Multimedia Tools and Applications (2019) 78:32437–32465
https://doi.org/10.1007/s11042-019-07945-y

Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition

Bogdan Kwolek[1] · Agnieszka Michalczuk[2] · Tomasz Krzeszowski[3] · Adam Switonski[4] · Henryk Josinski[4] · Konrad Wojciechowski[2]

Received: 30 July 2018 / Revised: 27 April 2019 / Accepted: 4 July 2019 /
Published online: 3 August 2019
© The Author(s) 2019

Abstract

We introduce synchronized and calibrated multi-view video and motion capture dataset for motion analysis and gait identification. The 3D gait dataset consists of 166 data sequences with 32 people. In 128 data sequences, each of 32 individuals was dressed in his/her clothes, in 24 data sequences, 6 of 32 performers changed clothes, and in 14 data sequences, 7 of the performers had a backpack on his/her back. In a single recording session, every performer walked from right to left, then from left to right, and afterwards on the diagonal from upper-right to bottom-left and from bottom-left to upper-right corner of a rectangular scene. We demonstrate that a baseline algorithm achieves promising results in a challenging scenario, in which gallery/training data were collected in walks perpendicular/facing to the cameras, whereas the probe/testing data were collected in diagonal walks. We compare performances of biometric gait recognition that were achieved on marker-less and marker-based 3D data. We present recognition performances, which were achieved by a convolutional neural network and classic classifiers operating on gait signatures obtained by multilinear principal component analysis. The availability of synchronized multi-view image sequences with 3D locations of body markers creates a number of possibilities for extraction of discriminative gait signatures. The gait data are available at http://bytom.pja.edu.pl/projekty/hm-gpjatk/.

Keywords Gait recognition · Covariate factors · Biometrics · Markerless 3D tracking

✉ Bogdan Kwolek
bkw@agh.edu.pl

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

[2] Research and Development Center of Polish-Japanese Academy of Information Technology, Aleja Legionow 2, Bytom, Poland

[3] Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszow, Poland

[4] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

1 Introduction

Gait is a complex function of body weight, limb lengths, skeletal and bone structures as well as muscular activity. From a biomechanics point of view, human walk consists of synchronized, integrated movements of body joints and hundreds of muscles. Although these movements share the same basic bipedal patterns across all humans, they vary from one individual to another in several aspects, such as their relative timing, rhythmicity, magnitudes and forces involved in producing the movements. Since gait is largely determined by its musculoskeletal structure, it is unique to each individual. Cues such as walking speed, stride length, rhythm, bounce, swagger as well as physical lengths of human limbs all contribute to unique walking styles [41].

The human can identify individual people from movement cues alone. In a study of the motion perception from a psychological point of view, Johansson used Moving Light Display (MLD) and shoved that the relative movements of certain joints in the human body carry information about personal walking styles and dynamics [35]. It turned out that humans can in less than one second identify that MLD patterns correspond to a walking human. Cutting and Kozlowski [22] shoved that friends can be recognized by their gait on the basis of reduced visual stimuli. Barclay et al. [8] showed that the identity of a friend and the person's gender can be determined from the movement of light spots only. They investigated both temporal and spatial factors in gender recognition on the basis of data from point light displays. They showed that in the spatial domain, the shoulder movement for males and hip movement for females are important factors in the recognition. Another research finding is that the duration of dynamic stimulus plays crucial role in gender recognition.

Much research in biomechanics and clinical gait analysis is dedicated to studying the inter-person and intra-person variability of gait, primarily to determining normal vs. pathological ranges of variation. Several measures have been proposed to quantify the degree of gait deviation from normal gait, stratify the severity of pathology and measure changes in gait patterns over time [18]. Although the pioneering research on gait analysis was performed using 2D techniques, such 2D analysis is currently rather rarely used in clinical practice, and 3D methods are now the standard. There are two main reasons for this: parallax errors and perspective errors [37].

In the last decade, gait recognition has been extensively studied as a behavioral biometric technique [17, 19, 20, 70]. Identity recognition on the basis of gait is non-invasive, can be done at a distance, is non-cooperative in nature, and is difficult to disguise. However, the recognition performance of existing algorithms is limited by the influence of a large number of nuisance or so-called covariate factors affecting both appearance and dynamics of the gait, such as viewpoint variations, variations in footwear and clothing, changes of floor surface characteristic, various carrying conditions and so on. Vast of present approaches to gait recognition make use of extracted human body silhouettes in image sequences from a single camera as input [50, 76]. In addition to clothing and carrying variations, the view angle is found to be the most influential covariate factor on recognition performance. Many methods have been elaborated to establish a cross-view mapping between gallery and probe templates [19]. However, their effectiveness is restricted to small variations of view angle. Thus, how to extract informative features, which of them are robust, and how to represent them in a form that is best suited for gait recognition is still an active research topic. Clearly, being a natural representation of human gait perceived by human, 3D gait data conveys more information than 2D data. Moreover, the research results obtained by clinicians [37] demonstrate that 2D data is often inadequate for carrying out an accurate and reliable

biomechanical assessment since many gait features are inherently unique to 3D data. However, until now, only limited research on 3D data-based biometric gait recognition has been done because of restricted availability of synchronized multi-view data with proper camera calibration [50]. Motivated by this, taking into account a forecast formulated at Stanford for an evolution of methods for the capture of human movement [52], promising results that were obtained using marker-less 3D motion tracking [39] and marker-based 3D motion tracking [5, 32] for gait recognition, in this work we introduce a dataset for marker-less 3D motion tracking and 3D gait recognition. Since gait abnormalities are often impossible to detect by eye or with video based systems, we also share data from marker-based moCap, which is synchronized and calibrated with marker-less motion capture system.

2 Background and relevant work

At present, the most common methods for precise capture of three-dimensional human movement usually require the attachment of markers or sensors to the body segments and can only be applied in laboratory environments [47, 64]. In [16], in order to determine disease-specific gait characteristics a marker-based motion capture system (moCap) was employed for investigating and understanding body movement. One of the major limitations of marker-based moCap systems is the need of attaching the markers on the person. The markers are placed manually on the skin with respect to anatomical landmarks. Since the markers are attached to the skin, their position is influenced by the movement of the soft tissue underneath. Benoit et al. [11] investigated the effect of skin movement on the kinematics of the knee joint during gait through comparing skin markers to pin in bone markers. They found that kinematic analysis is burdened with errors resulting from movement of the soft tissue, which should be considered when interpreting kinematic data. The accuracy of the measurements is also influenced by the placement of the markers. Differences can occur between different clinicians applying the markers on the same patient and also between placements made by the same clinician. In general, marker-based moCap systems are costly and thus they are not available in many clinical settings [57].

Recently, increasing interest in using the Kinect sensor for motion capture has emerged, including clinical and scientific analysis of gait [9, 21, 65]. The experimental results achieved in [74] showed that the Kinect sensor can follow the trend of the joint trajectories with considerable error. Gait analysis of healthy subjects using Kinect V2 in single-camera [9] or multi-camera [21] setups showed high accuracy of the motion sensors for estimation of gait speed, stride length, stride time but lower accuracy for other parameters like stride width or speed variability. A recently proposed 3D method [1] employs the spatiotemporal changes in relative distances and angles among different skeletal joints to represent the gait signature. A comparison of traditional marker-based motion capture technologies and Kinect-based technology for gait analysis is presented in [57]. The experiments demonstrated that for the Kinect and Vicon the correlation of hip angular displacement was very low and the error was considerable, whereas the correlation between Kinect and Vicon stride timing was high and the error was fairly small.

Over recent years, the research community has given much interest in gait as a biometric modality [19], primarily due to its non-intrusive nature as well as ease of use in surveillance [54]. As already mentioned, the pioneering works on human motion analysis and gait recognition fall into the category of marker based techniques [8, 22]. Since then, various methods [76] and modalities [10, 26, 77] were proposed to determine one's identity. Usu-

ally, the vision-based methods start with extracting the human silhouette in order to obtain the spatiotemporal data describing the walking person. The extracted silhouettes are then pre-processed, normalized and aligned. Afterwards, various computer vision and machine learning techniques are utilized to extract and model gait signatures, which are finally stored in a dictionary/database. During the authentication a test gait signature is calculated and compared with the dictionary formed in advance.

Human gait analysis and recognition techniques can be divided into main categories, namely model-free and model-based methods. The methods belonging to the first category characterize the entire human body motion using a concise representation without taking into account the underlying structure. They can be in turn divided into two major categories based on the way of preserving temporal information. The methods belonging to the first subcategory consider temporal information in the recognition. Liu et al. [42] utilized a population hidden Markov model (pHMM) defined for a set of individuals to model human walking and generated the dynamics-normalized stance-frames to identify pedestrians. For such probabilistic temporal models, a considerable number of training samples are generally needed to achieve satisfactory performance. The methods from the second subcategory convert an image sequence into a single template. Han and Bhanu [29] proposed the gait energy image (GEI) to improve the accuracy of gait recognition. The disadvantage of template-based methods is that they may lose the temporal information. Several limitations can arise in real scenarios since GEI relies heavily on shape information, which is significantly altered by changes of clothing types and carrying conditions. As demonstrated in many studies, e.g. [75], single-view GEI-based gait recognition performance can drop significantly when the view angle changes. Model-based methods infer gait dynamics directly by modeling the underlying kinematics of human motion. A method proposed in [12] uses a motion-based model and elliptic Fourier descriptors to extract the key features. A recently proposed method [25] combines spatiotemporal and kinematic gait features. The fusion of two different kinds of features gives a comprehensive characterization of gait dynamics, which is less sensitive to variation in walking conditions.

Generally speaking, gait-based person identification is achieved through extracting the image and/or gait signatures using the appearance or shape of the subject undergoing monitoring, and/or the dynamics of the motion itself [23, 42]. What constrains the application of biometric gait recognition is the influence of several of so-called covariate factors, which affects both appearance and dynamics. Those include not only viewpoint, footwear and walking surface, clothing, carried luggage, but also illumination and camera setup. Viewpoint is considered as the most crucial of those covariate factors [76]. Thus, view-invariance to achieve more reliable gait recognition has been studied by several research groups [19, 33, 34, 43, 58, 69]. Clothing and carrying conditions are other important covariate factors that are frequently investigated [2, 24, 53].

