IET Biometrics – ISSN 2047-4938 – www.ietdl.org
Published on 1st November 2016
DOI: 10.1049/iet-bmt.2015.0072

Human Gait Recognition from Motion Capture Data in Signature Poses

Michal Balazia · Konstantinos N. Plataniotis

Abstract Most contribution to the field of structure-based human gait recognition has been done through design of extraordinary gait features. Many research groups that address this topic introduce a unique combination of gait features, select a couple of well-known object classifiers, and test some variations of their methods on their custom Kinect databases. For a practical system, it is not necessary to invent an ideal gait feature – there have been many good geometric features designed – but to smartly process the data there are at our disposal. This work proposes a gait recognition method without design of novel gait features; instead, we suggest an effective and highly efficient way of processing known types of features. Our method extracts a couple of joint angles from two signature poses within a gait cycle to form a gait pattern descriptor, and classifies the query subject by the baseline 1-NN classifier. Not only are these poses distinctive enough, they also rarely accommodate motion irregularities that would result in confusion of identities. We experimentally demonstrate that our gait recognition method outperforms other relevant methods in terms of recognition rate and computational complexity. Evaluations were performed on an experimental database that precisely simulates street-level video surveillance environment.

Keywords Human gait recognition · Motion capture data · Signature poses · Features extraction · Gait pattern descriptor · Identity classification

1 Introduction

People clearly understand the importance of security monitoring and control for the purposes of national defense and public safety. Video surveillance technology records video footage for potential future recognition of suspicious individuals and activities in public areas. Many streets and airports already have surveillance cameras installed, but these require intelligent approaches to human recognition. A useful early-warning system would analyze the collected video footage and release an alert before an adverse event takes place. Triggered by detection of an abnormal behavior, the system could instantly identify all scene participants, rapidly investigate their previous activities, and launch tracking the suspects.

Public places of heavy flow, such as streets or airports, are usually equipped with video cameras. The following (non-exhaustive) list outlines a typical video surveillance environment:

1. Data are captured by a system of multiple cameras or a depth camera.
2. Tracking space is rather large – at least $25\,m^2$.
3. People walk various directions, speeds, and often in crowds. They wear various clothes and shoes and often carry large objects.
4. The system utilizes a database to store hundreds of subject identities and thousands of biometric samples.
5. People are encountered repeatedly, thus contribute with multiple biometric samples.
6. Identification has to be performed in real time, that is, in a few seconds.
7. Data acquired from video is all that one can work with. No pre-calculation or training is allowed.
8. Tracking and identification have to be automatic. Any intervention of an operator is slow and costly.

The goal of this work is to design a method for recognizing individuals from videos by their gait pattern. From the surveillance perspective, gait pattern biometrics is appealing because of its possibility of being performed at a distance and without body-invasive equipment or subject cooperation. This enables sample acquisitions possible even without a subject's consent. As the data are collected with high participation rate and surveilled subjects are not expected to claim their identities, the method is preferably employed for human identification rather than for authentication.

The introduced method uses a motion capture technology that provides video clips of walking individuals containing structural motion data. The format keeps an overall structure of the human body and holds estimated 3D positions of major anatomical landmarks as the person moves. These so-called *motion capture data* (MoCap) can be collected online by a system of multiple cameras (Vicon[1]) or a depth camera (Microsoft Kinect[2]). To visualize motion capture data (see Figure 1), a simplified stick figure representing the human skeleton (a graph of joints connected by bones) can be recovered from the values of body point spatial coordinates. With recent rapid improvement in MoCap sensor accuracy, we believe in an affordable MoCap technology that can be installed in the streets and identify people from MoCap data.

M. Balazia[1,2] (✉), *IEEE Student Member* · K.N. Plataniotis[1], *IEEE Member*
[1] The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada
[2] Faculty of Informatics, Masaryk University, Brno, Czech Republic
E-mail: xbalazia@mail.muni.cz

[1] `http://www.vicon.com/products/camera-systems`
[2] `https://dev.windows.com/en-us/kinect`

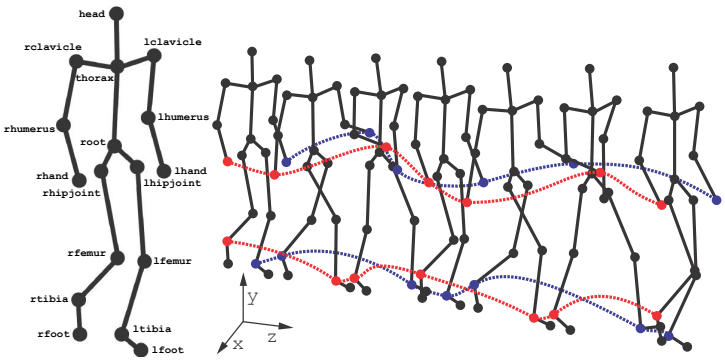

Fig. 1 Motion capture data. Skeleton is represented by a stick figure of 31 joints (only 17 are drawn here). Seven selected video frames of a walk sequence contain 3D coordinates of each joint in time. Red and blue lines track trajectories of hands and feet. [1]

People spotted in our tracking space do not walk all the time; on the contrary, they perform various activities. Recognizing people by gait requires processing video segments where they are actually walking. Clean gait cycles need to be first filtered out from the video sequences of general motion. Some methods focus on detecting gait cycles directly [2,3] or we can use a general action recognition method [4,5,6,7,8] that only need an example of a gait cycle to search general motion sequences.

Having a query motion clip where a person performs an action classified as a gait cycle, the system can proceed to the identification phase (see Figure 2). Gait sample of each recorded walker in a raw MoCap form is pre-processed to contain the discriminative gait information. A collection of extracted gait features, such as feet distance or elbow angle, builds a gait pattern descriptor. Descriptor, also referred to as gait pattern, serves as a walker's signature. Associated with the walker's identity, the descriptors are stored in a central database. To identify someone by gait means to classify an identity for their gait pattern that is unknown at the moment. The classifier composes a query to search this database for a set of similar gait patterns, retrieving a collection of candidate identities and reporting the most likely one as the classified identity. Here, the similarity of two gait patterns is expressed in a single number computed by a similarity/distance function of their descriptors.

