Benchmarking Human Motion Analysis Using Kinect One: an open source dataset

Daniel Leightley*, Moi Hoon Yap*, Jessica Coulson†, Yoann Barnouin† and Jamie S. McPhee†
*School of Computing, Mathematics and Digital Technology
†School of Healthcare Science
Manchester Metropolitan University, UK, M1 5GD.
Email: dleightley@ieee.org, {m.yap, j.coulson, y.barnouin, j.s.mcphee }@mmu.ac.uk

***Abstract*—There is a clear advantage to developing automated systems to detect human motion in the field of computer vision for applications associated with healthcare. We have compiled a diverse dataset of clinically-relevant motions using the Microsoft Kinect One sensor and release the dataset to the community as an open source solution for benchmarking detection, quantification and recognition algorithms. The dataset, namely Kinect 3D Active (K3Da), includes motions collected from young and older men and women ranging in age from 18 - 81 years. Participants performed standardised tests, including the Short Physical Performance Battery, Timed-Up-And-Go, vertical jump and other balance assessments which were recorded using depth sensor technology and extracted to generate motion capture data, sampled at 30 frames-per-second. Preliminary evaluations using Support Vector Machines, Random Forests, Artificial Neural Networks and Boltzmann Machines show age-related differences in many of the movements. These results demonstrate the relevance of the dataset to support benchmarking of algorithms associated and/or intended for use in a healthcare setting.**

***Index Terms*—Kinect One, motion capture, dataset, healthcare, benchmarking**

I. Introduction

Automatic methods for detection, recognition and quantification of human movements has become more accessible due to increased availability of low-cost multi-modality marker-less capturing devices [1], [2]. This provides potential to develop applications suitable for use in healthcare settings to detect problems that patients have in coordination of movements [3]–[6]. For example, Alankus *et al.* [7] and Wang *et al.* [5] devised techniques to characterise movements in stroke and musculoskeletal patients, respectively. However, both utilised publicly available datasets that were intended for use in gaming populations, restricting their broader application. Indeed, existing datasets (*e.g.* [8]–[10]) were captured for specific purposes, such as daily living, first person or gestures, principally for use in the entertainment and gaming industries. One of the most popular datasets, Carnegie Mellon University Motion Capture Database (CMU MoCap) [11] includes 2600 trials across 23 action categories captured using a marker-based Vicon system. While the numbers of trials and action categories were diverse, rigid recording protocols captured movements from a university student population. A similar dataset named G3D [12] (later complemented with G3Di [13]) captured 200 trials across 20 categories from 10 participants using a Microsoft Kinect 360 sensor. While this sensor is affordable and portable, the database does not provide enough diversity. Currently, none of the available datasets specifically includes movements based on common clinical assessments of patient groups and this limits the development of tools and applications for use in healthcare settings.

Movement problems experienced by a diverse group of patients include slow and altered gait, difficulties changing from standing-to-sitting or sitting-to-standing and balancing. These problems increase the risk of disability and falls which have major consequences for quality of life and healthcare provision. Specialist nursing staff, physiotherapists and geriatricians routinely assess movements using standardised tests such as the Short Physical Performance Battery (SPPB [14]), Timed-Up-And-Go (TUG [15]), Six-Minute Walk [16] and Balance (*e.g.* Tinetti [17]). However, manual-assessments require trained staff and variations between assessor ratings and experience may cause problems. Computer-based analysis of these movements can standardise the assessments and may be more resource-effective. Automated assessments requires algorithms to detect joint angles in different body segments, stride length and foot positioning, whilst also accounting for the diversity that exists across populations in terms of body size and shape.

Depth-sensors have been used several times for the assessment of balance [6], [18]–[21] by extracting simple gait-based vectors from a skeletal stream to provide basic stability-based single value scores. Using the Microsoft Kinect for Xbox 360, Yang *et al.* [21] showed that detection of the centre-of-mass was correlated with standard assessments performed on a force platform. Classifying, recognising and segmenting human balance can be defined into two categories: model-based and model-free [22]. Model-based approaches use pre-defined models of the underlying kinematics in order to represent the data and capture balance. This approach, while computationally more expensive, gives representations that are more informative while also being view point, style, anthropometric and subject invariant. Static and dynamic features consist of height, stride length, velocity and execution duration [23], [24]. Model-free approaches, otherwise referred to as appearance-based methods, rely on low-level processing of image data to extract the silhouette or motion-history-images of the segmented human body which can be used to encode human motion characteristics [25]. While these are computationally efficient, they have significant drawbacks in view point, lighting, occlusion and preservation of dynamic kinematics of human motion.

