3D Gait Recognition Using Spatio-Temporal Motion Descriptors

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

[1] AGH University of Science and Technology
30 Mickiewicza Av., 30-059 Krakow, Poland
bkw@agh.edu.pl
[2] Polish-Japanese Institute of Information Technology
Al. Legionów 2, 41-902 Bytom, Poland
{amichalczuk,hjosinski}@pjwstk.edu.pl
[3] Rzeszów University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
tkrzeszo@prz.edu.pl

Abstract. We present a view independent algorithm for 3D human gait recognition. The identification of the person is achieved using motion data obtained by our markerless 3D motion tracking algorithm. We report its tracking accuracy using ground-truth data obtained by a marker-based motion capture system. The classification is done using SVM built on the proposed spatio-temporal motion descriptors. The identification performance was determined using 230 gait cycles performed by 22 persons. The correctly classified ratio achieved by SVM is equal to 93.5% for rank 1 and 99.6% for rank 3. We show that the recognition performance obtained with the spatio-temporal gait signatures is better in comparison to accuracy obtained with tensorial gait data and reduced by MPCA.

1 Introduction

Gait is an attractive biometric feature for human identification at a distance, and recently has gained much interest from academia. Vision-based gait recognition has attracted increased attention due to possible applications in intelligent biometrics and visual surveillance systems [2]. Compared with conventional biometric features, such as face or iris, gait has many unique advantages since the identification techniques are non-contact, non-invasive, perceivable at a larger distance and do not require a cooperation of the individual.

In general, there are two main approaches for gait-based person identification, namely appearance-based (model free) and model-based ones [10]. Appearance-based approaches focus on identifying persons using shape, silhouette, geometrical measures, etc. On the other hand, model-based approaches are focused on identifying persons taking into account the kinematic characteristics of the walking manner. The majority of the approaches are based on appearance and rely on analysis of image sequences acquired by a single camera. The main drawback

of such approaches is that they can perform the recognition from a specific viewpoint. To achieve view-independent person identification, Jean et al. [4] proposed an approach to determine view-normalized body part trajectories of pedestrians walking on potentially non-linear paths. However, Yu et al. [15] reported that view changes cause a significant deterioration in gait recognition accuracy. Viewpoint dependence is still significant problem for many gait analysis techniques, since it is not easy to match between different orientations of the subject.

Model-based gait recognition is usually based on 2D fronto-parallel body models. In [14], a sequential set of 2D stick figures is utilized to extract gait patterns. Afterwards, a SVM classifier is used to classify gender using such gait signatures. The use of 3D technology for gait analysis [1] dates back to 1990. The 3D approaches for gait recognition model human body structure explicitly, often with support of the gait biomechanics [13]. They are far more resistant to view changes in comparison to 2D ones. In [11] 3D locations of markers were utilized, from which joint-angle trajectories measured from normal walks were derived. The recognition was performed using dynamic time warping on the normalized joint-angle information. It was done on two walking databases of 18 people and over 150 walk instances using nearest neighbor classifier with Euclidean distance. In [9] 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. In [12], 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 (moCap). To overcome the non-frontal pose problem, more recently a multi-camera based gait recognition method has been developed [3]. In the mentioned work, joint positions of the whole body are employed as a feature for gait recognition.

In controlled laboratory experiments, gait has been shown to be very effective biometric for distinguishing between individuals. However, in real world scenarios it has been found to be much harder to achieve good recognition accuracy. Most of gait analysis techniques, particularly neglecting 3D information, are unable to reliably match gait signatures from differing viewpoints. Moreover, they are also strongly dependant on the ability of the background segmentation and require accurate delineation between the subject and the background.

In this work we present an approach for 3D gait recognition. The motion parameters are estimated on the basis of markerless human motion tracking. They are inferred with the help of a 3D human model. The estimation takes place on video sequences acquired by four calibrated and synchronized cameras. We show the tracking performance of the motion tracking algorithm using ground-truth data acquired by a commercial motion capture (moCap) system from Vicon Nexus. The identification is done on the basis of the proposed spatio-temporal motion descriptors. We show that they allows us to achieve better results in comparison to an algorithm based on third order tensor and a reduction of the tensorial gait data by Multilinear Principal Components Analysis (MPCA) [7].

