AUTOMATING SENIOR FITNESS TESTING THROUGH GESTURE DETECTION WITH DEPTH SENSORS

A. R. Gonçalves*, †, E. R. Gouveia†, M. S. Cameirão*, †, S. Bermúdez i Badia*, †

*Madeira Interactive Technologies Institute, Portugal, † Universidade da Madeira, Portugal
afonso.goncalves@m-iti.org, erubiog@staff.uma.pt, {monica.cameirao, sergi.bermudez}@m-iti.org

Keywords: Senior Fitness Test, Older Adults, Assistive Technology, Gesture Detection, Kinect.

Abstract

Sedentarism has a negative impact on health, life expectancy and quality of life, especially in older adults. The assessment of functional fitness helps evaluating the effects of ageing and sedentarism, and this assessment is typically done through validated battery tests such as the Senior Fitness Test (SFT). In this paper we present a computer-based system for assisting and automating SFT administration and scoring in the elderly population. Our system assesses lower body strength, agility and dynamic balance, and aerobic endurance making use of a depth sensor for body tracking and multiple gesture detectors for the evaluation of movement execution. The system was developed and trained with optimal data collected in laboratory conditions and its performance was evaluated in a real environment with 22 elderly end-users, and compared to traditional SFT administered by an expert. Results show a high accuracy of our system in identifying movement patterns (>95%) and consistency with the traditional fitness assessment method. Our results suggest that this technology is a viable low cost option to assist in the fitness assessment of elderly that could be deployed for at home use in the context of fitness programs.

1 Introduction

According to current projections, nearly one third of European citizens will be aged 65 or over by 2060 [1]. In addition to these demographic changes, sedentarism is the 4th main factor in worldwide mortality, being associated with 21-25% of breast and colon cancer cases, 27% of diabetes, and 30% of ischemic strokes [2]. The combination of ageing with sedentary behaviours is a growing concern, and is putting a high strain on modern societies and their health systems. There is strong scientific evidence that regular (moderate-to-vigorous intensity) physical activity produces major and extensive health benefits in adults, particularly in older adults (aged 65 and above), as they suffer more frequently the consequences of inactivity [3].

In older adults, the assessment of multiple dimensions of physical function is commonly done using Senior Fitness Tests (SFT) [4]. These tests assess several physical parameters such as muscle strength, agility and dynamic balance, and aerobic endurance. Parameters that have high impact on people's ability to live independently which according to Fleg et al. "*is dependent largely on the maintenance of sufficient aerobic capacity and strength to perform daily activities*" [5]. The administration of SFT requires very specific training and elevated levels of concentration by a single test administrator who needs to simultaneously guide elderly through the tests, evaluate the quality of their movements, keep test scores and ensure safety.

Recent advances in information and communication technologies (ICT), specifically in the area of affordable and reliable motion tracking technology based on depth sensors, create new opportunities in the field of kinematic based assessment such as SFT. These systems can now be adapted to assess the quality of movement execution and measure task performance in a non-invasive manner. Such an approach does not only reduce the workload from the health and fitness professionals, but it also allows administration by non-experts and increases the accuracy of the results. Hence, movement kinematics and their quality can be quantified in an objective and reliable manner through machine based metrics. In this work we present a system for the automated administration of SFT that uses a Kinect V2 sensor for body tracking and gesture detectors to evaluate lower body strength, agility and dynamic balance, and aerobic endurance.

2 Related Work

In human body motion tracking, the most relevant sensing technologies are: marker-based optical systems; inertial and magnetic systems; and marker-less infrared systems.

Marker-based optical systems are the most precise with most of its usage being in motion capture applications, but the requirement of an elaborate and expensive multi-camera setup plus the use of markers distributed over the body makes them unpractical for domestic or low-cost applications.

