Automated Timed Up & Go Test for functional decline assessment of older adults

Martin Kampel
TU-Wien, Computer Vision Lab
Vienna, Austria
martin.kampel@tuwien.ac.at

Stefan Doppelbauer
TU-Wien, Computer Vision Lab
Vienna, Austria
e0656930@student.tuwien.ac.at

Rainer Planinc
Cogvis gmbh
Vienna, Austria
planinc@cogvis.at

ABSTRACT

Human mobility is an important health indicator, especially for older adults potentially transitioning to frailty. Currently, the analysis of human mobility is based on expensive or intrusive technologies. Depth camera devices, such as the Microsoft Kinect, have been demonstrated to be a valid low-cost alternative for assessing a persons̀ mobility. In this work, mobility assessment is approached based on the automated analysis of the Timed Up & Go (TUG) test. Two methods based on depth and on skeleton data are proposed. In order to evaluate the proposed mobility analysis approaches, human mobility datasets have been acquired and manually labeled. It is shown that human mobility analysis based on off the shelf 3d sensors have the potential to assess functional decline of older adults.

KEYWORDS

3d Vision/Gait recognition/ Ambient Assisted Living

ACM Reference Format:
Martin Kampel, Stefan Doppelbauer, and Rainer Planinc. 2018. Automated Timed Up & Go Test for functional decline assessment of older adults. In *PervasiveHealth '18: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '18), May 21–24, 2018, New York, NY, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3240925.3240960

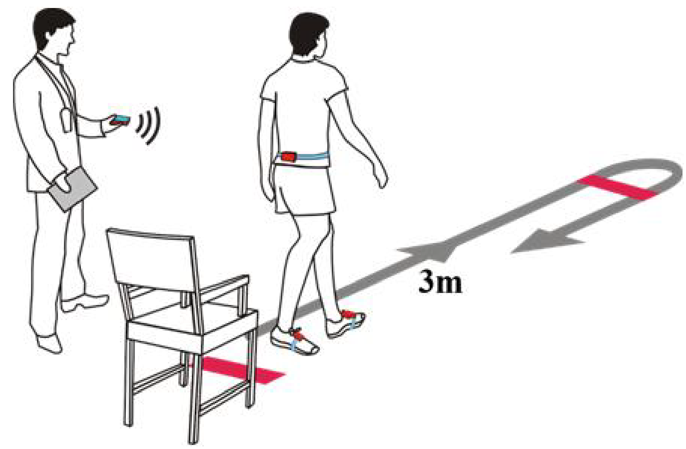


Figure 1: Illustration of the Timed Up & Go (TUG) test [42].

1 INTRODUCTION

Human populations are growing old at an accelerated speed. The world population of 65 years or older was estimated in 2004 at 461 million people [17] and in 2015 at 617 million [13]. Predictions for the next 10 years estimate an additional 236 million people aged 65 and older, further accelerating the worldwide phenomenon of older populations [13].

According to findings and evidences, frailty is not an irreversible process [26, 44]. Frail elderly who are receiving medical care have been shown to have less cognitive or functional decline, possess lower mortality rates and experience fewer falls [3]. Moreover, exercise programs and increased activity have been shown to reduce adverse health outcomes and to delay the progression of frailty [20, 28, 35]. Other potential treatments for frailty include caloric and protein support, vitamin D and reduction of polypharmacy [26].

Hence, the early diagnosis of frailty may reduce its severity. This does not just benefit the individuals, but also relieve the burden for their families and the society [3]. Moreover, dropping the need for healthcare services reduces iatrogenic disability [19], which is the effect of reduced independence after hospitalization.

The most comprehensively researched model to determine the capabilities of frail elderly persons is called comprehensive geriatric assessment (CGA) [43]. It is an interdisciplinary diagnostic process including medical, psychological, social and environmental components. However, it is a resource intensive process and therefore complex and expensive. New methods of frailty assessments are required to find equally reliable but more efficient ways for routine care [7].

An alternative approach for measuring frailty is through physical assessments. A well-known and simple test for assessing mobility is the Get-Up-and-GO test [24]. It consists of a subject standing up, walking three meters, turning around and returning to sit back down on the chair (see Figure 1). A clinical expert observes the person and rates the mobility on a scale from one to five.

Due to its subjective and imprecise scoring system, a modified version was proposed, the Timed Up & Go (TUG) test [29]. The score in the TUG test is computed based on the time taken for completing the test. There is no established norm for the TUG time cutoff for identifying frailty. Rockwood et al. [32] choose a 19-second cutoff based on the slowest 20% of TUG times from participants of the CSHA study [33], one of the largest population-based studies to measure the TUG. Other cutoff values in the literature are 20

PervasiveHealth '18, May 21–24, 2018, New York, NY, USA

ACM ISBN 978-1-4503-6450-8/18/05…$15.00
https://doi.org/10.1145/3240925.3240960

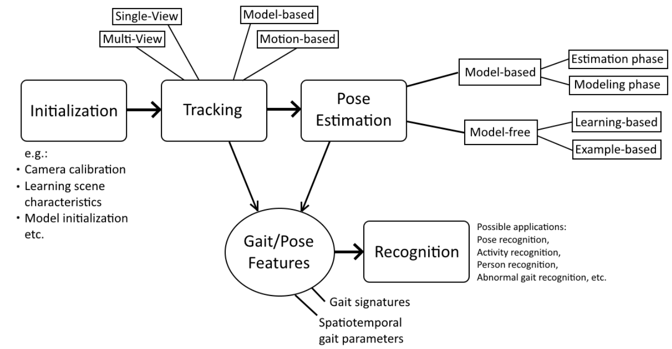


Figure 2: A general structure for human body motion analyis and its recognition applications.

seconds (for the Edmonton frail scale [34]) and 30 seconds [12]. The TUG test is widely considered to be a good indicator for frailty and future falls [37]. For example, having a high duration for the TUG test completion is most closely related (together with having any functional dependence) to needing institutionalized care after an operation [31].

The remainder of this manuscript is organized as follow: Section 2 describes the TUG analysis and presents the state of the art, Section 3 covers the methodological approach and in Section 4 results are shown. Section 5 closes the conclusions and future lines of research.

