Gait & Posture 90 (2021) 489–495

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Full length article

Extraction of gait parameters from marker-free video recordings of Timed Up-and-Go tests: Validity, inter- and intra-rater reliability

Anna Cristina Åberg [a,b,*], Fredrik Olsson [c], Hanna Bozkurt Åhman [a], Olga Tarassova [d], Anton Arndt [d,e], Vilmantas Giedraitis [a,b], Lars Berglund [a,b], Kjartan Halvorsen [b,f]

[a] Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, BMC, Box 564, SE-751 22 Uppsala, Sweden
[b] School of Health and Welfare, Dalarna University, SE-791 88 Falun, Sweden
[c] Department of Engineering Mechanics, KTH Royal Institute of Technology, SE-114 28 Stockholm, Sweden
[d] The Swedish School of Sport and Health Sciences, Lidingövägen 1, SE-11486 Stockholm, Sweden
[e] Karolinska Institutet, SE-17177 Stockholm, Sweden
[f] Department of Mechatronics, School of Engineering and Sciences, Campus Estado de Mexico, Tecnologico de Monterrey, Atizapan, Mexico, Carretera Lago de Guadalupe km 3.5, 52926 Atizapan, Estado de Mexico, Mexico

ARTICLE INFO

Keywords:
Gait analysis
Semi-automatic
2D pose estimation
Validity
Inter-rater reliability
Intra-rater reliability

ABSTRACT

Background: We study dual-task performance with marker-free video recordings of Timed Up-and-Go tests (TUG) and TUG combined with a cognitive/verbal task (TUG dual-task, TUGdt).
Research question: Can gait parameters be accurately estimated from video-recorded TUG tests by a new semi-automatic method aided by a technique for human 2D pose estimation based on deep learning?
Methods: Thirty persons aged 60–85 years participated in the study, conducted in a laboratory environment. Data were collected by two synchronous video-cameras and a marker-based optoelectronic motion capture system as gold standard, to evaluate the gait parameters step length (SL), step width (SW), step duration (SD), single-stance duration (SSD) and double-stance duration (DSD). For reliability evaluations, data processing aided by a deep neural network model, involved three raters who conducted three repetitions of identifying anatomical keypoints in recordings of one randomly selected step from each of the participants. Validity was analysed using 95 % confidence intervals (CI) and p-values for method differences and Bland-Altman plots with limits of agreement. Inter- and intra-rater reliability were calculated as intraclass correlation coefficients (ICC) and standard errors of measurement. Smallest detectable change was calculated for inter-rater reliability.
Results: Mean ddifferences between video and the motion capture system data for SW, DSD, and SSD were significant ($p < 0.001$). However, mean differences for all parameters were small (-6.4%–13.0% of motion capture system) indicating good validity. Concerning reliability, almost all 95 % CI of the ICC estimates exceeded 0.90, indicating excellent reliability. Only inter-rater reliability for SW (95 % CI = 0.892;0.973) and one rater's intra-rater reliability for SSD (95 % CI = 0.793;0.951) were lower, but still showed good to excellent reliability.
Significance: The presented method for extraction of gait parameters from video appears suitable for valid and reliable quantification of gait. This opens up for analyses that may contribute to the knowledge of cognitive-motor interference in dual-task testing.

1. Background

In the ongoing Uppsala-Dalarna Dementia and Gait project (UDDGait™) [1] we use marker-free video recordings to document Timed Up-and-Go test (TUG) [2] and TUG combined with a cognitive/verbal task (TUG dual-task TUGdt), which are studied as markers of early dementia disorder. TUG is a well-established test of a one-movement sequence: starting from a sitting position in a chair, standing up and walking 3 m, turning around, walking back to the chair and sitting down again. Promising results have been shown concerning the TUGdt outcome "words per time unit" [3–5]. Since gait disturbance may precede and indicate cognitive decline [6,7], gait parameters extracted from the UDDGait™ videos may give additional valuable information.

