The Use of Inertial Sensors for the Classification of Rehabilitation Exercises

Oonagh Giggins, Kevin T. Sweeney and Brian Caulfield, *IEEE Member*

***Abstract*— The benefits of exercise in rehabilitation after orthopaedic surgery or following a musculoskeletal injury has been widely established. Within a hospital or clinical environment, adherence levels to rehabilitation exercise programs are high due to the supervision of the patient during the rehabilitation process. However, adherence levels drop significantly when patients are asked to perform the program at home. This paper describes the use of simple inertial sensors for the purpose of developing a biofeedback system to monitor adherence to rehabilitation programs. The results show that a single sensor can accurately distinguish between seven commonly prescribed rehabilitation exercises with accuracies between 93% and 95%. Results also show that the use of multiple sensor units does not significantly improve results therefore suggesting that a single sensor unit can be used as an input to an exercise biofeedback system.**

I. Introduction

Exercise therapy is recommended and widely used in the treatment of a number of musculoskeletal and orthopaedic conditions as it has been shown to improve a patient's ability to return to full function [1]. Within the hospital or clinical environment, patients are closely supervised and consistently guided through their rehabilitation by their physical therapist/rehabilitation specialist. However, with the high number of people suffering from lower limb ailments (the OECD currently estimates that 10% of men and 18% of women aged over 60 years have symptomatic osteoarthritis [2]) and the increasing aging demographic there is a requirement for much of the therapy which is conducted within the hospital environment to now be performed externally.

Home-based exercise therapy is becoming more frequent, as there is a demand for a more efficient delivery of healthcare. Unlike performing therapy within a clinical environment, there is a considerable increase in commitment required from the patient for the home exercise therapy to be successful due to the lack of support. Due to this additional commitment, many patients do not fully adhere to their prescribed program of exercise [3]. In addition to poor adherence to an exercise program, without the supervision of their therapist, many patients perform their exercises incorrectly – with poor biomechanical alignment and errors in the rate, rhythm and range of movement. Accurate assessments of exercise performance and adherence to exercise programs are required to optimise the home exercise experience. There is therefore an urgent requirement for an exercise feedback system which could encourage the patient to continue with the program over time.

Many different measurement tools have been previously employed in clinical practice to deliver biofeedback to patients as they exercise, e.g. force plates, electromyography, and optical motion capture systems; however these systems are restricted to a clinical or laboratory environment. The use of inertial sensor based biofeedback platforms could provide an alternative as they are low cost, easy to use and are ubiquitous.

Previous work has evaluated the use of multiple inertial sensors to evaluate exercise quality [4-7]. However, all of the previous research employs multiple sensor units, therefore reducing their ability to be used in an in-home environment. Minimising the number of sensors used both reduces the cost and makes the platform more user friendly which is extremely important for older adults who may use this technology. Previous work has also attempted to classify between good and bad movements for particular exercises [5, 10], however this work is novel as it examines the use of machine learning techniques to classify between different performed exercises.

The work carried out in this paper aims at providing a validation for a novel feedback system by answering two questions. Firstly, can inertial sensor units be used to accurately classify between seven different commonly implemented lower limb exercises and secondly, can a single sensor provide similar results to those provided by multiple sensor setups. Results are compared between the use of one, two and three sensors combined.

II. Methodology

This section outlines the methodology employed to collect, manage and analyse the data gathered in this study. This study is a cross-sectional analytical study. The protocol of this study was approved by the Human Research Ethics Committee in University College Dublin.

A. Participants

Probability sampling techniques were not possible in this study therefore a sample of convenience of suitable participants was selected for this study from a local physiotherapy clinic. Male or female patients who were attending the clinic, aged between forty and eighty years, and who had performed or were performing lower limb exercises for a musculoskeletal/orthopaedic condition or

Research supported by Enterprise Ireland under the Commercialisation Fund.