2 TUG ANALYSIS BASED ON 3D SENSING

A common classification of motion analysis techniques distinguishes three categories: Image processing or machine vision, floor sensors and wearable sensors [8, 11, 27]. Approaches based on machine vision are markerless approaches based on either color or depth data. Gait recognition refers to the identification of a person from its walking style [36]. While gait recognition is a different area of application than vision-based human gait analysis, it can be seen that the pipeline of steps applied for both gait recognition and gait analysis is very similar. The main difference between both applications is that gait recognition applies a classifier on the extracted gait parameters (also called gait signature), while this is not necessarily true for gait analysis. Due to the complexity of gait recognition, current work focus on specific topics like system initialization and camera calibration, segmentation and tracking, pose estimation, feature extraction, classification and recognition [25] For a viable gait recognition system, all of these steps have to be considered. Figure 2 shows the general structure of a motion analysis system, based on several motion analysis surveys [25, 30, 41].

The automation of TUG is approached based on various sensor modalities, e.g. color cameras [1, 38], using wearable IMUs [14] or based on ambient sensors, which are for example integrated in a chair [10]. More information for automatizing TUG tests based on different sensor technologies can be found in the survey of Sprint et al. [39].

TUG automation based on depth sensors has the advantage that TUG can be recorded unobtrusive, no additional equipment is required and there is no need for persons to use body-worn sensors while performing the TUG test.

Vernon et al. [40] investigate the potential of Kinect v1 as a medical instrument for TUG test analysis. The Kinect sensor is placed off-center of the walking area, facing the armchair. Due to their camera setup, part of the walking phase and the turning phase are not in the sensor field of view. Their method recognizes the following seven TUG events: Standing, first step, first stride, both times the participant was 2 m from the camera and the final sitting position. They identify start and end as the beginning and end of trunk joint movement. The standing event is found when the shoulder center joint reached peak height. First step and first stride are detected from ankle velocity. Based on these seven TUG events, a set of seven TUG parameters is extracted: Peak trunk flexion angle and peak trunk angular velocity during standing, first step length, first stride length, gait speed and turn time. Due to the lack of vision during the turn event, they roughly estimate turn time as the time between reaching the point 2 meters from the camera twice. Thirty individuals with stroke participate in their evaluation. Test-rest reliability is assessed using intraclass correlation coefficient, redundancy using Spearman's correlation and score prediction using multiple regression. Except trunk flexion angle, all of their Kinect-based TUG parameters are considered reliable. Most parameters are deemed redundant with TUG time, except first step length and trunk flexion angle. They conclude that utilizing the Kinect for TUG analysis provides additional information that may be predictive for changes in performance over time. The accuracy of the detected TUG event is not assessed in their work.

Cippitelli et al. [5] apply a sensor fusion approach to combine data from the skeleton model of Kinect v1 and an inertial measurement unit (IMU). They use their time synchronization method on the extraction of parameters of the TUG test. Data from Kinect is utilized for obtaining step lengths and cadence, the duration of the sit-to-stand (STS) phase, the duration of the back-to-sit (BTS) phase, the duration of the turning phase and the total tug time. Step length and cadence are estimated from peaks of the distance between both feet joints. The STS phase is detected based on the assumption that standing up requires to lean forward to put the center of mass over the feet. The movement of leaning forward is detected as a minimum followed by a maximum from the y-coordinate (height) of the head joint trajectory. The BTS phase is identified analogue by looking for a maximum followed by minimum followed by returning to default height. Turning phase is identified from shoulders and head joints. An orientation vector is computed from these three points and the angle between the orientation vector and the reference direction is used to estimate the beginning and the end of the turning phase. For their evaluation three trials are conducted by 20 healthy persons and the mean and standard deviation for all obtained parameters were computed. However, the accuracy of their approach is not evaluated. Their recorded dataset is available online [1].

Table 1 shows a summary of the described TUG automation approaches. It is observed that none of the previous approaches use Kinect v2 as a depth sensor and only one out of five approaches

[1] http://www.tlc.dii.univpm.it/blog/databases4kinect#IDTUG, Accessed 2017-07-12

Article	Particp.	Data	TUG events	Feature	Study evaluation
Lohmann et al. [22]	4 healthy, 5 elderly	Kinect v1 skeleton	TUG time, 10 TUG events	velocity and acceleration of shoulder center joint; distance of left and right shoulder joints	mean difference of detected TUG time was -0.12 compared to -0.10 for stopwatch and 1.36 for aTUG [10]
Kitsunezaki et al. [18]	6 healthy adults	Kinect v1 skeleton	start and end (TUG time)	trajectories of head and hip; cubic box around chair	average TUG time error of 0.33 seconds compared to stopwatch
Kargar et al. [16]	12 elderly participants	Kinect v1 skeleton	average step number, step duration and turn duration	distance between sensor and hip joint; distance of elbow joints	classification of fallers and non-fallers
Cippitelli et al. [5]	20 healthy persons	Kinect v1 skeleton and IMU	TUG time and duration for STS, turning and BTS	trajectories of shoulder and head joints	mean and standard deviation of TUG events; dataset available online
Vernon et al. [40]	30 elderly with stroke	Kinect v1 skeleton	7 TUG events	trajectory of trunk joint and shoulder center joint	Spearman correlation of TUG time assessed with Kinect and stopwatch was 0.99; TUG parameters redundant; error of single TUG events not assessed

Table 1: Summary of TUG analysis approaches based on Kinect.

evaluate the error of individual TUG events. 3 out of 5 papers evaluate the results with a stopwatch and only one approach uses manual video annotation as a ground truth. All approaches use the skeleton model of the Kinect pose estimation. Four out of five approaches use the trajectories of specific joints, the distance between joints or the distance between a joint and the sensor as a feature for the extraction of TUG parameters. While Lohmann et al. [22] use the distance between two joints for the detection of turn events, all other TUG parameters are obtained by using the velocity or the acceleration of the shoulder center joint.

In this work the pose estimation of the Kinect v2 SDK is utilized for gait analysis. It fits a skeleton model with 25 joints on the subject within the field of view. It is important to note that a number of alternative pose estimation methods exist and may perform better in some scenarios, e.g. when sitting. For example, Cippitelli et al. [6] propose the use of anthropometric models to locate the position of 6 joints in side view. Evaluation with a marker based system shows lower errors of the trajectories compared to the corresponding trajectories obtained with the Kinect SDK. However, in this work a view-invariant pose estimation approach is required and therefore the Kinect v2 SDK is chosen.

