Evaluation of the Microsoft Kinect Skeletal Versus Depth Data Analysis for Timed-up and Go and Figure of 8 Walk Tests

Benjamin Hotrabhavananda, *Student Member, IEEE*, Anup K. Mishra, *Student Member, IEEE*, Marjorie Skubic, *Senior Member, IEEE,* Nijaporn Hotrabhavananda, and Carmen Abbott

***Abstract*—We compared the performance of the Kinect skeletal data with the Kinect depth data in capturing different gait parameters during the Timed-up and Go Test (TUG) and Figure of 8 Walk Test (F8W). The gait parameters considered were stride length, stride time, and walking speed for the TUG, and number of steps and completion time for the F8W. A marker-based Vicon motion capture system was used for the ground-truth measurements. Five healthy participants were recruited for the experiment and were asked to perform three trials of each task. Results show that depth data analysis yields stride length and stride time measures with significantly low percentile errors as compared to the skeletal data analysis. However, the skeletal and depth data performed similar with less than 3% of absolute mean percentile error in determining the walking speed for the TUG and both parameters of F8W. The results show potential capabilities of Kinect depth data analysis in computing many gait parameters, whereas, the Kinect skeletal data can also be used for walking speed in TUG and F8W gait parameters.**

I. Introduction

One of the extensively used mobility assessments in physical therapy is the Timed-up and Go (TUG) test [1]. The test consists of 5 components: stand up, walk 3m, turn around, walk 3m back towards the chair, and sit down. The total time is captured for the TUG test, but also physical therapists can observe balance deficits and stride information including stride length and stride time. Originally, physical therapists could only analyze the total time of the TUG using a stopwatch; other gait aspects have been observed and documented manually, which makes the whole process time consuming and error-prone.

The Figure of 8 walk (F8W) is another technique to assess walking skills of older adults [2-3]. A typical F8W walking path consists of both straight and curved paths, designed to represent everyday walking skills. See Fig. 1. Hess et al. have conducted experiments to validate the F8W, especially for older adults [2]. Their results show that the F8W time is correlated with gait, physical function, and movement control. Also, the number of F8W steps correlate with step width variability and fear of falling. Considering the correlations of several gait and balance parameters, the F8W test can be used to determine fall-risk in older adults [4]. This test is also done using a stopwatch, manually by clinicians.

After the release of the Microsoft Kinect (version 1) in 2010, researchers started utilizing the technology and developing systems to automate the TUG test. Lohmann *et al*. [5] is the first TUG study using a skeletal model captured by two Kinect sensors to compute 10 time parameters of the TUG. Kitsunezaki *et al*. [6] also used the Kinect to compare the differences between Kinect and human scored stopwatch times to evaluate the precision. As far as we know, none of the automated systems utilizes only one Kinect (version 2) to capture all TUG components. And also, none of the F8W studies have used the Kinect before.

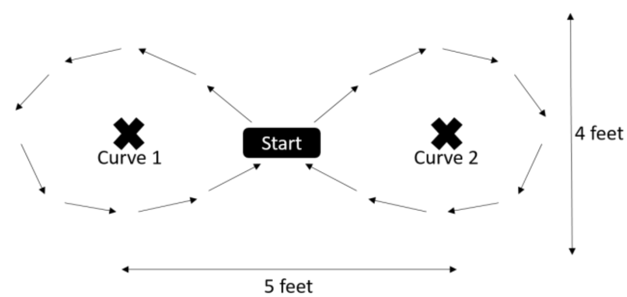

Figure 1. Figure-of-8 Walk Test Floor setup.

In this work, the Kinect skeletal model and depth analysis were evaluated for gait analysis during the TUG and F8W, which have been proven to be useful for evaluating gait impairment and other related functional limitations [1-4]. A Vicon motion capture system was used for ground truth. Furthermore, we plan to integrate the method with the higher accuracy for each gait parameter into the automated assessment application we are currently developing. The hybrid system that uses both Kinect skeletal data and depth data analysis will generate results in real-time for clinical usage.

