Automated Analysis and Quantification of Human Mobility using a Depth Sensor

Daniel Leightley*, *Member, IEEE,* Jamie S. McPhee and Moi Hoon Yap, *Member, IEEE*

***Abstract*—Analysis and quantification of human motion to support clinicians in the decision-making process is the desired outcome for many clinical-based approaches. However, generating statistical models that are free from human interpretation and yet representative is a difficult task. In this work, we propose a framework that automatically recognises and evaluates human mobility impairments using the Microsoft Kinect One depth sensor. The framework is composed of two parts. Firstly, it recognises motions, such as sit-to-stand or walking 4 metres, using abstract feature representation techniques and machine learning. Secondly, evaluation of the motion sequence in the temporal domain by comparing the test participant with a statistical mobility model, generated from tracking movements of healthy people. To complement the framework, we propose an automatic method to enable a fairer, unbiased approach to label motion capture data. Finally, we demonstrate the ability of the framework to recognise and provide clinically relevant feedback to highlight mobility concerns, hence providing a route towards stratified rehabilitation pathways and clinician led interventions.**

***Index Terms*—depth sensor, human motion, mobility, motion quantification, human action recognition**

I. Introduction

THERE is a clear advantage to developing automated systems to detect human motion for applications associated with healthcare [1]. The general population is living longer, therefore a factor in providing health and social care services is to quantify and continuously assess people's health. While many people remain healthy, active and engaged into later life, studies have indicated that a minority suffer from frailty and musculo-skeletal mobility disorders [2]. Frailty is not a single disease; but a combination of the natural ageing processes, during which neuromuscular systems decline, and the accumulation of medical conditions, which leaves a patient vulnerable to illness or trips and falls [3]. Frailty is an indicator of general health and well-being, and is usually assessed by asking the person to perform several standardised tests (*e.g.* walk back and forth, sit-to-stand) during which a clinician observes the activity for stability, duration, coordination and posture control.

Daniel Leightley was with the School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, England. He is now at the Centre for Military Health Research, King's College, London. e-mail: dleightley@ieee.org. *Corresponding author.

Jamie S. McPhee leads the Neuromuscluar and Skeletal Ageing Research Group with the School of Healthcare Science, Manchester Metropolitan University, England.

Moi Hoon Yap is with the School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, England.

Manuscript received XX X, X; revised X X, XX.

Although direct clinical assessment is vital, there is a need to construct more efficient clinical approaches to address limitations in assessment [3]–[5]. First, clinician-led assessments are dependent upon the skills, experiences and opinion of the clinician, so are not always objective. Second, clinical assessments are open to subjective bias and variance between assessments as well as interpretation of the results. Third, the entire process can be time consuming considering the patients' need to attend the appointment, undertake the assessment and the need for clinics to arrange appointments and oversee the assessments. Fourth, people with mobility impairments increase their risk of further trauma by having to attend specialist clinics, so it would be preferable to undertake the assessment at home. Fifth, patients may exhibit different behaviour as a result of the examination, which may alter the outcome and perceptions by the clinician.

Several studies [6], [7] indicate that reliable detection of a person with "poor mobility" can predict future declines in health, early identification could enable earlier remedial clinician-led interventions [2], [8]. Several attempts have been made to develop home-based monitoring systems for assessment and rehabilitation [7], [9]. While these systems have been validated and have the potential to support home-based monitoring of mobility, they fall short of assessing for mobility impairments. Further, these systems provide a single health indicator whereas an in-depth descriptive indicator could prove more useful to a clinician. In addition, existing systems have been evaluated using game-orientated datasets, and without clinical validation [5], [6], [10]. We utilise the newly released K3Da dataset [11], which consists of clinically relevant motions captured using the Kinect One sensor to validate the proposed framework.

In this work, we propose a reliable and non-invasive framework to recognise, assess and quantify the mobility of participant's. The system acquires motion capture (MoCap) from a single depth sensor, where the skeletal stream is decomposed into a set of novel joint-group features. Analysis techniques are employed to provide joint-group feedback highlighting the state of mobility, hence providing detailed insight for clinicians. To identify and label the MoCap data, we propose a method for automating the ground truth labelling of MoCap that is free from human bias or interpretation.

This paper is structured as follows: Section II presents and discusses current literature. Section III describes our proposed method to recognising, decomposing and analysing human motion. Section IV demonstrates the effectiveness of our method for classifying and analysing human motion. Finally, Section V provides a discussion, conclusions to our work and

presents future avenues of research.

II. Related Work

Several works have been proposed to measure and analyse human motion and stability. The most prevalent methodology suggests using one or a combination of intrusive sensors, such as body-based accelerometer or markers. Then, the clinical relevant indicators are extracted by analysing the patterns presented in time-series data. In recent years, the computer vision community has proposed an array of solutions to solve this challenge, but majority are not validated clinically. These works have predominantly focused on depth sensor technology, removing the need for intrusive sensors, which has been shown to be sufficiently accurate and responsive for tracking in both in-home and clinical settings [10], [12], [13].

A number of frameworks have been proposed to enable greater understanding of human motion. These frameworks follow a similar structure. They first seek to identify the motion, using recognition frameworks, then undertake quantitative analysis [14], [15]. The Rehabilitation Gaming System [9], [16], [17] extracts the stability of a participant using a combination of gloves and markers on the hands. After the participant performs several standardised tests, the data is processed to calculate compliance. Dolatabadi *et al.* [18] proposed a home-based system for assessing changes in gait and balance. The authors utilise a Microsoft Kinect 360 (Kinect 360) for XBox sensor to observe gait recovery in a participant that had undergone surgery. They were able to track the gait, and temporal change over a number of weeks. Gonzalez *et al.* [19] proposed a solution for *real-time* balance estimation by deriving Centre-of-Mass (CoM) feature from the Kinect 360 and a Wii Balance Board. The authors unite the CoM and angular momentum to quantify the stability of participant. While this work presents a novel solution to providing balance and stability measurements, it has been tested only on two participants.

There are few approaches that seek to unite recognition and quantitative analysis of human motion in a hierarchical context. Cary *et al.* [20] proposed a system to unite the Kinect 360 and Artificial Neural Network (ANN) to aid in recognition of physiotherapy assessment sessions. The authors design a feature vector based on joint groups. The first group is composed of the torso joints; with the second group the remaining joints. The vector is computed by extracting the associated angles between joints. The work employs a multi-level ANN that decomposes each joint into a separate model. This allows the recognition of complex motion sequences and assesses their correctness in relation to a predefined model. Kargar *et al.* [21] utilised a depth sensor to automatically measure the physical mobility of participants. They analyse and classify human gait in relation to "Get-up-and-Go-Test". Two types of features are extracted from the MoCap data provided by the sensor. The first type of feature is related to the human gait (*e.g.* number of steps, duration of each step, and turning duration), whereas the second type describes the anatomical configuration. The authors state that using these features provided a descriptor for characterising physical mobility. To enable classification of the imbalance severity, the authors implement a Support Vector Machine (SVM).