There are several techniques to determine the complex spatio-temporal features of a persons walking style [26], [27]. Sinha and Chakravarty [28] identified key poses relevant to gait using machine learning, although subjects were used for both learning and decision stages, making their predictions more favourable, but limiting the generalisation to a diverse population. Gianaria *et al.* [24] extracted static and dynamic features from a sequential time series of skeletal data to describe the subjects posture (*e.g.* height, stride length, sway). However, none of these techniques gives detailed quantitative feedback, as would be needed in a healthcare setting.

To address the limitations in available datasets that do not contain healthcare-related motions, we have established a dataset, Kinect 3D Active (K3Da), to capture balance, walking, sitting and standing from a diverse population of young and older adults. The dataset is readily available from http://www.k3da.leightley.com. The motions were based on common clinical assessments used to assess movements in disease and frailty. Here, our aim is to describe the study design and release of the dataset for benchmarking of human movement detection, quantification and recognition algorithms. We also demonstrate two possible applications in motion recognition as well as quantifying simple differences in movements between the young and older populations. Finally, we have released a basic toolset to facilitate novel data capture, viewing and motion analysis.

II. Related Datasets

There are a vast number of datasets that exists for benchmarking proposed techniques, these datasets can be defined into three categories. Firstly, those that are simply action datasets for use in recognition such as G3D [12] which contain simple, basic level action sequences obtained in a controlled environment. Secondly, security/surveillance datasets such as i-Lids [29] which are captured in realistic environments such as airports and bus stations. The third type are movie datasets, which are obtained from movie scenes such as Hollywood2 [30]. to the authors' knowledge, there are no healthcare-related datasets in existence for benchmarking clinically-related motions.

In the computer vision community there exists multiple datasets composed of different modalities such as Euler angles, RGB images and depth images. For MoCap data these are composed entirely of two modalities, namely marker-based systems such as Vicon and marker-less systems such as Microsoft Kinect (360/One) sensor.

Marker-based systems require markers to be placed on the user at anatomical significant locations. Using multiple cameras, these markers are tracked resulting in Euler angles that are relative to the camera coordinate system. CMU MoCap [11] is the most popular marker-based dataset in use. It consists of a large amount of game-orientated trials recorded in a lab-based setting. The HDM05 [31] contains a limited number of realistic fitness workout trials captured by five participants, the dataset contained a strict recording protocol resulting in each trial being similar in nature. Finally, TUM Kitchen dataset [32] consists of multi-modality dataset including video and MoCap. The dataset captures participants in a daily living scenario performing specific tasks, with participants asked to perform motions as they would in the home. However, these marker-based datasets lack the action sequences and interactions that are necessary in a clinical setting. Further, their rigid capture protocols results in data being similar across multiple trials by different subjects.

Recent technological advancements has led to the availability of low-cost and easy to use image sensor technology (*e.g.* Microsoft Kinect 360/One, ASUS Xtion). These systems are marker-less, where the human skeleton is extracted from depth or video sequences to provide x, y and z coordinates for specific locations. The G3D [12] dataset provides image, depth and skeleton data captured using a Microsoft Kinect 360. The dataset contains a range of gaming actions from 10 participants performing 20 gaming actions in a controlled setting. More recently, the authors of [12] introduced the G3Di [13] dataset which captured 12 participants split into 6 pairs in a multiplayer game setting. Finally, the MSRDailyActivity3D [33] consists of 10 participants performing daily living activities such as *eating* or *reading a book*. As is common with other datasets, the MSRDailyActivity dataset was captured using a rigid protocol to ensure uniformity in action trials.

While datasets exist to reflect daily living and gaming actions/activities. To our knowledge, this dataset is the first proposed to provide clinically supported motion sequences from both the young and elderly using depth sensor technology. Further, this is the one of the first dataset introduced using the latest edition of the Microsoft Kinect One sensor directly focused at benchmarking medical frameworks.

