DTW-based Gait Recognition from Recovered 3-D Joint Angles and Inter-ankle Distance

Tomasz Krzeszowski[3], Adam Switonski[2,4], Bogdan Kwolek[1], Henryk Josinski[4], and Konrad Wojciechowski[2]

[1] AGH University of Science and Technology
30 Mickiewicza Av., 30-059 Krakow, Poland

[2] Polish-Japanese Institute of Information Technology
Al. Legionów 2, 41-902 Bytom, Poland

[3] Rzeszów University of Technology
Al. Powst. Warszawy 12, 35-959 Rzeszów, Poland

[4] Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland

tkrzeszo@prz.edu.pl, aswitonski@pjwstk.edu.pl, bkw@agh.edu.pl,
henryk.josinski@polsl.pl, kwojciechowski@pjwstk.edu.pl

Abstract. We present a view independent approach for 3D human gait recognition. The identification of the person is done on the basis of motion estimated by our marker-less 3D motion tracking algorithm. We show tracking performance using ground-truth data acquired by Vicon motion capture system. The identification is achieved by dynamic time warping using both joint angles and inter-joint distances. We show how to calculate approximate Euclidean distance metric between two sets of Euler angles. We compare the correctly classified ratio obtained by DTW built on unit quaternion distance metric and such an Euler angle distance metric. We then show that combining the rotation distances with inter-ankle distances and other person attributes like height leads to considerably better correctly classified ratio.

1 Introduction

Gait is an attractive biometric feature for human identification at a distance [2]. Compared with other biometric modalities, such as face or iris, gait has many advantages since the identification techniques are non-contact, non-invasive, perceivable at a larger distance and do not require cooperation of the individual.

The existing methods for gait recognition can be divided in two main categories: appearance-based (model-free) and model-based [9]. Appearance-based gait recognition approaches consider gait as a holistic pattern, where the full-body of a human subject is represented by silhouettes or contours. Model-based approaches identify individuals on the basis of kinematic characteristics of the walking manner. Model based approaches fit a model to human body and represent gait using the parameters of the model that are estimated over time. Model based approaches are more complex and computationally more expensive than model free approaches. Thus, the majority of the approaches are based on appearance and rely on analysis of image sequences acquired by a single camera. The main limitation 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. However, as reported by Yu et al. [15], the appearance-based methods are view dependent and perform best when a side view is used.

The use of 3D gait analysis dates back to the nineties of the last century [1]. In 3D gait recognition the human body structures are modeled explicitly, often with support of the gait biomechanics [14]. As a result, they are far more resistant to view changes than 2D approaches. In [12] 3D markers locations were used to extract joint-angle trajectories. The recognition was achieved using dynamic time warping on the normalized joint-angle information and nearest neighbor classifier with Euclidean distance. It was evaluated on two walking databases of 18 people and over 150 walk instances. In [13], an approach relying on matching 3D motion models to images, and then tracking and restoring the motion parameters is proposed. The system was evaluated 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. The motion models were derived on the basis of Vicon motion capture system (moCap). Recently, in [3] a multi-camera based gait recognition method has been developed to overcome the non-frontal pose problem. The recognition was done using the recovered 3D human joints.

Despite its attractive features, gait-based person identification is still far from being ready to be deployed in practice. What limits the use of gait recognition systems in real-world scenarios is the impact of lots of covariate factors, which affect the dynamics of the gait. The most important covariate factors include camera setup (viewpoint), lightning, walking surface, footwear and clothing, carrying conditions. Thus, most of gait analysis techniques, particularly neglecting 3D information, are unable to reliably match gait signatures from differing viewpoints, but also in case of different walking surface, different clothing. 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, 3D-joint angles and locations are estimated on the basis of marker-less 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 system from Vicon Nexus. The person identification is done on the basis of dynamic time warping (DTW) using both joint angles and inter-joint distances. We show how to calculate approximate Euclidean distance metric between two sets of Euler angles. We compare the correctly classified ratio obtained by DTW built on unit quaternion distance metric and such an Euler angle distance metric. We then show that combining the rotation distances with the inter-ankle distances and other person attributes like height leads to considerably better correctly classified ratio of the person identification system.

