CrossMark

Dynamic time warping in classification and selection of motion capture data

Adam Switonski[1] · Henryk Josinski[1] · Konrad Wojciechowski[2]

Abstract

The paper is a comprehensive study on classification of motion capture data on the basis of dynamic time warping (DTW) transform. It presents both theoretical description of all applied and newly proposed methods and experimentally obtained results on real dataset of human gait with 436 samples of 30 males. The recognition is carried out by the classical DTW nearest neighbors classifier and introduced DTW minimum distance scheme. Class prototypes are determined on the basis of DTW alignment and chosen methods of averaging rotations represented by Euler angles and unit quaternions. In the basic classification approach the whole pose configuration space is taken into account. The influence of different rotation distance functions operating on Euler angles and unit quaternions, on an obtained accuracy of recognition is investigated. What is more, a differential filtering in time domain which approximates angular velocities and accelerations of subsequent joints is utilized. Because in the case of unit quaternions representing rotations classical subtraction is unworkable, the differential filtering based on a product with a conjugated quaternion is applied. The main contribution of the paper is also related to the proposed and successfully validated approach of an exploration of pose configuration space. It selects the most discriminative joints of a skeleton model in considered classification problem in a binary or fuzzy way. The introduced approach utilizes hill climbing and genetic search strategies as well as DTW transform based evaluation. The selection makes the recognition more efficient and reduces pose signatures.

1 Introduction

In motion capture measurements markers attached on a human body are tracked by a set of calibrated cameras. It allows to reconstruct markers' 3D coordinates in a global system. On

---

✉ Adam Switonski
    adam.switonski@polsl.pl

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

[2] Research and Development Center of Polish-Japanese Academy of Information Technology, Aleja Legionow 2, 41-902 Bytom, Poland

the basis of such data a skeleton model is determined. In the preliminary calibration stage, markers are clustered to groups of separate bone segments with estimated lengths in such a way that distances between instances from the same clusters are preserved. To precisely estimate skeleton segments motion data with movements of all joints are required. Thus, before a proper acquisition a person participating in experiments has to perform a special set of exercises, which takes into consideration all degrees of freedom of the specified model.

Finally, in subsequent time instants relative positions between adjacent segments are established and they are represented by rotations. Skeleton data have a tree like structure of kinematic chain. The root object placed by default in lower part of the spine is described by global rotation and position. The descendants are specified by rotations performed by subsequent joints with respect to a given reference pose. As a result, skeleton based motion capture data contain sequences of poses, described by rotations and global translation.

The problem of motion data classification is related to the analysis of multidimensional time series. The broad review of the proposed and successfully applied methods of motion data classification can be found in Aggarwal and Ryoo (2011). In the first of the most often used approaches feature extraction is carried out. It constructs motion descriptors containing a set of defined numeric attributes. Thus, well-known supervised or unsupervised classification can be applied. There are many different strategies of features extraction. For instance in Balazia and Plataniotis (2017) strike and clearance poses are introduced and taken to be gait descriptors. In Lu et al. (2008), Josiński et al. (2012), Balazia and Sojka (2018) generic feature extraction is carried out on the basis of Multilinear Principal Component Analysis, Linear Discriminant Analysis and Maximum Margin Criterion. Feature extraction approach of motion data is commonly used in medical applications. In such a case attributes usually correspond to medical diagnosis of a clinical state of a patient, useful in the assessment of specified motion disorders. A well-known technique successfully introduced to Parkinson's disease diagnosis is a set of gait indices. Among others, they are related to approximation of: step length symmetry, arm swing asymmetry, arm swing size asymmetry, and bilateral coordination of movements between a given pair of joints (Mian et al. 2011; Plotnik et al. 2007; Zifchock et al. 2008).

In another classical approach of motion data classification Hidden Markov Models (HMM) are utilized. In such a case motion time series are considered to be Markov chains and for every class a single model is built. In proper recognition stage the model with the greatest probability of generating given motion sequence is determined. Because of difficulty in probability estimation of high dimensional random variables, necessary in HMM modeling, the process is usually preceded by some kind of dimensionality reduction of pose or silhouette spaces (Bashir et al. 2005).

The motion similarity can be approximated on the basis of aggregated comparison of the corresponding poses. It is only directly applicable in the case of the time series with an equal number of frames. In the naive approach linear scaling of the time domain can be utilized to satisfy the requirement. However, much more adequate synchronization is obtained by dynamic time warping (DTW), which also takes into consideration local shifts between gait phases. In Gavrila and Davis (1995) selected motion activities are recognized on the basis of DTW applied to low precisely reconstructed skeleton data with 17 degrees of freedom corresponding to torso, upper limbs and head. In Muller and Roder (2006) DTW is carried out in respect to motion templates with binary features of subsequent poses. Another application of DTW is presented in Barnachon et al. (2014). The proposed method is based on histograms of action poses, extracted from MoCap (motion capture) data, that are computed according to the Hausdorff distance. The histograms are then compared with the Bhattacharyya distance and warped by DTW. In Krzeszowski et al. (2014) classification is related to kinematic motion

capture data and it is applied to human gait identification problem. The example results of application of dynamic time warping to quaternion time series data are presented in Jablonski (2012).

In recent years dynamic time warping has gained more attention in the analysis and classification of motion data. It was used:

– to recognize tennis shots and selected sport activities as well as to construct an index for quantitative upper-limb mobility evaluation in post-stroke rehabilitation with acquisition carried out by wearable IMU sensors (Srivastava and Sinha 2016; Margarito et al. 2016; Zhang et al. 2016),
– to discover behavioral patterns of attention deficit hyperactivity disorder (ADHD) on the basis of features extracted from RGB-D acquisition (Bautista et al. 2016),
– to develop user interface that tracks and recognizes hand gestures in real time and to train exemplar of the target action movement with immediately acquired motion instructions while learning, which are based on depth data collected by a Kinect sensor (Plouffe and Cretu 2016; Hu et al. 2015),
– to analyze gestures movements represented by EMG measurements (AbdelMaseeh et al. 2016),
– to classify motion activities characterized by wireless on-body creeping signal (Li et al. 2016).

DTW also has numerous applications in the analysis of life data which are not only limited to human motion analysis. For instance, it was used in the classification of imagined words coming from electrocorticographic signals (Martin et al. 2016) or proteomic data analysis (Liu et al. 2016).

Though numerous publications devoted to dynamic time warping, there is still lack of a comprehensive study on DTW focused on highly precised skeleton based motion capture data with exhaustive computational experiments, which is a subject of the paper. The main contribution of the presented work is related to the introduced method of selection of the most discriminative pose configuration parameters with respect to a specified classification task. It is based on chosen heuristic search strategies and dynamic time warping transform. Moreover, minimum distance classification scheme is proposed. It utilizes DTW alignment and chosen techniques of averaging rotations coded by Euler angles and unit quaternions. The considered methods are successfully applied to real dataset of human gait.

2 Dynamic time warping

Dynamic time warping, originally applied to spoken word recognition (Sakoe and Chuba 1978), is a very effective method of time series comparison and classification. It outperforms both simple lock-step measures as for instance Euclidean or Manhattan metrics and more sophisticated edit distance approaches—Longest Common SubSequence (Andre-Jonsson and Badal 1997), Edit Sequence on Real Sequence (Chen et al. 2005), Swale Morse and Patel (2007), Edit Distance with Real Penalty (Chen and Raymond 2004; Wang et al. 2013). Thus, its choice to the problem of motion data analysis seems to be justified.

DTW is an elastic measure type, which means that one-to-many mappings are allowed. It matches time instants of compared sequences by a nonlinear, monotonic transformation. As a result, the lowest total cost calculated as aggregation of dissimilarity between matched poses is determined and it corresponds to a dissimilarity between motion data.
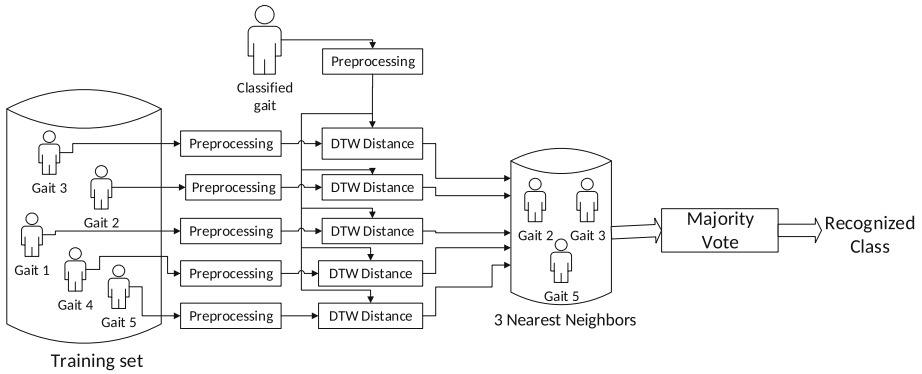
Fig. 1 Flowchart of 3NN-DTW gait classification

DTW requires to calculate a dissimilarity matrix containing distances between every pair of poses of analyzed motions as presented in Fig. 8. The matrix is used to determine monotonic path with minimal aggregated total distance across underlying pose dissimilarity values, connecting edge points corresponding to start and end of compared time series. To perform the computations, dynamic programming can be applied. The path is monotonic, which means that moving backward in a time domain is not allowed. Thus, the cost to reach the specified poses i and j of the first and second motions, respectively, can be determined on the basis of costs reaching to possible previous neighbors with indexes $(i-1, j)$, $(i, j-1)$, and $(i-1, j-1)$. Hence, it is sufficient to carry out the calculations in the proper order for subsequent rows or columns.

Listing 1 DTW Dynamic Programming Procedure

```
{cost − cost matrix}
{sim − dissimilarity matrix}
{m,n − resolution of cost and sim matrices}
cost[:,0]:=inf;
cost[0,:]:=inf;
cost[0,0]:=0;
for i:=1 to n do
  for j:=1 to m do
    cost[i,j]:=sim[i,j]+min(cost[i−1,j],cost[i,j−1],cost[i−1,j−1]);
pathCost:=cost(n,m);
```

The final cost of the path found corresponds to motions dissimilarity and the shape of the path—to performed synchronization. The shape can be reconstructed on the basis of backtracing of possible previous neighbors with minimal cost. Thus, the nearest neighbors classification scheme with DTW based motion distance assessment can be applied to recognize the data as illustrated in Fig. 1. It is also possible to utilize support vector machine (SVM) algorithm along with DTW distance to classify time series data (Mei et al. 2016). However, such DTW application is disputable, because triangle inequality condition is not satisfied by the DTW distance.

2.1 Dissimilarity of poses

In the first stage of DTW computations dissimilarity matrix has to be determined. It requires to specify how pairs of poses are compared. As previously described, pose configuration parameters are related to rotations performed by subsequent joints with respect to a given reference pose. The rotations are specified in local coordinate systems. Thus, in the most obvious way, dissimilarity between two poses P_1 and P_2 is determined by aggregated total distance of corresponding skeletal joints:

$$d(P_1, P_2) = \sum_{joint} d_{rotation}(P_1(joint), P_2(joint)) \tag{1}$$

Therefore, the crucial challenge is the proper choice of distance function $d_{rotation}$ responsible for assessment of dissimilarity between two rotations. By default, rotations are coded by three Euler angles (α, β, γ). Such data contain basic rotations performed around axes of the local coordinate system. Hence, any classical distance function of a vector space can be utilized, as for instance Euclidean or Manhattan metrics. However, periodicity of angle range has to be taken into account—the difference Δ between two angles α_1 and α_2 is calculated as: $\Delta(\alpha_1, \alpha_2) = \pi - ||\alpha_1 - \alpha_2| - \pi|$.

$$d_{Euc}((\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2)) = \sqrt{\Delta(\alpha_1, \alpha_2)^2 + \Delta(\beta_1, \beta_2)^2 + \Delta(\gamma_1, \gamma_2)^2} \tag{2}$$

$$d_{Man}((\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2, \gamma_2)) = |\Delta(\alpha_1, \alpha_2)| + |\Delta(\beta_1, \beta_2)| + |\Delta(\gamma_1, \gamma_2)| \tag{3}$$

Different, efficient and compact representation of rotations is given by quaternions. They are a natural extension of complex numbers with three-dimensional imaginary part:

$$q = a + i \cdot b + j \cdot c + k \cdot d \tag{4}$$

In the case of rotation with θ angle around axis represented by a unit length vector $\mathbf{u}$, the scalar part of a quaternion $a = cos(\frac{\theta}{2})$ and the vector imaginary part $(\mathbf{b}, \mathbf{c}, \mathbf{d}) = sin(\frac{\theta}{2}) \cdot \mathbf{u}$.

