Machine Vision and Applications (2015) 26:905–918
DOI 10.1007/s00138-015-0704-z

ORIGINAL PAPER

Person identity recognition on motion capture data using multiple actions

Ioannis Kapsouras[1] · **Nikos Nikolaidis**[1]

Received: 17 June 2014 / Revised: 25 June 2015 / Accepted: 7 July 2015 / Published online: 6 August 2015

Abstract In this paper, we introduce a novel method for person identity recognition (identification) on skeleton animation/motion capture data representing persons performing various actions. The joints positions or orientation angles and the forward differences of these quantities are used to represent a motion capture sequence. First K-means clustering is applied on training data to discover the most representative patterns on joints positions or orientation angles (dynemes) and their forward differences (F-dynemes). Each frame is then assigned to one of these patterns and the frequency of occurrence histograms for each movement are constructed in a bag-of-words fashion. Person identity recognition is done through a nearest neighbor classifier. The proposed method is experimentally tested on a number of datasets of motion capture data, with very good results.

Keywords Dynemes · Forward differences · Identity recognition · Bag of words · Motion capture data

1 Introduction

Motion capture (mocap) data provide a representation of human motion. Such data describe the locations of characteristic parts on the human body such as joints or the joint angles over time and can be obtained by using appropriate tracking devices (magnetic, ultrasonic, inertial, optical, mechanical) [5]. Motion capture can also be performed with the use of low-cost capturing devices such as the Microsoft Kinect sensor or other RGB+depth (RGB-D) sensors. Skeleton models used in mocap consist of nodes that represent the joints of the skeleton and arcs that represent the segments.

Usually all joints have 3 rotational degrees of freedom whereas the root node has also 3 translation parameters. The values of the angles of a certain frame form the n-dimensional pose or posture vector. Examples of mocap sequences are shown in Fig. 1.

Human identity recognition (identification) is a very active research field. A lot of biophysical features are used in biometrics to recognize the identity of a subject. Many of the approaches such as fingerprint, face or iris recognition require special hardware and/or the subject's cooperation. On the other hand, research on non-invasive biometrics has recently gained significant momentum due to the advantages offered by such approaches. One of the most commonly used non-invasive biometric is gait. The research area of gait recognition deals with the identification of subjects by the way they walk. Walk sequences can be captured either on video or by a motion capture system. Until recently, recording using a motion capture system was in most cases an invasive and costly to implement approach but now with the use of RGB-D sensors such as Kinect, motion capture can be made in a non-invasive and inexpensive way.

The fact that activity-based identity recognition systems are focusing only on the walking style is obviously a limiting factor. People perform various activities such as running, jumping, clapping, etc. and it would be beneficial if a method could recognize the subject's identity from such activities. In addition, a recognition system that can work with various movements/actions might provide better results, since other types of movements, more complex than walking, might pro-

✉ Ioannis Kapsouras
jkapsouras@aiia.csd.auth.gr

Nikos Nikolaidis
nikolaid@aiia.csd.auth.gr

1 Department of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloníki, Greece

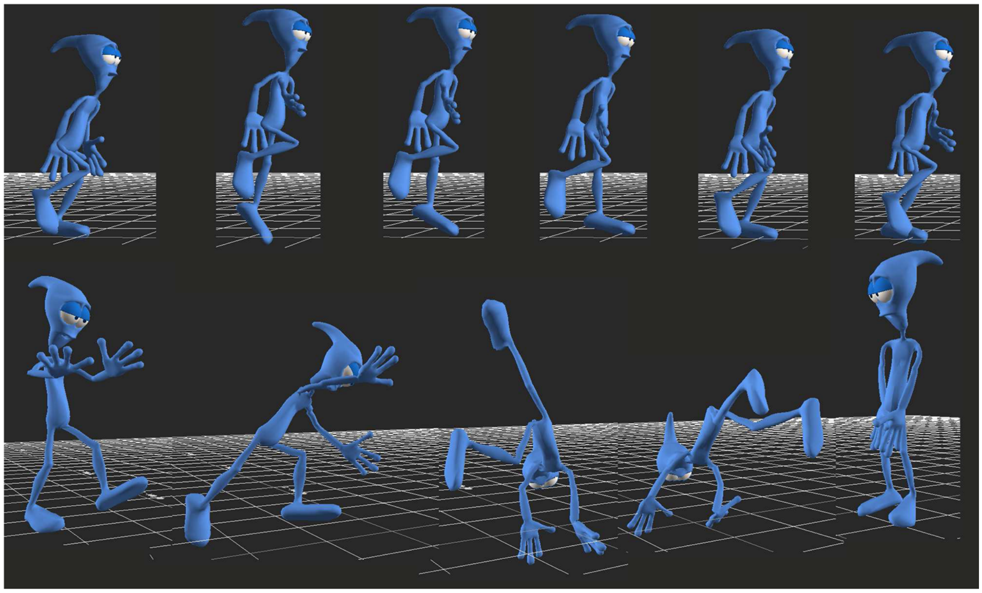

Fig. 1 Hop on left leg and cartwheel movement sequences from the HDM05 dataset [19]

vide more discriminative information with respect to person's identity.

Although a considerable amount of research has been performed on movement-based identity recognition from video data (mainly gait recognition), research dealing with skeletal animation or motion capture data is almost non existent. Identity recognition using various movements in skeleton animation data can have various applications such as access control, human computer interaction, person monitoring, surveillance and automatic annotation of skeletal animation data.

This paper presents an identity recognition method that operates on motion capture data. As already mentioned, in a motion capture sequence, a single frame is represented by the posture vector of the human body. The forward differences over various temporal scales for each joint angle are computed in order to capture the dynamics of each joint. K-means is applied separately on the postures and the forward differences of the training data to discover characteristic patterns. Then each posture and each forward difference vector is mapped to the closest pattern and histograms with the frequencies of appearance of the patterns for each sequence are created. Classification is implemented using a nearest neighbor approach. The method uses a modified K-means algorithm that can handle angular data such as the joint angles involved in mocap which are of a periodic nature. The proposed method, that adapts the activity recognition framework presented in [13] towards identity recognition, is essentially a Bag of Words (BoW) method, where the words are the posture and the forward difference vectors and the codebooks are computed by implementing K-means separately to postures and to forward differences vectors. The method can operate both on angle-based and position-based mocap data.

The rest of this paper is organized as follows. In Sect. 2, we present a review on previous work. The proposed method is described in detail in Sect. 3. In Sect. 4, experimental performance evaluation of the method is presented. Conclusions follow.

2 Previous work

The body of research for movement-based identity recognition on skeleton animation data is limited. Research on motion capture data focuses mainly on motion indexing and retrieval as well as movement/activity recognition rather than identity recognition.

Indeed, methods that use skeleton animation data for movement-based identity recognition are very few and focus on gait. In [26], the authors use the position of the projections in the walking plane of 3D markers of a motion capture system to estimate the planar offsets between the markers and the underlying skeleton during walking. Then the authors compute the joint angle trajectories using the offsets. Furthermore, time wrapping is used to normalize the trajectories. The classification is performed by using a nearest neighbor approach involving Euclidean distance. In [12], second-order tensor objects are used to represent mocap walking data. Multilinear principal component analysis (MPCA) is used to reduce the high-dimensionality of the tensor objects. Then k Nearest Neighbors (kNN), Naive Bayes, Multilayer Perceptron and Radial Basis Function Network are applied for the classification. In [16], the authors use three-dimensional kinematic and kinetic variables acquired for lower limb joints. The parameters are acquired by tracking marker positions using a motion capture system with force plates. A self-organizing map (SOM) neural network is then used for the classification. The authors in [8] perform identity recognition using gait while also performing recognition of other activities. A markerless pose recovery method to obtain 3D human joints from volume data is proposed. Combination of joint positions (configuration features) and position, orientation and height of the body (movement features) are used as features. A hidden Markov model is used to model the movement features and an exemplar-based HMM is used to model the configuration features. Finally, a maximum a posteriori classifier is adopted for classification.