3 METHODOLOGY

The goal of the proposed TUG automation approach is to extract TUG time and the start and end of six TUG phases without any user input. The following assumptions have been made regarding the setup of the TUG and the sensor.
Kinect is placed perpendicular to the 3 meter walkway on a table with medium height. The distance from the walkway to the Kinect is smaller than the maximum viewing range of the sensor. Figure 3 shows a schematic placement of Kinect in a scene based on the two assumptions. While the following approach is described based on the assumption that Kinect is placed perpendicular to the walking direction of the person, it can also be applied for a different setup by updating the coordinates based on the walking direction. A preprocessing routine that detects the initial setup is required in this case.

The six TUG phases that are detected from the described approach are chair rise C^1, walk #1 W^1, turn #1 T^1, walk #2 W^2, turn #2 T^2 and sit down C^2. The start and end event of each TUG phase are detected for a total of twelve detected TUG events. It should be noted that it is possible that these events overlap, e.g. turning and sitting down may be performed simultaneously by a particular person. Moreover, there may also be a significant pause between two events, e.g. a person may look back at the chair for several seconds to prepare for sitting down after turning.

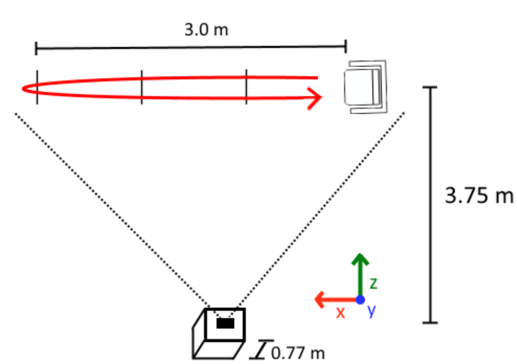


Figure 3: Scheme of the TUG recording scene; Kinect is placed perpendicular to the walking direction. *red:* x-axis, *blue:* y-axis, *green:* z-axis.

TUG automation approach. In order to detect twelve TUG events two generic functions have been defined. The first one, $\mathcal{G}(f(x), I_s, I_e)$, returns the beginning and the end of the longest consecutive occurrence of the logical condition $f(x)$ being true in the interval $[I_s, I_e]$. The second function $\mathcal{F}(f(x), n, I_s, I_e)$ returns the beginning and the end of the highest n peaks within the interval $[I_s, I_e]$. Unlike sTUG [22], events are not detected in the order they take place. Instead, events which are less challenging to detect, e.g. walking, are detected first followed by more challenging events, e.g. chair rise and sit down or turning. This has the advantage that the interval of occurrence can be narrowed down based on the information of previously detected events. Both skeleton and depth data of Kinect v2 are utilized for estimating the time to complete the TUG test and for the detection of six TUG phases.

Skeleton data. The derivative (velocity) of the trajectory of the spine shoulder joint ss' in x and y direction (ss'_x and ss'_y) is used for the detection of the walking, chair rise and sit down events. For turning, the distance between the left and the right shoulder joint in x direction $sd_x = |sl_x - sr_x|$ is used. While shoulders are close to parallel during walking, their distance in x direction (parallel to walking direction) reaches a maximum in the middle of turning. A Gaussian filter with $\sigma = 5.0$ is applied to both ss and its derivative ss'. A Median filter with window size 15 followed by a Gaussian filter with $\sigma = 5$ are applied to sd_x before the extraction of turn events. The median filter is used to remove outliers caused by noise from either shoulder joint. The start and end of the following three TUG phases are detected based on the following criteria:

- *walking:* The first walking event W^1 takes place after rising from the chair. It is caused by a steady movement away from the chair in x-direction. It is detected by searching for $W^1 = \mathcal{G}(ss'_x > T, ss_{\text{start}}, ss_{\text{end}})$ with $T = \frac{1}{2}\sigma_{ss'_x}$. $\sigma_{ss'_x}$ denotes the standard deviation of the derivative of ss in x-direction. During the second walking event W^2 the person returns to the chair. Similar to the first walking event, it is caused by a steady movement towards the chair in x-direction. It is detected by searching for $W^2 = \mathcal{G}(ss'_x < T, ss_{\text{start}}, ss_{\text{end}})$ with $T = -\frac{1}{2}\sigma_{s'_x}$.
- *chair rise and sit down:* The chair rise and sit down events mark the beginning and end of the TUG. Rising from the chair is detected as $C^1 = \mathcal{G}(ss'_y > 0, ss_{\text{start}}, W^1_{\text{start}})$. Sitting down is detected as $C^2 = \mathcal{G}(ss'_y < 0, W^2_{\text{end}}, ss_{\text{end}})$.
- *turning:* The first turn takes place between the two walks and the second turn is done right before sitting down. Turns are detected by looking for the two highest peaks in sd_x: $(T^1, T^2) = \mathcal{F}(sd_x, 2, W^1_{\text{start}}, C^2_{\text{end}})$. The peaks are ordered based on their time of occurrence and assigned accordingly.

Figure 4 illustrates detected TUG phases with the corresponding TUG movements, while Figure 5 is showing the detection of the TUG phases within the signal of the whole TUG sequence.

Depth data. For depth data, the center of mass com of the point cloud of the person is used instead of the trajectory of the joint ss. A median filter with a windows size of 1 second or 30 frames and a Gaussian filter with $\sigma = 5.0$ is applied on com'. A median filter with a large window size is used since com' is less stable and more noisy than ss' due to potential errors in the silhouette extraction.

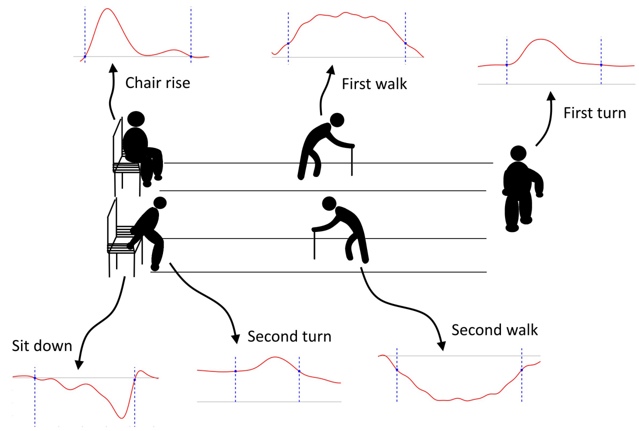


Figure 4: Detection of 6 TUG phases: enlarged snippets of detected TUG phases with the corresponding TUG movements.