While works exist to assess human balance and gait, the work is limited to extracting single-valued indicators which lack clinical validation. To the best of the authors' knowledge, this work is the first proposed method to detect, analyse and provide clinically relevant outcome measures using depth sensor technology. Our work relies on several key computer vision and pattern recognition techniques: stereo vision, pose estimation, feature representation, clustering and temporal segmentation.

III. Application Framework

The framework is divided into three parts. First, joint-group features are generated from skeletal MoCap data. Second, recognition of human motion using a range of machine learning classifiers to provide the clinician with the motion being observed. Finally, a motion analysis approach to providing clinically relevant outcome measures is proposed.

A. Feature Encoding

Identification and recognition of motions is not a trivial task. Wang *et al.* [6] used feature groups to detect human action with a Kinect 360, which yielded promising results. However, the same feature set would not provide the abstract level of detail required for subtle clinical outcome measures. This, in part is due to the way in which motion performance between humans differs slightly, making a single top-level outcome generalised in nature [13], [22]. Derived features have been shown to be more useful than raw MoCap data [23]–[25]. This led to the work of Du *et al.* [26], using a Hierarchical Recurrent Neural Network for human action recognition, the concept of dividing the skeleton into joint groups, based on anatomical significance to the motion sequence. A joint level group of features is encoded, representative of multiple motion types and shows promise in encoding subtle variations. As in [26], we employ the same joint group decomposition. A summary of the feature groups and encoding methodology is presented in Table I, with a visual illustration presented in Figure 1.

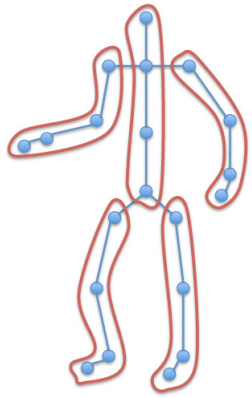


Fig. 1. An illustration of a MoCap skeleton divided into five joint groups as proposed in [26]. Each joint group represents motions to allow for representation of all types of human motion.

TABLE I
SUMMARY OF JOINT DECOMPOSITION AND DERIVED FEATURES FOR EACH GROUP AND THE CORRESPONDING DIMENSIONALITY OF THE FINAL FEATURE VECTOR. *Where I is the number of features*

Joint Group	Features	Length	Notation*
Left Arm ($LeftShoulder$, $LeftElbow$, $LeftWrist$, $LeftHand$)	Left arm Euler Angle (between left shoulder and left wrist), Euclidean distance between the left shoulder and left hand, x and y axis vectors.	12	$F_{LeftArm} = \{1 \dots, I\}$
Left Leg ($LeftHip$, $LeftKnee$, $LeftAnkle$, $LeftFoot$)	Left leg Euler Angle (between left hip and left ankle), Euclidean distance between the left hip and left foot, x and y axis vectors.	12	$F_{LeftLeg} = \{1 \dots, I\}$
Right Arm ($RightShoulder$, $RightElbow$, $RightWrist$, $RightHand$)	Right arm Euler Angle (between right shoulder and right wrist), Euclidean distance between the right shoulder and right hand, x and y axis vectors.	12	$F_{RightArm} = \{1 \dots, I\}$
Right Leg ($RightHip$, $RightKnee$, $RightAnkle$, $RightFoot$)	Right leg Euler Angle (between right hip and right ankle), Euclidean distance between the right hip and right foot, x and y axis vectors.	12	$F_{RightLeg} = \{1 \dots, I\}$
Torso ($SpineBase$, $SpineMid$, $Neck$, $Head$, $SpineShoulder$)	Torso Euler Angle (between the spin base and neck) relative to the body, Euclidean distance between the spine base and head, Body lean angle (relative to the floor with torso as a reference), Centre-of-Mass (between left shoulder, right shoulder, spine mid), x and y axis vectors.	15	$F_{Torso} = \{1 \dots, I\}$

There are multiple measurements which are capable of being extracted from a skeletal stream of MoCap [21], [25], [27]. The difficulty is to select the most appropriate features capable of describing the motion and subtle variations. Alongside these derived features, raw MoCap data itself is also utilised. The x and y coordinates are extracted for each joint to describe the posture change with respect to the axis [23].

Euler Angle: Any rigid body can be described as an angle around three orthogonal coordinates in fixed space. However, computing Euler Angles from marker-less MoCap is difficult. Joint angles are extracted using the approach by Lewandowski *et al.* [28]. Frames are normalised, and then three angles defining the reference joint (see Table I) are computed and represented by a three unit angle ($\mathbb{R}^3$).

Euclidean Distance: An important characteristic of human motion is the way in which the participant transitions over time in relation to a fixed point. For example, with the *Torso Group*, the Euclidean distance is computed between the base of the spine and head. While this value will remain constant for motions such as walking, when the participant performs a bend, or sit-to-stand the distance between the two joints differs. The change in distance is computed as:

$$distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (1)$$

where x_1, y_1, z_1 and x_2, y_2, z_2 are the reference joints.

Body Lean Angle: It is possible to represent a rotation group ($\mathbf{SO}(3)$), which is a rotation in Euclidean space as a set of two vectors (unit vector $\hat{\mathbf{e}}$ indicating the direction of lean, θ angle representing the magnitude of rotation about the axis). The Body Lean Angle represents the body in relation to the ground plane, see Figure 2. The angle is computed by the flexion of the spine in relation to ground floor plane, defined as the centre of the feet. The lean angle between the spine and the floor is defined as:

$$\hat{\mathbf{e}}, \theta = \arccos\left(\frac{S \cdot Q}{\|S\|\|Q\|}\right)^t \quad (2)$$