A recent systematic review and meta-analysis [8] indicated that gait analysis could contribute to the diagnosis and prognosis of dementia, since the results showed that several gait parameters are affected by the

* Corresponding author at: School of Health and Welfare, Dalarna University, SE-79188 Falun, Sweden.
E-mail address: aab@du.se (A.C. Åberg).

https://doi.org/10.1016/j.gaitpost.2021.08.004
Received 8 April 2021; Received in revised form 3 August 2021; Accepted 5 August 2021
Available online 14 August 2021

progression of dementia disorder. Additionally, the use of assessments of the cognitive-motor interference by dual-task testing and standardised methods are urgently needed in both research and clinical settings [9].

In the current study we extracted gait parameters from video recordings of TUG tests performed in a laboratory environment, in accordance with the UDDGait™study protocol [1], and evaluated their validity and reliability. Heel-strike and toe-off events are detected and a keypoint on each heel is located in the heel-strike images. This is done by visual inspection. The process is greatly aided by the identification of anatomical keypoints obtained from the deep neural network model OpenPose [10]. The keypoints obtained from OpenPose are not accurate enough for quantitative gait analysis [11]. Although, OpenPose has been used for quantitative movement analysis in several papers [12–15], the accuracy of the identified keypoints is not sufficient for detailed movement analysis relevant for clinical purposes [16]. However, their trajectories can be used to identify relevant parts of the video (gait), rough estimates of gait events, such as heel strike and toe off, and the location of the feet in the images. Still, marker-less techniques for video-based gait analysis are so far not well-established [17].

Validity and reliability are central characteristics that define the quality of measurement methods and the test results' potential of application in research and clinical practice. Validity can be determined by comparisons with a gold standard, which in the current study was obtained by an optoelectronic motion capture system (Qualisys AB, Gothenburg Sweden). Since some stages in our method for gait parameter extraction are manual and thus involve rater judgement, analyses of both inter- and intra-rater reliability are crucial. We therefore performed reliability analyses comprising the degree of agreement between results of data-processing of the same steps from different raters (inter-rater), as well as from repeated data-processing by the same rater (intra-rater) [17]. Our primary goal was to evaluate the measurement properties of our novel approach to inform future utility within the UDDGait™ project. The main research question was: Can gait parameters be accurately estimated from video-recorded TUG tests by a new semi-automatic method aided by OpenPose?

2. Methods

We conducted a study at the Biomechanics and Motor Control laboratory (BMC), at The Swedish School of Sport and Health Sciences (GIH), Stockholm, to evaluate measurement properties of the gait parameters step length (SL), step width (SW), step duration (SD), single-stance duration (SSD) and double-stance duration (DSD) extracted from video recordings [18]. The Swedish Ethical Review Authority approved the study.

2.1. Participants

Thirty persons without any diagnosed cognitive impairment were recruited among participants in a health study at GIH (Table 1). Exclusion criteria were need of an interpreter, inability to rise from a sitting position, inability to walk three meters back and forth, and indoor use of a walker.

2.2. Data collection

For descriptive purposes some standard tests of motor and cognitive functioning were used, i.e. 10 m standing start usual walking speed [19], hand grip strength measured by a dynamometer, and Mini-Mental State Examination (MMSE) [20]. Experimental conditions were TUG [2] as a single task test, as well as TUG combined with cognitive/verbal tasks [1,3,4,21] constituting three different TUGdt tests. Reliability and validity results were not investigated separately for TUG/TUGdt, with the assumption that condition does not affect reliability and validity of the proposed method. The participants were instructed by an experienced physiotherapist to complete all tests at their own speed, and to complete the mobility sequence even if they did not know what to say.