Quaternions representing rotations are of unit length, which means they are located only on a hypersphere S^3. Thus, to compare rotations, it is sufficient to calculate geodesic distance between related quaternions which is reflected by the angle between vectors formed by their components—the simplified 3D illustration is depicted in Fig. 2a. The scalar product $< q_1, q_2 >$ can be used to accomplish the task (Huynh 2009):

$$d_{geodesic}(q_1, q_2) = \frac{1}{\pi} \cdot arccos(< q_1, q_2 >) \tag{5}$$

In another approach instead of raw angle its cosine can be determined, which transforms geodesic distance in nonlinear way (Huynh 2009):

$$d_{cosine}(q_1, q_2) = \frac{1 - < q_1, q_2 >}{2} \tag{6}$$

The logarithm operator in unit quaternion space is defined as:

$$\log(q) = \mathbf{u} \frac{\alpha}{2} \tag{7}$$

It transforms an input value into three-dimensional vector $\mathbb{H}_1 \rightarrow \mathbb{R}^3$ of tangent space at the identity point of hypersphere S^3 located in $(1, 0, 0, 0)$. The analogous operation for 3D space is illustrated in Fig. 2b. Hence, any default distance metric in a tangent space can be utilized to compare rotations.

$$d_{TS}(q_1, q_2) = \|(log(q_1) - log(q_2))\| \tag{8}$$
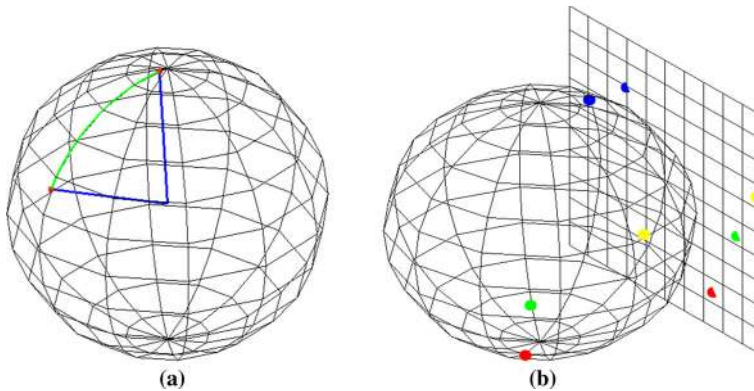
Fig. 2 Simplified 3D visualizations of selected quaternion distance functions. **a** Quaternion geodesic distance. **b** Projection to a tangent space

The product of two quaternions calculated on the basis of below stated multiplication formulas of imaginary coefficients, concatenates two rotations.

$$
\begin{aligned}
i^2 = j^2 = k^2 = i \cdot j \cdot k &= -1 \\
i \cdot j = k \qquad j \cdot i &= -k \\
j \cdot k = i \qquad k \cdot j &= -i \\
k \cdot i = j \qquad i \cdot k &= -j
\end{aligned}
\tag{9}
$$

What is more, the conjugate quaternion $\overline{q}$ with negative vector part $(-\mathbf{b}, -\mathbf{c}, -\mathbf{d})$ refers to an opposite rotation. Thus, the product $\overline{q_1} \cdot q_2$ specifies the rotation which has to be performed to obtain q_2 from q_1 ($q_2 = q_1 \cdot \overline{q_1} \cdot q_2$) and its real part approximates dissimilarity between q_1 and q_2.

$$
d_C(q_1, q_2) = 1 - re(\overline{q_1} \cdot q_2)
\tag{10}
$$

There is another issue which has to be taken into consideration. Because of duality of unit quaternion space (q and $-q$ describe the same rotation)—minimum distance between q_1 and $\pm q_2$ is a proper measure d_X^H (X denotes: geodesic, cosine, TS, C) of dissimilarity of specified rotations:

$$
d_X^H(q_1, q_2) = min(d_x(q_1, q_2), d_x(q_1, -q_2))
\tag{11}
$$

It ensures that only quaternions located on a single hypersphere are compared and such a process is called negative selection or hemispherization.

In computer graphics affine matrices are frequently used to represent some geometric transformations. The basic matrices $R_X(\alpha)$, $R_Y(\beta)$ and $R_Z(\gamma)$ which correspond to rotations of Euler angles α, β and γ, respectively, are as follows:

$$R_X(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_Y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_Z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 & 0 \\ \sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$(12)$$

Multiplication of matrices combines rotation similarly to quaternions. Thus, final matrix $R_{XYZ}(\alpha, \beta, \gamma)$ is calculated as a product of the basic ones $R_X(\alpha) \cdot R_Y(\beta) \cdot R_Z(\gamma)$. To compare such represented rotations Frobenius norm is usually applied:

$$d_F(R_{XYZ}(\alpha_1, \beta_1, \gamma_1), R_{XYZ}(\alpha_2, \beta_2, \gamma_2)) = ||R_{XYZ}(\alpha_1, \beta_1, \gamma_1) - R_{XYZ}(\alpha_2, \beta_2, \gamma_2)||_F$$

$$(13)$$

where: $||R||_F = \sum_i \sum_j r_{ij}^2$ and r_{ij} are elements of R.

The comparison of rotations performed by aforementioned distance functions in respect to hip joint movements is visualized in Fig. 3. The differences are easily to be noticed. Some of the functions ($d_{Man}, d_{geodesic}, d_{TS}, d_F$) enhance distances between more similar rotations, while the others (d_{cosine}, d_{Euc})—between more distant ones. However, at the current stage it is very difficult to state final conclusion of preferable approach to dissimilarity matrix calculation of DTW transform. An obtained accuracy of classification will be crucial.

In another way a pose can be described by features related to dynamics of the movements (Keogh and Pazzani 2001). In such a case angular velocities v and accelerations a have to be approximated. They correspond to the first and second derivatives of motion capture sequences. Differential filtering in time domain can be applied for the purpose of approximation of v and a. In the case of Euler angles simple approach can be utilized, which is based on differences of acquired angles and velocities at subsequent time instants t:

$$v_t = \Delta(angle_t, angle_{t-1}) \tag{14}$$

$$a_t = \Delta(v_t, v_{t-1}) = \Delta(\Delta(angle_t, angle_{t-1}), \Delta(angle_{t-1}, angle_{t-2})) \tag{15}$$

Such an approach is unworkable in the case of unit quaternions. First of all, requirement of unit length is not satisfied by the subtraction operator. Thus, its result does not represent any rotation, in particular the one which reflects the difference. To approximate an angular velocity the approach introduced in d_C distance function (10) utilizing product with conjugate quaternion can be applied once again. Such a product determines a rotation which, when concatenated with the first quaternion, is identical to the second one. This is exactly what is needed. Thus, final angular quaternion based velocities v^{quat} and accelerations a^{quat} are estimated as:

$$v_t^{quat} = q_t \cdot \overline{q_{t-1}} \tag{16}$$

$$a_t^{quat} = v_t^{quat} \cdot \overline{v_{t-1}^{quat}} = q_t \cdot \overline{q_{t-1}} \cdot \overline{q_{t-1} \cdot \overline{q_{t-2}}} \tag{17}$$

Similarly to basic configuration parameters, to compare a pose described by angular velocities or accelerations only proper distance functions have to be chosen. The choice depends on a rotation coding. In the case of default differential filtering of Euler angles d_{Euc} and d_{Man} functions are applicable and for quaternion based approaches $d_{geodesic}^H, d_{cosine}^H, d_{TS}^H$ and d_C^H distances are compatible.

In Jablonski (2012) complex measure of quaternion distance is proposed. It is determined by a product of distances raised to α, β and γ powers corresponding to rotations as well as their first and second derivatives calculated according to similar formula to (16) (the symbols
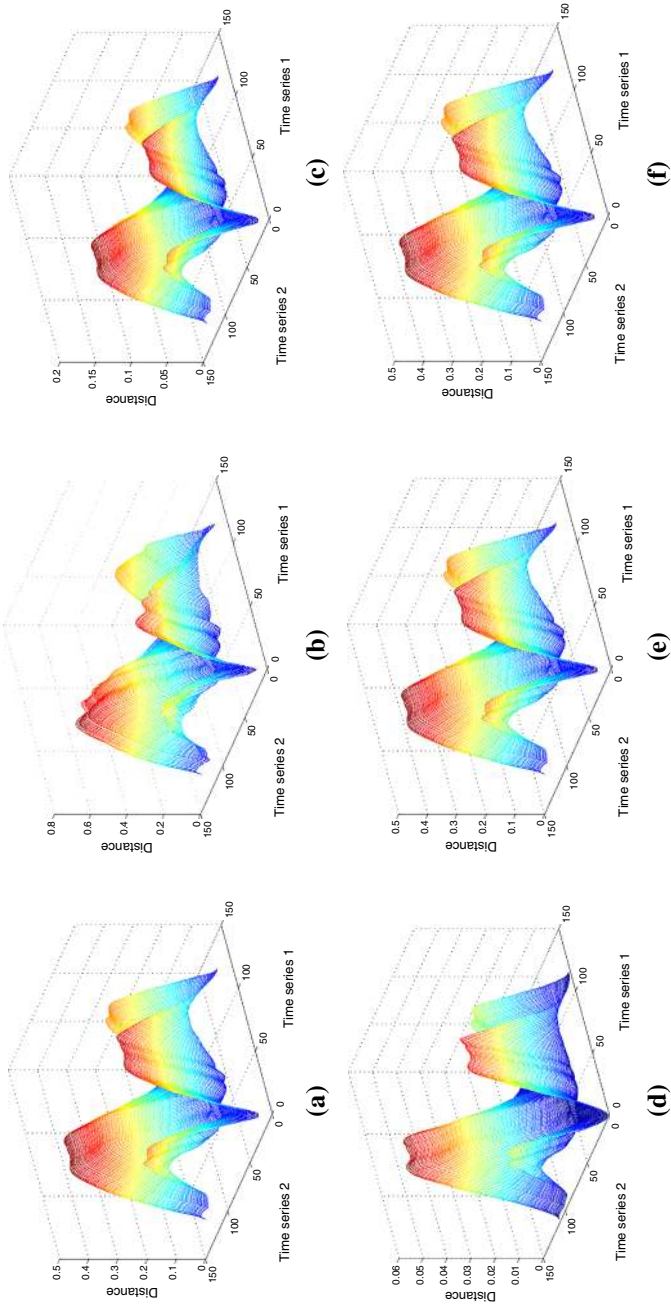
Fig. 3 Different distance functions applied to the comparison of left hip joint movements of two randomly chosen gaits performed by separate subjects. **a** d_{Euc}. **b** d_{Man}. **c** $d_{geodesic}^H$. **d** d_{cosine}^H. **e** d_{TS}^H. **f** d_F
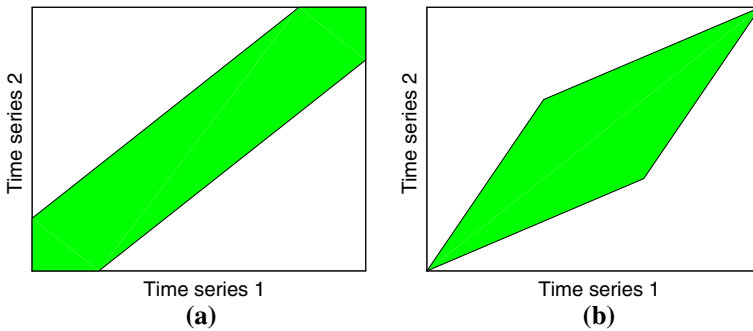
Fig. 4 Global constraints on time warping labeled by green color. **a** Neighborhood of constant size (Sakoe and Chuba 1978). **b** Neighborhood of variable size (Itakura 1990) (Color figure online)

α, β and γ in Jablonski (2012) do not mean Euler angles). However, another problem arises—how to set α, β and γ parameters? What is more, it does not allow for clear assessment of individual features related to joint rotations and dynamics of their movements.

