Segmentation of human upper body movement using multiple IMU sensors

Takashi Aoki, Jonathan Feng-Shun Lin, Dana Kulić and Gentiane Venture

Abstract— **This paper proposes an approach for the segmentation of human body movements measured by inertial measurement unit sensors. Using the angular velocity and linear acceleration measurements directly, without converting to joint angles, we perform segmentation by formulating the problem as a classification problem, and training a classifier to differentiate between motion end-point and within-motion points. The proposed approach is validated with experiments measuring the upper body movement during reaching tasks, demonstrating classification accuracy of over 85.8%.**

I. Introduction

The analysis of human body motion is useful in many fields such as medical diagnosis, sports, rehabilitation and in-home monitoring. Human motion observation is also used intensively for human-robot interaction. Recently, cleaning, home-care and security robots are beginning to be widely used in human environments such as at home and in the office [1][2], and these robots require accurate information about the location and the activities of any humans in their environment. Optical motion capture is commonly used for observing human movement in the laboratory setting [3][4][5], however, optical motion capture may not be suitable for many applications where low-cost, ambulatory sensors would be preferred. Lower cost optical solutions such as video cameras, stereo cameras and Kinect reduce the cost and setup time, but suffer from occlusion. Avoiding occlusion by the use of multiple cameras is commonly recommended. However, even if occlusion is reduced, the observation area is limited.

To enable low-cost observation without limitations to observation area, ambulatory sensors would be preferable. In this paper we propose to use three inertial measurement units (IMU) sensors to measure upper body motions. Classical methods of motion analysis utilize the joint angle information [6][7][8]. Computing the inverse kinematics from the IMU data and integrating the IMU sensor information is prone to drift, therefore we propose to use angular velocities directly. Without modeling the human body, we avoid the need to know the body dimensions, the exact placement of the sensors and the errors caused by joint state estimation. Therefore, for accurate pattern recognition of human body motion from IMU data, we use the angular velocity data directly [9]. The use of kinematics for modeling is also not necessary. Segmentation is an important issue for analysis of human motions. When human beings watch the motions of others, they recognize them as segmented motion patterns. The analysis of human motions is affected by the accuracy of segmentation, as motion patterns are generated from the segmented data.

This paper proposes a segmentation algorithm solely based on the IMU angular velocities.

The paper is organized as follows: in Section II we describe the segmentation algorithm. In Section III we describe the experimental set-up. In Section IV we discuss the obtained results, and in Section V we conclude the paper and overview directions for future work.

II. Proposed Segmentation algorithm

Segmentation points are defined as those points in the continuous data stream where the human motion is changing, e.g., from rest to movement, or from one movement type to another. By identifying the segment points, motion can be divided into different behaviors. We formulate segmentation as a classification problem by defining two classes: segment points, those instances in the data stream when the person is at rest or switching their motion type, and non-segment points, those instances where the person is executing a specific action [10]. This approach allows the classifier to learn the appropriate threshold differentiating the two types of points automatically, rather than relying on a priori knowledge or tuning.

In order to perform the classification, it is necessary to first define the features to be used by the classifier. Unlike previous work [11], where segmentation was performed following conversion to joint angle data using a detailed kinematic model and extended Kalman filtering, we propose to use the measured angular velocity directly to obtain the classifier features, avoiding the need for detailed kinematic modeling, or the need to handle IMU drift. Prior to feature computation, the signal is low-pass filtered using the MATLAB filtfilt function with a cut off frequency of 10 Hz. Then, the absolute value of each angular velocity is computed, as shown in Equation (1):

$$\|\omega_{\text{Position}}\| = \sqrt{\omega_x^2 + \omega_y^2 + \omega_z^2} \tag{1}$$

A vector of features is generated by concatenating a vector consisting of angular velocity magnitudes from IMUs attached to each body part:

$$\omega_{t=t_0} = \begin{bmatrix} \|\omega_{\text{Left}}\|_{t_0} \\ \|\omega_{\text{Body}}\|_{t_0} \\ \|\omega_{\text{Right}}\|_{t_0} \end{bmatrix}^T \tag{2}$$

To incorporate temporal information into the classifier, the feature vectors at each time step were further concatenated with feature vectors at adjacent time steps, using a windowing approach, centered at time *t*. The training data was also

T. Aoki and G. Venture are with the Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan.(e-mail: s128999s@st.go.tuat.ac.jp; venture@cc.tuat.ac.jp.).

