Improving Movement Analysis in Physical Therapy Systems Based on Kinect Interaction

Alina D. Călin
Department of Computer Science
Babeș-Bolyai University
1 Mihail Kogălniceanu Street
RO-400084 Cluj-Napoca, Romania
alinacalin@cs.ubbcluj.ro

Horia F. Pop
Department of Computer Science
Babeș-Bolyai University
1 Mihail Kogălniceanu Street
RO-400084 Cluj-Napoca, Romania
hfpop@cs.ubbcluj.ro

Rareș F. Boian
Department of Computer Science
Babeș-Bolyai University
1 Mihail Kogălniceanu Street
RO-400084 Cluj-Napoca, Romania
rares@cs.ubbcluj.ro

This study uses machine learning methods to analyse Kinect body gestures involved in the user interaction with exergaming systems designed for physical rehabilitation. We propose a method to improve gesture recognition accuracy and motion analysis, by extracting from the full body motion data recorded by the Kinect sensor three important features which are relevant to physical therapy exercises: body posture, movement trajectory and range of motion. By applying the Hidden Markov Model (HMM) and Dynamic Time Warping (DTW) algorithms, we obtained an improved accuracy by selecting specific features from the public UTD-MHAD full body gestures database (with up to 56% for HMM and 32% for DTW). Preliminary results show a positive correlation between the movement amplitude and the envelope feature ($r = 0.92$). Thus, this approach has the potential to improve gesture recognition accuracy and provide user feedback on how to improve the movement performed, in particular the movement amplitude. We propose further improvements and method validations to be the basis of creating an intelligent virtual rehabilitation assistant.

Movement Analysis; Kinect; Time Series Classification; Hidden Markov Models; Dynamic Time Warping;

1. INTRODUCTION

Exergames are video games specifically designed to incorporate exercises used in physical therapy and rehabilitation. One such example of a system is MIRA (see Mira (2017)), a software medical platform that gamifies physiotherapy, using the Microsoft Kinect sensor (see Kinect (2017)) for remote interaction and motion tracking. The system provides important statistics about the user's performance while exergaming (such as game points, time, involvement, number of repetitions), as well as direct feedback about the correctness of the exercises during game play.

In this study we aim to improve exergaming systems' interaction by performing an advanced gesture analysis which provides detailed feedback to the user regarding the correctness of the exercise performed, in particular on movement amplitude. The idea is to detect the correctness of a movement, providing also intelligently-derived information and recommendations on how to improve on the range of motion component of the exercise. The long-term purpose is to create a virtual rehabilitation assistant that would aid users while exercising.

In the following, we present the methods used to extract significant features from the movements performed, by separating from each gesture three main components: body pose (posture), movement amplitude (range) and movement pattern (trajectory). Next, we present testing results and their importance.

2. RELATED WORK

Literature reveals a number of studies using Kinect based gesture recognition. However, most of them are focused on multi-joint static postures or on one-joint time series. Capturing key poses which define a specific exercise/gesture, Wang (2015) obtains an overall classification accuracy of up to 94% and Călin (2016a) up to 99%. For one-joint time series gestures, Lin (2012) presents a feature-guided HMM algorithm to segment data on joint angles and angular velocity of rehabilitation movements, obtaining an accuracy of up to 91%. Călin (2016b) classifies right hand time series data obtaining an accuracy of up to 97% with DTW and HMM.

One study referring to multi-joint time series data is Grimm (2016), approaching a movement analysis based on a 7 degree gravity-compensating arm exoskeleton. It determines compensatory

movement strategies of patients and monitor their kinematic evolution throughout rehabilitation.

Our study is, to the best of our knowledge, among the first to approach movement analysis on multi-joint times series gestures using Kinect.

3. PROPOSED METHOD

In the following, we propose a method to process and analyse time series gestures in order to obtain a better accuracy of classification. This method was then tested on a publicly available dataset of 27 gestures collected with Kinect 1 (20 joints in 3D, which means 60 dimensions/features) named UTD-MHAD (see Chen (2015)).

