# Automation of the Timed-Up-and-Go Test Using a Conventional Video Camera

Patrick Savoie, James A. D. Cameron, *Student Member, IEEE,* Mary E. Kaye, *Life Member, IEEE,* and Erik J. Scheme, *Senior Member, IEEE*

***Abstract*—The Timed-Up-and-Go (TUG) test is a simple clinical tool commonly used to quickly assess the mobility of patients. Researchers have endeavored to automate the test using sensors or motion tracking systems to improve its accuracy and to extract more resolved information about its sub-phases. While some approaches have shown promise, they often require the donning of sensors or the use of specialized hardware, such as the now discontinued Microsoft Kinect, which combines video information with depth sensors (RGBD). In this work, we leverage recent advances in computer vision to automate the TUG test using a regular RGB video camera without the need for custom hardware or additional depth sensors. Thirty healthy participants were recorded using a Kinect V2 and a standard video feed while performing multiple trials of 3 and 1.5 meter versions of the TUG test. A Mask Regional Convolutional Neural Net (R-CNN) algorithm and a Deep Multitask Architecture for Human Sensing (DMHS) were then used together to extract global 3D poses of the participants. The timing of transitions between the six key movement phases of the TUG test were then extracted using heuristic features extracted from the time series of these 3D poses. The proposed video-based *vTUG* system yielded the same error as the standard Kinect-based system for all six key transitions points, and average errors of less than 0.15 seconds from a multi-observer hand labeled ground truth. This work describes a novel method of video-based automation of the TUG test using a single standard camera, removing the need for specialized equipment and facilitating the extraction of additional meaningful information for clinical use.**

***Index Terms*—Computer vision, automation, deep learning, R-CNN, Timed-up-and-go, gait analysis, neural networks, pose estimation**

## I. Introduction

The Timed-Up-and-Go (TUG) test [1] is a common clinical assessment used to quickly determine the mobility of elderly individuals. Since it is a simple and general test of movement, it has been widely employed with individuals with balance and gait impairments, or even cognitive decline, to assess their risk of falling [2], [3], [4]. The individual being tested starts in a seated position, stands, walks a set distance (typically 3 meters), turns, walks back, and sits back down. The total time taken for a user to complete the test, as measured by an observer using a stop-watch, is often used as a metric to describe a person's overall mobility [5], [6], [7]. It has been shown that changes in the relative timing of the individual subtasks of the TUG test (time seated, walking, and turning) may provide additional important information about the mobility of the individual [2], [8], [9]. In practice, however, this information has thus far been difficult to extract due to the inaccuracy and inconvenience of conventional hand-timed methods.

In recent years, there has been substantial interest from the research and health care communities in automating the TUG test. Automating the process could help to standardize the mobility assessment of elderly individuals, ensuring consistency across geography, clinical training, and time, and facilitate the extraction of these additional timings. Sensors of various modalities have been used to automate the test, each of them presenting their own set of advantages and challenges [2], [10].

Wearable inertial sensors (IMUs) comprise a large amount of the related literature [2], [11], [12]. IMUs are inexpensive, lightweight, and are not confined to measuring gait in a fixed volume, or within a line of site, making them suitable for a variety of human movement monitoring tasks. However, they also require setup and calibration for every new patient or test, must be recharged periodically, and require compensation from natural drift and interference from electromagnetic interference [2], [11], [13]. Smart phones have been leveraged as alternatives given that most now come with natively built-in IMUs, but remain limited by similar problems and in practice by placement, size, and number of sensors [2], [14]. Ambient sensors, such as pressure sensing floors, have also shown promise as they remove the need to instrument the patient. They, however, remain comparatively expensive, require specialized installation, and have limited range of use [2], [15].

Automating clinical tests such as the TUG test using video data has been heavily explored, citing its appeal as a minimally invasive method that requires no direct contact with the individual and little setup [2],[16]. Such video-based methods can be divided into two categories: those based on regular red-green-blue (RGB) video cameras [17], [18] and those that use systems that combine video data with depth sensors (RGBD) [2], such as the Microsoft Kinect [16].

Video cameras simply capture a series of RGB images over time, so many standard image processing techniques have been explored in the past for human movement [2]. For the TUG test, a silhouette of the individual is often extracted using the difference from some baseline frame, then

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, and the New Brunswick Innovation Foundation

P. Savoie, J. Cameron, and M. Kaye are with the Department of Electrical and Computer Engineering, University of New Brunswick, Fredericton, NB, Canada, E3B 5A3.

E. Scheme is with the Department of Electrical and Computer Engineering and the Institute of Biomedical Engineering, University of New Brunswick, Fredericton, NB, Canada, E3B 5A3 (506-453-4966; fax: 506-453-4827; e-mail: escheme@unb.ca).

Manuscript received January 11, 2019; revised July 3, 2019.

various heuristics are used to determine the time it took the individual to complete the various parts of the test. Generally, these techniques have required the use of several cameras to extract 3D information to yield acceptable results [8], [19]. In 2007, a single camera setup was proposed using basic image processing techniques [17], but neither the number of participants in the study nor numerical results were disclosed. The description of the algorithm in that work suggests that it struggled with accurate timing, and the figures showed that the algorithm lost track of participants multiple times throughout the test due to errors in the background subtraction algorithm used to determine the foreground in the video frames. As a result, the authors suggested using multiple sensors in addition to the camera to improve results.

The Microsoft Kinect V2, which was launched with the Xbox One in 2013, is a popular device for recording RGBD data. It is particularly appealing due to its accurate built-in skeleton tracking and silhouette extraction capabilities, which are desirable for gait analysis [20]. The Kinect sensor has been validated against other gold standards such as marker-based infrared cinematography systems and has been shown to be accurate enough to identify subtle differences in gait [21], [22], [23]. Several tests have been instrumented with the Kinect to diagnose a variety of conditions based on a user's gait [24], [25]. Despite this relative popularity within the research community, the Kinect has recently been discontinued [16]. Nevertheless, the Kinect is still being actively used in very modern works [26] and as recently as 2019 [27].

The Kinect, however, has several drawbacks and constraints that limit its use in real-world scenarios. For instance, Kinect does not work outdoors [28] and has an absolute limit of 4.5 meters for skeleton tracking and a working range of less than 3.5 meters [29]. Kinect is also prone to losing track of a subject's skeleton, causing failed trials and loss of data. During experiments, Kinect may also fail to accurately detect joints in certain conditions, such as when the participant is turning or facing perpendicular to the Kinect [30]. Furthermore, automating the full 3 meter TUG test is typically carried out using multiple Kinects, since the sensor can only reliably track the user for part of the test before they become out of the optimal range of the sensor [10], [30]. This is demonstrated later in this manuscript, and is due to the fact that the Kinect must be placed approximately 2 meters from the seated user's hip to ensure that the subject remains entirely within the field of view of the camera during the test. Longer range gait-related applications, such as the 10 meter walk [31], require up to 4 Kinects to reliably track subjects throughout the test. Using multiple Kinects or cameras requires calibrating the sensors, which is a time consuming process for setup. Despite these drawbacks, the convenience of the 3D skeleton generated from these systems in controlled settings has been leveraged to successfully automate versions of the TUG test [10], [32], [30], [33].