2 Articulated Motion Tracking

The purpose of motion capture systems is to measure the motion of bony segments during various activities of the performer. Optical marker motion capture (marker-based) relies on attaching reflective markers to be tracked using standard computer-vision techniques. One of the most popular commercially available systems is provided by Vicon. The Vicon system relies on infrared cameras and artificial reflective markers. The markers are attached to predefined body parts of a human subject. The markers reflect the light signal, which is emitted and then registered by a set of infrared cameras surrounding the subject. The data from each camera consisting of 2D coordinates of each recognized marker position is matched and then used by the triangulation algorithm, which computes the 3D position and the label of each visible marker. The motion data are stored in the C3D format, in which each frame of information is represented as a list consisting of Cartesian x, y, z coordinates for each marker. The motions are stored in ASF/AMC format with 19 defined segments, see also the skeleton on Fig. 1. They are calculated on the basis of 39 markers.

Most approaches for marker-less 3D motion tracking use a human body model to guide the pose estimation process, as the use of a model greatly increases the accuracy and robustness of the pose recovery. In our system [6], the articulated model of the human body is built on kinematic chain with 11 segments. Such a 3D model consists of truncated cones that model the pelvis, torso/head, upper and lower arm and legs, see Fig. 1. Its configuration is defined by 26 DOF and it is determined by position and orientation of the pelvis in the global coordinate system and the relative angles between the connected limbs. Each truncated cone is projected into 2D image plane via perspective projection. In this way we attain the image of the 3D model in a given configuration, which can then be matched to the person extracted through image analysis. A modified Particle Swarm Optimization (APSO) algorithm is used to estimate the 3D motion [5]. The motion is inferred using four calibrated and synchronized cameras.

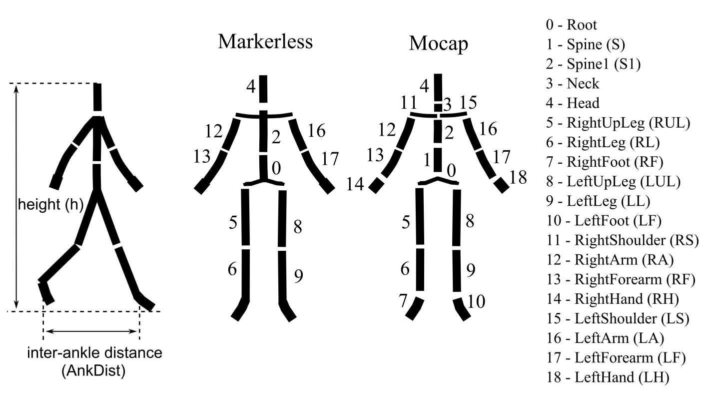

Fig. 1. Human attributes and joints used in gait recognition.

3 DTW on Joint Rotations and Geometric Relations

Dynamic time warping (DTW) [10] is algorithm for assessment the similarity between two temporal sequences, which may vary in time or speed. DTW is often used in motion analysis [7]. The similarities in walking patterns could be measured by DTW, even if one person was walking faster than the other, or if there were accelerations and decelerations during the movements. It determines an optimal match between two given sequences and returns distance-like measure between sequences. The sequences are warped non-linearly in the time dimension to match each other as closely as possible.

In the DTW algorithm, two motions R and R' of length N and M frames, respectively, are compared using local cost measure c. By evaluating c for all possible pairs of data we get a cost matrix C of size $N \times M$, which contains the matching costs. To find the best match between the given sequences we should find a path through the grid, which minimizes the total distance between them. The alignment minimizing the cost is represented by a so-called warping path, which, under certain constraints, optimally allocates the frame indices of the first motion with the frame indices of the second motion. Such optimal path can be found in $O(NM)$ using dynamic programming.