Motion sensing through inertial or magnetic systems uses accelerometers, gyroscopes and/or magnetometers attached to the body. These systems have the advantage of being able to work independently from an external setup. However, their main disadvantages are the presence of drift errors in the measurements and the fact that the sensors need to be "worn" by the user, which can be cumbersome or unpractical. These systems were used in [6], who developed a novel automatic tracking device for weight training and calisthenics. The system uses a 3 axis-accelerometer and 3 axis-gyroscope installed in an armband. It automatically segments periods of activity, recognizes the exercise and counts the repetitions, presenting high accuracy rates and both offline and online feedback. A different study compared the adhesion of elderly to a fall prevention program when it was done through wearable sensor exergames instead of the traditional instructional booklet approach; results suggested that adherence is improved in exergames through increased levels of engagement [7]. In [8] the authors presented the lessons learned from developing games for stroke rehabilitation using the Nintendo Wii™ inertial remote, and discussed what makes games playable, fun, challenging and useful from a therapeutically perspective. However, the body of research on the assessment of fitness indicators themselves is much more reduced than for exergames. One exception is the evaluation of standing balance using the Wii Balance Board [9][10]. Comparisons between Balance Board based exergames' scores and fitness indicators [11] showed significant correlations between game scores and aerobic fitness.

Marker-less infrared systems present the lowest cost alternatives. The adoption by Microsoft of this technology in their mass-produced motion controller Kinect has contributed to the widespread availability of such sensors. These specific devices are able to estimate human body poses by analysing the 3D depth information from a scene, also requiring minimal setup and no markers. The main disadvantage is the lower accuracy of the measurements when compared with the marker-based optical systems. Still, Kinect V1 is accurate enough to be used in rehabilitation [12], and improved accuracy has been shown for Kinect V2 [13]. These devices have been widely used in research, for example, for designing full-body interactions in exergaming for older-adults [14]; for motion tracking in gait evaluation [15],[16],[17]; as a guidance, correction and scoring prototype for shoulder abduction exercises [18]; for gesture detection associated with muscle and joint pain, common in older-adults [19]; or as a tool to assist in the medical diagnosis and monitoring of Parkinson's disease through movement analysis [20].

There is, however, limited work on the assistance or automation of SFT. To our knowledge, only one case has used such an approach [21], exploring the feasibility of a home-based solution through the combined use of a Kinect and inertial sensors to detect the correct performance of the SFTs. Hence, this gap in the application to fitness assessment in elderly combined with the recent release of higher resolution depth sensors (Kinect's second generation sensor V2) which provide a more accurate estimation of 25 skeleton joints [13], offers new opportunities for innovation.

3 Methods

3.1 Fitness Tests

The Senior Fitness Test (SFT) [4] is a valuable tool for evaluating and identifying risk factors, planning and assessing training programs, educating and setting goals, and motivating clients to be more active. The SFT is designed to be easy to administer by health and fitness professionals in common community settings without extensive time (20-30 minutes), equipment or space requirements. In this study, we considered the following domains and subtests of the SFT:

1- *Lower body strength* is an important aspect of muscular fitness with respect to health, namely, in retaining proficient functioning in most daily activities, especially with advancing age. It can be measured through the 30-second Chair-stand Test that consists on counting the number of times a participant can fully stand and sit from a chair, with the arms crossed, during a 30 seconds interval [4].

2- *Aerobic endurance* or *Cardio-respiratory Fitness* (CRF) is another key component of health-related fitness. Low levels of CRF have been associated with a markedly increased risk of premature death, while high levels are associated with higher levels of habitual physical activity, and consequently with many health benefits [22]. This fitness component is assessed with the 2-minute Step Test [4]. The test consists on having the participant step in place for 2 minutes, raising the knees up to a height marker placed halfway between the knee level and hip level. The number of times each knee reaches the target height is the score of the test.

3- *Agility* (the ability to move the body and change direction quickly) and *dynamic balance* (maintaining postural stability while moving) are good predictors of recurrent falls and independent living [23]. It can be measured with the 8-foot Up-and-go Test [4]. In this test, starting from a seated position, the user stands on a "go" signal, walks 2.4 m, turns around, walks back to the chair and sits. The participant practices once and then performs two trials. The score is the fastest time to the nearest tenth of a second of the two trials.

3.2 Setup

Our system was developed using the Unity 3D game engine (Unity Technologies, San Francisco, USA) making use of Kinect's V2 plugin for Unity, Kinect's SDK and its API (Microsoft, Redmond, USA). The Kinect V2 – a RGB-Depth sensor capable of tracking 25 body joints, per person, of up to 8 people simultaneously at a frequency of 30 Hz – was placed horizontally (no tilt angle) at a height of 0.74 m and facing a wall at 4.22 m distance (Figure 1). The Visual Gesture Builder from Kinect's SDK, which uses AdaBoostTrigger and RFRProgress detection technologies, was used to train gesture detection databases.