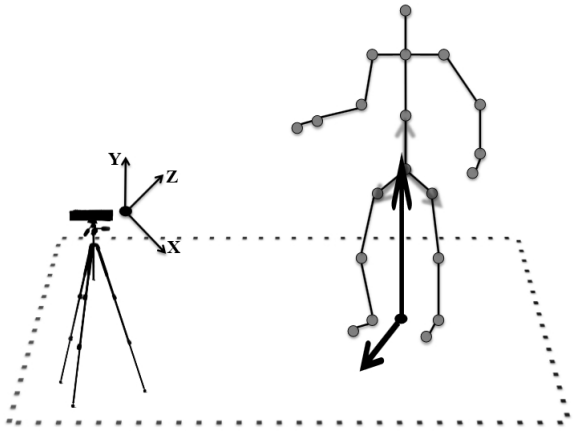


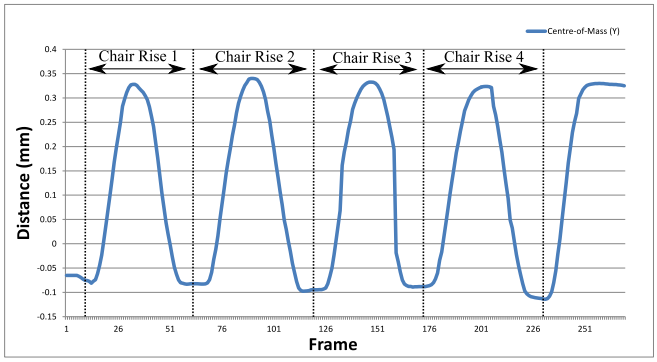
Fig. 2. Visual representation of the Body Lean Angle in relation to the Microsoft Kinect One sensor. The angle is computed by the intersection between the ground plane and spine.

where S^t is the spine vector (x, y, z) and Q^t is the floor vector - the centre of the feet.

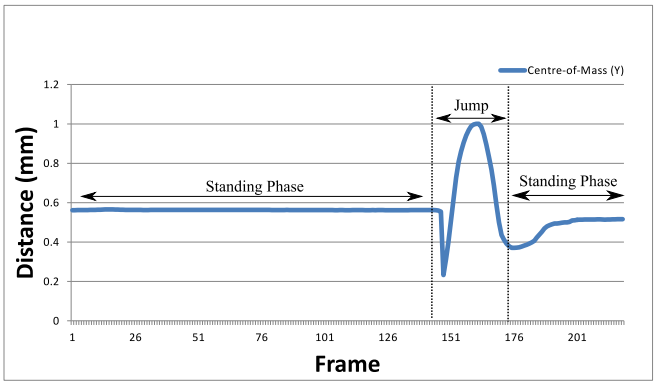
Centre-of-Mass: The CoM [19], [29] is extracted from MoCap data to describe the directional movement. Figure 3 demonstrates a 2D visual example for the CoM for two motion sequences. Vertical directional movements, such as *Chair Rise*, are identified by the Kinect One due to its ability to track millimetre postural changes [30]. In order to evaluate and measure stability, it is necessary to measure the movement of the body's CoM. Let com be the CoM at time t of pose p computed from three joints (*LeftHip*, *RightHip*, *SpineBase*) is given as:

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^{3} p'_{t,x_i}}{3} \\ \bar{y} &= \frac{\sum_{i=1}^{3} p'_{t,y_i}}{3} \\ \bar{z} &= \frac{\sum_{i=1}^{3} p'_{t,z_i}}{3} \\ com &= [\bar{x}, \bar{y}, \bar{z}]^t \end{aligned} \quad (3)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2558540, IEEE Journal of Biomedical and Health Informatics



a) Chair Rise



b) Maximum Jump

Fig. 3. Visual 2D representation of the CoM for two motion types. a) CoM (y axis) for *Chair Rise*. b) CoM (y axis) for *Maximum Jump*.

where $\bar{x}$, $\bar{y}$, $\bar{z}$ is the mean, i is the joint index of frame t and com is the concatenation of the mean values.

B. Recognition: Motion Identification

The framework for recognising trials (otherwise referred to as motions) is shown in Figure 4. Recognition enables the clinician to be aware of the motion performed by the participant. The framework is divided into two aspects. First, offline training of multiple machine learning classifiers based on an exemplar-based pose selection. Second, online detection and identification of motions in real-time.

1) Feature Reduction and Selection: For generalisation and consistency, $\mathcal{P}$ represents all skeletal poses, ordered in a time-sequential manner. Each feature encodes a motion characteristic, such as gait, or motion performance. To train any machine learning classifier, a unified training sample needs to be formed, given as:

$$\hat{F} = \{F_{LeftArm}, F_{LeftLeg}, F_{RightArm}, F_{RightLeg}, F_{Torso}\} \in \mathbb{R}^{63} \quad (4)$$

where $\hat{F}$ is a combined vector consisting of the features derived for each joint group.

The objective is to identify and extract only those features that contain enough descriptive information to describe the motion. A two tier clustering process is employed (Algorithm 1 illustrates the pseudo-code). Top-level process combines the features for each trial into a single matrix. An automated

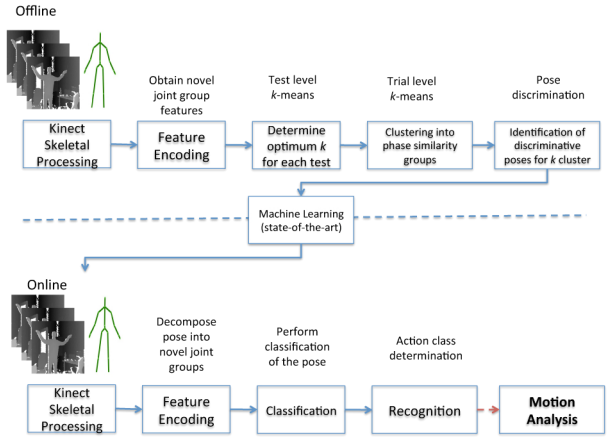


Fig. 4. Recognition Overview: Top row illustrates the training process. Bottom row illustrates the online recognition process.

clustering approach is then employed to identify the optimum number of clusters. k-means has been selected over other approaches as it is computationally faster when dealing with a large number of observations. Further, it is suitable for grouping and segmenting human motion compared hierarchical clustering methods. Then, sub-level clustering is undertaken on each feature group with the derived optimum clusters. This results in a feature set $\hat{F}$, is represented by k clusters. The next stage is to identify and extract key features. The clustering process results in a dynamic number of k clusters

Algorithm 1: Automatic k selection and grouping with k-means clustering

Input: $\mathcal{A}_x = \{\hat{F}_1, \hat{F}_2 \ldots, \hat{F}_N\}$ - training instances for all $\hat{F}_n$ in the motion class in concatenated form
$MaxIt$ - maximum number of convergence iterations

Output: $L = \{l(e)|1, 2, \ldots, E\}$ - set of cluster associate labels for $\hat{F}_n$