The natural way for representing human motion is to encode the rotations of each bone around three axes. One possible alternative is to utilize unit quaternions for encoding joint rotations. Quaternion-based pose distance is often used in DTW-based motion comparison [11]. However, quaternions can represent only rotations around a line through the origin of the coordinate system. As pointed out in [8], geometric relations between body key points are very important motion features. Thus, in our DTW-based algorithm for gait recognition we employ both individual distances between corresponding joint rotations as well as geometric relations between body parts.

Assume that B is a set of bones that are directed away from the root of the kinematic chain. The minimum angle between two angles $\Delta(\theta, \theta')$ was computed as follows: $\Delta(\theta, \theta') = \pi - ||\theta - \theta'| - \pi|$. The approximate measure of the distance between the two joint rotations is the sum of each of the Euler angle differences: $\rho_\theta = \sqrt{\Delta(\theta, \theta')^2 + \Delta(\phi, \phi')^2 + \Delta(\eta, \eta')^2}$. The local cost measure is equal to sum of the approximate distances between each two corresponding joint rotations: $c_\theta = \sum_{b \in B} \rho_\theta(b)$. For DTW employing both individual distances between corresponding joint rotations, geometric relations between body parts, and other body features, their weighted 3D Euclidean distance terms are included in the objective function: $c_{pose} = w_1 c_\theta + w_2(|h - h'| + |AnkDist - AnkDist'|)$, where h denotes persons's height, $AnkDist$ stands for inter-ankle distance, $w_1 = 57.3$ and $w_2 = 1.7$.

The quaternion based pose distance between two human poses has been calculated as sum of terms expressing distances between unit quaternions: $c_{\text{quat}} = \sum_{b \in B} 2/\pi \arccos(\langle q_b,\ q'_b \rangle)$, where $\langle \cdot,\ \cdot \rangle$ stands for standard dot product in $\mathbb{R}^4$. The term $\arccos(\langle q_b,\ q'_b \rangle)$ is known as the geodesic distance of the real three-dimensional projective space, which assumes it maximal value equal to $\pi/2$ for the case that q_b and q'_b are orthogonal.

4 Experimental Results

The marker-less motion tracking system was evaluated on video sequences with 22 walking individuals. In each 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 model was projected to 2D plane and then overlaid on the images. Figure 2 depicts some results that 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. The results were obtained by APSO consisting of 300 particles and executing 20 iterations per frame.

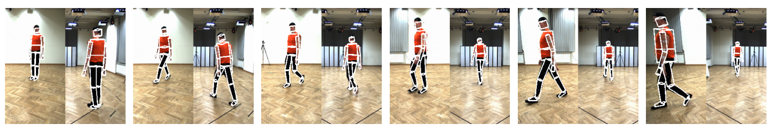

Fig. 2. 3D human body tracking in sequence p1s2. Shown are results in frames #0, 20, 40, 60, 80, 100. The left sub-images are seen from view 1, the right ones - view 2.

In Table 1 are shown some quantitative results that were obtained using the discussed image sequences. Given the human pose estimated by marker-less system, as well as the locations of physical markers, the virtual markers were generated on the 3D model utilized in marker-less moCap. The errors were calculated using the locations of the markers recovered by marker-based moCap system and locations of the virtual markers estimated by our marker-less moCap system. For each frame they were computed as average Euclidean distance between corresponding physical and virtual markers. 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 four image sequences. The images from sequence p1s2 are depicted on Fig. 2.

			Seq. p1s1	Seq. p1s2	Seq. p2s1	Seq. p2s2
	#particles	it.	error [mm]	error [mm]	error [mm]	error [mm]
APSO	100	10	60.0±42.9	51.3±25.5	59.8±30.4	55.8±23.2
	100	20	50.1±29.3	47.6±21.5	57.8±24.7	55.4±20.3
	300	10	48.4±29.9	48.4±24.7	58.5±26.6	56.2±20.5
	300	20	44.9±22.1	45.0±19.9	56.3±22.1	54.1±17.4

Table 2. CCR obtained by DTW built on angle distance metrics.