All TUG tests were documented by synchronous video recordings as well as by a marker-based opoelectronic motion capture system (Oqus4, Qualisys AB, Gothenburg, Sweden) as gold standard. For the video recordings, two cameras (Sony NEX-5 T) on tripods were used. One was placed 2 m in front of the line where the participant turned (front-view) and the other 4 m to the side of the line (side-view). The cameras' field of view was 61 degrees. Digital high-definition video (1080p) was recorded at 25 Hz. Six reference marks were placed at known locations on the floor for calibration of the video cameras. The calibration involved determining the mapping of pixel coordinates in the video images to global coordinates on the floor, with the x-direction defined as the horizontal direction from the chair to the 3 m-line, and the y-direction perpendicular and pointing to the left as seen from the chair. For the optoelectronic motion capture analysis reflective markers were placed at specified landmarks on the right and left foot: 1–2) lateral malleolus, 3–4) medial malleolus, 5–6) back of the shoe 7–8) front of the shoe, and 9–12) the head of the first and fifth metatarsal. Data were synchronized by a 'double clap' audio- and visual signal using a clapboard at the beginning and the end of each collection.

2.3. Data processing

Data processing was carried out by three raters: two researchers with expertise in motion capture and one physiotherapist, who met (virtually) on 22 occasions to discuss how to apply the definitions of the events and keypoints to be identified. The gait events *heel-strike* and *toe-off* [22] were identified in video recordings by visual inspection after approximate detection of the events using keypoints from OpenPose [10]. The video frame corresponding to the heel-strike event was defined as the first frame after the swing-phase for which there was a visible contact between the foot and the floor, and for which there was a visible decrease in the angle between the foot and the floor compared to the previous frame. The video frame corresponding to the toe-off event was the last frame of the stance phase before the toe lost contact with the floor, or the last frame before a visible reduction in the distance between the feet. To identify the heel-strike and toe-off events, the rater was first presented with a set of eight frames that were centered around an approximate event. The rater was then tasked with refining the approximated events by manually selecting the nearest frame that best matched the event definitions. Approximated heel-strike events were the frames where the

Table 1
Characteristics of study participants.

	Males	Females	Total
N	14	16	30
Age, years, median (min-max)	77 (69–83)	69.5 (60–85)	73 (60–85)
University educated, n	8	12	20
Body length, cm, median (min-max)	180.3 (171–185.5)	165.8 (151–174.5)	171.3 (151–185.5)
Walking speed, m/s, median (min-max)	1.3 (1–1.7)	1.2 (0.8–1.9)	1.3 (0.8–1.9)
Hand grip, lb, median (min-max)	84 (54–112)	58 (37–75)	64.5 (37–112)
MMSE score, median (min-max)*	29 (24–30)	29 (25–30)	29 (24–30)

* Mini-Mental State Examination.

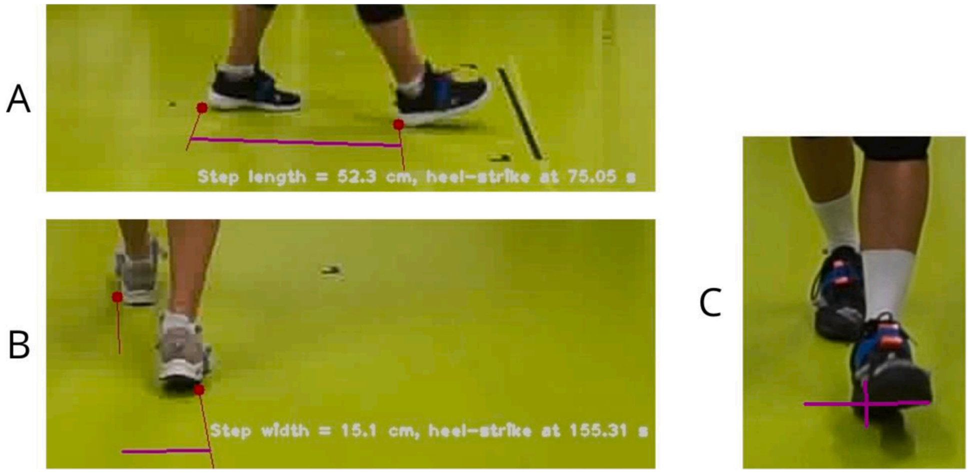

Fig. 1. Examples of identification of step length (A) and step width (B) from video recordings. Marking the heel points in the frontal-view video is guided by lines that show where the heel point was previously marked in the corresponding side-view video (C).

