Vision-based human motion analysis for the development of automated rehabilitation tools

Daniel Leightley*, Baihua Li*, Jamie S. McPhee*†, Moi Hoon Yap* and Kevin Tan*
*School of Computing, Mathematics and Digital Technology
†School of Healthcare Science
*{d.leightley, b.li, j.s.mcphee, m.yap, k.tan} @mmu.ac.uk

***Abstract*—Patients requiring physical rehabilitation, such as those suffering long-term chronic disease or recovering from illness or injury are reliant upon rehabilitation programmes to recover. Tele-rehabilitation has been proposed as a promising development to remove physical barriers of locality and place rehabilitation services within the patients home. We propose an interdisciplinary research approach to generate an assistive rehabilitative platform through the development of computer vision techniques, particularly *near-real-time* motion and gesture analysis. In this paper, we present our initial findings on the comparison of two popular machine learning algorithms, Support Vector Machines and Random Forests for their ability to accurately predict periodic kinematic actions.**

I. Introduction

Robust classification and interpretation of human activity in relation to recognition of actions is an active research area. The application of human recognition has received wide attention in recent years, in fields such as surveillance, healthcare, and the gaming industry. The release of the Microsoft Kinect (Kinect), which is a peripheral accessory intended for use with Xbox 360 gaming console enables natural human-machine interaction. The Kinect is able to track a users gestures and body movements to control the commands of a game instead of touching or moving a keyboard, control pad or mouse [1].

Kinect incorporates an infra-red and a RGB camera to create a three-dimensional (3D) map of the area in front of the device, and uses a randomised decision forest algorithm to automatically detect and determine anatomical joints on the body of the user (kinematic output) [1]. Going beyond the Kinects application in gaming, there are other promising capabilities of the depth sensor, RGB images and body tracking algorithms that could be exploited further in rehabilitation, particularly home-based rehabilitation [2], [3].

In this short paper, we present our initial experimental findings of the comparison of two training-based machine learning algorithms (MLAs), namely, Support Vector Machines (SVM) [4] and Random Forests (RF) [5] on how accurately they classify periodic actions sequences based on Kinect output.

II. Methods

Kinematic data used in this work was acquired by use of the Kinect and Kinect for Windows Software Development Kit. The application acquired the 3D positions of x, y, z axis position of 20 predefined feature joints at a rate of up to 30 frames-per-second [1].

A. Dataset

A set of 10 participants performing 10 periodic actions was recorded. The participants were asked to perform the actions periodically within the defined motion area for a 10 second period. Participants were asked to ensure that they had a neutral standing pose to which they would maintain at the start and the end of the action. In order to assess the classification accuracy, model training and classification times, the dataset was randomly split into two subsets, where five participants formed the training set and five formed the testing set. To conduct the training of the classifiers, one participant from the training set was randomly selected for training, with each classification trial introducing a new training participant until all five training participant were utilised. The testing set was tested against the classifiers at each trial.

B. Classifiers

Based on statistical learning theory, SVM is a supervised learning MLA first introduced in 1995 [4]. An SVM produces a model that represents the training data by learning the optimum separating hyperplane between classes; these hyperplanes are defined as the support vectors and represent each class.

RF is a MLA consisting of multiple decision trees constructed by supervised learning of a training set, introduced in 2001 [5]. An RF model is constructed by using the bootstrap method to generate n_{tree} of decision trees which are each provided with randomly selected samples of the training input and then all decision trees are combined into a decision forest. For each bootstrap, an unpruned classification tree with a random sample $mtry$ (e.g 3) of the training data, represents the number of samples used for each tree.

III. Results and Discussion

The training time for each classifier varied depending on the number of training participants and parameters selected (Table II). SVM overall, and for each training trial, was the quickest to train, while RF took considerably longer to train for each trial and overall, as demonstrated in Table II. With the incremental increase of training participants, the training of the classifiers becomes more complex, consequently the training time of the classifiers increases.

