ACTIVITY-BASED METHODS FOR PERSON RECOGNITION IN MOTION CAPTURE SEQUENCES

Eftychia Fotiadou, Nikos Nikolaidis

Aristotle University of Thessaloniki
Department of Informatics
{eftifot,nikolaid} @aiia.csd.auth.gr

ABSTRACT

In this paper we present two algorithms for efficient person recognition operating upon motion capture data, depicting persons performing various everyday activities. The first approach is driven from the assumption that, if two motion sequences depict a certain activity performed by the same person, then, consecutive frames (poses) of one sequence are expected to be similar to consecutive frames of the other. The proposed method constructs a pose correspondence matrix to represent the similarity between poses and utilizes an intuitive method for estimating a similarity score between two motion capture sequences, based on the structure of the correspondence matrix. The second algorithm is based on a Bag of Words model (BoW), where histograms are extracted from motion sequences, based on the frequency of occurrences of characteristic poses. This method is combined with the application of Locality Preserving Projections (LPP) on the data, in order to reduce their dimensionality. Our methods achieved more than 98% correct person recognition rate, in three different datasets.

Index Terms**—** person recognition, motion capture, classification, dimensionality reduction

1. INTRODUCTION

Motion capture has long been used in film and games industry in order to allow for more realistic renditions of human and animal motion. It is also useful in virtual reality applications, as well as in various disciplines that involve the study of human motion, such as ergonomic analysis, sports biomechanics and rehabilitation. The result of motion capture is a skeletal animation sequence, i.e. a series of skeleton configurations (poses) over time. Today, motion capture is becoming more affordable. As an example, the Kinect sensor with the accompanying human skeleton tracking software can deliver fairly good motion capture data with minimal cost. As a consequence, motion capture techniques gain more and more ground, giving boost to the development of diverse kinds of applications.

Person recognition refers to the process by which the identity of a person is recognized by a system, based on information that he or she carries. Examples of applications involving person recognition include security or surveillance systems, access control, patient monitoring, as well as a wide range of systems involving human-computer interaction. Traditionally, recognition is performed by means of credentials, supplied by the person, in form of IDs, smart cards or passwords. However, in the last decades, an increasing use of biometric features is observed [1]. These features may include physiological characteristics of a person, such as fingerprints, face/iris characteristics, palm prints or DNA, as well as behavioural characteristics [2], such as gait or style when performing a certain action, keyboard typing, and voice. Biometric characteristics show advantage over the aforementioned credentials, with respect to counterfeiting or loss risk.

To our knowledge, there do not exist methods for activity-based person recognition from motion capture data, other than those related to human gait analysis. Although those algorithms apply mostly on video motion data, several methods have been recently proposed, for gait-based person recognition (usually referred to simply as gait recognition) from motion capture sequences. In [3], the proposed method for gait recognition is based on the analysis of the trajectories of lower body joint angles, projected onto the sagittal plane. At the first step of the method, the joint angles are estimated by fitting a skeleton model to the sensor measurements. Thereafter, the trajectories are normalized with respect to duration and walk cycles, by means of a segmentation technique and Dynamic Time Warping. Finally, the trajectories are classified using a nearest-neighbor classifier, based on euclidean distance. In [4], motion capture data are combined with measurements from force plates, in order to combine both kinematic and kinetic data. As a result, the features representing the data include joint angles and angular velocities, as well as forces applied on joints. Classification of gait data is performed using Self Organizing Maps (SOMs). Additionally, the importance and contribution of each feature is investigated, in order for the factors that cause differences in gait to be determined.