III. Materials and Methods

A. Participants and ethical approval

The study was approved by the Research Ethics Committee at Manchester Metropolitan University (approval SE121308). All participants gave signed informed consent to take part in data collection and for their depth and skeleton data to be published. The acquisition sessions consisted of 13 tests based on the SPPB, TUG and additional tests of balance and power output (Table. I). These clinically relevant assessments of mobility and physical performance were led by experienced human physiologists and followed standardised protocols.

Fifty-four participants (32 men and 22 women) were recruited with a mean age of 25.53 (Standard Deviation (SD) of 23.54) and minimum/maximum of 18/81 yrs and a diverse range of body compositions.

Due to the large inter-individual variability in age and physical capabilities, the dataset contains large motion variation in the skeletal pose. For example, during the chair rise, some participants could easily perform five chair rises very quickly without losing balance or performance, while others (mainly older men and woman) experienced a deterioration of their

TABLE I
DETAILED CAPTURE PROTOCOL TEST DESCRIPTION FOR EACH TRIAL THAT IS CONTAINED WITHIN THE K3DA DATASET.

Test	Capture Protocol	Instructions or Constraints
Balance (open eyes)	The 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	Test terminated after 10 seconds
Balance (closed eyes)	The participant stood with their feet as close together as possible side-by-side. They balanced with their eyes closed and arms extended horizontally to be parallel with the floor	Test terminated after 10 seconds
Chair Stand	The participant started from a seated position. When instructed, 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	Perform five chair rises as quickly as possible. Test terminated after 60 seconds.
Jump (low power)	The participant stood with their legs fully extended and slightly less than shoulder width apart. When instructed, they produced a counter movement jump by bending at the knees and then performing a low-level jump	Perform and low-level jump
Jump (maximum power)	The participant stood with their legs fully extended and slightly less than shoulder width apart. When instructed, they produced a counter movement jump by bending at the knees and then performing a maximal-level jump	Perform a maximal-effort jump
One Leg Balance (closed eyes)	When instructed, the participant balanced with one leg (participant preference) 6 inches off the ground with their eyes closed and arms extended horizontally	Test terminated after 10 seconds or when the second leg touched the ground
One Leg Balance (open eyes)	When instructed, the participant balanced with one leg (participant preference) 6 inches off the ground with their eyes open and arms extended horizontally	Test terminated after 10 seconds or when the second leg touched the ground
Semi Tandem Balance	The participant was asked to place 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	Test terminated after 10 seconds
Tandem Balance	The 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	Test terminated after 10 seconds
Walk towards (towards Kinect)	The participant started from a standing position and walked forwards in a straight line towards the sensor at their usual walking speed	walk at 'usual' walking speed
Walk away (from Kinect)	The participant started from a standing position very close to the sensor and walked away from the sensor in a straight line at their usual walking speed	walk at 'usual' walking speed
Timed Stand Up and Go	The participant started in a seated position. They had to rise from the chair, walk 3 meters, turn around and walk back to sit on the chair again	walk at 'usual' walking speed
Hopping (One-Leg)	The participant was asked to hop with one leg (participant preference) on the spot multiple times	Test terminated after 10 seconds

performance throughout the test. A large number of motion variations also exist within the dataset. This is because no individual will perform the test in exactly the same way on each attempt, even when the test requires multiple repetitions. For example, small differences in gait cycle during walking or sway during balancing can be detected by the Kinect One.

Data collection and storage

The Microsoft Kinect One depth sensor was fixed horizontally to a tripod at a height of 0.7 m and all assessments were confined to within range of the sensor. Room furniture was removed to ensure maximum visibility and room lighting was standardised. The participants were provided with a maximum of three attempts to complete each short task (which is common in a clinical setting). A countdown timer was created to prompt the participant to start each test and sessions were recorded and stored automatically.

The Microsoft Kinect One sensor coupled with the Microsoft Windows Software Development Kit [34] synchronised capture of depth and skeleton streams at 30fps. Each data stream was retrieved and stored in a unique file for each time period with a unique millisecond timestamp. The raw storage format was selected for the depth stream, the raw information contains the depth of each pixel in millimetres. The 16-bits of depth data contain 13 bits for depth and 3 to identify the person-index. A text format was selected for storing the skeleton information with participants position, pose and relative depth map coordinates. The pose includes 25 joints and two action states as defined by Microsoft. The participants overall and joint positions are given as x, y and z coordinates in meters. These positions are also mapped into depth coordinates. The skeleton data includes a joint tracking state, shown as 'tracked', 'not tracked' and 'inferred'.