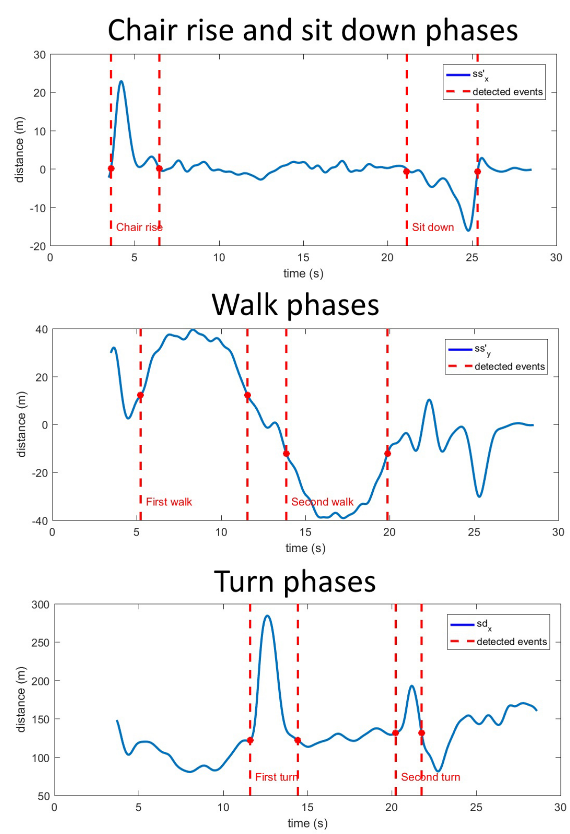


Figure 5: TUG phases within the signal of a whole TUG sequence. The right side shows enlarged snippets of detected TUG phases with the corresponding TUG movements the occurrence of detected TUG phases within the signal of a whole TUG sequence is shown..

For the detection of turns the movement direction during the first walk $\vec{d_W}$ and the normalized ground projected movement direction from frame to frame $\vec{d}(k)$ are computed.

$$\vec{d}(k) = \|com_{xy}(k+l) - com_{xy}(k)\| \tag{1}$$

$com_{xy}(k)$ denotes the ground-projected position of the center of mass com and l is a constant with $l = 2$ being used. Then the function $\Delta(k) \in [-1, 1]$ is computed, which represents the similarity of the current walking direction with the walking direction of the first walk.

$$\Delta(k) = \vec{d_W} \cdot \vec{d}(k) \tag{2}$$

$\Delta(k) = 1$ denotes the same walking direction as the first walk and $\Delta(k) = -1$ denotes the opposite walking direction. A median filter with window size 30 is applied on Δ before the extraction of turn events. Based on the center of mass trajectory com and the function for indicating turns Δ, the following criteria are used for the detection of TUG events:

- *walking:* The detection of walking events is done analogue to the skeleton data. The only difference is that com'_x is used instead of ss'_x.
- *chair rise and sit down:* Like walking, chair rise and sit down events are detected analogue to the skeleton data. Besides using com'_y instead of ss'_y a higher threshold is used to better cope with noise. Chair rise is detected as $C^1 = \mathcal{G}(com'_y > T_1, com_{\text{start}}, \frac{W^1_{\text{start}} + W^1_{\text{end}}}{2})$ with $T_1 = \frac{1}{2}\sigma_{com'_y}$. Sitting down is detected as $C^2 = \mathcal{G}(com'_y < T_2, \frac{W^2_{\text{start}} + W^2_{\text{end}}}{2}, com_{\text{end}})$ with $T_2 = -\frac{1}{2}\sigma_{com'_y}$.
- *turning:* Detecting turning events from a single trajectory is an ill-posed problem, since there is no or little movement when a person turns while standing on the same spot. In order to detect the first turn, the transition from the walking direction of the first walk to the walking direction of the second walk has to be found. This is achieved by $(-, T^1_{\text{start}}) = \mathcal{G}(\Delta > T_1, \Delta_{\text{start}}, \Delta_{\text{end}})$ with $T_1 = 0.9$ and $(T^1_{\text{end}}, T^2_{\text{start}}) = \mathcal{G}(\Delta < T_2, \Delta_{\text{start}}, \Delta_{\text{end}})$ with $T_2 = -0.9$. The end of the first turn is detected when the current walking direction reaches the walking direction of the second turn ($\Delta < -0.9$). The beginning of the second turn is found when the walking direction differs again from the walking direction of the second walk, which marks the end of the ($\Delta < -0.9$) condition. The end of the second turn is found when Δ reaches zero after the event T^2_{start} or the end of Δ is reached.

4 RESULTS

The proposed TUG automation approaches are based on depth and skeleton data are evaluated on self acquired elderly TUG test dataset[2]. No suitable datasets are publicly available, however three datasets have been acquired as a part of this work: Cai [4] discusses 46 RGBD datasets for object detection and tracking, human activity analysis, object and scene recognition, simultaneous localization and mapping and hand gesture analysis. Only three datasets are reviewed for human activity analysis and none of the reviewed datasets are relevant for this work. Firman [9] reviews 96 RGBD datasets for various applications and two can be considered relevant for this work, Kinect 3D Active and TST TUG.

Kinect 3D Active [21] includes recordings of standardized medical assessments including the TUG test. Participants are between 18 and 81 years old. TST TUG [5] contains a TUG test performed by 20 healthy participants and is available online [3]. While the authors use this data to extract several TUG parameters which are also relevant for this work, e.g. time duration of the sit-to-stand phase, step length, cadence and TUG time, they do not evaluate the accuracy of their results. Instead, only mean and standard deviation of each parameter is provided in the paper. Moreover, their dataset does not include persons with gait pathologies, making it unsuitable to assess the eligibility in real-world usage.

Kinect 3D Active and TST TUG do not have a labeling for TUG events and the accuracy of TUG event extraction is not yet assessed on these datasets. Since no comparable results exist for these datasets and manual labeling would be necessary, these datasets are not used for the evaluation of the TUG automation approach presented in section 3. For RGBD gait datasets, existing public datasets, e.g. TUM Gait [15] or DGait Database [2], are intended for gait recognition. While they include a large number of walking sequences, they do not have a labeling for the accuracy of spatiotemporal gait parameters.