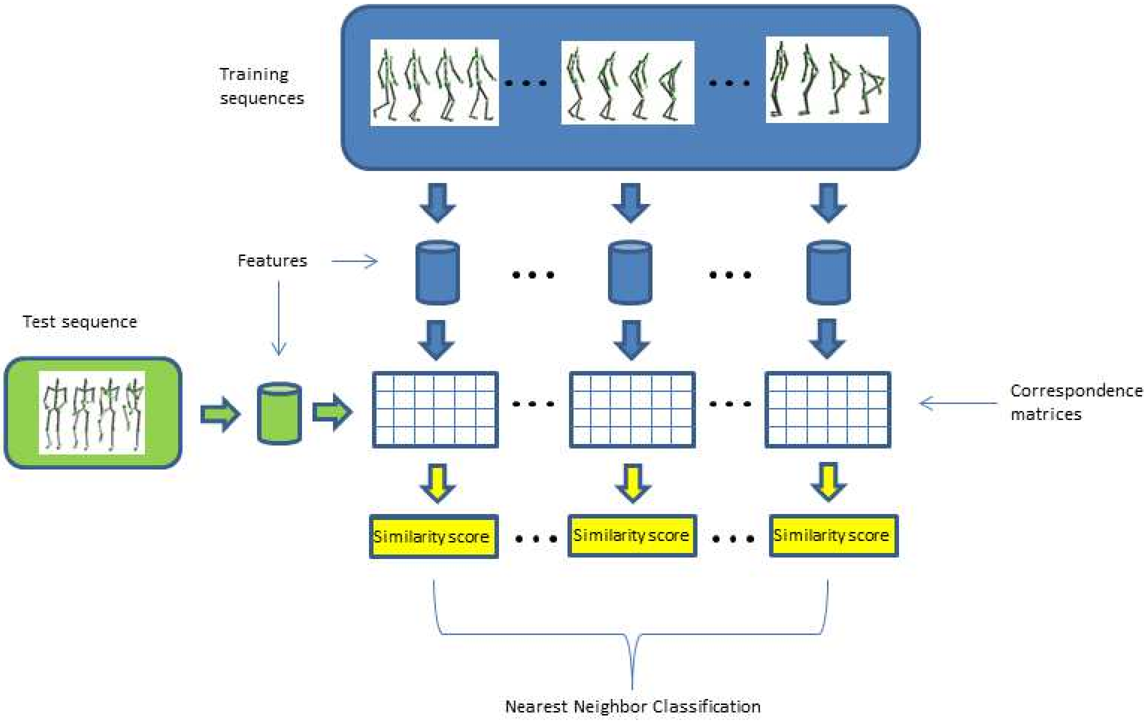

Fig. 1: Flowchart for the correspondence-based method.

The method presented in [5] is suitable for both classification of gait type (walking, running, jogging and limping) and person recognition. Deriving from the fact that the perception of human motion by an observer relies on the detection of specific ”motion features”, representing relative motion of body parts, a two-stage PCA scheme is applied on the motion data. The first stage of PCA is applied on the data, represented by joint angles and velocities, in order for a trajectory on a low-dimensional manifold to be extracted. The second stage of PCA detects the variability in the shape of this manifold across individuals or gait types. In [6], gait recognition is performed on motion data acquired from recordings with the Kinect sensor. The 3D positional joint data are used to extract both static (e.g. height, body part lengths) and dynamic (speed, step length) features. Gait data are classified by three different types of classifiers, namely Naive Bayes, 1R and C4.5. Also, the influence of the different features on the recognition rate is investigated. In [7] the trajectories of specific parts of the human body during walking, referred to as gait paths, are used for the extraction of features suitable for person recognition. From the motion capture data, four different kinds of features are extracted and subsequently classified using a Naive Bayes and a k-Nearest Neighbor approach.

The two methods proposed in this paper follow a more general approach to the person recognition problem, in comparison to the aforementioned methods. Recognition is based on motion capture data representing a repertory of different classes of human actions, such as waving, sitting down or standing up, and not solely on gait. Thus, the proposed approaches broaden the applicability of movement-based person recognition methods, to cover a large set of actions. As a matter of fact, the proposed approaches indicate that, other human actions apart from walking bear significant person-specific characteristics, that allow person recognition with high recognition rates. The first proposed algorithm is based on the hypothesis that, motion sequences of the same action performed by the same person, exhibit strong similarity between successive frames in one or more segments within them, which is expressed through specific patterns. In order to classify motion capture sequences to distinct humans, we developed a scheme for similarity estimation between such sequences. The second algorithm we propose consists a Bag of Words (BoW) approach, that combines dimensionality reduction of the motion data as a pre-processing step.

2. CORRESPONDENCE-BASED PERSON RECOGNITION

As already mentioned, the first proposed method for activity-based person recognition is based on the similarity between two motion capture sequences and comprises of two distinct steps:

1. The construction of a correspondence matrix, that de-

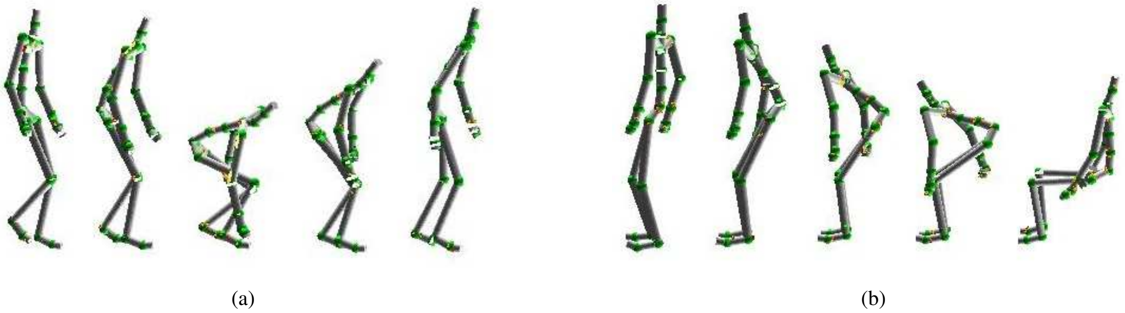

Fig. 2: Sample frames from motion capture sequences of the activity classes "deposit floor" (a) and "sit down chair" (b), from the HDM05 database.

scribes which frame in the second sequence is the most similar to each frame in the first sequence.

2. A process of calculating a similarity score between two sequences from the correspondence matrix.

Person recognition is subsequently performed, by classifying test motion sequences, using an 1-Nearest Neighbor classifier. The workflow of the correspondence-based method is summarized in Figure 1. In the following subsections the aforementioned steps are described in detail.

