhouette-based approaches. In general, the methods that are based on holistic space-time features or space-time shapes depend more on the training data in comparison to the model-based approaches. One of the earliest model-based approaches to gait-based human identification was developed by Cunado in 1997 [3]. The gait signature has been derived from the spectra of measurements of the thigh's orientation. A recognition rate of 90% on a database of 10 subjects was achieved.

Model-based gait recognition algorithms are usually based on 2D fronto-parallel body models and model human body structure explicitly, with support of the biomechanics of human gait [18]. Some work has also been done on coarser human body models. For example, in [10] several ellipses are fitted to different parts of the binary silhouettes and the parameters of these ellipses (e.g., location. and orientation) are used as gait features. Coarse models are usually employed in 3D approaches to gait recognition. The 3D approaches to gait recognition are far more resistant to view changes in comparison to 2D ones. In [17], an approach relying on matching 3D motion models to images, and then tracking and 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 motion capture system. To overcome the non-frontal pose problem, more recently multi-camera based gait recognition methods have also been developed [4]. In the mentioned work, joint positions of the whole body are employed as a feature for gait recognition.

In this work we present an algorithm for view-independent gait recognition, which is based on 3D articulated body model. The motion parameters are estimated using our algorithm for markerless human motion tracking. The algorithm

captures human motion on the basis of video sequences acquired by four calibrated and synchronized cameras. We show the tracking performance of the motion tracking algorithm using ground-truth data from a commercial motion capture (MoCap) system from Vicon Nexus. The captured motion was stored in ASF/AMC data format. Afterwards, a second order tensor was calculated. The tensorial gait data were reduced using Multilinear Principal Components Analysis (MPCA) algorithm and then classified.

2 Markerless System for Articulated Motion Tracking

2.1 3D Human Body Model

The human body can be represented by a 3D articulated model formed by 11 rigid segments representing key parts of the human body. The model of a human body specifies a kinematic chain, where the connections of body parts comprise a parent-child relationship, see Fig. 1. In our approach the pelvis is the root node

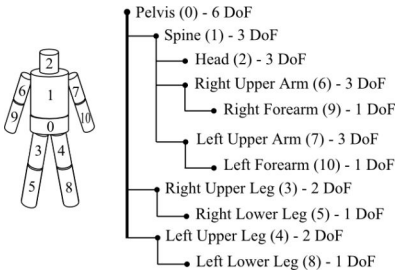

Fig. 1. 3D human body model. Human body consisting of 11 segments (left), hierarchical structure (right).

in the kinematic chain. Then the pelvis is the parent of the upper legs, which are in turn the parents of the lower legs. In consequence, the position of a particular body limb is partially determined by the position of its parent body part and partially by its own pose parameters. In this way, the pose parameters of a body part are described with respect to the local coordinate frame determined by its parent. The 3D geometric model is utilized to simulate the human motion and to recover the current position, orientation and joint angles. Although typical human bodies can be represented by such model, individuals have different body part sizes, limb lengths, and exhibit different ranges of motion. In our approach the individual parameters are pre-specified. For each degree of freedom the model has constraints beyond which movement is not allowed. The model is constructed from truncated cones and is used to generate contours, which are then compared with edge contours. 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.

2.2 Articulated Motion Tracking

Estimating 3D motion can be cast as a non-linear, high-dimensional optimization problem. The degree of similarity between the real and the estimated pose is evaluated using an objective function. Recently, particle swarm optimization (PSO) [7] has been successfully applied to body motion tracking [19][9]. In PSO each particle follows simple position and velocity update equations. Thanks to interaction between particles a collective behavior arises. It leads to the emergence of global and collective search capabilities, which allow the particles to gravitate towards the global extremum. Human 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. In this work the tracking of human motion is achieved using the Annealed Particle Swarm Optimization (APSO) [9].

2.3 Fitness Function

The fitness function expresses the degree of similarity between the real and the estimated human pose. It is calculated on the basis of following expression: $f(x) = 1 - (f_1(x)^{w_1} \cdot f_2(x)^{w_2})$, 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 model's into 2D image. The function $f_2(x)$ reflects the edge distance-based fitness. Figure 2 illustrates the extraction of the body and calculation of the edge distance. A background subtraction algorithm [9] is employed to extract the binary image of the person, see Figure 2b. The binary image is then utilized as a mask to suppress edges not belonging to the person, see Figure 2d.