In the last few years a number of datasets have been designed to study the effect of covariate factors [50] in gait recognition. A SOTON database with five covariates has been introduced in 2002 [63]. The USF database [61] has been specifically designed to investigate the influence of covariate factors on identity classification. CASIA Gait Database [13] is a newly developed challenge dataset for evaluation of gait recognition techniques. Most current gait recognition approaches performs the recognition on the basis of silhouettes captured in side-view, i.e. when the individual walks in a plane parallel to the camera. When the view angle deviates from the side-view, the gait representation on the silhouette is not so informative, and the recognition performance tends to degrade heavily. In [44, 48], view-invariance has been enhanced using 3D reconstructions.

Three-dimensional approaches to gait recognition are resistant to changes in viewpoint. In general, 3D data provides richer gait information in comparison to 2D data and thus has strong potential to improve the recognition performance. However, only a little research on 3D data-based gait recognition has been done due to the limited availability of synchronized multi-view data with proper calibration parameters [50]. Though the CMU MoBo [28] and the CASIA [75] multi-view databases have been available for a long time, there are no significant results on 3D data, because either the data was recorded on a treadmill and thus does not represent 3D gait or the calibration parameters of the multi-camera system are not available. In the MoBo multi-view dataset [28], six cameras are used to provide full view of the walking person on the treadmill. It comprises 25 individuals, where each one performed four gait types – slow walk, fast walk, ball walk, and inclined walk. Each sequence is recorded at 30 frames per second and is eleven seconds long in duration. An inherent problem associated with gait analysis on the basis of walking on a treadmill is that the human gait is not natural. The main reason for this is that the gait speed is usually constant, and the subjects cannot turn left or right. The INRIA Xmas Motion Acquisition Sequences (IXMAS) database [72], comprises five-view video and 3D body model sequences with eleven activities and ten subjects. The data were collected by five calibrated and synchronized cameras. However, this dataset cannot serve as a benchmark dataset for 3D gait recognition since the gait data has only been registered on closed circle paths. CASIA Dataset B is a multi-view gait database with 124 subjects, where the gait data was captured from eleven views [75]. Three covariate factors, namely view angle, clothing and carrying condition changes are considered separately. However, neither the camera position nor the camera orientation are provided for this frequently employed dataset. A multi-biometric tunnel [62] at the University of Southampton is a constrained environment similar to an airport for capturing multimodal biometrics of walking pedestrians, i.e., gait, face and ear in an unobtrusive way for automatic human identification. For gait recognition, the walking subject was recorded at 30 fps by eight synchronized cameras of resolution 640×480. The face and ear biometrics have been captured by two other high-resolution cameras, which had been placed at the exit of the tunnel. The walls of the tunnel were painted with a non-repeating rectangular lattice of different colors to support automatic camera calibration. However, the dataset has limited applicability for 3D tracking-based gait recognition. The main reason for this is that this dataset has no ground-truth data of 3D joints location. This means that the accuracy of 3D tracking of the joints and 3D motion estimation cannot be easily determined. 3D tracking [68], 3D motion descriptors [36], and 3D reconstructions [59] can provide useful information for gait recognition.

Since motion capture technology became recently more affordable, 3D structure-based gait recognition has attracted more interest from researchers [5, 7, 32, 66]. Recently, a new benchmark data and evaluation protocols for moCap-based gait recognition have been proposed in [6]. As recently shown in [60], marker-less motion capture systems can provide reliable 3D gait kinematics in the sagittal and frontal plane. In [38], a system for view-independent human gait recognition using marker-less 3D human motion capture has been proposed. The accuracy of the 3D motion reconstruction at the selected joints has been determined on the basis of marker-based moCap. In [14], a comparison of marker-less and marker-based motion capture technologies through simultaneous data collection during gait has been presented. A recently conducted study [56] showed that a marker-based and a marker-less systems have similar ranges of variation in the angle from the start of a squat to peak squat in the pelvis and lower limb in a single leg squat.

Table 1 Comparison of major gait recognition databases

Database	Covariate factors	#Subj.	#Seq.	Views	Environment	Synch.	Calib.	moCap	Year
HID-UMD [15]		25	100	side, front	outdoor	no	no	no	2001
CMU MoBo [28]	multi-view recognition, diff. walk cond.	25	100	6 views	indoor, treadmill	yes	yes	no	2001
CASIA A [71]		20	240	3 views	outdoor	no	no	no	2001
USF Human ID [61]	diff. clothing and carrying cond., time (6 months)	122	1870	2 views	outdoor	yes	yes	no	2001
SOTON small [55]	diff. clothing and carrying cond.	11	–	2 views	indoor, green background	no	no	no	2002
SOTON large [55]	multiple purposes	115	2128	2 views	in-outdoor, treadmill	no	no	no	2002
CASIA B [75]	multi-view recognition, diff. clothing and carrying cond.	124	13640	11 views	indoor	yes	yes[a]	no	2005
CASIA C [75]	diff. walk cond.	153	1530	side	outdoor, night, thermal camera	–	no	no	2005
TokyoTech DB [3]	speed variation	30	1602	side	indoor, treadmill	–	no	no	2010
TUM-IITKGP [31]	occlusions, diff. carrying cond.	35	840	side	indoor, occlusions	–	no	no	2011
SOTON Temporal [51]	diff. clothing, time (0, 1, 3, 4, 5, 8, 9 months)	25	2280	12 views	indoor	yes	yes	no	2012
OU-ISIR A [49]	speed variation	34	612	side	indoor, treadmill	–	no	no	2012
OU-ISIR B [49]	clothes variation	68	2746	side	indoor, treadmill	–	no	no	2012
OU-ISIR D [49]	gait fluctuation	185	370	side	indoor, treadmill	–	no	no	2012

Table 1 (continued)

Database	Covariate factors	#Subj.	#Seq.	Views	Environment	Synch.	Calib.	moCap	Year
AVA [43]	multi-view recognition	20	200	6 views	indoor	yes	yes	no	2013
GAID [30]		305	3370	side	indoor	–	–	no	2014
CMU MoCap [7]	marker-based mocap data	54	3843 gait cycles	3D data	indoor	–	–	yes	2017
GPJATK	multi-view recognition, marker-based mocap data, clothes variation, seq. with backpack	32	166	4 views, 3D data	indoor	yes	yes	yes	2017

[a]Only some geometry information on subjects could be reconstructed aided by some calibration equipment (four calibration taps were placed to support reconstruction of geometry information)

3 Databases for gait recognition

Table 1 contains a collective summary of major databases for gait recognition, which were evoked in the relevant literature. As we can observe, only a few datasets provide calibration data and/or were recorded using synchronized cameras. To our knowledge, only one publicly available dataset contains moCap data. However, the benchmark dataset mentioned above contains only moCap data.

Our dataset has been designed for research on vision-based 3D gait recognition. It can also be used for evaluation of the multi-view (where gallery gaits from multiple views are combined to recognize probe gait on a single view) and the cross-view (where probe gait and gallery gait are recorded from two different views) gait recognition algorithms. Unlike related multi-view and cross-view datasets, our dataset has been recorded using synchronized cameras, which allows to reducing the influence of motion artifacts between different views. Other application of our dataset is gait analysis and recognition on the basis of moCap data. For such research the data were stored in commonly used c3d and Acclaim asf/amc data formats. Last but not least, the discussed dataset can be used to evaluate the accuracy of marker-less motion capture algorithms as well as their performance in 3D gait recognition.

4 3D gait dataset

The motion data were captured by 10 moCap cameras and four calibrated and synchronized video cameras. Figure 1 depicts the placement of the cameras of both systems. During the recording session, the actor has been requested to walk on the scene of size 6.5m × 4.2m along a line joining the cameras C2 and C4 as well as along the diagonal of the scene.

In a single recording session, every performer walked from right to left, then from left to right, and afterwards on the diagonal from upper-right to bottom-left and from bottom-left to upper-right corner of the scene. Some performers were also asked to attend in additional

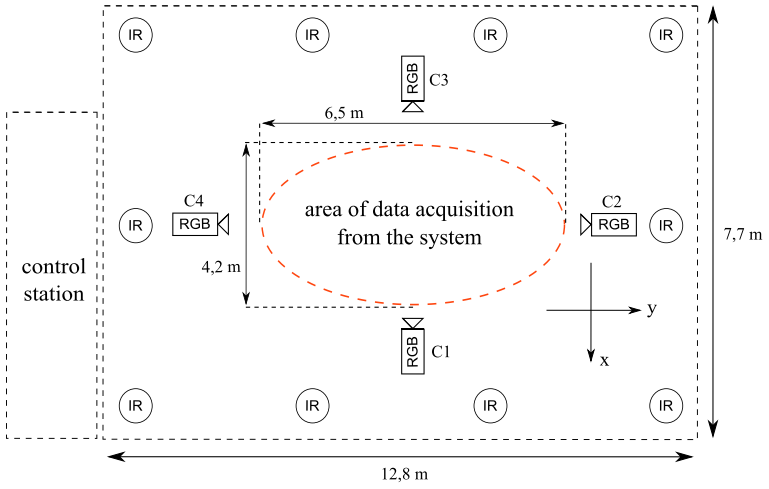

Fig. 1 Camera layout

recording sessions, i.e. after changing into other garment and putting on a backpack. Figure 2 demonstrates sample images from three sessions with the same person. The first row depicts a passage from left to right (the transition from right to left is not shown), the second row depicts a passage on the diagonal from upper-right to bottom-left (the transition from bottom-left to upper-right is not shown), the third and fourth rows illustrate the same passages as above but with different garment, whereas the fifth row illustrates sample images from the transition with the backpack (the remaining passage from the session with the backpack is not shown).

The 3D gait dataset consists of 166 data sequences. The data represents the gait of thirty-two people (10 women and 22 men). In 128 data sequences, each of thirty-two individuals was dressed in his/her clothes, in 24 data sequences, 6 of 32 performers (person #26 – #31) changed clothes, and in 14 data sequences, 7 of the performers attending in the recordings had a backpack on his/her back. Each sequence consists of videos with RGB images of size 960×540, which were recorded by four synchronized and calibrated cameras with 25 frames per second, together with the corresponding moCap data. The moCap data were registered at 100 Hz by Vicon system consisting of ten MX-T40 cameras of resolution 2352×1728 pixels. The synchronization between RGB images and moCap data has been realized using Vicon MX Giganet.