Identity and action recognition has also been attempted on data provided by RGB-D sensors such as Microsoft Kinect. The Kinect sensor and the accompanying software modules not only records depth video, but also provides information for the joints positions of the tracked skeletons. In [25], the authors propose a method for gait recognition by extending the Gait Energy Images (GEI) produced by averaging the silhouettes extracted over a gait cycle [9] towards Gait Energy Volumes (GEV) which is the 3D analogue of the GEI. Reconstructed voxel volumes substitute the temporally averaging segmented silhouettes. Preis et al. in [24] proposed a method for gait recognition based on Microsoft Kinect data. Their method is an appearance-based method that uses 13 biometric features such as the height, the length of legs, torso, etc., together with step-length and speed to characterize one person. The body features are extracted from the skeletal points detected by Kinect SDK. Three classifiers (1R, C4.5 and Naive Bayes) were used for classification. In [20], Munsell et al. present a method for person identification from Kinect videos. Their method applies in both walking and running sequences. The authors use initially an SVM to perform movement classification of an unknown sequence. Then a cost function is used on the recognized sequence to perform identity recognition. In [1], the authors used soft-biometric features, such as Euclidean distance between floor and head or Geodesic distance between torso center and left shoulder, extracted from depth data to create a signature that characterizes a subject. The computed signatures are used for the identification through a matching phase.

In the area of video-based identity recognition from movement characteristics, research focused mainly on gait recognition and numerous methods have been proposed in the literature. Although these methods are not directly comparable to the mocap-based approach proposed in this paper, we provide below a short review of some of these approaches. Han and Bhanu proposed a person identification method with the use of gait energy images (GEI). The authors use GEIs produced by real and synthetic data and they map them to subspaces using PCA and LDA. Recognition is performed by classifying the GEIs to the nearest centroid. In [2], the authors introduce gait entropy images (GEnI). A GEnI is produced by measuring the entropy at each pixel in the corresponding GEI. For gait recognition, a novel adaptive component and discriminant analysis (ACDA) algorithm is used. Lam et al. in [14] propose the motion silhouettes image (MSI) that is a grey-level image which embeds the critical spatio-temporal information. After representing the gait sequences with MSIs, they use PCA to reduce dimensionality so as to make the MSIs more distinguishable. A nearest neighbor approach is used to recognize an unknown gait sequence. Other methods for video-based gait recognition are presented in [3,4,22,27]. Surveys for video-based gait recognition can be found in [15,18,29].

A few video-based identity recognition methods use additional movements beyond gait to recognize an unidentified person. In [28], a person identification method that uses gait and run as movements is proposed. Movement sequences are represented by the magnitude of the Fourier transform of the angles of the lower leg rotation during the gait and running cycles over time. The angles of the lower leg are extracted from the model using an oscillator defined in [28]. A k-nearest neighbor classifier is used for the classification process. A disadvantage of this method is that it does not take into account other angles of the human skeleton. Thus, activities such as clapping cannot be recognized because they do not involve lower leg angles. Gkalelis et al. use various movements such as walk, run, skip, jump for video-based identity recognition in [7]. In this method, the authors use binary body posture masks to describe the body poses. Fuzzy vector quantization (FVQ) and linear discriminant analysis (LDA) are then used to recognize an unknown movement and then the person performing the movement is recognized from a movement-specific person classifier. One disadvantage of identity recognition in single-view video data is that the training and the test data have to be recorded from the same viewing angle, e.g., from a frontal or a side view. Iosifidis et al. [10] use a multicamera setup to overcome this disadvantage. Their method uses dynemes (in the image space) to represent the human body poses within the various movements. Then the human poses in a sequence are mapped to the dynemes using fuzzy distances. LDA is used to reduce the dimensionality of the mapped sequences. At the end, a Bayesian framework is used for identity and activity recognition.

3 Identity recognition on motion capture data

In the proposed method, two types of features within a Bag of Words framework are used to represent each movement and a nearest neighbor approach is used in order to recognize an unknown individual.

3.1 Feature extraction

A motion capture sequence is represented as a sequence of posture vectors $\mathbf{y}_i,\ i = 1, \ldots, M$, M being the number of frames. Each posture vector carries information either for the rotation angles in the skeleton joints or for the positions of the skeleton joints in the $3D$ space. In more detail,

$$\mathbf{y}_i = [\phi_{i1}, \phi_{i2}, \ldots, \phi_{im}], \tag{1}$$

where m is the number of rotation angles that form the posture vector, or

$$\mathbf{y}_i = [x_{i1}, y_{i1}, z_{i1}, x_{i2}, y_{i2}, z_{i2}, \ldots, x_{il}, y_{il}, z_{il}], \tag{2}$$

where l is the number of joints that form the posture vector. Motion capture with sensors attached to the human body usually leads to angular-based data, while motion capture with depth cameras such as Microsoft Kinect usually leads to positional-based data. The proposed method works both for angular-based and positional-based data and was tested on datasets from both cases.

A motion capture sequence can also be represented by vectors of forward differences evaluated over joint rotation angles or positions. Forward differences estimate the first derivative of a signal and thus, when applied on joints, carry information for the average angular or translational velocities of the skeleton joints. More specifically, forward differences in terms of a motion capture sequence can be defined as:

$$\mathbf{FD}_i^t = \Delta_t[\mathbf{y}_i] = \mathbf{y}_{i+t} - \mathbf{y}_i, \tag{3}$$

where $\mathbf{y}_i, \mathbf{y}_{i+t}$ are the posture vectors [defined by (1) or (2) depending on the data] in frames i and $i+t$, respectively. $\mathbf{FD}_i^t$ can be considered as a vector of the average velocities of the joints of a skeleton at frame i. The forward differences of the joints are computed in three different temporal scales namely for $t = 1$, $t = 5$ and $t = 10$. It should be noted that when the values in posture vectors $\mathbf{y}_{i+t}$ and $\mathbf{y}_i$ in (3) are angular values [i.e., when $\mathbf{y}_i, \mathbf{y}_{i+t}$ are defined as in (1)], the evaluation of the difference of these two vectors involves finding for each pair of corresponding elements their angular distance, instead of their Euclidean distance, due to the periodic nature of such data. The distance $\mathrm{arc}(\theta_i, \theta_j)$ between two angles θ_i, θ_j is the smallest arc between the two points that are defined by these angles on the unit circle. This distance can be evaluated as:

$$\mathrm{arc}(\theta_i, \theta_j) = \pi - \mid \pi - \mid \theta_i - \theta_j \mid\mid \tag{4}$$

In short, 4 groups of vectors are proposed to represent a motion capture sequence: posture vectors and forward differences vectors in three different temporal scales. Thus a sequence is represented by four sets of feature vectors $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4$:

$$\begin{aligned}
\mathbf{T}_1 &= \{\mathbf{y}_1, \ldots, \mathbf{y}_M\} \\
\mathbf{T}_2 &= \left\{\mathbf{FD}_1^1, \ldots, \mathbf{FD}_{M-1}^1\right\} \\
\mathbf{T}_3 &= \left\{\mathbf{FD}_1^5, \ldots, \mathbf{FD}_{M-5}^5\right\} \\
\mathbf{T}_4 &= \left\{\mathbf{FD}_1^{10}, \ldots, \mathbf{FD}_{M-10}^{10}\right\}
\end{aligned} \tag{5}$$

3.2 Codebook evaluation

After representing each training sequence in terms of the 4 sets presented in (5), the algorithm clusters the features in order to extract four codebooks (one for each set) consisting of characteristic features / words.

