
Gait recognition on the basis of markerless motion tracking and DTW transform

Adam Switonski¹ · Tomasz Krzeszowski² · Henryk Josinski¹ · Bogdan Kwolek³ · Konrad Wojciechowski⁴

Abstract In this paper, a framework for view-invariant gait recognition on the basis of markerless motion tracking and dynamic time warping transform is presented. The system consists of a proposed markerless motion capture system as well as introduced classification method of mocap data. The markerless system estimates the 3D locations of skeleton-driven joints. Such skeleton-driven point clouds represent the human poses over time. We align point clouds in every pair of frames by calculating the minimal sum of squared distances between the corresponding joints. A point cloud distance measure with temporal context has been utilized in kNN algorithm to compare time instants of motion sequences. In order to enhance the generalization of the recognition as well as shorten the processing time, for every individual a single multidimensional time-series among several multidimensional time-series describing the individuals gait is determined. Moreover, on the basis of recogni-

tion carried out with respect to successive markers the segment ranking has been constructed. The correct classification rate has been determined on the basis of real dataset of human gait. It was acquired in the Human Motion Laboratory of PJAiT and contains 230 gait cycles of 22 subjects. The tracking results on the basis of markerless motion capture are referenced to highly precise Vicon system, whereas the achieved accuracies of recognition are compared to the ones obtained by DTW that is based on rotational data of subsequent joints.

1 Introduction

Motion capture systems are commonly used to capture the detailed human locomotion. The locomotion data are in the form of time series of pose attributes consisting of joint rotations and global translation. In recent years, human motion analysis has been actively studied due to a high application potential in diagnosis of diseases [27,24], sports biomechanics [15], action recognition [8,6,9] and human identification [2,4,5]. The main challenges in gait analysis are related to acquiring motion data as well as description, and assessment of quantities characterizing human locomotion. Among these applications, significant number of studies are devoted to human identification. In relation to other biometric methods, gait recognition has numerous advantages, such as noninvasiveness, noncontactness, nonawareness, capability of being identified at a distance and difficulty to conceal [23,18]. Moreover, gait patterns can be extracted and measured even on the basis of low resolution videos. Hence, human gait identification is very important problem due to possible applications in medicine, biomechanics, biometrics and surveillance.

A. Switonski
Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland E-mail: adam.switonski@polsl.pl

T. Krzeszowski
Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, ul. W. Pola 2, 35-959 Rzeszow, Poland

H. Josinski
Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

B. Kwolek
Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, 30 Mickiewiczza Av., 30-059 Krakow, Poland

J. Wojciechowski
Research and Development Center of Polish-Japanese Academy of Information Technology, Aleja Legionow 2, Bytom, Poland

Emerging technique of motion data acquisition is markerless motion capture, which works on the basis of multicamera registration. A parameterized articulated body model is projected and then matched to image data. In contrast to traditional motion sensors that have quite limited maximal distance at which the motion can be estimated, the multi-camera markerless systems can estimate motions at far larger distances. It is worth noting that they can deal far better with occlusions and self-occlusions. Markerless systems can register motion data without awareness of human, which has crucial meaning for applications consisting in gait recognition.

There is a rich literature of various gait recognition techniques that can be roughly divided into two major categories, namely model-free (appearance-based) and model-based. The model-free approaches usually use a moving shape and combine it with motion. They strongly depend on the extracted silhouettes and are not resistant to different clothing or carried luggage. Moreover, the majority of the methods belonging to this group can achieve correct results only from a specific viewpoint, usually fronto-parallel (side-view) [20, 21, 19, 30]. The well known model-free method called gait energy image (GEI) [21] calculates an average of successive sequence frames with silhouettes. It represents both shape and movement of a gait cycle.

The model-based approaches focus on recovering a structural model of human motion, which is then used to establish gait patterns. Their main advantages are referred to handling of self-occlusion, invariance to scale and rotation as well as resistance to noise. Model based techniques are also less sensitive to individual's appearance and clothing [19].

In the most often used approach to motion data classification, the feature extraction is carried out. In such a case, motion sequences are transformed into fixed dimensional vector space and further recognized by a supervised or unsupervised machine learning algorithm. There are many different strategies to calculate features. In [1] structural gait parameters such as height as well as stride and footprint poses, whereas in [2] strike and clearance poses are stated to be motion descriptors. In medical applications the features usually correspond to clinical diagnosis. They can be related to gait symmetry [32], stride length, walking speed and phase coordination [24] as well as stance, double support and gait cycle [27]. There are also many generic feature extraction approaches in which attributes are determined by a specified transformation. Recently, Balazia & Sojka [3] presented algorithms for gait recognition, which transform input space in a linear way on the basis of Maximum Margin Criterion (MMC) as well as PCA and LDA (PCALDA) techniques. They reported that their

methods outperform thirteen relevant methods of motion data classification in the human gait identification task.

As shown in [16, 29], Dynamic Time Warping (DTW) is a very effective algorithm for motion sequence comparison. It dynamically scales the compared time series in the time domain, which results in minimization of influence of local shifts between motion phases with regard to total dissimilarity. There are numerous applications of DTW. It was used in gait identification [16, 29], gesture recognition [11], graphical symbols, handwritten characters and footwear prints recognition [13], assessment of tennis shots and selected sport activities [28, 22].

In time series analysis the Hidden Markov Models (HMM) are also broadly used [11, 8, 26, 31]. In this case, motion data are stated to be Markov chains with hidden states. Another approach to motion data classification is based on Recurrent Neural Networks (RNN) with long and short terms memory architectures [10, 6, 7].