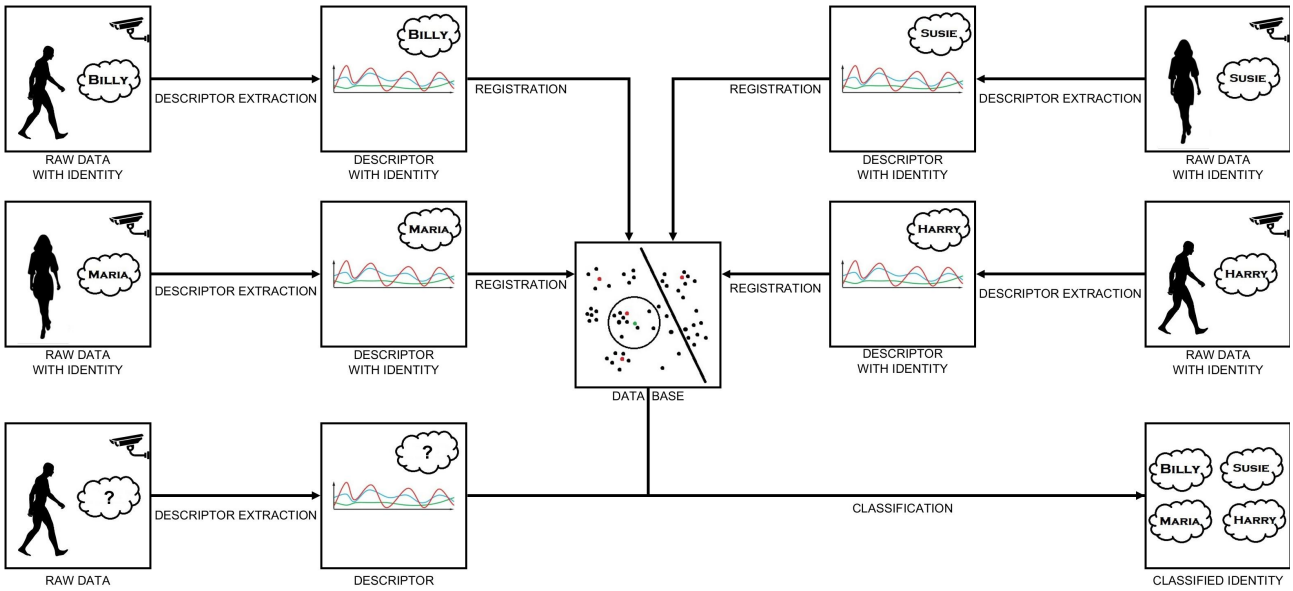

Fig. 2 A general pipeline for human identification by gait.

1.1 Related Methods

A scheme of extracted gait features and a classifier defines a gait recognition method. What follows is a to-date overview of gait recognition methods from MoCap data.

Ahmed et al. [9] present a method for gait recognition from horizontal and vertical distances of selected joint pairs. Temporally normalized within one gait cycle, the descriptors are classified by 1-NN queries with the CityBlock (Manhattan) distance. Their results reach 92% on their own database of 20 walking subjects with 10 walk cycles per subject on average.

Andersson et al. [10] calculate lower joint (hips, knees and ankles) angles as gait kinematic parameters. Mean and standard deviation in the resulting signals form the gait pattern descriptor. Walker identity is classified by the baseline 1-NN with L_1 metric. On their Kinect database of 160 individuals of about 5 walks each, the method achieves (10-fold cross-validation) 80% recognition rate.

Ball et al. [11] select mean, standard deviation and maximum of the signals of lower joint (hips, knees and ankles) angles to classify identities by K-means clustering algorithm. The accuracy of 43.6% is reached on their Kinect database of 71 walk samples of 4 people.

Derlatka et al. [12] propose a bimodal gait recognition system. Along with traditional skeleton parameters acquired by Kinect, they utilize so-called Ground Reaction Forces (GRF), which are dynamic gait measures that depend on the subject's weight, cadence, velocity and footwear. GRF are measured by Kistler platforms. To recognize a walker, they use a DTW-based 5-NN classifier with majority voting. This approach is not implemented as our approach does not support GRF input data.

Dikovski et al. [13] evaluate a broad spectrum of geometric features and classifiers. Static body parameters, joint angles and inter-joint distances aggregated within a gait cycle, along with various statistics, were combined to construct 7 different gait pattern descriptors. MultiLayer Perceptron (MLP), Support Vector Machine (SVM) with sequential minimal optimization, and J48 algorithms were used for classification. On their database of 15 walkers where each contributed with about 2 full gait cycles, the 89.8% recognition rate (10-fold cross-validation by MLP) was reached using a descriptor of 71 geometric features.

Gavrilova et al. [14] combine joint relative distances (JRD) and angles (JRA) and classify the identities by a DTW-based 3-NN classifier for each of the JRD and the JRA separately. The final decision is formed by majority voting of the 6 candidates. The method achieves the recognition rate of 92.1% (3-fold cross-validation) on their database of 60 samples (3 samples of each of 20 registered participants).

Jiang et al. [15] use fusion of static and dynamic gait features and the 1-NN classifier. Evaluations on a Kinect database of 10 subjects of 5 gait cycles reach 82%.

Krzeszowski et al. [16] extract many pose attributes, such as bone angles (bone rotations around three axes), inter-joint distances, and the person's height. Their DTW-based baseline 1-NN classifier uses a distance function that measures differences in Euler angles, step length and height. Experiments were carried out on their own structural database of 22 walking subjects and 230 gait cycles in total, achieving the perfect 100% rate.

Kumar et al. [17] propose a gait recognition algorithm which uses trajectory covariance of body points. The matrix of covariances between body point trajectories forms the gait model. Subject's identity is classified by computing the minimum dissimilarity measure between the covariance matrices if their gait. Recognition rate of 97.5% (2-fold cross-validation by 1-NN) is achieved for their dataset of 20 walking subjects, 10 gait samples each.

Kwolek et al. [18] uses the same database and descriptors as the previous author. Each gait cycle is linearly normalized in length to 30 frames and classified by MLP. Estimated with 10-fold cross-validation, the method achieves 96.1% recognition rate.

Preis et al. [19] define thirteen biometric features, eleven of them static body parameters and the other two dynamic parameters. Based on test data from 9 persons, the recognition rate of 91% was achieved by the Naïve Bayes classifier (7-fold cross-validation) on only four static skeleton parameters. We implemented this method with both static and dynamic features.

Sedmidubsky et al. [20] extract various inter-joint distances from linearly normalized gait cycles. The baseline 1-NN classifier uses a DTW-based similarity function. Evaluated on the CMU MoCap database of 131 gait cycles that belong to 24 walking subjects, the method achieved the recognition rate of 96%.

Sinha et al. [21] combine many gait features: areas of upper and lower body, various inter-joint distances and all features introduced by Ball et al. [11] and Preis et al. [19]. Classified by MLP, they reach 86% recognition rate (2-fold cross-validation) on their database of 5 individuals and 700 gait cycles in total.

Despite designed for an environment that does match video surveillance environment (e.g. Kinect-based recognition), the idea behind each reviewed method can be potentially successful. We have therefore implemented each of them for experimental comparisons in Section 3.2. However, the research groups used databases for evaluating their methods that did not contain data in a form that typical video surveillance systems can provide. The following paragraphs discuss some important aspects in which the experimental databases fail to demonstrate the quality of tested gait recognition methods if applied in video surveillance.

1.2 Experimental Databases

Techniques for pose estimation from MoCap vary from one technology to another. Depending on which objects are tracked, there are two main categories of pose estimation: (1) tracking bone rotations and calculating 3D joint coordinates using a pre-calibrated skeleton; and (2) tracking the actual 3D joint coordinates.

Vicon captures bone rotations. The bone rotational data of each walker are associated with their skeleton that is unique and constant across all their walk sequences. To obtain the skeletons, the database acquisition staff places reflective markers on the walker's body on the positions of tracked joints and measures the individual bone lengths. The skeletons are meant to aid at calculation of 3D joint coordinates; however, they are a unique piece of information about each subject and even a trivial skeleton check yields a 100% recognition rate. Therefore, the

skeleton parameters obtained this way should not be presumed, not even for calculating joint trajectories. Identity classification of any recognition system should only be based on data readily available from surveillance cameras without any additional off-scene measurement.