J. Lin and D. Kulić are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada.(e-mail: jf2lin@uwaterloo.ca; dkulic@uwaterloo.ca).

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

downsampled to ensure that the number of segmentation points and non-segmentation points are the same in order to avoid bias for the number of training data.

The classifier algorithm was *k*-NN (k=2) [11]. In the case of the initial data, since there is no previous data to form a full window, the first data was used in place of the previous data. In the same way, the final data was used to pad the rest of the window at the end of each trial.

III. EXPERIMENTS

A single participant was recruited for the experiment. Three IMU sensors were attached to the participant, on the left wrist, body trunk and right wrist as shown in Fig. 1. The IMU sensors are Shimmer2R; sensor specifications are shown in Table I.

The participant sat in front of a table with multiple cups in front and to the side of him, as illustrated in Fig. 2 (a). The participant grasped each cup on the table and brought in front of himself as shown in Fig. 2 (b). The experiments were performed with 5 cup located on the table. The participant was asked to use each hand to grasp the cup, as well as two handed grasping motions. The top view of the cup layout is shown in Fig. 3. The participant performed each operation three times, to grasp the cup in each position on the table with each arm (left hand, both hands and right hand). To analyze the performance of the automated segmentation algorithm, manual segmentation results are generated as ground truth. The data obtained by each sensor and the manual segment time are shown in Fig. 4. The manual segment data together with the measurement data is used as the training data set.

The manual segmentation is generated using the recorded video during experiments and observing the waveform corresponding to the experiments. From the video, we define a motion segment each time that the participant is at the start point, grasping cup point, putting cup point, and finishing point, as shown in Fig. 4. Those manual segmentation points are defined as the ground truth, knowing that there is a large inter subject variation in manual segmentation. To avoid this we had the data segmented by several observers. It is reported in [12] that inter-personal error was generally around ± 0.125 s. So a segment range was set to 20 frames (± 0.156 s).

We next compare the manual segmentation to the segmentation results obtained by the classification algorithm described in Section III. Classifier accuracy is determined using leave-one-out cross-validation, by training on all the trials except for the test trial in each fold.

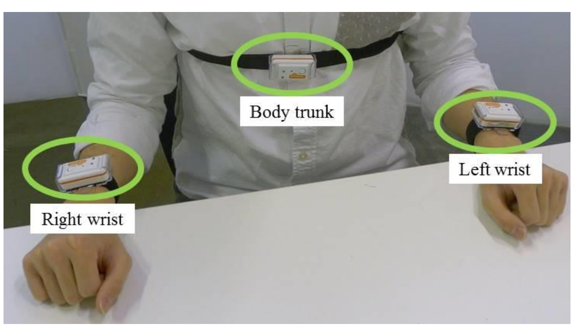

Figure 1. Attached position of 3 IMU sensors.

TABLE I. SPECIFICATION OF SHIMMER2R

		Specification
Dimension		53×32×16 mm
Mass		25 g
Accelerometer	**Range**	±60 g
	Sensitivity	187 mV/g
Gyroscope	**Range**	±500 deg/sec
	Sensitivity	2 mV/ (deg/sec)
Data frequency		128 Hz

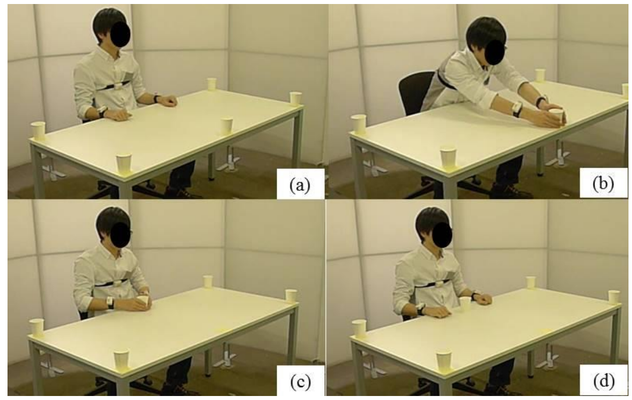

Figure 2. Condition of grasping experiments *(Both hands, center cup)*.

