**Automatic segmentation of therapeutic exercises motion data based on a predictive event approach**

Sofija Spasojevic\(^1\) and Rodrigo Ventura\(^2\)

\(^1\)Faculty of Electrical Engineering and Mihailo Pupin Institute, University of Belgrade
\(^2\)Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon

I. **INTRODUCTION**

This paper presents a predictive event approach which is applied for automatic segmentation of therapeutic exercise sequences. The main purpose of our study is to enhance the efficiency of movement analysis in patients with neurological disorders. When dealing with analysis of rehabilitation exercises it is often required to pre-process the collected data. In order to examine the movements and to carry out the relevant measurements and determine the parameters of interest, each movement must be segmented from a given sequence and analyzed separately. More generally, in case of gesture recognition tasks, after gesture acquisition, segmentation is the following and central step, which can have a huge impact on the classification rate of the gesture recognition system. Gesture segmentation and recognition systems have significant applications in many different fields such as virtual and augmented reality [1], industrial process control [2], physical rehabilitation [3], human-robot interaction [4], computer games [5] etc. Due to this, gesture segmentation is an active research topic and challenging scientific problem, fairly present in the latest research.

The predictive event approach is based on a principle of detecting an event when sensor data depart significantly from an adaptive model-based predictor. We have used portable, low-cost Kinect device with marker-free based technique. During exercise execution, the skeleton is continuously detected and 3D positions of characteristic human joints are collected for each frame (Fig. 2). From the original data set, which consists of all collected joints motion data, we have extracted the ones from interest for upper body movement therapeutic exercises (hand and elbow joint). Trajectories of selected joints are modelled as Gaussian processes. Based on this data set, a Gaussian process based predictor is adapted and used to detect significant changes in the exercise sequences. The results over the formed dataset are compared with commonly used technique and illustrate the superiority of our approach.

![Figure 1. Skeleton tracking (left) and collected joints (right)](image)

**Figure 1. Skeleton tracking (left) and collected joints (right)**

**Figure 2. Exercises acquisition using Kinect – RGB (a, b, c) and depth (d, e, f) stream**

II. **RELATED WORK**

Frequently used methods for gesture segmentation are based on the Dynamic Programming (DP) [6, 7], Dynamic Time Warping (DTW) [8, 9] and Hidden Markov Models (HMM) [10-13].

A technique based on the simple sliding window combined with simple moving average filter is used in [14]. The author defines the content of each gesture in the following form: starting static posture, dynamic gesture part and ending static posture. In addition, to obtain a more robust segmentation, the author observes also the length of each analyzed sequence to eliminate the appearance of the static part into dynamic part of the gesture.

In [15] the authors developed an algorithm for segmentation of dance sequences. This algorithm, called *Hierarchical Activity Segmentation*, is based on the division of the human body onto hierarchically dependent structures. They take into account relevant motion parameters for body segments (segmental force, kinetic energy and momentum) that characterize motion in the levels of defined hierarchy. In [16] the authors took a dynamical system approach for dynamic system identification, however, that approach did not account for sensor noise.

In this paper, we study an approach using Gaussian processes as a machine learning method [17] that provides information about both, value and uncertainty. This method has shown good properties related to complexity model, processing time and remarkable results in comparison with commonly used method.
III. EXPERIMENTAL GROUP AND ALGORITHM

Participants in this study are patients with Parkinson’s disease. Data collected using Kinect device on sampling frequency of 27 Hz consist of 3D position of characteristic skeleton joints, along with RGB and depth video sequences (Fig. 2). Acquired exercises are instructed by the therapist and give the insight of patient’s upper body functionality. In total, three different exercises (Fig. 2) for eight patients are collected. Each exercise is repeated at least seven times. In case of exercises performed by the hands (Fig. 2) only trajectories of the elbow and hand joint are relevant for the further processing. All coordinates are normalized with respect to the torso and filtered using a low pass filter of second order with cutoff frequency of 2 Hz due to measurement noise.

Trajectories of elbow and hand joint positions are modelled as Gaussian processes and three predictive Gaussian prediction model (each model for one coordinate) are formed. Number of hyperparameters which define the meaning and covariance functions of Gaussian process depends on the form of input and output training set samples. Let n be the number of frames in the exercise sequence and $x_i$ value of x-coordinate in i-th frame. Training input set (1) consists of samples that are organized as k dimensional vectors:

$$X = ([x_1; x_2; \ldots; x_{k-1}; x_k]; \ldots; [x_{n-(k-1)}; \ldots; x_n])$$  

(1)

Training output set (2) contains from following scalar samples of appropriate training input vector sample:

$$X_y = (x_{k+1}; x_{k+2}; \ldots; x_n)$$  

(2)

This procedure is repeated analogously in the case of the input and output training set for y coordinate, $Y$ and $Y_y$. Given this data set, corresponding mean functions of Gaussian models have per k, and covariance functions per two free parameters, which are determined in the process of hyperparameters optimization.