For a set of features $x \in \mathcal{A}$ **do**
foreach $k = 2 : N$ **do**
  randomly initialize k centroid location, C_i, for each cluster
  foreach $a_i \in \mathcal{A}$ **do**
    $l(e) \leftarrow argminDist||a_x - c_i||^2, i \in \{1, \ldots, k\}$
  end
  $it \leftarrow 0$
  repeat
    foreach $a_x \in \mathcal{A}$ **do**
      $minDist \leftarrow argminDist||a_x - c_i||^2, i \in \{1, \ldots, k\}$;
      if $minDist \neq l(e)$ **then**
        $l(e_n) \leftarrow minDist$
      end
    end
    $it++$;
  until $it \leq MaxIt$;
  $wcss_k \leftarrow argminDist||\mathcal{A}_{\hat{n}} - C_i||^2, i \in \{1, \ldots, I\}$
end
$est_k = E_I^*\{\log(wcss_k)\} - \log(wcss_k)\}$
then
foreach $\hat{F}_n \in \mathcal{A}$ **do**
  randomly initialize est_k centroid location, C_i, for each cluster
  do classify $\mathcal{F}_n$ samples according to nearest C_i
  recompute C_i
  until no change in C_i
  end
  return cluster identifications for each feature (n)
end

of any length. A "key" cluster is identified if it contains more than N/K, the average size of a cluster. For each key cluster, features are retained using an equivalence function.

The similarity between two features, a and b from $\hat{F}$ is computed as:

$$Similarity(\hat{F}_a, \hat{F}_b) = \min ||a_i - b_j||^2 \quad (5)$$

A Self-Similarity Matrix S for a key cluster KC from $\hat{F}$ can be computed using Eq. 5 and defined as:

$$S := (s_{i,j})_{N_z x N_z} = \{Similarity(\hat{F}_i, \hat{F}_j)\}_{N_z x N_z} \in KC \quad (6)$$

where S is the computed Similarity-Matrix with a dimensionality of $N_z \times N_z$ for cluster KC. The median element of the Similarity-Matrix S_{median} is selected, and a cost function is defined to identify features that are within a threshold, denoted as $hold$ are retained. This is computed as:

$$D(S_{median}, S_i) = hold \nleq \sum_{i=1}^{I} ||S_{median} - S_i||^2 \quad (7)$$

where $i \in \{1, 2, \ldots, I\}$ is the number of poses for each key cluster and D are the features which fall within the threshold $hold$. Hence, only those that are shown to be informative are retained.

2) Recognition: For recognition, several machine learning classifiers are evaluated (Table II). To train a model, each action class is represented by a set of key clusters with fine-tuning and parameter selection undertaken to improve the model stability. To classify, the skeletal stream is encoded in real-time using the features summarised in Table I, each frame is provided to the model for assignment of a class predictor.

C. Motion Analysis

The framework for analysing human mobility is presented in Figure 5. The framework is split into two parts. First, MoCap data is assigned a ground truth marker to identify if it contains "good" or "poor" mobility, then multiple SVMs are trained to detect mobility change. Second, identification and analysis of participant's mobility is undertaken to provide clinically relevant outcome measures. In the recognition stage, the joint groups are merged into a single feature vector, whereas for motion analysis each joint group is trained as a **separate model**.

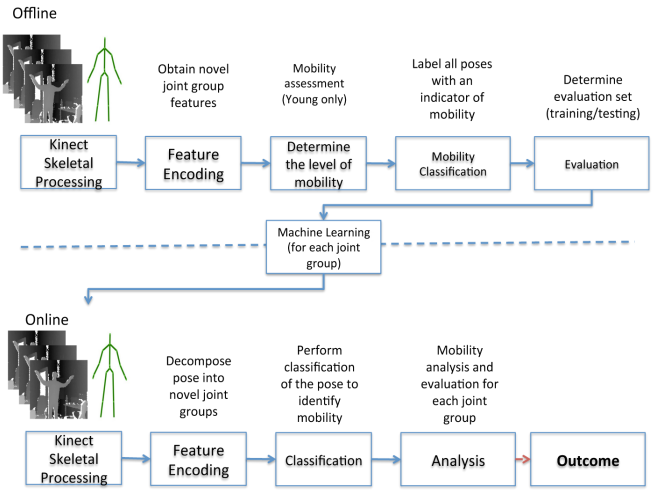


Fig. 5. Motion Analysis Overview: Top row illustrates the process undertaken to label, group and train a set of SVM models. Bottom row illustrates the analysis approach utilised to provide clinically relevant outcome measures.

1) Labelling and Computation of Human Mobility: For a typical recognition task, prior knowledge of the class label is required. This is usually straightforward to determine, for example a person *walking* or *jumping* can easily be defined with a single label [1]. However, the task becomes very difficult to identify and label in the context of different styles of the same motion. For example, attempting to group different

TABLE II
LIST OF MACHINE LEARNING CLASSIFIERS UTILISED, INCLUDING FINE-TUNING/PARAMETER SELECTION METHODOLOGY.

Machine Learning	Parameters Required	Parameters Selection Methodology
Support Vector Machines (SVM) [31]	C and γ	Parameter selection was undertaken using cross-validation.
Random Forests (RF) [32]	n_{tree} and m_{try}	n_{tree} selection is undertaken using cross-validation method iterating between n_{tree} = 100 and n_{tree} = 500 with m_{try} set at 3.
Artificial Neural Networks [32]	n_{layer}, neurons, rate	Parameter selection was undertaken using stratified k-fold cross validation with a validation set employed to present over-training.
Gaussian Restricted Boltzmann Machines (GRBM) [33]	$h_{variable}$	The number of hidden units, $h_{variable}$ was set at a default of 500.
Adaptive Boosting (AB) [34]	$n_{iterations}$	Parameter selection was undertaken using cross-validation.
LPBoost [35]	$n_{iterations}$	Parameter selection was undertaken using cross-validation.
RUSBoost [36]	$n_{iterations}$	Parameter selection was undertaken using cross-validation.
Total Boost (TB) [37]	$n_{iterations}$	Parameter selection was undertaken using cross-validation.
Bagging [38]	n_{tree}	Parameter selection was undertaken using cross-validation.

types of gait manually can result in subjective grouping and bias [18], [39]. There have been several approaches proposed to obtain and measure clinically supportive outcome(s), yet they have manually annotated motions with little clinically supportive reasoning for labelling [6], [13], [39]. However, there are methods for objectively identifying and grouping human motion in the clinical literature. Baumgartner *et al.* [40] introduced a normal distribution of motion values to derive the standard deviation (SD) of the mean to define groups of sarcopenia (loss of muscle mass with ageing). This has been used extensively within the medical community to identify different groups [3], [41], it has not been utilised within the computer science community.