2.1. Correspondence matrix construction

Let us consider two motion capture sequences denoted with $\mathbf{X}_s = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$ and $\mathbf{Y}_s = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}$, consisting of M and N frames respectively. A frame in a sequence consists of the rotation angles of each joint and describes the pose of the human body at a certain time instance. Some frames from motion sequences describing the activities "deposit floor" and "sit down chair" are illustrated in Figure 2. In order to construct the correspondence matrix, the distances from each pose $\mathbf{y}_i$ of the sequence $\mathbf{Y}_s$ to every pose $\mathbf{x}_i$ in sequence $\mathbf{X}_s$ are calculated.

For this purpose, we use a distance measure, based on the logarithmic representation of the rotational data of each joint. Let us assume that a pose is represented by J unit quaternions ($\hat{q}_k$), each of them describing the rotation on one of the J joints. A unit quaternion can be projected to the tangent space at some reference point of the 3-sphere that the unit quaternions lie on. The reference point we selected for the projection of the data, was the sample mean $\hat{q}_m$ of quaternions for each joint calculated over a number of frames, which was estimated as the quaternion that minimizes the sum of its squared distances from the other quaternions, as described in [8]. The aforementioned projection is performed by applying a logarithmic mapping:

$$\log^{(\hat{q}_m)}(\hat{q}) = \ln(\hat{q}_m^* \times \hat{q}), \tag{1}$$

where × denotes the quaternion multiplication and $\hat{q}_m^*$ is the conjugate of the unit quaternion representing the sample mean. In this way, quaternions can be mapped to 3D points in Euclidean space. Consequently, the distance between two rotations represented by quaternions, can be approximated by the Euclidean distance between two points in $\mathbb{R}^3$. Therefore, each joint rotation can be represented by a 3D point $\mathbf{P} = \{p_1, p_2, p_3\}$, and a pose of a skeleton consisting of J joints can be denoted as $\mathbf{x} = \{\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_J\} = \{p_1, p_2, ..., p_{3J}\}$, i.e. as a vector of $3J$ elements. The distance between two such poses $\mathbf{x}$ and $\mathbf{y}$ can then be estimated by the Euclidean distance between the two pose vectors:

$$d^{Log}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{3J}(p_i - q_i)^2} \tag{2}$$

The calculated distances between all pairs of poses in the two sequences $\mathbf{X}_s$, $\mathbf{Y}_s$ are used to construct a correspondence matrix of dimensionality $M \times N$, denoted with $\mathbf{C}$. The rows/columns of $\mathbf{C}$ correspond to poses of sequence $\mathbf{X}_s$,$\mathbf{Y}_s$ respectively. For each pose $\mathbf{x}_i$ of $\mathbf{X}_s$, the nearest pose $\mathbf{y}_j$ of $\mathbf{Y}_s$ is found and the element (i, j) of $\mathbf{C}$ is set to one, whereas all other elements $(i, k), k \neq j$ of the i-th row are set to zero.

The result of this process is, that $\mathbf{C}$ exhibits distinct structures depending on the similarity between the two sequences under examination. When the two compared sequences describe movements of the same class (e.g. two walking sequences) the correspondence matrix contains diagonal segments of ones, of various lengths, either continuous or interrupted, since successive poses from one sequence are in general most similar to successive poses from the other. Specifically, these diagonal segments extend from the upper left to the bottom right of the matrix. In case that the two sequences depict the same movements performed by different subjects, these diagonal segments tend to be smaller in length and weaker in terms of slope. When the sequences describe motions of different classes, irrespective of whether they come from the same subject or not, there are two possibilities: First, there may exist long vertical lines, implying that many poses in sequence $\mathbf{X}_s$ are matched to the same

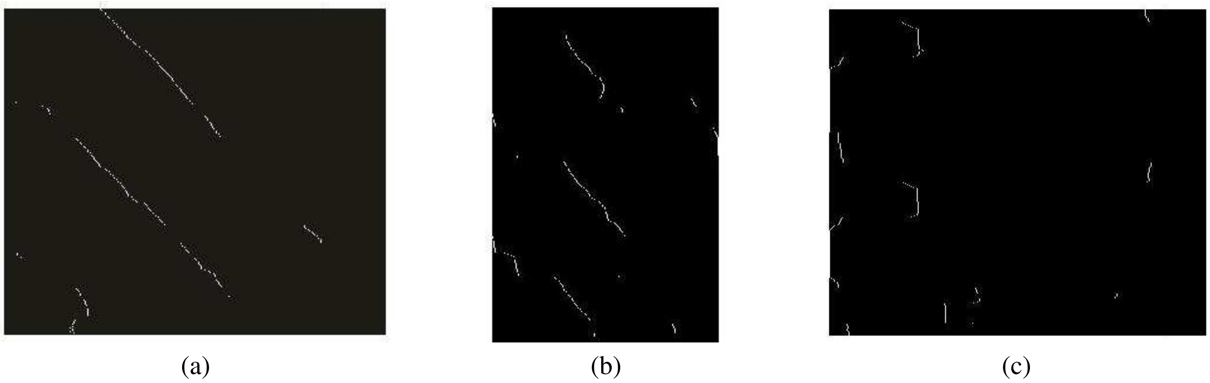

Fig. 3: Examples of correspondence matrices: two movements of the same class performed by the same (a) and different subjects (b). Two movements of different classes performed by the same subject (c).

pose in $\mathbf{Y}_s$. This is often the case, when the two movements described in the sequences are different, but share one or more similar poses. Second, the correspondence matrix may exhibit diagonal segments of limited length or the ones (units) may be arranged with no particular structure, a fact indicating, that the two sequences describe completely different movements. Examples of correspondence matrices are shown in Figure 3, where the unit entries are represented by white pixels.