For acquiring our evaluation dataset a total of eleven participants, 7 female and 4 male, volunteer to participate. Participants are between 85 and 95 years old with a mean age of $\mu_{\text{age}} = 89.3$ years and a standard deviation of $\sigma_{\text{age}} = 3.6$ years. All participants are asked to perform the TUG test once. Failed TUG tests, e.g. due to unclear instructions, are repeated until the test was performed smoothly. 9 out of 11 participants prefer to use a walking aid for walking and during the performance of the TUG test. No participant requires physical assistance in order to complete the TUG test and the two walking sequences. The experimental setup used for the recording and several sample images are shown in Figure 6. A Kinect v2 sensor was placed in front of the TUG walking path on a table with 77 cm height around 3.5 meters away from the path.

For the purpose of evaluation, the start and end events of six TUG phases are labeled using manual video inspection.

TUG time is estimated as the elapsed time between the first and last TUG event. Additionally, a stopwatch is used during the recording of the TUG test and manually obtained TUG times are compared with the automatically estimated TUG times.

4.1 Skeleton Timed Up and Go

For the purpose of evaluation, the TUG automation method based on the original approach from Lohmann [22] is called Skeleton Timed Up and Go (sTUG). The original approach detects a total of ten TUG events during the test, while our approach is based on nine TUG events. sTUG uses thresholds for the acceleration of joint trajectories to detect the events *start moving* M_s, *start walking* W_s, *end uprising* U_e, *start lowering* D_s and *end moving* M_e. The events *start rotating* R_s and *end rotating* R_e are detected based on the distance of the two shoulder joints. The events *start accelerating*

[2]Dataset will made available in May18

[3]http://www.tlc.dii.univpm.it/blog/databases4kinect#IDTUG, Accessed 2017-07-12

(a) Picture of the TUG setup. ***red:*** **x-axis,** ***green:*** **z-axis,** ***blue:*** **y-axis**

(b) Elderly participants during their performance of the TUG test.

Figure 6: TUG recording setup (left image) and samples during recording (right image). Red stick tape was used to mark the line 3 meters away from the armchair.

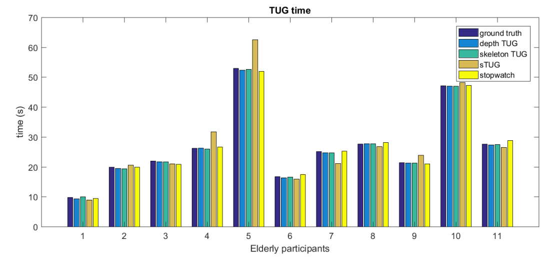


Figure 7: Comparison of TUG times obtained from depth TUG, skeleton TUG and sTUG with ground truth and stopwatch results for all eleven participants.

A_s and *end accelerating* A_e, which represent the end of the first walk and the beginning of the second walk, are detected based on thresholding the velocity in walking direction. In order to evaluate both walk times, an additional event *end walking* W_e is detected based on their approach. W_e is detected as the end of the second walk using the same method they used for detecting A_s and A_e.

The paper of Lohmann et al. [22] has inconsistencies and not all details are specified in their paper. Therefore the following assumptions have been made for the implementation:

- The usage of their coordinate system is inconsistent: While they originally define y-axis as walking direction and z-axis as upward vector, walking direction is later referred to as z-axis and upward direction is referred to as y-axis during their definition of gait events. Therefore y-axis and z-axis are interchanged, as it would not make sense otherwise. This concerns the detection of the events *start moving* M_s, *end uprising* D_s, *start walking* W_s, *start accelerating* A_s, *end accelerating* A_e, *end moving* M_e and start lowering D_s. The usage of x-axis is consistent with their original coordinate system definition.
- For the events *start rotating* and *end rotating* minimum and maximum are interchanged. This is assumed to be a mistake in the paper, since the derivation of a trough is trough followed by a peak, not the other way around.
- Both *end uprising* and *start lowering* are referred to as D_s. This is assumed to be a mistake in the paper and U_e is estimated based on the definition given in the description for the *end uprising* event.
- The definition of *end accelerating* A_e is equivalent with the definition of *start accelerating* A_s, *end accelerating* is even referred to as A_s in the paper. This is assumed to be a mistake in the paper and the definition for *end accelerating* is assumed to be $A_e = tmin(sc_y, T_m, S_e, m)$.
- No signals filters are specified, therefore the same filters have been applied as in the proposed method.
- Thresholds l_1, l_2 and l_3 are not specified and set to $\frac{3}{4}\sigma_{ss}$ with the exception of the threshold for the *end uprising* event U_e, which was set to $\frac{3}{20}\sigma_{ss}$. The only thresholds that is specified was the mean walking speed used for the detection of *start accelerating* A_s and *end accelerating* A_e.
- Shoulder center trajectory *sc* is replaced with spine shoulder trajectory *ss*. This is necessary since the authors used Kinect v1 and the skeletal model differs slightly compared to Kinect v2.
- The equivalent events for the start and end of turn #2 T_s^2 and T_e^2 are not defined in their paper so these events are not implemented.

4.2 TUG analysis results

The proposed approach based on skeleton data is referred to as Skeleton TUG and the proposed approach based on depth data as Depth TUG.

Figure 7 shows the results of the obtained TUG times, manually labeled ground truth and stopwatch in seconds for all 11 participants. The most notable outliers are participants #4, #5 and #7 for the sTUG method. The average error for TUG time is 0.294 s for Depth TUG, 0.227 s for Skeleton TUG, 0.536 s for the stopwatch and 2.549 s for sTUG.

Table 2 shows the mean absolute error for TUG time and the 6 TUG phases. Precision and recall for total TUG time and TUG phases are computed based on the overlap of the groundtruth interval and the detected interval. The resulting F1 scores are 0.824 for Depth TUG, 0.857 for Skeleton TUG and 0.695 for sTUG.

The mean absolute error of single TUG events and its standard deviation are shown in Table 3. The same data is also illustrated in Figure 8. From the 10 TUG events extracted from the three approaches, only the end of the first walking sequence W_e^1 has a lower error for sTUG compared to skeleton TUG and depth TUG. For all other events both the mean error and standard deviation are similar or higher than Depth TUG or Skeleton TUG.

