Automatic segmentation of therapeutic exercises motion data with a predictive event approach

S. Spasojevic[1, 2,3], R. Ventura[3], J. Santos-Victor[3], V. Potkonjak[1] and A. Rodic[2]

[1] *Faculty of Electrical Engineering and* [2]*Mihailo Pupin Institute, University of Belgrade, Serbia, e-mail: {sofija.spasojevic, aleksandar.rodic}@pupin.rs, potkonjak@yahoo.com*

[3] *Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Portugal, e-mail: {rodrigo.ventura, jasv}@isr.tecnico.ulisboa.pt*

Abstract. We propose a novel approach for detecting events in data sequences, based on a predictive method using Gaussian processes. We have applied this approach for detecting relevant events in the therapeutic exercise sequences, wherein obtained results in addition to a suitable classifier, can be used directly for gesture segmentation. During exercise performing, motion data in the sense of 3D position of characteristic skeleton joints for each frame are acquired using a RGBD camera. Trajectories of joints relevant for the upper-body therapeutic exercises of Parkinson's patients are modelled as Gaussian processes. Our event detection procedure using an adaptive Gaussian process predictor has been shown to outperform a first derivative based approach.

Key words: Gesture segmentation, Predictive event approach, Gaussian processes, Physical rehabilitation, RGBD camera

1 Introduction

Recent research in medical and service robotics has shown growing application of various robotic systems into the medical practice. Such systems have been mainly developed as a support for the traditional physical rehabilitation therapy and treatment, since the conventional techniques rely on the clinical assessment tools [1]. Robotic prostheses have been designed for the assistance during movements performing at patients with the reduced mobility. Other systems have been developed for the purposes of monitoring and evaluating patient's performance during different rehabilitation tasks. The advantages of their usage are the superiority of quantitative system outputs comparing to the subjective evaluations of therapists [1] and enhancement of patient's motivation. Our method can be potentially integrated into one such system for movement examination and characterization, since the main goal of our approach is to enhance the efficiency of motion data analysis in patients with neurological disorders. In addition, the Kinect device that has been used in our study for the movement data acquisition can be attached to the rehabilitation robot, which would increase patient's motivation through the interaction with the robot, and on the other side, motion data can be collected and analyzed in real time.

This paper presents a predictive event approach for automatic segmentation of therapeutic exercise sequences. 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.

In general, gesture recognition tasks require gesture acquisition followed by the central step of segmentation, 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 [2], industrial process control [3], physical rehabilitation [4], human-robot interaction [5], computer games [6] 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.

2 Related work

Frequently used methods for gesture segmentation are based on the Dynamic Programming (DP) [7, 8], Dynamic Time Warping (DTW) [9, 10] and Hidden Markov Models (HMM) [11-14].

A technique based on the simple sliding window combined with simple moving average filter is used in [15]. 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

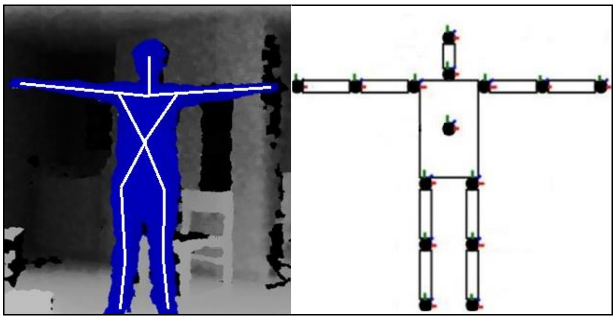

Fig. 1 Skeleton tracking (left) and collected joints (right).

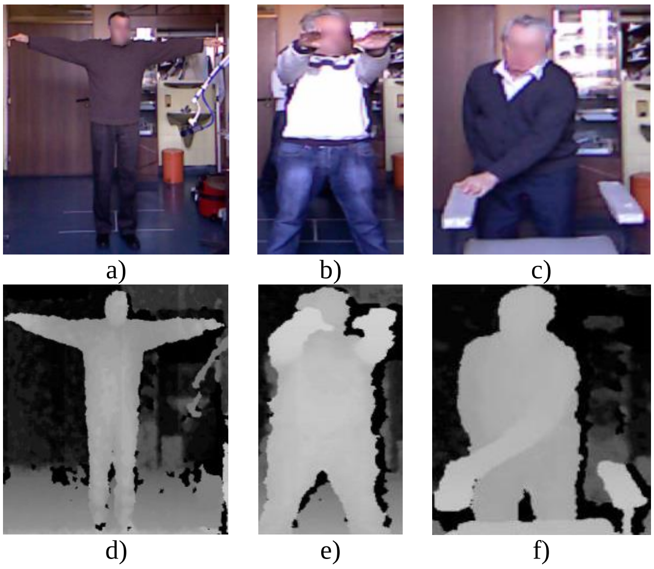

Fig. 2 Exercises acquisition using Kinect – RGB (a, b, c) and depth (d, e, f) stream.

the length of each analyzed sequence to eliminate the appearance of the static part into dynamic part of the gesture.