horizontal distance between the left and right foot keypoints in the x-direction attained a local maximum. Toe-off events were approximated as the frame at 28 % of the time interval between two subsequent heel-strike events. The value of 28 % was determined from event data obtained from the marker-data for the same sample population. Note that the approximate events use definitions different from the manually defined events, since they are estimated from the OpenPose keypoints. These keypoints are inaccurate and noisy, and require the use of simpler definitions that give more robust, albeit less accurate results. The heel-strike video frames for both the side-view camera and the frontal-view camera were automatically cropped to show the feet and shank only and then enlarged by a factor of three to improve the accuracy of manual marking. In the side-view video frame, the most posterior-inferior point of the heel was manually marked. Similarly, the most lateral-inferior point of the heel at heel-strike was marked in the frontal-view video frame. The coordinates of the heel point obtained from the side-view video was used to create visual guidelines for the approximate heel location in the frontal-view video (Fig. 1). The primary purpose of the guidelines was to aid the raters in locating the heel in video frames where the participant was walking towards the camera, where the heel was occluded by the forefoot. Image coordinates were transformed into 3D coordinates using the mapping calculated from known markings on the floor in the camera calculation step. Based on the identified events and heel point positions, the following gait parameters were determined: SL defined as the distance between posterior points on heels in the direction from chair to 3 m-line marking (horizontal x-direction) (Fig. 1) and SW defined as the distance between lateral heel points in the horizontal y-direction, calculated from the positions of markers on the heel and malleolus marker, and SD, SSD and DSD calculated from the times (s) of the identified gait events. For marker data the Foot velocity algorithm [18] was used to determine the gait events.

For evaluation of inter- and intra-rater reliability of the manual parts of the data processing, a program was developed to randomly choose a single step (among the four different TUG tests) from each of the 30 participants, resulting in 17 steps towards the 3 m-line and 13 steps back towards the chair. These steps were then processed (as described above), where each step required the rater to mark seven events: two sequential heel-strikes, the toe-off occurring in between them, and four heel events, i.e. two posterior in the side-view video frame and two lateral in the frontal-view video frame. This process was repeated three times by three different raters for all randomly selected steps. Processing of the randomised steps was carried out in one sequence, with sufficient pause in-between so that at least 30 min separated the repeated analysis of the same step. A quality control following the data extraction showed that three steps from one rater had to be excluded due to errors occurring during the data processing. Two steps were excluded for not being the same as processed by the other two raters. One step was excluded because of clearly non-valid durations. The excluded steps were not processed again as the procedure would have differed from that of the other raters. Learning might also affect the outcome if the rater were to process the same steps again.

2.4. Statistical analyses

Statistical analyses based on video results from 30 unique steps (processed by two raters and 27 processed by one rater) were carried out using Statistica® 13.4 (TIBCO Software Inc. CA, USA) and SAS® version 9.4 (SAS Institute Inc., Cary, NC, USA). The normality assumption of the gait parameters was evaluated using the Shapiro-Wilks test where values of the test statistic $w>0.95$ indicated normality. In case of $w<0.95$ a logarithmic transformation was performed.

Criterion validity data was calculated as the mean of all available gait parameters for each step compared to the value from the motion capture system. These data were analysed with Bland-Altman plots with limits of agreement [23], and 95 % confidence interval (CI) for differences with p values. A p-value less than 0.05 was deemed as statistically significant.

