Automated Timed Up and Go Test Segmentation via Pose Detection

Aram Sibar Algilani*, Sebastian Victor Aguilar*, Dario Salvi*, and Filippo Palumbo†
*Department of Computer Science and Media Technology, Sustainable Research Centre, Malmö University, Malmö, Sweden
†Information Science and Technologies Institute, National Research Council (ISTI-CNR), Pisa, Italy

***Abstract*—The Timed Up and Go test is commonly used for assessing gait, balance and function in people with limited mobility or low functional mobility. The test consists in measuring the time it takes for a subject to stand up, walk 3 meters, come back and sit down. Computer vision and pose estimation allows measuring the subject's performance more reliably than a human observer and to extract additional parameters that are potentially more sensitive for the assessment of symptoms than the total duration. We show an algorithmic approach for multi-camera settings and present accuracy results obtained in a pilot study with 13 older adults, establishing the methodology for larger studies.**

***Index Terms*—Timed Up and Go, Pose Estimation, Computer Vision, Mobility Assessment, Gait Analysis, Multi-camera System.**

I. Introduction

Assessing mobility and fall risk is a critical aspect of clinical care, especially for older adults and individuals with movement impairments [1]. The Timed Up and Go (TUG) test is a widely adopted clinical tool for evaluating functional mobility, balance, and fall risk [2], [3]. It measures the time taken for an individual to rise from a standard chair, walk $3\,\mathrm{m}$ at their normal pace, turn $180°$, walk back to the chair, and sit down, as illustrated in Fig. 1.

Traditional TUG assessments often rely on manual stop-watch timing by a human observer. This method can introduce inter-rater variability and lacks the precision to accurately segment the distinct movement phases within the TUG test (e.g., sit-to-stand, walking, turning, stand-to-sit) [4], [5]. This limitation hinders the objective comparison of patient performance over time and across different assessments, and it restricts the depth of analysis to the total time taken. Recent advancements in computer vision and pose estimation, particularly markerless 3D human pose estimation, offer a promising avenue to overcome these challenges by enabling automated, objective, and detailed analysis of the TUG test [6], [7].

Computer vision systems, especially those employing 3D pose estimation from multi-camera setups, can provide a more comprehensive analysis than simple timing. These systems allow for the extraction of various kinematic parameters, such as joint angles (e.g., knee flexion), segmental velocities, step characteristics, and postural stability during different phases of the test [7]. Recently, several efforts have been made in research towards such systems [8], [9]; these additional parameters may be more sensitive to assess subtle changes in symptoms, tracking rehabilitation progress, or identifying specific areas of difficulty for an individual [10].

This paper presents an algorithmic approach for the automated analysis of the TUG test using a multi-camera computer vision system. We describe the system architecture, data acquisition using standard RGB cameras, 3D pose estimation leveraging a real-time tracking algorithm, and the subsequent rule-based algorithmic segmentation of the TUG test into its constituent phases. We present detailed accuracy and agreement results from a pilot study involving 13 older adults, where our vision-based system was compared against a commercial inertial measurement unit (G-Walk) and manual video annotation (ground truth). This work establishes the methodology and demonstrates the potential for larger studies aiming for reliable and enhanced mobility assessment.

II. Related Work

The automation of the Timed Up and Go (TUG) test has been an active area of research, with various sensor modalities and computational approaches being explored.

Li et al. [11] proposed a video-based method for TUG sub-task segmentation in Parkinson's disease patients using a single 2D camera. Their approach involved 2D human pose estimation, followed by Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) networks for activity classification, with Dynamic Time Warping (DTW) to refine segmentation. While also video-based, our work differs by employing a multi-camera system to reconstruct 3D pose data, aiming for richer kinematic information, and uses a rule-based algorithm for segmentation rather than machine learning classifiers for this specific task. Our focus is on older adults generally, not specifically Parkinson's patients.