Over the last few years, advances in deep learning have made accurate silhouette extraction and skeleton tracking possible using regular RGB cameras [34], [28], [35], [36], [37], [38]. In light of the challenges associated with the Kinect and the ubiquity of conventional RGB cameras, exploration of the potential of these new systems to automate clinical tests of mobility is warranted. Li et al. [39] described an automated version of the TUG test using 2D skeletons created from RGB images. Their work, however, did not compare the results of such as system with those of the more commonly cited Kinect-based systems.

Indeed, despite the limitations of existing Kinect-based approaches, there is little work that has leveraged the emerging deep-learning based extraction of 3D skeletons from conventional video cameras. 3D skeletons are useful for extracting rich and low dimensional information from high dimensional images, which is useful for applications with relatively low amounts of training data. Moreover, the time-series signals generated from the key-points over time are highly interpretable, making it intuitive to determine the actions a person is performing over time and identify errors via observation of the time-series signals. Furthermore, similar time extraction techniques as those used on 3D Kinect skeletons could be applied to these deep-learning extracted 3D skeletons, enabling previous research and techniques to be leveraged. One of the first works to use deep learning-based 3D skeleton extraction for a clinical application was Marinoiu et al., which used 3D skeletons to detect the emotional state of children with autism for human-robot control [18]. We follow a similar approach to this manuscript for extracting a local 3D skeleton from RGB data.

In this work, we explore the use of this regular RGB video data to automate the TUG test and extract additional meaningful information about the participant's gait and mobility. We leverage recent advances in deep learning to transform a conventional RGB camera into a new state-of-the-art markerless video-based sensor for gait and mobility monitoring, and demonstrate it's viability using the popular TUG test. We have named this proof of concept the *vTUG*, as it employs recent deep learning advances [34] to automate the TUG test using conventional *video* data alone.

This work begins by describing a TUG test experiment designed to directly compare the performance of RGB and RGBD-based approaches. We then review our approach for creating 3D skeletons from a single RGB video using deep learning and post processing techniques. Finally, we compare the performance of the proposed deep learning-based and the Kinect-based RGBD skeleton in segmenting the TUG test using hand-labelled ground truth data. This work demonstrates that 3D skeletons extracted with CNN-based deep learning methods from conventional, and more widely availble, RGB cameras can be used to automate the TUG test.

## II. Methodology

### A. Participants

All experiments were conducted in the Health Technologies Laboratory at the University of New Brunswick as approved by the University of New Brunswick's Research Ethics Board (REB 2018-110). Thirty healthy subjects (18 male, 12 female, 26.1 $\pm$ 9.4 years old) with no known current or prior gait/mobility impairments were recruited from the general student and staff population at the University of New Brunswick.

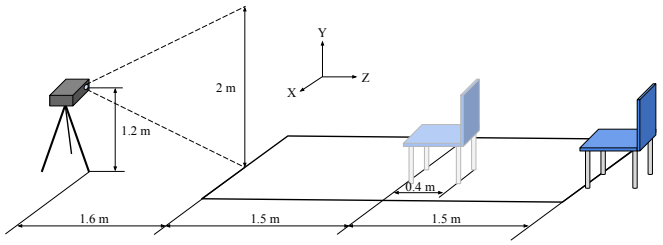



Fig. 1. Side view of the test setup. The Kinect is setup on the far left and the chair has two separate positions for the 3 m and 1.5 m test. A label of adopted axis orientation is included.

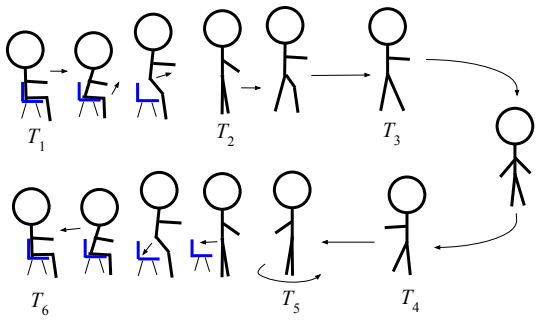



Fig. 2. The stages of the TUG test. A cartoon depiction of the complete TUG test with annotations for each stage.

All subjects gave written informed consent in accordance with the Declaration of Helsinki prior to their participation in the experiment.

### B. Camera Sensor Setup

For the RGBD configuration, a single Kinect V2 camera was used as described in [30]. Here, their protocol was repeated as closely as possible, with a few notable differences. In that work, a 2 meter TUG test was employed with the sensor placed at the end of the walking course at a height of 1.2 meters. In replicating their setup, it was found that the Kinect was performing skeleton tracking outside its optimal range of 3.5 meters when they were seated at the 2 meter position while still ensuring that the user would remain entirely within the frame when they were closest to the camera. Using the 2 meter configuration described in [30], the maximum approximate distance from the Kinect to the user's hip (when seated) was found to be approximately 4 meters. Therefore, although the performance of the proposed system was evaluated on the conventional 3 meter test, a 1.5 meter version was also tested to enable the Kinect to perform at its fullest capabilities. A diagram of this setup is shown in Fig. 1.

The RGB data was recorded at 28.3 FPS from the Kinect's 1080p raw RGB channel; ignoring the depth information. However, using the RGB stream from the Kinect ensured that both the RGB and RGBD data were synchronized and had an identical field of view.

### C. Procedure

After giving informed consent, the participants were invited to sit on the chair. Before sitting down, a calibration routine was conducted so that the NiTE middle-ware (the software used to extract the skeleton from the Kinect) could begin tracking [40]. The participant calibrated NiTE before each trial by moving close to the Kinect camera, and performing a brief series of quick movements with their arms and legs (i.e. jumping jacks). Once NiTE was calibrated, the subject sat down for a few seconds before rising to begin the TUG test. The Kinect video stream, the skeleton generated by NiTE, and the raw RGB data from the video camera were saved for each trial.

The participants performed two trials at the conventional 3 meter distance, as well as another two trials at the shortened 1.5 meter distance. These trials were performed sequentially with a minimal break in between them, sufficient to allow the subject to reset if needed. Participants were encouraged to walk at a regular and comfortable pace. There was no dress code for this experiment; participants wore whatever clothing they had with them on the day of the test.

### D. Segmentation

Prior to processing, the calibration phase of the data was removed by manually clipping the video and RGBD skeleton data streams to when the user was fully seated with their back against the chair. The user was encouraged to start the test whenever they chose, and the beginning of the test was defined as when the participant first leaned forward, moving their back away from the chair to stand up. It should be noted that the calibration phase was only required by the NiTE middleware [40], and not the proposed vTUG system.

To evaluate the effectiveness of the two automatic segmentation methods, results were compared to a baseline human-labeled assessment. The labelers, *five* independent university students, were coached how to perform the labeling, as in [30]. The location of the following key timing events were labeled based on the video data alone.

- $T_1$: The time when the participant began to stand.
- $T_2$: The time when the participant reached the standing position
- $T_3$: The time when the participant reached the end of the distance and their hips or trunk began to turn.
- $T_4$: The time when the participant completed the turn.
- $T_5$: The time when the participant reached the chair and began to turn around and sit.
- $T_6$: The time when the participant stopped moving and had successfully sat down with their back against the chair.