O. Giggins is with CLARITY centre, University College Dublin, Dublin Ireland (e-mail: Oonagh.giggins@ucdconnect.ie).

K. T. Sweeney and B.Caulfield are with the INSIGHT centre, University College Dublin, Dublin, Ireland (e-mail: kevin.sweeney@ucd.ie, brian.caulfield@insight-centre.org).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

injury were included in this study. The following exclusion criteria applied in this study; lower limb injuries that would limit ability to perform the study exercises, poor functional balance or mobility, any other medical condition that would limit ability to participate in exercise, and cognitive or language difficulties. Fifty-eight participants (19 male, 39 female, age: 53.9±8.5 years, height: 1.69±0.08 m, weight: 74.3±13.0 kg) took part in this investigation. Clinical information regarding the presenting condition or injury of the participants in the study is outlined in Table 1.

Table 1- Clinical information regarding the presenting condition of the study participants

Condition	N
Osteoarthritis of the knee joint	14
Osteoarthritis of the hip joint	9
Osteoarthritis of the knee and hip joint	4
Post meniscectomy	3
Knee ligament injury	4
Instability of knee joint	4
Non-specific low back pain	18
Unknown	2

B. Exercises

The experimental protocol consisted of seven different lower limb rehabilitation exercises commonly prescribed by physiotherapists following a knee or hip injury or surgery. These exercises were adapted from the Total Hip and Knee Replacement Exercise Guides of the American Academy of Orthopaedic Surgeons [8, 9]. The seven exercises studied were the heel slide (HS), the hip abduction (HA), the hip extension (HE), the hip flexion (HF), the inner range quadriceps (IRQ), the knee extension (KE) and the straight leg raise (SLR) exercises and are fully described by Giggins *et al.* in [10].

C. Experimental Procedure

Participants were required to attend a physiotherapy clinic for a once off measurement session. These measurements were performed by the same investigator for each participant. Participants were instructed to wear loose comfortable exercise attire during the measurement session to allow placement of the testing apparatus and to allow for free unrestricted movements during the exercises. Once informed consent had been obtained, demographic data including age, gender, body weight and height were gathered by self-report as well as a brief history of the presenting condition. The exercises were performed using the participants' affected limb. Where there was a bilateral lower limb injury/condition, as in cases with bilateral knee or hip osteoarthritis, the exercises were performed using the more affected side, provided the participant was comfortable to do so. A screening questionnaire was used to ensure that each participant was suitable for inclusion in this study and to perform the required exercises.

Participants each performed ten repetitions of each of the seven studied lower limb exercises. Three of these exercises were performed in standing (HA, HE and HF), one exercise was performed in a seated position on a standardised chair (KE) and three exercises were performed while lying supine on a plinth (HS, SLR and IRQ). Participants were given standardised verbal instructions and a demonstration by the investigator on how to perform each exercise correctly. The order in which the exercises were performed was not randomised. If the individual requested it, participants were allowed a practice trial of each exercise before the test performance, allowing for greater clarity. However, no feedback on their performance was offered during the practice trial, except in cases where an extremely erroneous movement was observed. Likewise, during the test performance of each exercise no feedback was given to participants about their performance.

Three inertial sensors units (Wireless 9DoF IMU Sensor, Shimmer, Dublin, Ireland) were secured to the leg that was being exercised for data collection; one on the thigh (T), one on the shin (S), and one on the foot (F) (Figure 1). The inertial sensors on the thigh and the shin were secured using a neoprene strap, which contained a pouch to house the sensor, while the foot sensor was secured using athletic tape. The orientation and positioning of each sensor was kept consistent across all measurement sessions. With a dimension of 5.3 cm x 3.2 cm x 1.5 cm and weight of 15 grams, these inertial sensors are unobtrusive, permitting unhindered subject movement.

Each of the employed sensors contained both a tri-axial accelerometer and a tri-axial gyroscope sampling at 100 Hz. The Shimmer 9DOF Calibration Application v1.0 (Shimmer, Dublin, Ireland) was used to calibrate the accelerometer and gyroscope sensors of each sensor unit prior to the start of data collection each day. The Multi Shimmer Sync application for Windows (Shimmer, Dublin, Ireland) was used to capture synchronised inertial sensor data over Bluetooth from the three sensors during each of the exercises. The raw inertial sensor data captured were saved onto the PC, as well as on an external hard-drive.

