Human Activity Recognition for Physical Rehabilitation

Daniel Leightley*, John Darby*, Baihua Li*, Jamie S. McPhee[†] and Moi Hoon Yap*

*School of Computing, Mathematics and Digital Technology
[†]School of Healthcare Science
Manchester Metropolitan University, M1 5GD, UK
dleightley@ieee.org, {j.darby, b.li, j.s.mcphee, m.yap}@mmu.ac.uk

Abstract—The recognition of human activity is a challenging topic for machine learning. We present an analysis of Support Vector Machines (SVM) and Random Forests (RF) in their ability to accurately classify Kinect kinematic activities. Twenty participants were captured using the Microsoft Kinect performing ten physical rehabilitation activities. We extracted the kinematic location, velocity and energy of the skeletal joints at each frame of the activity to form a feature vector. Principle Component Analysis (PCA) was applied as a pre-processing step to reduce dimensionality and identify significant features amongst activity classes. SVM and RF are then trained on the PCA feature space to assess classification performance; we undertook an incremental increase in the dataset size. We analyse the classification accuracy, model training and classification time quantitatively at each incremental increase. The experimental results demonstrate that RF outperformed SVM in classification rate for six out of the ten activities. Although SVM has performance advantages in training time, RF would be more suited to *real-time* activity classification due to its low classification time and high classification accuracy when using eight to ten participants in the training set.

Index Terms—Kinect, Machine Learning, Random Forests, Support Vector Machines.

I. Introduction

Robust interpretation and classification of human activity is an active research area which has received wide attention in recent years, with applications in surveillance, healthcare, and gaming [1]. There has been a long-standing interest in vision-based human motion and activity recognition for human-computer interaction (HCI). The Microsoft Kinect (Kinect) is a low-cost peripheral accessory intended for use with the Xbox 360 gaming console. The Kinect allows for real-time body detection and tracking of human activities and gestures. By incorporating infra-red and RGB camera technology, the underlying body detection algorithms create a three-dimensional (3D) depth map of the area in front of the device, randomised decision forest algorithms are then used to automatically detect and determine anatomical joints on the body of the user and stream the 3D coordinate location for each joint [2].

The innovation of the Kinect enables real-time HCI through recognition of user's gestures and body movements to, for example, control a character or gameplay elements. However, despite advances in vision-based HCI, the main constraint is machine understanding of the gesture, action and behavioural context, which still remains an open and ambitious problem to solve. Motivated by current limitations, we are focused on the promising application of the Kinect for assessment of physical rehabilitation activities for full-body tracking and gestured-based detection by an intelligent HCI system [3], [4].

In this study, we use kinematic data obtained from the Kinect, in particular vertical location, velocity and energy of skeletal joints at each frame to provide a detailed analysis of two popular machine learning algorithms (MLA). SVM [5] and RF [6] are assessed with how accurately they classify unseen test data. To our knowledge, no other paper has sought to compare MLA's in the context of Kinect kinematics. We seek to provide analysis on the average classification accuracy, model training and classification time to determine which is suitable for classification of Kinect kinematic data.

The remainder of this paper is organised in the following way: Section II, review of related work on activity recognition with Kinect. Section III, outline of the methods, dataset and discussion of MLA's. Section IV, presents experimental results on the recognition of human activity. Finally, we discuss and conclude our results in Section V and VI.

II. Related Work

Human pose, action and activity recognition has been extensively studied in the literature; we refer the reader to recent surveys for a detailed summary, [1], [7], [8]. The availability of low-cost depth cameras such as the Kinect and advances in motion tracking technology have created a high demand on adapting current and developing new techniques for real-time HCI vision-based activity recognition. In the field of activity tracking and recognition there are a limited number of studies that sought to exploit Kinect skeletal output for body tracking, posture analysis and action recognition. Clark et al. [9] undertook a review of the Kinect body tracking algorithms for posture control. The study compared Kinect with a VICON marker-based tracking system for joint accuracy. The authors concluded the potential use of the Kinect in posture analysis due to its high degree of joint accuracy.