In this paper, a complete system for human identification on the basis of gait is proposed. It consists of a developed markerless motion capture system as well as introduced and validated gait recognition method. Having on regard that the motion represented by skeleton-driven joint positions is unchanged if we translate it along the floor plane or rotate it about the vertical axis, we align the coordinate systems for every pair of frames. In the proposed approach, every pair of point clouds consisting of skeleton-driven 3D joint positions is aligned by calculating the minimal sum of squared distances between the corresponding joints. Gait recognition is achieved on the basis of Dynamic Time Warping (DTW) extended with classification scheme operating on aligned point clouds. In the experimental evaluations, a gait database containing data of 22 humans and 230 cycles has been collected and then utilized in evaluations of the system. Our classification method is inspired by the registration curves [14]. We demonstrate experimentally that DTW operating on aligned point clouds achieves superior results in comparison to results obtained by recently proposed methods.

The contribution of this work is a model-based method for gait recognition on the basis of motion data from a markerless system. In the proposed method the DTW is used to align and compare motion sequences, whose distance metric is calculated on the basis of aligned point clouds representing the skeleton-based 3D joint positions, whereas the nearest neighbor and minimum distance classifiers are used to identify individuals. In order to enhance the generalization of the recognition performance as well as decrease the computational overloads, for every individual a single multidimensional

time-series among several multidimensional time-series describing the individual's gait is determined. One of the advantages of our model-based approach to gait recognition, in which skeleton-driven 3D joint positions are estimated on the basis of multi-camera markerless motion capture is that the gait recognition system is view-invariant and can perform gait recognition at larger areas.

2 Markerless human motion tracking

Markerless motion tracking utilizes a baseline, calibrated multicamera acquisition. Successive 3D poses of a human are determined on the basis of an analysis of subsequent frames of video recordings. It is realized by matching an assumed articulated body model to silhouettes extracted from image data. Hence, there are two crucial challenges - the optimization technique, which is responsible for an effective search through a space of model parameters as well as an assessment of the projected model onto video data.

3D model of human body A 3D model is used to simulate the human motion and to provide the estimates of the current position, orientation and joint rotations. Our human body model, characterized by the kinematic tree, consists of 11 bone segments with limbs represented by truncated cones [12,17], as depicted in Fig. 1. There are three basic rotations performed by human joints, which are called flexion/extension, abduction/adduction and medial rotation. However, the number of degrees of freedom (DoF) of some segments is lower. It is caused by the anatomic limitations of movements range and assumed simplifications of the model. In total, the model is described by 26 parameters, whereas 23 of them are referred to joint and skeleton rotations and other three ones represent global translation.

Fitness Function In the first stage, background subtraction of input video frame (Fig. 2a) is carried out and binary silhouettes are extracted (Fig. 2b). The image edges are detected by the gradient operator (Fig. 2c) and they are masked by determined silhouettes. In the next stage, edge distance map is computed (Fig. 2d). It assigns to every pixel a value of its distance to the closest edge pixel.

The final fitness function $f_c(\mathbf{x}, \mathbf{y})$, which reflects the degree of similarity between the real \mathbf{x} and the estimated human pose \mathbf{y} in the camera c has two main components: $f_{1,c}(\mathbf{x}, \mathbf{y})$ and $f_{2,c}(\mathbf{x}, \mathbf{y})$, which are related to an extracted human silhouette and an edge distance

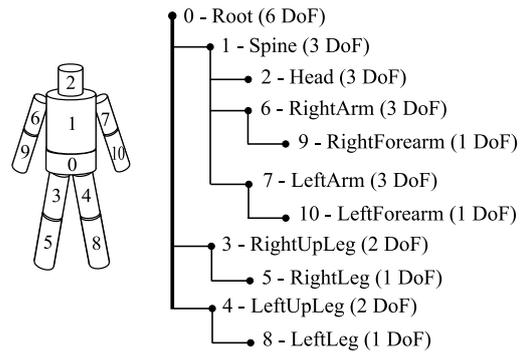


Fig. 1: 3D human body model (left), hierarchical structure (right)

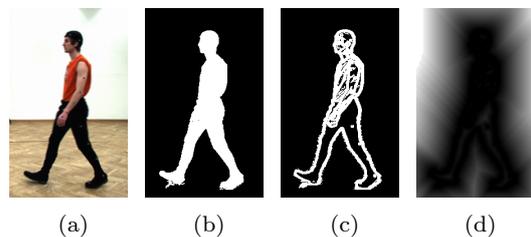


Fig. 2: Processing of image data, (a) - input image, (b) - silhouette, (c) - edges, (d) - distance map

map [12,17,18]. They are aggregated using weighting coefficients w_1 and w_2 in the following manner:

$$f_c(\mathbf{x}, \mathbf{y}) = 1 - (f_{1,c}(\mathbf{x}, \mathbf{y})^{w_1} + f_{2,c}(\mathbf{x}, \mathbf{y})^{w_2}) \quad (1)$$

The $f_{1,c}(\mathbf{x}, \mathbf{y})$ component defines the degree of overlap of the rendered 3D model with the extracted silhouette, whereas $f_{2,c}(\mathbf{x}, \mathbf{y})$ is calculated by comparison of the distance maps. The values of w_1 and w_2 were determined experimentally. The fitness function for all four cameras is calculated according to the following expression: $f(\mathbf{x}, \mathbf{y}) = \frac{1}{4} \sum_{c=1}^4 f_c(\mathbf{x}, \mathbf{y})$.