2.2. Similarity score evaluation

In order to quantify the similarity of two motion capture sequences $\mathbf{X}_s$,$\mathbf{Y}_s$, a score S is calculated over the respective correspondence matrix $\mathbf{C}$, based on the existence and the structure of diagonal segments. The higher the score, the more similar the two sequences are.

For each row of matrix $\mathbf{C}$ (which corresponds to a pose of sequence $\mathbf{X}_s$), the position of the unit entry (column index) is retrieved, in order to determine the relative position of the unit entries in subsequent poses and to identify possible diagonal segments. The relative position of a unit in the next row with respect to the unit in the current row defines whether the next unit lies in a "legal" position or not, according to the rules described below. These rules try to take into account the fact that, although for two matching sequences, units (matching poses) should ideally form a diagonal (45° slope) segment consisting of connected elements (i.e. units should be arranged in matrix cells (i, j), $(i + 1, j + 1)$, $(i + 2, j + 2)$ and so on), deviations from this ideal situation (e.g. gaps of limited extend) should be allowed.

Table 1: Legal (marked with a tick) and illegal (marked with an x) positions for a unit entry in the correspondence matrix.

1	0	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
x	√	√	√	√	x

First, a maximum of three consecutive points within a diagonal segment are allowed to lie in a vertical placement, such as (i, j),$(i+1, j)$,$(i+2, j)$. This means, that up to three consecutive poses of sequence $\mathbf{X}_s$ are allowed to be matched to the same pose of sequence $\mathbf{Y}_s$. Furthermore, if the current unit is at cell (i, j), then the unit in the next row can be either in the "ideal" position $(i + 1, j + 1)$, or in positions $(i + 1, j)$ (vertical placement), $(i + 1, j + 2)$, $(i + 1, j + 3)$. In other words, gaps of length 1 and 2 are allowed. Finally, the length of each diagonal segment, that is, the number of units lying on it, should be larger than a threshold T_l, in order for it to contribute to the final score. It was observed that, a threshold between 5 and 9 produces the best results. An example of legal and illegal positions for a unit entry in the 4-th row of a correspondence matrix, given the arrangement of units in the previous 3 rows, is shown in Table 1.

After the aforementioned process has been performed for every pose of sequence $\mathbf{X}_s$, all the units lying in "legal" positions, have been assigned to a diagonal segment. Each unit entry to the correspondence matrix lying in a diagonal segment, is assigned a weight $\alpha_{i,j}$. A weight equal to 1 is assigned if consecutive poses of sequence $\mathbf{X}_s$ are matched to exactly consecutive poses in sequence $\mathbf{Y}_s$, (i.e. units are in an arrangement (i, j), $(i+1, j+1)$) and a weight equal to 0.8 otherwise. These weights, whose role is to favor exactly diagonal and penalize "imperfect" diagonal unit arrangements, were determined by performing experiments with different values and selecting those that achieve the best recognition rate. However, there is a reasonable interval around the "optimal" weight values (1 and 0.8) that lead to similar results. Thus, the effect of the weight values to the method performance in not critical.

At the end of the procedure, the segments containing a number of points below threshold T_l are discarded as invalid, while the valid segments contribute to the calculation of the total score. The total score S is estimated by summing the weights of the units lying in valid diagonal segments:

$$S = \sum_{(i,j) \in V} \alpha_{i,j}, \tag{3}$$

where V is the set of all units lying in valid diagonal segments.

The actual classification of a test sequence, i.e. a sequence that has not been assigned a label with respect to the subject that performs it, is performed using an 1-Nearest Neighbor classifier. In other words, the test sequence is tested against all training sequences and is labeled with the subject label of the training sequence that yielded the biggest similarity score.

3. PERSON RECOGNITION USING A BAG OF WORDS APPROACH AND DIMENSIONALITY REDUCTION

The second person recognition method that has been developed is based on a Bag of Words (BoW) approach [9] and utilizes dimensionality reduction of the data as a pre-processing step. The proposed algorithm consists of three steps:

1. Application of dimensionality reduction on the motion capture data, using the Locality Preserving Projections (LPP) method [10].

2. A histogram-based representation of motion sequences, based on a set of characteristic poses.

3. Classification of the histograms calculated from test sequences, using a Support Vector Machine (SVM) classifier.

The basic steps of the method can be seen in the flowchart of Figure 4. In the following sections, the aforementioned steps are described in more detail.