2.2 DTW modifications

There are numerous modifications of the basic DTW algorithm. The first group is related to reduction of computational requirements. Classical implementation of DTW has $O(n \cdot m)$ complexity, where n and m correspond to the lengths of compared motion sequences. In fast DTW approaches, a global constraint is imposed, which limits how far a warping path may stray from the diagonal (Keogh and Ratanamahatana 2005). By default, neighborhood of a diagonal of a specified, constant size is applied as shown in Fig. 4a (Sakoe and Chuba 1978). However, it is also possible to construct neighborhood with variable width depending on a distance to the center of cost matrix as visualized in Fig. 4b (Itakura 1990). There is also a research on local constraints on the set of alternative steps (Keogh and Ratanamahatana 2005), where only specified shapes of the path are allowed. Independently of benefits related to lower computational complexity, global and local constraints prevent pathological warpings, where a relatively small section of one sequence is mapped onto a relatively large section of another (Keogh and Ratanamahatana 2005).

In addition, to speed up the computations some approximation techniques as for instance lower bounding or multiscale analysis are utilized (Muller 2007; Chu et al. 2002; Keogh and Ratanamahatana 2005). There are also GPU and FPGA accelerated implementations of DTW (Sart et al. 2010).

In another significant modification a time domain is warped independently for different body configuration parameters with a multidimensional DTW. In such a case dissimilarity tensor is described by $2 \cdot k$ different indexes and it has $(n \cdot m)^k$ elements, where k denotes the number of considered body parameters. The subsequent elements $(i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_k)$ contain data of dissimilarity corresponding to time instants i_1, i_2 and i_k of the first, second and k-th parameter of the first sequence and, similarly, j_1, j_2 and j_k of the second one. However, the computational complexity and memory requirements are growing exponentially in respect to k. For instance if in the case of motion data of 100 elements one-dimensional DTW is computed in 0.1s and it requires 78kB of memory, then two- and three-dimensional ones are determined in 16.5 min and 115 days, respectively, and dissimilarity tensor needs 762 MB and 7 TB of memory. There is one more important short-

coming of multidimensional DTW. It leads to unintuitive desynchronization of movements of human body parts, because their time domains are warped independently.

In the case of low capture frequency continuous version of DTW (Munich and Perona 1999) should be used. Another kind of improvement is canonical time warping, which determines the path with minimal cost from possible linear combinations of the body parameters (Zhou and De la Torre 2009, 2016). To find some specific short-lasting movement during the whole motion the similarity matrix has to be modified by assigning the zero distance value at the borders of the matrix as presented in Martin et al. (2010).

3 DTW minimum distance classifier

In a baseline minimum distance classification approach every class is represented by its prototype. By default it is stated to be a centroid of class instances belonging to a training set. To apply the classification scheme for motion data, a method of averaging time sequences has to be established. It turns out, that DTW alignment can be utilized to perform the task. It is sufficient to match time instants to reference motion sequence and finally, to calculate mean values of matched poses as described in Izakian et al. (2015).

However, it is straightforward only in the case of rotations coded by Euler angles for which the average can be calculated in classical way. If quaternions are applied, computing mean for their successive components is unworkable, because it does not satisfy the unit length requirement. Thus, there is another challenge to face—how to determine quaternion representing mean rotation? We have chosen two approaches from the literature.

The mean is defined as the value which has minimum total quadratic distance d_x to considered samples set $q_1, q_2, q_3, .., q_n$:

$$q_{mean}(q_1, q_2, q_3, .., q_n) = \arg\min_{|q|=1} \sum_{i=1}^{n} d_x^2(q, q_i) \tag{18}$$

Listing 2 Calculating the average of time series based on DTW alignment (Izakian et al. 2015)

```
a:=a₁; a₂; ...; aₙ  {the initial average of time series with length n}

x₁ := x₁₁; x₁₂; ...; x₁m₁  {the first time series of length m₁}
x₂ := x₂₁; x₂₂; ...; x₂m₂  {the second time series of length m₂}
}
x_N := x_N1; x_N2; ...; x_Nm_N  {the N–th time series of length m_N}

y := y₁; y₂; ...; yₙ   {Output: average of time series x₁; x₂; ...; x_N}

DTWavg(a, x₁; x₂; ...; x_N):
    Let assosTab := [ϕ; ϕ; ...; ϕ] {an empty vector with n elements}
    for k:=1,2,...,N do
        path:=path matrix in DTW(a, xₖ)
        i := n
        j := mₖ
        while i ≥ 1 and j ≥ 1 do
            assosTabᵢ := assosTabᵢ ∪ xₖⱼ;
            (i, j) := pathᵢ,ⱼ
        end
    end
    for i:=1,2,...,n do
        yᵢ := average of elements located in assosTabᵢ
```

Gramkow (2001) introduced the approach for the average related to the linearized geodesic distance $d_{geodesic}$ on the basis of Taylor expansion of cosine function. In such case the minimization problem from Eq. (18) is transformed to maximization of:

$$q_{mean}(q_1, q_2, q_3, .., q_n) = \arg\max_{|q|=1} q \cdot \sum_{i=1}^{n} q_i \tag{19}$$

Thus, finally the mean quaternion is the barycentric mean with renormalization:

$$q_{mean}(q_1, q_2, q_3, .., q_n) = \frac{\sum_{i=1}^{n} q_i}{\left|\sum_{i=1}^{n} q_i\right|} \tag{20}$$

In another approach Johnson (2002) proposed the method of finding the mean for quaternion samples q_i and cosine distance function d_{cosine}. The solution is based on the implicit constraint equation with Lagrange multiplier. The final algorithm is as follows:

---

1. Let the $4 \times N$ matrix Q be the data matrix with column i being the 4-vector for the i-th quaternion's components.
2. Let $S = Q \cdot Q^T$.
3. Let e_i be an eigenvector of S with real eigenvalue a_i.
4. Choose one of the two eigenvectors $\pm e_i$ associated with the maximal eigenvalue a_i as the estimate of the mean.

In a recognition mode of minimum distance classifier, the class with the closest prototype to the analyzed instance is determined. In our case, the prototypes are averaged time sequences which can be called human gait patterns. Hence, once again DTW can be utilized to approximate necessary dissimilarities between classified motion sequences and the prototypes.

4 Exploration of pose configuration space

In the basic approach DTW classification takes into consideration all the parameters of a skeleton model. However, some joints can have stronger impact on final pose comparison, while the influence of the others is much smaller or even imperceptible. Thus, Eq. (1) can be extended with weighted average:

$$d(P_1, P_2) = \frac{\sum_{joint} w_{joint} \cdot d_{rotation}(P_1(joint), P_2(joint))}{\sum_{joint} w_{joint}} \tag{21}$$

In the case when weights w_{joint} are binary, which means they can take only $\{0, 1\}$ values, it leads to a selection of pose configuration parameters.

The considered pose space contains 23 different rotations corresponding to skeleton segments and its global rotation, coded by Euler angles or unit quaternions. In a selection stage the subsets of rotations are discovered in respect to obtained accuracy of the classification. In general, there are two major challenges of selection tasks. The first one is a search strategy. In most cases exhaustive search which takes into consideration all possible combinations of subsets is unworkable, because the problem is NP-complete. For motion data with 23 different rotations there are more than eight million of different subsets.

In the most basic search approach every parameter is examined separately, which allows to construct their ranking. In a proper selection only top ranked attributes are taken into account. The main drawback of such an approach occurs in the case of attributes, which are discriminative only if considered together with others.

One of the most often used search techniques in selection tasks is a greedy hill climbing procedure (Witten and Frank 2005) with forward or backward directions. In such a case the search starts with empty or complete subsets and in successive iterations the best first attributes are added or removed depending on the direction. To avoid a convergence to the local maxima, more non-improving nodes are considered before terminating computations and the bidirectional search is carried on.

Other frequently used approaches to explore feature spaces are based on genetic algorithms. A single individual represents subset of features coded by binary genes. In the preliminary stage randomly chosen initial population is generated. Subsequent populations are determined on the basis of default genetic operators as mutation, crossover and selection in fixed number of iterations. It is also possible to utilize any other heuristic search algorithm to explore attributes spaces, as for instance IWO (Invasive Weed Optimization) (Mehrabian and Lucas 2006) or its extended variant (exIWO) (Josiński et al. 2014).

What is more, w_{joint} coefficients (21) can be real valued, which leads to fuzzy selection. In such a case only genetic search is directly applicable. It requires to code the coefficients with numeric genes.

The second challenge of selection is the way subsets are evaluated. In the most obvious case usually called wrapper approach (Witten and Frank 2005), it is sufficient to perform classification experiment, which takes only selected rotations into consideration. The overall assessment corresponds to the obtained accuracy of classification.

5 Collected gait database

There are some publicly available motion challenge datasets. For instance, TUM Kitchen Data Set is a collection of activity sequences captured in a kitchen environment equipped with multiple complementary sensors (Tenorth et al. 2009). It provides both the original video sequences and full body markerless motion capture data with action labels. MSR Action3D Dataset prepared by Microsoft Research, contains low resolution depth map sequences with 20 different action types performed by 10 subjects (Li et al. 2010). It is possible to estimate skeleton attributes on the basis of depth map representation. HDM dataset contains more than 3 h of systematically recorded and well documented MoCap data (Müller et al. 2007). Among others there are such activities as walking, running, jumping, sitting, laying down, grabbing, depositing, dancing, kicking and punching. Carnegie Mellon University Multimodal Activity Database (CMU-MMAC) stores multimodal (audio, video, accelerations, motion capture), long-lasting samples of human behaviors related to cooking several meals (De la Torre Frade et al. 2008). A very large gait database with instances coming from 744 subjects is introduced in Trung et al. (2014). Nevertheless, the data were captured by only three inertial measurement units (IMU) located around the waist, which is insufficient to determine all joint movements. The aforementioned datasets allow to obtain only coarse (TUM, MSR Action3D) or even incomplete (IMU database) skeleton data and in the case of CMU-MMAC and HDM they are not precisely divided and labeled. Except for the last two datasets, the recognition of different everyday life activities is taken into consideration.
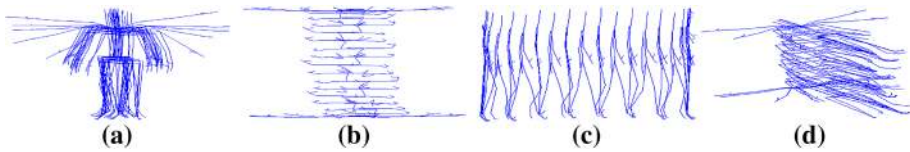
Fig. 5 Example gait. **a** Front view. **b** Top view. **c** Side view. **d** Mixed view

Moreover, the authors of Balazia and Plataniotis (2017) extracted a large number of gait cycles from the general MoCap database from CMU http://mocap.cs.cmu.edu. The data come from 48 subjects, which is very close to our expectations. However, the extracted database also has important drawbacks, which can have an influence on gait recognition. For instance, the number of gait cycles differs significantly across subsequent subjects, there is no information about calibration of motion capture system as well as it is not focused on typical gait—for selected subjects there are instances labeled as walk with anger, frustration or walk stealthily, muscular, heavyset person's walk, whistle, walk jauntily, navigate—walk forward, backward, on a diagonal, etc.