Kinect, on the other hand, tracks joint positions directly. However, today's technology for estimating 3D joint coordinates at a distance acceptable for visual surveillance is highly error-prone. The random error of depth measurements increases quadratically with increasing distance from the sensor and in the case of Kinect it reaches 4 cm at the range of 5 m [22]. Kinect data acquisition setups therefore have a small tracking area. The applicability of Kinect-based system is limited to narrow corridors.

We identify state-of-the-art benchmark databases as a major drawback of this track of research as their inappropriateness hides some unpleasant, yet fundamental statistics. The databases of all surveyed research groups are either acquired by Kinect tracking a small area [9,13,16,17,18,19] or contain pre-calibrated skeletons of all participants [20]. Another problem is their size: The largest structural gait database currently available has 230 gait cycles of 22 walking subjects. How much do we know about a 100% method evaluated on a database of 10 subjects? A human recognition method evaluated on a database of 15 or 20 subjects is not convincing. Their small size hides information about both effectiveness and efficiency. Also, many methods involve large descriptors or slow classifiers and often forget that a person in a video has to be recognized within a few seconds.

For this purpose we extract a large number of gait cycles from a general MoCap database acquired by the Vicon system (see Section 3.1) and construct a prototypical skeleton to shroud the unique walker skeleton parameters. This is a skeleton-robust solution as all bone rotational data are linked with a fixed, explicitly defined skeleton. All walking subjects are assumed physically identical, disabling their pre-calculated skeletons to unfairly help at identity classifications. This normalization overcomes all above-mentioned issues and, at the same time, separates gait as a behavioral trait from skeleton parameters as a physiological trait.

2 Gait Pattern Recognition

This section describes input data, denotes terminology, and explains a two-step methodology – descriptor extraction and identity classification – as outlined in Figure 2. The method is exceptionally fast and the gait samples are of minimalistic size, while keeping the recognition rate comparable to the state-of-the-art methods.

2.1 Motion Data Representation

Human motion is digitally represented as a sequence of frames containing motion (kinematic) data acquired from a video of moving person. These are in the form of 3-dimensional coordinates of joints all over human body $j \in \mathcal{J}$, such as a hand, shoulder, or foot. Joint coordinates are measured at frames of synchronized and regular time intervals during the entire motion. Recall Figure 1 with stick figures as joint configurations in video frames.

To describe characteristic aspects of gait pattern, various kinds of motion features can be extracted at the level of individual frames from the joint coordinates, such as relational features [23], joint angles [10,11,13,14,16,18], inter-joint distances [13,14,16,18,20,21], velocities or acceleration. Over a human body model (of any underlying joint structure) we denote *descriptor* as a set $\mathcal{D} = \{F_1, \ldots, F_m\}$ of *features* as real discrete-time signals $F_i : \mathbb{N} \to \mathbb{R}$ to indicate the development of particular aspects of motion over time. This kinematic data (see Figure 3 left) are extracted from the 3D joint coordinates in all frames, e.g. left elbow angle or feet distance. Individual *feature values* $F_1(t), \ldots, F_m(t)$ extracted from a frame t form an m-dimensional vector $\mathcal{P}_t = (F_1(t), \ldots, F_m(t))$ called *pose*. A *motion* of n frames is then represented as a sequence $\mathcal{M} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$ of poses. *Gait cycle* $\mathcal{G}$ is a motion where subject performs two steps, a left step followed by a right step.

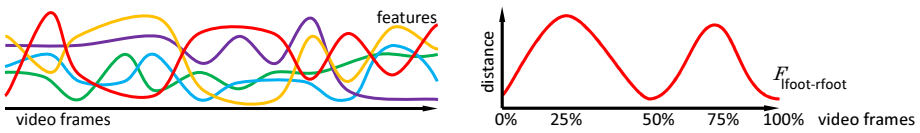

Fig. 3 Lines of different colors (left) indicate development of five feature signals through a motion. These signals contain kinematic information, such as elbow angle or torso joint velocity. Feet distance $F_{\texttt{lfoot-rfoot}}$ (right) as an inter-joint distance feature of a gait cycle where alternating high and low values indicate fluctuating leg spread.

Descriptors are designed on the basis of the underlying joint structure. Changing the structure or assigning different values to bone lengths may result in different classifications. Longer bone lengths amplify impact of associated geometric features. Any change in the skeleton parameters usually leads into decrease of recognition rate (see results of an experiment in Section 3.3.1).

Feature values are calculated from joint coordinates. For example, given spatial coordinates of 3 joints j_1, j_2 and j_3 as $c_{j_1}(t) = [x_t, y_t, z_t]$, $c_{j_2}(t) = [x'_t, y'_t, z'_t]$ and $c_{j_3}(t) = [x''_t, y''_t, z''_t]$ at frame t, respectively, the inter-joint

distance j_1–j_2, i.e. the Euclidean distance of the joints j_1 and j_2 in 3D, is the vector $\mathbf{v} = \overrightarrow{c_{j_1}(t)c_{j_2}(t)}$ magnitude

$$F_{j_1-j_2}(t) = \|\mathbf{v}\| = \sqrt{(x'_t - x_t)^2 + (y'_t - y_t)^2 + (z'_t - z_t)^2} \tag{1}$$

and joint angle j_1–j_2–j_3, is the angle given by the two vectors $\mathbf{v_1} = \overrightarrow{c_{j_2}(t)c_{j_1}(t)}$ and $\mathbf{v_2} = \overrightarrow{c_{j_2}(t)c_{j_3}(t)}$

$$F_{j_1-j_2-j_3}(t) = \frac{180}{\pi} \cos^{-1}\left(\frac{\mathbf{v_1} \cdot \mathbf{v_2}}{\|\mathbf{v_1}\|\|\mathbf{v_2}\|}\right). \tag{2}$$

Given a set of performer identities I and a recorded database of their gait patterns

$$D = \{(\mathcal{G}, id) \,|\, \text{performer of } \mathcal{G} \text{ has identity } id \in I\}, \tag{3}$$

we define *classifier* as a function of a query gait pattern $\mathcal{G}$ and the training database D that returns a classified identity $id \in I$. In the following we describe the steps of *descriptor extraction* and *identity classification* of our method in detail.

2.2 Descriptor Extraction

Inspired by many good geometric and relational gait features introduced by other authors, we extract a couple of joint angles. The strength of our method nests in the selection of only two particular poses, called *signature poses*, located within the gait sequence. Not all poses are necessary; on the contrary, some of them incorporate minor motion irregularities and act rather confusing at classification. Experiments show that these poses are discriminatory enough to serve as a biometric signature.

As said, motion capture data contain 3D joint coordinates of all tracked joints throughout the entire motion sequence. At each video frame t, the method first extracts the feet distance feature $F_{\texttt{lfoot-rfoot}}$ as the Euclidean distance between the feet coordinates (see Figure 3 right). This highly fluctuate feature most demonstratively shows the development of a gait cycle. Low values at roughly 0%, 50% and 100% signalize the feet passing by (single support), whereas high values at roughly 25% and 75% indicate heel strikes (double support). The signal is used to locate the two signature poses: The *strike pose* $\mathcal{P}_{strike}$, as the average of the poses at 25% and 75% gait cycle with leg stretch at maximum, and the *clearance pose* $\mathcal{P}_{clearance}$, as the average of the poses at 0%, 50% and 100% gait cycle with leg stretch at minimum. Signature poses can be obtained from the entire gait sequence, even if the walker provides multiple gait cycles. All residue poses are finally thrown away.