Articulated Motion Tracking In the motion tracking the annealed particle swarm optimization (APSO) [17] was used. The algorithm is initialized with a group of hypothetical solutions called particles with assigned values of parameters of human body model. Particles move through the solution space in successive time instants i and undergo evaluation according to the fitness function. They are described by current velocity $\mathbf{v}^{(i)}$ and position $\mathbf{x}^{(i)}$. There is a cooperation between individuals in a swarm and every particle knows the global best solution $\mathbf{gbest}^{(i)}$, which has been found by swarm members. The velocity and position of particle l are altered toward explored best global and local poses

according to the following formula:

$$\mathbf{v}_l^{(i+1)} = \chi^{(i)}[\mathbf{v}_l^{(i)} + c_1 \mathbf{r}_1(\mathbf{pbest}_l^{(i)} - \mathbf{x}_l^{(i)}) + c_2 \mathbf{r}_2(\mathbf{gbest}^{(i)} - \mathbf{x}_l^{(i)})] \quad (2)$$

$$\mathbf{x}_l^{(i+1)} = \mathbf{x}_l^{(i)} + \mathbf{v}_l^{(i+1)} \quad (3)$$

where $\chi^{(i)}$ is constriction factor, $\mathbf{pbest}_l^{(i)}$ denotes best solution found by a particle l till iteration i , \mathbf{r}_1 and \mathbf{r}_2 are uniformly distributed random numbers, c_1 and c_2 are positive constants, which are used to balance the influence of the individual's knowledge and that of the group. The value of $\chi^{(i)}$ depends on annealing factor $\alpha^{(i)}$ in the following manner: $\chi^{(i)} = -0.8 \cdot \alpha^{(i)} + 1.4$ where $\alpha^{(i)} = 0.1 + \frac{i}{I+1}$, $i = 0, 1, \dots, I$ and I denotes the number of iterations. The annealing factor is also used to smooth the fitness function - the larger the iteration number, the smaller the smoothing. In consequence, in the last iteration the smoothing is not carried out at all.

3 Recognition of markerless data

On the basis of our previous experiences on gait identification carried out with respect to highly precise marker based as well as markerless motion capture data [16, 29], Dynamic Time Warping algorithm was chosen for motion data classification.

In our earlier work [16], DTW was applied for the markerless motion capture data with sequences of skeleton joint angles. However, if training and testing sets contain gaits cycles acquired during different sessions with separate calibration stages, the precision of gait identification is unsatisfactory, as shown in Table 1. The performance of the identification can be improved if anthropometric data are involved in pose comparison. For instance, if pose description is extended by a left to right ankle distance and position of a head marker corresponding to height of the human, noticeable progress is achieved [16, 18].

However, fusion of pose attributes referred to rotations of skeleton joints and the ones corresponding to specified segments positions is troublesome. They have quite different scales, meanings and behaviors through a gait cycle. This is the reason we have decided to use a point cloud distance measure with an additional transformation $T_{\alpha,x,z}$. It matches compared poses by rotating one of them around axis Y (vertical direction) and translating by $(x, 0, z)$ vector [14]. It removes an influence of location and the direction of gait instances on the determined dissimilarity.

Moreover, pose descriptions are extended by some specified number τ of preceding time instants as shown in Fig. 3. It preserves pose temporal context.

In Figures 4, 5, 6 and Fig. 7 the matching of point cloud distance is visualized. In default case, see Fig. 4, which does not take into consideration a temporal context and engages complete set of markers located on human body, two poses are adjusted. In Fig. 5 the transformation $T_{\alpha,x(\alpha),z(\alpha)}$ also analyzes five preceding time instants. Thus, the number of markers involved in the comparison increases. There is also possibility to match only a diminished subset of markers as shown in Fig. 6 and Fig. 7. Hence, the distance is focused only on movements of selected body parts. A removal of markers with weak discriminating traits and/or markers that are influenced by acquisition noise could even improve the performance of classification. Moreover, the computational complexity is reduced.

Dynamic Time Warping only assesses dissimilarity of motion sequences, which is not sufficient to achieve gait-based person identification. In classical approach to DTW-based gait recognition, the recognition is achieved using nearest neighbors classification algorithm [16]. This means that on the basis of DTW distance, the k -nearest neighbors (kNN) for the identified individual are determined first, and finally a majority vote is carried out to identify the person.

In this work we follow DTW based approach to gait recognition and propose minimum distance classifier (MDC) that operates on aligned point clouds and is based on DTW distance. The resulting gait recognition algorithm is more resistant to overfitting and has lower run-time computational requirements. In the learning stage of the MDC, a single prototype for every individual in the gallery data is determined, whereas in the recognition phase the closest prototype to the considered instance prototype is determined in order to identify the person. A prototype is an average of subset of motion data sequences (multidimensional time-series) describing the gait of a single person in the gallery dataset. Thus, in run-time of kNN-based person identification, instead of determining the closest neighbor(s) for every motion data in the gallery dataset, we determine the neighbor(s) only for the representative prototypes. In such an approach, each person is represented by a single representative motion data. However, in case of motion data characterizing gait, which are usually of different length, computing such prototype motion data is not straightforward. Averaging by DTW is the problem of finding an average sequence for a set of sequences. The average sequence is the sequence that minimizes the sum of the squares to the set of multidimensional data points. When there are more

Table 1: Precision of gait identification on the basis of DTW using angular distance, inter-ankle distance and height [%] [16]

	Distance metric	RANK1	RANK2	RANK3
Angular distance	Euclidean	45.7	61.2	72.4
	Manhattan	45.7	64.7	71.6
Angular, height and inter-ankle distance	Euclidean	80.2	91.4	96.5
	Manhattan	80.2	90.5	96.5