This is the reason why we collected our own motion capture database. It is related to the human gait identification problem, which means that persons are recognized on how they walk. Such a classification is certainly more challenging in comparison to discrimination of different activities, because of very similar movements performed by every subject. What is more, the problem of gait identification is still gaining more practical importance, because of surveillance systems and security issues.

The acquisition process was carried out in the Human Motion Laboratory of the Polish-Japanese Academy of Information Technology (http://bytom.pja.edu.pl) equipped with the Vicon hardware and software. The default 100 Hz registration frequency rate was chosen. The database contains in total 436 gait instances of 30 different males at the age of 20–40 years. The gait route is a straight line, every recording starts and terminates with a reference T-pose, because of calibration requirements of the Vicon Blade software. The example data with different view perspective are shown in Fig. 5.

The default skeleton model of the Vicon Blade is applied, as presented in Fig. 6. It consists of 22 joints described by Euler angles triplets, global rotation and three-dimensional translation vector. In total it gives 72-dimensional pose space. However, to minimize the influence of direction and location of gait on recognition performance, the translation attributes are not taken into consideration in the classification, they are removed in preprocessing stage.

Gait is defined as cyclic coordination of movements which results in human locomotion (Boyd and Little 2005). Thus, the most representative gait cycle (the main cycle) should be extracted, which contains two middle adjacent steps performed by left and right lower limbs consecutively. To detect the cycle, tracking of distances between the left and right ankles is carried out, as it is visualized in Fig. 7a. The subsequent steps are separated by time instants with local maxima. To distinguish steps of the left and right limbs, distances of the ankles to their start and end points are compared. The left step starts and the right one terminates when the distance of the left ankle to its start point is lower than the analogous distance of the right ankle. At the same time, the distance of the left ankle to its end point is greater than the corresponding distance of the right ankle. An example of a detected main cycle is presented in Fig. 7b.

In the case of gait data cropped to a detected main cycle the classification seems to be more reliable, because directly corresponding motion data are analyzed. What is more, the T-pose which is quite untypical during a normal gait, is removed and, consequently, not taken into consideration by a subsequent recognition stage.

Fig. 6 Skeleton model
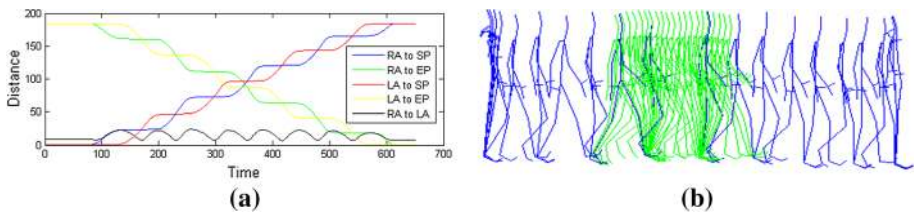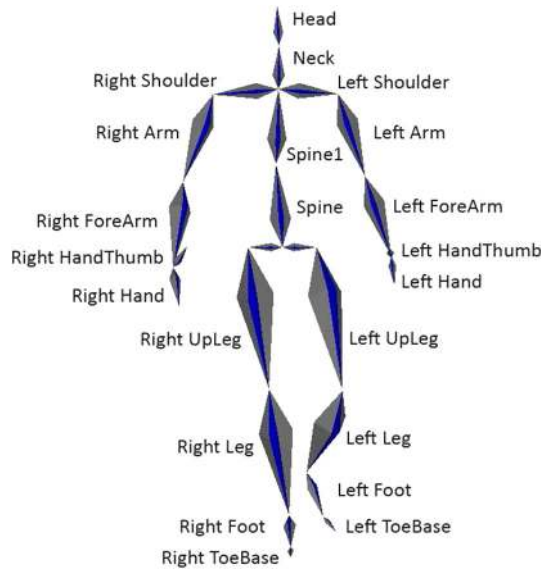
Fig. 7 Main cycle detection for example gait data. **a** Distances between: *RA* right ankle, *LA* left ankle, *SP* start point, *EP* end point. **b** Detected main cycle labeled with green color (Color figure online)

To minimize the influence of the strict markers locations and calibration process of the motion capture measurements on the final recognition, the acquisition is divided into sessions. Markers are attached on a human body and the calibration is carried out independently in every session.

The database contains highly precised, full body motion capture data related to the human gait identification problem. Thus, the clear assessment and tuning of a proposed classification method without a significant influence of an acquisition noise is feasible. To the authors' knowledge, it is the only database with such properties.

6 Results

Exemplary constructed warping paths and dissimilarity matrices for left hip joints movements of two randomly chosen gaits of different subjects, which utilize quaternion geodesic distance function are visualized in Fig. 8. As it is expected the path is adjusted to the regions of the matrix with the lowest distance values mainly located in the surroundings of its diagonal. The boundaries between different gait phases are strongest in the case of raw angle data.
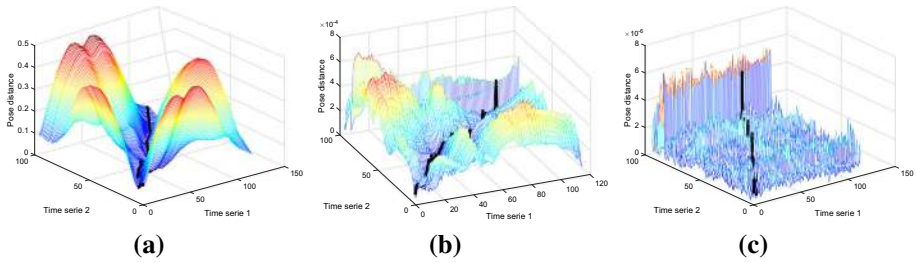
Fig. 8 DTW warping paths of the left hip joint movements of two randomly chosen gait instances (d_{geo}). **a** Angles. **b** Velocities. **c** Accelerations

Differential filtering enhances the acquisition noise, which in particular can be noticed for angular accelerations.

6.1 DTW classification on the basis of complete set of joints

In the basic validation experiment the classification is carried out with respect to a complete set of joints with uniform weights as it is presented by the Eq. (1). Two strategies of division of collected gait database into the training and testing parts are considered. In the first one, called TT, the database is split according to sessions into two subsets of approximately equal size—258 training and 178 testing samples. Thus, all measurements of a given session belong either to training set or to testing one. It means that recognition of any gait instance is carried out on the basis of data acquired in separate sessions. If there is an odd number of sessions of a given subject, the training set consists of one session more than the testing one. It is the reason for greater training set size. In the second division strategy (CV10) ten-fold cross validation is utilized. In this case the folds are generated randomly without respect to motion capture sessions. Thus, gait instances are also recognized by comparing them to data of the same sessions.

The obtained results for DTW kNN classifier representing correct classification rates (CCR), meaning of percentage of properly classified gait instances, are shown in Table 1 and in Fig. 9. As expected, calibration stage and strict markers' location on a human body have a great impact on performance of gait identification. In the case of CV10 strategy achieved values of recognition accuracies are on average 10% higher than for TT. What is more, the best CV10 precision is 99.77% which means only one misrecognized instance of 436 in contrast to 90.45% of TT with 17 incorrectly identified gait samples out of 178. It is mainly caused by slightly different orientation of local coordinate systems of subsequent joints, which are determined by motion capture system during aforementioned preliminary calibration stage. Estimated directions of flexion, abduction and rotation of bone segments are strictly related to precision of specified set of calibration movements performed by a human. It is confirmed by calculated average standard deviations of Euler angles values in two variants—in the first one standard deviations are calculated across motion capture sessions and in the second one across complete actor's data. The standard deviations of the first variant are usually a few degrees smaller in comparison to the second variant as shown in Table 2, which can be explained just by slightly different orientations of local coordinate systems.

This is extremely important in the case of direct analysis of raw angle data, because even slight differences are cumulated across warping path regardless of performed movements. If, in particular, velocities are compared, dynamics of the movements is crucial. For instance in

Table 1 DTW classification based on complete set of joints in respect to different number k of nearest neighbors and rotation distance functions

kNN	d_{Euc}		d_{Man}		$d^H_{geodesic}$		d^H_{cosine}		d^H_{TS}		d^H_C	
	TT	CV10	TT	CV10	TT	CV10	TT	CV10	TT	CV10	TT	CV10
Angles												
1NN	87.64	99.77	85.96	99.54	90.45	99.77	84.27	98.85	90.45	99.77	84.27	98.85
3NN	84.27	98.62	82.58	97.94	87.08	98.39	82.02	95.87	87.08	98.62	82.02	97.48
5NN	82.58	97.71	83.15	94.50	86.52	97.94	82.58	94.50	85.39	97.48	82.58	94.50
Velocities												
1NN	87.08	91.28	87.08	91.74	88.20	91.97	88.76	90.60	88.20	91.97	88.76	90.60
3NN	88.76	89.45	88.76	90.37	89.89	89.45	86.52	88.99	89.89	89.45	86.52	88.99
5NN	85.39	89.91	87.08	90.60	87.08	90.14	84.83	88.53	87.08	90.14	84.83	88.53
Acceler.												
1NN	49.44	62.39	49.44	60.55	46.63	61.70	50.00	60.78	46.63	61.70	50.00	60.78
3NN	41.57	52.98	38.76	50.92	43.26	51.83	43.82	49.08	43.26	51.83	43.82	49.08
5NN	38.20	48.85	34.27	46.10	38.76	48.85	41.57	49.31	38.76	50.00	41.57	49.31

(a)   (b)   (c)
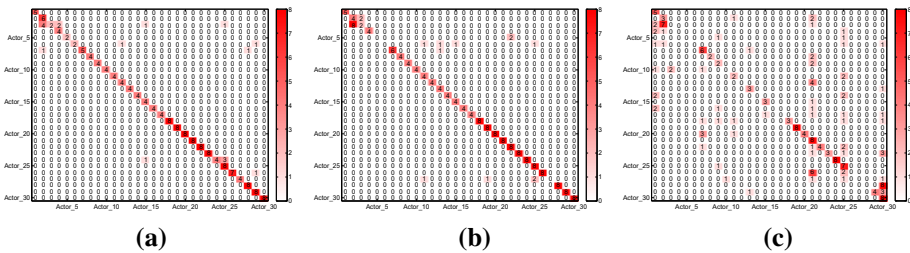
Fig. 9 Confusion matrices ($d_{geodesic}$, $1NN$, TT). **a** Angles. **b** Velocities. **c** Accelerations

Table 2 Average standard deviations of Euler angles values calculated separately for different sessions or for complete gait data of different actors

Segment	RX		RY		RZ	
	Actor	Sess.	Actor	Sess.	Actor	Sess.
RHand	16.65	6.70	7.88	4.70	14.19	6.85
LHand	13.91	6.79	7.88	4.79	12.37	6.46
Neck	7.49	2.42	5.89	2.24	5.71	2.18
Head	10.30	6.45	5.74	2.15	7.31	3.63
RShoulder	7.71	3.94	5.21	3.02	5.70	3.28
LFoot	10.84	9.89	7.66	6.33	9.56	6.30
LArm	13.03	9.85	8.44	6.15	8.94	6.68
Root	7.74	4.62	4.33	4.00	7.73	5.79
LShoulder	5.97	3.31	4.85	2.68	4.25	2.67
LToeBase	9.81	7.31	0.00	0.00	0.00	0.00

the case of static pose sequences the orientation does not have any impact. Thus, total influence of calibration is weaker, which explains more consistent results for both considered strategies. The obtained results for CV10 are obviously artificially inflated, because classification takes

into consideration not only individual traits, but also utilizes specific acquisition features, which makes the task more simpler. However on the other hand, in the case of TT strategy an imprecise estimation of local coordinates systems does not allow to directly compare only gait characteristics; thus, the results seem to be undervalued.