Bhattacharya et al. [10] analysed SVM and Decisions Trees (DT) for how they perform in detecting and classify gestures used in aircraft marshalling. The study found that SVM outperformed DT, though the authors noted that DT were susceptible to participant anatomical differences whereas SVM

was participant independent and not affected by differences in body posture. Additionally, Zhang et al. [11] presented a method for utilising kinematic output to recognise and segment the time-sequential postures of a golf swing. By transforming kinematic output to a symbol sequence through vector quantisation enabled a SVM GMM-KL kernel to score and classify golf swings accurately.

Lately, Patsadu et al. [12] performed a comparison of several popular MLA's to assess the performance in predicting human falling motions. The authors found that backpropagation neural network exceeded all other classifiers, with SVM second in predicting three activities, namely, sitting down, lying down and standing on a small training and testing dataset. However, closer inspection of this prior work reveals several important biases that may have affected the conclusions. Firstly, the study compared performance of the MLA's on a limited set of activities, which are distinctively different to one another. Secondly, the limited size of the dataset fails to factor in anatomical differences when scaled. Finally, the Kinect is not designed to capture lying down poses, leading to unpredictable inference of joints depending on the subject and activity [2].

SVM and RF are popular classifiers used in a number of domains, however in the field of 3D activity recognition, RF has had limited use ([1], [7], [13]). In other domains several studies have sought to investigate the performance differences in classification by SVM and RF. Nitze et al. [14] sought to provide a comparison for crop type classification (for agriculture) by use of image representations of different crop fields. Where as, Statnikov et al. [15] sought to compare for microarray-based cancer diagnosis and prediction based on gene profiling. Finally, Tang et al. [16] assessed for spam detection based on IP addresses. SVM was deemed by two of the studies ([14], [15]) to be the most accurate with its predictions, with one study finding RF more applicable [16]. A theme apparent in many comparisons is that SVM was more accurate due to it being less sensitive to the choice of input parameters than RF.

In the context of our own research, we build upon the shortcomings of Patsadu et al. [12] and the limited use of RF to determine which classifier, SVM or RF, performs reliably in classifying physical rehabilitation activities based on Kinect 3D kinematics. We believe that by introducing a larger dataset and increasing the number activities performed we can reliably attain which classifier is the most suited for our domain.

III. METHODS

In this section, we present our method in the following sub-sections: introduction to SVM and RF, dataset, PCA and kinematic reduction technique and model training.

A. Support Vector Machines

Based on statistical learning theory, SVM is a supervised learning classifier [5]. 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 symbolise each class.

A kernel is utilised by an SVM to minimise both the empirical risk and the model complexity. In this study, we use the radial basis function (RBF) kernel, which non-linearly maps samples into a higher dimensional space, the RBF can control the relationship between class labels and attributes when they are non-linear [17]. RBF can be defined as

$$K(x_i, y_i) = exp(-\gamma||x_i - y_i||^2), \gamma > 0 \qquad (1)$$

where x and y are the training and label respectively, γ is the kernel parameter. Parameter C is a user-defined parameter that controls the trade-off between model complexity and empirical error in SVM. In addition, the parameter γ determines the shape of the separating hyperplane in the RBF kernel.

When the trained SVM model is provided with data for classification, it is able to determine the best fit in relation to the support vectors leading to predicted categorisation of the data. The software used to implement SVM is libSVM Toolbox for Matlab [18] with the RBF kernel.

B. Random Forests

RF is a classifier consisting of multiple decision trees constructed by supervised learning of a training set [6]. An RF model is constructed by using the bootstrap method to randomly 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, a random $mtry$ (default 3) sample of the training data is used which determines the size of an unpruned classification tree.

When an RF model is provided with data for classification, the model predicts a corresponding class based on the voting of all trees, where the class with the greatest number of votes is selected [6]. RF only requires one parameter, n_{tree}, which sets the number of decision trees to grow. The software used to implement RF was randomForest Toolbox for Matlab [19].

C. Activities and Data Collection

The dataset used in this study was acquired by Kinect and Kinect for Windows Software Development Kit [20]. The application obtained the 3D coordinates of 20 body joints, namely: feet, ankles, knees, hips, torso, shoulders, elbows, wrists, hands and head at a rate of 30 fps, where: x extends from the left to right, y indicates vertical position and z extends in the direction in which the Kinect is facing.