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 the key parts of the body. The 3D model specifies a kinematic chain, where the connections of body parts comprise a parent-child relationship, see Fig. 1. The pelvis is the root node in the kinematic chain and at the same time it 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 persons's position, orientation and joint angles. To account for different body part sizes, limb lengths, and different ranges of motion we employ a set of pre-specified parameters, which express typical postures. For each degree of freedom there are constraints beyond which the movement is not allowed. The model is constructed from truncated cones and is used to generate contours, which are then matched with the image 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.
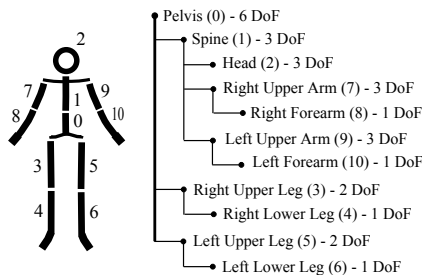

Fig. 1. 3D human body model consisting of 11 segments (left), hierarchical structure (right).

2.2 Articulated Motion Tracking

Estimating 3D motion can be cast as a non-linear optimization problem. The degree of similarity between the real and the estimated pose is evaluated using an objective function. Recently, particle swarm optimization (PSO) [5] has been successfully applied to full-body motion tracking [8]. In PSO each particle follows simple position and velocity update equations. Thanks to interaction between particles a collective behavior of the swarm arises. It leads to the emergence of global and collective search capabilities, which allow the particles to gravitate towards the global extremum. The 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 frame. In this work the 3D motion tracking is achieved through the Annealed Particle Swarm Optimization (APSO) [8].

2.3 Fitness Function

The fitness function expresses the degree of similarity between the real and the estimated human pose. Figure 2 illustrates the calculation of the objective function. It is determined in the following manner: $f(x) = 1 - (f_1(x)^{w_1} \cdot f_2(x)^{w_2})$, 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 3D model into 2D image. The function $f_2(x)$ reflects the edge distance-based fitness. A background subtraction algorithm [8] is employed to extract the binary image of the person, see Fig. 2b. The binary image is then utilized as a mask to suppress edges not belonging to the person, see Fig. 2d. The projected model edges are then matched with the image edges using the edge distance map, see Fig. 2g.
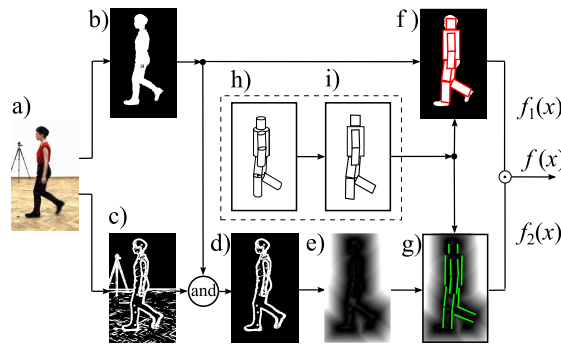

Fig. 2. Calculation of the fitness function. Input image a), foreground b), gradient magnitude c), masked gradient image d), edge distance map e), 3D model h) projected onto image 2D plane i) and overlaid on binary image f) and edge distance map g).

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.

A commercial motion capture system from Vicon Nexus was employed to provide the ground truth data. The system uses reflective markers and ten 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.

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 set of $M = 39$ markers attached to main body parts has been used. 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 determined by our APSO 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 in view-independent gait recognition. In particular, as we can observe, the gait cycle and the stride length can be determined with high precision.
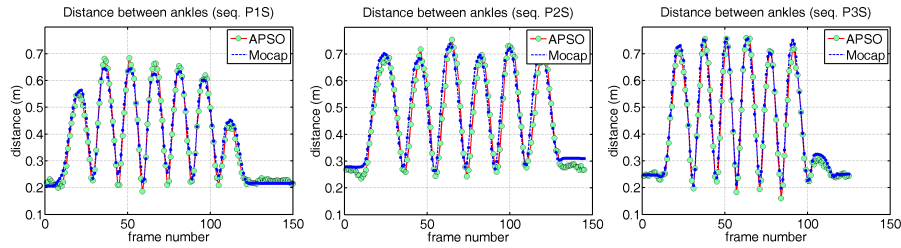

Fig. 4. Distance between ankles in sequences P1S, P2S and P3S (straight).