	System	Metrics	Rank 1 CCR	Rank 1 [%]	Rank 2 CCR	Rank 2 [%]	Rank 3 CCR	Rank 3 [%]
crossval	Motion	Euclid.	199	86.5	216	93.9	223	97.0
		Manhattan	202	87.8	217	94.3	222	96.5
		Quat. geod.	197	85.6	217	94.3	224	97.4
	moCap	Euclid.	230	100	230	100	230	100
		Manhattan	230	100	230	100	230	100
		Quat. geod.	230	100	230	100	230	100

Table 2 presents correctly classified ratio (CCR) that was obtained by DTW built on angle distance metrics only. It demonstrates the CCRs, which were achieved using data from both maker-based and marker-less motion capture systems. At the beginning we evaluated CCR in 10-fold evaluation using approximate Euclidean and Manhattan distance metrics between sets of Euler angles as well as quaternion geodesic distance. As we can observe, the DTW built on approximate Euclidean and Manhattan distances gave better CCR scores than DTW operating on geodesic quaternion distances. The CCR that was obtained using data from marker-based moCap is perfect.

Table 3 shows CCR that was obtained using 3-D joint angles, inter-ankle distance and height, recovered by our marker-less moCap system. As we can observe, the CCR ratio is much higher in comparison to CCR, which was obtained by DTW operating on rotation information only. Thanks to the use of the approximate angle metrics together with joint-to-joint distances and other human attributes like height we achieved considerably better identification performance. We calculated also CCR using separated test and validation data. As we can notice, for marker-less system, the CCR obtained on separate test and validation data is smaller in comparison to CCR, which was obtained in the 10-fold cross-validation.

Figure 3 depicts the confusion matrix both for DTW operating on rotation data only and rotation data together with inter-ankle distances and height. As we can see, the distance between joints as well as person attributes like height lead to far better identification scores.

Figure 4 illustrates the warping paths for LeftUpLeg joints as well as inter-ankle distance, see Fig. 1, which were obtained for person 1 and 2. The discussed figure demonstrates that both angles and inter-joints distances contribute towards better identification performance.

5 Conclusions

In DTW-based comparison of motion sequences the standard approach consists in the use of quaternions. However, the quaternions can not be utilized to measure the inter-joints distances. In order to comprise such inter-joints distances

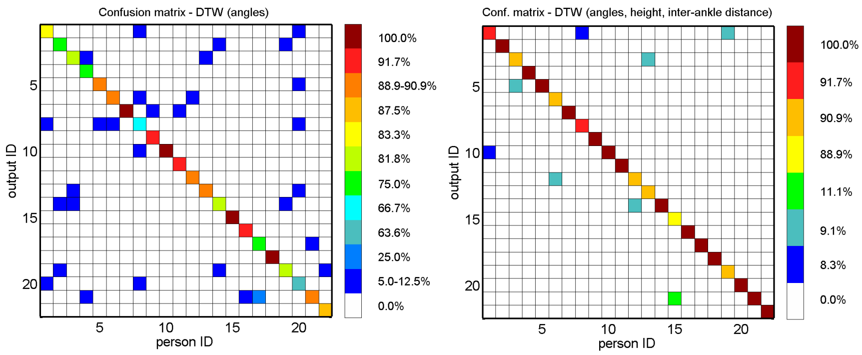

Fig. 3. Confusion matrix for DTW operating on angles only (left), DTW built on angles, inter-ankle distances, and height (right).