A diagram of these six stages can be seen in Fig. 2.

The start of a turn was noted to be the point at which the hips were no longer orthogonal to the chair. This can be seen as start of the dip before a large minima in Fig. 5(c). The end of the turn was identified when the hips were again orthogonal to the chair, but the subject was instead pointed in the opposite direction (the top of the curve shown in Fig. 5(c).)

The ground truth segmentation times were created by the human labelers watching the recorded video of the trials and clicking a button when the user completed each part of the

test. The video playback speed was slowed to approximately half speed to improve the accuracy of the labelers, with the option for labelers to move forward or backward frame by frame, or to restart a trial if they made an error. The test was segmented by a total of 5 human observers, all of which were graduate students who were familiar with the TUG test. The ground truth was selected as the mean of all the labelers. The variability in the ground truth labeling can be seen in Table I. The error for each part of the test was defined as the difference between the ground truth and the result of the automated segmentation.

### *E. RGBD-based Processing (Kinect)*

For the Kinect-based RGBD processing, the open-source PrimeSense NiTE middleware was used to extract a global 3D skeleton from the Kinect camera. It has been shown that there is no difference in accuracy between the Microsoft SDK and NiTE skeletons for regular tasks such as walking and sitting [41].

### *F. RGB-based Processing (vTUG)*

For the proposed vTUG system, the video footage was processed using the popular Mask R-CNN deep learning algorithm [34]. Mask R-CNN achieved state of the art results on the COCO dataset [42] in 2017 for semantic segmentation, bounding box detection, and 2D pose detection. The COCO dataset contains nearly 120,000 images for bounding box detection and semantic segmentation, as well as nearly 60,000 images with 2D keypoints for its training and validation set. Mask R-CNN was trained on this data to perform all three tasks using a multi-task loss function, and employs a two-stage architecture; one for detecting candidate regions for objects, and one for determining the bounding box/2D keypoints. Mask R-CNN is publicly available, and was implemented using Python by leveraging the Caffe2 deep learning framework [43]. This subject independent model was used to extract both the bounding box representing the general location of the participant and the 2D skeleton key points, as shown in Fig. 3. The Mask R-CNN algorithm processing speed was approximately 2 FPS using a Nvidia Tesla K40.

Using Python, the image was then cropped to the bounding box coordinates extracted by Mask R-CNN and subsequently processed by the deep multi-task architecture for human sensing (DMHS) [37] to extract the 3D pose of the participant. The DMHS model used to predict local 3D joint positions was trained by the original authors of [37] on the publicly available HumanEva [44], Human3.6M [45] and Leeds Sports Pose [46] These datasets contain a combined amount of more than 3.6 million cropped images of people, necessitating the automated cropping of the Mask R-CNN labeled images before being passed the network. DMHS uses a deep convolutional architecture with several recurrent stages to detect poses and semantic segmentation masks of all body parts. Each stage uses the results from the previous stage to refine the estimate of the pose and body part masks from the previous stage. DMHS generates local 3D joint positions with respect to the field of view of the cropped image, and a coordinate system centered at the hip. The DHMS algorithm processing speed was approximately 1.3 FPS using a Nvidia Tesla K40. An example of the output from DMHS can be seen in Fig. 3. While this provides useful information about the relative position of each joint, global information is still needed to determine how the subject is moving through the environment. We define a subject's global pose as the position of their joints with a coordinate system centered at the camera.

Estimates of the global 3D pose were created from the global 2D poses (provided by Mask R-CNN) and hip-centered 3D pose by modeling the transformation between the global 3D pose and global 2D pose as a projection from 3D space to image space [28], [47]. The simple and efficient method described by Zanfir et al. [47] was implemented in Matlab to obtain the global positions of the subjects' skeletal key-points. An overview of the entire method from raw RGB data to global 3D skeleton can be seen in Fig. 4. No additional intrinsic or extrinsic camera information was used in creating the skeleton data for the RGB-based vTUG system.

### *G. Transition Point Extraction*

For both of the 3D pose estimations created by the RGB (vTUG) and RGBD (Kinect), the TUG test was segmented into the subsections outlined in Section II-D. This included the times when the participant: reached the standing position, began to turn at the end of the distance, completed the turn, reached the chair and began to turn around and sit, and had successfully sat down with their back against the chair.

Examples of the time-series signals extracted from the 3D skeletons of both systems can be seen in Fig. 5. Fig. 5(a) shows the position of the centroid between the shoulders relative to the location of the camera. Fig. 5(b) shows the difference in height between the feet and nose. Fig. 5(c) shows the absolute difference between the right and left hip. It should be noted that the RGBD and RGB curves in these figures have been offset for easier visualization. Like colored dots appearing on the signal trace represent the location that the transition point extraction algorithm chose for each time segment, while the colored vertical stripes represent the average ($\mu \pm 1.5\sigma$) human labeler time.

*1) Transition Point ($T_1$):* In order to determine when the participant first stood up, the center of the shoulders in the Z axis was used. As seen in Fig. 5(a), both the RGBD and RGB signals begin to dip as soon as the user first stands, appearing in this example slightly before the human labeled time, $T_1$. This point occurs when the participant first leans to begin the forward motion to stand.

*2) Transition Point ($T_2$):* The signal used to ascertain when the participant had completed standing was found when the height of the person reached an initial local maximum. The height of the person was found by subtracting the Y coordinate of the nose from the Y coordinate of the center of the feet. The Y coordinate of the feet was extremely inaccurate for the Kinect-based system, therefore the Y coordinate of the nose alone was used as an estimate of the height. As highlighted as $T_2$ in Fig. 5(b), the full standing height occurs after a distinct local minimum that results from bending over to stand up.