Choi et al. [2] introduced a deep learning-based method for TUG sub-task segmentation using RGB-D cameras. They utilized a Temporal Convolutional Network (TCN) and demonstrated that pelvic motion data alone could achieve high accuracy in real-time. In contrast, our system uses multiple standard RGB cameras for 3D reconstruction via a proprietary algorithm for triangulation of 2D keypoints, and applies a rule-based segmentation logic rather than an end-to-end deep

Corresponding author: F. Palumbo (email: filippo.palumbo@isti.cnr.it).

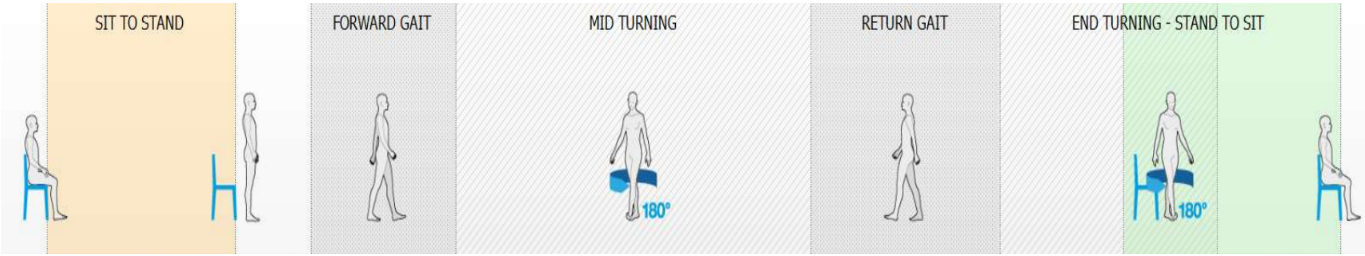

Fig. 1. Illustration of the Timed Up and Go (TUG) test protocol, showing the sequence of movements: sit-to-stand, walk forward, turn, walk back, and stand-to-sit (Figure obtained from the G-Studio software, BTS Bioengineering).

learning model for phase identification or relying solely on RGB-D data.

Savoie et al. [4] demonstrated TUG automation using a single conventional RGB video camera. They extracted 3D poses using Mask R-CNN for 2D keypoints and a Deep Multitask Architecture for Human Sensing (DMHS), then applied heuristic features for phase segmentation, achieving low error rates. Their approach of deriving 3D pose from RGB video and using heuristics for segmentation shares similarities with our methodology. Our system is designed around a multi-camera setup to potentially enhance 3D pose accuracy and robustness, particularly for complex movements, and uses a different pipeline for 3D triangulation. Dubois et al. [12] also utilized a depth camera and human pose estimation to automate the TUG test, focusing on the feasibility of such systems in clinical practice.

More recently, Wen et al. [13] proposed an Encoder-Decoder Graph Convolutional Network (GCN) for automatic TUG and Sit-to-Stand (STS) segmentation. This approach leverages advanced deep learning techniques by representing pose data as graphs and processing them with GCNs. This contrasts with our work, which relies on a more traditional computer vision pipeline for 3D reconstruction followed by an explicit rule-based algorithm for defining and detecting TUG phases, rather than a data-driven GCN model for segmentation.

These studies underscore the trend towards vision-based systems for objective mobility assessment and the diverse strategies for TUG automation. Our work contributes by detailing a multi-camera 3D vision system with rule-based segmentation, aiming to provide robust and interpretable kinematic analysis from standard RGB video.

III. Methodology

Our approach involves a multi-camera system for data acquisition, 3D pose estimation, and an algorithmic process for segmenting the TUG test into its constituent phases. An overview of the system pipeline is shown in Fig. 2.

A. Experimental Setup and Data Acquisition

The TUG tests were recorded using a multi-camera system composed of four AXIS M3045-V Full HD network cameras on a dataset involving 13 older adults ho repeated the test twice. The cameras were strategically positioned to ensure comprehensive coverage of the participant's movement throughout the 3-meter TUG path and were synchronized using Network Time Protocol (NTP) to ensure precise temporal alignment of video streams. Calibration procedures, including intrinsic and extrinsic calibration, were performed to establish a common coordinate system and enable accurate 3D reconstruction. Reference systems included a G-Walk inertial sensor (BTS Bioengineering) on the lower back (L5 vertebra) recording at $100\,\mathrm{Hz}$, and manual frame-by-frame video annotation for ground truth (GT). Facial anonymization using MTCNN and Gaussian blur was applied for protecting the privacy of the participants.

