

Action Recognition in Motion Capture Data Using a Bag of Postures Approach

Ioannis Kapsouras, Nikos Nikolaidis
Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
Email: {jkapsouras,nikolaid}@aiaa.csd.auth.gr

Abstract—In this paper we introduce a novel method for movement recognition in motion capture data. A movement is regarded as a combination of basic movement patterns, the so-called dynemes. Initially a K-means variant that takes into account the periodic nature of angular data is applied on training data to discover the most discriminative dynemes. Each frame is then assigned to one of these dynemes and a histogram that describes the frequency of occurrence of these dynemes for each movement is constructed. SVM classification and sparse representation based classification are used for movement recognition on the test data. The effectiveness and robustness of this method is shown through experimental results on a standard dataset of motion capture data.

I. INTRODUCTION

Three-dimensional motion capture (mocap) data provide a representation of the complex spatio-temporal structure of human motion. During a motion capture session, the locations of characteristic parts on the human body such as joints or the joint angles are recorded over time, using appropriate tracking devices [1]. Different tracking technologies (magnetic, ultrasonic, inertial, optical, mechanical) are in use today. Motion capture data, usually in the form of joint angles, are used in computer games and animation movies to animate a hierarchical structure (skeleton) representing a human [2], where the nodes model the joints of the skeleton and the arcs the segments (links). Some examples of mocap data are shown in Figure 1.

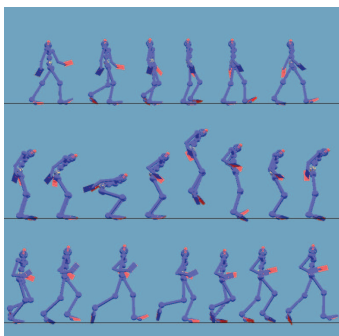


Fig. 1. Walk, hop and run movement sequences from the HDM05 database [3]

A node's degrees of freedom depend on the allowable rotations and translations of the corresponding joint in the skeleton. Usually all joints have 3 rotational degrees of freedom (with respect to each of the three axes) whereas the root node has

also 3 translation parameters. The angle values of a certain frame form the n -dimensional pose or posture vector.

This paper presents a movement recognition method that operates on motion capture data and is based on the bag of words paradigm. As already mentioned, in a motion capture sequence, a single frame is represented by the posture vector of the human body. K-means is applied on the postures space of the training data to discover characteristic posture patterns, called dynemes. Then each posture is mapped to a specific dyneme using a nearest neighbour approach and a histogram with the frequencies of appearance of the dynemes for each sequence is formed. Classification is implemented using two different approaches either by using support vector machines (SVM) or by applying the sparse representation based classification approach [4]. Principal Component Analysis (PCA) is applied before the SVM classification in order to project the histogram representation in a subspace of lower dimensionality. An important advantage of the proposed method is that it does not involve temporal information and thus it is not affected by speed variations among subjects performing the same movement. Despite the fact that ignoring the temporal information might seem as disadvantage, the experimental results verify that the method leads to very good recognition rates. In addition, the method does not require segmentation of motion capture sequences into "atomic" movements such as steps.

An additional novelty of this method is that it uses a modified K-means algorithm that can handle angular data such as the joint angles involved in mocap. The need for such a modification arises from the fact that angular data are periodic and their natural representation is on the unit circle. Thus, the notions of distance and mean value for angular data are different from those for data on the line.

The remaining of this paper is organized as follows. In Section II, we present a review on previous work. In Section III, the proposed method is described in detail. In Section IV, experimental performance evaluation of the method and comparison with other approaches is presented. Conclusions follow in Section V.

II. PRIOR WORK

Motion capture became popular only during the last years so the body of research for movement recognition on mocap data is not as extensive as for video data. Research on motion capture data focuses mainly in motion indexing and retrieval rather than movement recognition. The two tasks have many

differences but also bear important similarities such as the extraction of a suitable representation for each sequence and the need to define an appropriate similarity measure.