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

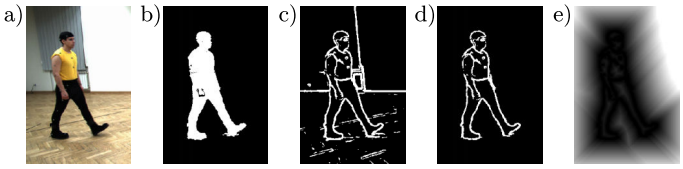

Fig. 2. Person segmentation. Input image a), foreground b), gradient magnitude c), masked gradient image d), edge distance map e).

3 Gait Characterization and Recognition

The markerless motion tracking was achieved using color images of size 960×540, 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. Figure 3 depicts the location of the cameras in the laboratory.

Fig. 3. Layout of the laboratory with four cameras. The images illustrate the initial model configuration, overlaid on the image in first frame and seen in view 1, 2, 3, 4.

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, which has been calculated by our algorithm, the corresponding positions of virtual markers on the body model were determined. Figure 4 illustrates the distances between ankles, which were determined by our markerless motion tracking algorithm and the MoCap system. High overlap between both curves formulates a rationale for the usage of the markerless 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.

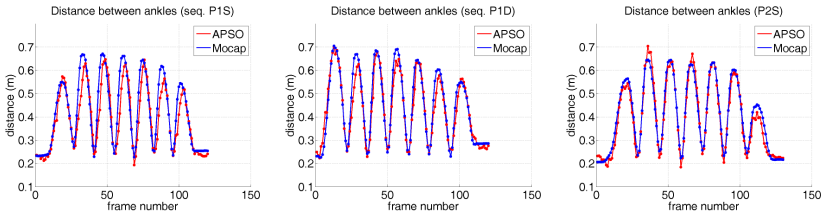

Fig. 4. Distance between ankles during walking in sequences P1 (straight and diagonal) and P2 (straight).

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. Thus, a gait sample is a second-order tensor, and the state space with the time space account for its two modes.

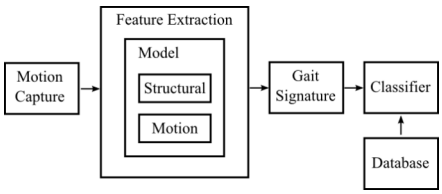

Fig. 5. Components of a typical model-based gait recognition system.

The data extracted by markerless motion tracking algorithm were stored in ASF/AMC data format. For a single gait cycle consisting of two strides a second order tensor was calculated. Before feeding the gait samples to MPCA, 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 subjected to normalization. The normalized time mode dimension was chosen to be 30, which was roughly the average number of frames in each gait sample. As mentioned in Subsection 2.1, 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. A set of the state variables plus the distance between ankles and the person's height account for the second mode of the tensor. Among the state variables there is roll angle of the pelvis and the angles between the connected limbs. In consequence, the dimension of the tensor is 30×33, where 30 is the number of frames and the second dimension is equal to the number of bones (excluding pelvis) times three angles plus three (i.e. pelvis roll angle, distance between ankles and the person's height). Such a gait signature was then reduced using Multilinear Principal Components Analysis (MPCA) [11] algorithm. MPCA is a multilinear extension of Principal Component Analysis (PCA) algorithm. There is one orthogonal transformation for each dimension (mode). 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. It determines a tensor-to-tensor projection that captures most of the signal variation present in the original tensorial representation.

4 Experimental Results

The markerless motion tracking system was evaluated on video sequences with 10 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 accuracy of the tracking.

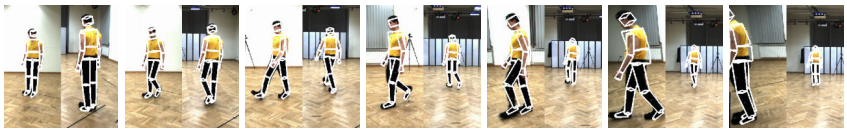

Fig. 6. Articulated 3D human body tracking in sequence P1D. Shown are results in frames #0, 20, 40, 60, 80, 100, 120. The left sub-images are seen from view 1, whereas the right ones are seen from view 2.