21 representative joint angles are extracted as gait features from the frames of each signature pose to form the gait pattern descriptor. These particular joint angles were selected because the interplay of their temporal variations adequately renders the motion during walking. They provide a complex description of each pose and there are no two different poses with identical descriptors. Adding extra joint angles in vision of a more precise motion representation does not make a difference in performance (see results of an experiment in Section 3.3.2). Also, joint angle is an appealing feature as current MoCap technology tracks 3D major bone rotations, marginalizing the errors of 3D joint coordinates' estimation. Each of the 21 triples j_1–j_2–j_3 in Table 1 determines the angle between vectors $\overrightarrow{c_{j_2}(t)c_{j_1}(t)}$ and $\overrightarrow{c_{j_2}(t)c_{j_3}(t)}$. Gait pattern is then described by the two signature poses

$$\mathcal{G} = (\mathcal{P}_{strike}, \mathcal{P}_{clearance}), \tag{4}$$

making a vector of length 42, i.e. 21 joint angles $F_{j_1-j_2-j_3}$ at both $\mathcal{P}_{strike}$ and $\mathcal{P}_{clearance}$. Note that this descriptor is of a very small size, which will have a noticeably positive effect on computational complexity.

j_1	j_2	j_3	j_1	j_2	j_3
lfoot	ltibia	lfemur	root	thorax	head
rfoot	rtibia	rfemur	lhumerus	thorax	rhumerus
lfoot	lfemur	lhipjoint	lhand	thorax	rhand
rfoot	rfemur	rhipjoint	root	lclavicle	lhumerus
lfemur	lhipjoint	root	root	rclavicle	rhumerus
rfemur	rhipjoint	root	thorax	lclavicle	lhumerus
ltibia	root	rtibia	thorax	rclavicle	rhumerus
lfemur	root	rfemur	lclavicle	lhumerus	lhand
lhumerus	root	rhumerus	rclavicle	rhumerus	rhand
lfoot	root	lhand	lclavicle	head	rclavicle
rfoot	root	rhand			

Table 1 21 joint angles $F_{j_1-j_2-j_3}$ extracted from both signature poses to form a gait pattern descriptor.

2.3 Identity Classification

Identity is classified by the baseline 1-NN classifier with L_1 as distance function. Note that gait cycles represented by this descriptor do not enclose temporal information of continuous character, thus using time-warping distance functions is unnecessary. Comparing values of individual signature pose parameters is only what is needed.

Conducted observations indicate that across a longer time span of surveillance, the set of one's gait patterns naturally forms a couple of dispersed clusters. This means that their new gait pattern will most likely be similar to a gait pattern they performed in the past. Provided that the system repeatedly encounters the same people, that is, each person has contributed with many gait samples, there is a very high probability for the nearest neighbor to be of the same identity. Attempts of using various machine learning techniques (see results of an experiment in Section 3.3.3) did not result in a significant qualitative improvement. The simple baseline 1-NN classifier achieves the recognition rate comparable to MLP, outperforming it quantitatively.

3 Experimental Evaluation

3.1 Experimental Database

The experimental set-up has to simulate video surveillance environment. All items listed in Section 1 are reflected in different parts of the system. Some of them, such as data capture constraints, size of the tracking space, ability to detect people walking in crowds or wearing various clothes and shoes, are addressed by choosing a proper data acquisition technology. Other, such as facility of data pre-calibration or training, variability in walkers' positions or directions of their walk, by a fitting pre-processing (normalization) technique. And more other, such as variability in walking speeds, multiplicity of subject enrollment, scale (numbers of subject identities and gait samples), computational time constraints or allowance for manual intervention, by designing a suitable recognition algorithm. This section describes the testing database as a collection of gait samples acquired in a laboratory that realistically simulates video surveillance environment, while disregarding data accuracy issues.

For the evaluation purposes we have extracted a large number of gait cycles from the general MoCap database from CMU [24] as a well-known and recognized database of structural human motion data. It contains numerous motion sequences, including a considerable number of gait sequences that can be extracted and used for evaluating MoCap-based gait recognition methods. Availability of such database is valuable to research groups so as not to spend resources on building their own.

Motions are recorded with an optical marker-based Vicon system. People wear a black jumpsuit and have 41 markers taped on. The tracking space of $30\,\mathrm{m}^2$ is surrounded by 12 cameras of sampling rate of 120 Hz in the height from 2 to 4 meters above ground. Please note that a typical video surveillance environment does not allow putting markers on the people, therefore it would be ideal to evaluate the methods by a marker-free technology. Unfortunately, to our best knowledge, the research on MoCap-based gait recognition starves from a reasonable benchmark database (see Section 1.2). The marker-based technology is understood to be more accurate; however, there has been a significant effort made in computer vision to advance marker-less motion capture software technology, such as iPi Soft[3], Organic Motion[4] or KinaTrax[5], but we lack an independent analysis on their accuracy. We understand this limitation and believe that evaluation on clean data, that is, without acquisition errors and noise, is still reasonable.

Motion videos are triangulated to get 3D data in the form of body point coordinates in each video frame and stored in the standard ASF/AMC data format. Each registered participant is assigned with their unique skeleton (ASF file) and their motions (AMC files) contain bone rotational data as instructions for this skeleton about how to deform over time while moving. To overcome the unique skeleton issue referred to in Section 1.2 and thus to use the collected data in a fairly manner, an explicit prototypical skeleton is constructed and used to represent bodies of all subjects. A skeleton with realistic parameters is calculated as the mean of all registered skeletons.

We calculate 3D joint coordinates of all joints $j \in \mathcal{J}$ using bone rotational data and the prototypical skeleton. The coordinates are necessary to calculate gait features for all tested methods. But raw joint coordinates refer to absolute positions in the tracking space and not all potential methods are invariant to person's position or walk direction. To ensure such invariance, center of the coordinate system is moved to the position of `root` joint $c_{\texttt{root}}(t) = [0, 0, 0]$ for each t and axes are adjusted to the walker's perspective: The X axis is from right (negative) to left (positive), the Y axis is from down (negative) to up (positive), and the Z axis is from back (negative) to front (positive). Normalized coordinates $\bar{c}_j(t)$ of joint j at frame t are calculated from the original $c_j(t)$ by subtracting the root coordinates and rotating around the Y axis by the angle θ between the X axis and the `lclavicle-rclavicle` line projected onto the XZ plane

$$\bar{c}_j(t) = \begin{bmatrix} \cos(\sigma\theta) & 0 & \sin(\sigma\theta) \\ 0 & 1 & 0 \\ -\sin(\sigma\theta) & 0 & \cos(\sigma\theta) \end{bmatrix} (c_j(t) - c_{\texttt{root}}(t)) \tag{5}$$

[3] `http://ipisoft.com`

[4] `http://www.organicmotion.com`

[5] `http://kinatrax.com`

with rotation direction $\sigma = 1$ if left shoulder has higher Z coordinate than right shoulder and $\sigma = -1$ otherwise.