In more detail, in order to recognize P different subjects performing various movements, we cluster feature vectors in each feature space into Q clusters using K-means. K-means is applied separately on the sequences of each subject, four times. First K-means is applied upon sets $\mathbf{T}_1$ of all training sequences, i.e., on the posture vectors $\mathbf{y}_{ij}, i = 1, \ldots, M_j, j = 1, \ldots, K$ of all movement sequences in the training set where M_j is the number of frames of the jth movement sequence and K the number of the training sequences. It should be noted that K-means is applied separately for the sequences of each subject in the training set resulting to $P \times Q$ cluster centers, where P is the number of subjects and Q the number of clusters for the sequences of each subject. Using specific "person" codewords allows to better capture the subtle differences between subjects performing the same action and thus to better recognition rates. Moreover, since our method applies to closed identification problems, the number of persons to be recognized is already known and the approach is valid. The classic approach of applying clustering to all movements of all subjects in the training dataset has been also tested, but the results were slightly worse. If the elements of posture vectors are joint angles, a K-means variant modified to work on angular data is applied (see Sect. 3.3). For each cluster created by the angular K-means algorithm, the centroid $\mathbf{d}_q, q = 1, \ldots, P \times Q$ is computed as the circular mean (Sect. 3.3) of all postures in this cluster. A centroid extracted by the K-means represents one Dyneme in a way similar to [6,11] that, however, operate on video data. Due to the averaging procedure, Dynemes correspond to "average", characteristic postures rather than specific postures from within the dataset. Some example Dynemes are shown in Fig. 2. Moreover, K-means is applied on the forward differences vectors $\mathbf{FD}_{ij}^t, i = 1, \ldots, M_j - t, j = 1, \ldots, K$, each time for a different temporal scale $t = 1, 5, 10$ where $M_j - t$ is the number of forward differences vectors in the jth sequence and t the temporal scale. K-means is applied, as before, separately for

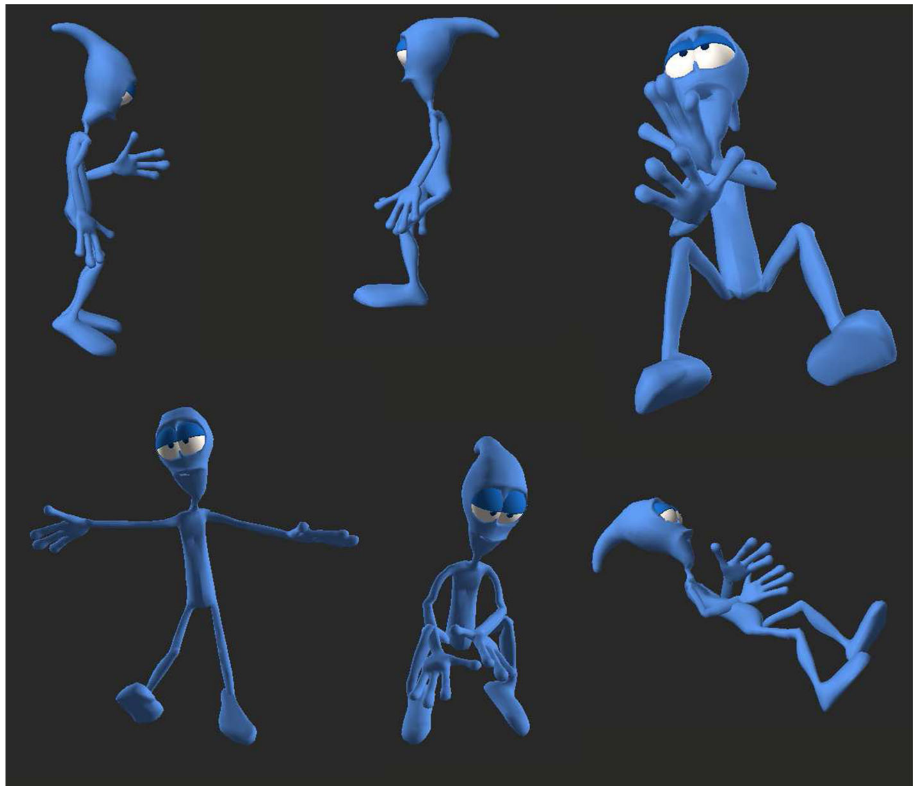

Fig. 2 Examples of dynemes extracted by the clustering algorithm

the sequences of each subject. This results to $3 \times P \times Q$ centroids, $\mathbf{f}_q^t, q = 1, \ldots, P \times Q, t = 1, 5, 10$. A centroid computed based on forward differences represents one F-Dyneme. At the end, the cluster centers of each subject for each feature space (posture space and forward differences space for each temporal scale) are concatenated to form 4 sets of $C = P \times Q$ clusters. The number C of clusters is selected empirically and depends on the number of subjects P that are to be recognized, the different ways a movement can be performed by different people, the different body types, etc.

3.3 Angular *K*-means

In certain types of motion capture data the posture vectors $\mathbf{y}_i$, describe rotation angles at the joints [cf. (1) in Sect. 3.1]. Due to the periodic nature of angular data the mean value estimator for data on the line cannot be used in such data. The circular mean [21] shall be used instead in this case.

The *circular mean* or *sample mean direction* $\bar{x_0}$ of k angular observations $\theta_1, \ldots, \theta_k$ represented by sample points $\mathbf{P}_1, \ldots, \mathbf{P}_k$ on a unit circle centered at point $\mathbf{O}$ is defined as the direction of the mean resultant vector $\mathbf{R}$ of the unit vectors $\mathbf{OP}_1, \ldots, \mathbf{OP}_k$. $\bar{x_0}$ is evaluated as follows:

$$\bar{x_0} = \arctan\left(\frac{\bar{S}}{\bar{C}}\right), \quad \bar{C} = \frac{1}{k}\sum_{i=1}^{k}\cos\theta_i, \quad \bar{S} = \frac{1}{k}\sum_{i=1}^{k}\sin\theta_i \tag{6}$$

Similarly, as already mentioned in Sect. 3.1 the angular distance (4) should be used instead of the Euclidean. Since the proposed method applies the K-means algorithm on motion capture sequences that involve angle data (posture vectors of the form (1)), a modified angular version of K-means was constructed by replacing the classical mean and Euclidean distance operators involved in the algorithm by the angular distance and the circular mean, respectively.

3.4 Projecting the features

In this step, we map the features to the codewords. As mentioned in Sect. 3.1, K-means is applied on each person's sequences separately resulting to $C = P \times Q$ cluster centers. The mapping of each posture vector $\mathbf{y}_{ij}$ to its closest center (dyneme) $\mathbf{d}_q$ is computed by taking into account all C centers:

$$q = \underset{c \in [1 \ldots C]}{\arg\min}\left(\sum_{k=1}^{m} \operatorname{arc}\left(d_{ck} - y_{ijk}\right)\right), \tag{7}$$

where m is the number of rotation angles that form the posture vector and y_{ijk} is the kth element of $\mathbf{y}_{ij}$. Similarly each forward differences vector $\mathbf{FD}_{ij}^t$ is mapped to an F-Dyneme $\mathbf{f}_q^t$, where the set of all F-Dynemes for a certain value of t includes all $P \times Q$ centers resulting by applying P times the K-means algorithm as described in Sect. 3.1:

$$q = \arg\min_{c\in[1\ldots C]} \left(\sum_{k=1}^{m} \left| f^t_{ck} - FD^t_{ijk} \right| \right), \tag{8}$$

where FD^t_{ijk} is the kth element of $\mathbf{FD}^t_{ij}$. Thus each movement sequence is represented in terms of Dynemes and F-Dynemes. More specifically, each frame is represented by one Dyneme and 3 F-Dynemes, i.e., those that correspond to the posture vector and the three forward differences vectors of this frame.

Next, we form for each sequence four histograms $\mathbf{m}^{\mathbf{y}}_j, \mathbf{m}^{\mathbf{FD}^1}_j, \mathbf{m}^{\mathbf{FD}^5}_j, \mathbf{m}^{\mathbf{FD}^{10}}_j$ by calculating the frequency of appearance of every Dyneme and F-Dyneme for each corresponding set of features. Each of the histograms of the jth sequence is a C-dimensional vector $\mathbf{m}_j = [m_{ij}]$, $j = 1, \ldots, K$, $i = 1, \ldots, C$:

$$m_{ij} = \frac{n_i}{M_j}, \tag{9}$$

where n_i is the number of occurrences of the ith center within the sequence and M_j the number of feature vectors of the sequence. Example histograms of posture vectors are shown in Figs. 3 and 4, whereas some histograms of forward differences vectors are shown in Fig. 5. Finally, the four histograms are concatenated to form the final vector having $4\cdot C$ elements that characterizes the movement:

$$\mathbf{m}_j = \left[\mathbf{m}^{\mathbf{y}}_j, \mathbf{m}^{\mathbf{FD}^1}_j, \mathbf{m}^{\mathbf{FD}^5}_j, \mathbf{m}^{\mathbf{FD}^{10}}_j\right] \tag{10}$$

3.5 Classification

The following procedure is used to classify an unknown motion capture sequence (test sequence) to one of the subjects that the algorithm has been trained to recognize.

