Person Identity Recognition on Motion Capture Data Using Label Propagation

Nikos Nikolaidis Charalambos Symeonidis
AIIA Lab, Department of Informatics
Aristotle University of Thessaloniki
Greece
email: nikolaid@aiia.csd.auth.gr

***Abstract*—Most activity-based person identity recognition methods operate on video data. Moreover, the vast majority of these methods focus on gait recognition. Obviously, recognition of a subject's identity using only gait imposes limitations to the applicability of the corresponding methods whereas a method capable of recognizing the subject's identity from various activities would be much more widely applicable. In this paper, a new method for activity-based identity recognition operating on motion capture data, that can recognize the subject's identity from a variety of activities is proposed. The method combines an existing approach for feature extraction from motion capture sequences with a label propagation algorithm for classification. The method and its variants (including a novel one, that takes advantage of the fact that, in certain cases, both activity and person identity labels might exist for the labeled sequences) have been tested in two different datasets. Experimental analysis proves that the proposed approach provides very good person identity recognition results, surpassing those obtained by two other methods.**

I. Introduction

Motion capture (mocap) data describe the locations of human body joints or the joint angles over time. Skeleton models used in mocap consist of nodes that represent the joints of the skeleton and arcs that represent the segments. Mocap data can be obtained by using various tracking devices such as magnetic, ultrasonic, inertial, optical, mechanical etc [1]. Such data can also be obtained with the use of the Kinect sensor or other RGB-D sensors. Most joints have 3 rotational degrees of freedom (DOF) except from the root node that has, in addition, 3 translation DOF. Examples of mocap sequences are shown in Fig. 1.

Person identification (identity recognition) is a very active area, face recognition being perhaps the most widely researched topic within this broad research area. Another category of identification approaches aim at recognizing the identity of a person by the way he or she performs one or more activities. One such approach is gait recognition that deals with the identification of subjects by their walking style. Walk sequences can be captured on video or by using a motion capture system, including the inexpensive and non-invasive Kinect device.

Obviously recognition of a subject's identity using only gait imposes limitations to the applicability of the corresponding methods. Indeed, humans are engaged in various activities such as running, sitting, waving etc, and a method capable of recognizing the subject's identity from such different activities would be much more widely applicable. Furthermore, an algorithm that can operate on various activities might yield better results compared to a gait recognition system. This is because activities other than walking, might provide more discriminant information, thus leading to higher recognition rates.

So far, considerable research efforts have been devoted to activity-based identity recognition on video data, although the vast majority of these methods focus on gait recognition. Activity-based person identification methods applied on skeletal animation or motion capture data are almost non existent, since research on mocap data focuses mainly on motion indexing and retrieval as well as activity recognition. Approaches that use mocap data for activity-based person recognition are very few and deal only with gait [2], [3], [4], [5]. The authors are aware of only two other approaches that perform activity-based person identification using skeletal animation / motion capture data. Indeed in [6] a method for person identity recognition using motion capture data depicting persons performing various actions is proposed. The joints positions or orientation angles and the forward differences of these quantities are used to represent a motion capture sequence. Initially, clustering using K-means 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 manner. Recognition is performed by using a nearest neighbor classifier. In [7] two algorithms for person recognition operating upon motion capture data, depicting persons performing various everyday activities are proposed. The first approach is based on the assumption that, if two motion capture sequences depict a specific activity performed by the same person, consecutive frames/poses of one sequence shall be similar to consecutive frames of the other. The method constructs a pose correspondence matrix to represent the similarity between poses and uses a method that is based on the structure of the correspondence matrix to estimate a similarity score between two sequences. The second approach is based on a Bag of Words model (BoW), where, similar to [6], histograms are extracted from motion sequences, based on the frequency of occurrences of characteristic poses.