Since the general motion database contains all motion types, we extracted a number of sub-motions that represent gait cycles. First, an exemplary gait cycle was identified, and clean gait cycles were then filtered out using the DTW distance over bone rotations. The similarity threshold was set high enough so that even the least similar sub-motion still semantically represents a gait cycle. Finally, subjects that contributed with less than 10 gait cycles were excluded.

The final database has 48 walking subjects that performed 4,188 gait cycles in total, which makes an average of about 87 gait cycles per subject. Out of this, we selected subsets 1 to 8 (see Table 2) to contain a fraction of gait cycles. Subset 1 has the fewest (73) and subset 8 has all (4,188) of them. The concept of subsets is to show how the size of an testing database affects performance. Even though a real application scales to much larger numbers, the current available MoCap gait databases are rather small and by extrapolating the results on this sequence one could develop a rough estimate on the performance beyond the scope of our experiments.

subset	1	2	3	4	5	6	7	8
subject identities	2	5	11	15	25	29	33	48
gait samples	73	303	585	941	1,393	2,017	2,951	4,188

Table 2 Sizes of the eight subsets extracted from the CMU MoCap database.

3.2 Results of Comparative Evaluation

In this section we provide comparative evaluation results of our recognition method (Balazia) against the related methods from Section 1.1. Some of these methods combine gait with skeleton parameters which we understand as a non-gait biometric. Recall that gait is a dynamic activity and thus a behavioral trait, whereas skeleton contains static parameters and thus is a physiological trait. Please also note that in the CMU database the skeleton data is a unique piece of information about each subject, such as the subject ID number, and need to be avoided from using at recognition. In our experiments we involve all MoCap-based gait recognition methods, including the ones that calculate physiological features, as relevant and potentially successful, even though they are not provided the skeleton data. With the same reasoning it would be appropriate to compare against methods that use a fusion with face or other available traits, yet to evaluate just the gait part.

All methods were implemented with their best-performance parameters and evaluated with 10-fold cross-validation as a unified recognition rate estimate. Figure 4 shows recognition rate, sample classification time of each implemented method evaluated on all subsets. Detailed performance statistics on subset 8 and comparative evaluation summary, as well as the configuration parameters are provided in Table 3.

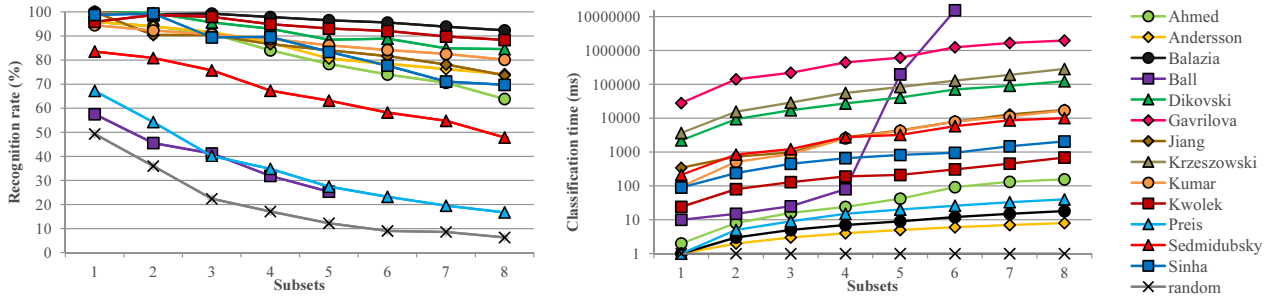

Fig. 4 Recognition rate in percent (left) and computational time of a sample classification in milliseconds (right) for all tested gait recognition methods evaluated the 8 subsets. Some statistics are not available due to extremely high computational costs.

Table 3 needs a discussion to explain the unacceptable time performance of some methods. Note that the largest database ever used for evaluating these methods had 230 gait samples, whereas ours has 4,188. Let us start with the Gavrilova's [14] half-an-hour-long classification time. This method uses a very large gait descriptor (20 inter-joint distances and 18 joint angles in 150 frames on average makes 5,700 real numbers) and classifies the identity using DTW (quadratic complexity with respect to number of poses and linear to number of features) to calculate two (one for inter-joint distances and one for joint angles) 3-NN sets of candidates. Given the 10-fold cross-validation on the set of 4,188 gait cycles, we have up to $4188 \cdot \frac{9}{10} \cdot 3 \cdot 38 \cdot 150^2 \approx 10^{10}$ elementary operations for classifying one gait sample. Analysis of Krzeszowski [16] is similar, except that their descriptor has 26 features per frame and uses the 1-NN classifier. The secret of Kumar's [17] computational time is hidden in the distance function: To compute similarity of two descriptors the method calculates eigenvalues of their covariance matrices. And finally,

Method	Recog. Rate	Class. Time	Descr. Size	Descriptor	Classifier
Ahmed [9]	63.8	156	6,080	HDF + VDF	L_1 1-NN
Andersson [10]	74.2	8	498	all features	L_1 1-NN
Balazia	92.4	18	576	joint angles at signat. poses	L_1 1-NN
Ball [11]	N/A	10^{10} (est.)	354	all features	K-means clustering
Dikovski [13]	84.5	123,066	1,489	Dataset 3	MLP
Gavrilova [14]	N/A	1,961,767	42,526	JRD + JRA	DTW(L_1)-based 3-NN + MV
Jiang [15]	73.7	17,497	4,751	DYN + STA	DTW(L_1)-based 1-NN
Krzeszowski [16]	N/A	284,435	30,721	all features	DTW(L_1)-based 1-NN
Kumar [17]	80.0	16,857	41,083	model covariance matrix	1-NN on gen. eigenvalues
Kwolek [18]	88.3	688	8,030	`g_all` features	MLP
Preis [19]	16.7	40	66	static and dynamic features	Naïve Bayes
Sedmidubsky [20]	47.8	10,073	2,391	$\mathcal{S}_{C_L H_L} + \mathcal{S}_{C_R H_R}$ with WN	DTW(L_1)-based 1-NN
Sinha [21]	69.7	2,055	840	all features + Ball + Preis	MLP
random	6.3	1	0	no features	random

Table 3 Comparison of individual methods in terms of recognition rate (%), sample classification time (milliseconds) and sample descriptor size (bytes) of all implemented methods for subset 8 only. Some statistics are not available (N/A) due to extremely high computational costs. High recognition rate and low classification time and descriptor size are desired. The color highlight describes acceptability of measured attributes: Green for excellent, yellow for sufficient, red for unacceptable, and black for not available. Two rightmost columns contain descriptor and classifier as configuration details of each implemented method.

the Ball's [11] method would be effective without the K-means clustering classifier where the initialization phase takes very long with such a large database.