B. 3D Pose Estimation and Processing

Each video frame was processed using a proprietary 3D pose real-time tracking system developed by Sony. The system has the ability to track multiple people in real time, it features automatic reconstruction in 3D of 18 keypoints per person, it provides robust output with 3 or more cameras, it is tolerant against occlusion and can measure absolute lengths/distances with a $5\,\mathrm{cm}$ accuracy. The keypoints correspond to major joints (e.g., hips, knees, ankles, shoulders) and follow a consistent anatomical order for each identified user.

The resulting 3D keypoint sequences were temporally aligned and interpolated to handle occasional dropped frames. A moving average filter of three frames was applied to critical features such as the x-position of the hips and the z-height of the hip joint to reduce motion jitter and enhance phase transition detection. Knee angles were computed using the scalar product of the thigh and shin vectors for each frame, offering a biomechanical basis for posture classification.

C. TUG Phase Segmentation Algorithm

A rule-based algorithm segments the TUG into five phases based on biomechanical markers from the 3D pose data:

- **Sit-to-Stand (SiSt):** *Start:* Knee flexion $< 100°$ AND hip height $< 0.68\,\mathrm{m}$. *End:* Knee flexion $> 120°$ AND hip height $> 0.77\,\mathrm{m}$.
- **Walk Forward (Walk1):** Starts after SiSt. Characterized by steady forward displacement of hip x-coordinate. Ends at turn initiation.
- **Turn:** Identified by reversal in hip x-direction combined with lateral displacement in y-direction.
- **Walk Back (Walk2):** Starts after Turn. Characterized by movement in reversed x-direction. Ends at sit initiation.

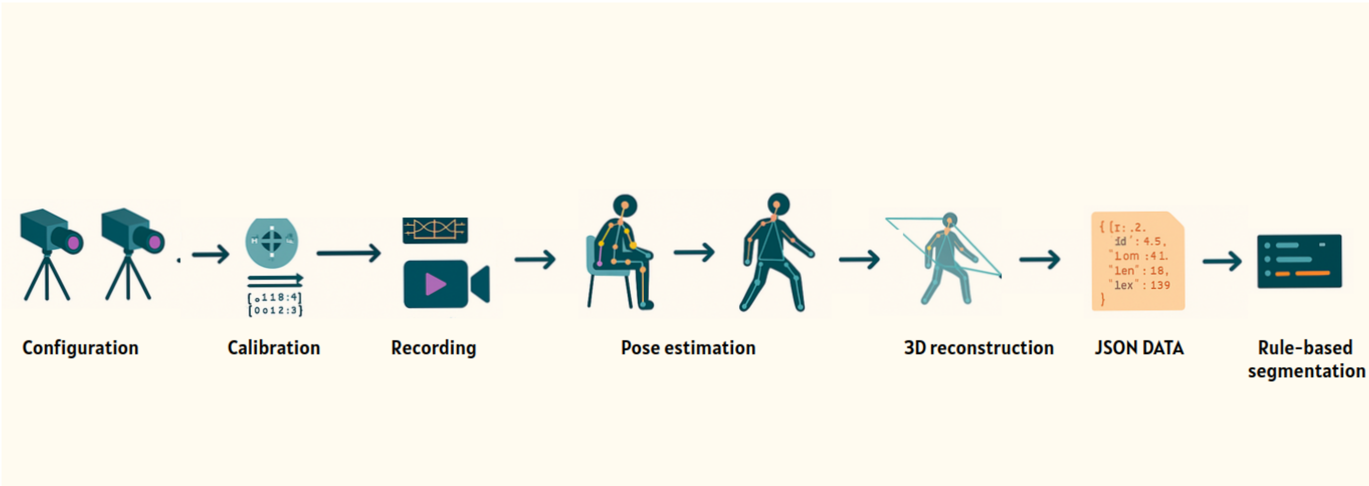

Fig. 2. Overview of the vision-based TUG analysis system pipeline, from multi-camera configuration and calibration recording, through pose estimation and 3D reconstruction (resulting in JSON data), to rule-based segmentation.