Table 4 shows the average difference between detected TUG events and the corresponding ground truth. It shows that the majority of TUG events are detected too early for sTUG. For Depth

Mean absolute error, precision and recall of TUG phases

		TUG time	Chair rise time	Walk #1 time	Turn #1 time	Walk #2 time	Turn #2 time	Sit down time
Depth TUG	mean error (s)	0.294	0.182	1.324	1.227	2.364	2.222	0.730
	Precision	0.998	0.874	0.900	0.929	0.729	0.467	0.783
	Recall	0.986	0.877	0.938	0.65	0.995	0.473	0.943
Skeleton TUG	mean error (s)	0.227	1.024	0.903	1.061	1.224	2.182	1.570
	Precision	0.997	0.647	0.961	0.793	0.831	0.832	0.593
	Recall	0.990	0.928	0.906	0.871	0.983	0.759	0.952
sTUG	mean error (s)	2.503	2.003	-	1.342	-	-	2.097
	Precision	0.941	0.294	0.809	0.769	0.719	-	0.438
	Recall	0.947	0.574	0.961	0.772	0.790	-	0.345

Table 2: Mean absolute error, precision and recall for TUG time and the six TUG phases obtained from depth TUG, skeleton TUG and sTUG [22].

Absolute mean error (s) and its standard deviation for 12 TUG events

	Depth TUG	Skeleton TUG	sTUG
Chair rise start	0.22 ± 0.16	0.16 ± 0.13	2.05 ± 3.80
Chair rise end	0.26 ± 0.18	1.10 ± 0.65	3.58 ± 4.42
Walk #1 start	0.65 ± 1.13	0.49 ± 0.89	2.945.52
Walk #1 end	0.89 ± 1.01	0.68 ± 1.03	0.46 ± 0.39
Turn #1 start	0.79 ± 0.40	0.61 ± 0.72	1.31 ± 2.87
Turn #1 end	0.53 ± 0.39	0.82 ± 0.62	0.79 ± 0.86
Walk #2 start	0.35 ± 0.31	0.37 ± 0.25	0.77 ± 0.83
Walk #2 end	2.08 ± 1.57	1.04 ± 0.80	2.01 ± 3.27
Turn #2 start	1.64 ± 1.46	1.33 ± 1.38	-
Turn #2 end	3.02 ± 1.87	1.13 ± 1.52	-
Sit down start	0.70 ± 0.64	1.61 ± 1.15	1.27 ± 1.07
Sit down end	0.094 ± 0.089	0.13 ± 0.08	1.51 ± 1.07

Table 3: Absolute mean error and standard deviation in seconds for all detected TUG events

TUG, the only event which has a clear bias is the end of walk #2 W_e^2, which is consistently detected too late. For Skeleton TUG, the end of chair rise C_e^1 and the end of walk #2 W_e^2 are detected too late by 1 second on average. The start of the sit down event C_s^2 is detected too early by around 1.5 seconds.

4.3 TUG analysis discussion

The average time for sit-to-stand is 1.7 s, for walk #1 3.3 s, for turn #1 1.8 s, for walk #2 3.4 s, for turn #2 1.6 s and for sitting down 1.0 s. This observation is consistent with the study by Manckoundia et al. [23], who show that sit-to-stand takes longer than stand-to-sit for elderly persons.

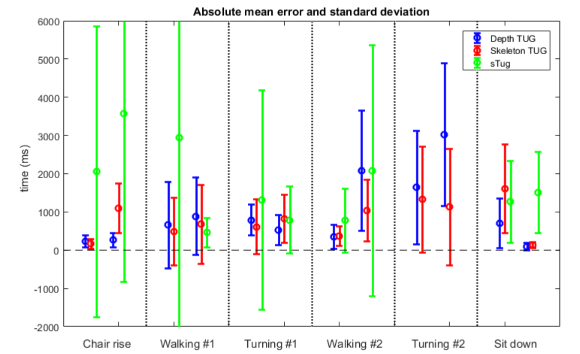


Figure 8: Absolute mean error and standard deviation in ms for 12 TUG events. The dotted line separates the five gait phases. For each phase, the first three values represent the error of the start event of the phase and the last three values represent the error of the end event of the phase.

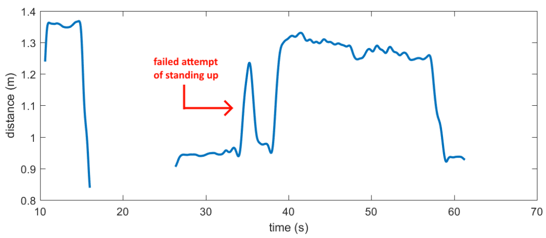
Average signed difference (s) of 10 detected TUG events

	C_s^1	C_e^1	W_s^1	W_e^1	T_s^1	T_e^1	W_s^2	W_e^2	C_s^2	C_e^2
Depth TUG	0.21	0.23	0.04	0.22	0.52	-	-	2.08	-	-
						0.08	0.28		0.45	0.05
Skeleton	0.1	1.1	0.43	-	-	0.47	-	1.01	-	-
TUG				0.19	0.08		0.17		1.57	0.10
sTUG [22]	-	-	-	0.18	-	-	-	-	0.12	-
	1.81	0.84	2.83		1.02	0.53	0.52	1.12		0.84

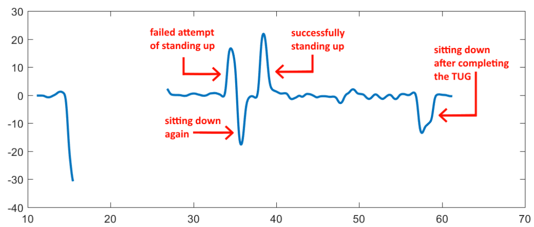
Table 4: Average signed difference of detected TUG event times and ground truth in seconds. The following abbreviations have been used: Chair rise start C_s^1 and end C_e^1, walk #1 start W_s^1 and end W_e^1, turn #1 start T_s^1 and end T_e^1 and sit down start C_s^2 and end C_e^2.

One reason for the poor performance of sTUG on this dataset is that their approach works on the assumption, that the person does not move before and after the TUG. However, this assumption is not necessarily true when working with elderly persons, who often need multiple attempts to perform a certain movement or additional instruction before rising from the chair. Moreover, Kinect skeleton tracking is noisy and not reliable when the person is sitting. This commonly causes sTUG to detect the *start moving* event too early. It can be seen that there are a lot of acceleration spikes at the beginning, but the participant only starts to rise from the chair at approximately 12 seconds for participant #4 and approximately 20 seconds for participant #5. Therefore joint acceleration as a sole feature as used by STUG is not robust enough for the detection of TUG events.

Moreover, sTUG detects the beginning of TUG as the point where the person starts to lean forward to move his center of gravity over his feet, followed by the person lifting the body from the chair. However, this approach does not consider that the person might



(a) *Blue line:* Position of spine shoulder joint in y-direction (height).