At designing gait features we are interested in finding an optimal feature space where a gait descriptor is close to those of the same walker and far from those of different walkers. In other words, the goal is to find a discriminant that maximizes the misclassification margin, which is reflected in class separability coefficients. We additionally evaluated Davies-Bouldin Index (DBI), Dunn Index (DI), Silhouette Coefficient (SC) and Fisher's Discriminant Ratio (FDR), and the results are illustrated in Figure 5. Feature extraction algorithms that produce samples of low intra-class distances and of high inter-class distances have a low DBI and high DI, SC and FDR. The class separability coefficients give an estimate on the potential of the extracted features and do not reflect eventual combination with a bad classifier.

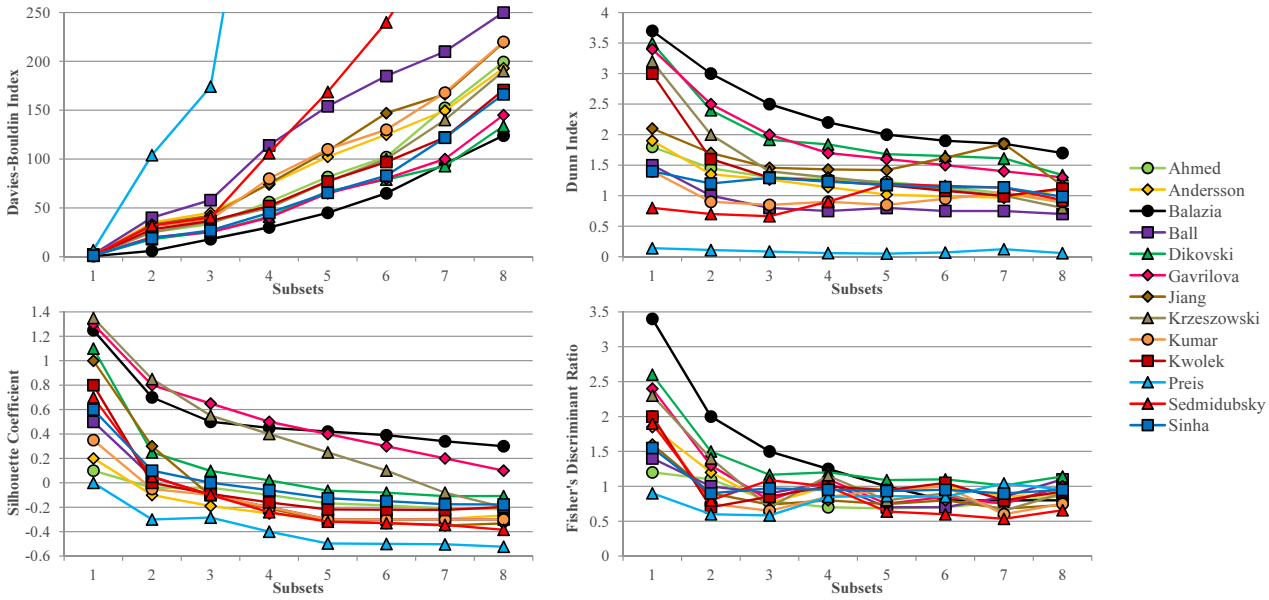

Fig. 5 Four class separability coefficients for extracted features of all tested gait recognition methods. Evaluated on the 8 subsets.

On top of this, on subset 1 we measure False Accept Rate vs. False Reject Rate (FAR/FRR), Receiver Operating Characteristic (ROC) as True Accept Rate (TAR) vs. False Accept Rate and finally recall vs. precision in order to evaluate the system based on distance distribution. Results are plotted in Figure 6. A desirable method intersects its FAR and FRR curves at a low value to minimize the Equal Error Rate (EER), pushes its ROC curve to the top left corner of the graph to maximize the Area Under Curve (AUC), and obtains high values of recall and precision simultaneously to get a high Mean Average Precision (MAP). Table 4 provides the supplementary statistics EER, AUC and MAP.

The introduced collection of features appears to be more effective than the features introduced by other authors: Figure 5 indicates the best results in DBI, DI, SC and FDR. Figure 6 shows that our method achieves the best

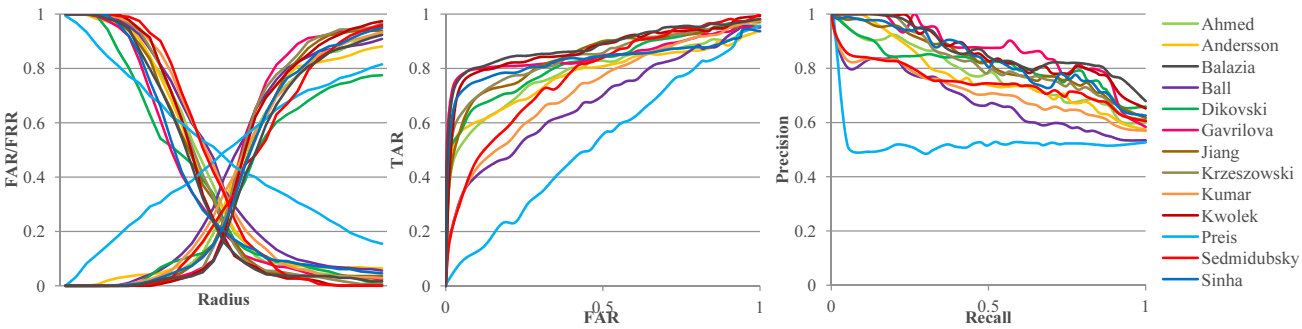

Fig. 6 FAR/FRR (left), ROC (middle) and recall/precision (right) metrics of all tested methods for subset 1.

Method	EER	AUC	MAP
Ahmed [9]	0.2779	0.7811	0.7948
Andersson [10]	0.2858	0.7755	0.7996
Balazia	0.1674	0.8902	0.8907
Ball [11]	0.3949	0.6714	0.6931
Dikovski [13]	0.2273	0.8288	0.8287
Gavrilova [14]	0.1922	0.8521	0.8904
Jiang [15]	0.2393	0.8414	0.8497
Krzeszowski [16]	0.2287	0.8464	0.8593
Kumar [17]	0.3545	0.6528	0.6197
Kwolek [18]	0.1839	0.8796	0.8840
Preis [19]	0.4916	0.5225	0.5356
Sedmidubsky [20]	0.2823	0.7726	0.7558
Sinha [21]	0.2194	0.8289	0.8493

Table 4 EER, AUC and MAP derived from FAR/FRR, ROC and recall/precision, respectively, plotted in Figure 6.

CMC, FAR/FRR, ROC and recall/precision scores along with all CCR, EER, AUC and MAP. We interpret the high scores as a sign of robustness. The recognition rate of 92.4% is the highest among all tested methods not only on subset 8 (see the recognition rate column in Table 3) but also keeps its first place on seven other smaller subsets (see the left graph in Figure 4). Apart from performance merits, the method is also efficient: The combination of its minimalistic descriptors (see the descriptor size column in Table 3) and the 1-NN classifier based on the L_1 descriptor distance function ensures fast classification time (see the right graph in Figure 4 and the classification time column in Table 3) and thus contribute to high scalability.

3.3 Method Variations

We present skeleton, descriptor and classifier variations to demonstrate effort for improving our method. While performing these experiments we frequently discovered configurations of increasing recognition rate and/or decreasing computational time. Finally, we proposed the optimal among all tested configurations.