Li et al. [5] propose a method for recognition and classification of motion capture data. Their method uses singular value decomposition (SVD) to extract feature vectors from motion data. SVM classifiers are used to segment and recognize motion streams. SVM classification applied on the vector of 3D locations of characteristic points on the human body is used by Wang et al. in [6] for human movement recognition. Kadu et al. [7] adopt the tree-structured vector quantization (TSVQ) method to represent human poses by codewords and approximate the dynamics of mocap sequences by a codeword sequence. For the classification, the authors use a spatial domain approach based on the histogram of codewords and a spatial-time domain approach via codeword sequence matching. An algorithm for sequence alignment and activity recognition, called IsoCCA, is described in [8]. IsoCCA extends the canonical correlation analysis (CCA) algorithm, by means of introducing a number of alternative monotonicity constraints. The activity classification task performed in this paper is based on a 1-NN classifier, that uses the alignment cost between sequences as distance metric, and yields improved classification rates in comparison to other alignment algorithms, such as CTW, DTW, Hungarian and CCA. In [9] the authors introduce a method for real-time classification of dance gestures from skeletal animation. An angular skeleton representation that maps the motion data to a smaller set of features is used. The full torso is fitted with a single reference frame. This frame is used to parametrize the orientation estimates of both the first-degree limb joints (joints adjacent to torso) and second-degree limb joints (tips of the wireframe extremities such as the hands and the feet). Then a cascaded correlation-based maximum-likelihood multivariate classifier is used to build a statistical model for each gesture class. The trained classifier compares the input data with the gesture model of each class and outputs a max-likelihood score. An input gesture is finally compared with a prototype one using a distance metric that involves dynamic time-warping. In [10], Lv et al. present a method for movement recognition where each movement is represented as a spatio-temporal template consisting of a set of channels with weights. The channels correspond to the 3D joints trajectories and the weights are learned according to the Neyman-Pearson criterion. Movements are recognized by comparing them with the templates. In [11], Deng et al. propose a method for human motion recognition. First the method partitions a human model in five parts, namely, torso, left upper limb, right upper limb, left lower limb and right lower limb and K-means is applied separately to each of these partitions. Then several trials from each K-means class are used to train a generalized model to represent that class. For isolated motion recognition the authors propose a voting scheme that can be used with common dynamic programming techniques and they also present a new penalty-based similarity measure for DTW. For continuous motion recognition, five body partition index maps are constructed and applied. Concepts from the theory of chaotic systems are used by the framework proposed by Ali et al. in [12] to model and analyze nonlinear dynamics of human actions. The authors use the trajectories of reference body joints to create time series by considering each data dimension separately. Mutual information and false nearest

neighbourhood algorithms are used to embed each time series in a phase space of an appropriate dimension. Phase space invariants are then used to represent the dynamical and metric structure of the phase space. The invariants from all time series are then used to generate a global feature vector of an action. These feature vectors are then used as an input in a K-nearest neighbor classifier.

Xiang in [13] proposed a method for motion retrieval in motion capture data based on ensemble HMM learning. First, 3D spatio-temporal features are extracted from training data and used for ensemble HMM learning. Then each movement class is learned with one HMM. Deng et al. [14] propose a method for human motion retrieval that employs a motion pattern discovery and matching scheme that breaks human motions into a part-based, hierarchical motion representation. Building upon this representation, a fast string matching algorithm is used for efficient runtime motion query processing. Liu et al. [15] construct a motion index tree based on hierarchical motion description. The motion index tree serves as a classifier to determine the sub-database that contains the most promising motions that are similar to the query sample in a motion retrieval context. The Nearest Neighbour rule-based dynamic clustering algorithm is adopted to partition the database and construct the motion index tree. A hierarchical indexing structure is also used by Pradhan et al. [16]. The proposed structure is based on the hierarchical structure of the human body, consisting of independent index trees each corresponding to a different sub-part of the body. Wu et al. [17] present an efficient motion data indexing and retrieval method based on self-organizing maps and the Smith - Waterman string similarity metric. An efficient motion retrieval system based on the query-by-example paradigm, which employs qualitative, geometric similarity measures is proposed by Demuth et al. [18]. Müller et al. in [19] propose a method for motion capture data annotation by deriving a motion template that captures the consistent and variable aspects of a motion class in an explicit matrix representation. Chiu et al. [20] introduce an affine invariant posture feature and propose an index map structure based on the posture distribution of raw data for content-based retrieval in human motion data. Forbes et al. [21] present a search algorithm for unsegmented motion data that is based on a weighted PCA-based pose representation that allows for flexible and efficient pose-to-pose distance calculations. Liu et al. [22] propose a method for analysing and indexing human-motion databases. They partition every body pose in the motion database into a hierarchy of low-dimensional local linear models. Data sequences are represented by their transitions through these local linear models. These transitions are called cluster transition signatures and are used for inter-sequence comparisons and sequence indexing.