For the purpose of extracting information regarding the correct amplitude of the movement, we created two open-source databases of time series gestures, Kinect DB (2016). It contains equally distributed classes, representing different variations (in terms of range of motion) of a specific movement:

- *Circles Database (CDB)* contains 6 classes of user created circle shapes: big, medium and small circles, vertical and horizontal ellipses, and a class of misshaped circles.
- *Flexion Database (FDB)* contains 2 classes with shoulder forward flexion: one performing at an angle of 90° and the other one at 180°.

We used Kinect (Windows SDK 2.0) to collect 3D gestures of the 25 body joints for the two databases CDB and FDB. We obtained 75 dimensions/features for each sample, with 15 samples per class.

The HMM and DTW algorithms, implemented by the Gesture Recognition Toolkit (GRT, Nick (2014)) to have multi-dimensional support, were utilized for multi-class classification, as they have presented good results in previous work (Călin (2016b)).

Several features were extracted (from the initial 75 in CDB and FDB, and from the initial 60 in the UTD-MHAD) as per our proposed method presented below. Validation was done using 7-fold cross validation, as the GRT library did not support cross validation with a higher number of folds, such as 10, on our large sample data. The statistical average of 30 tests was computed across the classes representing gestures.

We propose a method, as described in Figure 1, that is able to separate movement components of the full body gestures, by extracting two new derived features using two GRT feature extraction algorithms. The Movement Index Feature (MI) algorithm computes the amount of movement or variation for a N-dimensional signal over a time frame (as described in Nick (2014)). We use MI to separate the features representing active joints and passive joints respectively, using a threshold of 30% of the maximum MI value. The former define the Motion of the gesture (in our case, on the CDB, five 3D joints of the arm, resulting in 15 dimensions). The latter define the Pose of the gesture (the remaining 20 joints, resulting in 60 dimensions).

We also selected the joint (in 3D) with the maximum value of the MI as the most relevant one in describing the movement pattern (the Trajectory). As the UTD database contains various types of gestures, we have split them into 3 subsets, according to which joints compose the trajectory of the movement: RH (11 classes in which the right hand or the wrist is the joint with the MI value greater than 70% of the maximum MI value computed on all joints), BH (8 classes in which both right and left hands or wrists have MI values over 70% of the maximum MI) and FB (8 classes in which the significant joint is the spine or that have more than 30 active joints with MI over 20% of the maximum MI). From RH we have derived CH (a custom subset of RH with 6 coordination movements). The subsets are:

- *Coordination Hand Gestures (CH)*: wave, throw, draw X, draw circle clockwise, draw circle counter clockwise, draw triangle.
- *Right Hand Gestures (RH)*: CH gestures, swipe left, swipe right, tennis swing, knock, catch.
- *Both Hands Gestures (BH)*: clap, arm cross, basketball shot, boxing, baseball swing, arm curl, tennis serve, push.

By applying the Envelope Feature (EF) in CDB and FDB, we obtained data that was correlated with the amplitude of the movement. The EF used here is computed as the smooth curve outlining the upper extremes of the motion data signals. Next, we correlated the EF derived mean values with the mean amplitude of the corresponding movements for the two databases.

4. RESULTS AND DISCUSSION

4.1. Improving Gesture Classification Accuracy

Figure 2 displays the results obtained by classifying the 6 gestures of the CDB and the 2 gestures of the FDB respectively, based on all 25 skeleton joints provided by the Kinect and on the selection of features (15 of the right arm or 3 of the hand tip). In the two gestures, the right arm is the active body part that defines the movement and its amplitude, the hand provides the specific trajectory pattern

75 dimensions

60 dimensions   **15 dimensions**   **3 dimensions**

Posture
All joints except Right Arm joints

Movement (Gesture)
Right Arm (hand, hand tip, thumb, wrist, elbow)