- **Stand-to-Sit (StSi):** *Start:* Decrease in hip height and knee flexion after Walk2. *End:* Hip height $< 0.65\,\mathrm{m}$.

The total TUG duration is the sum of these phase durations.

D. Data Analysis

Agreement between the vision system, G-Walk, and Ground Truth (GT) for phase durations was assessed using:

- Mean Absolute Error (MAE).
- Standard Deviation (SD) of the absolute error.
- Intraclass Correlation Coefficient (ICC) (i.e., ICC(2,1) for absolute agreement).
- Bland-Altman plots for total TUG duration.

IV. Results

This section presents the accuracy and agreement of our vision-based system and the G-Walk sensor, both compared against manual video annotation (Ground Truth), for the 13-participant study, 2 tests each. In total 25 videos were analyzed (one video had to be discarded due to low quality).

A. Comparison of Phase Durations

Table I summarizes the MAE, SD of error, and ICC values for each TUG phase and the total duration when comparing our vision-based system and G-Walk against the Ground Truth.

The vision-based system demonstrated lower MAE and SD of error, and higher ICC values across all TUG phases and for the total duration compared to the G-Walk system when evaluated against the ground truth. For the Total TUG Duration, our system showed an MAE of $0.31\,\mathrm{s}$ and an excellent ICC of 0.89 , while G-Walk had an MAE of $0.59\,\mathrm{s}$ and an ICC of 0.77

The results across different sub-tasks also confirms the higher accuracy of the vision-based system, for example for Sit-to-Stand, the System achieved an MAE of $0.26\,\mathrm{s}$ (ICC=0.84) compared to G-Walk's MAE of $0.33\,\mathrm{s}$ (ICC=0.72). The vision system showed substantial improvement also in the Mid Turn, Return gait and End turn phases.

TABLE I
COMPARISON OF G-WALK AND VISION SYSTEM AGAINST GROUND TRUTH (GT) FOR TUG PHASE DURATIONS (N=25 VIDEOS, 13 PARTICIPANTS).

TUG Phase	**G-Walk vs GT**			**Vision System vs GT**		
	MAE (s)	**SD (s)**	**ICC**	**MAE (s)**	**SD (s)**	**ICC**
Sit-to-Stand	0.33	0.32	0.72	0.26	0.23	0.84
Walk Forward	0.78	0.67	0.61	0.42	0.48	0.85
Mid Turn	1.15	0.91	0.42	0.65	0.45	0.79
Return Gait	0.90	0.74	0.56	0.48	0.39	0.80
End Turn	0.71	0.62	0.64	0.41	0.38	0.82
Stand-to-Sit	0.40	0.36	0.68	0.35	0.33	0.73
Total Duration	**0.59**	**0.45**	**0.77**	**0.31**	**0.27**	**0.89**

B. Bland-Altman Analysis of Agreement

Bland-Altman plots were used to visualize the agreement for total TUG duration between each system and the ground truth, as shown in Fig. 3 and Fig. 4. The plot for G-Walk vs Ground Truth (Fig. 3) showed a mean difference (bias) of approximately $0.61\,\mathrm{s}$ (G-Walk tending to record longer durations on average), with 95% limits of agreement spanning a wider range. The plot for our vision-based System vs Ground Truth (Fig. 4) showed a smaller mean difference (bias) of approximately $0.39\,\mathrm{s}$ (the vision-based system also tending to record slightly longer durations) and narrower 95% limits of agreement, indicating less random error and tighter alignment with true timings.

C. Clinically Relevant Parameters

Beyond phase timing, the vision-based system allows extraction of other clinically relevant parameters:

- **Knee Flexion Angle:** Tracked throughout, crucial for defining SiSt and StSi transitions. Reduced/delayed extension can indicate weakness or joint issues [14].
- **Hip Vertical Displacement (Z-Coordinate):** Used to assess posture changes during SiSt and StSi.