The best performance of recognition is obtained for DTW applied to raw angles. However, what is a bit surprising—angular velocities are only slightly worse and they are much more resistant to calibration. Noticeably the lowest precision achieved by accelerations is probably caused by acquisition noise, which is enhanced twice by double differential filtering. The misclassification is distributed quite randomly across different actors as shown in Fig. 9. However, there are some actors with very poor specificity or sensitivity ratios as for instance Actor_5, Actor_6 and Actor_27 in the case of velocities and Actor_4, Actor_5, Actor_6, Actor_21, Actor_25, Actor_26, Actor_27 and Actor_28 for accelerations.

There is also a dependency of applied rotation distance function on performance of classification. The best accuracy is achieved by quaternion geodesic $d_{geodesic}^{H}$ measure, but tangent space based one d_{TS}^{H} is only slightly worse. Cosine d_{cosine}^{H} and the distance which utilizes conjugate quaternion d_{C}^{H}, the ones which enhance dissimilarity between more similar rotations (Fig. 3), are poorest in the case of raw angles analysis, but their performances improve for angular velocities. The functions directly applicable to Euler angles – Euclidean d_{Euc} and Manhattan d_{Man} metrics are very similar. They give noticeably lower precision of recognition, especially if compared to geodesic and tangent space ones. In all cases 1NN (the nearest neighbor) classifier obtains the best accuracy of classification, which can be explained by limited representativeness of the training sets. However, 3NN and 5NN are only about 4–6% worse and they are still acceptable.

The performances achieved by DTW minimum distance classifier (DTW MDC) are depicted in Table 3. The computations were carried out with aforementioned distance functions (d_{Euc}, d_{Man}, d_{geo}, d_{cosine}) and with the mean calculated in a classical way for Euler angles (Mean EA) as well as renormalized barycentric (QBCM) and Johnson means (QJM) for unit quaternions.

The accuracies of recognition are clearly lower in comparison to DTW kNN classifier, especially if raw rotation data are taken into account. It is not very surprising that instance based learning (kNN) is more efficient in motion data analysis. It can be explained by numerous variations appearing between gait cycles, which are difficult to preserve by a single averaged gait pattern. However, differences are much lower for angular velocities and in the case of accelerations minimum distance classification is even more effective. Once again it is consistent with our expectations. As mentioned before, differential filtering enhances acquisition noise and just for such data, using mean patterns usually reduces the noise and improves the performance of recognition. When comparing applied approaches of averaging rotational data, it can be noticed that classical mean used for Euler angles is more efficient for raw data, while the opposite observation concerns angular velocities and accelerations. For barycentric mean approach it can be partially explained by substantially narrowed ranges of rotations obtained after differential filtering—assumed linearization of geodesic distance is much more adequate for very similar rotations. The DTW MDC performance also relies on utilized pose distance function. In the case of rotational data d_{Man} gives the best results, even if quaternion based averaging techniques are chosen. On the other hand, for velocities and accelerations d_{cosine}^{H} achieves the highest precision of recognition. It is to some extent consistent with DTW kNN, when there is a similar improvement obtained by d_{cosine}^{H} if differential filtering is carried out.

Table 3 DTW Minimum distance classifier results for TT division strategy

	Mean EA	QBCM	QJM
Angles			
d_{Euc}	74.16	60.11	60.11
d_{Man}	76.40	67.42	65.73
$d_{geodesic}^{H}$	73.03	64.61	61.24
d_{cosine}^{H}	69.10	61.80	63.48
Velocities			
d_{Euc}	78.65	79.78	79.78
d_{Man}	79.78	80.90	80.90
$d_{geodesic}^{H}$	78.65	79.78	79.78
d_{cosine}^{H}	80.34	84.27	84.27
Accelerations			
d_{Euc}	42.70	43.82	43.82
d_{Man}	39.89	39.89	39.89
$d_{geodesic}^{H}$	42.70	43.26	43.82
d_{cosine}^{H}	53.37	53.93	53.93

Because of the highest obtained precision of classification, 1NN classifier is chosen to further computations related to joint selection issues.

6.2 Joint selection

The feature selection, which is a part of supervised learning, has to be carried out only on the basis of training data. If testing instances are also used during the learning stage, it means that classes of recognized objects are known, which is not adequate to real deployments. Hence, it leads to an overestimation of obtained classification results. Thus, the training set is split once again according to sessions into two parts, called initial training and testing sets (TT_0). They are directly utilized in validation stage of feature selection search and ranking construction.

On the basis of classification carried out for separate joints of TT_0 subsets ranking is determined, as presented in Table 4. In general, individual features of joints movements depend on their activity during gait. Thus, hip, knee, ankle, shoulder and arm joints are usually top ranked. Relatively high precision achieved by root and spine data is probably caused by some posture irregularities or even abnormalities. Mostly, dynamics of the movements expressed by angular velocities and even accelerations in the case of the analysis of single joints allows to obtain noticeably better results if compared to raw angle data. Surprisingly, top selected single joints perform similarly or even better than complete sets of them in that case. It can be explained by strong discriminative, individual features, which refer to dynamics of movements, and enhanced acquisition noise by differential filtering. Captured rotations of some of the joints which are not engaged in locomotion and perform free and uncontrolled movements, contain mainly noise. It disturbs strongly the rest of data. Thus, a complete set of joints with generally stronger discriminative traits gives lower accuracy because of the influence of the noise. The hypothesis can be stated—there are some subsets of

Table 4 Joint ranking (1NN, $d_{geodesic}^H$)

Angles			Velocities			Accelerations		
Segment	TT_0	TT	Segment	TT_0	TT	Segment	TT_0	TT
RightUpLeg	65.56	65.17	RightFoot	91.11	90.45	RightFoot	75.56	74.16
LeftUpLeg	63.33	68.54	LeftFoot	86.67	91.01	LeftFoot	67.78	74.72
RightLeg	50.00	48.88	RightUpLeg	85.56	90.45	RightUpLeg	51.11	48.88
Root	48.89	50.00	Root	84.44	95.51	Root	48.89	42.70
LeftArm	40.00	43.82	LeftUpLeg	84.44	93.82	LeftUpLeg	48.89	46.63
LeftShoulder	37.78	26.40	Spine	81.11	87.08	LeftLeg	44.44	50.56
RightForeArm	33.33	26.40	Spine1	71.11	82.58	RightLeg	40.00	37.64
LeftLeg	33.33	44.94	LeftArm	71.11	72.47	RightToeBase	37.78	34.27
Spine	33.33	38.76	RightArm	65.56	66.85	Spine1	36.67	39.33
RightArm	33.33	45.51	RightShoulder	63.33	61.80	LeftToeBase	34.44	36.52
Spine1	33.33	41.01	LeftShoulder	61.11	66.29	Spine	33.33	39.89
LeftFoot	31.11	56.18	LeftToeBase	57.78	47.75	LeftForeArm	27.78	20.79
RightFoot	31.11	42.13	RightLeg	57.78	53.37	LeftArm	25.56	23.60
Head	26.67	14.61	LeftLeg	50.00	61.80	LeftShoulder	22.22	17.42
LeftForeArm	26.67	23.60	RightToeBase	44.44	48.88	RightShoulder	18.89	17.98
LeftHand	24.44	22.47	LeftForeArm	41.11	36.52	RightArm	18.89	28.09
Neck	24.44	17.42	Neck	37.78	40.45	RightForeArm	16.67	16.29
RightShoulder	18.89	22.47	LeftHand	36.67	40.45	Head	15.56	8.99
LeftToeBase	14.44	20.79	Head	34.44	32.58	Neck	14.44	11.24
RightHand	14.44	14.61	RightHand	32.22	36.52	LeftHand	10.00	8.43
RightToeBase	13.33	26.40	RightForeArm	26.67	32.02	RightHand	10.00	13.48
LeftHandThumb	4.44	2.81	LeftHandThumb	4.44	2.81	LeftHandThumb	4.44	2.81
RightHandThumb	4.44	2.81	RightHandThumb	4.44	2.81	RightHandThumb	4.44	2.81

joints with noticeably better performance. This hypothesis will be experimentally examined through feature selection.

However, even for raw angle data, classification based on single joints is sufficient to achieve satisfactory results. In particular, segments with greater range of movements, for instance hips and shoulders or those related to global human posture, are top ranked.

The results of subsequent stages containing data of joints added or removed by forward and backward hill climbing procedures are shown in Tables 5 and 6, respectively. As it was mentioned, before search is carried out only on the basis of the training data. Thus, column TT_0 gives information about recognition accuracy with respect to the hill climbing validation and TT—to the proper testing set. The remaining, final feature selection results for d_{Euc} and d_{TS}^H functions are presented in Table 7. Though in validation TT_0 set is utilized, also analyzed subsets of segments in successive iterations of hill climbing search with the greatest precision in respect to TT validation are presented. It is shown in the column named "Best" in Table 7.

The selection certainly allows to reduce considered pose configuration parameters without the loss of the precision of classification. In the case of raw angle data, $d_{geodesic}^H$ function and default stop criteria of the hill climbing procedure which terminates calculations after the first non improving node, pose configuration space is diminished to nine joints with the same accuracy of gait identification. As expected, joint selection improves noticeably the performance of recognition if angular velocities are taken into consideration by DTW. In such a case CCR ratio is raised to 94.44% with respect to TT sets, but there is also a combination of five joints which achieves 97.19% precision with only 5 misclassified gaits of 178. Unfortunately, it does not correspond to globally or locally best result of validation sets TT_0. Thus, as mentioned before, it should not be considered to be properly discovered joints subset. In the case of angular accelerations, joint selection improves performance only a insignificantly. Once again it can be explained mainly by an enhanced acquisition noise.

Backward search direction is outperformed by the forward one—the determined subsets are of much greater size and the obtained final recognition accuracies are clearly lower.

The results differ considerably depending on the training and testing sets, as represented by TT_0 and TT columns. It is caused by an overfitting phenomenon of the selection process. The performance is adjusted to the hill climbing validation data and it becomes quite different in the case of unseen motion instances. However, this drawback is partially balanced by much greater size and representativeness of the training set of TT validation. Thus, the mean performances in both cases are very similar.

A genetic optimization algorithm was also used in the exploration of joints space. In a default approach binary genes are applied to code existence of successive skeleton segments in considered subsets. It corresponds to {0, 1} values of w_{joint} coefficients in Eq. (21). However, such limitation is assumed only to adapt classical feature selection search strategies to the considered problem of exploration of pose configuration space and to reduce computations. In the case of DTW classification coefficients can also be real valued, which results in weighted average applied to the calculated dissimilarity matrix. Thus, every joint has an assigned weight, which indicates its importance in global pose comparison. It leads to a fuzzy feature selection. The problem to determine w_{joints} can be solved just by genetic algorithm. It is sufficient to implement individuals with numeric floating genes, which correspond to w_{joint} values. Moreover, fuzzy selection allows to mix pose attributes related to raw angles, angular velocities and accelerations in single DTW classification. It is directly unworkable in the case of binary selection, because of quite different scales of those attributes as shown in Fig. 8. It would lead to complete domination of raw angles over velocities and accelerations in pose similarity assessment. Only real valued weights can be adjusted depending simultaneously on scale and explored individual features.