III. METHOD DESCRIPTION

Let each movement be represented as a sequence of posture vectors \mathbf{x}_i , $i = 1, \dots, N$ where N is the number of frames of the sequence. Each posture vector carries information for the rotation angles in all skeleton joints

$$\mathbf{x}_i = \{\theta_{i1}, \theta_{i2}, \dots, \theta_{in}\} \quad (1)$$

where n is the number of rotation angles that form the posture vector. It should be noted that the proposed method does

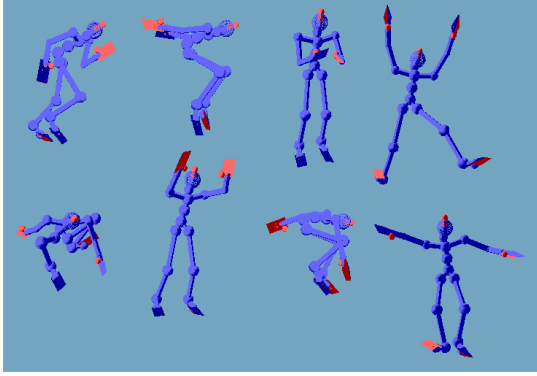


Fig. 2. Examples of dynemes extracted by the clustering algorithm.

not take into account information for the global rotation and translation of the body.

The basic building blocks of the method are presented in the following subsections.

A. Dyneme extraction

The first step of the algorithm concerns the quantization of the posture vectors space and the extraction, through clustering, of a codebook consisting of characteristic postures called dynemes.

Indeed, in order to recognize K different movement classes, we cluster the input space of posture vectors into C clusters. The clusters are identified by unsupervised clustering, using a K-means algorithm modified to work on angular data (see subsection III-B). Angular K-means is applied on all postures $\mathbf{x}_{ij}, i = 1, \dots, N_j, j = 1, \dots, L$ of all movement sequences in the training set where N_j is the number of frames of the j -th movement sequence and L the number of the training sequences. The number C of clusters is selected empirically and depends on the number of movements K that are to be recognized, the different ways a movement can be performed by different people, the different body types, etc. For each cluster created by the angular K-means algorithm, the centroid $\mathbf{v}_c, c = 1, \dots, C$ is computed as the circular mean of all postures in this cluster. This centroid represents one dyneme. Due to the averaging procedure, dynemes don't correspond to postures from the training set but rather on "average", characteristic postures. Examples dynemes are shown in Figure 2.

B. Angular K-means

Motion capture data, i.e. posture vectors \mathbf{x}_i , describe rotation angles at the joints. As already mentioned, due to the periodic nature of angular data neither the Euclidean distance nor the mean value estimator for data on the line can be used in such data. For example, an angle of 0 radians is the same as an angle of 2π radians but their Euclidean distance wouldn't be zero but 2π . Furthermore, the average of two angles 5° and 355° is 0° and not 180° as the classical average operator would entail. Two different measures, namely the distance between two angles and the circular mean [23] can be used instead. The distance 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:

$$\text{arc}(\theta_i, \theta_j) = \pi - |\pi - |\theta_i - \theta_j|| \quad (2)$$

The *circular mean* or *sample mean direction* \bar{x}_0 of N angular observations $\theta_1, \dots, \theta_N$ represented by sample points $\mathbf{M}_1, \dots, \mathbf{M}_N$ on a unit circle centred at point \mathbf{O} is the direction of the mean resultant vector \mathbf{R} of the unit vectors $\mathbf{OM}_1, \dots, \mathbf{OM}_N$. Its value is given by:

$$\bar{x}_0 = \arctan\left(\frac{\bar{S}}{\bar{C}}\right), \quad \bar{C} = \frac{1}{N} \sum_{i=1}^N \cos \theta_i, \quad \bar{S} = \frac{1}{N} \sum_{i=1}^N \sin \theta_i \quad (3)$$

Since the proposed algorithm applies the K-means algorithm on angle data, a modified angular version was constructed by replacing the classical mean and Euclidean distance by the above quantities.

C. Projection to dyneme space, evaluation of the bag of words

The next step of the method is to map all the posture vectors \mathbf{x}_{ij} in the training set to the dyneme space. The angular K-means algorithm will assign each posture vector to a class $c = 1, \dots, C$. Based on the clustering results we map each posture to its assigned dyneme (class centroid). Thus each movement sequence is represented in terms of dynemes, each frame/posture being represented by the dyneme it has been assigned to.