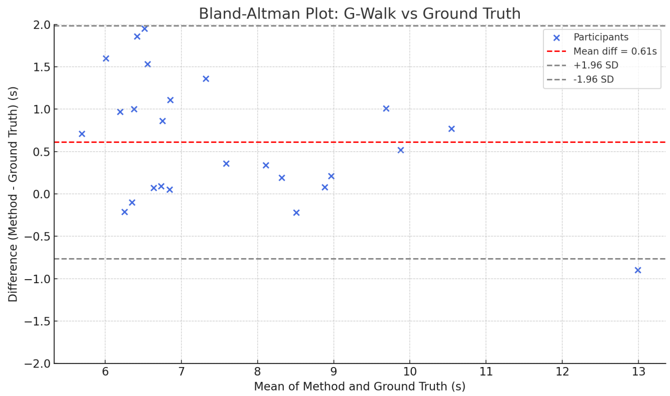

Fig. 3. Bland-Altman plot comparing total TUG duration from G-Walk system against Ground Truth (N=25). The blue points represent individual participants.

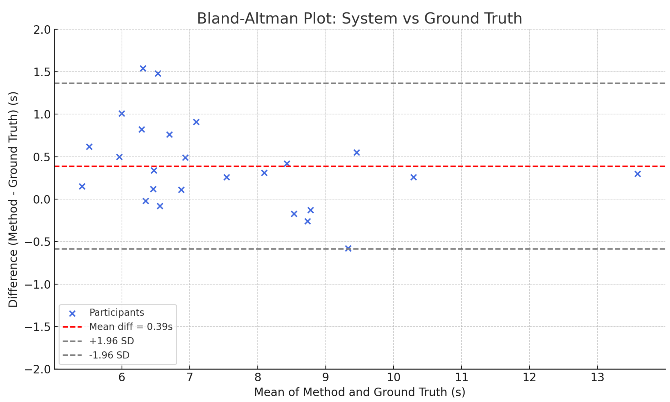

Fig. 4. Bland-Altman plot comparing total TUG duration from the vision-based System against Ground Truth (N=25). The blue points represent individual participants.

- **Gait Direction and Turning Detection:** Based on hip X-Y trajectory, important for evaluating balance and coordination, especially in conditions like Parkinson's disease [15].

The proposed vision-based system has the potential to extract further parameters such as stride length, cadence, step time asymmetry, postural sway, arm swing symmetry, turning radius/speed, and gait velocity, all of which have clinical significance, and with simple rules-based algorithms.

V. Discussion

The results from our study with 13 older adults demonstrate that the proposed multi-camera vision-based system can accurately and reliably segment the TUG test and measure phase durations.

Interpretation of Accuracy and Reliability: Our vision-based system consistently outperformed the G-Walk inertial sensor when compared to manual video annotation (Ground Truth) across all TUG phases and for the total duration (Table I). The System achieved lower MAEs, smaller SDs of error, and higher ICCs, indicating better accuracy and consistency. The ICC values for the proposed system were mostly in the "good" to "excellent" range (0.73-0.89), suggesting strong reliability for clinical use. This is a significant improvement over the G-Walk, which showed moderate agreement, particularly struggling with the Mid Turn phase (ICC=0.42). The proposed system's MAE for Total Duration was $0.31\,\mathrm{s}$ with an ICC of 0.89, supporting its potential as an alternative to manual or other sensor-based timing.

The Bland-Altman analysis (Fig. 3 and Fig. 4) further supported these findings, showing that our system had a smaller bias and narrower limits of agreement with the Ground Truth for total TUG time compared to the G-Walk. This indicates that the vision-based system provides measurements that are not only closer to the true values on average but also more consistent across participants. Importantly, the limits of agreement identified in our analysis are below the lowest of the minimum clinically significant differences (MCID) identified for the TUG test in several conditions: $3.40\,\mathrm{s}$ [16]–[18].