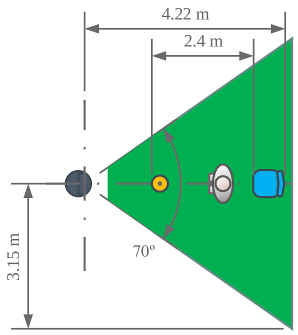

Figure 1: Top view of Kinect's V2 tracking area in green and chair and marker placement.

3.3 Gesture Detection Based Assisted SFT

In total, 5 gesture detectors were trained to compute the scores of the 3 SFT previously described, 2 discrete and 3 continuous detectors, as follows:

1. For the 30-second Chair-stand Test, 2 gesture detectors were trained: discrete detector of "arms crossed" to monitor the correct arms pose and the continuous progress "sit-stand" detector to trigger the counting of full stands.
2. To measure the 2-minute Step Test score, 2 continuous "step" progress gesture detectors were used, 1 for each leg, to monitor and count the number of steps.
3. The 8-foot Up-and-go Test timing was attained via the "sit" discrete detector acting as trigger to start and stop.

3.3.1 Discrete Detectors

The discrete gesture detectors use Adaptive Boosting (AdaBoost) [24] to construct a "strong" classifier as a linear combination of "weak" classifiers. The classifier output is a binary detection result of a gesture and its confidence level. For our system, two of these detectors were built:

1. "Arms crossed" detector, trained (ignoring the lower body: knees, ankles and feet) with 12 videos of 3 different people. Totalling 34658 labelled frames (obtained at 30 Hz) with a ratio of 1:1.132 positives to negatives;
2. "Sit" detector, trained (ignoring the arms: elbows, wrists and hands) using 8 videos of 1 person. Totalling 21300 labelled frames and ratio of 1:0.319 positives to negatives.

3.3.2 Continuous Detectors

Built with Random Forest Regression [25], the continuous detectors output is a continuous measure of progress according to a regression model constructed from training data. For our system, 3 continuous detectors were built:

1. "Sit-stand" progress detector, trained (ignoring the arms) with 12 videos of 3 people. Totalling 36126 labelled example frames;
2. Two "step" progress detectors (one for each leg), both trained (ignoring the arms) with 4 videos of 3 people. Totalling 10974 and 10976 labelled examples for left and right steps, respectively.

3.3.3 Automated SFT Score Computation

2-minute Step Test Automated Scoring: The system compares each step progress with two thresholds, one for detecting the knee raised to the target height, the other for detection of the foot on the ground. It requires the alternate triggering of right/left steps. Every time the right knee is detected above the top threshold and the opposite foot below the bottom threshold, the step count is incremented. From visual inspection of training data the detection thresholds were set to the regression values of afore mentioned situations, respectively 0.8 and 0.1.

30-second Chair-stand Test Automated Scoring: The output of the "sit-stand" progress detector is compared against two thresholds, a bottom one representing the regression output of a seating pose, and a top one representing the output of a standing pose. Every time the detector outputs alternate from sit to stand, the stand count is incremented. For this test, two different sets of thresholds were established. The first, 0.25 and 0.7, set according to the values given by the "sit-stand" detector when the subjects were considered to be on the limit of being correctly seated/standing (by SFT definition) using the training data. The second set, 0.584 and 0.8, was defined after the experiment using the evaluation by an expert as ground truth for what were considered valid sit/stand poses.

8-foot Up-and-go Test: In this chronometry test, our system measures the time that a subject takes to get up, walk 2.4 m forth and back, and sit. With the user seated the timer starts counting as soon as the sit gesture is no longer detected (user got up from the chair), the system then tracks the subject waist and once it detects that the user walked at least 2.4 m forward a validity variable is activated. The timer is stopped when sitting is once again detected if the 2.4 m walk was considered valid.

3.4 Participants

This study included 22 volunteers (15 females), 65±5.6 years old, who gave their informed consent. Participants were recruited from a physical activity program in a senior centre in Funchal, Portugal. The inclusion criteria were: (1) to be a community-dwelling older adult, aged 60 to 85 years old; (2) being able to walk independently and autonomy to perform normal everyday activities; and (3) absence of reported medical problems considered contraindications to exercise.