Next, we calculate for each movement the frequency of appearance of every dyneme, thus forming for each movement a histogram that characterizes it. Each histogram is a C -dimensional vector $\mathbf{s}_j = [s_i], j = 1, \dots, L, i = 1, \dots, C$:

$$[s_i] = \frac{n_i}{N_j} \quad (4)$$

where n_i is the number of occurrences of the i -th dyneme within the sequence and N_j the number of frames of the sequence. Some histograms are shown in Figures 3 and 4

It is obvious that such a "bag of words" (or, in this case, bag of postures) movement representation retains essentially no information regarding the temporal order of the various postures within the movement, its duration, speed, and start/end points, thus making it advantageous for recognition purposes, despite its apparent simplicity. The experiments presented in Section IV verify that this choice leads to very good movement recognition rates.

D. Classification

To classify an unknown mocap posture sequence to one of the movements that the algorithm has been trained to recognize the following procedure is used.

First we use the dynemes evaluated during the training stage to map each posture vector (frame) \mathbf{x} of the testing sequence into the nearest dyneme:

$$k = \arg \min_{c \in [1 \dots C]} \left(\sum_{i=1}^n \text{arc}(v_{ci} - x_i) \right) \quad (5)$$

Once all posture vectors have been mapped to dynemes we calculate the histogram \mathbf{q}_{test} for the testing sequence

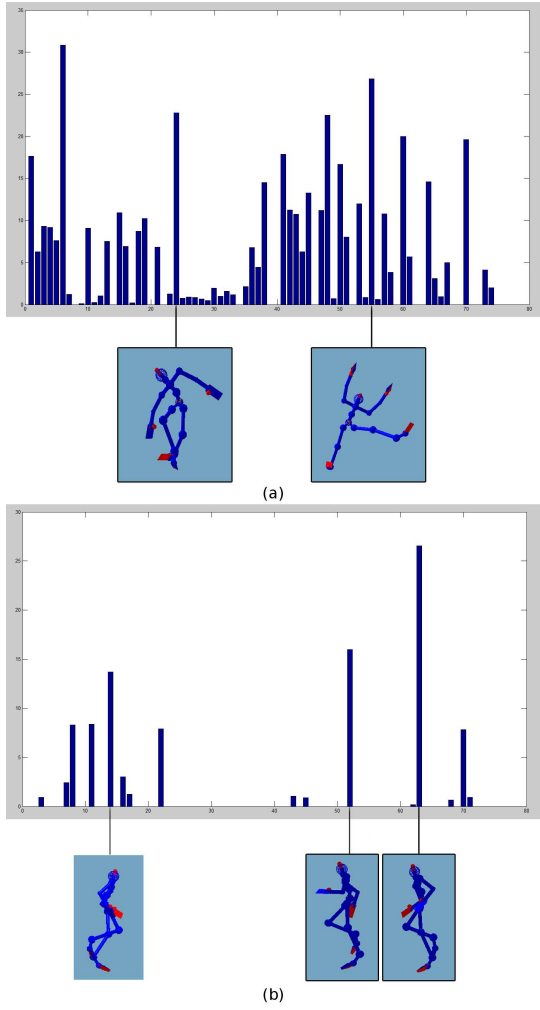


Fig. 3. Histograms of cartwheel (a) and run on place (b) movements generated by the proposed method, along with some dynemes that correspond to the most prominent peaks of the histograms

using (4). \mathbf{q}_{test} characterizes the sequence and is used for the classification/recognition. Two classifiers were used, a Support Vector Machine (SVM) classifier and a classifier based on sparse representation.

1) *SVM classification*: In this case, an SVM classifier is trained using the histograms of the labeled sequences of the training set. The trained SVM is then used to classify the histogram \mathbf{q}_{test} of an unknown sequence.

SVM is a widely used classification technique that stems from statistical learning theory. SVM minimize both a bound on the empirical error and the complexity of the classifier. SVMs can classify linearly or non-linearly separable data. For the non-linearly separable data SVM project the data into a higher dimensional Hilbert space using kernel functions and attempt to linearly separate them in this space [24].

Frequently used kernel functions include the polynomial kernel, $K(\mathbf{x}_i, \mathbf{x}_j) = (\beta \mathbf{x}_i^T \mathbf{x}_j + \eta)^q$, and the Radial Basis Function (RBF) kernel, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\gamma|\mathbf{x}_i - \mathbf{x}_j|\}$. Experiments conducted in this paper showed that the RBF kernel had the best results for the K-class SVM that was implemented.