Predictive models are defined using input and output training and input testing set, obtained hyperparameters and selection of appropriate inference method. Models are formed for x and y trajectories of hand joints, since the exercises are performed in x-y plane. The values of $y$-coordinate in this case did not give any contribution to the final result, therefore they are not taken into account.

Errors of prediction in the form of the difference between real ($x$, $y$) and predicted values () are calculated at each step. Since the Gaussian process based predictor predicts both, mean and variance, in order to obtain a normalized distance metric, Mahalanobis distance (3) is also calculated at each step. Using this metric, the method is more sensitive to small errors if a data point has high certainty.

$$MD = \left[ \begin{array}{c} x - \hat{x} \\ y - \hat{y} \end{array} \right] \left[ \begin{array}{cc} \sigma_x & 0 \\ 0 & \sigma_y \end{array} \right] \left[ \begin{array}{c} x - \hat{x} \\ y - \hat{y} \end{array} \right]$$  

(3)

where $\sigma_x$ and $\sigma_y$ are predictive variances for first and second Gaussian predictive model, respectively.

IV. RESULTS

We have observed changes of the Mahalanobis distance through exercise sequence. When Mahalanobis distance increases significantly for several successive time steps and then drops again, boundary points of that segment are marked as events. Mahalanobis distance for one sequence of exercises is shown on the Fig. 3. Peaks that have the greatest values represent points in the sequence where the values of $x$ and $y$ hand coordinate suddenly increase or decrease. More precisely, positions where Mahalanobis distance has greater value than a determined threshold (Fig. 3) are marked as events. As the threshold varies, positions and numbers of events are changing and this is the only parameter that is necessary to adjust. The hand trajectory of true and predicted $x$ and $y$ coordinates of hand joint together with detected events for $k=5$ (1, 2) are shown on Fig. 4.

Fig. 3 shows that detected events correspond to the characteristic points in the sequence where the values of $x$ or $y$ coordinate of hand joint start or stop to change significantly. In order to classify all events between the ones with meaningful information, for example real beginnings and ends of individual exercises, classification technique based on the Hidden Markov Model (HMM) will be applied.
The approach described in this paper is compared with standard technique for detecting characteristic or extreme points in the sequence – technique of the first derivative. Comparison of these two methods (Fig. 5) is based on the combined sensitivity and specificity criteria (4-6), commonly used statistical tool for measuring classifier performance [18]. Value P (y-axis on Fig. 5) is calculated using relations (4-6) for different values of the threshold in the case of five exercise sequences.

\[
P = \sqrt{sens \cdot spec} \quad (4)
\]

\[
sens = \frac{TP}{TP + FN} \quad (5)
\]

\[
spec = \frac{TN}{TN + FP} \quad (6)
\]

In relations (5) and (6), TP denotes the number of true positives, FN the number of false negatives, TN the number of true negatives and FP the number of false positives. According to the form of (4-6) it can be seen that greater values of sens and spec indicate better performances of the approach, hence figure 5 clearly illustrate the superiority and advantage of our approach.

![Figure 5. A comparison of our method and first derivative approach based on sensitivity and specificity criteria](image)

V. CONCLUSION AND FUTURE WORK

In this work we have presented one approach for therapeutic exercise segmentation based on a predictive Gaussian model and event detection principle. This approach has shown excellent results in the sense of correct detection of significant changes during therapeutic exercise performing and advantage in comparison with commonly used technique of the first derivative.

Next extension of this work will be oriented to the integration of this approach with HMM in order of meaningful event classification. Future work will be focused on the improvement of this method and generalization in case of larger and more diverse gesture sequences and implementation in real time environments.

Acknowledgement

This work was partially funded by bilateral project COLBAR, between Instituto Superior Técnico, Lisbon, Portugal and Mihailo Pupin Institute, Belgrade, Serbia, FCT [PEst-OE/EEI/LA0009/2013], III44008, TR35003 and SNSF IP SCOPES, IZ74Z0_137361/1.

References