Section II of this paper starts with a brief description of the Microsoft Kinect, followed by an explanation of the system setup for capturing the TUG and F8W, the gait measures used in this work and their definition, the skeletal data analysis, the depth analysis, and the Vicon analysis for ground truth. Section III contains the results of a pilot study conducted for evaluation purposes including mean error compared with the Vicon system. Section IV contains a brief discussion of the results. Section V contains a conclusion and future work.

*This work is supported by AHRQ, grant# R01HS022889.

B. Hotrabhavananda, A. K. Mishra, M. Skubic and N. Hotrabhavananda are with the Electrical and Computer Engineering Department at the University of Missouri, Columbia, MO 65211, USA (Phone: 573-239-9127; emails:bh6b6@mail.missouri.edu, akmm94@mail.missouri.edu, skubicm@missouri.edu, nhqp8@mail.missouri.edu).

C. Abbott is with the School of Health Professions, Physical Therapy, University of Missouri, Columbia, MO 65211, USA (e-mail: AbbottC@health.missouri.edu).

II. Methodology

A. Microsoft Kinect

The Kinect is one of the widely used sensors in the field of gait and balance analysis [5-10]. The version 2 Kinect depth camera uses a time-of-flight phase detection system to get depth data [11], generating depth images independent of ambient lighting. The Kinect comes with a software development kit (SDK) that fits a 25-point skeletal model to segmented human bodies in depth image frames [12]. Fig. 2(a) and 2(b) shows the Microsoft Kinect version 2.0 sensor and a skeletal model overlaid on a depth image.

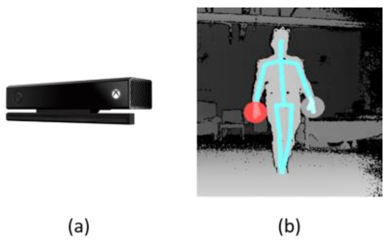

Figure 2: (a): Microsoft Kinect version 2.0; (b): Skeletal model overlaid on a depth image captured by the Kinect

B. System Setup

A Windows application has been developed to quantitatively assess the mobility of participants during the TUG and F8W. We set up one Kinect as in Fig. 3 to capture the entire walk of the TUG and F8W. Within the application, we created a guide to help users set up the Kinect, the chair, and the walking path in a standard configuration as shown in Fig. 3. A few benefits of using the application include real-time reports, ability to replay and analyze the video, the ability to save results in a database, and the ability to record trajectories of each skeletal joint location captured from the Kinect SDK transformed into world space coordinates.

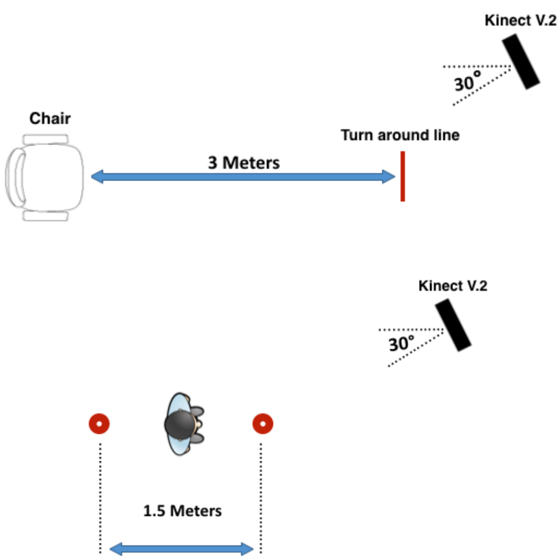

Figure 3. **Top:** Setup for TUG. **Bottom:** Setup for F8W

C. Gait Parameters

In this paper, we focus on measuring the stride length, stride time and walking speed during the TUG, and the number of steps and completion time during the F8W. For TUG, the gait parameters are separated into two portions, walking toward the Kinect and walking away from Kinect. To calculate the stride length, stride time, and walking speed, we follow the standard definition defined by physical therapists. The gait parameters of the TUG for walking toward the Kinect are captured when a subject takes the first step and before a subject starts rotating her body at the line. For walking away from Kinect, the gait parameters are evaluated when a subject has fully finished rotating the body at the line and right before the subject starts rotating her body at the chair to sit down.