Table 5 Forward hill climbing results (1NN, $d_{geodesic}^H$)

	Angles			Velocities			Accelerations		
	Segments	TT_0	TT	Segments	TT_0	TT	Segments	TT_0	TT
1	RightUpLeg	65.56	65.17	RightFoot	91.11	90.45	RightFoot	75.56	74.16
2	RightLeg	77.78	74.16	RightUpLeg	94.44	94.94	LeftLeg	76.67	76.97
3	LeftUpLeg	80.00	78.09	RightHandThumb	94.44	94.94	RightLeg	77.78	74.16
4	RightForeArm	82.22	80.90	LeftHandThumb	94.44	94.94	RightHandThumb	77.78	74.16
5	LeftShoulder	85.56	75.84	RightLeg	93.33	97.19	LeftHandThumb	77.78	74.16
6	Spine1	87.78	83.71	RightShoulder	92.22	94.38	LeftFoot	75.56	78.65
7	Root	90.00	83.71	LeftUpLeg	93.33	96.63	RightToeBase	75.56	74.16
8	LeftFoot	91.11	90.45	RightToeBase	93.33	93.82	Root	74.44	74.72
9	Spine	93.33	90.45	Head	93.33	92.13	RightForeArm	71.11	73.60
10	RightHandThumb	93.33	90.45	Spine1	93.33	92.70	RightUpLeg	70.00	70.22
11	LeftHandThumb	93.33	90.45	Neck	93.33	92.13	Spine	71.11	68.54
12	LeftToeBase	92.22	89.33	Spine	93.33	92.70	LeftShoulder	68.89	64.61
13	LeftLeg	94.44	89.89	RightForeArm	92.22	93.26	LeftToeBase	67.78	73.03
14	RightFoot	92.22	89.89	LeftFoot	92.22	94.38	LeftUpLeg	66.67	71.35
15	RightArm	94.44	88.76	LeftShoulder	92.22	93.82	Spine1	65.56	70.22
16	RightShoulder	94.44	88.20	Root	92.22	93.82	RightArm	64.44	67.98
17	RightToeBase	93.33	87.08	LeftLeg	91.11	93.26	RightShoulder	63.33	66.85
18	LeftForeArm	94.44	88.76	LeftForeArm	90.00	91.57	LeftArm	61.11	62.92
19	Head	91.11	88.76	LeftArm	88.89	91.01	LeftForeArm	61.11	61.80
20	RightHand	92.22	88.76	LeftToeBase	88.89	92.13	LeftHand	57.78	56.18
21	LeftHand	87.78	89.33	RightHand	87.78	89.33	Head	54.44	56.18
22	LeftArm	90.00	91.01	RightArm	87.78	89.89	Neck	53.33	52.81
23	Neck	83.33	90.45	LeftHand	87.78	88.20	RightHand	47.78	46.63

Table 6 Backward hill climbing results (1NN, $d_{geodesic}^H$)

	Angles			Velocities			Accelerations		
	Segment	TT_0	TT	Segment	TT_0	TT	Segment	TT_0	TT
22	Neck	90.00	91.01	RightHand	87.78	91.01	Neck	55.56	53.37
21	Root	91.11	88.76	LeftLeg	88.89	90.45	Root	56.67	56.18
20	LeftHandThumb	91.11	88.76	RightLeg	88.89	88.76	LeftHandThumb	58.89	56.18
19	RightHandThumb	91.11	88.76	LeftArm	88.89	88.76	RightHandThumb	60.00	57.87
18	LeftForeArm	91.11	88.20	LeftForeArm	90.00	90.45	LeftForeArm	60.00	57.87
17	RightLeg	91.11	87.64	LeftHand	91.11	92.70	RightLeg	60.00	57.87
16	RightToeBase	90.00	86.52	RightArm	92.22	93.82	RightToeBase	60.00	57.87
15	LeftShoulder	88.89	89.33	Spine1	92.22	93.82	LeftShoulder	64.44	66.85
14	LeftUpLeg	88.89	86.52	LeftShoulder	92.22	93.82	LeftUpLeg	64.44	69.66
13	Spine	87.78	86.52	LeftToeBase	92.22	92.70	Spine	65.56	70.79
12	LeftLeg	85.56	85.96	RightForeArm	93.33	93.26	LeftLeg	66.67	71.35
11	Head	85.56	86.52	LeftHandThumb	93.33	93.26	Head	67.78	66.85
10	RightShoulder	85.56	85.39	LeftUpLeg	93.33	91.57	RightShoulder	68.89	64.61
9	LeftFoot	87.78	83.71	Root	93.33	91.01	LeftFoot	68.89	65.73
8	LeftToeBase	85.56	81.46	Spine	93.33	90.45	LeftToeBase	68.89	70.79
7	RightHand	84.44	84.83	RightShoulder	93.33	91.01	RightHand	71.11	72.47
6	LeftHand	81.11	82.02	RightHandThumb	93.33	91.01	LeftHand	71.11	73.60
5	LeftArm	86.67	84.83	Neck	92.22	91.01	LeftArm	72.22	73.60
4	RightForeArm	84.44	78.65	RightForeArm	92.22	90.45	RightForeArm	68.89	76.97
3	RightFoot	70.00	81.46	RightToeBase	92.22	93.26	RightFoot	73.33	73.60
2	RightArm	55.56	70.22	Head	94.44	94.94	RightArm	75.56	74.16
1	Spine1	65.56	65.17	RightUpLeg	91.11	90.45	Spine1	4.44	2.81

Table 7 Remaining selected feature selection results

Meth	Fun.	Par	Best	Segments	TT_0	TT
HCF	d_{Euc}	Ang	TT_0	LeftUpLeg; RightLeg; LeftLeg; RightForeArm; RightFoot; Spine1; Spine; RightHandThumb; LeftHandThumb; LeftShoulder; RightArm	92.22	84.83
HCF	d_{Euc}	Ang	TT	LeftUpLeg; RightLeg; LeftLeg; RightForeArm; RightFoot; Spine1; Spine	88.89	90.45
HCF	d_{Euc}	Vel.	TT_0	RightFoot; RightUpLeg; RightHandThumb; LeftHandThumb; RightShoulder; LeftArm; RightToeBase; RightLeg	94.44	89.33
HCF	d_{Euc}	Vel.	TT	RightFoot; RightUpLeg	93.33	93.26
HCF	d_{Euc}	Acc	TT_0 & TT	RightFoot; RightLeg	76.67	71.91
HCF	d_{TS}^{H}	Ang	TT_0	RightUpLeg; RightLeg; LeftUpLeg; RightForeArm; LeftShoulder; Spine1; root; LeftFoot; Spine; RightHandThumb; LeftHandThumb; LeftLeg; LeftToeBase	94.44	89.33
HCF	d_{TS}	Ang	TT	RightUpLeg; RightLeg; LeftUpLeg; RightForeArm; LeftShoulder; Spine1; root; LeftFoot; Spine; RightHandThumb; LeftHandThumb; LeftLeg; LeftToeBase; RightArm; RightFoot; LeftForeArm; RightToeBase; LeftArm; Head; RightHand; LeftHand; RightShoulder	88.89	92.13
HCF	d_{TS}^{H}	Vel	TT_0	RightFoot; RightUpLeg; RightHandThumb; LeftHandThumb	94.44	94.94
HCF	d_{TS}^{H}	Vel	TT	RightFoot; RightUpLeg; RightHandThumb; LeftHandThumb; RightLeg; RightShoulder; LeftUpLeg	93.33	96.63
HCF	d_{TS}^{H}	Acc	TT_0	RightFoot; LeftLeg; RightLeg	77.78	74.16
HCF	d_{TS}^{H}	Acc	TT	RightFoot; LeftLeg; RightLeg; LeftHandThumb; RightHandThumb; LeftFoot	75.56	78.65
GB	d_{geo}^{H}	Ang	TT_0 & TT	root; LeftToeBase; RightShoulder; RightForeArm; RightHandThumb; RightUpLeg; LeftLeg; Spine; LeftUpLeg; RightArm; RightFoot; Spine1	94.44	88.76
GB	d_{geo}^{H}	Vel	TT_0	LeftHandThumb; LeftFoot; RightForeArm; RightHandThumb; RightUpLeg; RightToeBase; RightFoot; LeftArm	94.44	91.01
GB	d_{geo}^{H}	Vel	TT	LeftFoot; Neck; RightUpLeg; RightFoot; Spine1; LeftArm	92.22	95.51
GF	d_{geo}^{H}	Ang	TT_0	LeftShoulder (2.66%); Head (3.41%); root (4.99%); LeftHandThumb (6.15%); LeftForeArm (4.41%); LeftHand (3.51%); LeftToeBase (1.01%); RightShoulder (5.82%); LeftFoot (1.36%); Neck (0.21%); RightForeArm (5.79%); RightHandThumb (4.6%); RightUpLeg (4.1%); RightToeBase (5.22%); LeftLeg (6.82%); Spine (1.5%); LeftUpLeg (10.52%); RightHand (1.42%); RightArm (6.55%); RightFoot (3.12%); Spine1 (8.57%); LeftArm (0.34%); RightLeg (7.92%)	94.44	92.70

Table 7 continued

Meth	Fun.	Par	Best	Segments	TT_0	TT
GF	d_{geo}^H	Vel	TT_0	LeftShoulder (3.57%); Head (2.76%); root (4.15%); LeftHandThumb (1.76%); LeftForeArm (2.77%); LeftHand (0.02%); LeftToeBase (3.43%); RightShoulder (2.98%); LeftFoot (6.1%); Neck (0.29%); RightForeArm (4.04%); RightHandThumb (6.71%); RightUpLeg (0.64%); RightToeBase (0.94%); LeftLeg (7.22%); Spine (8.82%); LeftUpLeg (9.16%); RightHand (4.78%); RightArm (6.16%); RightFoot (7.05%); Spine1 (8.26%); LeftArm (0.74%); RightLeg (7.67%)	94.44	93.82
GF	d_{geo}^H	All	TT_0	DX1LeftToeBase (0.59%); DX1RightToeBase (1.53%); DX1LeftShoulder (0.74%); root (0.47%); LeftHandThumb (0.79%); DX2LeftUpLeg (2.02%); LeftHand (0.56%); DX1RightHand (2.83%); DX1LeftFoot (2.15%); DX1Spine1 (2.38%); DX2root (0.14%); DX1LeftHandThumb (2.52%); DX1LeftForeArm (0.02%); DX1LeftArm (2.71%); DX2RightShoulder (2.52%); DX2Spine1 (2.91%); Head (1.19%); DX2Head (1.74%); LeftForeArm (0.64%); LeftToeBase (0.02%); DX1Spine (2.4%); DX1LeftHand (0.26%); LeftFoot (1.65%); DX1RightForeArm (0.62%); Neck (2.21%); RightForeArm (3%); RightToeBase (1.51%); DX2LeftHandThumb (1.71%); DX1RightShoulder (1.63%); DX2LeftFoot (2.42%); DX2LeftForeArm (0.16%); DX2Neck (0.12%); RightFoot (2.28%); DX2RightHand (2.31%); DX1RightArm (1.89%); DX2LeftToeBase (0.82%); DX2LeftHand (0.05%); DX2Spine (2.75%); RightHandThumb (0.63%); DX2RightLeg (0.39%); DX2LeftArm (0.48%); DX2RightToeBase (2.82%); RightHand (0.18%); RightArm (1.94%); LeftArm (0.18%); DX1Neck (1%); RightLeg (2.28%); DX2LeftLeg (1.52%); DX1root (1.93%); DX1RightLeg (2.49%); LeftShoulder (0.74%); DX1LeftLeg (0.48%); RightShoulder (0.54%); DX2RightForeArm (2.22%); DX2RightArm (2.01%); DX2LeftShoulder (0.24%); RightUpLeg (2.39%); DX2RightFoot (1.55%); LeftLeg (1.07%); DX1Head (0.19%); Spine (2.72%); LeftUpLeg (3.01%); DX2RightHandThumb (0.21%); DX1RightFoot (2.11%); Spine1 (2.9%); DX1RightUpLeg (2.89%); DX1LeftUpLeg (0.76%); DX2RightUpLeg (1.09%); DX1RightHandThumb (0.86%)]	94.44	91.57

HCF forward hill climbing, *GB* genetic binary, *GF* genetic fuzzy

Because of computation expensiveness of DTW calculations and to allow to generate several subsequent populations by a genetic procedure a limited number of individuals had to be applied. It was set to 20. The results obtained by binary and fuzzy genetic selection are shown in Table 7. The real valued weights are presented relatively as a percentage of total weights sum. Unfortunately, there is no further improvement of obtained CCR ratio, though more sophisticated search strategy was applied which analyzed fuzzy combinations of attributes related to different joints described by angles, velocities and accelerations. It can be explained by still raised computation complexity caused by continuous optimization problem and higher dimensionality of input space. It results in insufficient pose exploration. What is more, another crucial limitation is related to the training set of TT_0 validation taken into account by selection search. The best recognition accuracy obtained with respect to TT_0 by any considered classification method is only 94.44%. It is also achieved by genetic selection. However, it does not always correspond to the best accuracy of TT.