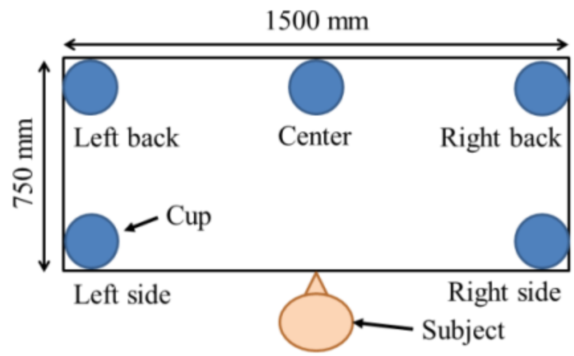

Figure 3. Cup layout.

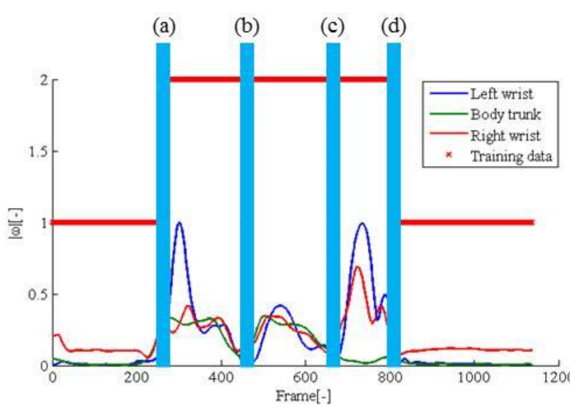

Figure 4. Segment points.

IV. Results

A. Effect of window size

We first investigate the influence of window size on the classification accuracy. Table II shows the classification accuracy for varying window sizes. A window size of 19 frames shows the highest accuracy rate, for both segment and non-segment points.

Fig. 5 shows the results for window size 1 for the experiment of both hands grasping the right side cup. The red cross points represent the training data. When this value is 1, it indicates that this point is a segment point. When this value is 2, it indicates that this point is a non-segment point. The blue circle points are the classifier generated result. When this value is -2, it indicates that this point is classified as a segment point. When this value is -1, it indicates that this point is classified as non-segment point. When the classifier generated result is same as the manual label for a given frame, classification is considered successful in this frame. Each line shows the normalized value of $\|\omega_{\mathrm{Position}}\|$. The blue line shows the left wrist, the green line shows the body trunk and the red line shows the right wrist angular velocity magnitude. Those lines are the testing data.

TABLE II. Classification average as a function of window size *(Both hands experiments)*

	Window size [frame]			
	1	**5**	**7**	**11**
Non-seg.	89.0%	89.0%	89.5%	89.5%
Seg.	83.9%	86.4%	87.3%	88.0%
	Window size [frame]			
	15	**17**	**19**	**21**
Non-seg.	89.5%	89.6%	**90.0%**	89.9%
Seg.	88.6%	88.6%	**88.7%**	88.7%

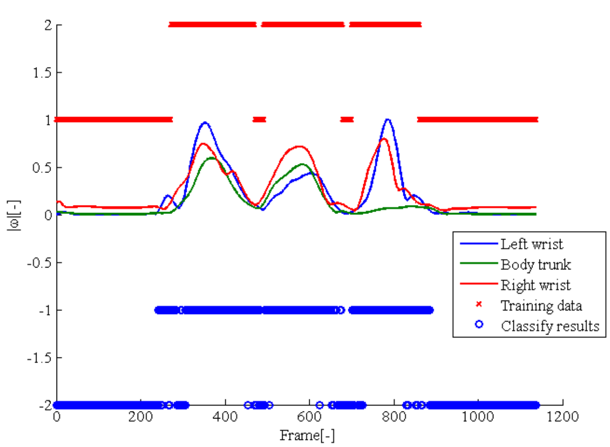

Figure 5. Both hands, grasping the right side cup *(Window size: 1)*.