The depth and skeletal streams were extracted from the Kinect One data stream while the participant performed the movements. The Microsoft Kinect One sensor provided a 512×424 depth image up to 30fps. Skeletal time series consisted of 25 3D orthogonal (x, y, z) locations. An example

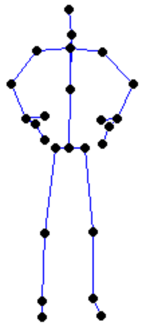

Fig. 1. A figure representing the skeleton structure extracted from the Kinect One. There are a total of 25 tracked joints using the algorithm presented in [35].

representation is shown in Fig. 1. Frame data were extracted in real time using the technique of Shotton *et al*. [35], which is part of the Microsoft SDK [34]. The developed dataset, K3Da, currently contains 525 tests from 54 participants, although more will be added over time. This has resulted in over 225,000 frames of depth and skeleton data. Each test has been encoded with age, height and a labelled matrix of outlier (noisy) frames.

Basic toolset

We have included a suite of tools to visualise and utilise the datasets. This includes a Kinect One Recorder (Desktop Application) capable of recording the skeletal and depth streams from the Microsoft Kinect One Sensor at 30fps; a Kinect One Visualiser which is a Matlab-based program to visualise the captured data (Fig. 2); and a Kinect One Analysis (Matlab-based program) interface that includes joint profiling and measurement of distance travelled.

IV. Dataset Evaluation

To identify and recognise which test case is being observed, we have adapted a method previously described [1] that uses the exemplar paradigm coupled with clustering techniques for recognition in real-time. To that end, it is important to identify the correct action before analysis on the action can be undertaken. Briefly, p_t^j denote the 3D position of a set of joints j at a given frame t *e.g.* $p_t^j = \{x_t^j, y_t^j, z_t^j\}_{j=1:J}^{t=1:T}$, with T represents the number of frames. Therefore, a motion can be seen as a set of poses *e.g.* $\mathbf{M} = \{p_1, p_2, \ldots, p_T\}, p \in \mathbb{R}^{3*J}$.

Instead of using a pseudo-invariant action space or looking for a reference system which is the most discriminative for each motion class, like subspace approaches, we have used a body relative coordinate system based on the skeletal pose itself. The relative motion of the body parts with respect to the torso is capable of describing the action. Therefore, we first selected a body relative reference joint and use it to redefine the 3D coordinates of all joints for each time period. The new frame of reference ($\bar{p^t}$) is described by the hip centre origin $\mathbf{e}$ of frame t and base vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ *e.g.* $\mathbf{e}_t^{\mathbf{x},\mathbf{y},\mathbf{z}}$, which is demonstrated in Fig. 3. The transformation of coordinates in the Kinect world to the body relative pose is undertaken by subtracting the $\mathbf{e}_t^{\mathbf{x},\mathbf{y},\mathbf{z}}$ from each joint of time t.

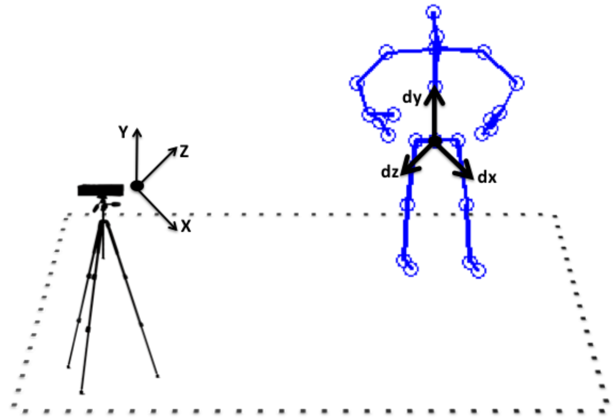

Fig. 3. A figure representing the global and local coordinate system of the Microsoft Kinect One.