The Kinect (tilt 0 degrees) was placed on a tripod at a height of 0.7 meters (m) with the participant standing 2m from the device in a defined movement area of 0.5m ×0.5m. Participants were asked to perform activities periodically within the defined movement area for a 10 second period directly facing the Kinect. These activities were chosen to reflect activities of daily living as well as movements a person would perform when undertaking a physical rehabilitation program. Activities were performed periodically to characterise temporal variations. Participants were asked to assume a neutral standing pose at the start and end of the activity, in which they stood

still with legs fully extended and arms extended and relaxed by the side of the body. The aim was to ensure consistency between the training and testing datasets and, to limit the anatomical variance between the participants. A group of twenty participants (12 men, 8 women) performed a set of ten activities, resulting in the capture of 200 activities with a total of 60,225 frames.

The activities performed were as follows; Jumping: arms by the side and feet together jumping approximately 10cm off the ground; Arm Movement: arms extended along the frontal plane moving to a side-by-side position; Pickup Object: from a standing position bending down to pick up an object off the floor with the right hand; Squats: arms by the side, bending down so that gluteals approximately 10cm off the ground; Walking: walking on the spot, with feet raised approximately 5cm. Jogging: jogging on the spot, with feet raised approximately 8cm; Bending to Toes: from a neutral standing position, bending over and keeping the legs and upper spine as straight as possible, the arms were extended until the toes have been touched; Standing to Seated: neutral standing position, bend knees and sit on a stool located behind; Upper Body Twist: both arms raised vertically in front of the torso and twist from left to right; Arm Stretch: with both feet flat on the floor raise both arms vertically as high as possible.

D. Kinematic Reduction and Pre-Processing

The dataset was visually checked to ensure correct recording of the 3D position data for each activity; a row vector of 20 body-joint coordinates represents each frame. Each activity captured is aligned to the "hip-centre" joint to create a coordinate system relative to the "hip-centre" of the first frame. Where the original coordinate $\mathbf{P}_{n,i}(x,y,z)$ of the nth joint at the ith frame is subtracted by "hip-centre", $\mathbf{P}_{hipcentre,1}(x,y,z)$ of the first frame, as defined by

$$\mathbf{A} = \{\mathbf{P*}_{n,i}(x,y,z)|n = 1,\dots,20,\ i = 1,\dots,I\}$$
$$\mathbf{P*}_{n,i}(x,y,z) = \mathbf{P}_{n,i}(x,y,z) - \mathbf{P}_{hipcentre,1}(x,y,z) \quad (2)$$

where $\mathbf{A}$ is $(60 \times I)$ matrix of the "hip-centre" aligned activity with total frames I, $\mathbf{P*}$ is the aligned joint position.

Location, velocity and energy are representative kinematic features. They are used to represent the dynamic variation of each activity in our study. The vertical location and kinematic properties are discriminative compared to horizontal left to right and forward to backwards directions. To reduce the dimensionality of the activity feature vector, y-position (y), y-velocity (v_y) and energy (e) are extracted to form the activity feature vector $\mathbf{F}_{n,i}$ of the nth joint at the ith frame. We compute the velocity and energy for each joint position as follows

$$\mathbf{F}_{n,i} = \{(y_{n,i}, v_{y(n,i)}, e_{n,i})|v_{y(n,i)} = y_{n,i} - y_{n,i-5},$$
$$e_{n,i} = (v^2_{x(n,i)} + v^2_{y(n,i)} + v^2_{z(n,i)})\} \quad (3)$$

where $y_{n,i}$ is the aligned y in $\mathbf{P*}_{n,i}$, as shown in Eq.2, energy is determined as a sum of energy in x, y, z of each

joint. The velocity v_y and energy (e) are calculated over the period of 5 frames (i_{-5}) to provide increased tolerance for the measurement error presented by the Kinect.

Principle Component Analysis (PCA) was applied to the aligned dataset at each λ participant to reduce dimensionality of the aligned activity feature vector $\mathbf{F}$ by projecting the data into a lower-dimensional space. Transposing to low-dimensional feature space provides a number of benefits, such as reduced computational complexity, stabilisation of data noise and improved accuracy. In this study, the variance has been set as 98%, meaning we keep the dimensions (eigenvectors) that contain 98% of the variance for the projected datasets. The projected dataset is defined in the following way

$$\mathbf{T} = \{(\mathbf{F}_{n,i}, l_a)|n = 1,\dots,N,\ i = 1,\dots,I, l_a = 1,\dots,L\} \quad (4)$$

where $\mathbf{F}_{n,i}$ is the projected activity feature vector defined in Eq. 3 for the nth joint at the ith frame and l_a is the class label for each activity (e.g. Jumping class: 1, Walking class: 2).