B. Effect of Motion Type

Table III shows the average classification accuracy as a function of the cup position when using both hands, and each of the left and right hands. A classification accuracy of 85.8% or higher is achieved in all experiments.

Fig. 6 shows the segmentation results for the experiment of both hands grasping the right side cup (window size: 19 frames). The three actions are clearly visible from the waveform. It can be seen that segmentation performance is improved as compared with Fig. 5.

TABLE III. Classification results *(The average rate of correct)*

		Left back [%]	Left side [%]	Center [%]	Right side [%]	Right back [%]
Left	Non-seg.	92.0	92.9	92.3	94.1	92.4
	Seg.	90.9	85.8	87.3	86.3	90.4
Both	Non-seg.	85.7	92.0	88.1	90.7	87.0
	Seg.	88.7	88.1	94.8	87.9	90.2
Right	Non-seg.	91.0	94.1	96.5	95.7	93.4
	Seg.	88.3	90.7	87.2	80.2	86.7
Total	**Non-seg.**	**90.6**	**91.7**	**94.5**	**92.5**	**92.0**
	Seg.	**88.3**	**89.5**	**87.5**	**85.8**	**88.2**

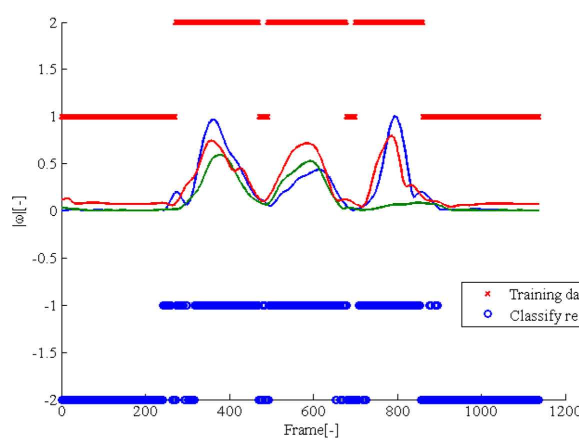

Figure 6. Both hands, grasping the right side cup.

C. Removal of small segmentation point

Fig. 7 shows the segmentation results for the experiment when both hands are grasping the right back cup. There are three segment point groups away from the correct segmentation points, circled in the figure. These points correspond to brief pauses in the motion during the performance of a motion primitive.

To remove these spurious short segment point sequences, only segment points in a continuous sequence of more than 4 time steps were considered valid. Fig. 8 shows the results when applying this simple post-processing step. The first and second separated point groups were removed. But the third point group was still present.

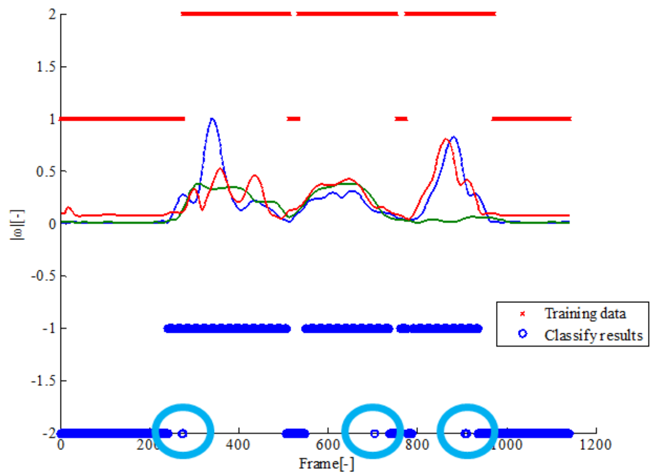

Figure 7. Both hands, grasping the right side cup *(Not adapted)*.