In [16] 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 [17] the authors took a dynamical system approach for dynamic system identification, however, that approach did not account for sensor noise.

Previously, a prediction-based approach to event segmentation was taken, based on an adaptive dynamical system approach [18, 19]. In this paper, we study a different approach employing a probabilistic model (Gaussian processes) as a machine learning method [20] 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.

3 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 hyper-parameters 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 (Eq. 1) consists of samples that are organized as *k* dimensional vectors:

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

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

$$X_* = (x_{k+1}; x_{k+2}; ... ; x_n) \tag{2}$$

This procedure is repeated analogously in the case of the input and output training set for y coordinate, Y and 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 hyper-parameters optimization.

Predictive models are defined using input and output training and input testing set, obtained hyper-parameters 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 the z-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 ($\hat{x}, \hat{y}$) 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 becomes more sensitive to small errors caused by highly uncertain data points.

$$MD = \sqrt{\begin{bmatrix} x-\hat{x} & y-\hat{y} \end{bmatrix} \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}^{-1} \begin{bmatrix} x-\hat{x} \\ y-\hat{y} \end{bmatrix}} \tag{3}$$

where σ_x and σ_y are predictive variances for first and second Gaussian predictive model, respectively.

4 Results

We observed changes of the Mahalanobis distance through exercise sequence. When the 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 [21]. 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.

$$P = \sqrt{sens \cdot spec} \tag{4}$$

$$sens = \frac{TP}{TP + FN} \tag{5}$$

$$spec = \frac{TN}{TN + FP} \tag{6}$$

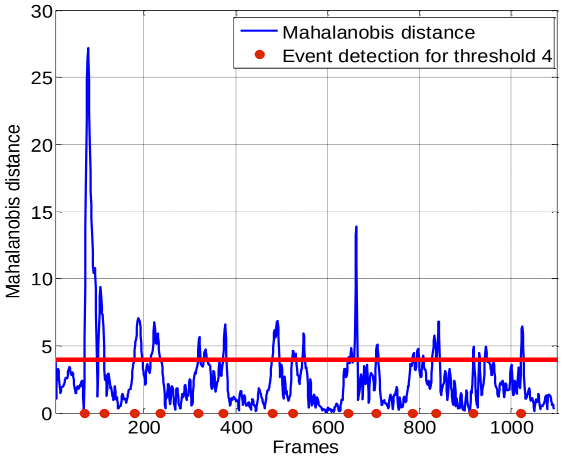

Fig. 3 Mahalanobis distance with event detection.

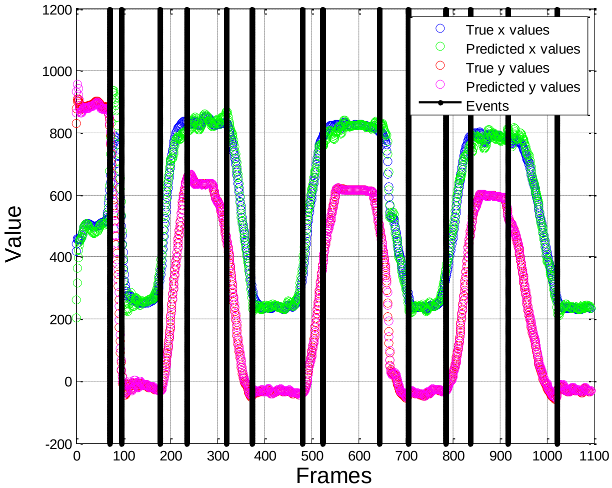

Fig. 4 True and predicted values with detected events.

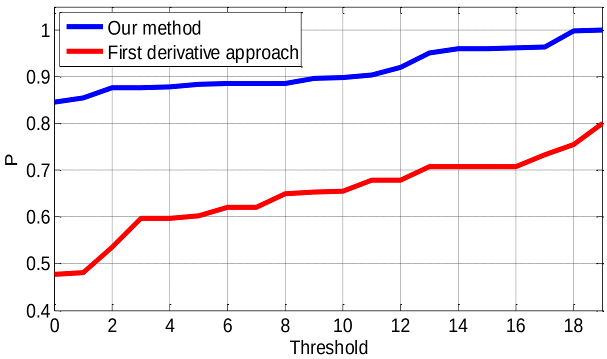

Fig. 5 A comparison of our method and first derivative approach based on sensitivity and specificity criteria.