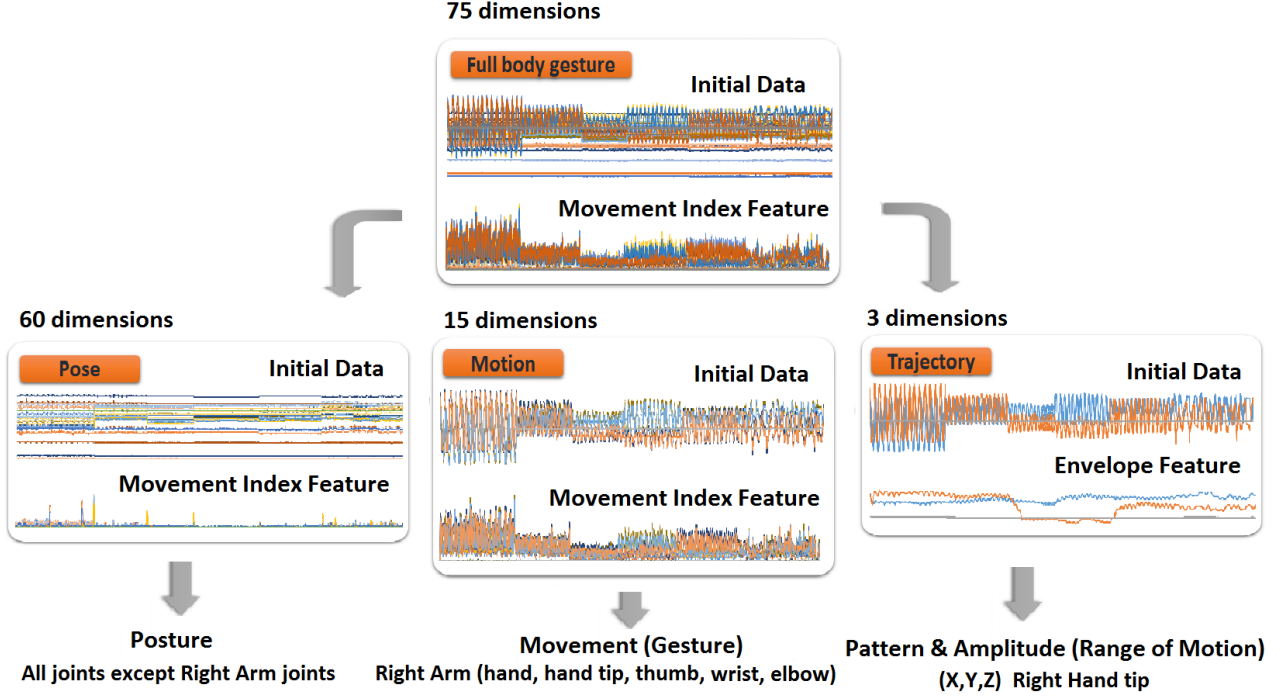Pattern & Amplitude (Range of Motion)
(X,Y,Z) Right Hand tip

Figure 1: The extraction of movement components exemplified on the CDB (with 6 classes), using the Movement Index Feature (MIF) and the Envelope Feature (EF).
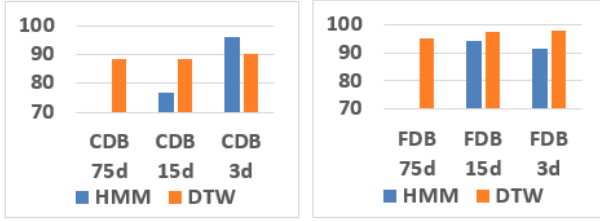

Figure 2: Multi-class classification accuracy for CDB and FDB, when using (1) all features, (2) the active joints (15D of the right arm joints) or (3) the trajectory joint having the maximum MI (3D of the right hand).
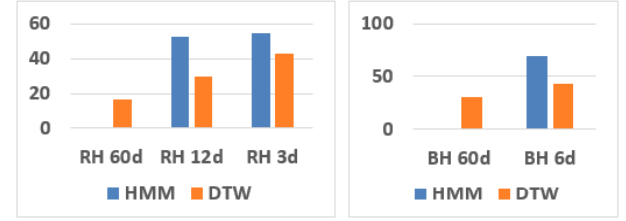

Figure 3: Classification accuracy on RH and BH. The active joints (12D of the right arm joints); maximum MI (3D of the right hand) for RH; both hands (6D for left hand and right hand) for BH.

joint which contains also information about the amplitude of the motion, and the rest of the body, excluding the right arm, is part of the base pose.