3.1. Dimensionality reduction

Dimensionality reduction methods are frequently used in machine learning applications, in order to project high-dimensional data onto a low-dimensional space. Apart from the reduction in the amount of data per se, which can be crucial in certain applications, machine learning tasks can also benefit from dimensionality reduction in other ways. For example, it can be used to extract more meaningful features, that carry useful information and describe better the data, to discover the structure of the data or to reduce noise and irrelevant information. Linear dimensionality reduction techniques such as Principal Component Analysis (PCA) [11] and Linear Discriminant Analysis (LDA) [12] are the most commonly used ones in machine learning applications, however, non-linear techniques, such as Isomap [13], Locally Linear Embedding (LLE) [14] and Laplacian Eigenmaps [15] are also used with increasing frequency.

In order to apply dimensionality reduction to motion capture data, we should consider that they are non-linear. Therefore, non-linear techniques could be more suitable for that type of data. In particular, local non-linear techniques, such as LLE and Laplacian Eigenmaps, have the property to preserve the structure of the local neighborhood around the datapoints, which is important when dealing with motion data. On the other hand, non-linear techniques are computationally more expensive and they lack a way of projecting new test samples, as the mapping they provide is defined only on the training samples. For the aforementioned reasons, we selected to use the Locality Preserving Projections (LPP) technique in conjunction with our recognition method. Although LPP is a linear dimensionality reduction technique, it combines properties of both linear and non-linear techniques: as a linear method, it is defined everywhere and can be applied effectively and fast in practical applications. Furthermore, like local non-linear techniques, it preserves the local structure of the data.

In the following sub-section, the LPP method and its application to motion capture data is briefly explained.

3.1.1. Locality Preserving Projections and application to motion capture data

The LPP method assumes that the data lie on a low dimensional manifold embedded in the high dimensional space. It calculates a linear transformation, which maps the data from the high dimensional space, with dimensionality D, to a subspace of dimensionality d, while preserving the local neighborhood information, in the sense of preserving the pairwise distances between neighboring points.

The LPP algorithm consists of the following steps:

1. **Construction of an adjacency graph**: Let us assume a graph $\mathbf{G}$ consisting of B nodes. An edge is put between two nodes i and j if the datapoints $\mathbf{x}_i$ and $\mathbf{x}_j$ are close. In order to determine how close two datapoints are, the k-nearest neighbors of a datapoint are considered: nodes i and j are connected with an edge if either the point $\mathbf{x}_i$ is among the k nearest neighbors of $\mathbf{x}_j$, or $\mathbf{x}_j$ is among the k nearest neighbors of $\mathbf{x}_i$.

2. **Choice of weights**: a sparse weight matrix $\mathbf{W}$ of size $B \times B$ is calculated. Each element W_{ij} has a value equal to the weight of the edge connecting the nodes i and j or equal to 0 if there is no edge between i and j. For the assignment of weights to the edges, the heat kernel function is used:

$$W_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}} \tag{4}$$

3. **Eigenmaps**: the requirement for close points in the initial space to be mapped to close points in the reduced

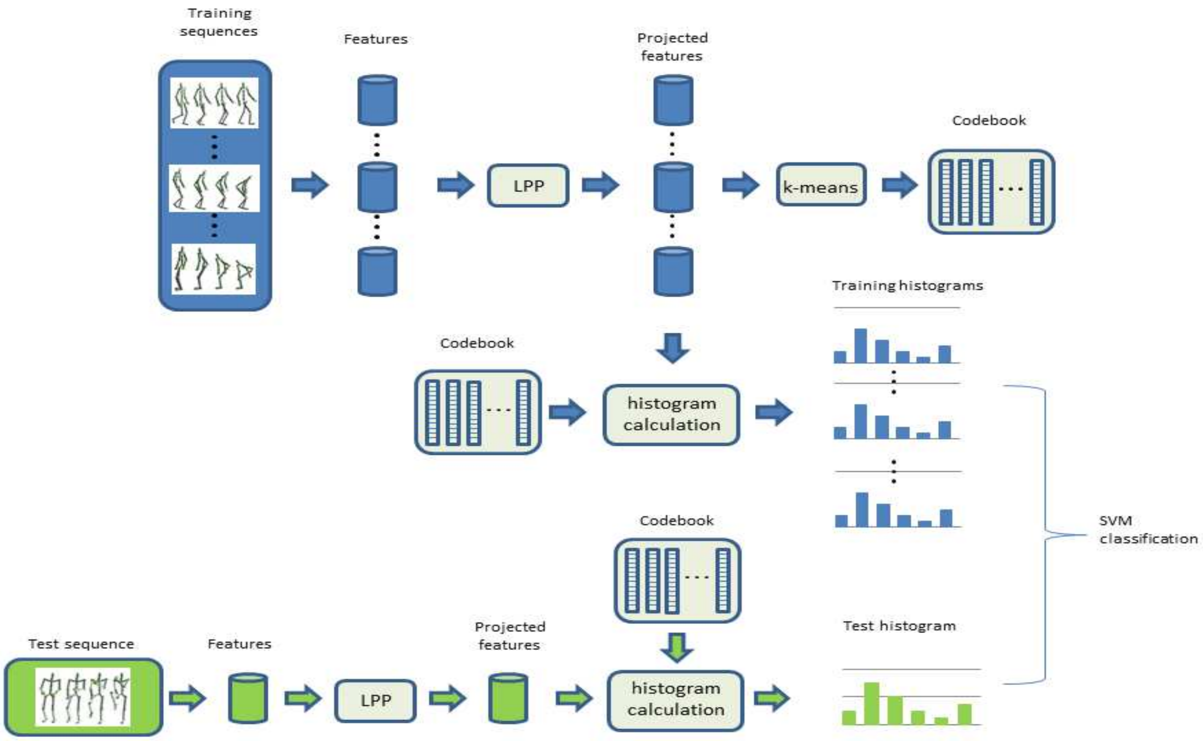

Fig. 4: Flowchart for the LPP/BoW method.