The calibration data of the vision system are stored in the xml data format. The camera model is the well known Tsai model [67], which has been chosen due its frequent use to calibrate multi-camera systems. Every data sequence consists of:

– four videos, see sample images on Fig. 2, which were compressed with Xvid video codec,

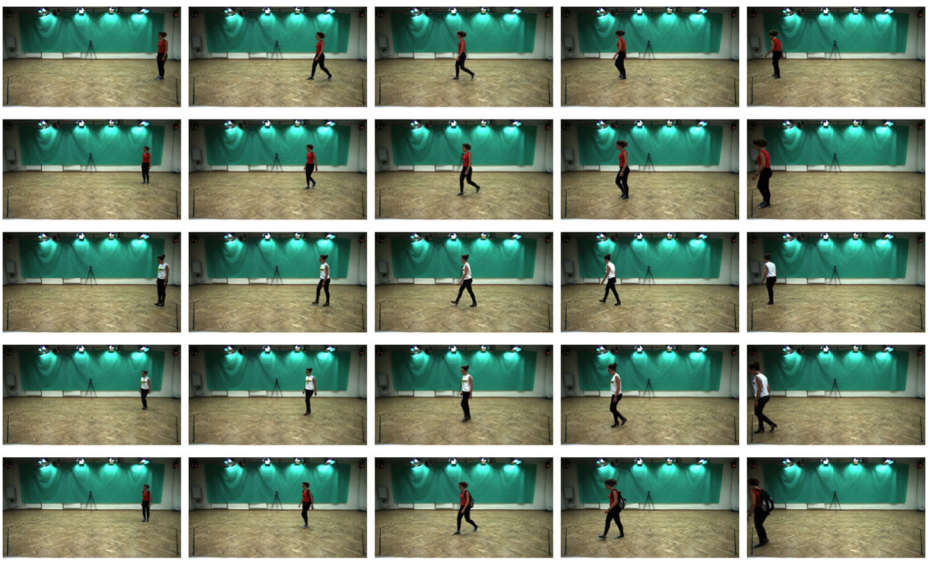

Fig. 2 Sample images from the dataset. First row: a walk from left to right, second row: a walk on diagonal from upper-right to bottom-left, third and four rows: walks in other clothes from left to right and on diagonal from upper-right to bottom-left, respectively, fifth row: a walk on diagonal from upper-right to bottom-left with a backpack

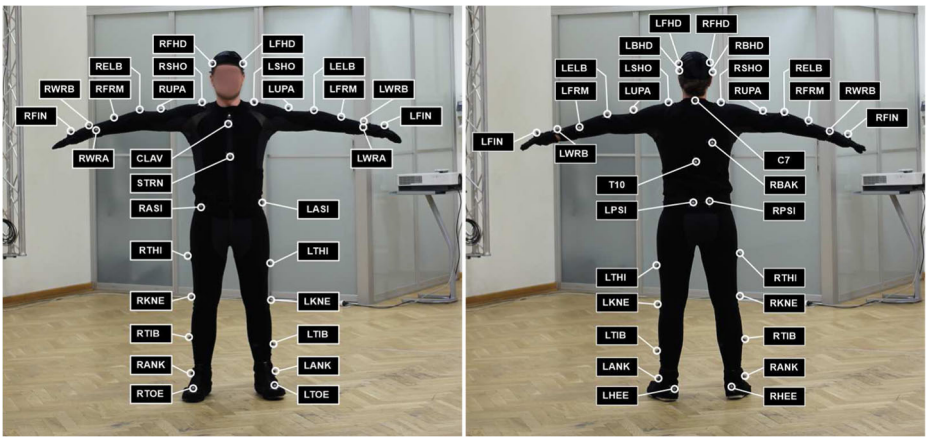

Fig. 3 Placement of the markers

– foreground images that were extracted by background subtraction [78], and which are stored in videos compressed by Xvid codec,
– images stored in the png image format with the extracted edges using Sobel masks,
– moCap data in c3d and Acclaim asf/amc data formats.

The moCap data contains 3D positions of 39 markers, which are placed at the main body joints, see Fig. 3. 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. The marker-less and marker-based systems have a common coordinate system, whose origin is located in the geometric center of the scene, see also Fig. 1.

The GPJATK dataset is freely available at http://bytom.pja.edu.pl/projekty/hm-gpjatk/. The dataset has size 6.5 GB and is stored in .7z format. The names of the data sequences are in the format $pXsY$, where X denotes person name, whereas Y stands for the sequence number. The assumed name convention is as follows:

– s1, s2 - straight walk, s1 from right to left, s2 from left to right, clothing 1
– s3, s4 - diagonal walk, s3 from right to left, s4 from left to right, clothing 1
– s5, s6 - straight walk, s5 from right to left, s6 from left to right, clothing 2
– s7, s8 - diagonal walk, s7 from right to left, s8 from left to right, clothing 2
– s9, s10 - walk with backpack, s9 from right to left, s10 from left to right

All data sequences, except s9 and s10, have corresponding marker-based moCap data.

5 Baseline algorithm for 3d motion tracking

At the beginning of this Section, we outline 3D motion tracking on the basis of image sequences from four calibrated and synchronized cameras. Afterwards, we explain how the accuracy of the system has been calculated using ground-truth from the marker-based moCap system. Finally, we present the accuracy of 3D motion tracking.

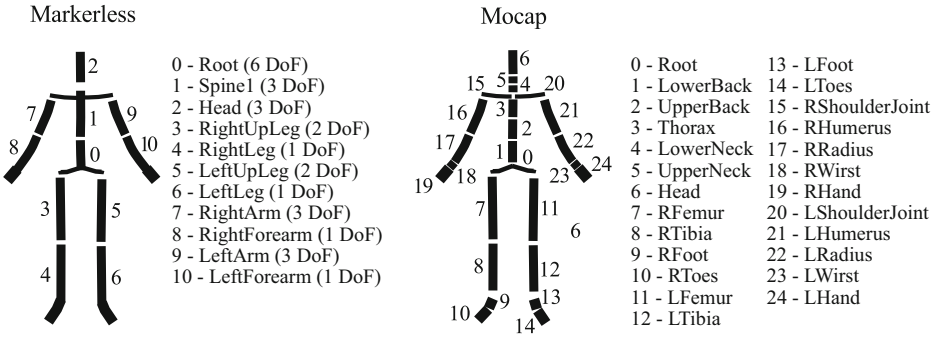

Fig. 4 3D models used by marker-less and marker-based motion capture systems

5.1 3D motion tracking

The human body can be represented by a 3D articulated model formed by 11 rigid segments representing the key parts of the body. 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. 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. Figure 4 (left) illustrates the 3D model utilized in marker-less motion tracking, whereas Fig. 4 (right) depicts the 3D model employed by the marker-based system.

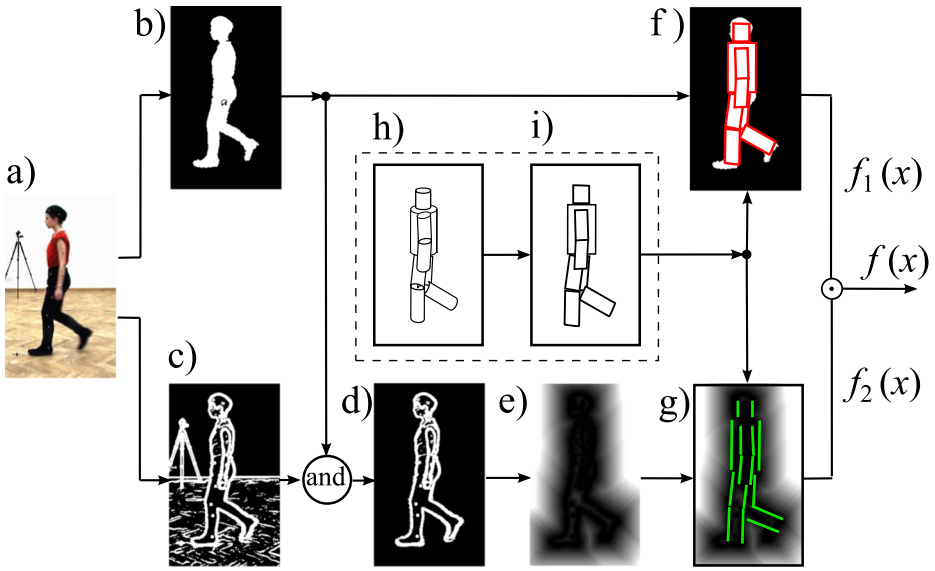

Fig. 5 Calculation of the fitness function. Input image **a**), foreground **b**), gradient magnitude **c**), masked gradient image **d**), edge distance map **e**), 3D model **h**) projected onto image 2D plane **i**) and overlaid on binary image **f**) and edge distance map **g**)

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. 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) [40]. The fitness function expresses the degree of similarity between the real and the estimated human pose. Figure 5 illustrates the algorithm for calculation of the objective function. For single camera c it is determined in the following manner: $f_c(\mathbf{x}) = 1-(f_{1,c}(\mathbf{x})^{w_1} \cdot f_{2,c}(\mathbf{x})^{w_2})$, where $\mathbf{x}$ stands for the state (pose), whereas w denotes weighting coefficients that were determined experimentally. The function $f_1(\mathbf{x})$ reflects the degree of overlap between the extracted body and the projected 3D model into a 2D image. The function $f_2(\mathbf{x})$ reflects the edge distance-based fitness value. A background subtraction algorithm [40] is employed to extract the binary image of the person, see Fig. 5b. The binary image is then utilized as a mask to suppress edges not belonging to the person, see Fig. 5d. The projected model edges are then matched with the image edges using the edge distance map, see Fig. 5g. Moreover, in multi-view tracking, the 3D model is projected and then rendered in each camera's view. The fitness value for all four cameras is determined as follows: $f(\mathbf{x}) = \sum_{c=1}^{4} f_c(\mathbf{x})$.