Figures 3 and 4 present the results obtained on the UTD-MHAD and its subsets (RH, BH and CH) for which the improvement is much more consistent for both HMM and DTW. The most significant increase for DTW is for CH, increasing accuracy from 22% to 54%. As for HMM, the probabilistic model does not support large time series data with a lot of features, being unusable, however by selecting the important features we obtain the highest improvement for BH reaching 68.7% accuracy. Results show an increase in accuracy of classification when selecting only the active body parts, with best results on the joint with the maximum value of the MI for RH. HMM performs better the more specific the extracted features are, as it is based on a state probability model, for which reason it performs very poorly on large
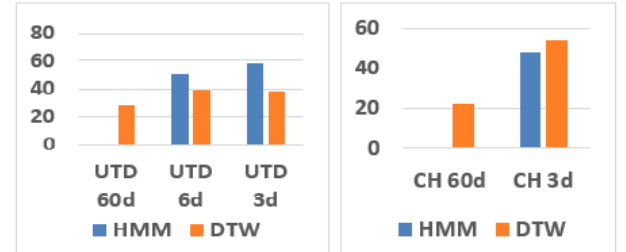

Figure 4: Classification accuracy results on CH and entire UTD-MHAD.

sample data with a lot of features, as it is in our case.

On the other hand, DTW, which works by finding the best mapping of each dimension on the time axis, may lose accuracy when certain features are removed (as in the case with the UTD, we have 38.85% for 6D and 38% for 3D). This is likely due to the fact that this initial analysis does not consider

the pose component for classification, but will be included in our future work.

4.2. Analysing Movement Amplitude

The EF values, obtained from the 3D joint (X,Y,Z) with the highest MI, show good correlation with the movement amplitude on each of the X, Y and Z features (Figures 5 and 6). For FDB, correlating EF with the movement range of motion (angle of 90° or 180°) resulted in the Pearson correlation coefficient r =0.8166937 for X, r =0.8296638 for Y and for Z a value r =0.864509. For CDB we correlated the EF of Y-X with the circle radius (initial correlation with X was negative and small r = -0.1092782, due to the existing negative values of the motion data, thus we combined Y and X) obtaining the Pearson correlation coefficient value r =0.8896081. We also correlated EF of Z with the circle radius, resulting in value r =0.9229759. The big circle had a radius of approximately 50cm, the medium circle 30cm and the small circle 15cm.
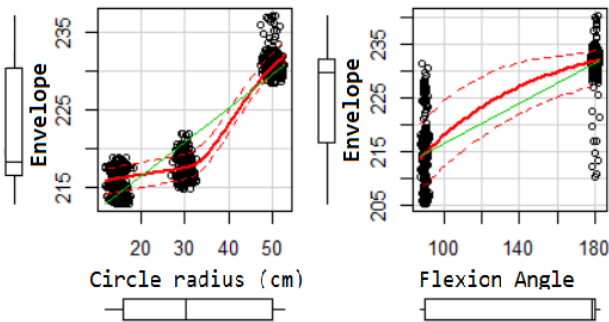

Figure 5: *EF of the Z feature showing correlation with the range of motion, in CDB (left) with the circle radius, having r=0.9229759; and in FDB (right) with the forward flexion angle, having r=0.8896081.*
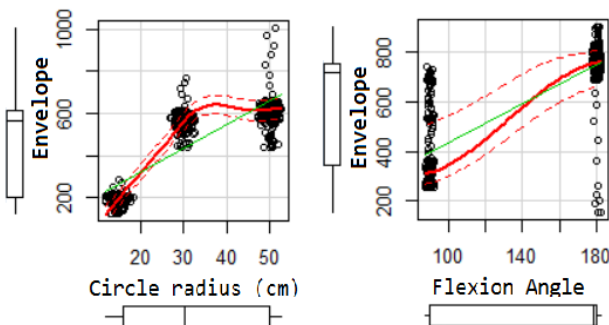