3.5 Experimental Protocol

Participants answered demographic and fitness questionnaires prior to the experimental session. Subsequently, they executed the 3 previously described SFT in the following order: 30-second Chair-stand Test; 8-foot Up-and-go Test; and 2-minute Step Test. These were administered and scored in real-time as follows:

- For the 30-second Chair stand test, a chair was placed against a wall facing the Kinect in the centre of its FOV.
- The 2-minutes Step test was performed with the participant stepping in place centred in Kinect's field of view at a distance of around 3 m from the sensor.
- In the 8-foot Up-and-go test, a chair was placed as in the 30-second Chair stand test. A cone was placed 2.4 m from the chair's front edge. The administrator accompanied the participant by his/her side through the trial to ensure his/her safety but without getting between him/her and the sensor.

In this study, to maximize the consistency of the assessment procedures, all the assessments were performed by the same trained fitness professional.

3.6 Data Analysis

Collected SFT score data consisted in 3 datasets: (1) "Traditional" – the standard assessment done by the professional live on site; (2) "Recordings" – a posterior assessment done by the same professional 5 weeks later by carefully replaying the Kinect recorded data (in a blinded and randomized fashion); and (3) "Automated" – the assessment done by our proposed system when replaying the Kinect recordings (emulated as real time data). Video data were

Method	Median	Percentile 25	Percentile 75
Recordings	19.00	15.75	24.00
Traditional	18.00	16.00	21.00
Lab. Trained Sys.	18.50	15.75	20.50
Exp. Trained Sys.	19.00	15.75	23.25

Table 1: 30-second Chair-stand Test scoring for the different assessment methods.

tagged by the expert for positives (correct movement patterns) and negatives (incorrect movement patterns). These data were then compared, using Mathworks software MatLab R2013b, with the detection outputs from the automated system.

A within-subjects design was used to compare the conditions. Normality of the distributions of differences was assessed using a Kolmogorov-Smirnov test. Because data deviated from normality, nonparametric statistical tests were used. For assessing the overall difference between assessments, a Friedman test was used on each dependent variable. For further pairwise comparisons, the Wilcoxon's T matched pairs signed ranks test was used. For all pairwise comparisons a Bonferroni correction was used to account for the number of comparisons. Additionally, the inter-rater reliability was measured via Intraclass Correlation. All statistical testing was done using IBM software SPSS Statistics 22.

4 Results

Here we present a comparative analysis, including the overall scoring performance metrics as well as a comparison between automated movement detection and expert tagged data. The later comparison resulted in the identification of correct detections (True Positives – TP), correct non-detections (True Negatives – TN), incorrect detections (False Positives – FP), and incorrect non-detections (False Negatives – FN). The TP and TN detection rates represent the ratio of correct detections to the respective total number of positives or negatives. FP and FN detection rates are the ratio of incorrect detections to the total number of both positives and negatives.

4.1 30-second Chair-stand Test

In the case of the 30-second Chair-stand Test, exceptionally, the system was tested with the 2 different sets of parameters. Here, “Laboratory trained system” refers to the thresholds obtained from laboratory training data and “Expert trained system” to the thresholds derived from the expert tagged experimental data, as explained in Subsection 3.3.3.

The scores, assessed as number of repetitions, are presented in Table 1. The number of counted full stands did not differ across assessments (Traditional: 18; Recordings: 19; Laboratory Trained System: 18.5; Expert Trained System: 19), $\chi^2(3) = 5.723$, $p > .05$. The Intraclass Correlation Coefficient, for absolute agreement definition, was ICC(3,1) = 0.858 and ICC(3,4) = 0.960, high values meaning that the different methods of measurement agree and are reliable between themselves.

		Lab. Trained Sys.		Exp. Trained Sys.	
		True %	**False %**	**True %**	**False %**
Stand	**Positive**	98.39	0.45	98.62	0.45
	Negative	95.16	0.78	98.90	0.67
Sit	**Positive**	95.98	0.45	99.55	0.45
	Negative	97.50	2.01	97.73	0.22

Table 3: 30-second Chair-stand Test detection rates for both the Laboratory Trained and Expert Trained systems.