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 APSO algorithm consisting of 300 particles.

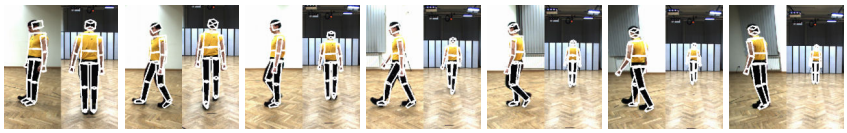

Fig. 7. Articulated 3D human body tracking in sequence P1S. Shown are results in frames #0, 20, 40, 60, 80, 100, 120. Left sub-images are seen from view 1, whereas the right ones are seen from view 2.

The plots 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 110 mm. The discussed results were obtained by APSO algorithm in 20 iterations using 300 particles.

In Tab. 1 are presented some quantitative results that were obtained in the discussed image sequences. The errors were calculated on the basis of 39 markers. For each frame they were computed as average Euclidean distance between

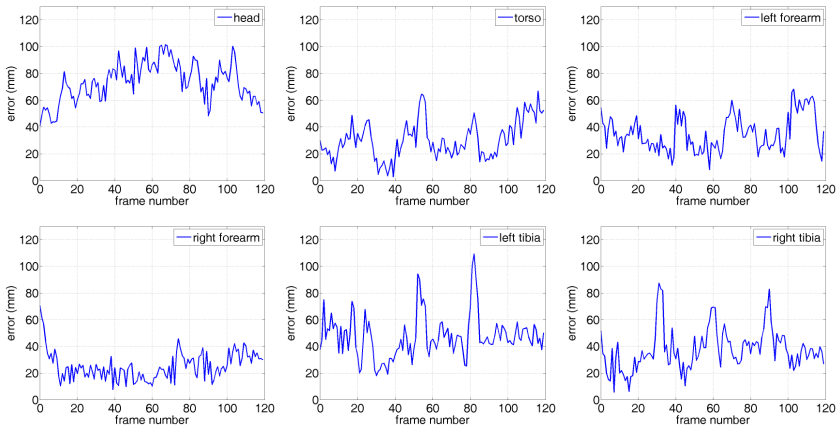

Fig. 8. Tracking errors [mm] versus frame number.

individual markers and the recovered 3D joint locations [15]. For each sequence they were then averaged over ten runs with unlike initializations.

Table 1. Average errors for $M = 39$ markers in four image sequences. 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	Seq. P2S	Seq. P2D
	#particles	it.	error [mm]	error [mm]	error [mm]	error [mm]
APSO	100	10	56.4±29.9	61.0±30.4	51.0±28.3	67.7±41.9
	100	20	51.4±28.3	53.6±24.9	47.1±27.1	62.1±40.7
	300	10	51.9±28.5	53.6±24.5	47.9±28.0	62.4±38.3
	300	20	50.0±27.3	48.6±20.5	45.5±25.7	56.5±38.2

Table 2 illustrates the recognition accuracy that was obtained on 10 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 52 gait cycles. 10-fold cross-validation was used to evaluate the performance of the proposed algorithm for gait recognition . In the person identification we employed Naïve Bayes (NB) and multilayer perceptron (MLP) classifiers. The parameter Q denotes the ratio of variations, 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 $A = 3$ attributes and rank 2 the NB classifier achieves 100% recognition accuracy, whereas the MLP gives 92% recognition accuracy. For Q equal to 0.95 and 17 attributes and rank 2 both classifiers allows us to obtain 90% recognition accuracy. As one can observe, for $Q = 0.99$, 111 attributes and rank 1 the MLP classifier gives somewhat better classification accuracy than 80%.

In [16] a far better gait recognition accuracy using data from maker-based MoCap system has been obtained. The above work demonstrates that given a sufficiently precise location of the joints it is possible to obtain high recognition accuracy of the gait. Therefore, in order to obtain better recognition accuracy of marker-less based gait recognition the precision of the motion tracking should be improved.

Table 2. Identification accuracy [%].

		Naïve Bayes			MLP		
Q	A	Rank 1	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
0.60	3	88	100	100	83	92	96
0.90	7	85	94	94	75	88	90
0.95	17	81	90	92	81	90	96
0.99	111	38	63	71	81	87	90

The complete human motion capture system was written in C/C++. The evaluation of the recognition performance was performed using WEKA software. The marker-less motion capture system runs on an ordinary PC. Our implementation of MPCA is based on Jama-1.0 library, supporting matrix operations and eigen decomposition.