First each posture vector $\mathbf{y}$ of the test sequence is mapped into the closest Dyneme $\mathbf{d}_q$ evaluated during the training stages:

$$q = \arg\min_{c\in[1\ldots C]} \left(\sum_{i=1}^{m} \operatorname{arc}\left(d_{ci} - y_i\right) \right) \tag{11}$$

Similarly each forward difference vector $\mathbf{FD}^t$ (for the three temporal scales t) of the testing sequence is mapped into the nearest F-Dyneme $\mathbf{f}^t_q$:

$$q = \arg\min_{c\in[1\ldots C]} \left(\sum_{i=1}^{m} \left| f^t_{ci} - FD^t_i \right| \right) \tag{12}$$

In both (11) and (12) m is the number of rotation angles that form the posture vector.

Once all vectors have been mapped to Dynemes and F-Dynemes, we calculate four histograms $\mathbf{m}^{\mathbf{y}}_{\text{test}}$, $\mathbf{m}^{\mathbf{FD}^1}_{\text{test}}$, $\mathbf{m}^{\mathbf{FD}^5}_{\text{test}}$, $\mathbf{m}^{\mathbf{FD}^{10}}_{\text{test}}$ for the test sequence using (9) and the four histograms are concatenated to form the final $\mathbf{m}_{\text{test}} = \left\{\mathbf{m}^{\mathbf{y}}_{\text{test}}, \mathbf{m}^{\mathbf{FD}^1}_{\text{test}}, \mathbf{m}^{\mathbf{FD}^5}_{\text{test}}, \mathbf{m}^{\mathbf{FD}^{10}}_{\text{test}}\right\}$ histogram. $\mathbf{m}_{\text{test}}$ characterizes the sequence and is used for the classification/recognition, performed by Nearest Neighbor (NN) classifier.

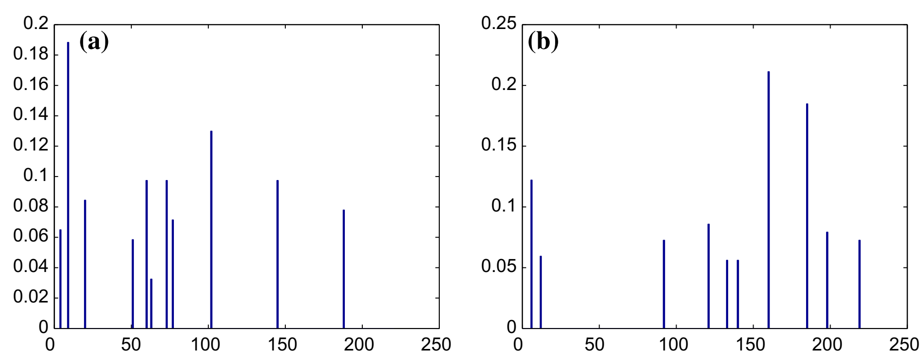

Fig. 3 Dynemes histograms of walk for two different subjects **a** and **b** generated by the proposed method. The most prominent *peaks* of the histograms are different for the two subjects

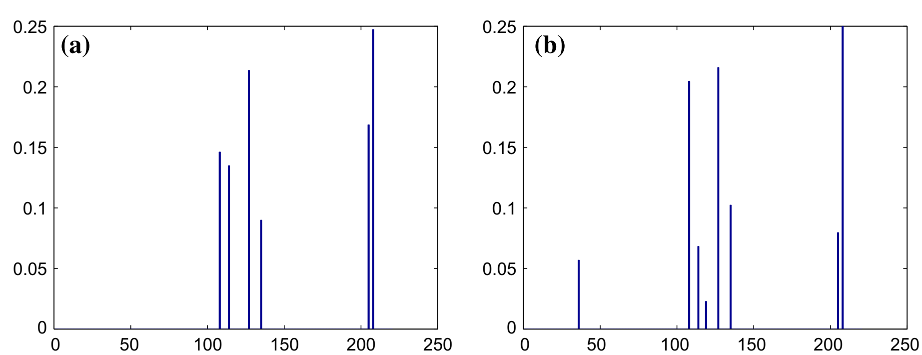

Fig. 4 Dynemes histograms of two run on place sequences **a** and **b** for the same subject generated by the proposed method. The most prominent *peaks* of the histograms appear at the same bins

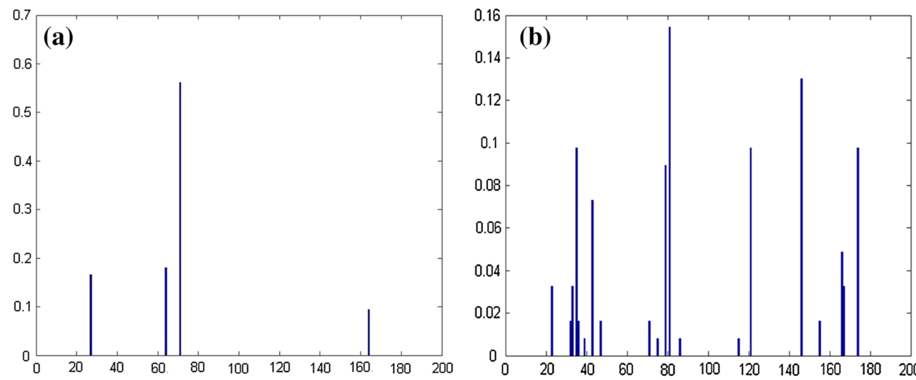

Fig. 5 F-Dynemes ($t = 1$) histograms of **a** deposit floor movement and **b** clap above head movement performed by two different subjects

The similarity of $\mathbf{m}_{\text{test}}$ with all labeled movement sequences of the training set, is evaluated using histogram intersection (13), cosine similarity (14) or a symmetric version of Kullback–Leibler divergence (15):

$$HI(\mathbf{m}_j, \mathbf{m}_{\text{test}}) = \sum_{i=1}^{4\cdot C} \min\left\{m_{j,i}, m_{\text{test},i}\right\} \tag{13}$$

$$CS(\mathbf{m}_j, \mathbf{m}_{\text{test}}) = \frac{\sum_{i=1}^{4\cdot C} m_{j,i} m_{\text{test},i}}{\sqrt{\sum_{i=1}^{4\cdot C}\left(m_{j,i}\right)^2}\sqrt{\sum_{i=1}^{4\cdot C}\left(m_{\text{test},\text{i}}\right)^2}} \tag{14}$$

$$KL\left(\mathbf{m}_j, \mathbf{m}_{\text{test}}\right) = \frac{1}{2}\left(\sum_{i=1}^{4\cdot C} m_{j,i} \ln \frac{m_{j,i}}{m_{\text{test},i}} + \sum_{i=1}^{4\cdot C} m_{\text{test},i} \ln \frac{m_{\text{test},i}}{m_{j,i}}\right), \tag{15}$$

where $m_{j,i}$ and $m_{\text{test},i}$ are the values of the ith element for the vector $\mathbf{m}_j$ of the jth training sequence and the test sequence vector $\mathbf{m}_{\text{test}}$, respectively.

The subject label of the training set sequence whose similarity with the test sequence is maximum (i.e., the training sequence with the maximum histogram intersection or cosine similarity, or the minimum Kullback–Leibler divergence) is assigned to the test sequence.

4 Experimental evaluation

Experiments have been conducted to test the proposed identity recognition method on 4 different datasets containing mocap data in angle-based or position-based representations.

Two variants were tested. In the first variant (1F), only posture vectors features were used to form the histogram of each sequence. In the second variant (4F), posture vectors and forward differences were used to form the histograms of each sequence. Histogram intersection (HIST), cosine similarity (COS) and Kullback–Leibler divergence (KL) were used as distance metrics in 1-NN classifier. A cross-validation procedure was used to assess the performance of the algorithm. Since the cross-validation may lead to overfitting and also in order to test the method in a more realistic and difficult setup, a 50/50 setup was also used. In this setup 50 % of the sequences of each dataset were used for training and the remaining 50 % formed the test set. A summary of the used datasets along with the type of data and the experimental setups applied can be seen in Table 1. Cross-validation was used only in balanced datasets, where all persons perform all movements the same number of times.

It should be noted here that comparison of the proposed method with other methods is very limited (only one comparison on gait data), because as far as we know there are no other methods that perform identity recognition on motion capture data using multiple movements. Regarding gait recognition methods operating in such data, comparison only with the method in [24] was possible since other methods were not tested on publicly available datasets (e.g., [20]).