Step		True %	False %
Right	**Positive**	98.67	0.16
	Negative	95.85	0.66
Left	**Positive**	96.15	0.55
	Negative	97.60	1.90

Table 4: 2-minute Step Test detection rates for the automated system.

For gesture detection, the positive gestures were fully seated and fully standing. Negatives were all other gestures, including positives of the opposite detector, *i.e.* a positive sit is a negative stand and a positive stand is a negative sit. Results of detection rates for both the Laboratory Trained System and the Expert Trained System show a high detection performance, above 95% for the Laboratory Trained and 98% for the Expert Trained, with false detections never exceeding 2% (Table 3).

4.2 2-minute Step Test

The next results encompass data from 21 (out of 22) subjects as one dataset was corrupted and excluded from the analysis. For this test, the number of counted complete steps differed significantly between methods used for counting (Traditional: 97; Recordings: 101; Automated: 96), $\chi^2(2) = 13.156$, $p < .05$ (Table 2). Interestingly, pairwise comparisons showed that the number of counted complete steps was significantly higher in the recordings than in the traditional measurements, $p = 0.001$, $T = 8.50$, effect size $r = -0.498$. However, no significant differences between the recordings and our automated system were found, $p = 0.109$, $T = 15$, $r = -0.248$, as well as between the traditional and the system method, $p = 0.359$, $T = 64.50$, $r = -0.142$. Average differences of about 4 steps (3.9%) can be seen between assessment methods. Intraclass Correlation, for the absolute agreement definition, was measured at ICC(3,1) = 0.790 and ICC(3,3) = 0.919, indicating an agreement between the measuring methods.

For individual detector performance, positive gestures consisted in the detection of knee elevation (one detector for each leg) up to the target height, and negatives all remaining cases. Step detection performance was above 95% with less than 2% false detections (Table 4).

	2-minute Step Test (nr rep)			8-foot Up-and-go (all Trials) (s)			8-foot Up-and-go (Score) (s)		
Method	**Median**	**Percentile 25**	**Percentile 75**	**Median**	**Percentile 25**	**Percentile 75**	**Median**	**Percentile 25**	**Percentile 75**
Recordings	101.00	91.50	116.50	4.70	4.13	5.10	4.65	4.08	5.03
Traditional	97.00	88.50	112.50	4.80	4.20	5.10	4.75	4.15	5.00
Automated	96.00	87.00	112.00	3.95	3.53	4.48	3.90	3.30	4.40

Table 2: Descriptive statistics for the different assessment methods and different tests.

4.3 8-foot Up-and-go Test

As explained in Section 3.1, the 8-foot Up-and-go Test requires the execution of two individual timed trials, where only the fastest is considered for the assessment. This enables us to analyse the results in two different ways. One where all the 44 timed trials for the 22 participants are considered. The second where only the fastest trial per subject is considered.

4.3.1 All trials

Measurements of execution time for each trial of the 8-foot Up-and-go Test differed significantly with the method used for timing (Traditional: 4.80; Recordings: 4.70; Automated: 3.95), $\chi^2(2) = 71.268$, $p < .05$ (Table 2). The average differences between traditional and computer mediated assessment amount to approx. 0.7 sec., which represent a 15% difference. Time measured did not significantly differ between the traditional measurements in situ and the one performed by inspection of the recordings, $p = 0.077$, $T = 173$, $r = -0.188$. However, both "traditional" and "recording" were significantly higher than our system, $p < 0.001$, $T = 0$, $r = -0.619$, and $p < 0.001$, $T = 0$, $r = -0.618$ respectively. The Intraclass Correlation for absolute agreement definition was ICC(3,1) = 0.661 and ICC(3,4) = 0.854, whereas using the consistency definition was ICC(3,1) = 0.957 and ICC(3,4) = 0.985. The low correlation values for the absolute agreement definition indicate that timing methods did not provide reliable absolute measures between themselves. However, very high values of Intraclass correlation according to the consistency definition indicate that the methods were precise, although inaccurate due to system's delay.