Our proposal extends [40] to define "good" and "poor" mobility using a digitalised labelling framework, based on the data itself. This approach is free from human interpretation, bias or subjectiveness except for providing a threshold value. The labelling can be summarised as follows: Frames that contain a value greater than the $\pm$1.5 SD of the mean are identified having "poor" mobility, whereas frames within $\pm$1.5 SD of the mean are identified having "good" mobility. Each joint group is labelled individually. For each group, each individual frame is labelled as having "good" mobility or "poor" mobility. The labelling is summarised as follows:

1) Using young adults from the K3Da dataset, each motion class and joint group is combined into a single matrix. This results in five matrices representing each joint group.
2) The $\pm$1.5 SD from the mean (computed as the average per feature row) is computed for each joint group.
3) Using the SD of the mean values computed at item 2, all trials including **young and elderly** from the K3Da dataset are labelled.
4) Frames that are within the $\pm$1.5 SD are labelled as having "good" mobility.
5) Frames that are greater than $\pm$1.5 SD are labelled as having "poor" mobility.

The mean $\pm$1.5 SD threshold value is computed from the young only, to represent the general population. As this work seeks to identify mobility, using the elderly may create a bias within the model and increase the rate of false positives. Table III provides a summary for the number of frames labelled as having "good" or "poor" mobility for each participant group.

TABLE III
FRAME LABELS FOR "GOOD MOBILITY" AND "POOR MOBILITY" FOR EACH JOINT GROUP, REPRESENTED FOR EACH PARTICIPANT CATEGORY.

Joint Group	Young		Elderly	
	Good (%)	Poor (%)	Good (%)	Poor (%)
Left Arm	31,382 (87)	4,131 (13)	12,516 (51)	12,851 (49)
Right Arm	32,145 (84)	5,146 (16)	13,728 (54)	11,639 (46)
Left Leg	30,367 (89)	3,368 (11)	18,089 (60)	7,278 (40)
Right Leg	31,344 (87)	4,169 (13)	19,725 (72)	5,642 (28)
Torso	31,355 (87)	4,158 (13)	16,089 (57)	9,278 (43)

The labelling of each frame provides information of the state of mobility at any period of time; A "mobility score" metric is derived to indicate the level of mobility the participant has compared to the statistical models derived earlier. The mobility score is an aggregate of the number of frames identified as having "good" mobility versus "poor" mobility for each joint group.

2) Analysing Mobility using Multiple SVMs: Over recent years, a number of classifiers have been proposed for the task of motion recognition, consistently SVM have yielded high accuracy results. They are computationally less expensive to train, and provide a low latency for recognition compared to others (*e.g.* Random Forest). The aim is to generate a detailed insight of a participant's mobility. A random sample of participant's from the K3Da dataset are extracted, and each joint group is modelled using an SVM with 10-fold cross-validation, Figure 6 demonstrates the training and evaluation pipeline. While it is possible to train a single SVM, indeed [1], [42] obtained high accuracy results for the task of recognition it is not suitable for clinical scenarios. If we follow these type of approaches, subtle motion variations would be overshadowed resulting in over generalisation (over fitting) leading to inter-/intra-class confusion between "good" mobility and "poor" mobility. Training an individual SVM for each joint group models subtle changes in motion, providing a greater motion context. Furthermore, it permits the framework to identify specific joint groups that may be of concern.

To compute a clinical outcome measure, test data is decomposed into joint-group based features and provided to the corresponding SVM. Each SVM provides a feature-level classification for "good" mobility or "poor" mobility, detailing

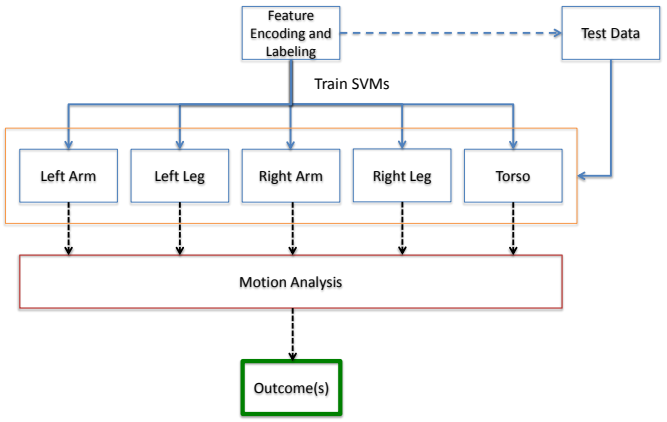


Fig. 6. Summary of the training and evaluation approach for analysing human mobility.

the state of mobility. Then, detailed analysis of the motion is undertaken on the complete labelled sequence to provide feedback on the participant's mobility.

For practical utilisation, each joint group is assessed based on the number of frames classified as having "good" mobility versus "poor" mobility. If any joint group has more than a predefined number of frames labelled as "poor" mobility, such as 30%, an outcome is generated highlighting that further investigation is required.

IV. Experimental: Motion Detection and Quantification

The validation of the framework is undertaken in two parts. First, evaluate the ability of the framework to perform motion recognition. Second, evaluate the ability of the framework to detect mobility concerns and impairment. The evaluation is performed on "unseen" test participant's, meaning that no test data has been included during the modelling phase. Using the K3Da Dataset [11], we selected five trials scenarios from the dataset namely; *balance (eyes open)*, *chair rise*, *semi-tandem balance*, *tandem balance*, *walk (4 metres)*.

A. Evaluation: Motion Detection

We evaluated the performance of a range of machine learning classifiers using our joint groups. These classifiers, together with parameter selection methodology are presented in Table II. For each technique, a 10-fold cross-validation using the random 'leave-one-out' principle was used to compute recognition accuracy. Table IV illustrates the performance of each machine learning algorithm for motion recognition.

We were able to achieve acceptable recognition rates (Table IV), particularly when we consider the similarity in motions using our proposed joint groups and features. Random Forests produced the highest average recognition rate of 87.10%, GRBM producing the lowest average result of 73.08%. The recognition rates fluctuated due to the leave-one-out approach employed. Encouragingly, similar motions such as *semi-tandem* and *tandem* balance were identifiable, with little inter-/intra-class variation.