Exemplar selection

Human poses may be semantically, but not necessarily numerically similar, yet represent the same time instance of a motion. To further complicate matters, to identify and group similar poses manually is a time consuming and arduous task. In the approach adopted from [1], we have selected a centroid method, namely k-means clustering [36] to group a set of poses in $\mathbf{M}$, into similar groups. The k-means algorithm is iterative in nature, starting with an initial estimation of the centroid for each cluster and continues until convergence of a motion sequence into an assigned number of k clusters. However, while k-means is robust for unravelling compact sequences, such as MoCap, accurate estimation of the number of clusters is crucial. There is no single solution to estimating the optimum k value, with several works selecting k manually (*e.g.* [37], [38]) or using automatic selection methods (*e.g.* [39]–[41]). To select the optimum k, the Elbow method [41] was used to represent the within-cluster-sum-of-squares (WSS). The purpose is to cluster a sequence for a single test class into k clusters. A useful aspect of k-means process is that each cluster characterises a phase of the test case, as demonstrated in Fig. 5. In order to select a delegate for each k cluster, a ranking scheme for each pose according to the City Block metric (also referred to as Manhattan distance) is utilised. The equivalence D between any two poses $\bar{p}_m$ and $\bar{p_n}$ in a cluster is measured using the total distance amongst corresponding joints, defined as

$$D(\bar{p}_m, \bar{p_n}) = \sum_{j=1}^{J} [\bar{p}_m^j - \bar{p}_n^j] \tag{1}$$

where D is the distance between any two poses of the same k cluster. The delegate exemplar is a pose which has the lowest distance average between all of the poses within the cluster. Therefore, for each test class, a set of exemplars representing the different phases is selected to form the training set.

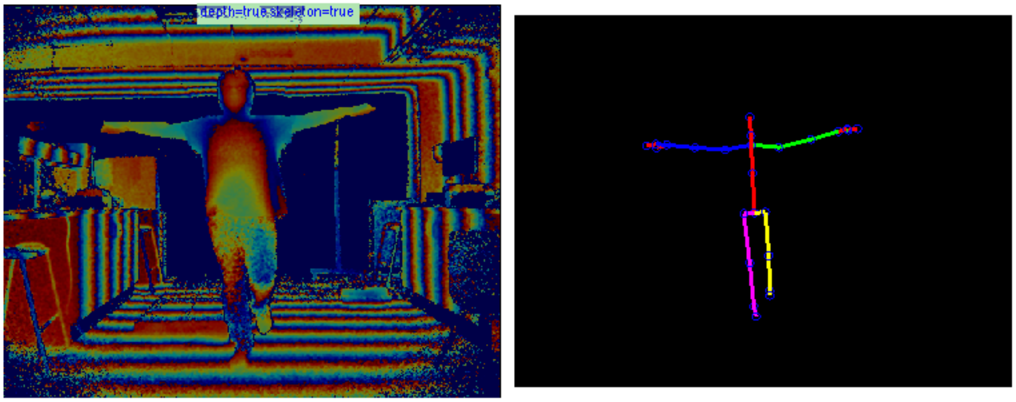

Fig. 2. Left-to-right: raw depth image (512 x 424) and skeleton representation (25 tracked joint locations)

TABLE II
BENCHMARK IDENTIFICATION RESULTS: RECOGNITION RESULTS FOR EACH MACHINE LEARNING METHOD, MODEL TRAINING AND RECOGNITION TIME. WHERE *s* IS SECONDS, *ms* IS MILLISECONDS.

Method	Rate (%)	Model training (s)	Recognition execution (ms)
SVM	85.53	1.89	4.9
RF	62.89	2.45	1.5
ANN	82.42	4.56	1.78
GRBM and SVM	79.64	3.87	7
GRBM and RF	71.11	4.09	8

V. RESULTS

For preliminary evaluation of the dataset and the ability to identify each test case, we utilised Support Vectors Machines (SVM) [42], Random Forest (RF) [43], Artificial Neural Networks (ANN) and Gaussian Restricted Boltzmann Machines (GRBM) [44] with 10-fold cross validation on each method. Unlike other approaches that use "leave one out" method, we have sought to generate a more representative result by randomly selecting a 40/60 training and testing set split. This reflects real-world situations where testing will always outweigh training data. This method relies on the appropriate selection of the k parameter for k-means. Initial results are shown in Table II. This approach removed irrelevant information, resulting in a very small training set.

Therefore, complexity and in turn model training times were reduced to only a few seconds. SVM and ANN had high recognition success rates, with the other techniques obtaining respectful results. SVM had much shorter training time than ANN, while conversely recognition execution time was shorter for ANN than for SVM. Conversely, each GRBM approach had a significantly longer recognition time. The k, the number of clusters for each sequences was automatically generated using the methoed defined above, the average k across the model was 13 (SD = 4). Fig. 4 shows an example of k-means segmenting a MoCap sequence into distinct cluster groupings.