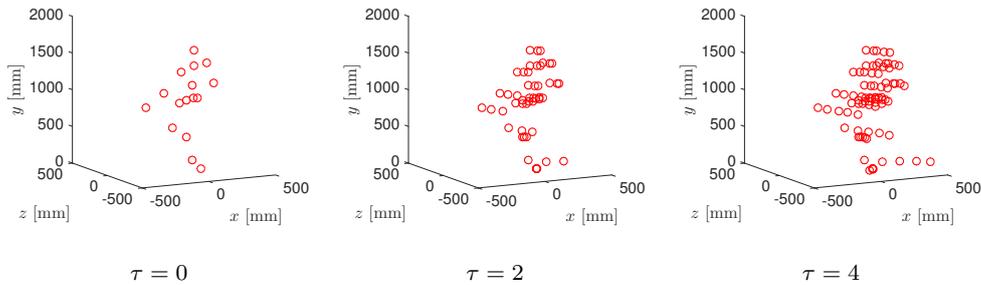


Fig. 3: Temporal context

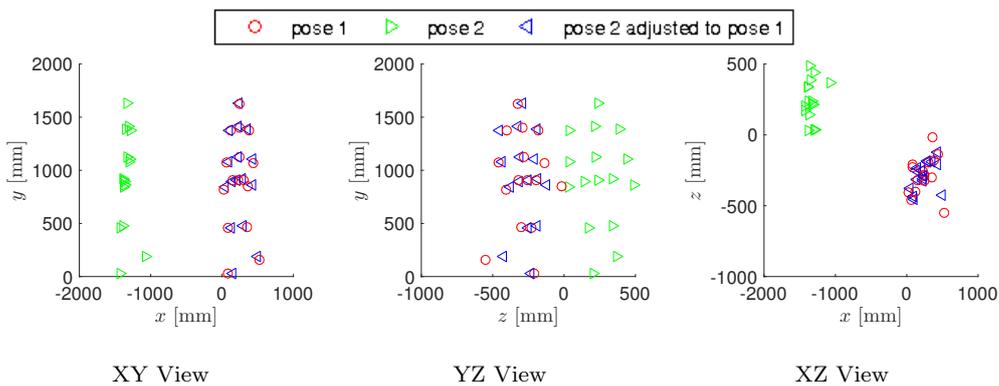


Fig. 4: Example transformation: pose 2 is adjusted to pose 1, complete set of markers, no temporal context

than two data sequences, the problem is related to the one of the multiple alignment and necessitates heuristics. In [25], a method called DBA, which is currently the reference method to average a set of sequences consistently with the DTW has been proposed. It aligns motion sequences in time domain on the basis of DTW transform and calculates mean for corresponding time instants. There is one more challenge to face, namely, it is the way in which the mean is calculated for a set of point clouds. Similarly like in the DTW alignment, the naive, direct approach in which centroids are determined is influenced by the location and the direction of gait. Thus, again transformation $T_{\alpha, x(\alpha), z(\alpha)}$ that aligns the corresponding poses of reference motion and averaged motions is applied.

4 Dataset

To assess performance of the proposed system a real dataset of human gait was collected. It contains 230 gait cycles extracted from 88 video sequences. The acquisition took place in the Human Motion Laboratory of the Polish-Japanese Academy of Information Technology (<http://bytom.pja.edu.pl>) equipped with Vicon software and hardware. The vision subsystem consists of four color, synchronized and calibrated cameras with full HD resolution. The cameras register front, back, left and right views of a gait as illustrated in Fig. 8. The acquisition frequency was set up to 25 fps.

In the recordings, in total 22 volunteers participated - 8 females and 14 males. Two gait paths were designed - the first one joining two opposite cameras and second one joining two nonconsecutive laboratory corners. In total, there are four directions of the gait, two for every

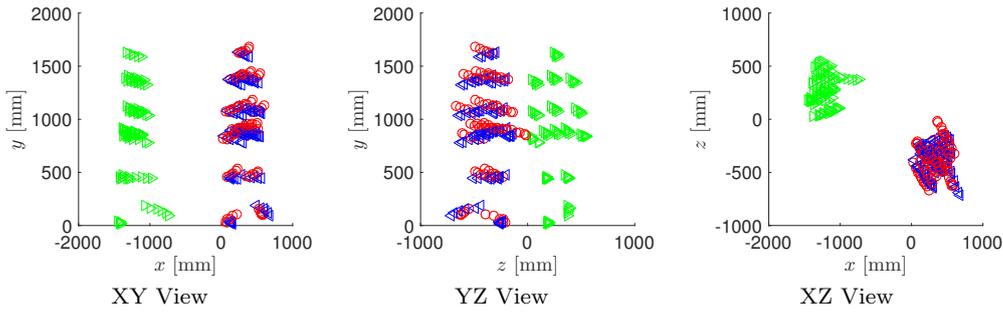


Fig. 5: Example transformation: pose 2 is adjusted to pose 1, $\tau = 5$ for temporal context, complete set of markers

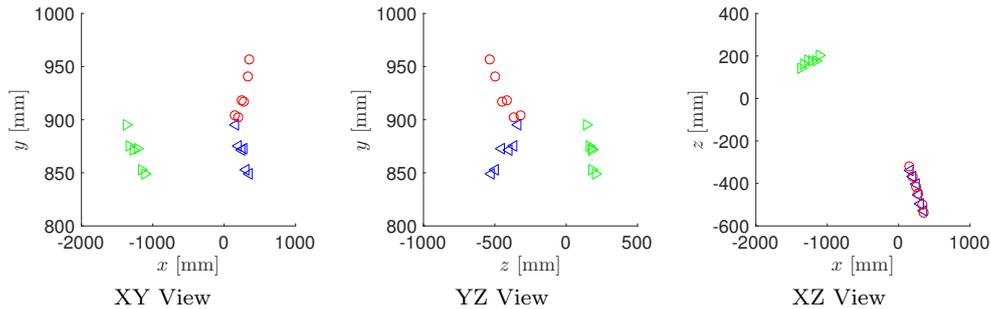


Fig. 6: Example transformation: pose 2 is adjusted to pose 1, $\tau = 5$ for temporal context, only LeftUpLeg marker (see Fig. 1)

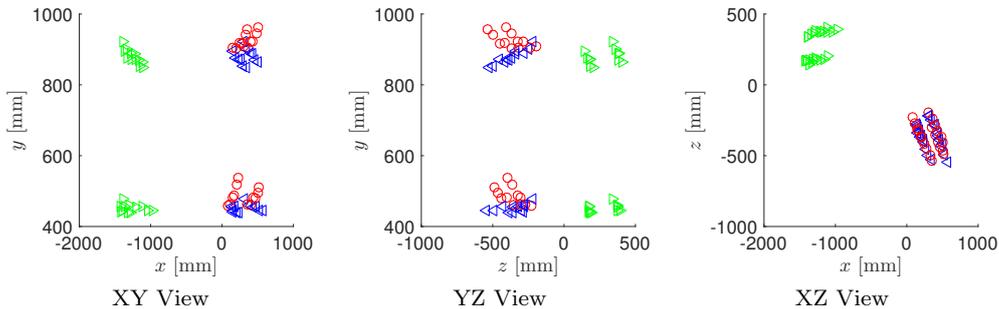


Fig. 7: Example transformation: pose 2 is adjusted to pose 1, $\tau = 5$ for temporal context, only markers located on lower limbs (LeftUpLeg, LeftLeg, RightUpLeg, RightLeg)

path, as shown in Fig. 9. Single video sequence is related to uninterrupted acquisition of gait performed along a specified path and direction. Disjoint gait cycles are extracted from video sequences and they contain two successive steps performed by left and right lower limbs.