4.1 HDM05

At first, the proposed method was tested on a subset of HDM05 dataset [19]. The HDM05 dataset contains five persons (non-professional actors) performing several movements. Fifteen of these movements namely, run on place (runop), walk (walk), hop with both legs (hopboth), hop with left leg (hopl), hop with right leg (hopr), clap (clap), clap with

Table 1 Datasets and experimental setups used

Database	HDM05	MHAD	HDM05-MHAD	MSR	GaitD
Data type	mocap	mocap	mocap	kinect	kinect
Setup	CV and 50/50	CV and 50/50	CV and 50/50	50/50	CV

hands above head (clapa), left elbow to right knee (eltoknee), right elbow to left knee (eltokneer), walk sideways (walks), walk sideways crossing the legs (walksc), kick with left leg (kickll), kick with right leg (kickrl), punch with left hand (punchlh) and punch with right hand (punchrh) were used in this paper. In order to have a balanced dataset, the same number of sequences from a specific movement was selected for all persons. This number was equal to the smallest number of available sequences for this movement, throughout the list of the five persons. The number of sequences used for each person in each movement is presented in Table 2.

The selected subset contains in total 680 sequences, in ASF/AMC format. The dataset contains rotation angles at the corresponding joints for 29 body parts, i.e., the posture vectors are of the form (1). Since some of these body parts, for example the thumb or the toes, bear no significant information for identity and movement recognition, we have retained only information for the following body parts: lower back, upper back, thorax, right humerus, right radius, left humerus, left radius, right femur, right tibia, right foot, left femur, left tibia, left foot. In addition, the global translation and rotation information (i.e., the rotation and translation of the root node) were not considered by the algorithm.

For the cross-validation (CV) procedure five sequences of the same movement, one from each person, were retained to form the test set whereas all the remaining sequences were used to form the training set at each cycle. This was repeated until all available sequences were used once as test data. The number of correctly identified sequences in all cycles (136 cycles in total) was used to compute the classification rate. The best identity recognition rates for the two variants (1F and 4F) are shown in Table 3.

As can be seen in this table, the method achieved the best recognition rate when all features are used and when histogram intersection is used as distance measure for the classification (HIST+4F). All other variants also achieved very high performances. The number of K-means clusters for which the best recognition rates has been achieved for a certain variant is also mentioned in this and all subsequent tables. The identity recognition confusion matrix for the best performing variant is shown in Fig. 6.

Table 2 Number of sequences for each movement and subject in the HDM05 dataset

Movement	clap	clapa	eltoknee
Num of sequences	6	5	3
Movement	eltokneer	hopboth	hopl
Num of sequences	3	5	5
Movement	hopr	tunop	walk
Num of sequences	5	14	16
Movement	walks	walksc	kickll
Num of sequences	6	4	13
Movement	kickrl	punchlh	punchrh
Num of sequences	18	18	15

Table 3 Correct identity recognition rates for all method variants (HDM05 dataset, cross-validation)

Variant	Recognition rate	Clusters
HIST+1F	99.12	215
COS+1F	98.97	185
KL+1F	99.26	185
HIST+4F	**99.71**	205
COS+4F	99.56	210
KL+4F	99.41	205

Bold values indicate the best result achieved

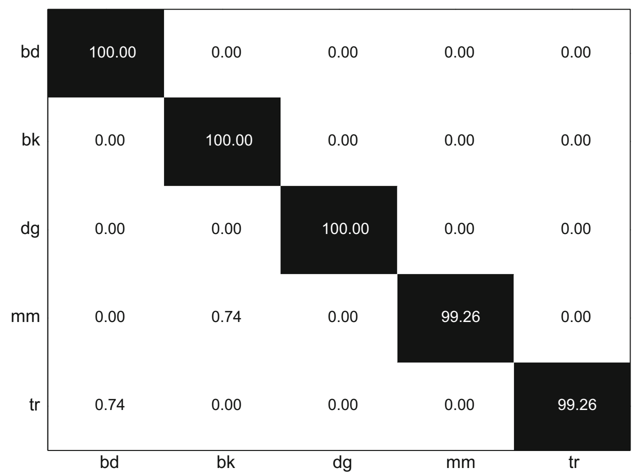

Fig. 6 Identity recognition confusion matrix (5 subjects: bd, bk, dg, mm, tr, subject labels as provided by the dataset creators) for the best performing variant, i.e., HIST+4F with $C = 205$; 99.71 % overall correct recognition rate in HDM05 dataset, using cross-validation

Moreover, the identity recognition rate per movement, i.e., the percentage of correctly recognized identities when a specific type of movement is considered, is shown in Table 4. By observing this table, one can see that all movements lead to very high (93–100 %) identity recognition rate. As a matter of fact, all movements but two lead to 100 % recognition rates. This finding leads to the important conclusion that the focus of the computer vision community on gait-based identity recognition (gait recognition) is not well justified since it is obvious that other movements bear information that allows to distinguish one person from another.

The results for this dataset, when using a 50 % training/50 % testing setup, are shown in Table 5.

As can be seen in this table, even in this more challenging setup the proposed method achieves slightly lower but still very high identification rates.

Table 4 Identity recognition rates per movement for HIST+4F with $C = 205$; 99.71 % overall correct recognition rate (HDM05 dataset)

Movement	Recognition rate (%)
runop	100
walk	100
hopboth	100
hopl	100
hopr	100
kickll	100
kickrl	100
punchlh	100
punchrh	100
clap	100
clapa	96
eltoknee	100
eltokneer	93.33
walks	100
walksc	100

Table 5 Correct identity recognition rates for all method variants in a 50/50 train/test split (HDM05 dataset)

Variant	Recognition rate	Clusters
HIST+1F	**98.82**	245
COS+1F	97.94	230
KL+1F	97.94	260
HIST+4F	95.30	225
COS+4F	95.59	225
KL+4F	95	285

Bold values indicate the best result achieved

4.2 MHAD

The proposed method has been also tested on the Berkeley Multimodal Human Action Dataset (MHAD) [23]. The MHAD dataset contains 12 persons (non-professional actors) performing 11 movements namely, jump in place (jplace), jumping jacks (jjacks), bend (bend), punch (punch), wave two hands (wavtwo), wave one hand (wavone), clap hands (clap), throw a ball (throw), sit down then stand up (sitstand), sit down (sit) and stand up (stand). There are 5 sequences of each movement for each person in the dataset, therefore the dataset contains $12 \times 11 \times 5 = 660$ sequences in BVH format. Information for the following parts has been retained in this dataset: hips, spine, neck left and right shoulder, left and right arm, left and right arm roll, left and right fore arm, left and right up leg, left and right up leg roll, left and right leg and left and right foot. In addition, the global translation and rotation information (i.e., the rotation and translation of the root node) were not considered by the algorithm. A cross-validation procedure like the one used in the HDM05 dataset was used in this case. The results (best overall correct recognition rates) for the two variants and the 3 distance/similarity measures used in the NN classifier are shown in Table 6.

Table 6 Correct recognition rates for all method variants (MHAD dataset, cross-validation)

Variant	Recognition rate	Clusters
HIST+1F	99.55	240
COS+1F	98.94	240
KL+1F	**100**	240
HIST+4F	99.85	240
COS+4F	99.39	240
KL+4F	99.85	240

Bold values indicate the best result achieved

Table 7 Correct identity recognition rates for all method variants in a 50/50 train/test split (MHAD dataset)

Variant	Recognition rate	Clusters
HIST+1F	96.06	264
COS+1F	**96.36**	264
KL+1F	95.76	264
HIST+4F	93.94	216
COS+4F	94.85	240
KL+4F	93.94	228

Bold values indicate the best result achieved

The method achieved the best recognition rate (100 %) for the KL+1F variant and for 240 clusters. The 100 % recognition rate reinforces the observation that more movements beyond gait can be used for person identity recognition.

The 50/50 % train/test split experimental setup was also applied in the MHAD database. The results for all variants in the MHAD for this setup are shown in Table 7.

As can be seen in this table, the proposed method achieves also very high identification rates despite the more challenging experimental setup.

4.3 MSR

Motion capture data recorded by special devices such as those contained in the HDM05 and MHAD datasets are usually very accurate and noise free. In order to test the robustness of the method, we also used two more challenging datasets that contain skeleton data obtained by the Kinect device. Kinect data usually suffer from noise and therefore are more difficult to use for identity recognition.