5.2 Evaluation metrics

Given the placement of the markers on the human body in the marker-based system, see Fig. 3, the virtual positions of the markers were determined for the marker-less system. A set of $M = 39$ markers has been defined for the 3D articulated model that is utilized in the marker-less moCap. Given the estimated 3D human pose by the algorithm discussed in Section 5.1, as well as the position of the virtual markers with respect to the skeleton, the virtual 3D positions of the markers were determined and then utilized in calculating the Euclidean distance between the corresponding 3D position of the markers as follows:

$$dist(\mathbf{p}, \hat{\mathbf{p}}) = \sqrt{(x-\hat{x})^2 + (y-\hat{y})^2 + (z-\hat{z})^2} \tag{1}$$

where the $x-$, $y-$ and $z-$variables represent the X-, Y- and Z-coordinates of the physical markers, whereas the $\hat{x}-$, $\hat{y}-$ and $\hat{z}-$ are estimates of the X-, Y- and Z-coordinates of the virtual markers. On the basis of 3D Euclidean distance between ground truth and estimated joint positions, Root Mean Squared Error (RMSE), which is also referred as the average joint error over all joints has been calculated in the following manner:

$$\text{RMSE}(\mathbf{P}, \hat{\mathbf{P}}) = \frac{1}{MF} \sum_{i=1}^{M} \sum_{j=1}^{F} dist\left(\mathbf{p}_i^{(j)}, \hat{\mathbf{p}}_i^{(j)}\right) \tag{2}$$

where $\mathbf{p}_i^{(j)} \in \mathbf{P}, \hat{\mathbf{p}}_i^{(j)} \in \hat{\mathbf{P}}, \mathbf{P} = \{\mathbf{p}_1^{(1)}, \ldots, \mathbf{p}_M^{(1)}, \ldots, \mathbf{p}_M^{(F)}\}$ represents the set of ground-truth joint positions, $\hat{\mathbf{P}} = \left\{\hat{\mathbf{p}}_1^{(1)}, \ldots, \hat{\mathbf{p}}_M^{(1)}, \ldots, \hat{\mathbf{p}}_M^{(F)}\right\}$ represents the set of estimated joint positions, M represents the total number of joints, whereas F represents the total number of frames.

Table 2 RMSE for $M = 39$ markers in four sample image sequences

#part.	it.	Seq. p1s1 error [mm]	Seq. p1s2 error [mm]	Seq. p28s1 error [mm]	Seq. p28s2 error [mm]
100	10	42.6±21.6	48.4±25.4	58.7±34.1	67.1±33.2
100	20	41.7±23.6	45.4±22.5	52.6±25.0	63.6±30.0
300	10	39.7±20.1	44.8±23.4	52.7±25.4	63.1±27.5
300	20	43.2±21.0	48.8±22.8	40.8±22.8	60.9±24.7

For each marker i the average Euclidean distance $\bar{d}_i$ between the physical and virtual markers has been calculated in the following manner:

$$\bar{d}_i = \frac{1}{F}\sum_{j=1}^{F} dist(\mathbf{p}^j, \hat{\mathbf{p}}^j) \tag{3}$$

For each marker i the standard deviation σ_i has been used to measure the spread of joint errors around the mean error. It has been determined as follows:

$$\sigma_i = \sqrt{\frac{1}{F-1}\sum_{j=1}^{F}\left(dist(\mathbf{p}^j, \hat{\mathbf{p}}^j) - \bar{d}_i\right)^2} \tag{4}$$

The standard deviation for all M markers has been obtained through averaging σ_i values.

5.3 Accuracy of 3D motion tracking

Table 2 presents RMSE errors that were achieved on sequences p1s1, p1s2, p28s1 and p28s2. The results were obtained in ten runs with unlike initializations.

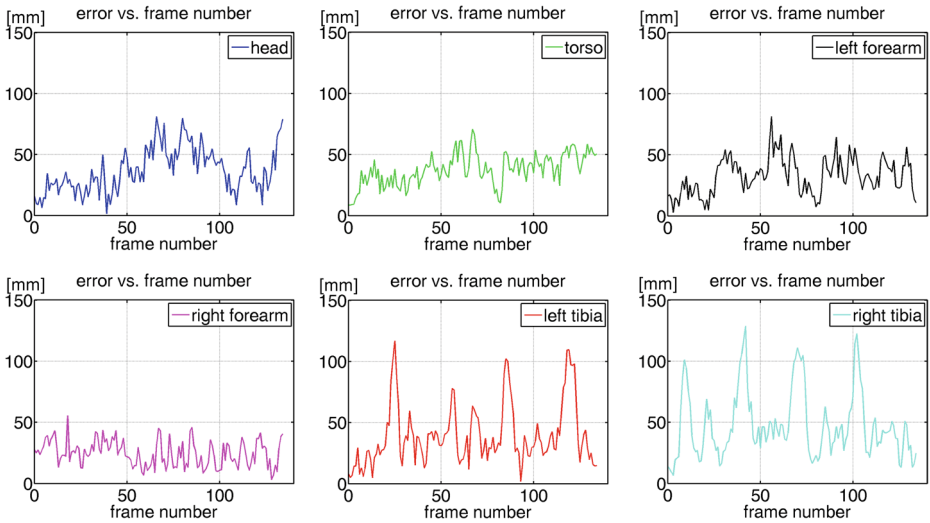

Fig. 6 Tracking errors [mm] versus frame number for p1s1 sequence, achieved by APSO consisting of 300 particles and executing 20 iterations

Figure 6 shows sample plots of Euclidean distance (1) over time for head, torso, left forearm, right forearm, left tibia and right tibia, which were obtained on the sequence p1s1. The average Euclidean distances $\bar{d}_i$ and the standard deviations for mentioned above body parts are equal to: 36.8 ± 18 mm (head), 37.7 ± 13.0 mm (torso), 32.6 ± 14.9 mm (left forearm) 24.5 ± 10.7 mm (right forearm), 39.3 ± 25.0 mm (left tibia) and 47.5 ± 27.4 mm (right tibia). The discussed results were obtained by APSO consisting of 300 particles and executing 20 iterations.

The plots in Fig. 7 depict the distance between ankles for marker-less and marker-based motion capture systems, which were obtained on p1s1 and p1s2 sequences.

6 Performance of baseline 3D gait recognition

At the beginning, we discuss the evaluation methodology. Then, we outline the Multilinear Principal Component Analysis algorithm, afterwards in Section 6.3 we present the evaluation protocol. Afterwards, in Section 6.4 we present the recognition performance, which has been achieved by the baseline algorithm as well as a convolutional neural network. Finally, in Section 6.5 we analyze the performance of individual identification.

6.1 Methodology

The single gait cycle is a basic entity describing the gait during ambulation, which includes the time when one heel strikes the floor to the time at which the same limb contacts the floor again. In our approach, gait is recognized on the basis of a single gait sample, which consists of two strides. Since the number of frames registered in gait samples differs slightly from the average number of frames, the time dimension was chosen to be 32, which roughly equals to the average number of video frames in each gait sample. Having on regard that the marker-based system has four times higher frame rate, the time dimension for moCap data was set to 128. The data extracted by the motion tracking algorithm were stored in ASF/AMC data format. For such a single gait cycle, a third order tensor $32 \times 11 \times 3$ for marker-less data is determined, whereas for marker-based a tensor $128 \times 25 \times 3$ is calculated. The marker-less data was filtered by applying a running mean of length nine samples to the original data. For data from the marker-less system, the second dimension of the tensor is equal to the number

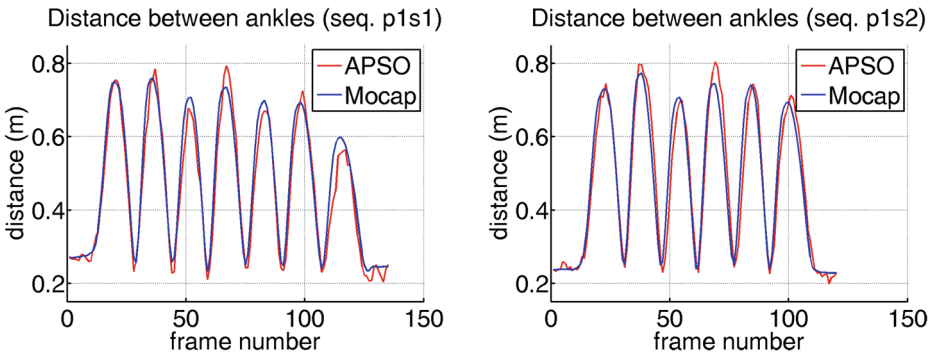

Fig. 7 Distance [m] over time between ankles for marker-less and marker-based motion capture systems

of bones (excluding pelvis), i.e. 10 plus one element in which the distance between ankles and person height are stored. The third mode accounts for three angles, except the eleventh vector that contains distance between ankles and person's height. For data from marker-based system, the second dimension of the tensor is equal to the number of bones, see also Fig. 4, whereas the third mode accounts for three angles. Such a gait signature was then reduced using Multilinear Principal Components Analysis (MPCA) algorithm [45], which is overviewed in Section 6.2.

A benchmark dataset for gait recognition should have two subsets: the gallery and the probe. Gait samples in the gallery set are labeled with person identities and they are utilized in training, while the probe set encompasses the test data, which are gait samples of unknown identities, and which are matched against data from the gallery set. The introduced GPJATK dataset has gallery and probe sets. The gallery set contains a known class label, and a model is learned on this data in order to be generalized to probe set later on. The evaluation methodology is discussed in Section 6.3

6.2 Feature extraction using multilinear principal component analysis

Tensors are denoted by calligraphic letters and their elements are denoted by indexes in brackets. An N^{th}–order tensor is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$. The tensor is addressed by N indexes i_n, where $n = 1, \ldots, N$, and each i_n addresses the n-mode of $\mathcal{A}$. The n-mode product of a tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$ is a tensor $\mathcal{A} \times_n \mathbf{U}$ with elements: $(\mathcal{A} \times_n \mathbf{U})(i_1, \ldots, i_{n-1}, j_n, i_{n+1}, \ldots, i_N) = \sum_{i_n} \mathcal{A}(i_1, \ldots, i_N) \cdot \mathbf{U}(j_n, i_n)$. A rank-1 tensor $\mathcal{A}$ is equal to the outer product of N vectors $\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$, which means that for all values of indexes, $\mathcal{A}(i_1, i_2, \ldots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdot \cdots \cdot \mathbf{u}^{(N)}(i_N)$.