Afterwards, the precision of applied markerless motion tracking was evaluated. In capturing, APSO executing 20 iterations and consisting of 300 particles was used. The ground truth data are provided by Vicon commercial marker based motion capture system that is synchronized with vision cameras. The visual assessment for randomly selected person from the dataset is depicted in Fig. 10. As we can observe, the degree of overlap of the projected 3D body model with the per-

son's silhouette on images data appears to be satisfactory, and the differences are not significant.

Table 2: Average errors [mm] for six persons in four image sequences

Person id.	Direction 1	Direction 2	Direction 3	Direction 4
p1	57.4 ± 25.4	63.6 ± 26.6	45.3 ± 19.1	51.8 ± 20.0
p2	48.1 ± 23.4	58.4 ± 27.3	52.4 ± 22.5	59.3 ± 28.6
p3	42.9 ± 20.7	38.8 ± 18.1	46.5 ± 18.4	44.9 ± 18.3
p4	53.4 ± 31.2	47.6 ± 22.9	46.7 ± 22.2	52.5 ± 23.8
p5	62.1 ± 28.4	56.4 ± 23.4	57.1 ± 20.1	56.4 ± 22.8
p6	35.4 ± 18.4	55.7 ± 30.9	35.8 ± 16.0	39.7 ± 18.7

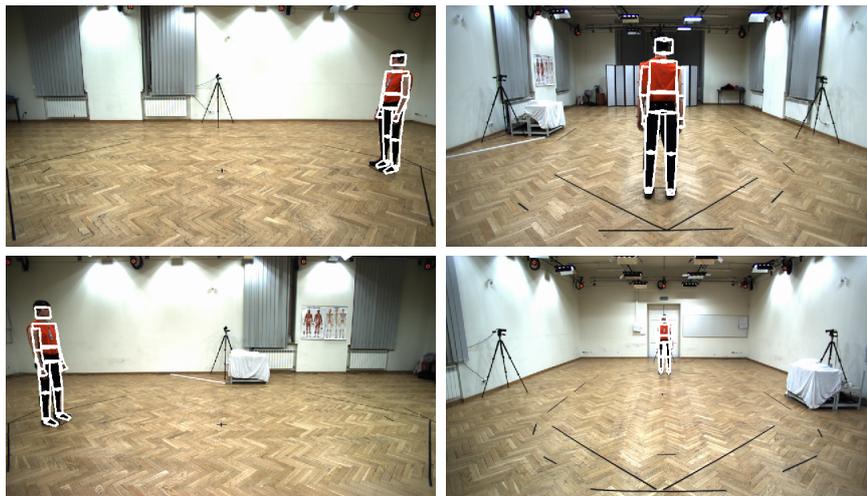


Fig. 8: HML layout and camera setup

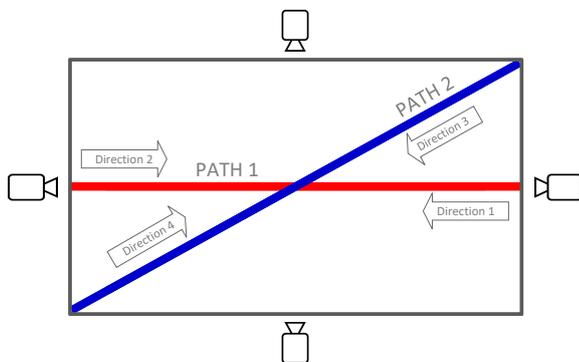


Fig. 9: Paths and directions of gaits, cameras localization

Moreover, quantitative measures were calculated. The average Euclidean distances over whole video sequences between poses estimated by markerless system and locations of physical markers determined by the Vicon are computed. The results obtained for six randomly selected persons are presented in Table 2. The achieved mean distances have to be related to segments lengths. Taking that into consideration the performance of developed markerless motion capture system is acceptable. However, the most important issue is usability of the system and it is assessed by its application in human gait identification.

5 Evaluation protocol and metrics

The developed markerless motion capture system estimates bone segments lengths and their rotations. To obtain compatible results with our previous work that was carried out on the basis of rotational data [16], in

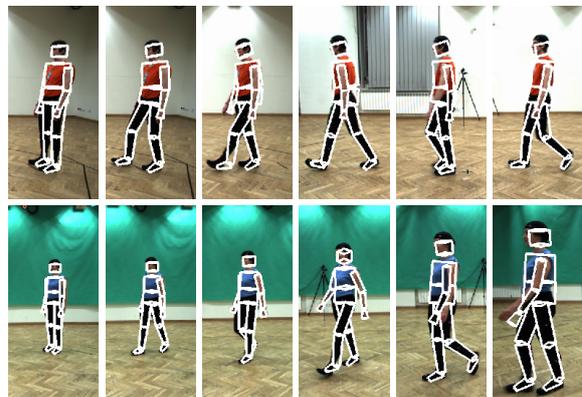


Fig. 10: Tracking results on the two image sequences: p1s1 (first row) and p2s1 (second row) in frames: #0, 20, 40, 60, 80, 100

the classification the virtual markers are used. They are located at the ends of bone segments and their names correspond to a preceding joint (see Fig. 1).