4.3.2 Fastest trial

The fastest measured time for performing both trials (for each participant) of the 8-foot Up-and-go Test also differed significantly with the method used for timing (Traditional: 4.75 s; Recordings: 4.65 s; Automated: 3.90 s), $\chi^2(2) = 35.877$, $p < .05$ (Table 2). The measured time did not significantly differ between the traditional measurements in situ and the one performed by inspection of the recordings, $p = 1.000$, $T = 60$, $r = 0$. However, and consistent with the previous data, time was found significantly higher in both "recordings" and "traditional" methods than by our system, with $p < 0.001$, $T = 0$, $r = -0.622$ in both cases. Intraclass Correlation for the absolute agreement definition was ICC(3,1) = 0.674 and ICC(3,4) = 0.861, when calculated using the consistency definition was ICC(3,1) = 0.962 and ICC(3,4) = 0.987. Values are identical to the previous case.

5 Discussion and Conclusions

In this work, we developed and evaluated a low cost system to support health and fitness professionals in the assessment of physical function in the elder population, with the potential to be used autonomously at home by non-experts. Not many researchers have addressed this issue, particularly in real scenarios and with end users [21]. Here we presented a comparative study with 22 elderly community dwelling participants using 3 standard SFT performed in a real world scenario. The results confirmed that the proposed system can be used to score as accurately as an expert in 2 of the 3 tests. The 8-foot Up-and-go Test presented a systematic error, underestimating time due to our experimental setup. Our system would only measure the time to perform the actions, and not the reaction time to the instructions. Thus, this error would not exist if the go signal had been given by the system itself. Low Intraclass Correlation for the absolute agreement but very high consistency values support the possibility of a systematic delay. The overall performance of the system in gesture detection was very high, with TP and TN rates over 95%. This individual gesture detection results are better (3-6%) than what was presented in [19], and while the comparison is hard to make, for exercise repetition counting they are alike to what was presented in [6].

In the case of the 30-second Chair stand Test we observed that despite very high rates for TP and TN with the Laboratory Trained System (~95%), results were further improved with the Expert Trained System (~99%). This sensitivity of score accuracy relative to the threshold values, confirmed by our data, shows the importance of using training data collected in realistic settings as opposed to laboratory conditions. This adaptation can be done by a large enough training sample in real scenarios (which is very demanding and time consuming) or alternatively by introducing a system calibration phase immediately before each individual subject assessment. The scores obtained for our system were not significantly different neither from the traditional assessment done in situ nor by inspecting the recordings. This was corroborated by a high Intraclass Correlation Coefficient (ICC(3,4) = 0.960) indicating a high absolute agreement.

We identified a high rate of FN for the left step detector in the 2-minute Step Test. The main contributor for this asymmetry was an occlusion introduced by a height marker the test administrator was using in front of the subjects' left knee. A significant difference was found between the traditional assessment and that from the recordings. This could be attributed to the high attention levels this test requires from the administrator when performed live without a tally counter. For a test administrator, it is challenging to correctly evaluate every single step validity (at rates sometimes over 2Hz), count them mentally, and ensure safety of the participant, what further supports the need for a system such as the one proposed here. In fact, our system produced scores that did not differ from those of the expert, indicating an accuracy equivalent to that of the post analysis of the video recordings.

We are currently integrating our automated fitness assessment system in a stand-alone spatial augmented reality guiding system for SFT assessment. By introducing visual (projected on the ground) and audio guidance for each test, we intend to deliver an objective assessment as well as to provide the patients with live feedback of the movement quality compared to the test goals. In a later phase, the system will be modified to provide continuous assessment during gaming exercise purposely designed to improve activity levels and overall fitness of elderly.

Acknowledgements

This work was supported by the Fundação para a Ciência e Tecnologia through the AHA project (CMUP-ERI/HCI/0046/2013) and LARSyS – UID/EEA/50009/2013. We are grateful to Ginásio de Santo António of Câmara Municipal do Funchal for their cooperation.

References

[1] European Commission, Economic and Financial Affairs, 'The 2012 Ageing Report', 2012.

[2] World Health Organization, 'WHO | The top 10 causes of death', *WHO*, 2014.

[3] World Health Organization, 'Global recommendations on physical activity for health'. 2010.

[4] R. E. Rikli and C. J. Jones, *Senior Fitness Test Manual*. Human Kinetics, 2013.