The first such dataset is the MSR Action3D dataset [17]. The dataset consists of 10 subjects performing 20 actions with 2 or 3 repetitions of each action. Some actions are not performed by all the subjects. The actions performed by the 10 subjects are *high arm wave (HighArmW), horizontal arm wave (HorizArmW), hammer (Hammer), hand catch (HandCatch), forward punch (FPunch), high throw*

Table 8 Correct identity recognition rate for all method variants in a 50/50 train/test split (MSR dataset)

Variant	Recognition rate	Clusters
HIST+1F	96.76	190
COS+1F	**97.84**	200
KL+1F	94.6	190
HIST+4F	86.69	210
COS+4F	91.73	210
KL+4F	86.69	210

Bold values indicate the best result achieved

(HighThrow), draw x (DrawX), draw tick (DrawTick), draw circle (DrawCircle), hand clap (Clap), two hand wave (TwoHandW), side-boxing (Sidebox), forward kick (FKick), side kick (SKick), jogging (Jog), tennis swing (TSwing), tennis serve TServe, bend (Bend), golf swing (Golf) and pickup & throw (PickT). It should be noted that the positions of joints are available in the MSR dataset instead of rotation angles as in HDM05 and MHAD datasets, thus the classic K-means was used to compute the Dynemes and F-Dynemes in this case. All joints of the skeletons were used by the algorithm. 50 % of the sequences were used for testing and 50 % for training. The correct recognition rates for all variants is shown in Table 8.

The method achieved the best recognition rate (97.84 %) for the COS+1F variant and for 200 clusters. The corresponding confusion matrix is shown in Fig. 7. The high recognition rate shows that the method achieves very good results even when noisy and more challenging data are considered.

Table 9 Identity recognition rate per movement for COS+1F with $C = 200$; 97.84 % overall correct classification rate (MSR dataset)

Movement	Recognition rate (%)
High arm wave	100
Horizontal arm wave	100
Hammer	100
Hand catch	100
Forward punch	100
High throw	100
Draw x	100
Draw tick	100
Draw circle	100
Hand clap	100
Two hand wave	100
Side-boxing	100
Bend	92.31
Forward kick	100
Side kick	100
Jogging	95
Tennis swing	100
Tennis serve	100
Golf swing	76.92
Pick up and throw	92.31

The correct recognition rate per action for the MSR dataset is shown in Table 9.

As can be seen in this table, in 12 out of 20 movements the correct recognition rate is 100 %. The correct

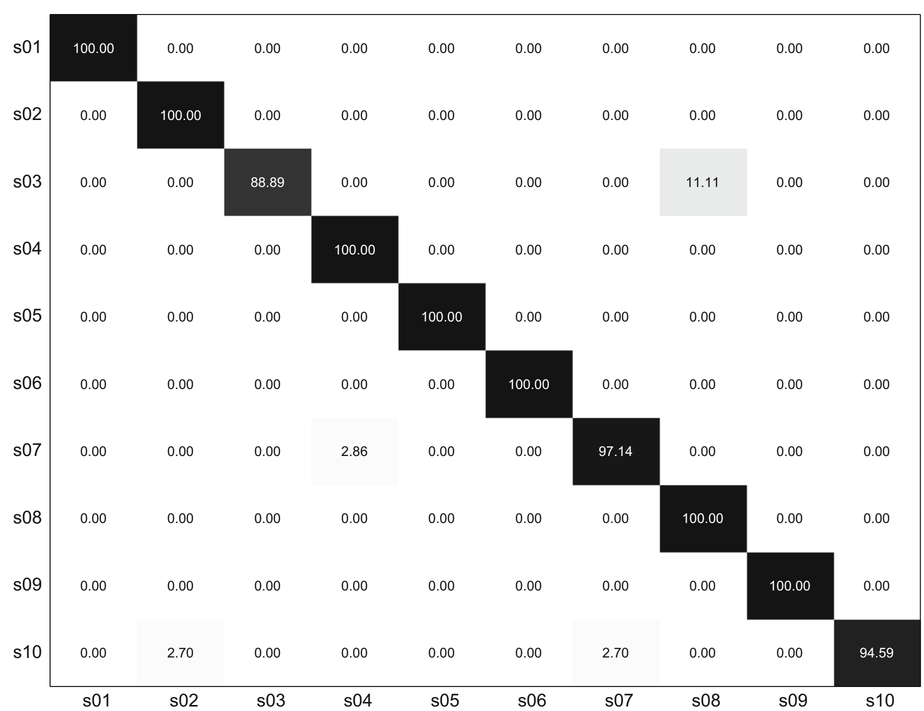

Fig. 7 Identity recognition confusion matrix (10 persons: $s01, \ldots, s10$) for the best performing variant, i.e., COS+1F with $C = 200$; 97.84 % overall correct classification rate in MSR dataset

Table 10 Correct recognition rate for all method variants (GaitD dataset) and comparison with [24]

Variant	Recognition rate	Clusters
HIST+1F	92.19	184
COS+1F	85.94	240
KL+1F	92.19	248
HIST+4F	98.44	248
COS+4F	**100**	200
KL+4F	98.44	248
[24]	91	

Bold values indicate the best result achieved

identification rate drops below 85 % only for one action (Golf swing).

4.4 GaitD

The second dataset that consists of sequences recorded by Kinect and used for the evaluation of the proposed method was GaitD, proposed in [24]. This dataset contains only gait sequences, thus it was used to test the proposed method in gait recognition only. The GaitD dataset consists of 8 subjects and each one of them is performing walking 8 times, resulting to 64 sequences. Joint positions are recorded in the GaitD dataset as in MSR. A cross-validation procedure proposed in [24] was used to assess the performance of the method in the GaitD dataset. In each cross-validation cycle a sequence from each subject is used for testing and the rest for training. The correct recognition rate and a comparison with the method proposed in [24] is shown Table 10.

As can be seen in this table, the proposed method outperforms the method proposed in [24] for all variants but COS+1F. Moreover, the COS+4F variant leads to 100 % correct identification rate.

4.5 HDM05+MHAD

One can note that the datasets used for experimental evaluation consist of only a few subjects each and thus are not very suitable for identity recognition experiments. Unfortunately, to the best of our knowledge, there are no publicly available datasets with a large number of people performing various movements. In an attempt to somehow overcome this problem, HDM05 and MHAD datasets were united to one dataset with 17 subjects. All the sequences from the subset of HDM05 dataset described in Sect. 4.1 and the MHAD dataset were used to form the HDM05+MHAD dataset. The merging of these two datasets was possible since they consist of angular-based skeleton representations. Unfortunately, the MSR dataset consist of positional data and its merging with HDM05 and MHAD in order to create an even larger dataset was not feasible. The correct identification rate of all variants of the proposed method can be seen in Table 11. The 50/50 train/test split was used in this case.

Table 11 Correct identity recognition rates for all method variants in a 50/50 train/test split (HMD05-MHAD dataset)

Variant	Recognition rate	Clusters
HIST+1F	**97.31**	425
COS+1F	97.01	425
KL+1F	**97.31**	510
HIST+4F	96.42	476
COS+4F	**97.31**	442
KL+4F	97.01	408

Bold values indicate the best result achieved

Table 12 Best correct recognition rates for the two variants in all datasets, using a 50/50 test/train split, except for the GaitD dataset where a cross-validation was used

Variant	HDM05	MHAD	GaitD	MSR	HDM05+MHAD
1F	**98.82**	**96.36**	92.18	**97.84**	**97.31**
4F	95.59	94.85	**100**	91.37	**97.31**

Bold values indicate the best result achieved in each dataset

As can be seen in this table, the proposed method achieves slightly lower but still very high identification rates in the case of the combined dataset (17 persons).

Table 12 summarizes the best recognition results for the two variants (1F and 4F) in all datasets.

It can be easily observed, that the correct recognition rates are very high in all tested datasets regardless of the choice of features.