E. Training

In order to assess classification accuracy, model training and classification time, the dataset was randomly split into two subsets, ten participants formed the training set and the remaining formed the testing set. With each experiment introducing another participant (denoted by λ) from the training set until all the participants were used. By training the classifiers in this way, the study was able to determine the suitable number of participants for training to achieve a stable classification rate.

TABLE I
OPTIMUM PARAMETERS FOR MODEL TRAINING FROM λ

λ participants/ Parameter	2	4	6	8	10
	SVM				
C	28	28	32	28	30
γ	8	8	8	6	6
	RF				
n_{tree}	400	700	800	800	800

To ensure optimal performance of each classifiers, parameter optimisation was performed as demonstrated in Table I.

For C and γ in SVM, the selection was undertaken according to the cross-validation method [17]. To perform cross-validation, the training set was segregated into two subsets of equal size. Then the classifier was trained on one subset (training data) and accuracy is tested with the introduction of the second subset. The optimisation process was repeated for each of the possible parameter in exponential steps for both C and γ between 10^{-4} to 10^5 and 10^{-6} to 10^3 respectively.

In RF, n_{tree} represents the number of trees to be generated requires optimisation. To perform optimisation, the range of trees has been tested with incremental increases of 100 between 100 up to 1000 trees. The optimised number of trees required for each experiment are shown in Table I. The experiment results suggest a consistent number of 800 n_{tree} was sufficient for training.

IV. Experimental Results for Recognition

In this section, we present our experimental results. The proposed method has been implemented in Matlab 2012a on a workstation with an IntelCore i7 processor and 8GB RAM.

A. Classification Accuracy

Our first experiment illustrated the benefits of using the reduced feature space of PCA to train the classifiers. We compare the classification accuracy versus not applying PCA, our experimental results are demonstrated in Fig. 1. The standard procedure for calculating classification accuracy is by assessing the number of correctly classified frames versus the total number of frames in the activity. Furthermore, classification accuracy was calculated as the average over all ten activities from each of the testing participants. We observe that by using 6 to 10 participants in the training dataset, use of PCA improved classification accuracy considerably, where 1 to 5 participants were used each classifier achieved similar low classification rate. All results presented hereafter are with PCA applied.
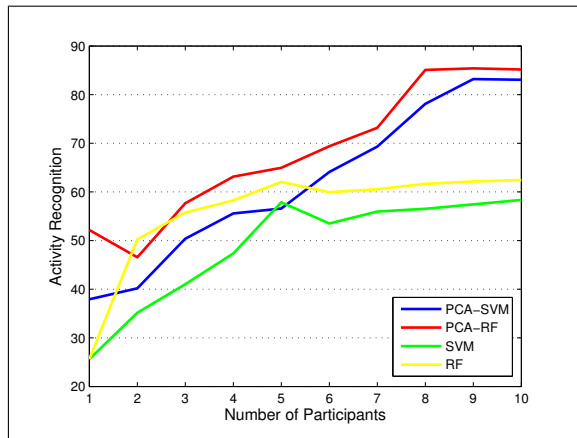

Fig. 1. Average activity classification rate for SVM and RF trained from an increasing number of λ)

We demonstrate in Fig. 2 the standard deviation (SD) at each incremental λ) increase. The standard procedure for calculating SD was the deviation of the average classification accuracy for each participant's activities. Observing Fig. 2, the deviation of error reduced, represented by an error bar, considerably for both SVM and RF, between participants 1 to 10 from, 24.03, 22.7 to 8.98, 7.07 respectively.