[5] J. L. Fleg, C. H. Morrell, A. G. Bos, L. J. Brant, L. A. Talbot, J. G. Wright, and E. G. Lakatta, 'Accelerated Longitudinal Decline of Aerobic Capacity in Healthy Older Adults', *Circulation*, vol. 112, no. 5, pp. 674–682, Aug. 2005.

[6] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, 'RecoFit: Using a Wearable Sensor to Find, Recognize, and Count Repetitive Exercises', in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2014, pp. 3225–3234.

[7] S. Uzor and L. Baillie, 'Investigating the Long-term Use of Exergames in the Home with Elderly Fallers', in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2014, pp. 2813–2822.

[8] G. Alankus, A. Lazar, M. May, and C. Kelleher, 'Towards Customizable Games for Stroke Rehabilitation', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2010, pp. 2113–2122.

[9] W. Young, S. Ferguson, S. Brault, and C. Craig, 'Assessing and training standing balance in older adults: A novel approach using the "Nintendo Wii" Balance Board', *Gait Posture*, vol. 33, no. 2, pp. 303–305, Feb. 2011.

[10] R. J. Reed-Jones, S. Dorgo, M. K. Hitchings, and J. O. Bader, 'WiiFit™ Plus balance test scores for the assessment of balance and mobility in older adults', *Gait Posture*, vol. 36, no. 3, pp. 430–433, Jul. 2012.

[11] B. Haddock, S. Siegel, P. Costa, S. Jarvis, N. Klug, E. Medina, and L. Wilkin, 'Fitness Assessment Comparison Between the "Jackie Chan Action Run" Videogame, 1-Mile Run/Walk, and the PACER', *Games Health J.*, vol. 1, no. 3, pp. 223–227, Apr. 2012.

[12] A. Fernández-Baena, A. Susin, and X. Lligadas, 'Biomechanical Validation of Upper-Body and Lower-Body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments', in *2012 4th International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, 2012, pp. 656–661.

[13] X. Xu and R. W. McGorry, 'The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures', *Appl. Ergon.*, vol. 49, pp. 47–54, Jul. 2015.

[14] K. Gerling, I. Livingston, L. Nacke, and R. Mandryk, 'Full-body Motion-based Game Interaction for Older Adults', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2012, pp. 1873–1882.

[15] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, 'Full body gait analysis with Kinect', in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 1964–1967.

[16] E. E. Stone and M. Skubic, 'Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing', in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 6491–6494.

[17] A. Chaaraoui, J. Padilla Lopez, and F. Florez Revuelta, 'Abnormal gait detection with RGB-D devices using joint motion history features', presented at the 1st International Workshop on Understanding Human Activities through 3D Sensors (UHA3DS'15), Ljubljana, Slovenia, 2015.

[18] A. Da Gama, T. Chaves, L. Figueiredo, and V. Teichrieb, 'Guidance and Movement Correction Based on Therapeutics Movements for Motor Rehabilitation Support Systems', in *2012 14th Symposium on Virtual and Augmented Reality (SVR)*, 2012, pp. 191–200.

[19] S. Saha, M. Pal, A. Konar, and R. Janarthanan, 'Neural Network Based Gesture Recognition for Elderly Health Care Using Kinect Sensor', in *Proceedings of the 4th International Conference on Swarm, Evolutionary, and Memetic Computing - Volume 8298*, New York, NY, USA, 2013, pp. 376–386.

[20] S. Spasojević, J. Santos-Victor, T. Ilić, S. Milanović, V. Potkonjak, and A. Rodić, 'A Vision-Based System for Movement Analysis in Medical Applications: The Example of Parkinson Disease', in *Computer Vision Systems*, Springer International Publishing, 2015, pp. 424–434.

[21] C. Chen, K. Liu, R. Jafari, and N. Kehtarnavaz, 'Home-based Senior Fitness Test measurement system using collaborative inertial and depth sensors', in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 4135–4138.

[22] American College of Sports Medicine, *ACSM's Guidelines for Exercise Testing and Prescription*. Lippincott Williams & Wilkins, 2013.

[23] D. J. Rose, *Fallproof!: A Comprehensive Balance and Mobility Training Program*. Human Kinetics, 2010.

[24] Y. Freund and R. E. Schapire, 'A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting', *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[25] L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.