6.3 Classification on the basis of selected state-of-the-art methods

The basic classification related to motion sequences represented by Euler angles and with TT strategy of division of collected gait instances into training and testing sets was also carried out with usage of selected five state-of-the-art methods. Primarily, for comparative analysis, two classical techniques are chosen—Hidden Markov Models and Multilinear Principal Component Analysis (Lu et al. 2008). Moreover, recently published classification approaches are investigated, which are reported to outperform other nine (Balazia and Plataniotis 2017) and thirteen (Balazia and Sojka 2018) relevant methods of motion data classification in the gait-based human identification task.

The crucial challenge in HMM modeling is the proper choice of a method of pose probability estimation required for both training and recognition stages. It is extremely troublesome for motion capture data, because the estimated pose is described by 69 Euler angles. By default multivariate Gaussian distribution is assumed and parametric estimation is carried out. However it is unworkable for such high dimensional space, because of limited precision of floating point operations. The probability of generating specified observation sequence in successive time instants decreases, because it is multiplied by an estimated probability of the pose in a given state. As a result it is too close to zero value to be effectively represented by floating point data types. Even calculating the logarithm of the probability, which is quite common in HMM computations, does not solve the problem. On the basis of the performed numerical experiments we can notice that multivariate Gaussian distribution is an effective technique for maximum of two joints taken into consideration. Adding other ones to pose description worsen the classification accuracy.

This is the reason why we proposed to use some other pose estimation approaches. In the first one, similar to Naive Bayes classification scheme, it is assumed that pose configuration parameters are independent. Thus, inverse covariance matrix does not need to be determined. It is sufficient to estimate one-dimensional normal distributions and to calculate the product of the probabilities for subsequent Euler angles. In a more restrictive simplification, it is assumed that Euler angles have also the unit variance. It preserves impact of angle scale on final estimation, which seems to be reasonable—slight movements usually are indiscriminative. In the last approach kernel based estimation with Gaussian kernel and Euclidean metric with different values of smoothing parameter were applied.

The results obtained with respect to different number of hidden states and utilized estimation techniques are presented in Table 8. In general, they are noticeably worse than the

Table 8 HMM results (in computations JAHMM library https://github.com/KommuSoft/jahmm was used)

Pose estimation	Number of states								
	2	3	4	5	6	7	8	9	10
Naive Gaussian	63.48	63.48	63.48	63.48	63.48	63.48	63.48	63.48	63.48
Naive Gaussian unit variance	73.03	73.03	73.03	73.03	73.03	76.40	74.72	75.28	75.28
Kernel based	78.65	78.65	79.21	79.21	78.65	79.21	79.21	78.65	78.65

Table 9 MPCA kNN results

Q	L1	L2	Angle	MMD	ML1	ML2	MAD
80	61.24	70.79	66.29	39.89	64.04	66.85	64.61
85	64.61	71.91	73.03	51.69	63.48	66.85	66.85
90	66.85	73.03	76.97	65.17	65.73	67.98	71.91
91	67.98	74.72	78.09	68.54	65.73	73.60	74.16
92	77.53	75.84	79.78	60.67	84.83	82.02	74.16
93	76.40	76.97	77.53	66.85	84.27	82.58	74.72
94	78.65	75.84	79.21	68.54	84.27	82.58	78.65
95	80.34	76.97	80.34	71.35	87.08	84.83	81.46
96	80.90	76.97	80.90	76.97	90.45	85.39	83.15
97	85.96	76.97	81.46	83.71	89.33	85.39	86.52
98	85.96	76.97	82.02	85.96	91.01	87.08	87.64
99	88.20	79.21	82.58	89.89	92.70	87.64	89.89

ones achieved by means of DTW technique—the maximum recognition performance is less than 80%. We attribute it to the impact of aforementioned troublesome estimation of high dimensional pose space.

In the MPCA classification approach, time sequences of motion data, linearly scaled into equal size, are stated to be two-mode tensors. They are reduced by MPCA technique, which determines a multilinear projection that captures most of the original tensorial input variation Q (Lu et al. 2008). Afterwards, tensors are vectorized and classified by the nearest neighbor and the naive Bayes schemes. The MPCA performances are presented in Tables 9 and 10. Seven distance measures $L1, L2, Angle, MMD, ML1, ML2$ and MAD (Lu et al. 2008) were taken into consideration. The results are satisfactory, in the case of $ML1$ distance function the accuracy of recognition is even slightly better than for DTW operating on complete pose description, but still worse than for DTW with the introduced selection of pose attributes. It is worth to notice that the results obtained by Naive Bayes classifier are much worse. It is consistent with our previous observations that instance based learning is a more efficient technique for motion data classification. However, this time the difference is much greater than for DTW kNN and DTW MD classifiers.

In the next chosen state-of-the-art method, motion descriptors are determined and classified once again by baseline kNN and Naive Bayes classifiers. The descriptors contain the strike pose as the average of the poses at 25% and 75% of a gait cycle, and the clearance pose as the average of the poses at 50% and 100% gait cycle (Balazia and Plataniotis 2017). The obtained results are presented in Table 11. Primarily, they are clearly worse than reported in Balazia and Plataniotis (2017), even though the number of recognized individuals was

Table 10 MPCA Naive Bayes results

Q	Normal distribution	Kernel based estimation
80	32.02	35.96
85	29.21	37.64
90	32.02	38.20
91	32.58	38.20
92	46.07	53.37
93	47.19	53.93
94	50.56	55.06
95	61.80	62.36
96	61.80	64.61
97	62.36	67.98
98	62.92	67.42
99	64.04	66.85

Table 11 Results obtained by motion descriptors proposed in Balazia and Plataniotis (2017)

Classifier	CCR
1NN/L1	83.71
3NN/L1	76.97
5NN/L1	73.03
Naive Bayes (normal distribution)	52.25
Naive Bayes (kernel based estimation)	56.18

lower. It confirms that our dataset is more challenging as mentioned in the previous section, because it contains different motion capture sessions and is focused on typical gait performed with various speed. What is more, the results are also worse in comparison to DTW classification—a descriptor with only strike and clearance poses is oversimplified and insufficient for effective gait identification.

In the last investigated approach, proposed recently by Balazia and Sojka (2018), the features are extracted by linear transformation of the input space. The recognition is carried out on the basis of 3D data representing locations of markers attached to the human body in the global coordinate system as well as rotations of the skeleton joints. In the case of 3D data, to ensure invariance to person's position and walk direction, the center of the coordinate system is moved to root location, whereas the axes are adjusted to the walker's perspective. This was obtained by resetting root attributes according to global translation and rotation of the pose. In all gait sequences a single mean skeleton has been applied. Similarly to MPCA in the preprocessing stage, the motion sequences are also normalized with respect to the number of frames. Thus, it is feasible to construct common input space with pose attributes of successive time instances placed one by one. In the first proposed feature extraction approach that is called PCALDA, the key transformation is performed by supervised Fisher Linear Discriminant Analysis. However, to ensure numerical stability, the discussed method requires low dimensionality of the samples in comparison to the number of samples. In order to alleviate this, the measured data are initially projected to a lower dimensional space using Principal Component Analysis with compression ratio related to the number of recognized individuals. The second, a bit more sophisticated MMC feature extraction, enhances class separability, which similarly to the Vapnik's Support Vector Machines (SVM), is measured by the Maximum Margin Criterion. Thus, linear transformation that maximizes between-class

Table 12 Results obtained by generic motion descriptors proposed in Balazia and Sojka (2018)

	PCALDA			MMC		
	1NN	3NN	5NN	1NN	3NN	5NN
Markers	88.76	88.76	84.83	92.70	92.13	91.57
Angles	79.21	75.84	76.97	87.64	88.20	87.64

margins is determined. The authors formulate the optimization problem and solve it using Lagrangian multipliers. Finally, the proper recognition is carried out by the nearest neighbor classifier with Mahalanobis distance function operating on extracted motion features.

The performances obtained by kNN classifiers operating on PCALDA and MMC feature spaces are presented in Table 12. If, analogously to our approach, rotational data are taken into consideration, accuracies of recognition are noticeably worse. For PCALDA it is only 79.21%, for MMC—88.20%. There is an improvement if 3D data are analyzed, which can be explained by dependency of estimation of markers' locations on the calibration stage of motion capture system, which is weaker than in the case of joints rotations. However, DTW with selection of pose configuration space achieves still almost 2% higher precision of gait identification. Similarly to the previous investigated method, the results are worse than reported in the original paper (Balazia and Sojka 2018). The reason is the same—our collected dataset is more challenging.

7 Summary and conclusions

The problem of skeleton based motion data classification is related to the analysis of multidimensional time sequences of poses represented by rotations of successive joints. Quite an obvious and straightforward way to recognize such data is to compare corresponding time instants. However, if applied directly it has serious limitations, which refer to identical lengths of analyzed data series and possible local shifts between motion phases. What is more, some joints possess stronger discriminative features, while the others are characterized by weaker ones or they are even irrelevant. Thus, to overcome aforementioned problems, dynamic time warping and selection of pose attributes are carried out.

The main contribution of the paper is related to the proposed and successfully validated approaches of minimum distance classification scheme of time series data and an exploration of pose configuration space as well as performed exhaustive numerical experiments for motion capture data of human gait. It constitutes substantial progress beyond state-of-the-art results.

The main conclusions of the presented work are as follows:

- Dynamic time warping is an effective technique in motion data analysis.
- Minimum distance classification scheme of time series data can be accomplished on the basis of DTW alignment and properly chosen techniques for averaging time instants.
- Quaternions persist strong discriminative features of rotations and they are efficient in joints movements comparison.
- The product of normal and conjugate quaternions referring to successive time instants of motion sequence constitutes robust approach to differential filtering. It can be used to estimate angular velocities and accelerations.
- Greedy hill climbing heuristic is a fast and valuable procedure in exploration of pose configuration space.
- DTW based classification relies on applied distance function between rotations.

- Instance based learning is an efficient technique of motion data classification.
- A calibration stage of motion capture measurements is ambiguous. It depends on strict markers location and a set of exercises performed by a human. It leads to different orientations of local coordinate systems, which makes the classification more challenging.
- The selection of pose space improves the performance of the classification noticeably and it reduces computational expensiveness.
- The individual traits of successive joints depend on their activity during gait.
- There are very strong discriminative features related to dynamics of joints movements.
- An acquisition noise is enhanced by differential filtering. It results in poor precision of recognition if estimated angular accelerations are taken into consideration.
- The performance of motion data classification deeply relies on representativeness of a training set.

There are still possibilities to improve the proposed methods. A troublesome acquisition noise can be partially removed by filtering of time domain. In the case of Euler angles, the most well-known technique is Savitzky–Golay filter (Orfanidis 1995) and for unit quaternions multiresolution analysis can be carried out. Moreover, differential filtering can utilize signal of higher scales. Once again it reduces a noise and it supplies more general angular velocities and accelerations obtained by the analysis of more distant gait phases.