In relations (Eq. 5) and (Eq. 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 sensitivity and specificity indicate better performances of the approach, hence figure 5 clearly illustrate the superiority and advantage of our approach.

5 Conclusion and future 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.

Acknowledgments 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

1. C. Goetz, B. Tilley, S. Shaftman et al. Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results. Movement Disorders, vol. 22, pp. 2129-70, 2008.
2. C. Cruz-Neira, D. Sandin, T. DeFanti et al. The Cave: Audio Visual Experience Automatic Virtual Environment. Communications of the ACM, vol. 35, no. 6, pp. 64-72, 1992.
3. T. Starner, B. Leibe, D. Minnen et al. The Perceptive Workbench: Computervision-Based Gesture Tracking, Object Tracking, and 3d Reconstruction of Augmented Desks. Machine Vision and Applications, vol. 14, no. 1, pp. 59-71, 2003.
4. A. Gama, T. Chaves, L. Figueiredo et al. Guidance and Movement Correction Based on Therapeutics Movements for Motor Rehabilitation Support Systems, 14th Symposium on Virtual and Augmented Reality, pp. 191 – 200, 2012.
5. S. Lee. Automatic Gesture Recognition for Intelligent Human-Robot Interaction. Proc. 7th International Conference on Automatic Face and Gesture Recognition, pp. 645-650, 2006.
6. H. Park, D. Jung and H. Kim. Vision-Based Game Interface Using Human Gesture, Advances in Image and Video Technology, pp. 662-671, 2006.
7. J. Alon, V. Athitsos and S. Sclaroff. Accurate and efficient gesture spotting via pruning and subgesture reasoning. Proc. IEEE ICCV Workshop on Human Computer Interaction, pp. 189–198, 2005.
8. R. Oka. Spotting method for classification of real world data. The Computer Journal, vol. 41, no. 8, pp. 559–565, 1998.
9. T. Darrell, I. Essa and A. Pentland. Task-specific gesture analysis in real-time using interpolated views. Pattern Analysis and Machine Intelligence, vol. 18, no. 12, pp. 1236–1242, 1996.
10. J. Kruskall and M. Liberman. The symmetric time warping algorithm: From continuous to discrete. Time Warps, String Edits and Macromolecules, pp. 125–162, 1983.
11. P. Morguet and M. Lang. Spotting dynamic hand gestures in video image sequences using Hidden Markov Models. Proc. IEEE International Conference on Image Processing, pp. 193–197, 1998.
12. T. Starner, J. Weaver and A. Pentland. Real-time American Sign Language recognition using a desk and wearable computer based video. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375, 1998.
13. F. Chen, C. Fu and C. Huang. Hand gesture recognition using a real-time tracking method and Hidden Markov Models. Image and Video Computing, vol. 21, no. 8, pp. 745–758, 2003.
14. A. Wilson and A. Bobick. Parametric Hidden Markov Models for gesture recognition. Pattern Analysis and Machine Intelligence, vol. 21, no. 9, pp. 884–900, 1999.
15. D. Y. Kwon (2008), A Design Framework for 3D Spatial Gesture Interfaces, PhD, ETH, Switzerland.
16. K. Kahol, P. Tripathi and S. Panchanathan. Automated Gesture Segmentation from Dance Sequences. Proc. IEEE 6th International Conference on Automatic Face and Gesture Recognition, pp. 883 – 888, 2004.
17. J. Kocian, A. Girard, B. Banko et al. Dynamic system identification with Gaussian processes. Mathematical and Computer Modelling of Dynamical Systems, vol. 11, no. 4, pp. 411-424, 2005.
18. Bruno Nery and Rodrigo Ventura. A dynamical systems approach to online event segmentation in cognitive robotics. Paladyn. Journal of Behavioral Robotics, vol. 2, no. 1, pp. 18–24, 2011.
19. Bruno Nery and Rodrigo Ventura. On the scalability and convergence of simultaneous parameter identification and synchronization of dynamical systems. Complex Systems, vol. 22, no. 3, pp. 203–219, 2013.

20. C. Rasmussen and C. Williams. Gaussian processes for machine learning. The MIT Press, 2006.
21. M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. Information Processing and Management, vol. 45, no. 4, pp. 427-437, 2009.