The investigator, who is a chartered physiotherapist, observed each participant as they performed each repetition of each exercise. The investigator evaluated performance during each exercise using a rating scale that was developed beforehand.

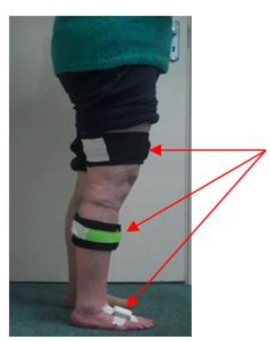

Figure 1 – Sensor position on affected limb

D. Data Analysis

This section will first detail the required pre-processing steps performed on the acquired data, followed by details of the feature extraction and selection techniques performed prior to classification. As stated previously, each of the sensors were first calibrated [12] and synced prior to initiating recording. This ensured that all sensors outputs could be compared post recording. Following data recording and labeling by the onsite investigator, post analysis was performed using MATLAB (2012, The MathWorks, Natwick, USA). For each sensor, six distinct signals were available; namely acceleration X, Y and Z, and gyroscope X, Y and Z. To allow for cross subject analysis, the accelerometer X signal and the gyroscope Y and Z signals were first inverted, prior to further processing, if the subject performed the exercises using their left leg. Following this inversion, three additional signals were calculated; namely overall acceleration magnitude, pitch and roll. Pitch and roll were calculated using a Kalman filter [13]. This filter calculated the orientation using information from the gyroscope signals and used the accelerometer signals to correct for drift in the signal. The nine available signals were then filtered using a 4^{th} order low-pass Butterworth filter with a cut-off frequency of 20 Hz.

To allow for the classification of the exercises, a number of features were first extracted from the nine available signals for each individual trial for all subjects. A number of both time and frequency domain features were extracted to represent the signals. These features were namely signal mean, standard deviation, skewness, kurtosis, signal energy, level crossing rate, signal range, 25^{th} percentile, 75^{th} percentile and the variance of the wavelet coefficients using the Daubechies 5 mother wavelet to level 6. This resulted in 14 features for each of the nine available signals producing a total of 126 features per sensor unit.

Although each of these features could be useful to represent the data, it is not good practice to employ a large number of features when only a small number of trials are available as, by doing so, it is possible to over-fit the model, producing very good classification results during training but significantly poorer results during testing. In order to reduce the number of employed features principle component analysis (PCA) was performed [11]. PCA converts the set of features from a 126 dimensional matrix, with possibly correlated variables, into a set of principle components which are linearly uncorrelated. During analysis, the components which accounted for 99% of the variance were selected as the features. However, these new "features" no longer have any physical meaning (such as max, min etc.). It should also be noted that this process of feature selection using PCA is only performed on the training data, with the test data remaining unseen to the setup to again refrain from biasing the system. The test data is reduced using the coefficients found using the training data.

Classification was performed using leave-one-subject-out-cross-validation (LOSOCV) and the simple logistic regression classifier. One classifier was trained for each of the exercises using a one-vs-all approach, i.e. HS vs non-HS, HF vs non-HF etc., resulting in a bank of seven separate classifiers. For each fold of the cross validation, once the classifiers have been trained, they are presented with the data from the test subject. For each individual test trial, the classifier which outputs the highest probability is then chosen as the determined exercise for that trial. The use of LOSOCV ensures that there is no biasing of the classifiers, in such that the test subjects data is completely unseen by the classifier prior to testing. Previous research by Taylor *et al.* [5] has shown that not employing this method of testing can skew results by up to 21%.