Strengths of the Multi-Camera Vision Approach:

- **Enhanced Accuracy in Complex Phases:** The system particularly excelled in segmenting spatially complex phases like turning, where single IMUs often struggle. This is likely due to the rich 3D spatial information captured by multiple cameras.
- **Objective and Detailed Measurement:** The automated nature reduces observer bias, and phase segmentation provides a granular view of mobility not available from total time alone.
- **Rich Kinematic Data Extraction:** The 3D pose data allows for extraction of numerous kinematic parameters beyond timing (e.g., joint angles, trajectories, gait characteristics). This offers potential for a more comprehensive mobility assessment.
- **Non-intrusive and Scalable:** Being markerless and not requiring body-worn sensors, the system is less obtrusive for participants and potentially more scalable for clinical or home-based settings.

Limitations and Challenges:

- **Dependence on Pose Estimation Quality:** System accuracy is tied to the underlying Sony 3D pose estimation algorithm. Factors like occlusions, lighting, and participant clothing can still impact keypoint detection, although the multi-camera setup and the Sony algorithm's robustness mitigate this to a large extent.
- **Rule-Based Segmentation Sensitivity:** While effective, rule-based algorithms may require tuning for different populations or environments and can be challenged by highly atypical movement patterns. The Mid Turn phase, despite improvement, remained the most challenging.
- **Setup and Calibration Complexity:** Multi-camera systems inherently require more setup and calibration effort than single sensors, though the Sony system aims to simplify this.
- **Generalizability:** The current detailed results are from 13 older adults. While promising, performance in larger

participant studies and across more diverse populations (e.g., specific neurological conditions) needs to be confirmed.

Clinical Implications: The ability to accurately segment TUG phases and extract detailed kinematics offers significant clinical potential. It could help identify specific impairments (e.g., difficulty initiating movement, slow turning, instability during transitions), guide targeted interventions, and provide more sensitive outcome measures for rehabilitation or disease progression monitoring. The extraction of parameters like step asymmetry or turning velocity could offer insights not captured by traditional TUG timing.

VI. Conclusion

This paper presented a multi-camera computer vision system for automated segmentation and analysis of the Timed Up and Go test. Based on a study with 13 older adults, our system, which leverages a sophisticated 3D pose tracking algorithm and rule-based phase detection, demonstrated superior accuracy and reliability compared to a commercial G-Walk inertial sensor when evaluated against manual video annotation. The vision system achieved lower errors (e.g., MAE of $0.31\,\mathrm{s}$ for total TUG time) and higher agreement (e.g., ICC of 0.89 for total TUG time).

The findings confirm that computer vision, specifically multi-camera 3D pose estimation, can provide precise, objective, and detailed TUG assessments. The system's ability to accurately segment phases, especially complex ones like turning, and to enable the extraction of rich kinematic parameters beyond simple timing, highlights its potential to enhance clinical understanding of mobility impairments.

Future work will focus on reporting the results from larger cohorts, further refining segmentation algorithms (potentially incorporating machine learning), and exploring the clinical utility of the additional extracted biomechanical parameters in diverse populations. The continued development of such non-intrusive, vision-based assessment tools holds considerable promise for advancing mobility evaluation in both clinical and remote monitoring settings.

VII. Acknowledgments

This research is partially funded by the Swedish Knowledge Foundation and the Sustainable Digitalization Research Center (SDRC) through the Synergy project Intelligent and Trustworthy IoT Systems and by the European Union - Next Generation EU, in the context of The National Recovery and Resilience Plan, Investment 1.5 Ecosystems of Innovation, Project Tuscany Health Ecosystem (THE), CUP: B83C22003930001.

References

[1] J. D. Blasco-García, G. García-López, M. Jiménez-Muñoz, J. A. López-Riquelme, J. J. Feliu-Batlle, N. Pavón-Pulido, and M.-T. Herrero, "A computer vision-based system to help health professionals to apply tests for fall risk assessment," *Sensors*, vol. 24, no. 6, p. 2015, 2024.