ISBN 978-0-9928626-7-1 © EURASIP 2017

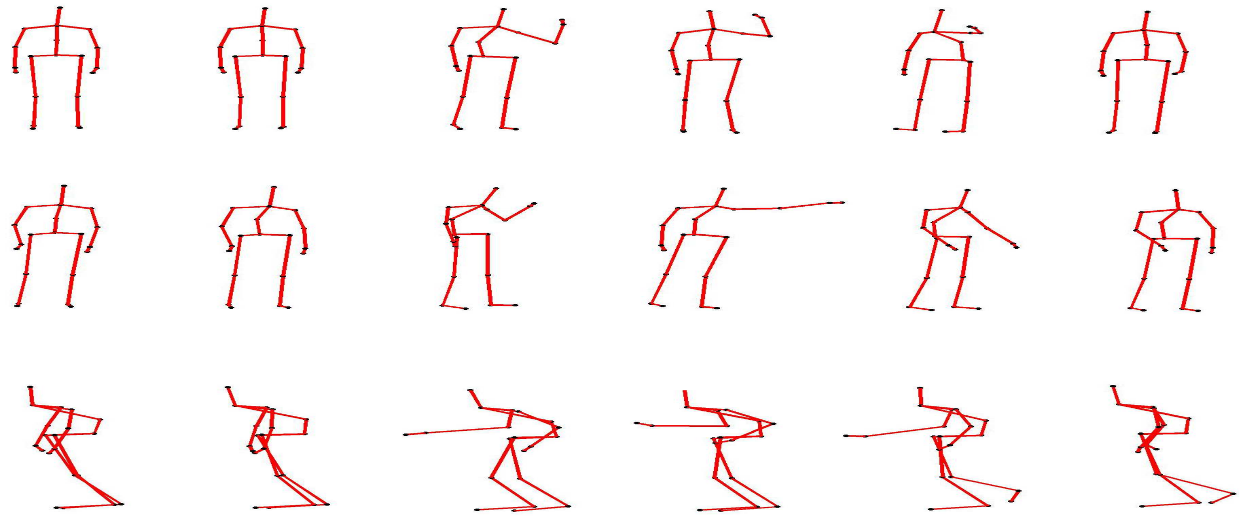

Fig. 1. Selected frames from three action sequences of the MSR Action3D dataset: side-boxing (upper row), high throw (middle row), jogging (lower row).

The method uses Locality Preserving Projections (LPP) on the data, in order to reduce their dimensionality. Recognition is performed by using a Support Vector Machine (SVM).

This paper presents a novel activity-based identity recognition method that operates on motion capture data depicting humans in various activities. The method uses the features proposed in [6] to represent each mocap sequence and subsequently applies a label propagation approach proposed in [8] for the classification. Label propagation aims at propagating label information (in our case person identity labels) from a number of labeled data to data without labels, based on data similarity. Label propagation approaches have been used in various digital media (images, videos) related tasks [9]. The proposed approach, which, as far as we know, is the first one to use label propagation in this task, has been applied to two motion capture datasets and has been shown to provide superior recognition results compared to those obtained by both [6] and [7]. A novel variant, that utilizes knowledge of the activity depicted in the motion capture sequences, in case such information is available, is also presented.

The rest of the paper is organized as follows. The proposed method is described in detail in Section II. In Section III, experimental performance evaluation of the method is presented. Conclusions follow.

II. Proposed Method

The proposed person identity recognition method consists of two building blocks: extraction of suitable representations (features) for the motion sequences and classification of the sequences into one of the person classes through label propagation. The aforementioned modules are described in detail sections II-A and II-B, respectively.

A. Data Representation

In order to extract a representation for the motion capture data, we follow the Bag of Words (BoW) - based approach proposed in [6]. According to this approach, motion data are represented by histograms of occurrence of codewords from a codebook of characteristic poses (dynemes and F-dynemes). As already mentioned, motion capture data provide information about the configuration of the moving body (positions or rotations of specific joints in the human body) in discrete time steps. The motion information used in our method is in the form of rotation angles on the joints of a skeletal hierarchy or joint positions. Therefore, the body configuration (pose) in the i-th frame of a motion capture sequence is described by a posture vector comprising of either a set of rotation angles:

$$\mathbf{p}_i = \{\theta_{i1}, \theta_{i2}, ..., \theta_{ir}\}, \quad i = 1, ..., M \tag{1}$$

where r is the number of rotation angles of the skeletal hierarchy joints and M the number of frames in the sequence, or by a set of joints positions:

$$\mathbf{p}_i = \{x_{i1}, y_{i1}, z_{i1}, ..., x_{il}, y_{il}, z_{il}\}, \quad i = 1, ..., M \tag{2}$$

where l is the number of joints.