space is met by the following cost function, the minimization of which results to the optimal projection:

$$\Phi(\mathbf{y}) = \sum_{ij} \left\| \mathbf{y}_i - \mathbf{y}_j \right\|^2 W_{ij}, \tag{5}$$

where $\mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i$ is the projection of a datapoint $\mathbf{x}_i$. It can be proven, that the transformation matrix $\mathbf{A}$ is calculated by solving the generalized eigenvalue problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a} = \lambda \mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{a}, \tag{6}$$

where $\mathbf{M}$ is the degree matrix, while $\mathbf{L} = \mathbf{M} - \mathbf{W}$ is the graph Laplacian. The transformation matrix $\mathbf{A}$ is constructed by the eigenvectors corresponding to the first d eigenvalues $\lambda_0 < \lambda_1 < ... < \lambda_{d-1}$.

From the above discussion, it can be observed, that the LPP algorithm involves up to three parameters, that need to be determined: the dimension d of the low-dimensional space, the number of neighboring datapoints k in the construction of the adjacency graph and the parameter t of the heat kernel function. In our case, the optimal parameters were defined as those that lead to the highest recognition rate and were found using grid search.

In order to apply the LPP method to motion capture data, the frames of a training set of motion sequences are concatenated in a matrix. In more detail, let us assume that there are k training sequences, $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k$, consisting of $N_1, N_2, ..., N_k$ frames (poses) respectively, and that each frame (sample) is represented by a D-dimensional feature vector. Then, a matrix $\mathbf{X}$, with $\sum_i N_i$ rows and D columns is formed by concatenating all frames from all training sequences. For the representation of the rotational data, we used the logarithmic mapping of the quaternions, as described in section 2.1. Thus, D equals the number of the joints times three, as there are three components describing the information for each joint in the logarithmic mapping representation. By applying the LPP technique to the training data matrix $\mathbf{X}$, a transform matrix $\mathbf{A}$ is calculated, which can be used to project the training and testing sequences. Subsequently, the projected data are used as input by the classification method presented in the next section.

3.2. Bag of Words approach

In order to apply a Bag of Words approach, a codebook of characteristic poses, called dynemes, is calculated from the set of training sequences. This is achieved by applying the k-means algorithm on the projected frames of the training set, in order to cluster them into C clusters. Each dyneme represents the centroid of a cluster. Subsequently, for each training motion sequence a histogram can be calculated as follows: a motion sequence $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ is transformed to a sequence $\mathbf{X}_D = \{\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_N\}$, where each frame $\mathbf{d}_i$

is derived by replacing the frame $\mathbf{x}_i$ with the dyneme which is closest to it, i.e. the center of the cluster $\mathbf{x}_i$ has been assigned to. By calculating the frequency of occurrence for each one of the C dynemes in the sequence $\mathbf{X}_D$, a histogram $\mathbf{s} = \{s_1, s_2, ..., s_C\}$ of the dyneme appearances can be constructed.

A similar procedure is followed in order to calculate the histogram for a test sequence. For each frame of a test sequence, the distances to all dynemes are estimated, and the frame is replaced with the dyneme which is closest to it. Again, the histogram is constructed by calculating the frequencies of occurrence of each dyneme in the sequence.

Once the histograms for all training sequences have been calculated, they are used along with their labels, which in our case refer to the person that performs the movement depicted in the sequence, to train a Support Vector Machine (SVM) classifier [16]. In order to classify test sequences, the corresponding histograms are extracted and subsequently fed to the SVM classifier. The SVM classifier used involves a chi-squared kernel function [17] given by the following formula:

$$K(\mathbf{s}_i, \mathbf{s}_j) = \exp(-\frac{1}{A} D(\mathbf{s}_i, \mathbf{s}_j)) \tag{7}$$

where $D(\mathbf{s}_i, \mathbf{s}_j)$ denotes the χ^2 distance between $\mathbf{s}_i = \{s_{i1}, s_{i2}, ..., s_{iC}\}$ and $\mathbf{s}_j = \{s_{j1}, s_{j2}, ..., s_{jC}\}$:

$$D(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2} \sum_{k=1}^{C} \frac{(s_{ik} - s_{jk})^2}{s_{ik} + s_{jk}}, \tag{8}$$

while A is a scaling factor, calculated as the mean of χ^2 distances between all training samples.

4. DATASETS

In order to test the two proposed methods, we used the HDM05 and the MHAD datasets, that were mainly devised to test the performance of activity recognition and motion retrieval algorithms, as in [18], [19], [20], [21], [22], as well as a gait dataset, used in gait recognition experiments.