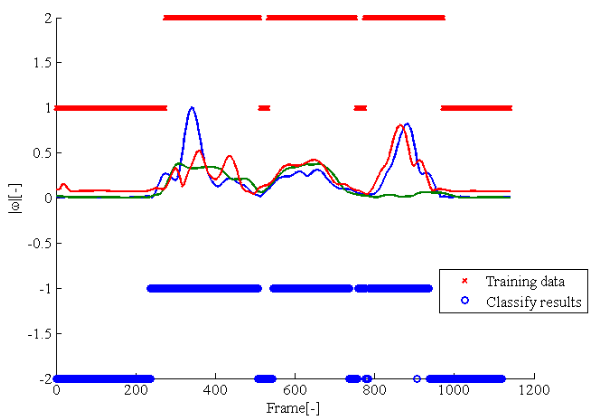

Figure 8. Both hands, grasping the right side cup *(Adapteda)*.

V. CONCLUSIONS AND FUTURE WORK

This paper proposed an approach for automated segment point classification from direct IMU measurements. An experiment was conducted focused on upper body movements. Using the data obtained by multiple IMU sensors, classification into segment and non-segment points was performed, showing an overall accuracy of over 85.8%.

For future work, we would like to identify what kind of movement is being performed within the non-segmentation points, and explicitly identify any motion pauses. The use of such a system in daily life, would allow the robot to recognize human actions and offer appropriate assistance.

REFERENCES

[1] M. Dunbabin, and L. Marques, ”Robotics for Environmental Monitoring,” IEEE Robotics and Automation Magazine, March, 2012.

[2] M. Okamura, M. J. Matari_c, and H. I. Christensen, ”Medical and Health-Care Robotics,” IEEE Robotics and Automation Magazine, September, 2010.

[3] G. Venture, K. Ayusawa, and Y. Nakamura, “Motion capture based identification of the human body intertial parameters,” International IEEE EMBC, pp. 4575-4578, 2008.

[4] G. Venture, K. Ayusawa, and Y. Nakamura, “Optimal estimation of human body segmentation dynamics using realtime visual feedback,” In Proceeding of the 2009 IEEE/RSJ International Conf. on Int. Robots and Systems, pp. 1627-1632, 2009.

[5] K. Ayusawa, G. Venture, and Y. Nakamura, ”Real-time implementation of physically consistent identification of human body segments,” ICRA 2011, pp. 6282-6287,2011.

[6] D. Kulić, W. Takano, and Y. Nakamura, “On-line segmentation and clustering from continuous observation of whole body motions,” IEEE/Transactions on Robotics, vol. 25, No. 5, pp.1158-1166, 2009.

[7] D. Kulić, and Y. Nakamura, “Scaffolding on-line segmentation of full body human motion patterns,” In Proceeding of the IEEE/RSJ International Conf. on Int. Robots and Systems, pp. 2860-2866, 2008.

[8] D. Kulić, C. Ott, D. Lee, J. Ishikawa, and Y. Nakamura, “Incremental learning of full body motion primitives and their sequencing through human motion observation,” The International Journal of Robotics Research 31(3), pp. 330-345, Nov. 2012.

[9] T. Aoki, G. Venture, and D. Kulić, “Segmenatation of human body movement using internal measurement unit,” IEEE International Conference on systems Man and Cybemetics, SMC2013, pp. 1181-1186, 2013.

[10] J. F. S. Lin, V. Joukov and Dana Kulić, “Full-body multi-primitive segmentation using classifiers,” IEEE-RAS International Conference Humanoid Robots 2014, pp. 874-880. November, 2014,

[11] J. F. S. Lin, V. Joukov, and Dana Kulić, “Human Motion Segmentation by Data Point Classification,” IEEE International Conference EMBC 2014, pp. 9-13. August, 2014,

[12] N. Doshida,“Classification of Grasping Motions for Smooth Communication between Human and Robots,”Graduation thesis of Tokyo University of Agriculture and Technology, March, 2015.