In addition to the posture vector, vectors of forward differences between posture vectors of the current and subsequent frames are evaluated, so as to capture the dynamics of motion. A forward difference vector for frame i is calculated as:

$$\mathbf{d}_i^t = \mathbf{p}_{i+t} - \mathbf{p}_i. \tag{3}$$

Forward differences vectors for $t = 1, 5$ and 10 were used. As a result, a motion capture sequence consisting of M frames is

ISBN 978-0-9928626-7-1 © EURASIP 2017

described by four different types of feature vectors:

$$\begin{aligned}
&\mathbf{V}_1 = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_M\}\\
&\mathbf{V}_2 = \left\{\mathbf{d}_1^1, \mathbf{d}_2^1, ..., \mathbf{d}_{M-1}^1\right\}\\
&\mathbf{V}_3 = \left\{\mathbf{d}_1^5, \mathbf{d}_2^5, ..., \mathbf{d}_{M-5}^5\right\}\\
&\mathbf{V}_4 = \left\{\mathbf{d}_1^{10}, \mathbf{d}_2^{10}, ..., \mathbf{d}_{M-10}^{10}\right\}
\end{aligned} \tag{4}$$

In order to construct a BoW model, codebooks are first calculated from the training data. For each feature vector type $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_4)$, a separate codebook consisting of C codewords (dynemes for the case of posture vectors, F-dynemes for forward differences) is calculated using the standard k-Means algorithm in the case of mocap data where joint positions are provided, or the angular k-means algorithm proposed in [6], in case of data were rotation angles are given. Dynemes and F-dynemes are the centers of the clusters discovered by the k-means, Due to the calculation procedure, dynemes correspond to "average", characteristic postures rather than specific postures from within the dataset.

Subsequently, features are mapped to the codewords. Each posture vector is mapped to its closest dyneme wereas each forward differences vector is mapped to its closest F-dyneme. Thus each sequence is represented in terms of Dynemes and F-Dynemes. More specifically, each frame is represented by one dyneme and 3 F-dynemes. Then for a specific motion sequence, a set of four C-dimensional histograms $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4$ are calculated. These histograms are obtained by calculating the frequency of appearance of every dyneme and F-dyneme for each corresponding set of features.

The final feature vector describing a motion capture sequence is formed by concatenating the histograms corresponding to the four feature types:

$$\mathbf{x} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4]\,. \tag{5}$$

B. Classification by Label Propagation

In order to classify motion capture sequences to different person-identity classes, we employed the label propagation algorithm proposed in [8]. This algorithm propagates the label information from a set of initially labeled data samples, referred to as seeds, to samples with unknown labels. A brief description of this algorithm is provided below.

Let us assume a dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ consisting of N samples and a set of different labels $\mathcal{L} = \{1, 2, ..., L\}$, that can be assigned to the samples. We consider that the first k samples have known labels $y_i, i = 1, ..., k$, while the remaining $N - k$ samples have no labels. In our case, the samples are the motion capture sequences, represented by the features (concatenation of histograms calculated according to the BoW model) described in Section II-A. The labeled samples correspond to the training set, while the unlabeled ones to the testing set. Let us define a matrix $\mathbf{Y}$ of size $N \times L$, which contains the initial labels and is given by:

$$Y_{ij} = \begin{cases} 1, & \text{if sample } i \text{ has label } j \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In addition, we consider a $N \times L$ matrix $\mathbf{F} = \left[\mathbf{F}_1^T, ..., \mathbf{F}_N^T\right]^T$, which assigns labels to each sample, according to $y_i = \arg\max_j F_{ij}, 1 \leq j \leq L$. The label propagation is performed following the steps below [8]:

1) Matrix $\mathbf{W}$ of size $N \times N$ is constructed. $\mathbf{W}$ contains the similarities between pairs of samples:

$$\mathbf{W}_{ij} = \begin{cases} \exp(s * HI()), & \text{if } i \neq j \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where s is a parameter taking values in the range $[4, 6]$ and HI is the histogram intersection metric calculated, for two histogram vectors $\mathbf{x}_i, \mathbf{x}_j$ by:

$$HI(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{4C} \min\{x_{il}, x_{jl}\}\,. \tag{8}$$