[2] Y. Choi, Y. Bae, B. Cha, and J. Ryu, "Deep learning-based subtask segmentation of timed up-and-go test using rgb-d cameras," *Sensors*, vol. 22, no. 17, p. 6323, 2022.

[3] 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, 1991.

[4] P. Savoie, J. A. Cameron, M. E. Kaye, and E. J. Scheme, "Automation of the timed-up-and-go test using a conventional video camera," *IEEE journal of biomedical and health informatics*, vol. 24, no. 4, pp. 1196–1205, 2019.

[5] Y. Higashi, K. Yamakoshi, T. Fujimoto, M. Sekine, and T. Tamura, "Quantitative evaluation of movement using the timed up-and-go test," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 4, pp. 38–46, 2008.

[6] S. Bi, "Computer vision based 3d reconstruction and motion tracking algorithm in video processing," in *2024 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE)*. IEEE, 2024, pp. 887–893.

[7] A. Avogaro, F. Cunico, B. Rosenhahn, and F. Setti, "Markerless human pose estimation for biomedical applications: a survey," *Frontiers in Computer Science*, vol. 5, p. 1153160, 2023.

[8] C. Francia, F. Motta, L. Donno, M. Covarrubias, C. Dornini, A. Madella, and M. Galli, "Validation of a mediapipe system for markerless motion analysis during virtual reality rehabilitation," in *Extended Reality*, L. T. De Paolis, P. Arpaia, and M. Sacco, Eds. Cham: Springer Nature Switzerland, 2024, pp. 40–49.

[9] K. Song, T. J. Hullfish, R. Scattone Silva, K. G. Silbernagel, and J. R. Baxter, "Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements," *Journal of Biomechanics*, vol. 157, p. 111751, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021929023003214

[10] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, I. Maidan, N. Giladi, T. Gurevich, and J. M. Hausdorff, "Can an accelerometer enhance the utility of the timed up & go test when evaluating patients with parkinson's disease?" *Medical engineering & physics*, vol. 32, no. 2, pp. 119–125, 2010.

[11] T. Li, J. Chen, C. Hu, Y. Ma, Z. Wu, W. Wan, Y. Huang, F. Jia, C. Gong, S. Wan *et al.*, "Automatic timed up-and-go sub-task segmentation for parkinson's disease patients using video-based activity classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2189–2199, 2018.

[12] A. Dubois, T. Bihl, and J.-P. Bresciani, "Automating the timed up and go test using a depth camera," *Sensors*, vol. 18, no. 1, p. 14, 2017.

[13] B. Wen, C. Du, and T. Q. Nguyen, "Encoder-decoder graph convolutional network for automatic timed-up-and-go and sit-to-stand segmentation," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[14] F. Cicuttini, A. Wluka, J. Hankin, and Y. Wang, "Longitudinal study of the relationship between knee angle and tibiofemoral cartilage volume in subjects with knee osteoarthritis," *Rheumatology*, vol. 43, no. 3, pp. 321–324, 2004.

[15] M. Mancini, A. Weiss, T. Herman, and J. M. Hausdorff, "Turn around freezing: community-living turning behavior in people with parkinson's disease," *Frontiers in neurology*, vol. 9, p. 18, 2018.

[16] O. P. Gautschi, M. N. Stienen, M. V. Corniola, H. Joswig, K. Schaller, G. Hildebrandt, and N. R. Smoll, "Assessment of the minimum clinically important difference in the timed up and go test after surgery for lumbar degenerative disc disease," *Neurosurgery*, vol. 80, no. 3, pp. 380–385, 2017.

[17] J. D. Ries, J. L. Echternach, L. Nof, and M. Gagnon Blodgett, "Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with alzheimer disease," *Physical therapy*, vol. 89, no. 6, pp. 569–579, 2009.

[18] S.-L. Huang, C.-L. Hsieh, R.-M. Wu, C.-H. Tai, C.-H. Lin, and W.-S. Lu, "Minimal detectable change of the timed "up & go" test and the dynamic gait index in people with parkinson disease," *Physical therapy*, vol. 91, no. 1, pp. 114–121, 2011.