4.1. HDM05 database

The HDM05 motion capture database [23] consists of motion capture files, for various types of activities, performed by five actors. Our person recognition task included 16 classes of activities (393 files in total) for the five subjects: depositFloorR, elbowToKnee3RepsLelbowStart, grabHighR, hopBothLegs3hops, jogLeftCircle4StepsRstart, kickRFront1Reps, lieDownFloor, rotateArmsBothBackward3Reps, sneak4StepsR start, squat1Reps, throwBasketball, hopBothLegs1hops, jumpingJack1Reps, throwStandingHighR, sitDownChair and standUpSitChair. Motion files in ASF/AMC format were used.

From the motion capture clips, we selected the rotation information for a subset of 13 joints, namely lower back, upper back, thorax, right humerus, right radius, left humerus, left radius, right femur, right foot, left femur, left tibia, left foot and right tibia, since these joints were observed to be the most informative, and therefore more discriminant for our recognition task.

Since the number of sequences of each motion for each subject in the dataset is different, we selected a fixed train (276 motion sequences) and test (117 motion sequences) data set for our experiments.

4.2. Berkeley Multimodal Human Action Database (MHAD) database

The MHAD [24] database contains motion capture data for a set of 11 motion classes, namely Jumping in place, Jumping jacks, Bending - hands up all the way down, Punching (boxing), Waving - two hands, Waving - one hand (right), Clapping hands, Throwing a ball, Sit down then stand up, Sit down, and Stand up. These motions are performed by 12 different subjects. For each subject, there are 5 sequences of each motion, resulting to 659 (there is a missing file) sequences in total. Motion files in BVH format were used in our experiments. Only rotations of the following joints were considered: spine, spine1, spine2, RightShoulder, RightArm, RightForeArm, LeftShoulder, LeftArm, LeftForeArm, RightUpLeg, RightLeg, RightFoot, LeftUpLeg, LeftLeg and LeftFoot.

The fact that the dataset contains equal number of trials for each motion by each subject, allows us to use two distinct configurations for our experiments: (a) we build fixed train (396 sequences) and test (263 sequences) sets and (b) we use a cross validation scheme, where in each fold the test set consists of a sequence by each subject, for the same motion (12 samples in total) and the training set includes the remaining sequences.

4.3. Gait data

In addition to the databases presented above, we also performed experiments using gait data from [6]. The dataset consists of 67 files, each of them containing the 3D coordinates of specific joints, recorded with Microsoft's Kinect sensor. Gait data are collected for nine different persons and each of them (except for one) performs 8 trials. Since the data refer to the position rather than the rotation of each joint, we used the 3D coordinates for our experiments. As a pre-processing step, we canceled out the global translation and rotation of the root, from all the skeleton joints. Through this process, each pose in a walking sequence becomes independent of the orientation and position of the person in the 3D space. Subsequently, a frame (pose) is treated as a vector consisting of 3D points, thus, having a dimensionality equal to the number of

Table 2: Correct person recognition rates

Database	**Correspondence**	**BoW initial dimensionality**		**LPP/BoW**		
	Rate	Rate	Dimensionality (D)	Rate	Dimensionality (d)	Parameters
HDM05 (fixed sets)	91.45%	95.73%	39	**98.29%**	14	k=40, t=3.7, C=110
MHAD (fixed sets)	96.2%	97.72%	45	**98.1%**	10	k=60, t=3.8, C=110
MHAD (folds)	98.02%	98.77%	45	**98.79%**	14	k=40, t=3.7, C=110
Gait data	85.4%	94.44%	18	**98.57%**	8	k=60, t=3.9, C=80

selected joints times three. As with the two aforementioned datasets, we selected to use the positional data for a subset of joints, located on the legs, namely: HipLeft, KneeLeft, AnkleLeft, HipRight, KneeRight and AnkleRight. For our experiments we used a 7-fold cross validation scheme, as the one described in [6].

5. EXPERIMENTAL RESULTS

Several parameters have to be defined, for the approach that combines LPP dimensionality reduction and Bag of Words. Different values for the number of clusters C (and consequently the number of dynemes) in the k-means algorithm were tested, whereas a grid search was performed for the values of the parameters d, k and t of the LPP technique, described in section 3. Correct recognition rates for the two methods are presented in Table 2. Furthermore, in order to show the effect of the dimensionality reduction step, we also display the recognition rates for the BoW method when applied on the data of initial dimensionality, i.e. without dimensionality reduction.

It can be observed, that both methods provide high recognition rates, with the LPP/BoW method always performing better than the other one. In the cases of HDM05 database and the gait data, the LPP/BoW method achieved a significantly better recognition rate, while in MHAD database both methods achieved similar performance. Additionally, although lower recognition rates for the MHAD database would be expected, since it includes more subjects, results in this datasets were equally good or better than in the HDM05. This is partially due to the fact that, the MHAD dataset contains more sequences for each movement for the subjects, allowing for better training. As far as dimensionality reduction is concerned, it can be observed, that LPP significantly improved the recognition rates of the BoW approach in the case of HDM05 and gait data, while it had a minor effect in the case of MHAD data, for which high recognition rates were already obtained without dimensionality reduction and thus not much room for improvement existed.