2) Matrix $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is calculated, where $\mathbf{D}$ is the diagonal matrix with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$.
3) Matrix $\mathbf{F}$ is calculated as:

$$\mathbf{F} = (\mathbf{I} - \alpha\mathbf{S})^{-1}\mathbf{Y}, \tag{9}$$

where α is a parameter taking values in $(0, 1)$. The optimal value for α is determined through experimentation.
4) The final label is assigned to sample $\mathbf{x}_i$ according to:

$$y_i = \arg\max_j F_{ij}, 1 \leq j \leq L \tag{10}$$

Another novel approach that was tested in order to recognize the identity of each person was to take advantage of the fact that in certain cases both activity and person identity labels might exist for the labeled sequences, i.e. those that belong to the training set, In such a case, one can use the following approach: a) Use the features described in Section II-A and the label propagation algorithm described above in order to propagate activity labels from the labeled data (training set) to the unlabeled ones (test set). By doing so, all data (samples) obtain an activity label. b) Perform person identity label propagation separately on sequences depicting the same activity. In other words, if the dataset contains L different activity classes, it is split into L subsets (each containing sequences of the same activity) and identity label propagation is performed separately in each subset. This approach will be subsequently called double label propagation (DLP).

III. Experimental Evaluation

The performance of the proposed method was tested on two publicly available datasets: Berkeley Multimodal Human Action Database (MHAD) and MSR Action3D Dataset. Both datasets contain sequences including multiple repetitions of each activity performed by a number of subjects. The datasets were split into two sets in a 50%-50% manner.

A. Berkeley MHAD

The Berkeley MHAD dataset [10] contains motion capture data depicting the following 11 activities: jumping in place, jumping jacks, bending with hands up all the way down, punching, waving with two hands, waving with one hand(right), clapping hands, throwing a ball, sitting down and then standing up, sitting down, standing up. Those activities are performed by 12 subjects; seven males and five females. Each subject performs every activity five times, yielding to 660 action sequences. Each of these sequences is labeled with the activity it represents and the subject who performed the activity. The dataset has been divided into two equal subsets, namely a training set and a testing set, both containing 330 sequences each. For each subject and activity 2 or 3 repetitions were assigned to the training set and the remaining ones to the test set. The experiments were performed using either the posture vectors features $\mathbf{h}_1$ only, or the features $\mathbf{x}$ derived by the combination of the posture vectors and the forward differences (FD). The double label propagation (DLP) approach presented in Section II-B was also tested The best identity recognition results for the four different variants are shown in table I, along with the results obtained in this dataset from the methods in [6] and [7], which, as already stated in the Introduction, are the only ones that perform activity-based identity recognition on motion capture data depicting various activities.

TABLE I
CORRECT RECOGNITION RATES: MHAD DATASET.

Algorithm	Recognition rate
Proposed, Posture	99.7
Proposed, Posture+FD	99.7
Proposed, Posture, DLP	**100**
Proposed, Posture+FD, DLP	99.7
[6]	99.39
[7]	98.1

As can be seen in this table the identity recognition rates achieved by all four variants of the proposed method are superior to those achieved by the methods in [6] and [7]. The best results (perfect identity recognition) were obtained by using posture features only, along with double label propagation. It should be noted that the results of the method [7] are not directly comparable with those obtained by the proposed methods because a different split of the dataset into training and test set (training: 396 sequences, testing: 263 sequences) was used in [7]. It should be also noted that results presented in [6] for this dataset were lower (96.36%) than the ones in the table above, since in [6] the dataset was split into a 50%-50% manner using a different combination of sequences.

The percentages of correctly recognized identities when a specific type of movement is considered have been also evaluated. As expected, due to the very large overall correct identity recognition rate, the method achieves 100% correct recognition rate for all movements/actions with the exception of ”clapping hands” (97.14% when only postures are used) and ”sit down” (97.14% when both postures and forward differences are utilized).

B. MSR Action3D Dataset

The MSR Action3D dataset contains motion capture sequences from the following 20 activities: high arm wave (HighArmW), horizontal arm wave (HorizArmW), hammer (Hammer), hand catch (HandCatch), forward punch (FPunch), high throw (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), pickup and throw (PickT). Those actions are performed by 10 subjects, each performing every activity two or three times. Thus, the dataset contains a total of 567 action sequences.