The MPCA operates on third-order gallery gait samples $\{\mathcal{X}_1, \ldots, \mathcal{X}_L \in \mathbb{R}^{I_1 \times I_2 \times I_3}\}$, where L stands the total number of data samples in the gallery subset. The MPCA algorithm seeks for a multilinear transformation $\{\tilde{\mathbf{U}}^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, 2, 3\}$, where $P_n < I_n$ for $n = 1, 2, 3$, which transforms the original gait tensor space $\mathbb{R}^{I_1}\mathbb{R}^{I_2}\mathbb{R}^{I_3}$ into a lower-dimensional tensor subspace $\mathbb{R}^{P_1}\mathbb{R}^{P_2}\mathbb{R}^{P_3}$, in which the feature tensor after the projection is obtained in the following manner: $\mathcal{Y}_l = \mathcal{X}_l \times_1 \tilde{\mathbf{U}}^{(1)^T} \times_2 \tilde{\mathbf{U}}^{(2)^T} \times_3 \tilde{\mathbf{U}}^{(3)^T}$, where $\times_n$ is the n-mode projection, $l = 1, \ldots, L$, and $\tilde{\mathbf{U}}^{(1)} \in \mathbb{R}^{I_1 \times P_1}$ is a projection matrix along the first mode of the tensor, and similarly for $\tilde{\mathbf{U}}^{(2)}$ and $\tilde{\mathbf{U}}^{(3)}$, such that the total tensor scatter $\Psi_{\mathcal{Y}} = \sum_{l=1}^{L} ||\mathcal{Y}_l - \bar{\mathcal{Y}}||_F^2$ is maximized, where $\bar{\mathcal{Y}} = \frac{1}{L}\sum_{l=1}^{L} \mathcal{Y}_l$ denotes the mean of the training samples. The solution to this problem can be obtained through an iterative alternating projection method. The projection matrixes $\{\tilde{\mathbf{U}}^{(n)}, n = 1, 2, 3\}$ can be viewed as $\prod_{n=1}^{3} P_n$ so-called EigenTensorGaits [45]: $\tilde{\mathcal{U}}_{p_1 p_2 p_3} = \tilde{\mathbf{u}}_{p1}^{(1)} \circ \tilde{\mathbf{u}}_{p2}^{(2)} \circ \tilde{\mathbf{u}}_{p3}^{(3)}$, where $\tilde{\mathbf{u}}_{p_n}^{(n)}$ is the p_n^{th} column of $\tilde{\mathbf{U}}^{(n)}$.

In the Q-based method [45], for each n, the first P_n eigenvectors are kept in the n-mode such that $Q^{(1)} = Q^{(2)} =, \ldots, Q^{(N)} = Q$, where Q is a ratio determined as follows: $Q^{(n)} = \sum_{i_n=1}^{P_n} \lambda_{i_n}^{(n)*} / \sum_{i_n=1}^{I_n} \lambda_{i_n}^{(n)*}$, and where $\lambda_{i_n}^{(n)*}$ is the i_nth full-projection n-mode eigenvalue.

Having on regard, the distance between two tensors $\mathcal{A}$ and $\mathcal{B}$, which can be measured by the Frobenius norm $dist(\mathcal{A}, \mathcal{B}) = ||\mathcal{A} - \mathcal{B}||_F$, equals the Euclidean distance between their vectorized representations $vect(\mathcal{A})$, $vect(\mathcal{B})$, the feature tensors extracted by the MPCA were vectorized. They were then utilized in the gait recognition using k-Nearest Neighbors (kNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Multilayer Perceptron (MLP) classifiers. To the best of our knowledge, MPCA and its extended versions were only used in conjunction with kNN classifiers, c.f. [46].

6.3 Evaluation protocol

From 166 video sequences, 414 gait samples were extracted (325 – clothing 1, 58 – clothing 2, 31 – backpack). From every sequence, 2 or 3 gait samples were determined and then employed in the evaluations. The performance of the system has been evaluated as follows:

- clothing 1
  - 10–fold cross–validation on 325 gait samples
  - gallery–probe: 164 gait samples in training set (sequences s1 and s2), 161 gait samples in probe set (sequences s3 and s4)
- clothing 2
  - gallery–probe: training data – 325 gait samples (clothing 1: sequences s1, s2, s3 and s4), test set – 58 gait samples (clothing 2: sequences s5, s6, s7 and s8 with persons p26–p31).
- backpack
  - gallery–probe: training data – 325 gait samples (clothing 1: sequences s1, s2, s3 and s4), test set – 31 gait samples (backpack: sequences s9 and s10 with persons p26–p32).

The Correct Classification Rate (CCR) has been used to quantify the deviations between real outcomes and their predictions. The CCR is a ratio of correctly classified number of subjects to the total number of the subjects in the test subset. The classification rates for rank 2 and rank 3 were also determined. They correspond to percentages of correctly recognized gait instances in the first two and first three indications of a classifier, respectively.

6.4 Evaluation of recognition performance

Gait-based person identification has been performed using Naïve Bayes (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP) and k-Nearest Neighbor (1NN, 3NN and 5NN) classifiers, operating on features extracted by the Multilinear PCA. The classification and the performance evaluation were performed using WEKA package. Sequential Minimal Optimization (SMO) is one of the common optimization algorithms for solving the Quadratic Programming (QP) problem that arises during the training of SVMs. SMO operates by breaking down the dual form of the SVM optimization problem into many smaller optimization problems that are more easily solvable. The C parameter of the SMO algorithm has been determined using cross–validation and grid search.

6.4.1 Cross–validation on gait data

In order to assess how the results of statistical analyses will generalize to independent data set we performed evaluations on the basis of cross–validation. In the ten-fold cross–validation, the 325 gait samples are randomly partitioned into ten equal size sample subsets. Nine of ten sample subsets are used as training data, and the remaining one is retained as the validation data. The cross-validation process is then repeated ten times, with each of ten sample subsets used exactly once as the validation data. Ten results were then averaged to produce a single evaluation. To evaluate the recognition performance we determined the classification accuracies for rank 1, 2 and 3.

Table 3 Correctly classified ratio [%] in 10–fold cross-validation using data from marker-less motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	79.69	84.31	89.85	56.92	50.46	52.62	58.15	52.92	56.92
2	88.92	92.92	95.69	59.08	65.54	66.15	60.00	70.15	70.77
3	91.08	95.38	96.92	60.31	77.23	73.23	61.23	82.15	77.23

Table 3 shows the classification accuracies that were obtained in 10-fold cross-validation on the basis of data from marker-less motion capture. As we can observe, the MLP achieves the best results for ranks 1–3. At this point, it is worth emphasizing the high classification accuracy that was achieved by the baseline algorithm on data from markerless motion capture system, where almost 90% classification accuracy has been achieved for rank 1, and almost 97% classification accuracy has been obtained for rank 3. These promising results are a strong argument for the necessity to introduce this dataset for the pattern recognition community.

The experimental results in Table 4 were obtained in 10-fold cross-validation on the basis of data from marker-based motion capture. As we can observe, the best classification accuracies are obtained by the MLP classifier, which achieves CCR higher than 98% for rank 1 and classification accuracy better than 99% for rank 2. The SMO classifier achieves competitive results in comparison to the MLP, whereas the kNN classifiers, which are frequently used in biometric systems as baseline algorithms, achieve far worse results. The practical advantage of kNN classifiers is that they deliver the closest identities, and thus they permit analysis of the causes of miss classification. The availability of precise motion data for several joints allows identification of most important body segments for gait recognition.

As seen from the above presented results, marker-less motion capture systems are able to deliver useful information for view-invariant gait recognition. They were obtained for $Q = 0.99$, for which the number of attributes in the vectorized tensor representations is equal to 144. Such value of Q gives the best results and it will be utilized in remaining evaluations.

6.4.2 Train/test split - clothing 1

In this subsection, experimental results that were obtained in train/test data split are presented. In the following tables, we present classification accuracies that were obtained on data from marker-less and marker-based motion capture systems.

Table 4 Correctly classified ratio [%] in 10–fold cross-validation using data from marker-based motion capture

			MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	96.62	97.54	98.15	92.62	87.38	84.31	93.85	89.85	88.31
2	96.92	97.54	99.38	92.92	92.62	91.38	94.15	93.23	92.31
3	97.85	97.85	99.38	92.92	94.77	93.23	94.15	95.38	93.85

Table 5 Correctly classified ratio [%] for train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	55.77	67.95	80.13	47.44	44.23	44.23	50.64	46.79	49.36
2	64.74	89.10	84.62	50.00	60.26	59.62	52.56	63.46	64.74
3	73.08	94.23	89.74	50.64	71.79	71.15	53.21	77.56	78.21

Table 5 illustrates the classification accuracy for train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture system. As we can observe, for rank 1 the best classification accuracy is slightly better than 80% and it was achieved by the MLP, whereas for ranks 2 and 3 the best results were achieved by the SMO and they are equal to 89.1% and 94.2%, respectively. In this multi-view scenario, where a straight walk that was perpendicular to a pair of cameras was used in gallery/training, whereas diagonal walk was used in probe/testing, promising results were obtained despite noisy data from marker-less system. Slightly better than 94% classification accuracy for rank 3 is promising since the gait data was recorded at significantly different observation perspectives.

Table 6 shows the classification accuracy for train/test (train - clothing 1, test - clothing 1) split of data from marker-based motion capture system. As we can observe, the classification accuracies are slightly worse in comparison to results demonstrated in Table 4.

6.4.3 Train - clothing 1 / test - clothing 2

This subsection is devoted to the analysis of classification results for the clothing covariates. We present results that were achieved using clothing 1 in gallery/training and clothing 2 in probe/testing data split.

Table 7 presents the classification accuracy for train/test (train - clothing 1, test - clothing 2) split of data from marker-less motion capture. In such a scenario the best classification accuracy was obtained by the MLP classifier. As we can observe, for rank 3 the classification accuracy is better than 96%.

The experimental results in Table 8 illustrate the classification accuracies, which were obtained on data from marker-based system. As expected, the classification accuracy does not change noticeably with respect to previously considered scenarios with data from marker-based system.