4.6 Other experiments

4.6.1 CMC curves

Another way to assess the performance of an identity recognition (identification) method is the Cumulative Match Characteristic (CMC) curve. The CMC curve is used for closed-set identification setups. Every subject in the test sequences in such a setup has a corresponding match in the training set. A CMC curve shows the probability of identification for various ranks, i.e., the curve demonstrates the cumulative probability of correct identification of a test subject. In more detail, the proposed method computes a distance or similarity metric (depending on the variant as explained in Sect. 3) between a movement sequence of an unknown subject and all the training sequences (that include other sequences of the test subject). The CMC curve value for rank x equals the probability that the label of the test subject is included within the labels of the x most similar training sequences returned by the method. The CMC curves for the

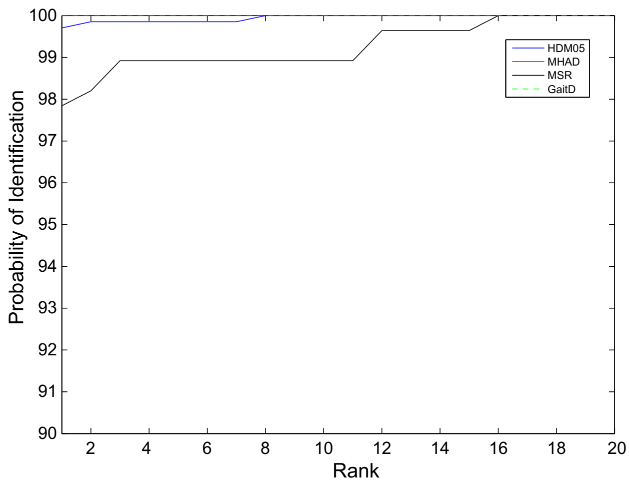

Fig. 8 CMC curves for the presented datasets

best performing variant of the proposed algorithm for each of the four presented datasets are shown in Fig. 8. According the CMC definition, the rank 1 value of these curves equals the best identification rate for the corresponding dataset, as presented in Table 12.

As can be seen in Fig. 8, the probability of correct identification reaches 100 % in rank 1 for the MHAD and GaitD datasets, in rank 8 for the HDM05 dataset and in rank 16 for MSR dataset.

4.6.2 Effect of the number of K-means clusters

The most noticeable parameter of the proposed method is the number of K-means clusters in the bag of features framework. Experiments with different number of clusters have been conducted for the HDM05, MHAD and MSR datasets. The identification rates for a wide range of cluster numbers are shown in Fig. 9a–c. As can be seen in these figures, the proposed method achieves high identification rates regardless of the number of clusters used in the K-means algorithm. Thus the method is not sensitive to this parameter. Moreover, the identification rates for all variants, in all examined datasets with the same number of clusters within each dataset is shown in Table 13. It can also be seen in this table that the identification rates are only slightly affected, compared to the results presented in Tables 5, 7 and 8.

4.6.3 Identity recognition on unseen actions

Finally, an experiment was conducted in order to test whether the algorithm can recognize persons by using some actions of these persons it has not been trained on. In this experiment, performed on the HDM05 dataset, all sequences of one action for all subjects were left out for testing and the method was trained using the remaining data. Results are presented in Table 14. Since the proposed method performs identity recognition by recognizing small differences in the way the persons

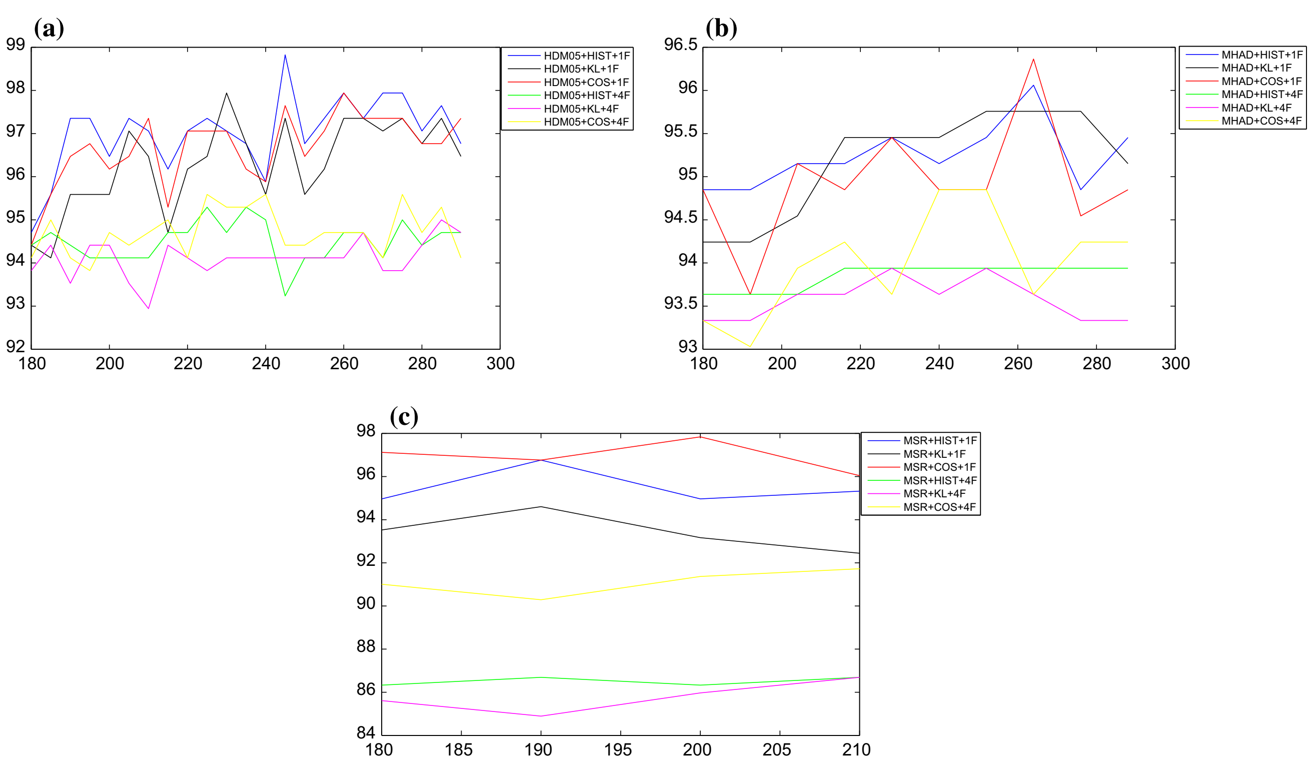

Fig. 9 Identification rates for different number of clusters. **a** HDM05 dataset, **b** MHAD dataset and **c** MSR dataset

perform various actions it is natural that if the algorithm is not trained on the actions used for identity recognition in the test phase, its performance will decrease significantly. Indeed, in a "leave one action out" cross-validation the best overall identification rate in HMD05 is only 58.24 %.

5 Conclusions and discussion

In this paper, a novel method for activity-based identity recognition in motion capture data was proposed. The method utilizes characteristic postures (Dynemes) and characteristic forward differences (F-Dynemes) derived through the K-means algorithm, along with a bag of words approach and a nearest neighbor classifier. Experimental analysis verifies that the proposed approach provides very good person identity recognition results and highlights the fact that the way humans perform various actions can be used for identity recognition. The purpose of using other movements/actions was not to improve the identification rate with respect to using only walking. The actual purpose was to show that other movements can be also used for identification. As a matter of fact, by looking Table 3 one can see that by using only walking in the HDM05 dataset results into slightly better results than using all actions (100 vs 99.7 %). In a real scenario, acquiring walking sequences of a subject might not always be feasible. Allowing the identification algorithm to operate on a set of actions increases its versatility and applicability in real usage scenario: the more the actions used, the higher the chances that a subject will be captured performing one of these actions. And this increase in versatility is not causing any serious performance deteriorations.

It should be noted that the dependence of the feature dimension $C = P \times Q$ to the number of persons P in the dataset is a mild disadvantage of the proposed method since the feature dimensionality would increase as the number of persons increases. However, as mentioned in Sect. 3.2, the results are only slightly worse when K-means is applied only once in all the sequences of all subjects. For example, when using the KL+1F variant of the method on the HDM05 dataset, the performance of the method is 97.74 %. In comparison, if K-means is applied to all sequences from all subjects, the performance drops only slightly to 96.76 %. Thus in the case where there is a large number of persons to be recognized, one application of K-means could be preferred.

Finally, It should be stressed that since the available datasets contain only a limited number of subjects, creation of larger datasets is a necessity. Testing the proposed method in such datasets, wherever they become available, would be needed in order to check its performance in a more realistic environment.