(b) *Blue line:* Velocity of spine shoulder joint in y-direction (upwards).

Figure 9: An older person needing two attempts of standing up. The failed attempt can be easily identified in the joint height trajectory (top image), but not in its derivative (bottom image).

move his upper body back and forth prior to starting the TUG. Therefore it is necessary to verify that the person stands up from the chair shortly after leaning forward. An approach to achieve this is the cubic space approach used by Kitsunezaki et al. [18]. They count the number of joints inside a virtual cubic space around the chair in order to find the transition from sitting on the chair to standing up and walking away from the chair.

Another reason why sTUG tends to identify the *start moving* event too early is that an examiner starts the stopwatch only when the patient's hips are no longer touching the chair [18] instead of when the patient is leaning forward. As pointed out by Kitsunezaki et al. [18], it is challenging to detect standing up or sitting down movements due to the very slow movements of old adults. Even the proposed methods are probably not robust enough to perform well on a larger dataset. One option to increase the robustness of the approach is to additionally consider the trajectory of the joint position in addition to the velocity. By looking at the current position of the person, the current TUG state can be verified and certain state transitions are only allowed if the person is in the correct position. Moreover, this can help to resolve cases like the one depicted in Figure 9. One person attempts to rise from the chair, but is unsuccessful and sits down again before making a second attempt. When only a derivative of the joint trajectory is considered, the resulting peaks can easily be falsely classified.

In addition to the increased robustness, future methods should also consider the additional challenges present in a real-world setting. It is common during a TUG test that a second person is walking around, e.g. a medical assistant being nearby to assist the patient if necessary [18]. This creates additional challenges for consistently tracking the person and a second sensor might be be necessary to consistently resolve this problem. Moreover, it is common for older adults to execute the TUG incorrectly. Future methods should also be able to determine whether the TUG is executed correctly or whether certain movements are missed and the trial should be repeated.

5 CONCLUSION

In this work, approaches for the analysis of the Timed Up & Go (TUG) test were presented and evaluated on the acquired elderly TUG test dataset. The experiments in this work showed, that the TUG test can be automatically performed by a single Kinect sensor. Moreover, they showed that TUG analysis based on joint acceleration as a sole feature is not able to robustly measure TUG times and detect TUG events. It was shown that the proposed approach worked well on the acquired TUG test dataset. However, it may not be robust enough for future real-world datasets that include failed movement attempts or other atypical behavior. Potential improvements of the proposed approach should consider the position of the person in relation to the chair instead of solely considering the movement of the person. This ensures that the movement state of the person can be continuously tracked and mistakes - caused by the person's movement being different than expected - can be avoided. Furthermore, a large, manually labeled TUG test dataset is needed for the evaluation of future TUG automation approaches.

ACKNOWLEDGMENTS

This work was partly supported by the European Active and Assisted Living Prgramme (AAL-JP) and the Austrian Research Promotion Agency (FFG) under grant AAL-2015-056.

REFERENCES

[1] Dounia Berrada, Mario Romero, Gregory Abowd, Marion Blount, and John Davis. 2007. Automatic administration of the get up and go test. In *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*. ACM, 73–75.
[2] Ricard Borràs, Àgata Lapedriza, and Laura Igual. 2012. Depth information in human gait analysis: an experimental study on gender recognition. In *International Conference Image Analysis and Recognition*. Springer, 98–105.
[3] Fanny Buckinx, Yves Rolland, Jean-Yves Reginster, Céline Ricour, Jean Petermans, and Olivier Bruyère. 2015. Burden of frailty in the elderly population: perspectives for a public health challenge. *Archives of Public Health* 73, 1 (2015), 19.
[4] Ziyun Cai, Jungong Han, Li Liu, and Ling Shao. 2017. RGB-D datasets using microsoft kinect or similar sensors: a survey. *Multimedia Tools and Applications* 76, 3 (2017), 4313–4355.
[5] Enea Cippitelli, Samuele Gasparrini, Ennio Gambi, Susanna Spinsante, Jonas Wåhslény, Ibrahim Orhany, and Thomas Lindhy. 2015. Time synchronization and data fusion for RGB-depth cameras and inertial sensors in AAL applications. In *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 265–270.
[6] Enea Cippitelli, Samuele Gasparrini, Susanna Spinsante, and Ennio Gambi. 2015. Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view. *Sensors* 15, 1 (2015), 1417–1434.
[7] Andrew Clegg, John Young, Steve Iliffe, Marcel Olde Rikkert, and Kenneth Rockwood. 2013. Frailty in elderly people. *The Lancet* 381, 9868 (2013), 752–762.
[8] Mohammad Omar Derawi, Claudia Nickel, Patrick Bours, and Christoph Busch. 2010. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*. IEEE, 306–311.
[9] Michael Firman. 2016. RGBD datasets: Past, present and future. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 19–31.
[10] Thomas Frenken, Björn Vester, Melina Brell, and Andreas Hein. 2011. aTUG: Fully-automated timed up and go assessment using ambient sensor technologies. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 55–62.

[11] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. 2007. Gait authentication and identification using wearable accelerometer sensor. In *2007 IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 220–225.

[12] Asma Hassani, Alexandre Kubicki, Vincent Brost, and Fan Yang. 2014. Real-time 3D TUG test movement analysis for balance assessment using Microsoft Kinect. *Aiia2014. Di. Unipi. It* (2014), 1–10.

[13] Wan He, Daniel Goodkind, and Paul R Kowal. 2016. *An aging world: 2015*. United States Census Bureau.

[14] Yuji Higashi, Kenichi Yamakoshi, Toshiro Fujimoto, Masaki Sekine, and Toshiyo Tamura. 2008. Quantitative evaluation of movement using the timed up-and-go test. *IEEE Engineering in Medicine and Biology Magazine* 27, 4 (2008), 38–46.

[15] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. 2014. The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation* 25, 1 (2014), 195–206.

[16] B Amir H Kargar, Ali Mollahosseini, Taylor Struemph, Wilson Pace, Rodney D Nielsen, and Mohammad H Mahoor. 2014. Automatic measurement of physical mobility in Get-Up-and-Go Test using kinect sensor. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3492–3495.

[17] Kevin G Kinsella and David R Phillips. 2005. *Global aging: The challenge of success*. Vol. 60. Population Reference Bureau Washington, DC.