Table 6 Correctly classified ratio [%] for train/test (train - clothing 1, test - clothing 1) split of data from marker-based motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	87.18	92.95	92.95	91.03	82.69	77.56	92.31	86.54	82.05
2	89.74	96.15	95.51	91.67	89.10	85.90	92.95	91.67	90.38
3	91.67	98.72	98.08	91.67	91.67	91.03	92.95	94.23	92.31

Table 7 Correctly classified ratio [%] for train/test (train - clothing 1, test - clothing 2) split of data from marker-less motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	56.90	63.79	75.86	37.93	24.14	31.03	29.31	15.52	27.59
2	74.14	84.48	87.93	37.93	36.21	37.93	29.31	31.03	36.21
3	82.76	91.38	96.55	37.93	56.90	41.38	29.31	51.72	44.83

6.4.4 Train - clothing 1 / test - backpack

In this subsection, we present the classification accuracies that were achieved in a scenario in which clothing 1 was used in gallery/training, whereas in probe/testing the backpack data was used. As shown in Table 9, the system achieves 77.4%, 87.1% and 93.55% accuracies for rank 1, 2 and 3, respectively.

6.4.5 End-to-end biometric gait recognition by convolutional neural network

Recently, deep learning methods such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown vast potential for automatically learning features for action recognition and motion analysis [4]. Several CNN-based methods have been applied to biometric gait recognition [70]. Deep learning-based methods usually benefit from big training datasets. Due to limitations of currently available datasets for multi-view [20] and 3D biometric gait recognition [7], limited research has been dedicated to this research topic. In [73] a 3D convolutional neural network has been applied to recognize gait in multi-view scenarios. A method proposed in [27] can be applied to cross-view gait recognition. Heatmaps extracted on the basis of CNN-based pose estimates are used to describe the gait in one frame. An LSTM recurrent neural network is then applied to model gait sequences. Below we show that promising results can be achieved on the proposed dataset in end-to-end biometric gait recognition by a neural network built on 1D convolutions. We evaluated the gait recognition performance in clothing 2 and backpack scenarios.

One of the benefits of using CNNs for sequence classification is that they can learn from the raw time series data directly, and thus do not necessitate a feature extraction for the subsequent classification. The input of a neural network built on 1D convolutions are time series with a predefined length. The input data is 33-dimensonal and can consist of 25 or 32 time steps. The output consists of a probability distribution over number of persons ($C =$

Table 8 Correctly classified ratio [%] for train/test (train - clothing 1, test - clothing 2) split of data from marker-based motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	51.72	82.76	94.83	63.79	53.45	55.17	63.79	53.45	55.17
2	55.17	89.66	98.28	63.79	60.34	62.07	63.79	60.34	62.07
3	58.62	93.10	98.28	63.79	74.14	65.52	63.79	74.14	65.52

Table 9 Correctly classified ratio [%] for train/test (train - clothing 1, test - knapsack) split of data from marker-less motion capture

Rank	NB	SMO	MLP	Euclidean distance			Manhattan distance		
				1 NN	3 NN	5 NN	1 NN	3 NN	5 NN
1	70.97	67.74	77.42	38.71	32.26	41.94	51.61	32.26	58.06
2	74.19	80.65	87.10	38.71	38.71	48.39	51.61	41.94	61.29
3	90.32	93.55	93.55	38.71	61.29	48.39	51.61	77.42	64.52

32) in the dataset, i.e. it is a softmax classifier with C neurons. The CNN model consists of a convolutional block and a fully connected layer. The convolutional block comprises two 1D CNN layers, which are followed by a dropout layer for regularization and then a pooling layer. The length of the 1D convolution window in the first layer is set to three, whereas the number of output filters is set to 256. The length of the 1D convolution window in the second layer is set to three and the number of filters is set to 64. The fraction of units to drop is equal to 0.5, whereas the size of max pooling windows is equal to two. After the pooling, the learned features are flattened to one long vector and pass through a fully connected layer with one hundred neurons. Glorot's uniform initialization, also called Xavier uniform initialization, is used to initialize all parameters undergoing learning. The parameters are learned using the Adam optimizer (with learning rate set to 0.0001, and the exponential decay rates of the first and second moment estimates set to 0.9 and 0.999, respectively) and the categorical cross-entropy as the objective cost function.

The experimental results that were achieved by the convolutional neural network are shown in Table 10. The neural network has been trained on 325 gait samples belonging to Clothing 1 covariate. Comparing results in Tables 7 and 10 for Clothing 1 – Clothing 2 covariate, we can observe that for the rank 1 the CCR is better than CCR achieved by NB, equal to CCR achieved by SMO, worse than CCR obtained by the MLP, and better in comparison to CCRs achieved by k-NNs. On the Clothing 1 – Knapsack covariate, the CCR achieved by the CNN is better in comparison to CCRs achieved by k-NNs, equal to CCR achieved by SMO, and worse in comparison to CCRs achieved both by the NB and the MLP.

As seen in Table 11, on precise motion data acquired by the marker-based system, the convolutional neural network achieved better results in comparison to results achieved by classical classifiers operating on MPCA features, c.f. Table 8. It is worth noting that in order to relate results achieved on the basis of marker-less and marker-based data, the discussed evaluation has been performed on time-series of length equal to 32, i.e. sub -sampled

Table 10 Correctly classified ratio [%] achieved by 1D convolutional neural network on data from marker-less motion capture

Covariate	Rank		
	1	2	3
Clothing 1 - Clothing 2	63.79	75.86	82.76
Clothing 1 - Knapsack	67.74	77.42	87.10

Table 11 Correctly classified ratio [%] achieved by 1D convolutional neural network on data from marker-based motion capture

Covariate	Rank		
	1	2	3
Clothing 1 - Clothing 2	94.83	98.28	100.00

motion data. The temporal convolutional neural network demonstrated that it is capable of extracting characteristic spatiotemporal 3D patterns generated by human motions.

6.5 Individual identification based on gait

In this subsection, we analyze the performance of individual identification on the basis of data from marker-less system. Figure 8 depicts the confusion matrix for train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture, that was determined on the basis of MLP classification results.

The left plot in Fig. 9 depicts the recognition rates for persons p1-p32, which were achieved by the MLP classifier in train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture, see also confusion matrix in Fig. 8. The right plot shows recognition rates for persons p26–p31 in clothing 1/clothing 1, clothing 1/clothing 2 and clothing 1/backpack data splits, respectively.

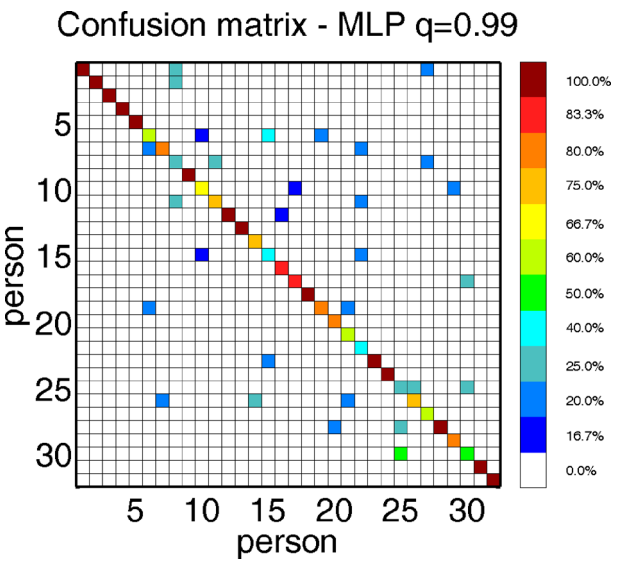

Fig. 8 Confusion matrix for train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture, achieved by the MLP classifier, see also Table 5

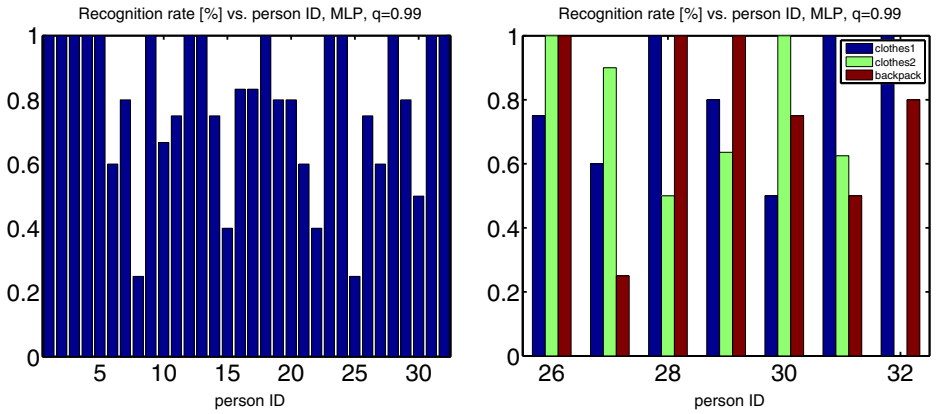

Fig. 9 Recognition rate [%] vs. person ID, (left) recognition rate for train/test (train - clothing 1, test - clothing 1) split of data from marker-less motion capture, achieved by the MLP classifier, corresponding to confusion matrix depicted in Fig. 8, (right)) recognition rate for persons p26–p31 in clothing 1/clothing 1, clothing 1/clothing 2 and clothing 1/backpack data splits

7 Conclusions and discussion

We have introduced a dataset for analysis of 3D motion as well as evaluation of gait recognition algorithms. This is a comprehensive dataset that contains synchronized video from multiple camera views with associated 3D ground truth. We compared performances of biometric gait recognition, which were achieved by algorithms on marker-less and marker-based data. We discussed recognition performances, which were achieved by a convolutional neural network and classic classifiers operating on handcrafted gait signatures. They were extracted on the basis of 3D gait data (third-order tensors) using multilinear principal component analysis. All data are made freely available to the research community. The experimental results obtained by the presented algorithms are promising. Much better results achieved by algorithms operating on marker-based data suggest that the precision of motion estimation has strong impact on performance of biometric gait recognition. The availability of synchronized multi-view image sequences with 3D locations of body joints creates a number of possibilities for extraction of gait signatures with high identification power.