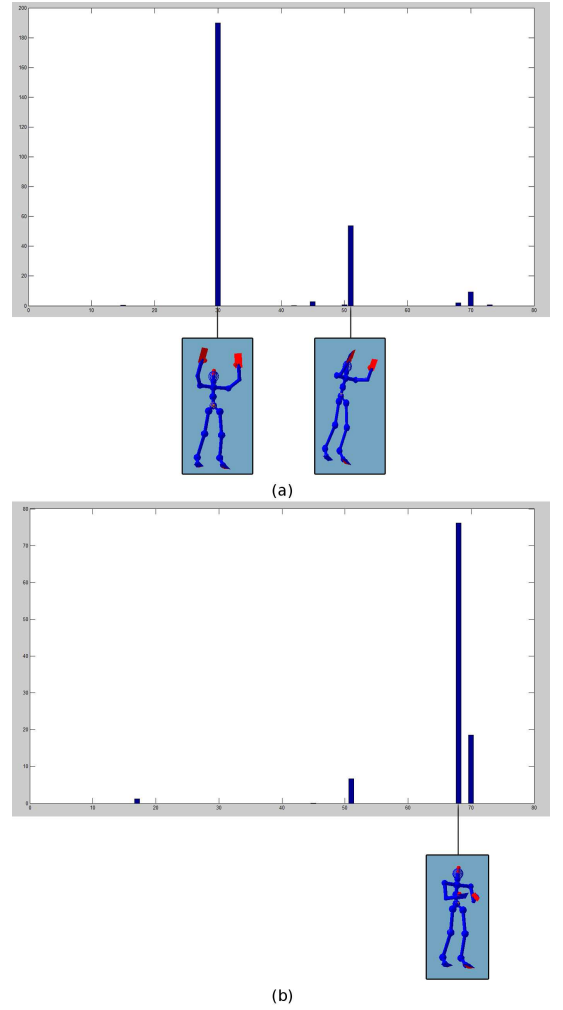


Fig. 4. Histograms of clap above head (a) and clap (b) movements generated by the proposed method, along with some dynemes that correspond to the most prominent peaks of the histograms

Instead of using directly the sequences histograms as input to the SVM classifier, PCA was used to reduce the dimensionality of the movement histograms before the SVM classification. In our case we keep the eigenvectors that correspond to 90% of the energy and we form the PCA matrix (\mathbf{W}_{pca}). Then we project the histograms in the PCA space as follows:

$$\mathbf{y}_i = \mathbf{W}_{pca}^T \mathbf{s}_i \quad (6)$$

2) *Sparse representation classification*: In this case, the classification is performed using the SRC algorithm proposed in [4]. Initially, the training histograms form the matrix $A = [A_1, A_2, \dots, A_K] \in \mathbb{R}^{m \times n}$ for K movement classes where $A_i = [\mathbf{s}_{i,1}, \mathbf{s}_{i,2}, \dots, \mathbf{s}_{i,n_i}]$ is the matrix formed by the histograms of the i -th class. Then the columns of A are normalized to have unit ℓ_2 -norm and the following ℓ_1 -minimization problem is solved:

$$\hat{\mathbf{x}}_1 = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{q}_{test}\|_2 < \varepsilon \quad (7)$$

The next step is to compute the residuals:

$$r_i(\mathbf{q}_{test}) = \|\mathbf{q}_{test} - A\delta_i(\hat{\mathbf{x}}_1)\|_2, i = 1, \dots, K \quad (8)$$

where $\delta_i(\hat{\mathbf{x}}_1)$ keeps only the elements of $\hat{\mathbf{x}}_1$ associated with the i -th class, whereas all other elements are zeroed.

The histogram \mathbf{q}_{test} is then classified in the class with the smallest residual and the movement is recognized as the movement of the associated class.

IV. EXPERIMENTAL RESULTS

The proposed method has been tested on the HDM05 database [3]. The database contains five persons performing several movements. Fourteen of these movements namely, run on place (runop), walk (walk), cartwheel (cartwheel), hop with both legs (hpboth), hop with left leg (hopl), hop with right leg (hopr), clap (clap), clap with hands above head (clapa), sit in chair (sitc), sit in floor (sif), left elbow to right knee (eltoknee), right elbow to left knee (eltokneer), walk sideways (walks) and walk sideways crossing the legs (walksc) were used in this paper. The selected dataset contains in total 641 sequences, in ASF/AMC format. The database contains rotation angles at the corresponding joints for 29 body parts. Since some of these body parts, for example the thumb or the toes, bear no significant information for 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.