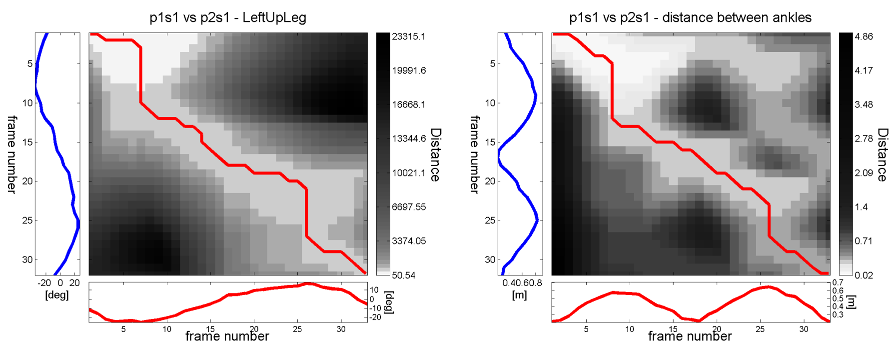

Fig. 4. Warping paths for a joint angle (left) and inter-ankle distance (right).

Table 3. CCR obtained by DTW using angle distance, inter-ankle distance and height.

			Rank 1		Rank 2		Rank 3	
	System	Metrics	CCR	[%]	CCR	[%]	CCR	[%]
crossval	Motion	Euclid.	222	96.5	229	99.6	230	100
		Manhattan	223	97.0	227	98.7	230	100
	moCap	Euclid.	230	100	230	100	230	100
		Manhattan	230	100	230	100	230	100
sep. test/val	Motion	Euclid.	93	80.2	106	91.4	112	96.5
		Manhattan	93	80.2	105	90.5	112	96.5
	moCap	Euclid.	114	98.3	230	100	230	100
		Manhattan	115	99.1	230	100	230	100

as well as other human attributes we proposed to use approximate Euclidean distance metric between two sets of Euler angles in dynamic time warping. Owing to the use of such a metric we can combine the rotations with inter-joint distances and other features. We demonstrated that the combined features allow us to obtain better classification scores.

Acknowledgment. This work has been partially supported by the Polish Ministry of Science and Higher Education within a grant for young researchers U-529/DS/M and the National Science Center (NCN) within the research projects N N516 483240 and DEC-2011/01/B/ST6/06988.

References

1. Areblad, M., Nigg, B., Ekstrand, J., Olsson, K., Ekström, H.: Three-dimensional measurement of rearfoot motion during running. J. Biom. 23(9), 933 – 940 (1990)
2. Boulgouris, N., Hatzinakos, D., Plataniotis, K.: Gait recognition: a challenging signal processing technology for biometric identification. Signal Proc. Magazine, IEEE pp. 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. Krzeszowski, T., Kwolek, B., Wojciechowski, K., Josinski, H.: Markerless articulated human body tracking for gait analysis and recognition. Machine Graphics & Vision 20(3), 267–281 (2011)
6. Krzeszowski, T., Kwolek, B., Michalczuk, A., Switonski, A., Josinski, H.: View independent human gait recognition using markerless 3D human motion capture. In: Computer Vision and Graphics, LNCS, vol. 7594, pp. 491–500. Springer (2012)
7. Kulbacki, M., Bak, A.: Unsupervised learning motion models using dynamic time warping. In: Intelligent Information Systems, Advances in Soft Computing, vol. 17, pp. 217–226. Physica-Verlag HD (2002)
8. Müller, M., Röder, T., Clausen, M.: Efficient content-based retrieval of motion capture data. ACM Trans. Graph. 24(3), 677–685 (Jul 2005)
9. Nixon, M.S., Carter, J.: Automatic recognition by gait. Proc. of the IEEE 94(11), 2013–2024 (2006)
10. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. on Acoustics, Speech and Signal Processing 26(1), 43–49 (Feb 1978)
11. Switonski, A., Polanski, A., Wojciechowski, K.: Human identification based on the reduced kinematic data of the gait. In: Int. Sym. Sig. Pr. Anal. pp. 650–655 (2011)
12. Tanawongsuwan, R., Bobick, A.: Gait recognition from time-normalized joint-angle trajectories in the walking plane. In: CVPR. vol. 2, pp. 726–731 (2001)
13. 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)
14. 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)
15. Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: ACCV, LNCS, vol. 3851, pp. 807–816. Springer (2006)