Figure 6: *EF of the Y feature with r=0.8660295 for CDB (left) and r=0.8296638 for FDB (right).*

These results show a good potential to determine the amplitude of the movement and inform the user on how their motion stands relative to the amplitude of the correct exercise to be performed. The system can determine using EF if the motion amplitude is too small/large and inform the user accordingly on how to improve their performance. This method provides a generalised approach that is able to improve the classification accuracy and

provide the user with feedback on how to adjust their movement amplitude on a large range of exercises, not just arm or hand movements, but also lower limb, trunk or neck exercises. Yet further improvement and testing are necessary, by using a clustering algorithm to aggregate passive joints in order to combine the posture feature with motion trajectory and amplitude.

This way the system would be able to generate other user feedback, besides that regarding movement amplitude, referring to trajectory (user should concentrate on the correct trajectory of the joint as required in the exercise) or posture (user posture should be corrected as instructed, for example standing, sitting, keeping the left arm in abduction at 90°), which we aim to implement further.

5. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a generalised movement analysis method on time series gestures for physical rehabilitation. We tested this method on our own databases collected with Kinect 2 and on the UTD-MHAD publicly available database of gestures with Kinect 1. By extracting selective features and computing derived features (MI and EF) we separated three main components of the movement: pose, trajectory and range of motion. This way, we obtained better classification accuracy, with up to 56% for HMM and 32% for DTW. We also found a positive correlation between movement amplitude and the EF extracted feature (r =0.92), which can generate user feedback to help users improve on their exercise physical performance.

However, the model requires further optimisation and validation on larger datasets. We intend to improve on the combination of extracted features, by using the Principal Components Analysis and Cluster Analysis methods for aggregating the pose component of the movement.

Our aim is to extend and generalise this model for different types of physical therapy exercises in order to construct an intelligent virtual rehabilitation assistant for Kinect systems. This agent would provide users feedback on the exercises they perform to help them improve on different movement components (e.g. amplitude, trajectory, posture).

3. REFERENCES

Călin, A. D. (2016) Variation of Pose and Gesture Recognition Accuracy Using Two Kinect Versions. *International Symposium on INnovations in Intelligent SysTems and Applications*. 1–7.

Călin, A. D. (2016) Gesture Recognition on Kinect Time Series Data Using Dynamic Time Warping and Hidden Markov Models. *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*. 264–271.

Călin, A. D. (2016) Time Series Kinect Gestures Database. Available from: https://goo.gl/ZzmE0a (Dec 2016).

Chen, C., Jafari, R. and Kehtarnavaz, N. (2015) UTD-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and a Wearable Inertial Sensor. *Proceedings of IEEE International Conference on Image Processing*, Canada. 168–172.

Grimm, F., Naros, G. and Gharabaghi, A. (2016) Compensation or Restoration: Closed-loop Feedback of Movement Quality for Assisted Reach-to-grasp Exercises with a Multi-joint Arm Exoskeleton. *Frontiers in Neuroscience*, 10.

Lin, J. F. and Kulic, D. (2012) Segmenting Human Motion for Automated Rehabilitation Exercise Analysis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2881–2884.

Microsoft Corporation (2017) Microsoft Kinect for Windows, Version 2. Available from: https://developer.microsoft.com/en-us/windows/kinect (Mar 2017).

MIRA Rehab Limited (2017) Available from: http://www.mirarehab.com (May 2017).

Nicholas, G., Joseph, A. P. (2014) The Gesture Recognition Toolkit. Available from: https://github.com/nickgillian/grt (Nov 2016).

Wang, Q. et al. (2015) Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect. *International Conference on Healthcare Informatics*. 380–389.