For F8W, the number of steps are the total number of steps that a subject takes to complete the walk, and the completion time is calculated from the time when a subject takes the first step until the time a subject returns to the starting point.

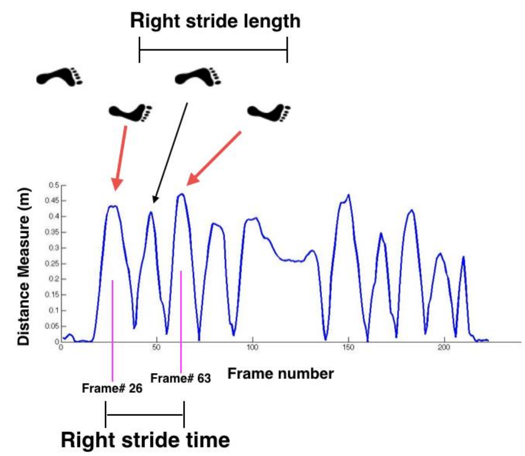

Figure 4. A plot of distance measure between left foot and right foot showing the key frames pointed by red arrows to calculate the right foot stride length and stride time

D. Skeletal Data Analysis

The Kinect versions 1 and 2 suffer from a limited range to capture the entire TUG reliably using the skeletal model, especially for the lower part of the body with only one Kinect. The further the sensor is positioned from the subject, the more joint instability occurs. However, with our current setup, we are able to capture the skeletal model during the TUG and F8W reliably enough to compare with the ground truth data.

To extract the stride length, stride time, number of steps, and the completion time from the Kinect skeletal data, we use the left foot and right foot joint locations as inputs to our algorithm. The algorithm finds the distances between left foot and right foot as participants walk during the TUG and F8W. These are plotted by frame number to visualize the key frames when the two feet are most separated from each other as shown in Fig. 4. The figure illustrates the estimation of the frame numbers used in computing stride length. In this example, the position of the right foot at frame 26 and frame 63 are used to compute the right stride length at this stride. For the stride time, we find left stride time (left footfall to left footfall) and right stride time (right footfall to right footfall) by calculating the number of frames between the first peak and the third peak divided by the frame rate of the Kinect, 30 frames per second. For the number of steps during the F8W, the system counts the number of peaks during the walk. One more foot step is added to the final result, for the subject returning to the starting location, as the last step has no corresponding distance peak.

To calculate the walking speed, the algorithm uses the spine base joint. Distance traveled is estimated and divided by the time for each portion.

E. Depth Data Analysis

For comparison, we also tested the depth images directly, for extraction of gait parameters. The subject in a depth image frame was segmented using the body index frames, provided by the Kinect SDK [12]. The body index frames provide a unique identity code for each pixel in the depth frame that represents a human. For multiple people in view, a unique identity represents each person. This helps in distinguishing between the physical therapy (PT) client and the therapist who may walk next to a frail client.

Stone and Skubic [10] have successfully used the Kinect depth images to extract stride parameters in the home. In this study, we have adapted their algorithm with a few modifications to make it robust for our scenario. As the first step, each depth frame was processed to get the 3D point cloud of the subject. The centroids of the 3D point clouds for each frame were projected on the ground. Then, a moving mean filter was applied to the centroid time series to smooth the centroid locations and reduce noise. To extract the gait parameters, such as stride length and stride time, the 3D points with a height between 18 to 2 inches were projected onto the floor plane. The projected 3D points were normalized and centered for each frame. Considering each frame has *N* points, the correlation coefficient for each frame was computed as:

$$\rho = \frac{\sum_{n=1}^{N} x_n y_n}{N} \tag{1}$$

where x_n and y_n represent the X and Y coordinates of the n^{th} point in the 3D point projection. The correlation coefficient time series is used to determine the steps in a walking sequence. The signal was filtered using a moving mean filter with a window size given by:

$$w = \frac{f * k}{v} \tag{2}$$

where f is the frame rate of depth data collection (f =30), k is a constant parameter (5.6), and v is walking speed. The walking speed is obtained by the total distance travelled by the centroid, dividing by elapsed time, excluding the turnarounds, sit-to-stand, and stand-to-sit frames. The correlation time signal was again filtered with a moving mean filter with a smaller window size to remove the remaining minor local extrema. Fig. 5 shows examples of right and left foot step point projections 5(a), along with the corresponding correlation time signal 5(b), showing the raw and filtered signal in red and blue, respectively. Moreover, in 5(b) a gray box represents the turnaround frames. Before the turnaround time, the local maxima and the local minima correspond to right and left footsteps, respectively, whereas, after the turn, maxima and minima correspond to left and right footsteps, respectively.

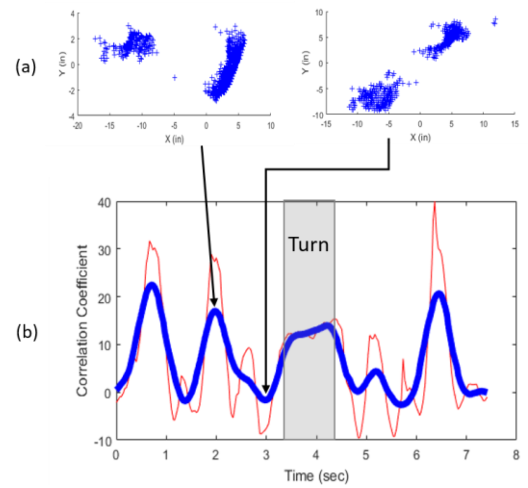

Figure 5. (a): Centered and Normalized ground plane projection of points with a height between 18 to 2 inches at two different time instances. (b): Raw (red) and Filtered (blue) Correlation Coefficient time series for a walk. The gray area shows the TUG turn around. Before the turn, local maxima correspond to right footsteps and the local minima correspond to the left footsteps, whereas, after the turn, this is reversed.

The stride parameters, such as stride length and stride time, are estimated considering the correlation coefficient time series and foot fall locations. The stride time is measured as the time between successive footfalls of the same foot in the correlation coefficient time series. The stride length is approximated calculating the distance travelled by the body centroid between two successive footfalls of the same foot.

F. Vicon Analysis

A marker-based Vicon motion capture system is used to validate both skeletal data and depth data analysis. The stride length and stride time are extracted using the same approach as the Kinect skeletal data. The markers were placed at the toe of participants to mimic the foot joint location of the Kinect skeletal model. For the walking speed, we use the mid-point between the Right Posterior Superior Iliac (RPSI) and the Left Posterior Superior Iliac (LPSI), and find its total distance travel and divided by the elapsed time.

III. Results

To evaluate the accuracy of the Kinect skeletal model and the depth analysis for capturing the gait parameters, 5 healthy participants were recruited to take part in an IRB approved human subjects study, with ages from 18 to 54 years. The Kinect setup is shown in Fig. 3. Locations of the feet and spine base from Kinect skeletal model and the locations of feet, RPSI, and LPSI from the Vicon system were recorded and processed. Tables I-III show the results of the stride length, the stride time, and the walking speed for the Kinect skeletal model and Kinect depth images, compared to the Vicon as ground truth during the TUG. Table IV shows the results of the number of steps and the completion time during the F8W; μ is the average percentage error compared with the Vicon and σ is the percentage standard deviation.

TABLE I. AVERAGE % ERROR: TUG RIGHT STRIDE

	Skeletal Model		*Kinect Depth*	
	μ	σ	μ	σ
Stride Length (%)	13.62	4.77	10.14	2.74
Stride Time (%)	7.11	6.93	5.56	5.76

TABLE II. AVERAGE % ERROR: TUG LEFT STRIDE

	Skeletal Model		*Kinect Depth*	
	μ	σ	μ	σ
Stride Length (%)	16.86	12.15	4.66	11.02
Stride Time (%)	11.40	7.39	-2.29	7.42