III. RESULTS

One study participant only performed the three exercises in supine lying due to time constraints, three subjects were not able to perform the SLR exercise and data were lost for one participant during the HS exercise and for another during the KE exercise due to sensor failure. This resulted in a total of 570 trials for the HS exercise and the three exercises in standing, 550 trials for the SLR and IRQ exercises, and 560 trials for the KE exercise.

The results of the paper are presented in Table 2 and 3. Table presents the sensitivity and specificity scores for each of the individual sensors as well as all combinations of multiple sensors. F=Foot, S=Shin, T=Thigh, HS=heel slide, HA=hip abduction, HE=hip extension, HF=hip flexion, IRQ=inner range quadriceps, KE=knee extension, SLR= straight leg raise.

Table then presents the overall accuracy values obtained using each of the sensor combinations. Accuracy measures the overall effectiveness of a classifier and is computed by taking the ratio of correctly classified examples and the total number of examples available. Sensitivity measures the effectiveness of a classifier at identifying a desired label, while specificity measures the classifiers ability to detect negative labels [5].

Table 2- Classification results: Sensitivity (Sens) and Specificity (Spec) scores across all exercises. Looking at all possible sensor combinations.

		HS	HA	HE	HF	IRQ	KE	SLR
All	**Sens**	0.97	0.92	0.92	0.91	0.95	0.97	0.95
	Spec	0.99	0.99	0.98	0.99	0.99	0.99	0.99
F&S	**Sens**	0.97	0.94	0.92	0.84	0.96	0.89	0.92
	Spec	1.00	0.99	0.98	0.97	0.99	0.98	0.99
S&T	**Sens**	0.95	0.94	0.96	0.97	0.99	0.97	0.96
	Spec	0.99	1.00	0.99	0.99	1.00	1.00	0.99
F&T	**Sens**	0.93	0.95	0.94	0.94	0.99	0.97	0.95
	Spec	0.99	0.99	0.98	0.99	1.00	1.00	0.99
F	**Sens**	0.95	0.94	0.89	0.84	0.94	0.93	0.95
	Spec	0.99	0.99	0.98	0.98	0.99	0.98	0.99
S	**Sens**	0.97	0.95	0.95	0.88	0.99	0.94	0.93
	Spec	1.00	0.99	0.99	0.99	0.99	0.98	0.99
T	**Sens**	0.88	0.98	0.95	0.94	0.99	0.98	0.78
	Spec	0.97	1.00	0.98	1.00	1.00	1.00	0.98

F=Foot, S=Shin, T=Thigh, HS=heel slide, HA=hip abduction, HE=hip extension, HF=hip flexion, IRQ=inner range quadriceps, KE=knee extension, SLR= straight leg raise.

Table 3 - Accuracy (Acc) Results: Overall accuracy using the different sensor combinations

	All	F&S	S&T	F&T	F	S	T
Acc	0.94	0.92	0.96	0.95	0.95	0.95	0.93

F=Foot, S=Shin, T=Thigh.

IV. Discussion

It can be observed from the tables of results that the employed classification protocol is very efficient at separating the various investigated exercises when only a single sensor is employed, with the accuracy score only varying between 0.93 and 0.95. However, when the sensitivity results are also accounted for it can be seen that the various sensor positions provide higher classification results for different exercises. For example, when employing the foot sensor, it can be seen that the classifier has some trouble in separating the HF exercise and the KE exercise. This finding is as expected however due to the similar nature of the two exercises when examined from the perspective of the foot. However, the thigh sensor has no such problem with these two particular exercises as they have significantly different movement patterns when examined from the thigh. It does however have minor difficulties separating the HS and the SLR exercise due again to the similar morphologies of the signals. When employing the shank sensor, the exercises with the lowest classification accuracies are the HF and the SLR. Therefore, depending on the exercises to be classified, the optimal position of the sensors may change between the three examined positions. These results answer the question as to whether a single inertial sensor is able to accurately distinguish between seven commonly performed lower limb rehabilitation exercises.