In a typical system for gait-based person identification a gait signature is extracted in advance. 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 consists of some attributes describing a person, like height, stride length and joint angles estimated by marker-less motion capture.

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 gait signature was determined. 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 dimension was chosen to be 30, which was roughly the average number of frames in each gait cycle. As mentioned in Subsection 2.1, the body configuration is parameterized by the position and the

orientation of the pelvis in the global coordinate system and the angles between the connected limbs. Among the state variables there is roll angle of the pelvis and the angles between the connected limbs. A set of the mentioned above state variables plus the distance between the ankles and the person's height account for the gait sample. In consequence, the dimension of single data sample is 32, and it is equal to the number of bones (excluding pelvis) times three angles plus two (i.e. distance between ankles and the person's height). From such a gait data sample we excluded the angles not used and the resulting dimension of the gait sample was equal to 22. Thus, a single gait pattern was of size 30×22.

In order to comprise the variation of the gait attributes over time we evaluated a number of spatio-temporal motion descriptors. The first group of gait descriptors was calculated on the basis of the normalized cumulative sums and the corresponding variances, which were calculated for each attribute in some time intervals. In the first gait signature g_cum5 they were calculated in every fifth frame, in the second signature g_cum10 they were calculated every tenth frame, whereas in third one g_cum15 they were calculated every fifteenth frame. Given the dimension of the gait sample, the lengths of the gait signatures were 264, 132 and 88. The second group of gait signatures was determined analogously, but instead of the normalized cumulative sums and variances, the averages and the corresponding variances were calculated for a specified range of adjacent frames, i.e. 5, 10 and 15. That means that the first gait signature g_ti5 consisted of the averages and the variances of the attributes corresponding to frames 1-5, 6-10, ..., 25-30. Such gait signatures were then utilized by the classifiers.

4   Experimental Results

The markerless 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 estimated pose, the 3D model was projected to 2D plane and then overlaid on the images. Figure 5 depicts some results, which were obtained for person 1 in a straight walk. The degree of overlap of the projected 3D body model with the performer's silhouette reflects the accuracy of motion tracking.
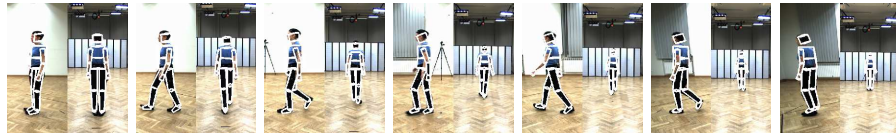

Fig. 5. Articulated 3D human body tracking in sequence P1S. Shown are results in frames #0, 20, 40, 60, 80, 100, 120. The left sub-images are seen from view 1, whereas the right ones are seen from view 2.

The plots depicted in Fig. 6 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 results presented above were obtained by APSO algorithm in 20 iterations using 300 particles.
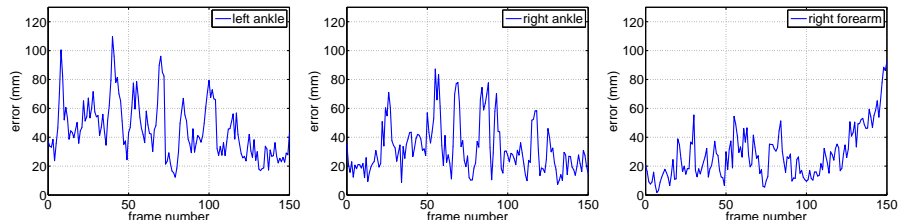

Fig. 6. Tracking errors [mm] versus frame number.

In Table 1 are presented some quantitative results that were obtained using 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 individual markers and the recovered 3D joint locations. For each sequence they were then averaged over ten runs of the APSO with unlike initializations.

Table 1. Average errors for $M = 39$ markers in three image sequences. The images from sequence P1S are depicted on Fig. 5.