The considered classification problem is related to human gait identification, where individuals are recognized on the basis of their way of walking. The collected database was split in a random way into separate training and testing sets according to video sequences with separate calibration stages of the markerless system (cameras are calibrated only once directly after their setup). This means, that all cycles extracted from a given video sequence belong either to training or to testing set. It guarantees that recognition does not utilize specific features referred to the performed calibration, which estimates orientation of local 3D coordinate systems of skeleton segments. It is consistent with a real applications in which the classification is carried out only on disjoint video recordings.

The results are described by a correct classification rate – percentage of properly identified gait instances of testing set (column RANK1). Moreover, metrics RANK2, RANK3, RANK4 and RANK5 are calculated. They correspond to percentages of correctly recognized gait instances in the first two, three, four and five indications of a classifier, respectively. The iterative procedure is used to determine a subset of class indications. In successive steps instances of already recognized classes are removed from a training set and the recognition is repeated.

6 Results

The classification performance for complete set of markers is presented in Table 3. The discussed table presents the results achieved by the considered classification schemes as well as the influence of the interval τ of temporal context on the precision of gait identification. The obtained results are promising, particularly in comparison to results achieved by kNN/DTW classifier with angular distance measure (see Table 1). As expected, the precision of the classification is significantly improved. It is caused by influences of anthropometric features on markers locations and lack of direct relationships between strict orientation of local coordinates systems of successive skeleton joints and markers positions. Moreover, in reference to hybrid distance, which combines rotation data with inter-ankle distance and height value, there is a substantial progress since the best obtained accuracy of gait identification exceeds 90%, which is more than 10% better. If, for instance, for safety reasons, some specific persons are recognized and it is crucial not to reject their identifications at the expense of greater number of false positives, it is necessary to take some number of the most probable classifier indications. If four are taken into consideration, sensitivity achieves 100% (RANK4), which means no false negatives. In case of three and two, sensitivity is reduced to 99.14% and 97.41%, respectively. The best results are obtained by 1NN classifier, but MDC is only less than 3% worse. Temporal context does not have a great impact on performance of classification if complete set of markers is analyzed.

We compared results achieved by our method with results obtained by PCALDA and MMC algorithms [3]. Similarly to [3], the motion sequences were normalized in time domain. Moreover, two different variants of preliminary motion preprocessing named Variant 1 and Variant 2 were investigated. The second one, more restrictive, directly corresponds to the preprocessing applied in the experiments presented in [3]. It means that

all motion sequences are described by a single skeleton, whereas the root attributes are reset. However, point cloud alignment does not normalize bone segments lengths, which is satisfied in Variant 2. For this reason, in the Variant 1 only root attributes are reset and motion sequences preserve their custom skeletons estimated by motion capture system. This is more adequate to our approach. Thus, it allows comparison of the recognition techniques rather than influence of data preprocessing. Moreover, 1NN DTW classifier with naive Euclidean distance function comparing point clouds for initially preprocessed data in the both variants was examined.

Table 4 presents the results achieved by algorithms proposed by [3] as well as the best classification result, which has been achieved by DTW-kNN with point cloud alignment, see also Tab. 3. There are substantial differences between compared approaches to markerless motion data recognition. In the best case – MMC extraction for data preprocessed in the Variant 1, CCR is only 81.03% which is almost 10% worse than DTW with point cloud alignment. PCALDA and raw DTW are weaker by another 11% and 15%, respectively. As expected, skeleton parameters contain strong individual features, which are helpful in gait identification. Thus, removing them in Variant 2 causes significant decrease of classification accuracy.

Finally, on the basis of prepared segments rankings and in order to reflect the movements of different body parts through a gait cycle, arbitrarily selected markers are utilized in a 1NN classification, as shown in Table 5. It is sufficient to analyze only hand, feet and head segments to obtain 89.66% precision of identification, which is just one misrecognized gait instance more in comparison to the best case of the classification carried out in respect to complete set of markers. Head marker corresponding directly to human height improves the performance noticeably. The analysis of only upper and lower limbs, which are mainly responsible for movements in typical gait, is inadequate for more precise classification.

The experiments were conducted on mobile computer equipped with Intel i7-7820HQ, 2.9 GHz (4 cores) and 16 GB of RAM. The entire tracking process takes approximately 1.3 s per frame. Therefore, processing a single gait cycle (about 30 frames) will take approximately 39 seconds. It should be noted that if necessary, tracking can be realized in real-time [17]. The gait recognition process of single motion sequence with complete set of markers using the kNN/DTW and MD-C/DTW classifiers without temporal context takes approximately 171 ms and 32 ms, respectively. The time of computations depends on the number of samples in the

Table 3: Classification accuracies achieved by kNN and DTW operating on aligned point clouds for complete set of markers [%]

Classifier	τ	RANK1	RANK2	RANK3	RANK4	RANK5
1NN/DTW	0	89.66	96.55	99.14	99.14	100.00
	1	88.79	97.41	99.14	99.14	100.00
	2	89.66	94.83	96.55	99.14	100.00
	3	88.79	94.83	97.41	98.28	100.00
	4	87.07	94.83	97.41	98.28	100.00
	5	86.21	92.24	96.55	98.28	100.00
3NN/DTW	0	90.52	95.69	97.41	99.14	100.00
	1	90.52	93.10	97.41	99.14	100.00
	2	89.66	93.10	95.69	96.55	99.14
	3	87.93	93.10	95.69	97.41	98.28
	4	87.93	93.10	94.83	96.55	98.28
	5	85.34	92.24	93.97	97.41	98.28
5NN/DTW	0	90.52	93.97	97.41	100.00	100.00
	1	88.79	92.24	96.55	100.00	100.00
	2	87.93	92.24	95.69	99.14	100.00
	3	87.07	91.38	94.83	95.69	99.14
	4	85.34	90.52	93.97	96.55	98.28
	5	81.03	89.66	94.83	96.55	97.41
MDC/DTW	0	87.93	95.69	99.14	100.00	100.00
	1	86.21	94.83	99.14	100.00	100.00
	2	85.34	91.38	98.28	100.00	100.00
	3	82.76	89.66	96.55	100.00	100.00
	4	78.45	87.07	95.69	99.14	99.14
	5	77.59	84.48	94.83	96.55	97.41