Walk towards (towards Kinect): Participants started from a standing position and walked forward in a straight line towards the Kinect One sensor at their usual walking speed. A total of 24 test cases were used to compile results. The average time taken, walking speed and distance travelled in the groups of young and older people are presented in Table III. These results demonstrate a small divide between young and older people.

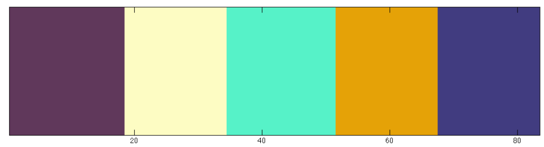

Fig. 4. Action label for each frame composing of the *Walk* action segmented into stride from a Support Vector Machine.

TABLE III
WALKING TOWARDS KINECT ONE - INITIAL ANALYSIS: AVERAGE HUMAN MOTION ANALYSIS RESULTS FOR YOUNG AND OLDER PARTICIPANT GROUPS UNDERTAKING A WALK. WHERE *s* IS SECONDS, *m* IS METERS AND *mps* IS METER-PER-SECOND.

Group (test case)	Time taken (s)	Distance Travelled (m)	Speed (mps)
Young (14)	2.22 (SD=0.046)	3.06 (SD=0.27)	1.41 (SD=0.21)
Older (10)	2.40 (SD=0.34)	3.08 (SD=0.19)	1.3 (SD=0.16)

One Leg Balance (closed eyes): Participants balanced with one leg (participant preference) 6 inches off the ground. Then, they balanced with their eyes closed and arms extended horizontally to be parallel with the floor for a maximum of 10 seconds. A total of 70 test cases were used to compile the results. Balance and stability are important factors in the vast majority of daily tasks. Detecting an identifying potential issues in relation to balance, between young and older groups, but also within group differences can aid in identifying age-/gait-related deficits. Table IV shows the data from trials in which subjects balanced on one leg with their eyes closed. Older people had larger movements compared with young and fewer older people were able to achieve the target of 10 seconds.

TABLE IV
ONE LEG BALANCE - EYES CLOSED - INITIAL ANALYSIS:AVERAGE HUMAN MOTION ANALYSIS RESULTS FOR YOUNG AND OLDER PARTICIPANT GROUPS STANDING ON ONE LEG WITH THEIR EYES CLOSED. WHERE *s* IS SECONDS AND *cm* IS CENTIMETRES.

Group (test case)	Average Time (s)	Failed Attempts	Movement Area (cm)	Number Passed
Young (26)	7.6 (STD=3.01)	12	11.85 (SD=9.08)	14
Older (44)	5.58 (SD=3.21)	38	23.79 (SD=19.39)	7

TABLE V
SIT TO STAND - INITIAL ANALYSIS: MOTION ANALYSIS ON THE ESTIMATED AND ACTUAL NUMBER OF CHAIR RISES FOR EACH PARTICIPANT GROUP, WITH AVERAGE TIME PER RISE.

Group (test case)	Chair Rise (Actual)	Chair Rise (Estimated)	Avg Time Per Rise (s)
Young (16)	80	78	1.44 (SD=0.27)
Older (17)	83	84	1.72 (SD=0.23)

Sit to Stand: Participants started from a seated position and 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 stand/seat cycles. A total of 33 test cases were used to compile the results. The number of chair rises, as well as the time taken to perform a rise can indicate mobility issues with the lower limbs. Using existing peak detection algorithms we are able to detect the chair rise for each participant with a high degree of accuracy (as demonstrated in Table V). In addition, subtle time differences between young and the older people were evident.