Acknowledgements This work was supported by Polish National Science Center (NCN) under research grants 2014/15/B/ST6/02808 and 2017/27/B/ST6/01743 as well as Statutory Research funds of Institute of Informatics, Silesian University of Technology Poland (BK/204/RAU2/2019).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Ahmed F, Paul PP, Gavrilova ML (2015) DTW–based kernel and rank–level fusion for 3D gait recognition using Kinect. Vis Comput 31(6):915–924. https://doi.org/10.1007/s00371-015-1092-0

2. Al-Tayyan A, Assaleh K, Shanableh T (2017) Decision–level fusion for single–view gait recognition with various carrying and clothing conditions. Image Vis Comput 61:54–69
3. Aqmar MR, Shinoda K, Furui S (2010) Robust gait recognition against speed variation. In: 20th Int. conf. on pattern recognition, pp 2190–2193. https://doi.org/10.1109/ICPR.2010.536
4. Asadi-Aghbolaghi M, Clapes A, Bellantonio M, Escalante HJ, Ponce-Lopez V, Baro X, Guyon I, Kasaei S, Escalera S (2017) A survey on deep learning based approaches for action and gesture recognition in image sequences. In: IEEE Int. conf. on automatic face gesture recognition, pp 476–483. https://doi.org/10.1109/FG.2017.150
5. Balazia M, Plataniotis KN (2017) Human gait recognition from motion capture data in signature poses. IET Biom 6(2):129–137. https://doi.org/10.1049/iet-bmt.2015.0072
6. Balazia M, Sojka P (2017) An evaluation framework and database for MoCap–based gait recognition methods. Springer Int. Publ., Cham, pp 33–47. https://doi.org/10.1007/978-3-319-56414-2_3
7. Balazia M, Sojka P (2018) Gait recognition from motion capture data. ACM Trans Multimedia Comput Commun Appl 14(1s):22:1–22:18. https://doi.org/10.1145/3152124
8. Barclay CD, Cutting JE, Kozlowski LT (1978) Temporal and spatial factors in gait perception that influence gender recognition. Percept Psychophys 23(2):145–152. https://doi.org/10.3758/BF03208295
9. Behrens J, Pfüller C, Mansow-Model S, Otte K, Paul F, Brandt AU (2014) Using perceptive computing in multiple sclerosis – the short maximum speed walk test. J NeuroEngineering and Rehabilitation 11(1):89. https://doi.org/10.1186/1743-0003-11-89
10. Benedek C, Galai B, Nagy B, Janko Z (2018) Lidar-based gait analysis and activity recognition in a 4D surveillance system. IEEE Trans Circuits Syst Video Technol 28(1):101–113. https://doi.org/10.1109/TCSVT.2016.2595331
11. Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, Renstroem P (2006) Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. Gait & Posture 24(2):152–164. https://doi.org/10.1016/j.gaitpost.2005.04.012
12. Bouchrika I, Nixon MS (2007) Model–based feature extraction for gait analysis and recognition. In: Proceedings of the 3rd int. conf. on computer vision/computer graphics collaboration techniques, MIRAGE'07. Springer-Verlag, Berlin, pp 150–160. http://dl.acm.org/citation.cfm?id=1759437.1759452
13. Center for Biometrics and Security Control: Chinese Academy of Sciences (CASIA) gait database. http://www.cbsr.ia.ac.cn/english/Gait
14. Ceseracciu E, Sawacha Z, Cobelli C (2014) Comparison of markerless and marker–based motion capture technologies through simultaneous data collection during gait: Proof of concept. PLoS ONE 9(3):e87640. https://doi.org/10.1016/j.medengphy.2014.07.007
15. Chalidabhongse T, Kruger V, Chellappa R (2001) The UMD database for human identification at a distance. University of Maryland, Tech. rep.
16. Chester VL, Tingley M, Biden EN (2006) A comparison of kinetic gait parameters for 3–13 year olds. Clin Biomech 21(7):726–732. https://doi.org/10.1016/j.clinbiomech.2006.02.007
17. Choi S, Kim J, Kim W, Kim C (2019) Skeleton-based gait recognition via robust frame-level matching. IEEE Trans on Information Forensics and Security. https://doi.org/10.1109/TIFS.2019.2901823
18. Cimolin V, Galli M (2014) Summary measures for clinical gait analysis: A literature review. Gait & Posture 39(4):1005–1010. https://doi.org/10.1016/j.gaitpost.2014.02.001
19. Connie T, Goh KO, Beng Jin Teoh A (2015) A review for gait recognition across view. In: 3rd Int. conf. on information and communication technology (ICoICT), pp 574–577. https://doi.org/10.1109/ICoICT.2015.7231488
20. Connor P, Ross A (2018) Biometric recognition by gait: A survey of modalities and features. Comput Vis Image Underst 167:1–27
21. Coolen DJ, Geerse BH, Roerdink M (2015) Kinematic validation of a multi–Kinect v2 instrumented 10–meter walkway for quantitative gait assessments, vol 10. https://doi.org/10.1371/journal.pone.0139913
22. Cutting JE, Kozlowski LT (1977) Recognizing friends by their walk: Gait perception without familiarity cues. Bull Psychon Soc 9(5):353–356. https://doi.org/10.3758/BF03337021
23. Das Choudhury S, Tjahjadi T (2012) Silhouette–based gait recognition using procrustes shape analysis and elliptic Fourier descriptors. Pattern Recogn 45(9):3414–3426. https://doi.org/10.1016/j.patcog.2012.02.032
24. Das Choudhury S, Tjahjadi T (2016) Clothing and carrying condition invariant gait recognition based on rotation forest. Pattern Recogn Lett 80(C):1–7. https://doi.org/10.1016/j.patrec.2016.05.009
25. Deng M, Wang C, Cheng F, Zeng W (2017) Fusion of spatial–temporal and kinematic features for gait recognition with deterministic learning. Pattern Recogn 67:186–200. https://doi.org/10.1016/j.patcog.2017.02.014 http://www.sciencedirect.com/science/article/pii/S0031320317300560
26. Devanne M, Wannous H, Daoudi M, Berretti S, Bimbo AD, Pala P (2016) Learning shape variations of motion trajectories for gait analysis. In: 23rd int. conf. on pattern recognition (ICPR), pp 895–900. https://doi.org/10.1109/ICPR.2016.7899749

27. Feng Y, Li Y, Luo J (2016) Learning effective gait features using LSTM. In: 23rd int. conf. on pattern recognition (ICPR), pp 325–330. https://doi.org/10.1109/ICPR.2016.7899654
28. Gross R, Shi J (2001) The CMU motion of body (MoBo) database. Tech. Rep. CMU-RI-TR-01-18, Pittsburgh PA
29. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322. https://doi.org/10.1109/TPAMI.2006.38
30. Hofmann M, Geiger J, Bachmann S, Schuller B, Rigoll G (2014) The TUM gait from audio, image and depth (GAID) database. J Vis Comun Image Represent 25(1):195–206
31. Hofmann M, Sural S, Rigoll G (2011) Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In: Proc. of int. conf. on computer graphics, visualization and computer vision, Plzen, Czech Republic, pp pp 99–104
32. Hosni N, Drira H, Chaieb F, Ben Amor B (2018) 3D Gait recognition based on functional PCA on Kendall's shape space. In: Int. Conf. on Pattern Rec. (ICPR) pp 2130–2135. Beijing, China
33. Hu H (2013) Enhanced Gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition. IEEE Trans Circuits Syst Video Technol 23(7):1274–1286. https://doi.org/10.1109/TCSVT.2013.2242640
34. Isaac ER, Elias S, Rajagopalan S, Easwarakumar KS (2017) View–invariant gait recognition through genetic template segmentation. IEEE Signal Process Lett 24(8):1188–1192. https://doi.org/10.1109/LSP.2017.2715179
35. Johansson G (1973) Visual perception of biological motion and a model for its analysis. Percept Psychophys 14(2):201–211. https://doi.org/10.3758/BF03212378
36. Khokhlova M, Migniot C, Dipanda A (2016) 3D visual–based human motion descriptors: A review. In: 12th Int. Conf. on Signal–Image Technology Internet–Based Systems (SITIS), pp 564–572. https://doi.org/10.1109/SITIS.2016.95
37. Kirtley C (2006) Clinical Gait analysis. Theory and Practice. Churchill Livingstone, Edinburgh
38. Krzeszowski T, Kwolek B, Michalczuk A, Świtoński A, Josiński H (2012) View independent human gait recognition using markerless 3D human motion capture. In: Int. conf. on computer vision and graphics, lecture notes in computer science, vol 7594. Springer-Verlag, Inc., New York, pp 491–500. https://doi.org/10.1007/978-3-642-33564-8_59
39. Kwolek B, Krzeszowski T, Michalczuk A, Josinski H (2014) 3D gait recognition using spatio-temporal motion descriptors. In: 6th Asian conf. on intelligent information and database systems, lecture notes in computer science, vol. 8398, pp 595–604. Springer Int. Publ. https://doi.org/10.1007/978-3-319-05458-2_61
40. Kwolek B, Krzeszowski T, Wojciechowski K (2011) Swarm intelligence based searching schemes for articulated 3D body motion tracking. In: Int. conf. on advanced concepts for intell. vision systems, lecture notes in computer science, vol 6915, pp 115–126. Springer
41. Levine D, Richards J, W Whittle M (2012) Whittle's gait analysis. In: Whittle's Gait analysis, fifth edn. Elsevier Health Sciences
42. Liu Z, Sarkar S (2006) Improved gait recognition by gait dynamics normalization. IEEE Trans Pattern Anal Mach Intell 28(6):863–876. https://doi.org/10.1109/TPAMI.2006.122
43. López-Fernández D, Madrid-Cuevas F, Carmona-Poyato Á, Marín-Jimenez MJ, Muñoz-Salinas R (2014) The AVA multi–view dataset for gait recognition. In: Activity monitoring by multiple distributed sensing, lecture notes in computer science, pp 26–39. Springer Int. Publ. https://doi.org/10.1007/978-3-319-13323-2_3
44. López-Fernández D, Madrid-Cuevas F, Carmona-Poyato A, Marin-Jimenez M, Munoz-Salinas R, Medina-Carnicer R (2016) Viewpoint–independent gait recognition through morphological descriptions of 3D human reconstructions. Image Vision Comput 48(C):1–13. https://doi.org/10.1016/j.imavis.2016.01.003
45. Lu H, Plataniotis KN, Venetsanopoulos AN (2008) MPCA: Multilinear principal component analysis of tensor objects. Trans Neur Netw 19(1):18–39. https://doi.org/10.1109/TNN.2007.901277
46. Lu H, Plataniotis KN, Venetsanopoulos AN (2011) A survey of multilinear subspace learning for tensor data. Pattern Recogn 44(7):1540–1551. https://doi.org/10.1016/j.patcog.2011.01.004
47. Lu TW (2012) Biomechanics of human movement and its clinical applications. Kaohsiung J Med Sci 28(2):S13–S25. https://doi.org/10.1016/j.kjms.2011.08.004. http://www.sciencedirect.com/science/article/pii/S1607551X11001835
48. Luo J, Tang J, Tjahjadi T, Xiao X (2016) Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. Pattern Recogn 60(C):361–377. https://doi.org/10.1016/j.patcog.2016.05.030
49. Makihara Y, Mannami H, Tsuji A, Hossain M, Sugiura K, Mori A, Yagi Y (2012) The OU-ISIR gait database comprising the treadmill dataset. IPSJ Trans Comput Vis Appl 4:53–62