Initial findings demonstrate that classification accuracy improved and error reduced when the number of participants in the training set was incrementally increased, levelling off between 8 to 10 participants (Fig. 1 & Fig. 2). Table IV summarises classification accuracy for each activity from our proposed method. As demonstrated, there is a variation between SVM and RF with the number of training participants affecting average classification accuracy. RF exhibited the highest average overall accuracy with 85.17%, outperforming SVM, which obtained 83.05%. With SVM and RF, increasing

the number of training participants improved classification accuracy considerably, with both having a similar linear increase in accuracy for 6 to 10 participants.
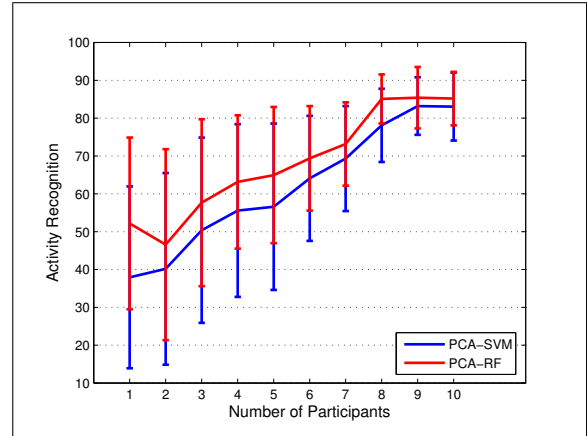

Fig. 2. Activity classification and its standard deviation for classifiers trained from λ

The classification accuracy of several activities fluctuated due to the number of training participants used and the classifier. We observed in Table IV, using two, four and six training participants resulted in low accuracy across the range of activities, the use of eight and ten training participants saw a significant improvement and levelling in accuracy across the range of activities.

The most notable class confusion was observed amongst Walking, Jogging, Upper Body Twist and Arm Movement with misclassification present in Pickup Object.

Due to anatomical similarities between Upper Body Twist and Arm Movement, class confusion was observed, both SVM and RF over classifying Upper Body Twist. As observed in Table IV, classification accuracy between the aforementioned activities, there is a difference for two, four and six training participants. SVM and RF use a probability technique to determine the class of each frame, Fig. 3 demonstrates the class probability for Arm Movement. We observe (Fig. 3) that RF and SVM both experience confusion at different points in the activity sequence for Arm Movement, this inconsistency is observed throughout the study. However, for eight and ten training participants the classifiers stabilised and provided correct classification, with SVM producing the highest accuracy from all results for Arm Movement with 92.25%.

Further confusion between Walking and Jogging was encountered due to the similarity in limb rotation and movement. As observed in Table IV, classification accuracy for the aforementioned activities with SVM and RF had a notable difference in classification rate and confusion. For each λ increase SVM struggled to classify Walking correctly, with an over confidence in Jogging observed. In addition, RF observed a similar over confidence in Jogging, however for both SVM and RF when ten training participants were used, we observed a harmonising of both Walking and Jogging.

Finally, the Pickup Object activity suffered consistent mis-

classification throughout our experiments. Due to the similarity and overlap of posture with a number of other activities, we observed misclassification for two, four and six training participants. Furthermore, even for eight and ten training participants classification accuracy for SVM and RF was below the average classification accuracy, with misclassification not being reduced with increase in the training set.
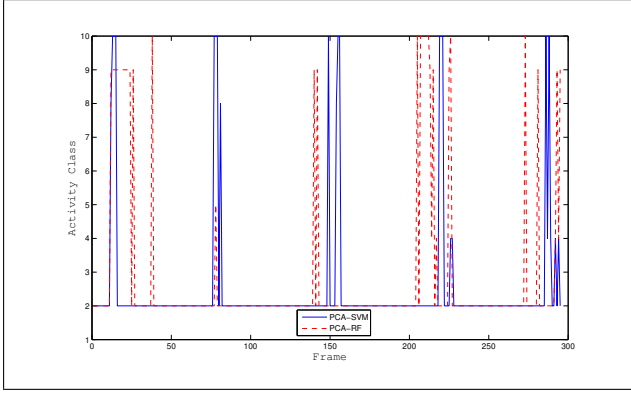

Fig. 3. An example class estimate for SVM and RF by a participant performing Arm Movement. Expected class is 2

To conclude, RF provided the highest average classification accuracy for each λ increase in dataset size when compared with SVM. Conversely, SVM provided improved classification results for a few activities, namely, Arm Movements, Pickup Object, Standing to Seated and Upper Body Twist, with Arm Movements seeing a 7.02% improvement on the result when compared to RF.