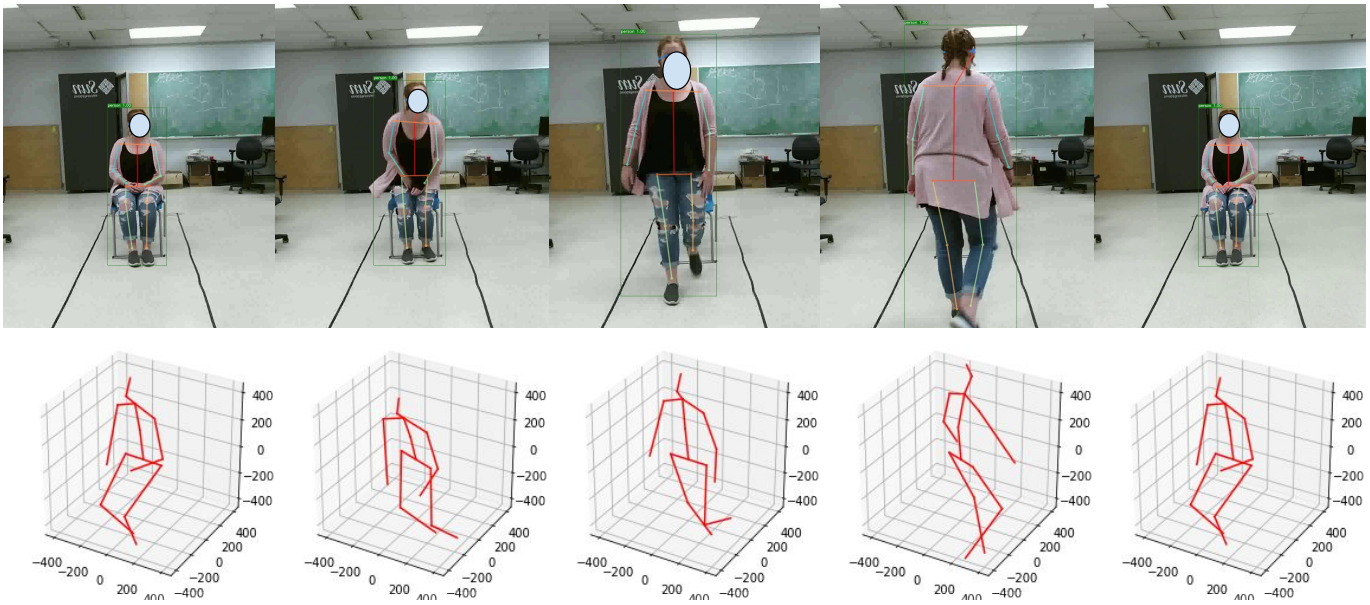


Fig. 3. (Top row of images) An example of the 2D Skeleton features and bounding box output created by the Mask R-CNN algorithm [34] and (bottom row of images) the RGB 3D skeleton created with DHMS [37].

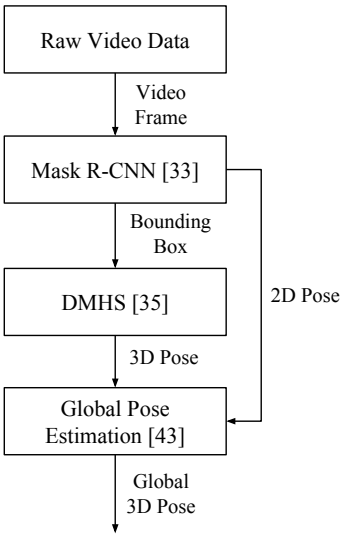



Fig. 4. Flowchart of the algorithms used to extract global 3D poses from a single camera in the vTUG system

*3) Transition Point ($T_3$):* The point at which the participant reached the end of the distance and began to turn around was segmented by observing the absolute difference between the X coordinates of the right and left hips. When the signal begins to dip (shown by the $T_3$ label in Fig. 5(c)), the participant was determined to have begun to turn.

*4) Transition Point ($T_4$):* The end of the turn was determined by when the same hip difference signal had completed the dip and risen back to a peak, as shown by the $T_4$ label in Fig. 5(c).

*5) Transition Point ($T_5$):* When the participant reached the chair, they would turn around in order to sit down. Using the same hip feature as in $T_3$, representing a dip in the absolute hip difference signal. This is highlighted by the $T_5$ label in Fig. 5(c).

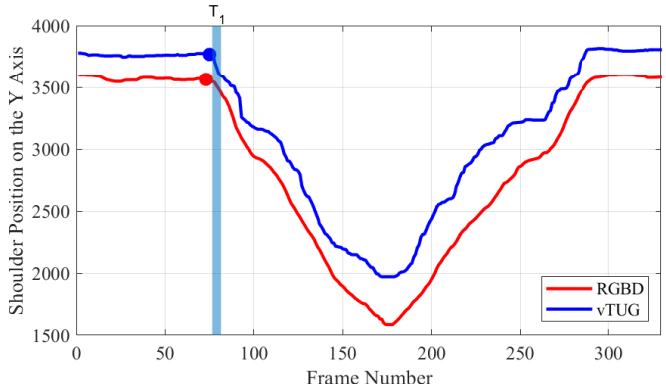
*6) Transition Point ($T_6$):* Finally, the trial was noted as complete when the participant had completely sat down as measured using the same height signal as outlined for $T_2$. As denoted by the $T_6$ label in Fig. 5(b), a local peak occurs after the dip and is followed by a flat line, representing that the participant is no longer changing in height.

The raw RGB and RGBD skeleton time-series data were filtered using an 8 sample median filter to smooth the signal and reduce the overall noise. A heuristic algorithm was then used to extract two main types of points of interest. The first was a local maximum after a meaningful rise in the signal (as in $T_2$, $T_4$ and $T_6$) and the second was a plateau in the signal before a sudden drop (as in $T_1$, $T_3$, and $T_5$).

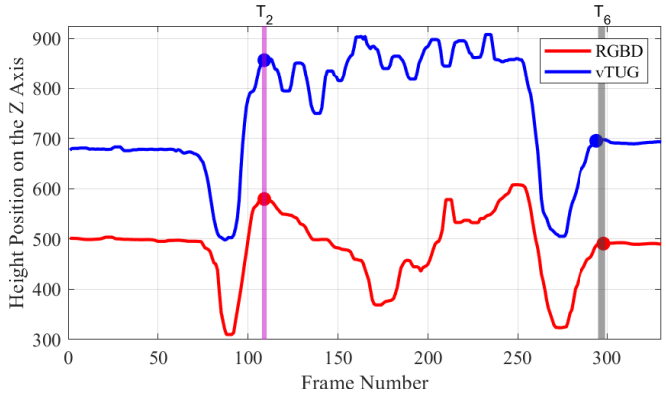
To find the local maxima after a rise event, the preceding rise was required to exceed a minimum percentage ($\delta$) of the dynamic range of the signal for that trial. The steady state after the local maxima was determined by when the timing magnitude of the derivative of the next $N$ samples was found to be less than some threshold ($\epsilon$).

Similarly, the requirement for a sudden drop was a decrease of at least ($\delta$) percent of the dynamic range of the signal for that trial. The preceding plateau was again identified by the period when the mean value of the magnitude of the derivative of the last $N$ consecutive samples was less than $\epsilon$.

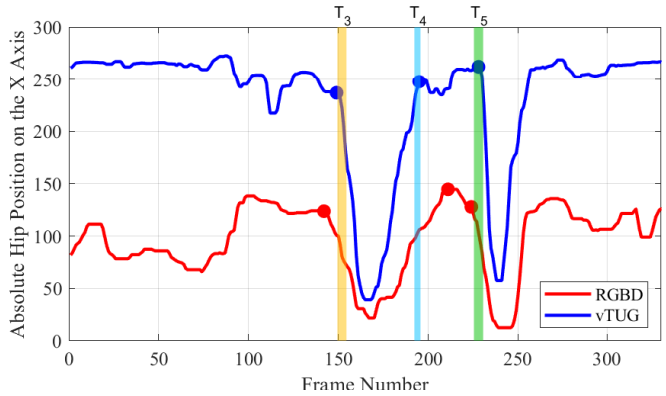
Due to the low number of parameters in the model, the search space of hyperparameters for each transition point was small. The three hyper-parameters ($\delta$, $\epsilon$ and $N$) were therefore determined using a brute force parameter search with five-fold cross validation that minimized the mean absolute error on the training set for each transition point ($T_1$-$T_6$) within each type of test (3m vTUG, 1.5m vTUG, and 1.5m RGBD). Results shown here represent the corresponding average performance




(a) Example of the average shoulder signals in the Y Axis (distance from the camera). The average human labeled time for $T_1$ is shown by the purple stripe. The corresponding algorithm labels are shown by the colored dots centered on the signal.