5 Conclusions

We have presented a view-independent approach to gait recognition. The recognition is done using gait data obtained by our markerless 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.

Acknowledgments. This work has been supported in part 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 Ministry of Science and Higher Education within the grant U-8604/DS/M.

References

1. Bodor, R., Drenner, A., Fehr, D., Masoud, O., Papanikolopoulos, N.: View-independent human motion classification using image-based reconstruction. Image Vision Comput. 27(8), 1194–1206 (2009)
2. Boulgouris, N.V., Hatzinakos, D., Plataniotis, K.N.: Gait recognition: a challenging signal processing technology for biometric identification. Signal Processing Magazine, IEEE 22, 78–90 (2005)
3. Cunado, D., Nixon, M.S., Carter, J.N.: Using gait as a biometric, via phase-weighted magnitude spectra. In: Proc. of the First Int. Conf. on Audio- and Video-Based Biometric Person Authentication. pp. 95–102 (1997)
4. Gu, J., Ding, X., Wang, S., Wu, Y.: Action and gait recognition from recovered 3-D human joints. IEEE Trans. Sys. Man Cyber. Part B 40(4), 1021–1033 (2010)
5. Hofmann, M., Sural, S., Rigoll, G.: Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In: 19th Int. Conf. on Computer Graphics, Visualization and Computer Vision (WSCG) (2011)
6. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perceptron & Psychophysics 14, 201–211 (1973)
7. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of IEEE Int. Conf. on Neural Networks. pp. 1942–1948. IEEE Press, Piscataway, NJ (1995)
8. Kozlowski, L.T., Cutting, J.E.: Recognizing the sex of a walker from a dynamic point-light display. Perception Psychophysics 21(6), 575–580 (1977)
9. Kwolek, B., Krzeszowski, T., Wojciechowski, K.: Swarm intelligence based searching schemes for articulated 3D body motion tracking. In: Int. Conf. on Advanced Concepts for Intell. Vision Systems, LNCS. pp. 115–126. Springer, vol. 6915 (2011)
10. Lee, L., Dalley, G., Tieu, K.: Learning pedestrian models for silhouette refinement. In: Proc. of the Ninth IEEE Int. Conf. on Computer Vision. pp. II:663–670 (2003)
11. Lu, H., Plataniotis, K., Venetsanopoulos, A.: MPCA: Multilinear principal component analysis of tensor objects. IEEE Trans. on Neural Networks 19(1), 18–39 (2008)
12. Nixon, M.S., Carter, J.: Automatic recognition by gait. Proc. of the IEEE 94(11), 2013–2024 (2006)
13. Niyogi, S.A., Adelson, E.H.: Analyzing and recognizing walking figures in xyt. In: Proc. of the Int. Conf. on Computer Vision and Pattern Rec. pp. 469–474 (1994)
14. Seely, R., Goffredo, M., Carter, J., Nixon, M.: View invariant gait recognition. In: Handbook of Remote Biometrics: for Surveillance and Security, pp. 61–82. Springer (2009)
15. Sigal, L., Balan, A., Black, M.: 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)
16. Świtoński, A., Polański, A., Wojciechowski, K.: Human identification based on the reduced kinematic data of the gait. In: 7th Int. Symp. on Image and Signal Processing and Analysis. pp. 650–655 (2011)
17. Urtasun, R., Fua, P.: 3D tracking for gait characterization and recognition. In: Proc. of IEEE Int. Conf. on Automatic Face and Gesture Rec. pp. 17–22 (2004)
18. Yam, C., Nixon, M.S., Carter, J.N.: Automated person recognition by walking and running via model-based approaches. Pattern Rec. 37(5), 1057 – 1072 (2004)
19. Zhang, X., Hu, W., Wang, X., Kong, Y., Xie, N., Wang, H., Ling, H., Maybank, S.: A swarm intelligence based searching strategy for articulated 3D human body tracking. In: IEEE Workshop on 3D Information Extraction for Video Analysis and Mining in conjuction with CVPR. pp. 45–50. IEEE (2010)