B. Model Training and Classification Time

TABLE II
MODEL TRAINING TIME FOR MODELS TRAINED FROM λ

λ participants/ Classifier	2	4	6	8	10
	Training time [Sec]				
SVM	2.381	14.69	30.39	42.24	55.96
RF	5.08	19.35	49.06	79.21	123.17

SVM overall, λ increase was the quickest to train, while RF took considerably longer to train, as demonstrated in Table II. The training time for each classifier was affected by the number of training participants and parameters selected (Table I). Parameter optimisation time was not factored into model training time. With the increase of training participants, we observed the training of the classifiers becomes more complex; consequently the training time of the classifiers tends to increase exponentially.

Average classification time was significantly reduced to millisecond predictions, compared with training time, with on average RF performing faster than SVM as demonstrated in Table III. SVM, as with training, predicting classification was computationally more expensive than RF, however the

TABLE III
CLASSIFICATION TIME FOR MODELS TRAINED FROM λ

λ participants/ Classifier	2	4	6	8	10
	Classification time [Sec]				
SVM	0.036	0.07	0.114	0.146	0.179
RF	0.007	0.022	0.031	0.034	0.037

duration for classification is directly linked to the n_{tree}, C, γ. The average classification time of each activity also increased with the introduction of new participants, albeit on a lower scale (100 up to 300 milliseconds) as demonstrated in Table III. Classification times were calculated by the average of the ten testing participants and on all activities.

V. DISCUSSION

In this study, we analysed SVM and RF on their ability to predict and correctly classify physical rehabilitation activities. The study found that by using 8 to 10 participants in the training set stable classification accuracy for both SVM and RF was achieved. In addition, results presented in this paper illustrate that RF offers performance advantages compared to SVM in the classification of activities based on our 3D kinematic dataset. As in [10], we also note that there has to be an understanding of the anatomical differences between training and testing participants which could influence the result if no prior normalisation is undertaken.

We have found that normalising with the "hip-centre" joint of the first frame, before computing $y_{n,i}, v_{y(n,i)}, e_{n,i}$ and applying PCA reduces anatomical differences and aids in improved classification accuracy (Fig. 1 & Table III). In addition, the computed $y_{n,i}, v_{y(n,i)}, e_{n,i}$ is high-dimensional, containing twenty body joints and while SVM and RF are capable of handling high-dimensional data, feature space reduction by PCA has aided in providing higher accuracy results, improved model training and classification times.

The results reveal that both SVM and RF are suited towards classification of kinematics, although RF was capable of classification significantly faster than SVM. RF was computationally more expensive when training, with a considerable difference versus SVM. RF struggled to detect continuous movement such as Arm Movement resulting in a high misclassification rate (Fig. 3), if misclassification of continuous movement can be overcome it could potentially increase accuracy.

The average SD for achieved accuracies reduced as further training participants were introduced (Fig. 2), suggesting harmonisation between the training and testing datasets. Finally, we have found that parameter selection, both in terms of C, γ for SVM and n_{tree} for RF has an impact on training and classification time due to the introduction of computational complexity (Table I, Table II). For SVM, C affected the number of support vectors, leading to an increase in classification time, whereas RF, n_{tree} increased model training times when a high number of decision trees needed generating.

The Kinect has sensitivity of joint rotation, with the Kinect designed to detect activities of a participant who is standing face forward to the Kinect, large rotation could prove difficult