(b) Example of the height signal (difference between feet and nose) in the Z Axis. The position of the feet were very inaccurate for the Kinect-based system, therefore the position of the nose was used instead (Which is why the signals appear quite different, although the peaks remained the same). The human labeled times for $T_2$ and $T_6$ are shown by the pink and grey stripes respectively. The corresponding algorithm labels are shown by the colored dots centered on the signal.




(c) Example absolute hip difference signal in the X Axis. The human labeled times for $T_3$, $T_4$ and $T_5$ are shown by the yellow, cyan, and green stripes respectively. The corresponding algorithm labels are shown by the colored dots centered on the signal.)

Fig. 5. Examples of key skeletal point time-series. These points were used to calculate time labels for a representative 1.5 m test. Note that each of these signals represent the 11 sample median filtered version.

on the unseen test folds using the computed $\delta$, $\epsilon$ and $N$ parameters. The limits of the search space were determined heuristically by testing a range of hyperparameters that yielded meaningful results. During training, it was ensured that chosen hyperparameters were not at the boundaries of the search space, otherwise the the search space was extended.

### H. Statistics

Both systems were compared based on their mean, absolute mean, and median errors with respect to the average human labels. A KolmogorovSmirnov normality test was used to determine that the distribution of errors for each key point of the vTUG and Kinect-based systems were not normal. The performance of the two systems were therefor compared using a non-parametric Mann-Whitney U, with p-values of $p<0.05$.

## III. Results

### A. Skeleton Tracking

Both the Kinect-based and RGB-based models provided tracking of key body segments in close agreement with those in [30], as shown by the example signals in Fig. 5. In these figures, the RGB signals show the output of the Kinect-based NiTE middle-ware, while the RGB signals denote the outputs obtained from the proposed RGB-based 3D skeleton method outlined in Fig. 4.

Each participant performed each of the 3 meter and 1.5 meter trials twice, resulting in a total of 60 TUG test sequences of each. For ten of the sixty 1.5 meter trials, the skeleton data created by the RGBD data was incorrect or corrupted. In these cases, NiTE stopped tracking the participant at some point during the trial and did not recover. This lead to a flat-line signal that was clearly not representative of the participant's movements and the RGBD data from these trials were therefore discarded. It should be noted that the RGB-based models from these trials had no such errors, thus these 10 trials were included with the RGB results. It should also be re-iterated that the Kinect-based system was unable to track the participants successfully for any of the 3 meter trials.

### B. Segmentation Results

In order to evaluate and compare the vTUG and RGBD systems, the results from the corresponding automated segmentations were measured against the results of the human labelers. Using the transition point extraction method as described in Section II-G, the times for the six transition points were found for each trial, and compared to the average times found by the five human labelers. Table I shows the standard deviation across all trials for each transition point. The timing of the five human observers, who used a slowed version of the video to label, agreed closely resulting in consistently low standard deviations below 65 milliseconds. Using these locations as the baseline for comparison, Table I shows the average ($\mu$), median ($\tilde{m}$), and standard deviation ($\sigma$) of the error achieved by the RGBD and vTUG automated segmentation approaches. Since the Kinect-based system was unable to extract the skeleton information for the 3 meter test, only the results for the

TABLE I
TIMING ERROR FROM HUMAN LABELED GROUND TRUTH
(MEAN ERROR, $\mu$, MEAN ABSOLUTE ERROR, $|\mu|$,
MEDIAN ERROR, $\tilde{m}$, AND STANDARD DEVIATION, $\sigma$)

| | Metric | GT | RGBD(1.5m) | vTUG(1.5m) | vTUG(3m) |
|---|---|---|---|---|---|
| $T_1$ | $\mu$ (ms) | - | -156.2 | -20.0 | -27.7 |
| | $\tilde{m}$ (ms) | - | -17.7 | 0 | -70.7 |
| | $\|\mu\|$ (ms) | - | 328.6 | 196.7 | 306.8 |
| | $\sigma$ (ms) | 58.4 | 436.6 | 257.7 | 479.4 |
| $T_2$ | $\mu$ (ms) | - | 60.8 | 27.1 | 45.9 |
| | $\tilde{m}$ (ms) | - | 35.3 | 35.3 | 53.0 |
| | $\|\mu\|$ (ms) | - | 127.2 | 157.8 | 162.5 |
| | $\sigma$ (ms) | 31.3 | 257.1 | 203.4 | 200.3 |
| $T_3$ | $\mu$ (ms) | - | 86.2 | 50.6 | 73.6 |
| | $\tilde{m}$ (ms) | - | 70.7 | 35.3 | 35.3 |
| | $\|\mu\|$ (ms) | - | 188.0 | 169.6 | 183.2 |
| | $\sigma$ (ms) | 59.4 | 234.3 | 301.8 | 397.8 |
| $T_4$ | $\mu$ (ms) | - | 12.0 | 52.4 | 183.7 |
| | $\tilde{m}$ (ms) | - | 53.0 | 35.3 | 141.3 |
| | $\|\mu\|$ (ms) | - | 228.3 | 210.2 | 255.6 |
| | $\sigma$ (ms) | 40.4 | 298.3 | 376.0 | 488.0 |
| $T_5$ | $\mu$ (ms) | - | 220.0 | 203.2 | 257.4 |
| | $\tilde{m}$ (ms) | - | 141.3 | 141.3 | 141.3 |
| | $\|\mu\|$ (ms) | - | 290.4 | 227.9 | 268.0 |
| | $\sigma$ (ms) | 60.4 | 498.4 | 431.3 | 386.1 |
| $T_6$ | $\mu$ (ms) | - | 263.6 | 80.7 | -9.42 |
| | $\tilde{m}$ (ms) | - | 88.3 | 88.3 | 17.6 |
| | $\|\mu\|$ (ms) | - | 519.4 | 319.8 | 260.3 |
| | $\sigma$ (ms) | 46.0 | 820.8 | 447.8 | 328.3 |

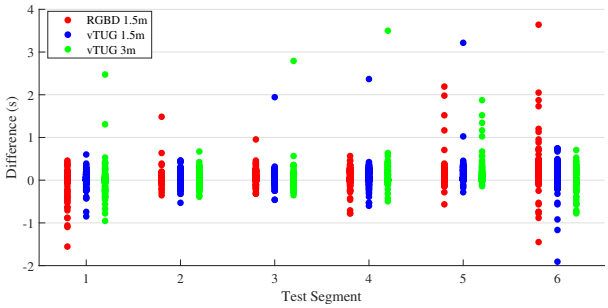



Fig. 6. A scatter plot of the segmentation error for both the RGBD and RGB skeleton signals for all of the trials

proposed vTUG system are included for this more common version of the test. The results of the statistical testing showed no significant differences between the performances of the vTUG and the Kinect-based system, nor between between the 1.5 meter and 3 meter version tests using the vTUG.