Table 4: Results achieved by 3NN/DTW and by gait recognition algorithms [3] for complete set of markers [%]

Classifier	RANK1	RANK2	RANK3	RANK4	RANK5	
3NN/DTW $\tau = 0$ (ours)	90.52	95.69	97.41	99.14	100.00	
Variant 1	1NN/PCALDA	69.83	75.86	84.48	88.79	90.52
	1NN/MMC	81.03	88.79	90.52	91.38	92.24
	1NN/Raw DTW	64.66	74.14	80.17	84.48	89.66
Variant 2	1NN/PCALDA	47.41	62.93	72.41	77.59	81.90
	1NN/MMC	67.24	76.72	81.90	86.21	88.79
	1NN/Raw DTW	48.28	57.76	62.07	68.10	74.14

Table 5: Classification results for arbitrary selected markers [%]

Segments	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
LeftHand; RightHand; LeftFoot; RightFoot; Head	88.79	89.66	89.66	88.79	86.21	87.93
LeftHand; RightHand; LeftFoot; RightFoot	79.31	80.17	81.03	80.17	81.03	79.31
LeftArm; RightArm	78.45	78.45	78.45	76.72	75.86	74.14
LeftUpLeg; RightUpLeg	50.00	50.86	47.41	45.69	44.83	43.97

training set for kNN and number of classes for MDC. Average time of single-thread DTW point cloud alignment is 6 ms.

7 Discussion

Multicamera markerless motion capture measurements permit acquisition of motion data without human aware-

ness. There is no requirement to attach markers on human body and to perform calibration movements before every acquisition. It is much more convenient, flexible, mobile and less expensive technique in comparison to marker based motion capture. Thus, quite new opportunities for practical applications are given - for instance, medical or biometrical. However, markerless measurements are influenced by acquisition noise. Due

to noise the analysis of captured data poses additional challenges. In case of classification of such data, features with stronger discriminative power have to be explored. Thus, we incorporated in a recognition both states of successive skeleton joints and anthropometric traits that are mainly related to bone lengths. The results are promising, we obtained substantial progress in comparison to previously investigated approaches. Moreover, the local and global alignment of points clouds outperforms noticeably the analyzed up to date state of the arts methods in the problem of markerless gait identification.

The following observations can be drawn as a result of our work:

1. The proposed markerless motion capture system is characterized by acceptable acquisition noise. The tracking errors are small enough so the system can be used to perform human gait analysis.
2. Dynamic Time Warping is an efficient tool for markerless motion data classification.
3. Point cloud distance is not only coherent approach that merges rotational data and anthropometric features, but furthermore it achieves high performance of classification.
4. Though, complete markers set contains the strongest discriminative features, there are selected subsets of markers for which accuracy of recognition is very similar.
5. Temporal context provides additional movement description and it is useful in case of pose described by reduced markers set.
6. Minimum distance classification in comparison to kNN achieves only bit worse, but still acceptable performance of recognition. On the other hand, it has essential advantage over kNN - it has much less demands of computational power. It can be really important in production systems with extremely large training sets containing numerous data samples.
7. Classification of markerless motion capture data is challenging task. It is not guaranteed that methods, which are efficient in recognition of marker based data as for instance the analyzed PCALDA and MMC, have similar performance if combined with markerless acquisition.

8 Conclusions

This paper proposes a method for view-invariant gait recognition on the basis of motion data from markerless system. The applied point cloud-based distance measure takes into account the temporal context of the

pose, which depends on velocities and accelerations of markers in successive frames. We demonstrate experimentally that kNN and MDC classifiers operating on aligned point clouds achieve superior results in comparison to results obtained by recently proposed methods.

9 Acknowledgments

This work was partially supported by Polish National Centre of Research and Development (NCBiR) under a research grant UOD-DEM-1-183/001, Polish National Science Center (NCN) under grant 2014/15/B/ST6/02808 and a grant DEC-2011/01/B/ST6/06988.

References

1. Ariyanto, G., Nixon, M.S.: Model-based 3d gait biometrics. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–7 (2011). DOI 10.1109/IJCB.2011.6117582
2. Balazia, M., Plataniotis, K.N.: Human gait recognition from motion capture data in signature poses. *IET Biometrics* **6**(2), 129–137 (2017)
3. Balazia, M., Sojka, P.: Gait recognition from motion capture data. *ACM Transactions on Multimedia Computing, Communications, and Applications* **Special issue on Representation, Analysis and Recognition of 3D Human** (2018)
4. Benedek, C., Gálai, B., Nagy, B., Jankó, Z.: Lidar-based gait analysis and activity recognition in a 4d surveillance system. *IEEE Transactions on Circuits and Systems for Video Technology* (2016)
5. Castro, F., Marín-Jiménez, M., Mata, N., Muñoz-Salinas, R.: Fisher motion descriptor for multiview gait recognition. *International Journal of Pattern Recognition and Artificial Intelligence* **31**(01), 1756,002 (2017). DOI 10.1142/S021800141756002X
6. Du, Y., Wang, W., Wang, L.: Hierarchical recurrent neural network for skeleton based action recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1110–1118 (2015)
7. Fragkiadaki, K., Levine, S., Felsen, P., Malik, J.: Recurrent network models for human dynamics. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4346–4354 (2015)
8. Gu, J., Ding, X., Wang, S., Wu, Y.: Action and gait recognition from recovered 3-d human joints. *Trans. Sys. Man Cyber. Part B* **40**(4), 1021–1033 (2010). DOI 10.1109/TSMCB.2010.2043526
9. Hachaj, T., Ogiela, M.R., Koptyra, K.: Human actions recognition from motion capture recordings using signal resampling and pattern recognition methods. *Annals of Operations Research* pp. 1–17 (2016)
10. Harvey, F.G., Pal, C.: Semi-supervised learning with encoder-decoder recurrent neural networks: Experiments with motion capture sequences. *CoRR* **abs/1511.06653** (2015). URL <http://arxiv.org/abs/1511.06653>
11. Ibañez, R., Soria, Á., Teyseyre, A., Campo, M.: Easy gesture recognition for kinect. *Advances in Engineering Software* **76**, 171–180 (2014)