TABLE IV
AVERAGE CLASSIFICATION ACCURACY PER ACTIVITY FOR λ FOR RF AND SVM

λ /	2		4		6		8		10	
Overall Accuracy [%]	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF
Jumping	19.23	36.19	41.97	55.22	60.14	69.49	76.56	85.24	85.07	86.45
Arms Movement	41.03	48.64	66.42	68.64	62.67	73.93	80.5	85.76	92.25	85.23
Pickup Object	36.64	55.79	39.61	58.47	58.25	57.93	74.85	79.71	78.47	76.24
Squats	29.45	26.88	34.36	46.08	57.92	55.09	78.86	85.92	82.22	84.03
Walking	8	13.37	40.96	54.99	59.75	68.1	74.16	84.3	79.59	86.65
Jogging	42.2	61.16	73.08	78.81	70.48	79.12	78.41	87.42	80.95	85.84
Bending to Toes	49.35	42.48	64.84	63.61	67.37	71	77.59	86.02	81.13	85.86
Standing to Seated	46.86	41.97	59.03	59.75	64.25	69.19	76.57	83.09	85.15	84.58
Upper Body Twist	74.33	78.18	82.02	79.77	80.13	81.62	81.58	84.09	87.71	87.55
Arm Stretch	54.64	61.03	55.18	66.17	80.06	85.04	90.03	91.02	89.16	89.32
Average [%]:	40.17	46.57	55.75	63.15	64.09	69.38	78.1	85.09	83.05	85.17

to detect. Classification between the range of activities was reliable, even rotation and subtle posture changes between similar activities of Walking and Jogging could be classified accurately. The Standing to Seated activity presented a further challenge, due to occlusion of the chair and natural limb movement, yet both classifiers were capable of achieving acceptable classification results. Nevertheless, misclassification was an issue for a number of activities that have similar movements, with both classifiers finding it difficult to classify individual frames without any information about past frames. In our study, we found that by increasing the number of participants enabled an enhanced representative tolerance of the variations in anatomical joints and resulted in stable classification results.

VI. CONCLUSION

In this paper, we introduce SVM and RF to classify human activities by use of projected Kinect kinematic data. We utilise PCA to reduce the dimensionality, demonstrating the potential improvement to classification accuracy. We show that even with a small training set the classification results produced by incrementally increasing the number of participants results in steady increase classification accuracy. We conclude that with eight and ten training participants, the classifiers produces a steady average classification results of 80% and greater.

Our study has demonstrated that transposing Kinect kinematics to a lower-dimensional space and training the classifiers on 8 to 10 participants that both SVM and RF can be reliable for accurate classification. Between them, RF was overall more accurate, but more expensive in terms of model training, while SVM was more expensive in terms of classification rate performance. Finally, the results revealed that RF would be suited to our domain, with a three-fold increase in classification rate when compared with SVM. To conclude, our study has presented a number of potential uses for using RF with the Kinect and 3D output for use in *real-time* classification of activities and gestures related to physical rehabilitation.

REFERENCES

[1] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 16:1–16:43, April 2011.

[2] J. Shotton, A. Fitzgibbon, 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 *CVPR*, 20 - 25 June 2011, pp. 1297 – 1304.

[3] S. Obdrzalek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel, "Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population," in *Engineering in Medicine and Biology Society*, San Diego, California, August 2012, pp. 1188 – 1193.

[4] 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.

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

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

[7] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, June 2010.

[8] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473 – 1488, 2008.

[9] 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.

[10] S. Bhattacharya, B. Czejdo, and N. Perez, "Gesture classification with machine learning using kinect sensor data," in *Emerging Applications of Information Technology*, December 2012, pp. 348–351.

[11] L. Zhang, J. Hsieh, T. Ting, Y. Huang, Y. Ho, and L. Ku, "A kinect based golf swing score and grade system using gmm and svm," in *CISP*, 16-18 Oct 2012, pp. 711–715.

[12] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using kinect camera," in *JCSSE*, June 2012, pp. 28–32.

[13] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognition*, vol. 44, no. 2, pp. 330–349, Feb 2010.

[14] I. Nitze, U. Schulthess, and H. Asche, "Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification," in *GEOBIA*, Rio de Janeiro, 2012, pp. 35 – 40.

[15] A. Statnikov, L. Wang, and C. F. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC Infomatics*, vol. 9, June 2008.

[16] Y. Tang, S. Krasser, Y. He, W. Yang, and D. Alperovitch, "Support vector machines and random forests modeling for spam senders behavior analysis," in *GLOBECOM*, 2008, pp. 2174–2178.

[17] C. Hsu, C. Chang, and C. Lin. (2010, April) A practical guide to support vector classification. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/papers/guide/guide.pdf

[18] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[19] A. Jaiantilal, "Classification and regression by randomforest for matlab," Available at http://code.google.com/p/randomforest-matlab, 2009.

[20] Microsoft Coporation. (2012) Kinect for windows software development kit. [Online]. Available: http://www.microsoft.com/en-us/kinectforwindows