Below average recognition rates were observed for several iterations, this, in part due to the formation of the training and testing sets or cross-validation parameter selection may have been impacted due to inter-class similarity. Being able to correctly identify a motion is important to ensure the correct model is applied for motion analysis and quantification. Another consideration is the time in which it takes to perform classification, an average recognition per frame was below $1ms$, suggesting real-time recognition is viable. Leightley *et al.* [11] reported a peak accuracy of 85.53%, with our proposals obtaining a marginally increased peak accuracy of 87.10%.

B. Experimental: Motion Analysis

The ground truth labels derived in this work are used to evaluate the proposed framework for the task of detecting mobility. Figure 7 provides an overview for the success of the overall framework in identifying features of concern in relation to the ground truths.

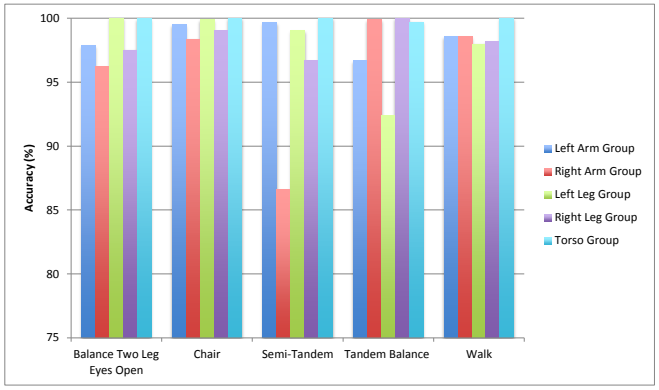


Fig. 7. Success of the framework when compared to ground truth labelling.

For evaluation, a random sample of participant's were used for training and the remainder used for testing, using leave-one-out methodology. Across trials evaluated, a high true-positive rate is obtained (Figure 7). This indicates that the framework is capable of correctly identifying mobility concerns based on the ground truths.

We were able to detect a large number of participant's who had been identified at labelling as *possibly* having a mobility concern. Overall, the confusion matrix (Table V) for each joint group performed strongly, with a high rate of true positives, and a small rate of false positives with an overall sensitivity of 0.98, specificity of 0.95 and Matthew Correlation Coefficient (MCC) score of 0.94 supporting this conclusion.

Balance - Two Legs (Eyes Open): Each participant stood with their feet as close together as possible side-by-side. They balanced with their eyes open and arms extended horizontally to be parallel with the floor. Each joint group could be correctly identified, with only *Left Leg* providing the lowest detection rate of 96.23%. The framework detected large amounts of mobility concern, most notably in the *Left Arm*. A

TABLE IV
MACHINE LEARNING RECOGNITION RATE FOR EACH ITERATION AND CLASSIFIER.

Iteration	SVM	RF	ANN	GRBM	AdaBoost	LPBoost	RUSBoost	Total Boost	Bagging
1	80.78%	86.75%	**83.92%**	**87.83%**	78.36%	82.58%	89.50%	84.19%	79.98%
2	79.94%	82.85%	78.38%	81.01%	86.22%	88.12%	85.15%	87.92%	82.47%
3	82.99%	89.04%	82.46%	79.87%	91.08%	**90.89%**	87.67%	83.29%	87.15%
4	76.59%	**93.48%**	80.92%	82.10%	**92.13%**	84.76%	82.73%	84.77%	89.86%
5	86.43%	78.47%	82.19%	80.56%	81.79%	79.36%	83.89%	**88.91%**	85.78%
6	76.49%	83.67%	70.71%	82.37%	79.37%	83.85%	83.74%	86.07%	84.37%
7	**86.65%**	92.16%	80.68%	76.10%	83.58%	85.52%	**90.17%**	86.55%	**89.98%**
8	84.24%	87.92%	82.83%	78.34%	82.01%	84.71%	86.24%	87.72%	87.72%
9	81.05%	90.11%	74%	56.28%	74.18%	86.62%	85.71%	88.23%	88.47%
10	77.14%	86.60%	78.99%	74.09%	87.61%	88.09%	84.37%	83.47%	89.91%
Average	81.23%	87.10%	79.50%	73.08%	83.63%	85.45%	85.91%	86.11%	86.56%

TABLE V
OVERALL CONFUSION MATRIX FOR THE PERFORMANCE OF EACH JOINT GROUP IN IDENTIFYING MOBILITY CONCERNS. WHERE TRUE POSITIVE INDICATES THE PARTICIPANT HAS GOOD MOBILITY AND TRUE NEGATIVE INDICATES PARTICIPANT'S WITH A MOBILITY CONCERN.

Left Arm - Predicted Outcome

actual value	p	n
p′	5230	103
n′	9	2549

Left Leg - Predicted Outcome

actual value	p	n
p′	5676	17
n′	113	2148

Right Arm - Predicted Outcome

actual value	p	n
p′	5717	239
n′	138	1860

Right Leg - Predicted Outcome

actual value	p	n
p′	6658	67
n′	73	1155

Torso - Predicted Outcome

actual value	p	n
p′	6423	0
n′	6	1498

sensitivity of 0.99, a specificity of 0.96 and a MCC of 0.96 was obtained. Mobility was in line with expectations, with the framework performing reliably across all joint groups.

Chair Rise: Each participant started from a seated position. When promoted, they had to stand up so that the legs were fully extended, and then sit down again. This was repeated five times with the aim to complete five complete stand/seat cycles. The arms were held across the chest so that all of the power needed to stand and sit was produced by the legs muscles. Each joint group could be correctly identified, with only *Left Leg* providing the lowest detection rate of 98.38%. A sensitivity of 0.99, specificity of 0.96 and an MCC of 0.96 was obtained. Overall, mobility across the participant range was good, with only a minority of features falsely classified as having "concern" across the joint groups.

Semi-Tandem Balance: Each participant placed one foot behind the other so that the big toe of the back foot was touching the side of the heel of the front foot. Their arms were fully extended horizontally for a period of 10 seconds. Each joint group could be correctly identified, with only *Right Arm* providing the lowest detection rate of 86.60%. A large number of features were identified as having a "concern" in *Right Arm*; this may be due to the incorrect recognition of features. This false classification is further observed with a relatively low sensitivity of 0.95, a specificity of 0.99 and a MCC of 0.91.

Tandem Balance: Each participant placed one foot directly behind the other so that the big toe of the back foot was touching the back heel of the front foot. The arms were fully extended horizontally for a period of 10 seconds. Each joint

group could be as being correctly identified, with only *Left Arm* providing the lowest detection rate of 96.69%. For *Right Leg*, all features were classified correctly. This may be in part due to most participant's using their left leg for the balancing resulting in the sensor being obscured from the right leg; therefore the Right Joint Group may be hidden for duration of the motion. A high sensitivity of 0.99 was achieved, however low scores for specificity of 0.91 and MCC of 0.92.