3.3.1 Skeleton Variations

The prototypical skeleton is defined by the underlying structure and bone lengths. Features of all registered walkers are mapped onto this skeleton, even if their real parameters are different due to a measurement error or physical disproportions of the walker. Joint angles are selected to form the gait pattern descriptor for a good reason: Not only that the motion acquisition technology directly and more accurately captures bone rotations, whereas the inter-joint distances differ from the real ones if associated with another skeleton, but also they are sufficiently skeleton-robust. This means that the model does not collapse even if data are mapped wrongly because of a failure in 3D tracking technology.

Since changing the skeleton transforms the descriptors non-linearly and their distances are not preserved, we cannot claim that the model is perfectly robust. However, the joint angles are convenient here because taking different skeleton parameters makes the feature signals only shift their range while keeping their temporal variations unchanged. Therefore, these features make changes in skeleton parameters have negligible effect on recognition rate. In order to find out how skeleton-robust the model is, we conducted a series of experiments where we assumed four skeleton mapping errors and measured their drop in recognition rate. The only assumption is that the model registers the same type of error at all times. Figure 7 illustrates recognition rate of the model with the following errors: original (no error), half-sized, double-sized, disproportional (double-sized left side only), and random (all bone lengths set at random). Our experiments show that mapping on even a random skeleton does not result in a major drop in performance.

3.3.2 Descriptor Variations

Our descriptor has been tested on whether adding or removing some features improves recognition rate. In addition to the 21 original joint angles (see Table 1), we consider six new joint angles $F_{\texttt{ltibia-lfemur-lhipjoint}}$, $F_{\texttt{rtibia-rfemur-rhipjoint}}$, $F_{\texttt{lfoot-root-rfoot}}$, $F_{\texttt{lhand-root-rhand}}$, $F_{\texttt{lhand-head-rhand}}$ and $F_{\texttt{lhumerus-head-rhumerus}}$. $\binom{6}{1} = 6$ new descriptors are built by taking the original 21 joint angles plus one of the 6 new joint angles, $\binom{6}{2} = 15$ new descriptors by taking the original plus two of the new joint angles, up until $\binom{6}{6} = 1$ new descriptor of all $21 + 6$ joint angles, building a total of 63 new descriptors. Each new descriptor contains the original 21 joint angles and a subset of the new joint angles. In the same experiment, we construct 231 other new descriptors with up to 2 original joint angles less, that is, $\binom{21}{1} = 21$ new descriptors of one original joint angle less and $\binom{21}{2} = 210$ new descriptors of two original joint angles less. Figure 7 illustrates the recognition rates of the proposed method with all new descriptors categorized by the number of changes in joint angles.

The proposed 21 joint angles are selected as the final feature set as we see that neither adding nor removing joint angles seems to improve recognition rate. We tested just a limited number of new joint angles as we believe that the model cannot be significantly improved in this aspect. Evaluating all subsets of 13,485 potential joint angles would be more sound, but hardly achievable in a reasonable time.

We tested our approach with one more descriptor variation: Instead of taking the two signature poses we took all poses through the gait cycle and measured whether having the full temporal information helps at recognition. The results illustrated in Figure 7 demonstrate that the large portion of data that causes confusion of identities. Taking the signature poses slightly improves the performance, decreases classification time about 20 times, and reduces the descriptor size from 25,440 to just 576 bytes.

3.3.3 Classifier Variations

The purpose of the following experiments is to try various classifiers that can potentially improve our method. Figure 8 shows the recognition rates and sample classification time of the proposed method with six well-known classifiers: Baseline k-NN classifier plus Majority voting with k up to 15, MultiLayer Perceptron (MLP) with learning rate 0.1 and momentum 0.1, linear classifier, logistic regression, Support Vector Machines (SVM), and J48 decision trees with C4.5 algorithm. We worked with the WEKA suite of machine learning software [25].

It might appear surprising that the baseline 1-NN classifier outperforms the remaining classifiers that are based on machine learning principles. Our reasoning here is that gait cycles of individual walkers do not naturally form just a single cluster – they form multiple clusters. When a lady owns two pairs of shoes, one with high heels and the other without, her gait cycles naturally form two clusters. Also one could sometimes carry a suitcase or they could be in a hurry from time to time. A person often has multiple ways of walking and when they are spotted on a camera they tend to walk in a way they did before. The 1-NN classifier is designed exactly for this: It finds the closest gait pattern (the way they walked before) and propagates its identity as the result.

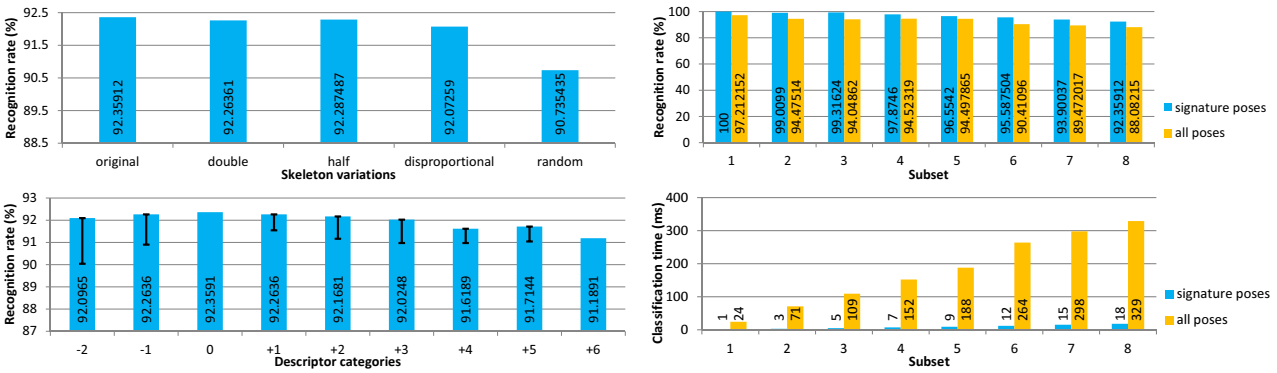

Fig. 7 Recognition rate for five skeleton variations (top left). Nine categories of descriptor variations (bottom left): Descriptors in each category are created by adding (positive) up to 6 or removing (negative) up to 2 joint angles. Classification time and descriptor size differences are negligible. The thin black bars show the range of recognition rates across all descriptors in respective category, e.g., the zero range (no bar) of the single descriptor in the +6 category and a large range (tall bar) of the 210 descriptors of the −2 category. Recognition rate (top right) and sample classification time (bottom right) of our original method with signature poses and of a descriptor variation with processing all poses. All is evaluated on subset 8.