The final results presented in Table I demonstrate the comparability of the RGBD data and the proposed RGB 3D skeleton-based vTUG for use in automating the TUG test. For reference, the average total times for the TUG test are shown in Table II. These times, as labeled by the human labelers, were calculated as the difference between $T_1$ and $T_6$.

Fig. 7 presents the average times of the 1.5 m test for the three labeling methods. The six time key points were aligned at $T_1$ (set to zero for readability) while the rest of the times follow in order afterwards (ie. orange represents $T_2 - T_1$, yellow $T_3 - T_2$, purple $T_4 - T_3$, etc.). The error bars centered on each key time point represent one standard deviation of the result.

TABLE II
TOTAL TUG TEST TIMES BASED ON HUMAN LABELS
(MEAN, $\mu$, STD. DEV., $\sigma$)

| Method | Distance | $\mu$ (s) | $\sigma$ (s) |
|---|---|---|---|
| GT | 1.5 m | 7.84 | 1.84 |
| RGBD | 1.5 m | 8.04 | 1.97 |
| RGB | 1.5 m | 7.94 | 1.80 |
| GT | 3 m | 10.19 | 2.21 |
| RGBD | 3 m | - | - |
| RGB | 3 m | 10.20 | 2.23 |

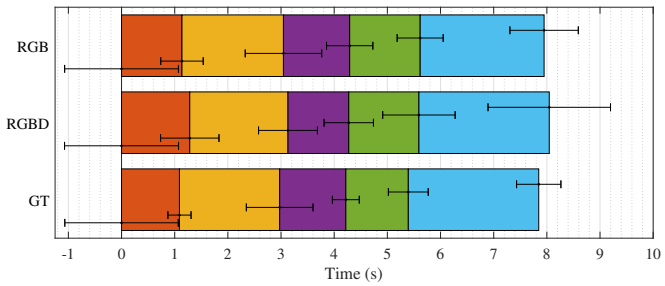



Fig. 7. A chart of the average times of the 1.5 m test. Error bars are $\pm$ the standard deviation of the values. $T_1$ is set to zero while the rest of the time follow in order after (ie. orange represents $T_2 - T_1$, yellow $T_3 - T_2$, purple $T_4 - T_3$, etc.)

## IV. DISCUSSION

Overall, the RGB and RGBD-based systems obtained similar performance in mean and median errors for the 1.5 meter tests, despite some differences standard deviation. As shown in Fig. 7, the overall time series of the vTUG approach closely matched that of the ground truth labels. In Fig. 6, outliers can be observed in the RGBD-based test (for example, for $T_6$, which skewed the distribution resulting in a larger standard deviation. These outliers were caused by frames where the NiTE middleware provided erroneous results for the skeleton, leading to errors in the segmentation.

The RGBD-based system also completely failed to track the subject in *ten* of the sixty 1.5 meter tests, yielding no data from which to extract timing information. These trials were, therefore, necessarily excluded from the calculation of the mean and standard deviation, possibly biasing the results in favor of the RGBD-based system. The limited range of the RGBD-based system also prevented the comparison of results for the 3 meter test altogether. Despite the protocol being the same for the conventional 3 meter test and the shortened 1.5 meter test, RGBD-based system could not track the subject during the 3 meter test. After successful calibration, as the subject moved towards the chair to start the test, the sensor would lose calibration before the user reached the chair to start the test. This highlights the range limitations of the Kinect sensor, and the potential of the vTUG system in scenarios where Kinect sensor fails entirely. Indeed, the 1.5 meter is not commonly used clinically, emphasizing the benefit of the

proposed vTUG test, which is able to handle the 3M test and could be extended for even longer tests.

Through inspection of the time series plots, the time-series signals of the RGBD-based were more noisy when the user was further from the Kinect. This explains why the system had trouble detecting movement when some participants began to stand up or sit down, accounting for the high absolute error and standard deviation of $T_1$ and $T_6$, as the user is furthest from the sensor during both.

There were a few notable outliers for $T_5$ of the 3 meter vTUG tests. These were caused by DMHS [37] loosing track of the pose of individuals for a short period (8-16 frames). In these cases, this caused enough noise to create a false drop in the position of the hips. These erroneous frames were limited to two 3 meter trials from two individuals (4 trials total) who took very long strides on their way back to the chair, and whose clothing blended in with the background. The use of loose clothing limits the 3D pose extraction's overall ability to precisely identify body segments - particularly those around the hips. These factors, coupled with the fact that they were facing away from the camera caused DMHS to momentarily loose track of the individual. These frames were likely different from the data used to train DMHS, as the HumanEva [44] and Humans3.6m [45] datasets were created in a laboratory setting. A 3D pose neural network trained on more relevant images and with more pose and clothing variability, such as [28], may be better suited to accommodate such issues. Despite this challenge, the vast majority of poses for RGB data were qualitatively observed to be very accurate and caused no issues other than for these two subjects.

The standard deviations for the ground truth labels ranged between 30 and 60 ms. This is in part due to the discrete nature of video, where an error of a single frame corresponds with 35 ms (based on the frame rate of 28.3 FPS). The variance of the automated segmentation algorithm was similar for both systems, but is higher than those of the ground truth, partially due to a small number of outliers. For example, for $T_4$ of the 3M test, the standard deviation is 488ms. If one outlier (out of 60 trials) were to have been removed, the standard deviation would have dropped to 223ms. Nevertheless, these results were included for completeness and, even when included, the observed variances are lower than those previously reported for manual timing. Lohmann et al. [10] compared the timing between stopwatch measurements and video analysis and found a mean difference ($T_{watch} - T_{video}$) of -100ms with a standard deviation of 680ms. Furthermore, the ability to extract the timing of relevant key sub-actions, like how long it took to stand up ($T_2$-$T_1$), may provide additional diagnostic information that is difficult to time manually. Additional information about a patients gait can also be extracted from the 3D skeleton, such as variation in gait patterns and speed [48].

Although the heuristic transition point extraction method used here was relatively resilient to noise, it did not always identify the precise value of the transition point. The features chosen here to represent the transition points ($T_1$ through $T_6$) may not correspond directly with what a human labeler may interpret as the start of a movement (ie. standing or sitting). This may account for some of the mean offset of some of the segmentation times. By looking at several of the body joints and fusing their signals predictions into a single feature, the robustness of the segmentation could be increased. Nevertheless, this work serves as a proof of concept to compare the RGB-based deep learning skeletons (vTUG) and the Kinect-based NiTE skeletons for the use in automating the TUG test. This simply means that even greater performance may be achievable with an improved segmentation scheme.

Another important point of consideration is that all subjects in this experiment were able bodied. As shown in Table II, participants performed the 3 meter test in 10.19 seconds, on average, and the 1.5 meter test in 7.81 seconds. This is significantly faster than most populations of interest for the TUG test [49]. It is likely that the speed and trajectories of seniors or individuals with pathological gait may be less periodic and this could cause anomalies in the segmentation process. However, Li et al. [39] found little difference in accuracy between patient groups of different degrees of severity, suggesting that the impact of moving to patient populations may be minimal. Their work used a LSTM/SVM to classify each movement phase [39], whereas the proposed vTUG uses a heuristic approach for simplicity and robustness. While heuristic approaches can generalize well, it is yet to be determined how these two approaches compare in their generalization to new populations. This will be further explored in future works.