Walk (4 metres): Each participant started from a standing position and walked forwards in a straight line towards the sensor at their 'usual' walking speed. Each joint group could be correctly identified, with only *Right Arm* providing the lowest detection rate of 98.10%. Across the joint groups, features were classified correctly, with only a few features being identified as a concern requiring further invention, this was further supported with a sensitivity of 0.99. However, when considering specificity of 0.94 and a MCC of 0.95, the framework does yield low results for this type of trial when compared to the others.

With the proposed framework it is possible to use the recognition of frames to identify if any intervention or rehabilitation is required. A threshold of 70% was selected through experimentation, if any joint group contained more than 70% of frames classified as "concern" it would be acknowledged as requiring investigation by a clinician. Of the participant's used in these experiments, 16 were highlighted as having at least one joint group of concern. In a clinical context, these participant's would be examined further by a clinician to determine why poor mobility was observed. In accuracy terms, this is a 94% success rate in detecting mobility concerns between participant groups based on ground truth labelling.

V. Discussion and Conclusion

The release of the Kinect One has presented new capacities for innovation within the healthcare sector. The ability to deploy the sensor in a wide range of locations, as well as its low-cost are important highlights. Further, the Kinect One is capable of providing detailed measurements extracted from motion sequences and encoded features. Standard approaches, such as SPPB [43] provide a single score measurements with no contextual information whereas the Kinect One is able to provide finite kinematic information. The extraction of joint groups provides an abstract level of detail and insights into how the joint group is performing in relation to the motion as a whole, this led to an improved insight for clinicians to make a recommendation.

In this work, we utilise the Kinect One for detecting mobility concerns to aid in stratified clinical intervention. The motions used for evaluation are commonplace and necessary parts of typical daily living, such as walking, sitting, standing and balancing. These same movements become problematic for persons with a mobility concerns of any age. Due to the large inter-individual variability in age and physical capabilities, the K3Da dataset is well suited for evaluating the proposed framework. However, it contains only a limited number of participant's which makes general population modelling difficult. If the number of participant's for modelling is increased, a more reliable and representative population model can be computed. Future work will explore this in more detail.

While analysis is a key theme of this work, it is important to consider detection of motions as they occur, this enables the correct outcome measure to be applied. We have compared several classifiers, presenting a detailed comparison of their ability to detect a range of subtly different motions. They have been able to detect subtle differences between similar motions, for example *semi-tandem* and *tandem* balance with the aid of our proposed feature set. This work has united recognition and motion analyse to provide a united decision making process.

We propose a framework which unites human motion recognition techniques with motion analysis. The framework has been shown to be reliable and accurate for evaluation of mobility. By utilising low-cost depth sensor technology the application framework is deployable in a large number of scenarios and environments, resulting in real world practical benefits. Future work will focus on clinical validation of the proposed framework with an increased population size using power analysis.

References

[1] D. Leightley, J. Darby, B. Li, J. S. Mcphee, and M. H. Yap, "Human activity recognition for physical rehabilitation," in *IEEE Conference on Systems, Man and Cybernetics*, Manchester, UK, October 2013, pp. 261–266.

[2] R. Collard, H. Boter, R. A. Schoevers, and R. C. Oude Voshaar, "Prevalence of frailty in community-dwelling older persons: a systematic review," *Journal of the American Geriatrics Society*, vol. 60, pp. 1487–1492, 2012.

[3] General Practitioners, "Fit for frailty: Consensus best practice guidance for the care of older people living with frailty in community and outpatient settings," 2014.

[4] M. Fuhrer, "Subjectifying quality of life as a medical rehabilitation outcome," *Disability Rehabilitation*, vol. 22, no. 11, pp. 481–489, 2000.

[5] P. Gregory, J. Alexander, and J. Satinksky, "Clinical telerehabilitation: Applications for physiatrists," *American Academy of Physical Medicine and Rehabilitation*, vol. 3, no. 7, pp. 647 –656, 2011.

[6] R. Wang, G. Medioni, C. J. Winstein, and C. Blanco, "Home monitoring musculo-skeletal disorders with a single 3d sensor," in *IEEE Conference on Computer Vision and Pattern Recognition (Workshop)*, June 2013.

[7] J. Bae and M. Tomizuka, "A tele-monitoring system for gait rehabilitation with an inertial measurement unit and shoe-type ground reaction force sensor," *Mechatronics*, vol. 23, no. 6, pp. 646–651, September 2013.

[8] L. P. Fried, C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, and M. A. McBurnie, "Frailty in older adults: Evidence for a phenotype," *Gerontology: Biological Sciences*, vol. 56, no. 3, pp. 808–813, March 2001.

[9] M. S. Cameirao, S. Bermudez, E. D. Oller, and P. F. Verschure, "The rehabilitation gaming system: a review," *Studies in health technology and informatics*, vol. 145, pp. 65– 83, 2009.

[10] D. Webster and O. Celik, "Experimental evaluation of microsoft kinect's accuracy and capture rate for stroke rehabilitation applications," in *Haptics Symposium*, Feb 2014, pp. 455–460.

[11] D. Leightley, M. H. Yap, J. Coulson, Y. Barnouin, and J. S. McPhee, "Benchmarking human motion analysis using kinect one: an open source dataset," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015.

[12] 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, 2012.

[13] B. Galnaa, G. Barrya, D. Jacksonb, D. Mhiripiria, P. Olivierb, and L. Rochestera, "Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease," *Gait and Posture*, vol. 39, no. 4, pp. 1062–1068, 2014.

[14] N. Vernadakis, V. Derri, E. Tsitskari, and P. Antoniou, "The effect of xbox kinect intervention on balance ability for previously injured young competitive male athletes: A preliminary study," *Physical Therapy in Sport*, vol. 15, no. 3, pp. 148–155, 2014.