Table 13 Identification rates for the same number of clusters per dataset in a 50/50 train/test split

Variant	HDM05		MHAD		MSR	
	Rate	Clusters	Rate	Clusters	Rate	Clusters
HIST+1F	**98.82**	245	96.06	264	94.96	200
COS+1F	97.64	245	**96.36**	264	**97.84**	200
KL+1F	97.35	245	95.76	264	93.16	200
HIST+4F	93.93	245	93.94	264	86.33	200
COS+4F	94.41	245	93.64	264	91.37	200
KL+4F	94.12	245	93.64	264	85.97	200

Bold values indicate the best result achieved in each dataset

Table 14 Identity (HDM05 dataset) recognition rates in a leave one action out cross-validation experimental setup

Movement	Recognition rate (%)
runop	47.14
walk	35
hopboth	56
hopl	32
hopr	56
kickll	64.62
kickrl	73.33
punchlh	88.89
punchrh	72
clap	10
clapa	32
eltoknee	33.33
eltokneer	33.33
walks	80
walksc	60
Overall rate	58.24

All the sequences of the action denoted in the first column form the test set in each cycle. Histogram intersection was used as a distance function in the classification in this experiment

References

1. Barbosa, I.B., Cristani, M., Del Bue, A., Bazzani, L., Murino, V.: Re-identification with rgb-d sensors. In: Proceedings of the 12th International Conference on Computer Vision, ECCV'12, vol. Part I, pp. 433–442. Springer, Berlin, Heidelberg (2012)
2. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. Lett. **31**(13), 2052–2060 (2010)
3. Boulgouris, N., Hatzinakos, D., Plataniotis, K.: Gait recognition: a challenging signal processing technology for biometric identification. Signal Process. Mag. IEEE **22**(6), 78–90 (2005)
4. Boyd, J.E.: Synchronization of oscillations for machine perception of gaits. Comput. Vision Image Underst. **96**(1), 35–59 (2004)
5. Burdea, G., Coiffet, P.: Virtual reality technology, 2nd edn. Wiley, New York (2003)
6. Gkalelis, N., Tefas, A., Pitas, I.: Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. Circuits Syst Video Technol IEEE Trans **18**(11), 1511–1521 (2008). doi:10.1109/TCSVT.2008.2005617

7. Gkalelis, N., Tefas, A., Pitas, I.: Human identification from human movements. In: Proceedings of the 16th IEEE International Conference on Image Processing, ICIP'09, pp. 2557–2560. IEEE Press, Piscataway (2009)
8. Gu, J., Ding, X., Wang, S., Wu, Y.: Action and gait recognition from recovered 3-D human joints. IEEE Trans. Syst. Man Cybern. Part B **40**(4), 1021–1033 (2010)
9. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
10. Iosifidis, A., Tefas, A., Pitas, I.: Activity-based person identification using fuzzy representation and discriminant learning. IEEE Trans. Inf. Forensics Secur. **7**(2), 530–542 (2012)
11. Iosifidis, A., Tefas, A., Pitas, I.: Dynamic action recognition based on dynemes and extreme learning machine. Pattern Recogn. Lett. **34**(15), 1890–1898 (2013) (smart Approaches for Human Action Recognition)
12. Josiński, H., Świtoński, A., Jędrasiak, K., Kostrzewa, D.: Human identification based on gait motion capture data. In: Proceedings of the 2012 International MultiConference of Engineers and Computer Scientists, IMECS'12 (2012)
13. Kapsouras, I., Nikolaidis, N.: Action recognition on motion capture data using a dynemes and forward differences representation. J. Vis. Commun. Image R **25**, 1432–1445 (2014)
14. Lam, T.H.W., Lee, R.S.T.: A new representation for human gait recognition: motion silhouettes image (msi). In: Proceedings of the 2006 International Conference on Advances in Biometrics, ICB'06, pp. 612–618. Springer, Berlin, Heidelberg (2006)
15. Lee, T., Belkhatir, M., Sanei, S.: A comprehensive review of past and present vision-based techniques for gait recognition. Multimed. Tools Appl. **72**, 2833–2869 (2014)
16. Lin, Y.C., Yang, B.S., Lin, Y.T., Yang, Y.T.: Human recognition based on kinematics and kinetics of gait. J. Med. Biol. Eng. **31**(4), 255–263 (2011)
17. Liu, F., Zhuang, Y., Wu, F., Pan, Y.: 3D motion retrieval with motion index tree. Comput. Vis. Image Underst. **92**(2–3), 265–284 (2003)
18. Liu, L.F., Jia, W., Zhu, Y.H.: Survey of gait recognition. In: Huang, D.S., Jo, K.H., Lee, H.H., Kang, H.J., Bevilacqua, V. (eds.) Emerging intelligent computing technology and applications. With aspects of artificial intelligence, lecture notes in computer science, vol. 5755, pp. 652–659. Springer, Berlin, Heidelberg (2009)
19. Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., Weber, A.: Documentation mocap database HDM05. Tech. Rep. CG-2007-2. Universität Bonn (2007)
20. Munsell, B.C., Temlyakov, A., Qu, C., Wang, S.: Person identification using full-body motion and anthropometric biometrics from kinect videos. In: Proceedings of the 12th international conference on Computer Vision, ECCV'12, vol. Part III, pp. 91–100. Springer, Berlin, Heidelberg (2012)
21. Nikolaidis, N., Pitas, I.: Nonlinear processing and analysis of angular signals. IEEE Trans. Signal Process. **46**(12), 3181–3194 (1998)
22. Nixon, M.S., Carter, J.N.: Advances in automatic gait recognition. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, FGR'04, pp. 139–144. IEEE Computer Society, Washington, DC (2004)
23. Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., Bajcsy, R.: Sequence of the most informative joints (smij): a new representation for human skeletal action recognition. J. Vis. Commun. Image Recogn. **25**(1), 24–38 (2014) (visual Understanding and Applications with RGB-D Cameras)
24. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: Gait recognition with kinect. In: Proceedings of the First Workshop on Kinect in Pervasive Computing (2012)
25. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Gait energy volumes and frontal gait recognition using depth images. In: Proceedings of the 2011 International Joint Conference on Biometrics, IJCB '11, pp. 1–6. IEEE Computer Society, Washington, DC (2011)
26. Tanawongsuwan, R., Bobick, A.: Gait recognition from time-normalized joint-angle trajectories in the walking plane. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001., vol. 2, pp. II–726 (2001)
27. Wang, J., She, M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification. In: Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications, DICTA '10, pp. 320–327. IEEE Computer Society, Washington, DC (2010)
28. Yam, C.Y., Nixon, M.S., Carter, J.N.: Extended model-based automatic gait recognition of walking and running. In: Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA '01, pp. 278–283. Springer, London (2001)
29. Zhang, Z., Hu, M., Wang, Y.: A survey of advances in biometric gait recognition. In: Proceedings of the 6th Chinese Conference on Biometric Recognition, CCBR'11, pp 150–158. Springer, Berlin, Heidelberg (2011)

Ioannis Kapsouras was born in Larisa, Greece, received the Degree in Informatics and MSc in Digital Media, both from the Aristotle University of Thessaloniki. He is currently a Ph.D. candidate at the Artificial Intelligence and Information Analysis (AIIA) laboratory of the Department of Informatics in Aristotle University of Thessaloniki. He has participated in one research project financed by European funds. Ioannis Kapsouras has co-authored 1 journal paper and 4 conference papers. His current research interests include computer graphics, video processing, computer vision and pattern recognition.

Nikos Nikolaidis is Assistant Professor at the Department of Informatics, Aristotle University of Thessaloniki. He has co-authored 1 book, 15 book chapters, 48 journal papers and 152 conference papers and co-edited one book and two special issues in journals. The number of citations to his work by third authors exceeds 3000 (h-index 24). He has participated in 24 research projects funded by the EU and national funds. His areas of interest/expertise include anthropocentric video analysis (human detection and tracking, activity recognition), computer vision, digital image/video processing, stereoscopic/multiview video processing/analysis, computer graphics and visualization. Dr. Nikolaidis is currently serving as associate editor for the EURASIP Journal on Image and Video Processing and 4 more international journals. He served as Exhibits chair of IEEE ICIP 2001, Technical Program chair of IEEE IVMSP 2013 workshop and is currently serving as Publicity co-chair of EUSIPCO 2015. He will be publicity chair of IEEE ICIP 2018. He is an IEEE Senior Member.