TABLE III. AVERAGE % ERROR: TUG WALKING SPEED

	Skeletal Model		*Kinect Depth*	
	μ	σ	μ	σ
Walking Speed (%)	-0.99	6.54	2.54	1.81

TABLE IV. AVERAGE % ERROR: F8W

	Skeletal Model		*Kinect Depth*	
	μ	σ	μ	σ
Number of Steps (%)	1.91	5.61	-2.96	2.75
Time (%)	-1.34	2.75	2.75	8.93

IV. DISCUSSION

The result of this study indicates that the Kinect depth data analysis offers a better accuracy for evaluating the stride length and stride time during the TUG, and also offers a comparable accuracy with the Kinect skeletal model for evaluating walking speed for the TUG, and the number of steps and the completion time for F8W. However, the skeletal model is slightly better than the depth data analysis for evaluating the walking speed during the TUG, and the number of steps and the completion time during the F8W. But, due to the distance between the subject and the Kinect and the occlusion of the lower right side of the body during the walk, both the Kinect skeletal model and the depth data analysis tend to have a higher % error for the right stride length during the TUG: 13.62±4.77 and 10.14±2.74 respectively. The possible solution to this problem is to change the angle of the Kinect. However, this will result in inaccuracies when capturing movement at the ends of the walking path (i.e., at the chair and the turn). This option will be investigated further in future studies. For the walking speed, even though the % error of the skeletal model is better than the Kinect depth analysis, the standard deviation shows that the Kinect skeletal model has less consistency than the Kinect depth analysis, which may be caused by a reduced reliability of the Kinect skeletal tracking system.

Finally, the main limitation of this study includes the sample size, and the small number of older adults. We anticipate that the gait results may be different for older adults, as they tend to take a shorter stride length and a longer stride time than the participants in this study.

V. CONCLUSION

We reviewed our methods to capture gait parameters in the TUG and F8W using both Kinect skeletal data and depth data. We performed an experiment with 5 healthy subjects to validate the stride length and stride time during the TUG test extracted from Kinect skeletal data and Kinect depth data against a Vicon Motion capture system. Through the use of a depth data algorithm, we can improve the accuracy of the stride length and stride time extraction during the TUG test.

In the future, we plan to include a larger and more varied sample for further validation of Kinect skeletal model and depth analysis for TUG and F8W. Also, we plan to inspect the scenarios where participants use walking aids and a second person during the tests.

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, pp. 142-148, 1991.

[2] R. J. Hess, J. S. Brach, S. R. Piva, and J. M. VanSwearingen, "Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test," *Physical Therapy*, vol. 90, pp. 89-99, 2010.

[3] J. VanSwearingen, J. Brach, and R. Hess, "Clinical correlates of motor control in walking: the Figure-of-8 Walk Test," in *59th Annual Scientific Meeting of the Gerontological Society of America*, 2006, pp. 16-20.

[4] C. A. Odonkor, J. C. Thomas, N. Holt, N. Latham, J. S. Brach, S. G. Leveille, et al., "A comparison of straight-and curved-path walking tests among mobility-limited older adults," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 68, pp. 1532-1539, 2013.

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

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

[7] E. Stone, M. Skubic, M. Rantz, C. Abbott, and S. Miller, "Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders," *Gait & Posture*, vol. 41, pp. 57-62, 2015.

[8] A. K. Mishra, M. Skubic, and C. Abbott, "Development and preliminary validation of an interactive remote physical therapy system," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 190-193.

[9] B. F. Mentiplay, L. G. Perraton, K. J. Bower, Y.-H. Pua, R. McGaw, S. Heywood, et al., "Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables," *Journal of Biomechanics*, vol. 48, pp. 2166-2170, 2015.

[10] E. Stone and M. Skubic, "Evaluation of an inexpensive depth camera for in-home gait assessment," *Journal of Ambient Intelligence and Smart Environments*, vol. 3, pp. 349-361, 2011.

[11] J. Sell and P. O'Connor, "The Xbox One System on a Chip and Kinect Sensor," *Micro, IEEE*, vol. 32, pp. 44-53, 2014.

[12] M. Corporation, "Kinect for Windows SDK 2.0 Documentation," 2014.