The aforementioned results indicate that there also exist various human movements, apart from gait, which bear discriminative information and can therefore be suitable for the person recognition task. For example, all motion classes of the HDM05 dataset achieved 100% person recognition rate, except for sneak4StepsRstart and standUpSitChair where rates 80% and 83.33% were achieved respectively, when the BoW/LPP method was applied for person recognition. Most probably, person recognition rates in these two motion classes were lower because these classes bear less discriminant information for person recognition, than the other classes.

Regarding the gait dataset, the LPP/BoW method outperformed the recognition method proposed in [6], that achieves a maximum 91% correct recognition rate. It should be noted that no comparisons with other methods that perform person recognition on motion capture data depicting various movements are provided, because, to the best of our knowledge, no such methods exist in the literature.

6. CONCLUSIONS AND FUTURE WORK

Two methods for person recognition on motion capture data have been presented in this paper. The first algorithm is based on the structure of a correspondence matrix between motion capture sequences, whereas the second algorithm adopts a Bag of Words approach combined with dimensionality reduction on the motion capture data, using the LPP method. Both methods achieved high person recognition rates on the test datasets. An important finding of this work is that human movements other than walking (gait) can be successfully used for person recognition. Regarding future work, additional features, such as the velocity and acceleration of joints could be explored. Testing the algorithms on other datasets will be also attempted.

Acknowledgments

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operation Program ”Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS-UOA-ERASITECHNIS MIS 375435. The authors would like to thank Mr. Ioannis Kapsouras for his valuable assistance in code development.

7. REFERENCES

[1] A.K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[2] R. V. Yampolskiy and Venu Govindaraju, “Behavioural biometrics: a survey and classification,” *Int. J. Biometrics*, vol. 1, no. 1, pp. 81–113, June 2008.

[3] R. Tanawongsuwan and A. Bobick, “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, 2001. CVPR 2001.*, 2001, vol. 2, pp. II–726–II–731 vol.2.

[4] Y.-C. Lin, B.-S. Yang, and Y.-T. Yang, “People Recognition by Kinematics and Kinetics of Gait,” in *13th International Conference on Biomedical Engineering*, vol. 23 of *IFMBE Proceedings*, pp. 1996–1999. Springer Berlin Heidelberg, 2009.

[5] S.R. Das, R.C. Wilson, M.T. Lazarewicz, and L.H. Finkel, “Gait Recognition by Two-Stage Principal Component Analysis,” in *7th International Conference on Automatic Face and Gesture Recognition, 2006. FGR 2006.*, 2006, pp. 579–584.

[6] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, “Gait Recognition with Kinect,” in *1st International Workshop on Kinect in Pervasive Computing*, Newcastle, 2012.

[7] A. Šwitoński, A. Polański, and K. Wojciechowski, “Human Identification Based on Gait Paths,” in *Proceedings of the 13th international conference on Advanced concepts for intelligent vision systems*, Berlin, Heidelberg, 2011, ACIVS’11, pp. 531–542, Springer-Verlag.

[8] M. P. Johnson, *Exploiting Quaternions to Support Expressive Interactive Character Motion*, Ph.D. thesis, MIT Media Lab, 2003.

[9] H. Wang, M. M. Ullah, A. Klaser, I. Laptev, and C. Schmid, “Evaluation of local spatio-temporal features for action recognition,” in *BMVC 2009 - British Machine Vision Conference*, Sept. 2009.

[10] X. He and P. Niyogi, “Locality Preserving Projections,” in *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[11] I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, 1986.

[12] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 7, pp. 179–188, 1936.

[13] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[14] S. T. Roweis and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *SCIENCE*, vol. 290, pp. 2323–2326, 2000.

[15] M. Belkin and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation,” *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, June 2003.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, Wiley-Interscience, 2000.

[17] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study,” in *Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW ’06.*, 2006, pp. 13–13.

[18] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, “Sequence of the Most Informative Joints (SMIJ): A new representation for human skeletal action recognition,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2012, pp. 8–13.

[19] X. Chen and M. Koskela, “Classification of RGB-D and Motion Capture Sequences Using Extreme Learning Machine,” in *Image Analysis*, pp. 640–651. Springer Berlin Heidelberg, 2013.

[20] Mohamed E. Hussein, Marwan Torki, Mohammad A. Gowayyed, and Motaz El-Saban, “Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations,” in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. 2013, IJCAI’13, pp. 2466–2472, AAAI Press.

[21] Thibaut Naour, Nicolas Courty, and Sylvie Gibet, “Fast motion retrieval with the distance input space,” in *Motion in Games*, vol. 7660 of *Lecture Notes in Computer Science*, pp. 362–365. Springer Berlin Heidelberg, 2012.

[22] T. Huang, H. Liu, and G. Ding, “Motion Retrieval Based on Kinetic Features in Large Motion Database,” in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. 2012, ICMI ’12, pp. 209–216, ACM.

[23] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation Mocap Database HDM05," Tech. Rep. CG-2007-2, Universität Bonn, June 2007.

[24] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley MHAD: A Comprehensive Multimodal Human Action Database," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, 2013, pp. 53–60.