	#particles	it.	Seq. P1S error [mm]	Seq. P2S error [mm]	Seq. P3S error [mm]
APSO	100	10	50.9±31.4	54.6±30.3	55.6±34.1
	100	20	46.9±26.5	49.0±27.1	49.7±26.5
	300	10	46.7±26.6	50.4±28.5	50.5±26.1
	300	20	44.4±25.9	46.6±24.8	48.2±25.2

Table 2 shows the recognition accuracy that was obtained on 230 gait cycles from both straight and diagonal walks. Each of 22 persons performed 2 walks in each direction, and each crossing consisted of 2 or 3 full gait cycles. 10-fold cross-validation was used to evaluate the performance of the proposed algorithm for view independent gait recognition. In the evaluation of the system we employed Naïve Bayes (NB), multilayer perceptron (MLP) and a linear SVM. The first column of the table contains acronym of the utilized gait signature, whereas the second one contains the number of the attributes of the gait signature. In order to show the usefulness of the spatio-temporal gait signatures we show also the correctly classified ratio (CCR), which was obtained by the use of all attributes,

see signature g_all, and a signature g_avg consisting of the averages and the variances for each attribute. As we can observe, for rank 1 the best correctly classification ratio was obtained by MLP classifier operating on all attributes. However, taking into account that this result was obtained on large number of attributes, the usefulness of MLP classifier can be reduced in practice due to its computational and memory requirements, tendency to overfitting, and reduced generalization abilities on such set of attributes. Unlike MLP, the computational complexity of SVMs does not depend on the dimensionality of the input space. Moreover, SVM classifiers are usually much quicker. As we can observe, the correctly classified ratios achieved by SVM are better than corresponding MLP CCRs for almost all remaining gait signatures (apart from g_ti5). In particular, for the SVM the best CCRs have been achieved using spatio-temporal features. For rank 3 the SVM gives better results than MLP, excluding the g_all gait signature for which the CCR is equal to the best CCR of the SVM.

Table 2. Correctly classified ratio [%] using data from markerless motion capture.

gait sig.	# att.	rank 1			rank 3		
		NB	MLP	SVM	NB	MLP	SVM
g_all	660	84.8	**96.1**	90.0	94.6	**99.6**	98.3
g_avg	44	87.0	90.9	91.6	95.2	97.0	**99.6**
g_cum5	264	82.6	91.3	**93.5**	92.6	97.8	99.1
g_cum10	132	83.0	90.0	**93.5**	92.6	97.8	98.7
g_cum15	88	83.9	88.7	91.7	93.9	97.0	99.1
g_ti5	264	81.7	91.3	90.4	91.7	98.3	98.3
g_ti10	132	80.4	90.9	92.2	93.6	98.3	98.3
g_ti15	88	83.0	91.3	93.0	95.2	98.7	**99.6**

The discussed experimental results were obtained using data, which were employed in [7] and were produced by our marker-less motion tracking system. For rank 1, the best identification performance obtained by SVM and operating on spatio-temporal features is better about 4% in comparison to results reported in [7], i.e. obtained by MLP operating on tensorial gait data that were reduced using Multilinear Principal Components Analysis (MPCA).

Figure 7 depicts the confusion matrix, which was obtained by SVM and g_cum10 signature. As we can see, eleven persons were classified with 100% probability. The lowest probability associated with true label is equal to 72.7%.

In Tab. 3 are shown results that were obtained using motion data from marker-based motion capture. As we can observe, the 3D data allows us to obtain very promising correctly classified ratios. In particular, the results demonstrate a potential of 3D techniques in view-independent gait analysis. The data obtained by motion capture systems are available at: `http://home.agh.edu.pl/~bkw/research/data/3DGaitData.7z`.
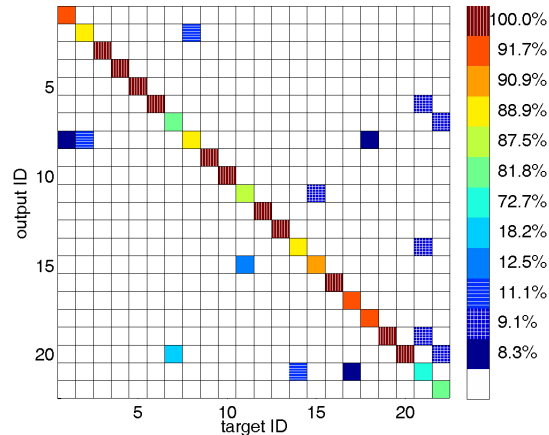
Fig. 7. Confusion matrix for SVM using g_cum10 (CCR=93.5%).