VI. DISCUSSION AND POTENTIAL APPLICATIONS

There is evidence that the most important risk factors for falls and mobility limitations are muscle weakness, particularly thigh muscles [45]. However medical imaging such as MRI is expensive and alternative non-intrusive system to automatically detect and classify movements is of great importance in healthcare. For instance, early identification of people most at risk of deterioration of physical function gives more time for remedial interventions, such as lifestyle or physical rehabilitation, before the impairments are irreversible. Such systems also give the opportunity for long-term monitoring of patients to observe effects of illness or ageing or monitor effectiveness of rehabilitation programmes. The movements that we captured were designed by healthcare professionals and the data collection sessions were conducted according to standardised protocols. The movements are commonplace and necessary parts of typical daily living, such as walking, sitting, standing and balancing. These same movements become problematic in disabled and in older people, leading to frailty that affects around 9% of the population [46]. Thus, the movements that we adapted are common clinical assessments. Due to the large inter-individual variability in age and physical capabilities, the dataset contains large motion variation in the skeletal pose. For example, some participants could easily perform five chair rises very quickly without losing balance or performance, while others (mainly older people) experienced a deterioration of their performance throughout the test. It is also possible to see intra-individual variations where the same subjects performed the same trials more than once, resulting in slight differences in movements and timings.

The increased availability of low-cost and user-friendly image sensor technology such as the Microsoft Kinect 360/One or ASUS Xtion, enable marker-less data collection. Marker-based systems, such as VICON, use multiple cameras to track body segments and calculate Euler Angles that are relative to the camera coordinate system. This is not possible with marker-less technology, so other solutions will need to be developed. It can be difficult to determine which machine learning approach to utilise to detect and classify movements when considering MoCap [1]. The methods that we applied extracted the human skeleton from depth or video sequences to provide x, y and z coordinates. Each of the methods showed acceptable recognition rates, with varying model training times.

VII. CONCLUSION

There are numerous datasets that are publicly available for the research community, however these were intended to be used for entertainment-based algorithms. A clinically orientated dataset to enable benchmarking does not currently exist. The objective of the K3Da dataset is to introduce a clinically supported dataset to enable benchmarking of healthcare-based applications, methods and techniques. To the best of our knowledge, the K3Da dataset is the first such Microsoft Kinect One dataset. Further, the dataset contains one of the largest diverse sample of both young and elderly available presently. Future work will involve the expansion of the dataset to increase the number of participants and improve the age diversity.

ACKNOWLEDGEMENT

We would like to thank the participants for completing the assessments and Dr Baihua Li for discussions when planning this work.

REFERENCES

[1] 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.

[2] D. Leightley, J. Darby, B. Li, J. S. Mcphee, and M. H. Yap, “Human activity recognition for physical rehabilitation,” in *International Conference on Systems, Man and Cybernetics*, Manchester, UK, Oct 2013.

[3] 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,” *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 48, 2010.

[4] D. González-Ortega, F. J. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez, “A kinect-based system for cognitive rehabilitation exercises monitoring,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 620–631, 2014.

[5] R. Wang, G. Medioni, C. J. Winstein, and C. Blanco, “Home monitoring musculo-skeletal disorders with a single 3d sensor,” in *Computer Vision and Pattern Recognition Workshops*, June 2013.

[6] O. Celiktutan, A. Burak, C. Wolf, and B. Sankur, “Graph-based analysis of physical exercise actions,” in *ACM Multimedia Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, 2013.

Fig. 5. A figure representing a participant running in a straight line, with each colour denoting a *k*-means cluster.

[7] G. Alankus, A. Lazar, M. May, and C. Kellecher, “Towards customizable games for stroke rehabilitation,” *SIGCHI Conference on Human Factors in Computing Systems*, pp. 2113 – 2122, April 2010.

[8] J. Sung, C. Ponce, B. Selman, and A. Saxena, “Unstructured human activity detection from rgbd images,” in *International Conference on Robotics and Automation*, 2012.

[9] B. I. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, “Re-identification with rgb-d sensors,” in *First International Workshop on Re-Identification*, 2012.

[10] S. Fothergill, H. M. Mentis, P. Kohli, and S. Nowozin, “Instructing people for training gestural interactive systems,” in *CHI*, 2012, pp. 1737–1746.

[11] Carnegie Mellon University Motion Capture Dataset. The data used in this project was obtained from mocap.cs.cmu.edu. The database was created with funding from NSF EIA-0196217.

[12] V. Bloom, D. Makris, and V. Argyriou, “G3d : A gaming action dataset and real time action recognition evaluation framework,” in *Computer Vision and Pattern Recognition Workshops*, 2012.

[13] Bloom,V and Makris, D and Argyriou, V, “G3di: A gaming interaction dataset with a real time detection and evaluation framework,” in *European Conference on Computer Vision Workshop*, 2014.