[18] Naofumi Kitsunezaki, Eijiro Adachi, Takashi Masuda, and Jun-ichi Mizusawa. 2013. KINECT applications for the physical rehabilitation. In *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*. IEEE, 294–299.

[19] C Lafont, S Gérard, T Voisin, M Pahor, B Vellas, et al. 2011. Reducing ÃćâĈňÅŞiatrogenic disabilityÃćâĈňÂİ in the hospitalized frail elderly. *The journal of nutrition, health & aging* 15, 8 (2011), 645–660.

[20] Francesco Landi, Angela M Abbatecola, Mauro Provinciali, Andrea Corsonello, Silvia Bustacchini, Luca Manigrasso, Antonio Cherubini, Roberto Bernabei, and Fabrizia Lattanzio. 2010. Moving against frailty: does physical activity matter? *Biogerontology* 11, 5 (2010), 537–545.

[21] Daniel Leightley, Moi Hoon Yap, Jessica Coulson, Yoann Barnouin, and Jamie S McPhee. 2015. Benchmarking human motion analysis using Kinect One: an open source dataset. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1–7.

[22] Okko Lohmann, Thomas Luhmann, and Andreas Hein. 2012. Skeleton timed up and go. In *2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 1–5.

[23] P Manckoundia, F Mourey, P Pfitzenmeyer, and C Papaxanthis. 2006. Comparison of motor strategies in sit-to-stand and back-to-sit motions between healthy and AlzheimerÃćâĈňâĎćs disease elderly subjects. *Neuroscience* 137, 2 (2006), 385–392.

[24] S Mathias, US Nayak, and Bernard Isaacs. 1986. Balance in elderly patients: the" get-up and go" test. *Archives of physical medicine and rehabilitation* 67, 6 (1986), 387–389.

[25] Thomas B Moeslund and Erik Granum. 2001. A survey of computer vision-based human motion capture. *Computer vision and image understanding* 81, 3 (2001), 231–268.

[26] John E Morley, Bruno Vellas, G Abellan van Kan, Stefan D Anker, Juergen M Bauer, Roberto Bernabei, Matteo Cesari, WC Chumlea, Wolfram Doehner, Jonathan Evans, et al. 2013. Frailty consensus: a call to action. *Journal of the American Medical Directors Association* 14, 6 (2013), 392–397.

[27] Alvaro Muro-de-la Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. 2014. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* 14, 2 (2014), 3362–3394.

[28] Matthew J Peterson, Carol Giuliani, Miriam C Morey, Carl F Pieper, Kelly R Evenson, Vicki Mercer, Harvey J Cohen, Marjolein Visser, Jennifer S Brach, Stephen B Kritchevsky, et al. 2009. Physical activity as a preventative factor for frailty: the health, aging, and body composition study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 64, 1 (2009), 61–68.

[29] Diane Podsiadlo and Sandra Richardson. 1991. The timed Up & Go: a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society* 39, 2 (1991), 142–148.

[30] Ronald Poppe. 2007. Vision-based human motion analysis: An overview. *Computer vision and image understanding* 108, 1 (2007), 4–18.

[31] Thomas N Robinson, Jeffrey I Wallace, Daniel S Wu, Arek Wiktor, Lauren F Pointer, Shirley M Pfister, Terra J Sharp, Mary J Buckley, and Marc Moss. 2011. Accumulated frailty characteristics predict postoperative discharge institutionalization in the geriatric patient. *Journal of the American College of Surgeons* 213, 1 (2011), 37–42.

[32] Kenneth Rockwood, Melissa Andrew, and Arnold Mitnitski. 2007. A comparison of two approaches to measuring frailty in elderly people. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62, 7 (2007), 738–743.

[33] Kenneth Rockwood, Xiaowei Song, Chris MacKnight, Howard Bergman, David B Hogan, Ian McDowell, and Arnold Mitnitski. 2005. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal* 173, 5 (2005), 489–495.

[34] Darryl B Rolfson, Sumit R Majumdar, Ross T Tsuyuki, Adeel Tahir, and Kenneth Rockwood. 2006. Validity and reliability of the Edmonton Frail Scale. *Age and ageing* 35, 5 (2006), 526–529.

[35] Laurence Z Rubenstein, Karen R Josephson, Peggy R Trueblood, Steven Loy, Judith O Harker, Fern M Pietruszka, and Alan S Robbins. 2000. Effects of a group exercise program on strength, mobility, and falls among fall-prone elderly men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 55, 6 (2000), M317–M321.

[36] Suvarna Shirke, SS Pawar, and Kamal Shah. 2014. Literature review: Model free human gait recognition. In *2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)*. IEEE, 891–895.

[37] Anne Shumway-Cook, Sandy Brauer, and Marjorie Woollacott. 2000. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. *Physical therapy* 80, 9 (2000), 896–903.

[38] Zoran Skrba, Brian O'Mullane, Barry R Greene, Cliodhna Ni Scanaill, Chie Wei Fan, Aaron Quigley, and Paddy Nixon. 2009. Objective real-time assessment of walking and turning in elderly adults. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 807–810.

[39] Gina Sprint, Diane J Cook, and Douglas L Weeks. 2015. Toward Automating Clinical Assessments: A Survey of the Timed Up and Go. *IEEE reviews in biomedical engineering* 8 (2015), 64–77.

[40] Stephanie Vernon, Kade Paterson, Kelly Bower, Jennifer McGinley, Kimberly Miller, Yong-Hao Pua, and Ross A Clark. 2015. Quantifying individual components of the timed up and go using the kinect in people living with stroke. *Neurorehabilitation and neural repair* 29, 1 (2015), 48–53.

[41] Jessica JunLin Wang and Sameer Singh. 2003. Video analysis of human dynamicsÃćâĈňâĂİa survey. *Real-time imaging* 9, 5 (2003), 321–346.

[42] Wei-Hsin Wang, Pau-Choo Chung, Guo-Liang Yang, Chien-Wen Lin, Yu-Liang Hsu, and Ming-Chyi Pai. 2015. An inertial sensor based balance and gait analysis system. In *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*. IEEE, 2636–2639.

[43] TJ Welsh, Adam L Gordon, and JR Gladman. 2014. Comprehensive geriatric assessment–a guide for the non-specialist. *International journal of clinical practice* 68, 3 (2014), 290–293.

[44] Qian-Li Xue. 2011. The frailty syndrome: definition and natural history. *Clinics in geriatric medicine* 27, 1 (2011), 1–15.