Overall average classification times were significantly reduced to millisecond predictions, compared with training

TABLE I
AVERAGE OVERALL CLASSIFICATION ACCURACY PER TRIAL

Trial:	1		2		3		4		5	
Overall Accuracy [%]:	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF
Jumping	19.66	39.97	11.66	15.31	57.7	37.53	73.23	76.65	83.34	77.25
Arms Movement	53.29	54.75	46.31	68.97	61.73	67.37	71.89	67.68	81.76	83.68
Pickup Object	45.75	53.75	41.02	51.09	62.25	67.82	79.37	75.81	89.68	87.81
Squats	27.65	47.27	19.82	30.64	54.22	44.17	72.98	64	80.51	84.18
Walking	9.09	15.3	27.16	64.68	61.22	80.67	77.29	73.89	81.6	90.31
Jogging	20.88	51.15	39.79	33.93	65.34	75.78	61.85	69.33	83.33	86.91
Bending to Toes	34.56	8.25	65.08	69.79	69.92	61.8	81.19	71.79	80.73	86.99
Standing to Seated	40.79	49.49	38.39	52.15	60.2	57.27	73.11	65.01	89.19	80.47
Upper Body Twist	76.3	74.65	82.04	77.31	87.7	85.37	84.91	85.04	86.77	86.64
Arm Stretch	58.91	64.06	57.94	63.93	62.42	60.08	64.44	78.61	79.78	82.68
Average [%]:	38.69	45.86	42.92	52.78	64.22	63.79	74.03	75.02	83.67	84.69

TABLE II
COMPUTATIONAL TRAINING TIME FOR EACH CLASSIFIER

Trial	1	2	3	4	5
Classifier	Training time [Sec]				
SVM	0.71	1.92	4.16	7.84	11.44
RF	1.56	5.78	7.33	10.63	15.26

times, with on average RF performing faster than SVM for each trial, as demonstrated in Table III. SVM, as with training the models, predicting classification is computationally more expensive than RF. The average classification time of each activity also increased at each trial with the introduction of additional training set, albeit on a lower scale (100 up to 400 milliseconds) as demonstrated in Table III.

TABLE III
AVERAGE CLASSIFICATION TIMES PER CLASSIFIER AND TRIAL

Trial	1	2	3	4	5
Classifier	Testing time [Sec]				
SVM	0.013	0.021	0.037	0.045	0.052
RF	0.007	0.014	0.014	0.012	0.016

Classification accuracy differed between SVM and RF. Furthermore, the number of participants in the training set greatly affected classification accuracy. As demonstrated in Table I, RF exhibited the highest average overall with 84.69%, however, only marginally outperforming SVM which obtained 83.67%. The largest overall margin produced between the two classifiers was using two training participants, where a margin of 9.86%, with RF at 52.78% and SVM with 42.92%. With both SVM and RF, increasing the number of training participants improved classification accuracies considerably with SVM and RF having similar overall mean accuracies from trial 3 to 5.

SVM and RF were compared on their ability to predict action sequences. The results presented in this paper illustrate that RF offers classification performance advantages compared to SVM in the classification of actions based on Kinect kinematic data and potential use in rehabilitation. The results show that both SVM and RF were suited towards classification of kinematics, with RF capable of predicting accuracies in a significantly shorter time than SVM (between $100 - 400$ milliseconds). In contrast, RF was computationally more expensive when training compared with SVM.

Classification between the broad range of actions was reliable when all five training participants, with even subtle changes between Walking and Jogging classified accurately. However, confusion remained in trial 1 for the aforementioned actions, with misclassification between the two actions having a negative affect on accuracy (Table I c1.). These actions were repeatedly misclassified until trial 3 where the accuracy stabilised. The Standing to Seated action presented a challenge, due to occlusion of the chair and other limbs, yet both classifiers were capable of achieving acceptable accuracy results. In our study, we found that by increasing the number of participants enabled a better tolerance on the variation in anatomical joints amongst the different participants, leading to improved classification results. A larger training set could improve the results further and remove misclassification of actions due to anatomical differences.

IV. CONCLUSION

Our initial results showed that both SVM and RF could reliably classify human actions and gestures. RF was marginally more accurate, but more expensive in terms of training time, however SVM was more expensive in terms of prediction times. Our study has presented a number of possibilities for using SVM or RF with Kinect kinematics in a rehabilitation setting, such as near-real-time classification of actions and multiple gestures for analysis.

REFERENCES

[1] J. Shotton, A. Fitzgibon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Patter Recognition*. IEEE Press, 20 - 25 June 2011, pp. 1297 – 1304.

[2] H. Aoki, M. Miyazaki, H. Nakamura, R. Furukawa, R. Sagawa, and H. Kawasaki, "Non-contact respiration measurement using structured light 3-d sensor," in *SICE Annual Conference*, Akita University, Akita, Japan, August 20 - 23 2012, pp. 614 – 618.

[3] R. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. Webster, L. Denehy, and A. Bryant, "Validity of the microsoft kinect for assessment of postural control," *Gait and Posture*, vol. 36, no. 3, pp. 372 – 377, July 2012.

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273 – 297, September 1995.

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 25, no. 1, pp. 5 – 32, 2001.