[14] 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,” *Journal of Gerontology*, vol. 49, no. 2, pp. 85 – 93, March 1994.

[15] D. Podsiadlo and S. Richardson, “The timed up and go: a test of basic functional mobility for frail elderly persons,” *Journal of the American Geriatrics Society*, vol. 39, no. 2, pp. 142–148, 1991.

[16] P. Enright, “The six-minute walk test,” *Journal of Respiratory Care*, vol. 48, no. 8, pp. 783–788, 2003.

[17] M. Tinetti, “Performance-oriented assessment of mobility problems in elderly patients,” *Journal of the American Geriatrics Society*, vol. 34, no. 2, pp. 119–126, 1986.

[18] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. Rizzo, and M. Bolas, “Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor,” in *International Conference of Engineering in Medicine and Biology Society*, Aug 2011, pp. 1831–1834.

[19] 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.

[20] 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.

[21] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li, “Reliability and validity of kinect rgb-d sensor for assessing standing balance,” *Sensors*, pp. 1633–1638, 2014.

[22] R. Zhang, C. Vogler, and D. Metaxas, “Human gait recognition,” in *Computer Vision and Pattern Recognition*, 2004, pp. 18–22.

[23] 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.

[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] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,” *Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.

[26] M. Gabel, E. Renshaw, A. Schuster, and R. Gilad-Bachrach, “Full body gait analysis with kinect,” in *Engineering in Medicine and Biology Society*, Aug 2012, pp. 1964–1967.

[27] C. Newman. (2012, June) Recent advances in kinect-based action and event recognition. [Online]. Available: https://nanohub.org/resources/14383

[28] A. Sinha and K. Chakravarty, “Pose based person identification using kinect,” in *International Conference on Systems, Man and Cybernetics*, no. Oct, Oct 2013, pp. 497–503.

[29] H. Office, “Imagery library for intelligent detection systems,” *The i-Lids User Guide*, 2011.

[30] M. Marszalek, I. Laptev, and C. Schmid, “Actions in context,” in *Computer Vision and Pattern Recognition*, 2009, pp. 2929–2936.

[31] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, “Documentation mocap database HDM05,” Universität Bonn, Tech. Rep. CG-2007-2, 2007.

[32] M. Tenorth, J. Bandouch, and M. Beetz, “The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition,” in *ICCVW*, 2009.

[33] J. Wang, Z. Liu, Y. Wu, and J. Yuan, “Mining actionlet ensemble for action recognition with depth cameras,” in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[34] Kinect for Windows Software Development Kit, “http://www.microsoft.com/en-us/kinectforwindows.” [Online]. Available: http://www.microsoft.com/en-us/kinectforwindows

[35] 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.

[36] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proc. of the fifth Berkeley Symposium on Mathematrical Statistics and Probability*. University of California Press, 1967, pp. 281–297.

[37] M. Barnachon, S. Bouakaz, B. Boufama, and E. Guillou, “Ongoing human action recognition with motion capture,” *Pattern Recognition*, vol. 47, no. 1, pp. 238–247, 2013.

[38] R. Melfi, S. Kondra, and A. Petrosino, “Human activity modeling by spatio temporal textural appearance,” *Pattern Recognition Letters*, 2013.

[39] F. Zhou, F. de la Torre, and J. K. Hodgins, “Hierarchical aligned cluster analysis for temporal clustering of human motion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 582–596, 2013.

[40] M. Brusco and J. Cradit, “A variable-selection heuristic for k-means clustering,” *Psychometrika*, vol. 66, no. 2, pp. 249–270, 2001.

[41] D. Ketchen and C. Shook, “The application of cluster analysis in strategic management research: An analysis and critique,” *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996.

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

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

[44] G. W. Taylor, G. E. Hinton, and S. Roweis, “Modeling human motion using binary latent variables,” in *Advances in Neural Information Processing Systems*, 2006, p. 2007.

[45] E. Ahmad, M. H. Yap, H. Degens, and J. S. McPhee, “Atlas-registration based image segmentation of mri human thigh muscles in 3d space,” in *SPIE Medical Imaging*, 2014, pp. 90 371L–90 371L.

[46] C. RM, S. R. Boter H, and O. V. RC., “Prevalence of frailty in community-dwelling older persons: a systematic review,” *Journal of the American Geriatrics Society*, vol. 60, pp. 1487–1492, 2012.