As the proposed vTUG system was created as a proof of concept, there is room for improvement in several areas. With the current fast pace of research in computer vision, it is reasonable to expect continued improvements in deep learning based 3D skeleton tracking technologies. Leveraging these new advancements will likely further reduce the error in the skeletal signal.

In this work, the algorithm was developed using a camera placed directly in front of the individual performing the test. Since the user moves in a relatively straight line during the TUG test, it is possible that the test could be modified to work with cameras at different angles. This could be achieved using the same algorithm by simply rotating the coordinate system with respect to the line of travel. Future work could explore a camera angle independent algorithm of automating the test with a single camera. Longer distance TUG tests are also becoming more common in clinical practice, as a way to place more emphasis on walking speed. The proposed vTUG, which should be capable of such tests, should therefore be explored for these longer versions.

Finally, it is worth noting that although the RGB video data used in this work was collected using the Kinect system, the result is completely generalizable to other conventional video capture systems. The same video data, or better, could readily be collected using a webcam or modern cellphone camera that operates at $\geq 28.3FPS$ or higher. The increased flexibility of using such deep learning systems to extract valuable information from such conventional RGB cameras could open the door to new applications of the TUG test that were previously unfeasible with the RGBD-based Kinect hardware. The fact that these techniques work with a regular RGB camera may enable them to work with smart phone

cameras, with processing offloaded to the cloud. Extensions of this work could therefore facilitate the use of the TUG test to enable mobility assessment in the home or in long term care facilities on a regular basis.

## V. Conclusion

This work presents a novel system for automating the time-up-and-go test using a conventional video camera. By using deep learning methods, global 3D skeletons were extracted that closely resemble those previously obtained only using Kinect-based RGBD. The time-series of the skeletons of both the Kinect-based NiTE and the new RGB skeletons were processed by the same segmentation algorithm to further evaluate the comparability of these preprocessing methods for use in automating the TUG test. By comparing the transition points extracted from both of these signal to the ground truth of human labelers, it was shown that global 3D skeletons created with CNN-based deep learning methods can serve as a replacement for systems that require the specialized, but limited, Kinect system. These results may facilitate the exploration of automating other clinical tests that were previously impractical due to hardware limitations.

## References


[1] D. Podsiadlo and S. Richardson, "The timed up & go: A test of basic functional mobility for frail elderly persons," *Journal of the American Geriatrics Society*, vol. 39, no. 2, pp. 142–148. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.1991.tb01616.x

[2] G. Sprint, D. Cook, and D. Weeks, "Towards Automating Clinical Assessments: A Survey of the Timed Up and Go (TUG)," *IEEE reviews in biomedical engineering*, vol. 8, pp. 64–77, 2015. [Online]. Available: https://doi.org/10.1109/rbme.2015.2390646

[3] B. R. Greene and R. A. Kenny, "Assessment of Cognitive Decline Through Quantitative Analysis of the Timed Up and Go Test," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 988–995, Apr. 2012.

[4] M. Katz-Leurer, H. Rotem, O. Keren, and S. Meyer, "Balance abilities and gait characteristics in post-traumatic brain injury, cerebral palsy and typically developed children," *Developmental Neurorehabilitation*, vol. 12, no. 2, pp. 100–105, Jan. 2009. [Online]. Available: https://doi.org/10.1080/17518420902800928

[5] J. C. Wall, C. Bell, S. Campbell, and J. Davis, "The Timed Get-up-and-Go test revisited: measurement of the component tasks," *Journal of rehabilitation research and development*, vol. 37, no. 1, 2000.

[6] A. Shumway-Cook, S. Brauer, and M. Woollacott, "Predicting the Probability for Falls in Community-Dwelling Older Adults Using the Timed Up & Go Test," *Physical Therapy*, vol. 80, no. 9, pp. 896–903, Sep. 2000.

[7] M. E. Tinetti, "Performance-Oriented Assessment of Mobility Problems in Elderly Patients," *Journal of the American Geriatrics Society*, vol. 34, no. 2, pp. 119–126, Feb. 1986. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.1986.tb05480.x

[8] F. Wang, M. Skubic, C. Abbott, and J. M. Keller, "Quantitative analysis of 180 degree turns for fall risk assessment using video sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 7606–7609.

[9] R. C. v. Lummel, S. Walgaard, A. B. Maier, E. Ainsworth, P. J. Beek, and J. H. v. Dien, "The Instrumented Sit-to-Stand Test (iSTS) Has Greater Clinical Relevance than the Manually Recorded Sit-to-Stand Test in Older Adults," *PLoS ONE*, vol. 11, no. 7, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938439/

[10] O. Lohmann, T. Luhmann, and A. Hein, "Skeleton Timed Up and Go," in *2012 IEEE International Conference on Bioinformatics and Biomedicine*, Oct. 2012, pp. 1–5.

[11] J. Beyea, C. A. McGibbon, A. Sexton, J. Noble, and C. OConnell, "Convergent Validity of a Wearable Sensor System for Measuring Sub-Task Performance during the Timed Up-and-Go Test," *Sensors (Basel, Switzerland)*, vol. 17, no. 4, Apr. 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426930/

[12] T. Frenken, B. Vester, M. Frenken, and A. Hein, "aTUG: Fully-automated Timed Up and Go Assessment Using Ambient Sensor Technologies," Jan. 2011, pp. 55–62.

[13] J. Williamson, Q. Liu, F. Lu, W. Mohrman, K. Li, R. Dick, and L. Shang, "Data sensing and analysis: Challenges for wearables," in *The 20th Asia and South Pacific Design Automation Conference*, Jan. 2015, pp. 136–141.

[14] M. Milosevic, E. Jovanov, and A. Milenkovic, "Quantifying Timed-Up-and-Go test: A smartphone implementation," in *2013 IEEE International Conference on Body Sensor Networks*. Cambridge, MA, USA: IEEE, May 2013, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/6575478/

[15] C. Ni Scanaill, C. Garattini, B. R. Greene, and M. J. McGrath, "Technology Innovation Enabling Falls Risk Assessment in a Community Setting," *Ageing International*, vol. 36, no. 2, pp. 217–231, Jun. 2011. [Online]. Available: https://doi.org/10.1007/s12126-010-9087-7

[16] Microsoft Corporation, "Kinect - Windows app development." [Online]. Available: https://developer.microsoft.com/en-us/windows/kinect

[17] D. Berrada, M. Romero, G. Abowd, M. Blount, and J. Davis, "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 - HealthNet '07*. San Juan, Puerto Rico: ACM Press, 2007, p. 73. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1248054.1248075

[18] E. Marinoiu, M. Zanfir, V. Olaru, and C. Sminchisescu, "3d Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children With Autism," 2018, pp. 2158–2167.

[19] Z. Skrba, B. O'Mullane, B. R. Greene, C. N. Scanaill, C. W. Fan, A. Quigley, and P. Nixon, "Objective real-time assessment of walking and turning in elderly adults," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2009, pp. 807–810.