The dataset is divided into two subsets of equal size (50%-50% partition), as in Section III-A; a 284 sequences training set and a 283 sequences test set. The results of identity recognition of all variants of the proposed method can be seen in table II. The table includes also the results obtained by the method in [6]. No results are provided for the method in [7], since no experiments were conducted in this paper on the MSR dataset.

TABLE II
CORRECT RECOGNITION RATES: MSR DATASET.

Case	Recognition rate
Proposed, Posture	98.94
Proposed, Posture+FD	**99.29**
Proposed, Posture, DLP	97.17
Proposed, Posture+FD, DLP	98.94
[6]	97.84

By observing this table one can see that the proposed method performs very well in this dataset, surpassing the method in [6] in all its variants but the one that involves posture vectors and double label propagation.

Similar to the MHAD dataset, the percentages of correctly recognized identities when a specific type of movement is considered have been also evaluated. Again, as expected, due to the very large overall correct recognition rate, the method achieves 100% correct recognition rate for all movements/actions with the exception of ”side-boxing” (96.67%), ”pick-up and throw” (97.14%) and ”high arm wave” (96.30%, when only posture features are used).

IV. CONCLUSION

In this paper, a new method for activity-based identity recognition in motion capture data, a subject that has barely been researched so far, has been proposed. The method can recognize persons from various types of activities and combines the approach proposed in [6] for feature extraction with a label propagation algorithm for classification. The method and its variants have been tested in two different datasets. Experimental analysis proved that the proposed approach provides very good person identity recognition results, surpassing those obtained by the other two existing methods that perform activity-based identity recognition in motion capture data the authors are aware of. Although the improvements are not dramatic, they show that label propagation can indeed be used for activity-based person identity recognition or similar tasks. It should be noted that, since the currently available datasets contain only a limited number of subjects, creation of larger datasets is needed. Evaluating the performance of the proposed method in such datasets, whenever they become available in the research community, would be needed in order to check its performance in a more realistic environment and derive more definitive conclusions. Future work will also include using other spatio-temporal features for the description of the motion capture sequences, applying label propagation for activity recognition on motion capture data and applying the proposed approach on gait-only motion capture datasets in order to compare it with identity recognition methods that operate on gait data only.

References

[1] G. Burdea and P. Coiffet, *Virtual Reality Technology*, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 2003.

[2] 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. CVPR 2001.*, vol. 2, 2001, pp. II–726 – II–731 vol.2.

[3] H. Josiński, A. Świtoński, K. Jędrasiak, and D. Kostrzewa, "Human identification based on gait motion capture data," in *Proceedings of the 2012 International MultiConference of Engineers and Computer Scientists*, ser. IMECS'12, 2012.

[4] Y.-C. Lin, B.-S. Yang, Y.-T. Lin, and Y.-T. Yang, "Human recognition based on kinematics and kinetics of gait," *Journal of Medical and Biological Engineering*, vol. 31, no. 4, pp. 255–263, 2011.

[5] J. Gu, X. Ding, S. Wang, and Y. Wu, "Action and gait recognition from recovered 3-D human joints," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 40, no. 4, pp. 1021–1033, Aug. 2010.

[6] I. Kapsouras and N. Nikolaidis, "Person identity recognition on motion capture data using multiple actions." *Machine Vision and Applications*, vol. 65, no. 7-8, pp. 905–918, 2015.

[7] E. Fotiadou and N. Nikolaidis, "Activity-based methods for person recognition in motion capture sequences," *Pattern Recognition Letters*, vol. 49, pp. 48 – 54, 2014.

[8] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, ser. NIPS'03, 2003, pp. 321–328.

[9] O. Zoidi, E. Fotiadou, N. Nikolaidis, and I. Pitas, "Graph-based label propagation in digital media: A review," *ACM Computing Surveys*, vol. 47, no. 3, pp. 48:1–48:35, Apr. 2015.

[10] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley MHAD: A comprehensive multimodal human action database." in *Proceedings of the IEEE Workshop on Applications on Computer Vision (WACV)*, 2013.

ISBN 978-0-9928626-7-1 © EURASIP 2017