The leave-one-person-out cross-validation (LOPOCV) procedure was used to assess the performance of the algorithm. At each cycle of the procedure all movement sequences of one person are retained to form the test set whereas all sequences of the remaining four persons were used to form the training set. The number of correctly classified movement sequences in all cycles (5 cycles in total) were used to compute the classification rate.

Experiments have been performed to identify the number of clusters C (dynemes) and the variant that provides the best results. In more detail, the two variants have been tested for clusters/dynemes C ranging between 80 and 100: SVM applied on histograms processed by PCA (SVM+PCA) and sparse representation-based classification (SRC). For SVM an RBF kernel and $\gamma = 0.0001$ was used. These SVM parameters were selected through additional experimentation. The results are shown in Figure 5. SVM+PCA classification provided the best overall result, for $C = 91$ clusters/dynemes. The correct classification rate achieved in this case was 88.92%. Sparse representation classification provided 84.38% for $C = 93$ clusters/dynemes. In more detail, 8 out of 14 movements achieved classification rate higher than 90% when SVM+PCA was used for classification.

The best overall correct classification rate for the method for both variants and a comparison with [8] and [5] is presented in Table I. The implementation for the method [8] is provided by the authors whereas the method in [5] was implemented using rotation angles as input, as in [14]. Both variants of the proposed method achieved better results than the methods in comparison as can be seen in Table I. The variant that uses SVM+PCA for classification achieved better results than the SRC-based variant. It should be noted that the performance of

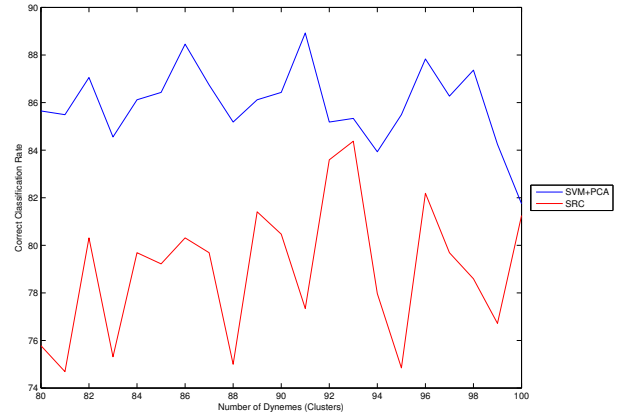


Fig. 5. Correct classification rate for SVM+PCA and SRC for various values of C (number of dynemes)

TABLE I. OVERALL BEST CORRECT CLASSIFICATION RATE AND COMPARISON WITH [8] AND [5]

| | 14 classes |
|-------------------|--------------|
| SVM+PCA | 88.92 |
| SRC | 84.38 |
| ISOCCA [8] | 49.14 |
| SVD [5] | 83.08 |

the isoCCA method is significantly inferior from that reported in [8] which is probably due to the fact that the experimental evaluation of the paper in that paper was conducted in a leave-one-out setting and not in a leave one out person setup, as in our case. Obviously the latter is a much more fair experimental setup than the former.

V. CONCLUSIONS

In this paper, a novel method for movement recognition in motion capture data was proposed. The method utilizes characteristic postures (dynemes) derived through a novel variant of the K-means algorithm, along with a bag of postures approach and SVM or SRC classifier. Experiment results verify that the proposed approach provides good movement recognition results surpassing other methods. In the future, we plan to extend this method so as to operate in a temporal window moving over a sequence so that we can achieve continuous movement recognition, namely recognition on sequences containing multiple successive movements. Extensions towards motion indexing and retrieval will also be considered.

ACKNOWLEDGMENT

The authors wish to thank Liqun Deng for his help in the implementation of the algorithm described in [5].

REFERENCES

- [1] G. Burdea and P. Coiffet, *Virtual Reality Technology*, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 2003.
- [2] R. Parent, *Computer Animation, Algorithms and Techniques*. Morgan Kaufmann, 2002.