12. John, V., Trucco, E., Ivekovic, S.: Markerless human articulated tracking using hierarchical particle swarm optimisation. *Image and Vision Computing* **28**(11), 1530–1547 (2010)
13. K.C., S., Lamiroy, B., Wendling, L.: DTW-Radon-based Shape Descriptor for Pattern Recognition. *International Journal of Pattern Recognition and Artificial Intelligence* **27**(3) (2013). DOI 10.1142/S0218001413500080
14. Kovar, L., Gleicher, M.: Flexible automatic motion blending with registration curves. In: *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pp. 214–224 (2003)
15. Krzeszowski, T., Przednowek, K., Wiktorowicz, K., Iskra, J.: Estimation of hurdle clearance parameters using a monocular human motion tracking method. *Computer Methods in Biomechanics and Biomedical Engineering* **19**(12), 1319–1329 (2016). DOI 10.1080/10255842.2016.1139092. PMID: 26838547
16. Krzeszowski, T., Switonski, A., Kwolek, B., Josinski, H., Wojciechowski, K.: DTW-based gait recognition from recovered 3-d joint angles and inter-ankle distance. In: L. Chmielewski, R. Kozera, B.S. Shin, K. Wojciechowski (eds.) *Computer Vision and Graphics, LNCS*, vol. 8671, pp. 356–363. Springer International Publishing (2014). DOI 10.1007/978-3-319-11331-9_43
17. Kwolek, B., Krzeszowski, T., Gagalowicz, A., Wojciechowski, K., Josinski, H.: Real-time multi-view human motion tracking using particle swarm optimization with resampling. In: F. Perales, R. Fisher, T. Moeslund (eds.) *Articulated Motion and Deformable Objects, Lecture Notes in Computer Science*, vol. 7378, pp. 92–101. Springer Berlin Heidelberg (2012). DOI 10.1007/978-3-642-31567-1_9. URL http://dx.doi.org/10.1007/978-3-642-31567-1_9
18. Kwolek, B., Krzeszowski, T., Michalczyk, A., Josinski, H.: 3D Gait Recognition Using Spatio-Temporal Motion Descriptors. *Lecture Notes in Computer Science*, vol. 8398, pp. 595–604. Springer International Publishing, Cham (2014). DOI 10.1007/978-3-319-05458-2_61
19. Li, X., Chen, Y.: Gait recognition based on structural gait energy image. *J. Comput. Inf. Syst.* **9**(1), 121–126 (2013)
20. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4, pp. 211–214 Vol.4 (2004). DOI 10.1109/ICPR.2004.1333741
21. Man, J., Bhanu, B.: Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(2), 316–322 (2006). DOI 10.1109/TPAMI.2006.38
22. Margarito, J., Helaoui, R., Bianchi, A.M., Sartor, F., Bonomi, A.G.: User-independent recognition of sports activities from a single wrist-worn accelerometer: A template-matching-based approach. *IEEE Transactions on Biomedical Engineering* **63**(4), 788–796 (2016)
23. Matovski, D.S., Nixon, M.S., Mahmoodi, S., Carter, J.N.: The effect of time on gait recognition performance. *IEEE Transactions on Information Forensics and Security* **7**(2), 543–552 (2012). DOI 10.1109/TIFS.2011.2176118
24. Mian, O.S., Schneider, S.A., Schwingenschuh, P., Bhatta, K.P., Day, B.L.: Gait in swedds patients: comparison with Parkinson’s disease patients and healthy controls. *Movement Disorders* **26**(7), 1266–1273 (2011)
25. Petitjean, F., Ketterlin, A., Ganarski, P.: A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* **44**(3), 678 – 693 (2011). DOI <https://doi.org/10.1016/j.patcog.2010.09.013>. URL <http://www.sciencedirect.com/science/article/pii/S003132031000453X>
26. Raheja, J., Minhas, M., Prashanth, D., Shah, T., Chaudhary, A.: Robust gesture recognition using kinect: A comparison between dtw and hmm. *Optik-International Journal for Light and Electron Optics* **126**(11), 1098–1104 (2015)
27. Salarian, A., Russmann, H., Vingerhoets, F.J., Dehollain, C., Blanc, Y., Burkhard, P.R., Aminian, K.: Gait assessment in Parkinson’s disease: toward an ambulatory system for long-term monitoring. *IEEE Transactions on Biomedical Engineering* **51**(8), 1434–1443 (2004)
28. Srivastava, R., Sinha, P.: Hand movements and gestures characterization using quaternion dynamic time warping technique. *IEEE Sensors Journal* **16**(5), 1333–1341 (2016)
29. Switonski, A., Josinski, H., Zghidi, H., Wojciechowski, K.: Motion data classification on the basis of dynamic time warping with a cloud point distance measure. In: *Proceedings of International Conference of Numerical Analysis and Applied Mathematics*, vol. 1738, p. 180019. AIP Publishing (2016)
30. Wang, H., Fan, Y., Fang, B., Dai, S.: Generalized linear discriminant analysis based on euclidean norm for gait recognition. *International Journal of Machine Learning and Cybernetics* (2016). DOI 10.1007/s13042-016-0540-0
31. Wen, R., Tay, W.L., Nguyen, B.P., Chng, C.B., Chui, C.K.: Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer methods and programs in biomedicine* **116**(2), 68–80 (2014)
32. Zifchock, R.A., Davis, I., Higginson, J., Royer, T.: The symmetry angle: a novel, robust method of quantifying asymmetry. *Gait & Posture* **27**(4), 622–627 (2008)