In order to determine if an increase in the number of sensors employed had a significant effect on the classification results, all combinations of sensors were also tested. It was found that, with the inclusion of the additional sensors, the minimum sensitivity observed across all exercises generally increased, but the overall average sensitivity score did not change dramatically. Similarly, the accuracy of the system did not change significantly. Therefore is can be stated that an increase in the number of employed sensors does not improve the classification results significantly enough to warrant the use of more than a single sensor.

V. Conclusion

With the changing demographics of the world's population, the number of patients undergoing exercise rehabilitation continues to increase. As a consequence home-based exercise therapy is becoming more frequent, with the demand for a more efficient delivery of healthcare.

This paper endeavored to determine if simple inertial sensors could be employed to accurately distinguish between seven commonly implemented rehabilitation exercises and whether the addition of additional sensors would significantly improve results.

Results have shown that the inertial sensors are capable of classifying between the analysed exercises with a high level of accuracy and they also support the hypothesis that the addition of extra sensor units does not significantly improve results. These findings therefore prompt the development of a simple biofeedback system using a single inertial sensor for use in rehabilitation to monitor adherence to exercise programs in the home. A single sensor approach is desirable as not only does it reduce the cost of the system but also avoids cumbersome set up and calibration procedures. Using a single sensor is also desirable as many mobile phones nowadays are equipped with inertial sensor technology, which means a mobile phone could be used as an input to a biofeedback system. Future work will examine the use of these inertial sensors for the classification of various commonly occurring deviations observed while performing the seven rehabilitation exercises as well as the development of a biofeedback system using the sensors in a smart phone.

Acknowledgment

The authors would like to thank "The Physio Company" for their help in recruiting the patients for the current study.

References

[1] Gilbey HJ, Ackland TR, Wang AW, Morton AR, Trouchet T, Tapper J (2003). Exercise Improves Early Functional Recovery After Total Hip Arthroplasty. Clinical Orthopaedics and Related Research 408

[2] OECD (2011) Health at a Glance 2011. OECD Publishing, URL: http://tinyurl.com/6ofnas6.

[3] Doyle J, Kelly D, Caulfield B (2011). Design Considerations in Therapeutic Exergaming. Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare pp 389–393.

[4] Crowell HP, Milner CE, Hamill J, Davis IS (2010). Reducing impact loading during running with the use of real-time visual feedback. The Journal of orthopaedic and sports physical therapy 40(4):206–13.

[5] Taylor PE, Almeida GJM, Kanade T, Hodgins JK (2010). Classifying human motion quality for knee osteoarthritis using accelerometers. Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2010:339–43.

[6] Davis JR, Carpenter MG, Tschanz R, Meyes S, Debrunner D, Burger J, Allum JHJ (2010). Trunk sway reductions in young and older adults using multi-modal biofeedback. Gait & posture 31(4):465–72.

[7] Verhoeff LL, Horlings CGC, Janssen LJF, Bridenbaugh SA, Allum JHJ (2009). Effects of biofeedback on trunk sway during dual tasking in the healthy young and elderly. Gait & posture 30(1):76–81.

[8] American Academy of Orthopaedic Surgeons (2013) Total Hip Replacement Exercise Guide. Available from: http://orthoinfo.aaos.org/topic.cfm?topic=a00303.

[9] The American Academy of Orthopaedic Surgeons. Total Knee Replacement Exercise Guide 2011. Available from: http://orthoinfo.org/topic.cfm?topic=A00301.

[10] Giggins O, Kelly D, Caulfield B (2013). Evaluating Re-habilitation Exercise Performance using a Single Inertial Measurement Unit. In: 7th International Conference on Pervasive Computing Technologies for Healthcare.

[11] Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23(19): 2507-2517.

[12] Ferraris F, Grimaldi U, Parvis M (1995). Procedure for Effortless In-Field Calibration of Three-Axis Rate Gyros and Accelerometers. Sensors and Materials 7(5): 311–330.

[13] Luinge H J, Veltink P H (2005). Measuring orientation of human body segments using miniature gyroscopes and accelerometers. Medical Biology Engineering and Computing 43(2): 273-282.