[15] S. Gauthier and A. Cretu, "Human movement quantification using kinect for in-home physical exercise monitoring," in *IEEE Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, 2014.
[16] M. S. Cameirão, S. Bermúdez, E. D. Oller, and P. F. Verschure, "Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation," *NeuroEngineering and Rehabilitation*, vol. 7, no. 48, 2010.
[17] M. da Silva Cameirao, S. B. i Badia, E. Duarte, and P. F. Verschure, "Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: A randomized controlled pilot study in the acute phase of stroke using the rehabilitation gaming system," *Restorative Neurology and Neuroscience*, vol. 29, no. 5, pp. 287–298, 2011.
[18] E. Dolatabadi, B. Taati, G. S. Parra-Dominguez, and A. Mihailidis, "A markerless motion tracking approach to understand changes in gait and balance: A case study," in *Rehabilitation Engineering and Assistive Technology Society of North America*, 2013.
[19] A. Gonzalez, P. Fraisse, and M. Hayashibe, "A personalized balance measurement for home-based rehabilitation," in *IEEE Conference on Neural Engineering*, 2015, pp. 711–714.
[20] F. Cary, O. Postolache, and P. Silva Girao, "Kinect based system and artificial neural networks classifiers for physiotherapy assessment," in *MeMeA*, June 2014.
[21] A. Kargar, A. Mollahosseini, T. Struemph, W. Pace, R. Nielsen, and M. Mahoor, "Automatic measurement of physical mobility in get-up-and-go test using kinect sensor," in *IEEE Conference on Engineering in Medicine and Biology Society*, 2014.
[22] V. Bloom, D. Makris, and V. Argyriou, "Clustered spatio-temporal manifolds for online recognition," in *IEEE Conference on Pattern Recognition*, August 2014, pp. 3963–3968.
[23] A. Sinha and K. Chakravarty, "Pose based person identification using kinect," in *IEEE Conference on Systems, Man and Cybernetics*, October 2013, pp. 497–503.
[24] E. Gianaria, N. Balossino, M. Grangetto, and M. Lucenteforte, "Gait characterization using dynamic skeleton acquisition," in *International Workshop on Multimedia Signal Processing*, Sept 2013, pp. 440–445.
[25] B. Dikovski, G. Madjarov, and D. Gjorgjevikj, "Evaluation of different feature sets for gait recognition using skeletal data from kinect," in *International Convention on Information and Communication Technology, Electronics and Microelectronics*, 2014, pp. 1304–1308.
[26] Y. Du, W. Wang, and L. Wang, "Hierarchial recurrent neural network for skeleton based action recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1110–1118.
[27] L. Zhang, J. C. Hsieh, T. T. Ting, Y. C. Huang, Y. C. Ho, and L. K. Ku, "A kinect based golf swing score and grade system using gmm and svm," in *International Congress on Image and Signal Processing*, 16-18 Oct 2012, pp. 711–715.
[28] M. Lewandowski, J. Martinez-Del-Rincon, D. Makris, and J. C. Nebel, "Temporal extension of laplacian eigenmaps for unsupervised dimensionality reduction of time series," in *IEEE Conference on Pattern Recognition*, August 2010, pp. 161–164.
[29] A. Gonzalez, M. Hayashibe, and P. Fraisse, "Estimation of the center of mass with kinect and wii balance board," in *Conference on Intelligent Robots and Systems*, 2012, pp. 1023–1028.
[30] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake, "Efficient human pose estimation from single depth images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, p. 1, 2012.
[31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273 – 297, September 1995.
[32] L. Breiman, "Random forests," *Machine Learning*, vol. 25, no. 1, pp. 5 – 32, 2001.
[33] G. W. Taylor, G. E. Hinton, and S. Roweis, "Modeling human motion using binary latent variables," in *Advances in Neural Information Processing Systems*, 2006, pp. 1345–1352.
[34] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 15, no. 4, pp. 771–780, 1999.
[35] H. T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, second edition, Ed. New York: Spriner, 2008.
[36] S. C., T. Khoshgoftaar, J. Hulse, and A. Napolitano, "Rusboost: Improving clasification performance when training data is skewed," in *IEEE Conference on Pattern Recognition*, 2008, pp. 1–4.
[37] M. Warmuth, J. Liao, and G. Ratsch, "Totally corrective boosting algorithms that maximize the margin," in *Proc. 23rd Int'l. Conf. on Machine Learning*, 2006, pp. 1001–1008.
[38] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 26, pp. 123–140, 1996.
[39] A. Amini Maghsoud Bigy, K. Banitsas, A. Badii, and J. Cosmas, "Recognition of postures and freezing of gait in parkinson's disease patients using microsoft kinect sensor," in *IEEE Conference on Neural Engineering*, April 2015, pp. 731–734.
[40] R. N. Baumgartner, K. M. Koehler, D. Gallagher, L. Romero, S. B. Heymsfield, R. R. Ross, P. J. Garry, and R. F. Lindeman, "Epidemiology of sarcopenia among the elderly in new mexico," *American Journal of Epidemiology*, vol. 147, no. 8, pp. 755–763, 1998.
[41] R. M. Dodds, H. E. Syddall, R. Cooper, M. Benzeval, I. J. Deary, E. M. Dennison, G. Der, C. R. Gale, H. M. Inskip, C. Jagger, T. B. Kirkwood, D. A. Lawlor, S. M. Robinson, J. M. Starr, A. Steptoe, K. Tilling, D. Kuh, C. Cooper, and A. A. Sayer, "Grip strength across the life course: Normative data from twelve british studies," *PLoS One*, 2014.
[42] D. Leightley, B. Li, M. H. Yap, J. McPhee, and J. Darby, "Exemplar-based human action recognition with template matching from a stream of motion capture," in *Lecturer Notes in Computer Science*, October 2014.
[43] J. Guralnik, E. Simonsick, L. Ferrucci, R. Glynn, L. Berkman, D. Blazer, P. Scherr, and R. Wallace, "A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission," *Gerontology*, vol. 49, no. 2, pp. 85 – 93, 1994.

Daniel Leightley received his B.Sc (Hons.) degree in Information and Communications, M.Sc. degree in Computing and Ph.D degree in Computer Science from Manchester Metropolitan University. He is currently a Post-Doctoral Research Assistant with the Centre for Military Health Research at King's College London. His research interests are human action recognition, motion analysis, human quantification and depth sensor processing.

Jamie S. McPhee is Reader of Muscle and Exercise Physiology at Manchester Metropolitan University and leads the university's Neuromuscluar and Skeletal Ageing Research Group. His main research interests are focussed on understanding the physiological mechanisms of muscle wasting and weakness in older age, how this relates to mobility problems faced by older people, and in identifying effective physical rehabilitation or other interventions to keep people healthy and independent for longer.

Moi Hoon Yap received her B.Sc (Hons.) degree in statistics and M.Sc. degree in Information Technology from Universiti Putra Malaysia, and Ph.D. degree in Computer Science from Loughborough University, in 2009. After her Ph.D, she was a Post-Doctoral Research Assistant with the Centre for Visual Computing, University of Bradford. She is currently Senior Lecturer with Manchester Metropolitan University. Her research interests are face and gesture analysis, medical imaging analysis and software development.