The shortcomings related to a calibration stage of motion capture measurements can be diminished by matching local coordinate systems. Thus, global rotation of referenced system has to be determined by selected optimization technique. It is not clear if matching would improve the performance of recognition, because it may also remove some individual, discriminative features, as for instance the ones related to human posture. Further research has to be carried out to test the stated hypothesis.

Normalizable distance approach may be utilized in order to balance the influence of joints or all angles, velocities and accelerations taken together, on global pose dissimilarity. In such a case w_{joint} weights of Eq. (21) are adjusted to maximum distances between instances of a training set calculated with respect to successive pose attributes.

The introduced method analyzes only joints movements without strict marker positions. It is reliable, resistant to clothing, surroundings and conditions during acquisition. In the case of practical deployments only a proper markerless motion capture system (Krzeszowski et al. 2014) has to be chosen. What is more, further improvements of recognition performance can be achieved by usage of features related to anthropometric data, as for instance human height, or the analysis of inter-ankle distances.

The validation experiments were carried out using highly precised motion capture data related to challenging gait identification problem. Thus, the influence of an acquisition noise is comparatively low and the complete skeleton model of a human was taken into consideration. The obtained results and conclusions are general and the proposed approach of exploration of pose configuration space has a practical meaning. For instance, it can be used to determine the most effective placements of IMU sensors on the human body with respect to any classification problem.

References

AbdelMaseeh, M., Chen, T. W., & Stashuk, D. (2016). Extraction and classification of multichannel electromyographic activation trajectories for hand movement recognition. *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, *26*(6), 662–673.

Aggarwal, J.K., & Ryoo, M.S. (2011). Human activity analysis: A review. ACM Comput. Surv. 43(3), 16. http://dblp.uni-trier.de/db/journals/csur/csur43.html#AggarwalR11.

Andre-Jonsson, H., & Badal, D. Z. (1997). Using signature files for querying time-series data. *Principles of Data Mining and Knowledge Discovery -Lecture Notes in Computer Science*, *1263*, 211–220.

Balazia, M., & Plataniotis, K. N. (2017). Human gait recognition from motion capture data in signature poses. *IET Biometrics*, *6*(2), 129–137.

Balazia, M., & Sojka, P. (2018). Gait recognition from motion capture data. In *ACM transactions on multimedia computing, communications, and applications special issue on representation, analysis and recognition of 3D human*.

Barnachon, M., Bouakaz, S., Boufama, B., & Guillou, E. (Jan 2014). Ongoing human action recognition with motion capture. *Pattern Recognition,47*(1), 238–247. http://liris.cnrs.fr/publis/?id=6192.

Bashir, F., Qu, W., Khokhar, A., & Schonfeld, D. (2005). Hmm-based motion recognition system using segmented pca. In IEEE international conference on image processing, 2005 (ICIP 2005) (Vol. 3, pp. III–1288). IEEE.

Bautista, M. A., Hernández-Vela, A., Escalera, S., Igual, L., Pujol, O., Moya, J., et al. (2016). A gesture recognition system for detecting behavioral patterns of adhd. *IEEE Transactions on Cybernetics*, *46*(1), 136–147.

Boyd, J., & Little, J. (2005). Biometric gait recognition. In Tistarelli, M., Bigun, J., Grosso, E. (Eds.), *IEEE international conference on advanced studies in biometrics, lecture notes in computer science* (Vol. 3161, pp. 19–42). Springer, Berlin. https://doi.org/10.1007/11493648_2.

Chen, L., & Raymond, N. (2004). On the marriage of lp-norms and edit distance. In *VLDB '04 proceedings of the thirtieth international conference on very large data bases* (Vol. 30, pp. 792–803).

Chen, L., Zsu, M.T., & Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD international conference on management of data* (pp. 491–502).

De la Torre Frade, F., Hodgins , J. K., Bargteil, A. W., Martin Artal, X., Macey, J. C., Castells, A. C., & Beltran, J. (April 2008). Guide to the carnegie mellon university multimodal activity (CMU-MMAC) database. *Techical report CMU-RI-TR-08-22, Robotics Institute, Pittsburgh, PA*.

Chu, S., Keogh, E. J., Hart, D. M., & Pazzani, M. J. (2002). Iterative deepening dynamic time warping for time series. In *Proceedings of the second SIAM international conference on data mining, Arlington, VA, USA, April 11–13, 2002* (pp. 195–212). https://doi.org/10.1137/1.9781611972726.12

Gavrila, D. M., & Davis, L. S. (1995). Towards 3-d model-based tracking and recognition of human movement: a multi-view approach In *International workshop on automatic face- and gesture-recognition* (pp. 272–277). IEEE Computer Society.

Gramkow, C. (2001). On averaging rotations. *Journal of Mathematical Imaging and Vision*, *15*(1), 7–16.

Hu, M. C., Chen, C. W., Cheng, W. H., Chang, C. H., Lai, J. H., & Wu, J. L. (2015). Real-time human movement retrieval and assessment with kinect sensor. *IEEE Transactions on Cybernetics*, *45*(4), 742–753.

Huynh, D. Q. (2009). Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, *35*(2), 155–164.

Itakura, F. (1990). Readings in speech recognition, chap. In *Minimum prediction residual principle applied to speech recognition* (pp. 154–158). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. http://dl.acm.org/citation.cfm?id=108235.108243

Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, *39*, 235–244.

Jablonski, B. (2012). Quaternion dynamic time warping. *IEEE Transactions on Signal Processing*, *60*(3), 1174–1183.

Johnson, M. P. (2002). *Exploiting quaternions to support expressive interactive character motion*. Ph.D. thesis, Massachusetts Institute of Technology.

Josiński, H., Kostrzewa, D., Michalczuk, A., & Świtoński, A. (2014). The expanded invasive weed optimization metaheuristic for solving continuous and discrete optimization problems. *The Scientific World Journal*, *2014*, 831691. https://doi.org/10.1155/2014/831691.

Josiński, H., Świtoński, A., Michalczuk, A., & Wojciechowski, K. (2012). Motion capture as data source for gait-based human identification. *Przeglad Elektrotechniczny*, *88*, 201–204.

Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems*, *7*(3), 358–386.

Keogh, E. J., & Pazzani, M. J. (2001). Derivative dynamic time warping. In *Proceedings of the 2001 SIAM international conference on data mining* (pp. 1–11). SIAM.

Krzeszowski, T., Switonski, A., Kwolek, B., Josinski, H., & Wojciechowski, K. (2014). Dtw-based gait recognition from recovered 3-d joint angles and inter-ankle distance. In *Computer vision and graphics* (pp. 356–363). Springer.

Li, W., Zhang, Z., & Liu, Z. (2010). Action recognition based on a bag of 3d points. In *IEEE computer society conference on computer vision and pattern recognition—workshops*.

Li, Y., Xue, D., Lee, G., Forrister, E., Garner, B., & Kim, Y. (2016). Human activity classification based on dynamic time warping of an on-body creeping wave signal. *IEEE Transactions on Antennas and Propagation*, *64*(11), 4901–4905.

Liu, H., Zhang, F., Mishra, S. K., Zhou, S., & Zheng, J. (2016). Knowledge-guided fuzzy logic modeling to infer cellular signaling networks from proteomic data. *Scientific Reports*, *6*, 35652.

Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). Mpca: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, *19*(1), 18–39.

Margarito, J., Helaoui, R., Bianchi, A. M., Sartor, F., & Bonomi, A. G. (2016). User-independent recognition of sports activities from a single wrist-worn accelerometer: A template-matching-based approach. *IEEE Transactions on Biomedical Engineering*, *63*(4), 788–796.

Martin, M., Maycock, J., Schmidt, F., & Krämer, O. (2010). Recognition of manual actions using vector quantization and dynamic time warping. In *Proceedings of the 5th conference on hybrid artificial intelligence systems* (Vol. 6067, pp. 221–228). Springer.

Martin, S., Brunner, P., Iturrate, I., Millán, J. D. R., Schalk, G., Knight, R. T., et al. (2016). Word pair classification during imagined speech using direct brain recordings. *Scientific Reports*, *6*, 25803.

Mehrabian, A. R., & Lucas, C. (2006). A novel numerical optimization algorithm inspired from weed colonization. *Ecological Informatics*, *1*(4), 355–366.

Mei, J., Liu, M., Wang, Y. F., & Gao, H. (2016). Learning a mahalanobis distance-based dynamic time warping measure for multivariate time series classification. *IEEE Transactions on Cybernetics*, *46*(6), 1363–1374.

Mian, O. S., Schneider, S. A., Schwingenschuh, P., Bhatia, K. P., & Day, B. L. (2011). Gait in swedds patients: comparison with parkinson's disease patients and healthy controls. *Movement Disorders*, *26*(7), 1266–1273.

Morse, M. D., & Patel, J. M. (2007). An efficient and accurate method for evaluating time series similarity. In *Proceedings of the 2007 ACM SIGMOD international conference on management of data* (pp. 569–580).

Muller, M. (2007). *Information Retrieval for Music and Motion*. Secaucus: Springer.

Muller, M., & Roder, T. (2006). Motion templates for automatic classification and retrieval of motion capture data. In *ACM SIGGRAPH/eurographics symposium on computer animation* (pp. 137–144).

Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., & Weber, A. (June 2007). Documentation mocap database hdm05. *Technical report CG-2007-2*, Universität Bonn.

Munich, M. E., & Perona, P. (1999). Continous dynamic time warping for trnaslation-invariant curve alignment with applicati omn to signature verification. In *7th international conference on computer vision*.

Orfanidis, S. J. (1995). *Introduction to Signal Processing*. Upper Saddle River: Prentice Hall.

Plotnik, M., Giladi, N., & Hausdorff, J. M. (2007). A new measure for quantifying the bilateral coordination of human gait: effects of aging and parkinson's disease. *Experimental Brain Research*, *181*(4), 561–570.

Plouffe, G., & Cretu, A. M. (2016). Static and dynamic hand gesture recognition in depth data using dynamic time warping. *IEEE Transactions on Instrumentation and Measurement*, *65*(2), 305–316.

Sakoe, H., & Chuba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *8*, 43–49.

Sart, D., Mueen, A., Najjar, W. A., Keogh, E. J., & Niennattrakul, V. (2010). Accelerating dynamic time warping subsequence search with GPUS and FPGAS. In G. I. Webb, , 0001, B.L., C. Zhang, D. Gunopulos, X. Wu (Eds.), *ICDM* (pp. 1001–1006). IEEE Computer Society. http://dblp.uni-trier.de/db/conf/icdm/icdm2010.html#SartMNKN10.

Srivastava, R., & Sinha, P. (2016). Hand movements and gestures characterization using quaternion dynamic time warping technique. *IEEE Sensors Journal*, *16*(5), 1333–1341.

Tenorth, M., Bandouch, J., & Beetz, M. (2009). The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition. In *IEEE international workshop on tracking humans for the evaluation of their motion in image sequences (THEMIS), in conjunction with ICCV2009*.

Trung, N. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., & Yagi, Y. (2014). The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, *47*(1), 228–237.

Wang, X., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. J. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, *26*, 275–309.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Burlington: Morgan Kaufmann.

Zhang, Z., Fang, Q., & Gu, X. (2016). Objective assessment of upper-limb mobility for poststroke rehabilitation. *IEEE Transactions on Biomedical Engineering*, *63*(4), 859–868.

Zhou, F., & De la Torre, F. (December 2009). Canonical time warping for alignment of human behavior. In *Advances in neural information processing systems conference (NIPS)*.

Zhou, F., & De la Torre, F. (2016). Generalized canonical time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 279–294.

Zifchock, R. A., Davis, I., Higginson, J., & Royer, T. (2008). The symmetry angle: a novel, robust method of quantifying asymmetry. *Gait & posture*, *27*(4), 622–627.