4 Conclusion

Gait as a behavioral biometric trait comprises temporal aspects of walk. Abstracting from individual walker skeletons highlights the walk dynamics over the static body parameters. We introduce the concept of a prototypical

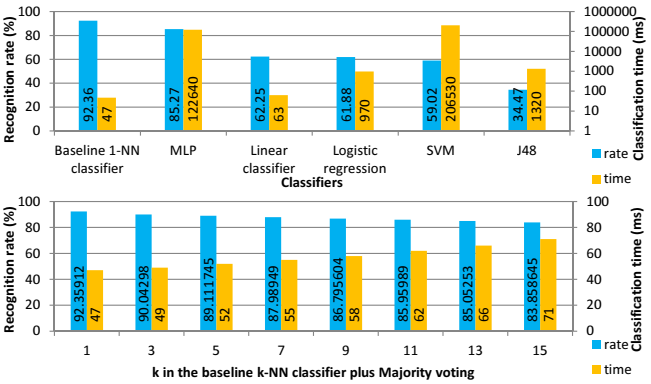

Fig. 8 Recognition rate and sample classification time on subset 8 for five additional classifiers (top) and for the baseline k-NN plus Majority voting with k up to 15 as the sixth tested classifier (bottom).

skeleton that is kept fixed for all motions performed by all registered participants. Note that one should not dismiss other potentially usable data just because they are not purely associated with gait; on the contrary, it is highly encouraged to enhance a gait-based recognition system by combining it with additional distance-capturable traits, such as the mentioned skeleton parameters or face.

Our gait recognition method is designed to process all known numerical types of gait features. Minimalistic gait pattern descriptors are extracted from people's walk sequences and their identity is classified by a baseline 1-NN classifier, making it extremely fast while keeping the recognition rate comparable to the state-of-the-art methods. In order to potentially service video surveillance applications, our method has been evaluated on a gait database that precisely simulates street-level video monitoring environment. The collection of introduced gait features achieves leading scores in four class separability coefficients and therefore has a strong potential in gait recognition applications. This is demonstrated by outperforming other methods in numerous classification metrics.

Acknowledgments

The authors thank to the reviewers and editor for their detailed commentary and suggestions. This research was partially supported by the specific research project MUNI/A/1206/2014 of Faculty of Informatics at Masaryk University and by the grant program Študenti do sveta 2015SDS074 of Nadácia Tatra banky. The data used in this project was created with funding from NSF EIA-0196217 and was obtained from `mocap.cs.cmu.edu`.

References

1. Valcik, J., Sedmidubsky, J., Zezula, P.: 'Assessing Similarity Models for Human-Motion Retrieval Applications', Computer Animation and Virtual Worlds, 2016, **27**, (5), pp. 484–500
2. Auvinet, E., Multon, F., Aubin, C.E., Meunierd, J., Raison, M.: 'Detection of Gait Cycles in Treadmill Walking using a Kinect', Gait & Posture, 2015, **41**, (2), pp. 722–725
3. Valcik, J., Sedmidubsky, J., Balazia, M., Zezula, P.: 'Identifying Walk Cycles for Human Recognition', Proc. Pacific Asia Workshop on Intelligence and Security Informatics (PAISI), 2012, pp. 127–135
4. Choensawat, W., Choi, W., Hachimura, K.: 'Similarity Retrieval of Motion Capture Data Based on Derivative Features', Advanced Computational Intelligence and Intelligent Informatics, 2012, **16**, (1), pp. 13–23
5. Hu, M.C., Chen, C.W., Cheng, W.H., Chang, C.H., Lai, J.H., Wu, J.L.: 'Real-Time Human Movement Retrieval and Assessment With Kinect Sensor', IEEE Transactions on Cybernetics, 2015, **45**, (4), pp. 742–753
6. Kapsouras, I., Nikolaidis, N.: 'Action Recognition in Motion Capture Data Using a Bag of Postures Approach', Int. Conf. Pattern Recognition (ICPR), 2014, pp. 2649–2654
7. Leightley, D., Li, B., McPhee, J.S., Yap, M.H., Darby, J.: 'Exemplar-Based Human Action Recognition with Template Matching from a Stream of Motion Capture', Image Analysis and Recognition, 2014, LNCS 8815, pp. 12–20
8. Vantigodi, S., Radhakrishnan, V.B.: 'Action Recognition from Motion Capture Data Using Meta-Cognitive RBF Network Classifier', IEEE Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014, pp. 1–6
9. Ahmed, M., Al-Jawad, N., Sabir, A.: 'Gait Recognition Based on Kinect Sensor', Proc. SPIE, Real-Time Image and Video Processing, 2014, **9139**, pp. B:1–B:10
10. Andersson, V., Dutra, R., Araujo, R.: 'Anthropometric and Human Gait Identification using Skeleton Data from Kinect Sensor', Proc. ACM Symp. Applied Computing, 2014, pp. 60–61
11. Ball, A., Rye, D., Ramos, F., Velonaki, M.: 'Unsupervised Clustering of People from 'Skeleton' Data', Proc. ACM/IEEE Int. Conf. Human-Robot Interaction, 2012, pp. 225–226
12. Derlatka, M., Bogdan, M.: 'Fusion of Static and Dynamic Parameters at Decision Level in Human Gait Recognition', Pattern Recognition and Machine Intelligence, 2015, LNCS 9124, pp. 515–524
13. Dikovski, B., Madjarov, G., Gjorgjevikj, D.: 'Evaluation of Different Feature Sets for Gait Recognition Using Skeletal Data from Kinect', Information and Communication Technology, Electronics and Microelectronics, 2014, pp. 1304–1308
14. Ahmed, F., Paul, P.P., Gavrilova, M.L.: 'DTW-Based Kernel and Rank-Level Fusion for 3D Gait Recognition using Kinect', The Visual Computer, 2015, **31**, (6-8), pp. 915–924
15. Jiang, S., Wang, Y., Zhang, Y., Sun, J.: 'Real Time Gait Recognition System Based on Kinect Skeleton Feature', ACCV Workshops on Computer Vision, 2015, LNCS 9008, pp. 46–57

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', Computer Vision and Graphics, 2014, **8671**, pp. 356–363
17. Kumar, M.S.N, Babu, R.V.: 'Human Gait Recognition Using Depth Camera: A Covariance Based Approach', Computer Vision, Graphics and Image Processing (ICVGIP), 2012, pp. 20:1–20:6
18. Kwolek, B., Krzeszowski, T., Michalczuk, A., Josinski, H.: '3D Gait Recognition Using Spatio-Temporal Motion Descriptors', Intelligent Information and Database Systems (ACIIDS), 2014, **8398**, pp. 595–604
19. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: 'Gait Recognition with Kinect', International Workshop on Kinect in Pervasive Computing, 2012
20. Sedmidubsky, J., Valcik, J., Balazia, M., Zezula, P.: 'Gait Recognition Based on Normalized Walk Cycles', Int. Symp. Visual Computing (ISVC), 2012, Springer, pp. 11–20
21. Sinha, A., Chakravarty, K., Bhowmick, B.: 'Person Identification using Skeleton Information from Kinect', Advances in Computer-Human Interactions, 2013, pp. 101–108
22. Khoshelham, K.: 'Accuracy Analysis of Kinect Depth Data', ISPRS Workshop Laser Scanning, 2011, **38**
23. Müller, M., Baak, A., Seidel, H.: 'Efficient and Robust Annotation of Motion Capture Data', ACM SIGGRAPH/Eurographics Symp. Computer Animation (SCA), 2009, pp. 17–26
24. Carnegie Mellon University: 'Carnegie-Mellon Motion Capture (MoCap) Database', `http://mocap.cs.cmu.edu`, 2003
25. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: 'The WEKA Data Mining Software: An Update', SIGKDD Explorations, 2009, **11**, (1), pp. 10–18