[20] K. Otte, B. Kayser, S. Mansow-Model, J. Verrel, F. Paul, A. U. Brandt, and T. Schmitz-Hbsch, "Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function," *PLOS ONE*, vol. 11, no. 11, p. e0166532, Nov. 2016. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166532

[21] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3d motion capture for gait analysis," *Journal of Medical Engineering & Technology*, vol. 38, no. 5, pp. 274–280, Jul. 2014.

[22] B. Mller, W. Ilg, M. A. Giese, and N. Ludolph, "Validation of enhanced kinect sensor based motion capturing for gait assessment," *PloS One*, vol. 12, no. 4, p. e0175813, 2017.

[23] D. J. Geerse, B. H. Coolen, and M. Roerdink, "Kinematic Validation of a Multi-Kinect v2 Instrumented 10-Meter Walkway for Quantitative Gait Assessments," *PLOS ONE*, vol. 10, no. 10, p. e0139913, Oct. 2015. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0139913

[24] S. Bei, Z. Zhen, Z. Xing, L. Taocheng, and L. Qin, "Movement Disorder Detection via Adaptively Fused Gait Analysis Based on Kinect Sensors," *IEEE Sensors Journal*, vol. 18, no. 17, pp. 7305–7314, Sep. 2018.

[25] Q. Li, Y. Wang, A. Sharf, Y. Cao, C. Tu, B. Chen, and S. Yu, "Classification of gait anomalies from kinect," *The Visual Computer*, vol. 34, no. 2, pp. 229–241, Feb. 2018. [Online]. Available: https://doi.org/10.1007/s00371-016-1330-0

[26] A. P. Rocha, H. M. P. Choupina, M. d. C. Vilas-Boas, J. M. Fernandes, and J. P. S. Cunha, "System for automatic gait analysis based on a single RGB-D camera," *PLOS ONE*, vol. 13, no. 8, p. e0201728, Aug. 2018. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201728

[27] X. Cui, Z. Zhao, C. Ma, F. Chen, and H. Liao, "A Gait Character Analyzing System for Osteoarthritis Pre-diagnosis Using RGB-D Camera and Supervised Classifier," in *World Congress on Medical Physics and Biomedical Engineering 2018*, ser. IFMBE Proceedings, L. Lhotska, L. Sukupova, I. Lackovi, and G. S. Ibbott, Eds. Springer Singapore, 2019, pp. 297–301.

[28] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3d Human Pose Estimation with a Single RGB Camera," *ACM Trans.*

*Graph.*, vol. 36, no. 4, pp. 44:1–44:14, Jul. 2017. [Online]. Available: http://doi.acm.org/10.1145/3072959.3073596

[29] Microsoft Corporation, "Human Interface Guidelines v2.0," 2014. [Online]. Available: http://download.microsoft.com/download/6/7/6/676611b4-1982-47a4-a42e-4cf84e1095a8/kinecthig.2.0.pdf

[30] A. H. K. B., A. Mollahosseini, T. Struemph, W. Pace, R. D. Nielsen, and M. H. Mahoor, "Automatic Measurement of Physical Mobility in Get-Up-and-Go Test Using Kinect Sensor," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3492–3495, Aug. 2014, arXiv: 1511.03603. [Online]. Available: http://arxiv.org/abs/1511.03603

[31] A. Dubois, T. Bihl, and J.-P. Bresciani, "Automating the Timed Up and Go Test Using a Depth Camera," *Sensors (Basel, Switzerland)*, vol. 18, no. 1, Dec. 2017.

[32] E. Cippitelli, S. Gasparrini, E. Gambi, and S. Spinsante, "A depth-based joints estimation algorithm for get up and go test using Kinect," in *2014 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2014, pp. 226–227.

[33] N. Kitsunezaki, E. Adachi, T. Masuda, and J. Mizusawa, "KINECT applications for the physical rehabilitation," in *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2013, pp. 294–299.

[34] K. He, G. Gkioxari, P. Dollr, and R. Girshick, "Mask R-CNN," *arXiv:1703.06870 [cs]*, Mar. 2017, arXiv: 1703.06870. [Online]. Available: http://arxiv.org/abs/1703.06870

[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 3431–3440.

[36] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[37] A.-I. Popa, M. Zanfir, and C. Sminchisescu, "Deep Multitask Architecture for Integrated 2d and 3d Human Sensing," *arXiv:1701.08985 [cs]*, Jan. 2017, arXiv: 1701.08985. [Online]. Available: http://arxiv.org/abs/1701.08985

[38] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "Monocular 3d Human Pose Estimation In The Wild Using Improved CNN Supervision," *arXiv:1611.09813 [cs]*, Nov. 2016, arXiv: 1611.09813. [Online]. Available: http://arxiv.org/abs/1611.09813

[39] T. Li, J. Chen, C. Hu, Y. Ma, Z. Wu, W. Wan, Y. Huang, F. Jia, C. Gong, S. Wan, and L. Li, "Automatic timed up-and-go sub-task segmentation for parkinson's disease patients using video-based activity classification," vol. 26, no. 11, pp. 2189–2199.

[40] PrimeSense Inc., "Prime Sensor NITE 1.3 Framework Programmer's Guide," 2010.

[41] A. Cosgun, M. Bunger, and H. I. Christensen, "Accuracy analysis of skeleton trackers for safety in HRI," in *Proceedings of the Workshop on Safety and Comfort of Humanoid Coworker and Assistant (HUMANOIDS), Atlanta, GA, USA*, 2013, pp. 15–17.

[42] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollr, "Microsoft COCO: Common Objects in Context," *arXiv:1405.0312 [cs]*, May 2014, arXiv: 1405.0312. [Online]. Available: http://arxiv.org/abs/1405.0312

[43] F. Research, "Caffe2 and PyTorch join forces to create a Research + Production platform PyTorch 1.0," May 2018. [Online]. Available: http://caffe2.ai/blog/2018/05/02/Caffe2_PyTorch_1_0.html

[44] L. Sigal, A. O. Balan, and M. J. Black, "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated HumanMotion," *International Journal of Computer Vision*, vol. 87, no. 1, p. 4, Aug. 2009. [Online]. Available: https://doi.org/10.1007/s11263-009-0273-6

[45] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large Scale Datasets and Predictive Methods for 3d Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.

[46] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *Proceedings of the British Machine Vision Conference*, 2010, doi:10.5244/C.24.12.

[47] A. Zanfir, E. Marinoiu, and C. Sminchisescu, "Monocular 3d Pose and Shape Estimation of Multiple People in Natural Scenes - The Importance of Multiple Scene Constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2148–2157.

[48] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 828–838, May 2005.

[49] R. W. Bohannon, "Reference Values for the Timed Up and Go Test: A Descriptive Meta-Analysis," *Journal of Geriatric Physical Therapy*, vol. 29, no. 2, p. 64, Aug. 2006.

https://dspace.org/

# Automation of the Timed-Up-and-Go test using a conventional video camera

**Savoie, Patrick; Cameron, James A. D.; Kaye, Mary E.; Scheme, Erik J.**

**2019-08-09**
IEEE
https://unbscholar.lib.unb.ca/handle/1882/37614

Downloaded from a DSpace repository. https://dspace.org/