- [3] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation mocap database hdm05," Universität Bonn, Tech. Rep. CG-2007-2, Jun. 2007.
- [4] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [5] C. Li, P. R. Kulkarni, and B. Prabhakaran, "Segmentation and recognition of motion capture data stream by classification," *Multimedia Tools and Applications*, vol. 35, no. 1, pp. 55–70, Oct. 2007.
- [6] J. Wang and H. Lee, "Recognition of human actions using motion capture data and support vector machine," in *Proc. of the 2009 WRI World Congress on Software Engineering (WCSE '09)*, vol. 1. Washington, DC, USA: IEEE Computer Society, 2009, pp. 234–238.
- [7] H. Kadu, M. Kuo, and C.-C. J. Kuo, "Human motion classification and management based on mocap data analysis," in *Proc. of the 2011 Joint ACM Workshop on Human Gesture and Behaviour Understanding (J-HGBU '11)*. New York, NY, USA: ACM, 2011, pp. 73–74.
- [8] S. Shariat and V. Pavlovic, "Isotonic cca for sequence alignment and activity recognition," in *International Conference on Computer Vision (ICCV '11)*, 2011.
- [9] M. Raptis, D. Kirovski, and H. Hoppe, "Real-time classification of dance gestures from skeleton animation," in *Proc. of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '11)*. New York, NY, USA: ACM, 2011, pp. 147–156.
- [10] F. Lv, R. Nevatia, and M. W. Lee, "3D human action recognition using spatio-temporal motion templates," in *Proc. of the 2005 International Conference on Computer Vision (ICCV '05)*. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 120–130.
- [11] L. Deng, H. Leung, N. Gu, and Y. Yang, "Generalized model-based human motion recognition with body partition index maps," *Computer Graphics Forum*, vol. 31, no. 1, pp. 202–215, 2012.
- [12] S. Ali, A. Basharat, and M. Shah, "Chaotic invariants for human action recognition," in *IEEE 11th International Conference on Computer Vision, 2007 (ICCV '07)*, Oct. 2007, pp. 1–8.
- [13] J. Xiang, "Motion retrieval with temporal-spatial features based on ensemble learning," in *Proc. of the 3rd International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence (ICIC '07)*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 300–308.
- [14] Z. Deng, Q. Gu, and Q. Li, "Perceptually consistent example-based human motion retrieval," in *Proc. of the 2009 Symposium on Interactive 3D Graphics and Games (I3D '09)*. New York, NY, USA: ACM, 2009, pp. 191–198.
- [15] F. Liu, Y. Zhuang, F. Wu, and Y. Pan, "3D motion retrieval with motion index tree," *Computer Vision and Image Understanding*, vol. 92, no. 2-3, pp. 265–284, Nov. 2003.
- [16] G. N. Pradhan, C. Li, and B. Prabhakaran, "Hierarchical indexing structure for 3D human motions," in *Proc. of Multimedia Modeling (MMM '07)*, vol. 1, 2007, pp. 386–396.
- [17] S. Wu, S. Xia, Z. Wang, and C. Li, "Efficient motion data indexing and retrieval with local similarity measure of motion strings," *The Visual Computer*, vol. 25, no. 5-7, pp. 499–508, Apr. 2009.
- [18] B. Demuth, T. Röder, M. Mäijler, and B. Eberhardt, "An information retrieval system for motion capture data," in *Proc. 28th European Conference on Information Retrieval (ECIR '06)*, vol. 3936. Springer, 2006, pp. 373–384.
- [19] M. Müller, A. Baak, and H. Seidel, "Efficient and robust annotation of motion capture data," in *Proc. of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '09)*. New York, NY, USA: ACM, 2009, pp. 17–26.
- [20] C. Chiu, S. Chao, M. Wu, S. Yang, and H. Lin, "Content-based retrieval for human motion data," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 446–466, Sep. 2004.
- [21] K. Forbes and F. E., "An efficient search algorithm for motion data using weighted PCA," in *Proc. of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '05)*. New York, NY, USA: ACM, 2005, pp. 67–76.
- [22] G. Liu, J. Zhang, W. Wang, and L. McMillan, "A system for analyzing and indexing human-motion databases," in *Proc. of the 2005 ACM SIGMOD International Conference on Management of Data (SIGMOD '05)*. New York, NY, USA: ACM, 2005, pp. 924–926.
- [23] N. Nikolaidis and I. Pitas, "Nonlinear processing and analysis of angular signals," *IEEE trans. on Signal Processing*, vol. 46, no. 12, pp. 3181–3194, Dec. 1998.
- [24] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.