Table 3. Correctly classified ratio [%] using data from marker-based moCap (rank 1).

gait sig.	# att.	NB	MLP	SVM	1NN	3NN	1NN	3NN
					Euclidean dist.		Manhattan dist.	
g_all	8880	100	-	100	100	100	100	100
g_avg	148	99.1	100	100	100	100	100	100
g_cum20	888	99.6	100	100	100	100	100	100
g_cum40	444	99.6	100	100	100	100	100	100
g_cum60	296	99.6	100	100	100	100	100	100
g_ti20	888	99.6	100	100	99.6	100	100	100
g_ti40	444	99.6	100	100	100	100	100	100
g_ti60	296	99.6	100	100	99.6	100	100	100

The complete motion capture system was implemented in C/C++. The recognition performance was evaluated using WEKA software. The marker-less motion capture system runs on an ordinary PC. As demonstrated in [6], the full body motion tracking can be realized in real-time on modern GPUs.

5 Conclusions

We have presented an approach for view-independent gait recognition. The motion parameters were estimated using markerless human motion tracking. The person identification was done on the basis of the proposed spatio-temporal motion descriptors. The classification performance was determined on a dataset consisting of 230 gait cycles that were performed by 22 persons. The correctly classified ratio achieved by SVM is equal to 93.5% for rank 1 and 99.6% for rank 3. The identification accuracy is better than accuracy reported in [7], where

tensorial gait data reduced by Multilinear Principal Components Analysis were classified by a multilinear perceptron. The correctly classified ratio of SVM, MLP and k-NN classifiers on data obtained by the moCap system is equal to 100%.

References

1. Areblad, M., Nigg, B., Ekstrand, J., Olsson, K., Ekstrm, H.: Three-dimensional measurement of rearfoot motion during running. J. of Biomechanics 23(9), 933 – 940 (1990)
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. 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)
4. Jean, F., Albu, A.B., Bergevin, R.: Towards view-invariant gait modeling: Computing view-normalized body part trajectories. Pattern Recogn. 42(11) (Nov 2009)
5. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of IEEE Int. Conf. on Neural Networks. pp. 1942–1948. IEEE Press, Piscataway, NJ (1995)
6. Krzeszowski, T., Kwolek, B., Wojciechowski, K.: GPU-accelerated tracking of the motion of 3D articulated figure. In: Int. Conf. on Computer Vision and Graphics, Lecture Notes in Computer Science, vol. 6374, pp. 155–162. Springer-Verlag (2010)
7. Krzeszowski, T., Michalczuk, A., Kwolek, B., Switonski, A., Josinski, H.: Gait recognition based on marker-less 3D motion capture. In: 10th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance. pp. 232–237 (2013)
8. 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)
9. 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)
10. Nixon, M.S., Carter, J.: Automatic recognition by gait. Proc. of the IEEE 94(11), 2013–2024 (2006)
11. Tanawongsuwan, R., Bobick, A.: Gait recognition from time-normalized joint-angle trajectories in the walking plane. In: IEEE Comp. Society Conf. on Computer Vision and Pattern Recognition. vol. 2, pp. 726–731 (2001)
12. 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)
13. 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)
14. Yoo, J.H., Hwang, D., Nixon, M.S.: Gender classification in human gait using support vector machine. In: Proc. of the 7th Int. Conf. on Advanced Concepts for Intelligent Vision Systems. pp. 138–145. LNCS, Springer-Verlag (2005)
15. Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: 7th Asian Conf. on Computer Vision, Lecture Notes in Computer Science, vol. 3851, pp. I:807–816. Springer Berlin Heidelberg (2006)