50. Makihara Y, Matovski DS, Nixon MS, Carter JN, Yagi Y (2015) Gait Recognition: Databases, representations and applications. https://doi.org/10.1002/047134608X.W8261
51. Matovski DS, Nixon MS, Mahmoodi S, Carter JN (2012) The effect of time on gait recognition performance. IEEE Trans Inf Forensics Secur 7(2):543–552. https://doi.org/10.1109/TIFS.2011.2176118
52. Mündermann L., Corazza S, Andriacchi TP (2006) The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. J. of NeuroEngineering and Rehabilitation 3:6. https://doi.org/10.1186/1743-0003-3-6
53. Nandy A, Chakraborty R, Chakraborty P (2016) Cloth invariant gait recognition using pooled segmented statistical features. Neurocomputing 191:117–140. https://doi.org/10.1016/j.neucom.2016.01.002. http://www.sciencedirect.com/science/article/pii/S0925231216000497
54. Neves J, Narducci F, Barra S, Proença H (2016) Biometric recognition in surveillance scenarios: A survey. Artif Intell Rev 46(4):515–541. https://doi.org/10.1007/s10462-016-9474-x
55. Nixon J, Carter J, Grant M (2001) Experimental plan for automatic gait recognition. Tech. rep., Southampton
56. Perrott MA, Pizzari T, Cook J, McClelland JA (2017) Comparison of lower limb and trunk kinematics between markerless and marker–based motion capture systems. Gait & Posture 52:57–61. https://doi.org/10.1016/j.gaitpost.2016.10.020 http://www.sciencedirect.com/science/article/pii/S0966636216306233
57. Pfister A, West AM, Bronner S, Noah JA (2014) Comparative abilities of microsoft Kinect and Vicon 3D motion capture for gait analysis. J Med Eng Technol 38(5):274–280. https://doi.org/10.3109/03091902.2014.909540
58. Portillo-Portillo J, Leyva R, Sanchez V, Sanchez-Perez G, Perez-Meana H, Olivares-Mercado J, Toscano-Medina K, Nakano-Miyatake M (2017) Cross view gait recognition using joint–direct linear discriminant analysis . Sensors 17(1):6. https://doi.org/10.3390/s17010006. http://www.mdpi.com/1424-8220/17/1/6
59. Sandau M, Heimbuerger RV, Jensen KE, Moeslund TB, Aanaes H, Alkjaer T, Simonsen EB (2016) Reliable gait recognition using 3D reconstructions and random forests – An anthropometric approach. J Forensic Sci 61(3):637–648. https://doi.org/10.1111/1556-4029.13015
60. Sandau M, Koblauch H, Moeslund T, Aanaes H, Alkjaer T, Simonsen E (2014) Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. Med Eng Phys 36(9):1168–1175. https://doi.org/10.1016/j.medengphy.2014.07.007
61. Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The HumanID gait challenge problem: Data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell 27(2). https://doi.org/10.1109/TPAMI.2005.39
62. Seely RD, Samangooei S, Lee M, Carter JN, Nixon MS (2008) The University of Southampton multi–biometric tunnel and introducing a novel 3D gait dataset. In: IEEE sec. int. conf. on biometrics: theory, applications and systems, pp 1–6. https://doi.org/10.1109/BTAS.2008.4699353
63. Shutler JD, Grant MG, Nixon MS, Carter J (2004) On a large sequence–based human gait database. Springer, Berlin, pp 339–346. https://doi.org/10.1007/978-3-540-45240-9_46
64. Stoddart AJ, Mrazek P, Ewins D, Hynd D (1999) Marker based motion capture in biomedical applications. In: IEE Colloquium on motion analysis and tracking, pp 4/1–4/5. https://doi.org/10.1049/ic:19990574
65. Sun J, Wang Y, Li J, Wan W, Cheng D, Zhang H (2018) View-invariant gait recognition based on Kinect skeleton feature. Multimedia Tools Appl 77(19):24,909–24,935. https://doi.org/10.1007/s11042-018-5722-1
66. Świtoński A, Polański A, Wojciechowski K (2011) Human identification based on Gait paths. Springer, Berlin, pp 531–542. https://doi.org/10.1007/978-3-642-23687-7_48
67. Tsai R (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J Robot Autom 3(4):323–344. https://doi.org/10.1109/JRA.1987.1087109
68. Urtasun R (2004) Fua, P.: 3D tracking for gait characterization and recognition. In: 6th IEEE int. conf. on automatic face and gesture recognition, pp 17–22. https://doi.org/10.1109/AFGR.2004.1301503
69. Verlekar T, Correia P, Soares L (2016) View–invariant gait recognition exploiting spatio–temporal information and a dissimilarity metric. In: Int. conf. of the biometrics special interest group (BIOSIG), pp 1–6. https://doi.org/10.1109/BIOSIG.2016.7736937
70. Wan C, Wang L, Phoha VV (2018) A survey on gait recognition. ACM Comput Surv 51(5):89:1–89:35. https://doi.org/10.1145/3230633
71. Wang L, Tan T, Ning H, Hu W (2003) Silhouette analysis-based gait recognition for human identification. IEEE Trans Pattern Anal Mach Intell 25(12):1505–1518. https://doi.org/10.1109/TPAMI.2003.1251144

72. Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comput Vis Image Underst 104(2):249–257. https://doi.org/10.1016/j.cviu.2006.07.013
73. Wolf T, Babaee M, Rigoll G (2016) Multi-view gait recognition using 3d convolutional neural networks. In: IEEE int. conf. on image processing (ICIP), pp 4165–4169. https://doi.org/10.1109/ICIP.2016.7533144
74. Xu X, McGorry R, Chou LS, Hua Lin J, Chi Chang C (2015) Accuracy of the Microsoft Kinect for measuring gait parameters during treadmill walking. Gait & Posture 42(2):145–151. https://doi.org/10.1016/j.gaitpost.2015.05.002. http://www.sciencedirect.com/science/article/pii/S0966636215004622
75. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th int. conf. on pattern recognition (ICPR'06), vol. 4, pp 441–444. https://doi.org/10.1109/ICPR.2006.67
76. Zhang Z, Hu M, Wang Y (2011) A survey of advances in biometric Gait recognition, pp 150–158. Springer. https://doi.org/10.1007/978-3-642-25449-9_19
77. Zhao G, Liu G, Li H, Pietikainen M (2006) 3D gait recognition using multiple cameras. In: 7th int. conf. on automatic face and gesture rec., pp 529–534. https://doi.org/10.1109/FGR.2006.2
78. Zivkovic Z, van der Heijden F (2006) Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recogn Lett 27(7):773–780. https://doi.org/10.1016/j.patrec.2005.11.005

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Bogdan Kwolek received the Ph.D. degree in Computer Science from AGH University of Science and Technology in Kraków, Poland and D.Sc. from at AGH University of Science and Technology in Kraków. His research interests concern computer vision, machine learning and pattern recognition. He was awarded DAAD Scholarships to Bielefeld University and Technische Universitt München, a scholarship from the French government to INRIA, and a scholarship from Polish Government to Stanford University. He worked as an associate professor at Nagoya Institute of Technology in Japan. He is an associate professor of computer science at AGH University of Science and Technology in Krakow.

Agnieszka Michalczuk received the M.Sc. degree in Computer Science from the Silesian University of Technology in Gliwice, Poland. She is an assistant at the Silesian University of Technology, Institute of Informatics and a PhD candidate at this University. Her current research and professional projects concern AR/VR application in various industries as well as machine learning, data mining and computer vision with special interest in human identification based on gait.

Tomasz Krzeszowski received the M.Sc. (Eng.) degree in Computer Science from the Rzeszow University of Technology in 2009. In 2013, he received his Ph.D. in Computer Science at the Silesian University of Technology. He is currently an assistant professor in the Faculty of Electrical and Computer Engineering at the Rzeszów University of Technology. His areas of interest lie in computer vision, human motion tracking, machine learning, and particle swarm optimization algorithms.

Adam Switonski received the Ph.D. in Computer Science from the Silesian University of Technology, Gliwice, Poland and the D.Sc. from the Polish-Japanese Academy of Information Technology, Warsaw, Poland. His scientific activity is related to analysis and classification of multimodal motion data, processing and recognition of multi- and hyperspectral images and computer vision. His latest research concerns assessment of gait abnormalities, human gait identification and application of multispectral imaging in retinal and photodynamic diagnoses.

Henryk Josiski received the M.Sc. degree in Computer Science and the 2 Ph.D. degree in Computer Science from the Silesian University of Technology in Gliwice, Poland. His scientific activity focuses on modeling of dynamical systems, data exploration, biometrics, computer vision, optimization algorithms, artificial intelligence, and databases.

Konrad Wojciechowski received the M.Sc. Diploma in Electrical Engineering from the Academy of Mining and Metallurgy, Kraków, Poland in 1967 and the Ph.D., D.Sc. in Control Theory from the Silesian University of Technology in Gliwice, Poland, in 1976 and 1991, respectively. He received the title of professor in 1999. His area of scientific activity is linear and nonlinear control theory, neural nets, image processing and pattern recognition, computer vision, computer graphics, animation and games. He has published 180 papers in refereed journals and conference proceedings on control theory, image processing and computer vision. Currently, prof. Wojciechowski is Director of the Centre for Research and Development of the Polish-Japanese Academy of Information Technology in Bytom, Poland.