Inter- and intra-rater agreements were calculated as intraclass correlation coefficient (ICC) and standard error of measurement (SEM). For inter-rater reliability, smallest detectable change (SDC) was calculated as $1.96 * \sqrt{2} * \text{SEM}$. SDC can be regarded as the smallest change between any two steps that cannot be attributed to measurement errors.

Raters and analysed steps were regarded as random samples from the corresponding populations. Our aim was to measure absolute agreement. The intended measurement protocol will be based on a single rater. Thus, for inter-rater reliability, ICC estimates with 95 % CI were calculated using a single-rating, absolute-agreement, 2-way random-effects model, and for intra-rater reliability, ICC estimates with 95 % CI were calculated using a single rating, absolute agreement, one-way random-effect model for each rater [24]. Furthermore, inter- and intra-rater reliability were illustrated in Bland-Altman plots. In these plots, intra-rater reliability for each rater and step was presented as intra-rater standard deviations versus mean of all measurements and inter-rater reliability for each step was presented as standard deviations between the raters' mean values versus mean of all measurements.

3. Results

Sixteen women and 14 men, aged 60–85 years (median = 73 years) participated in the study (Table 1). The participants had an MMSE score

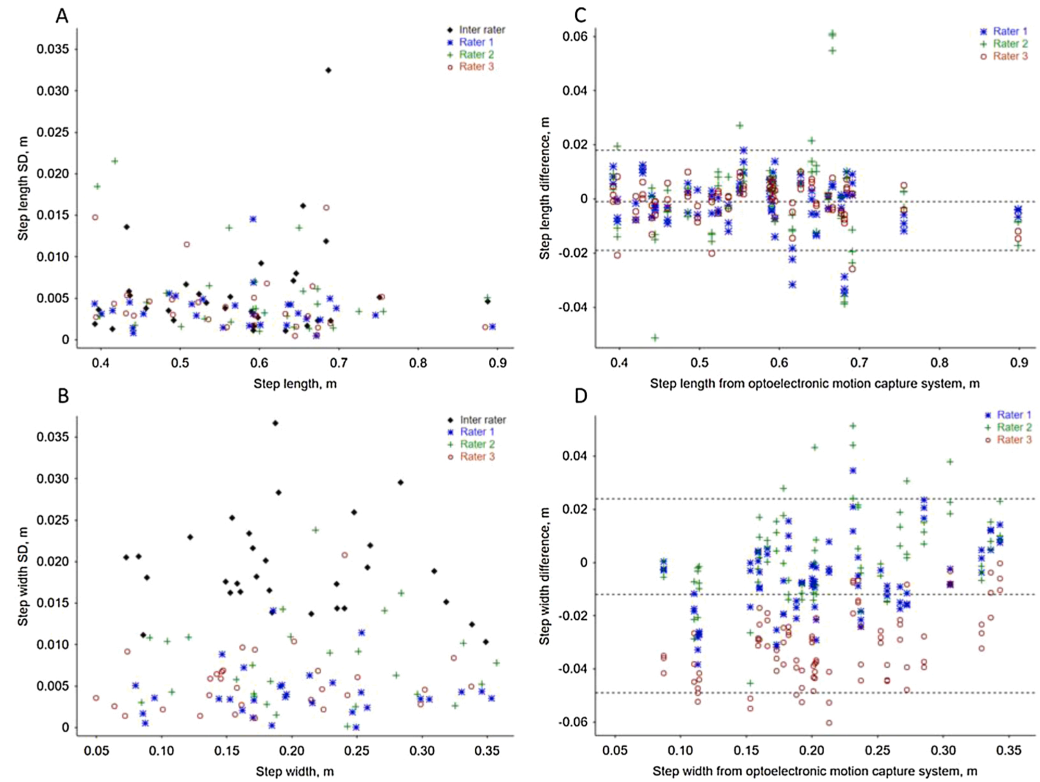

Fig. 2. Inter- and intra-rater standard deviations versus mean values of step length (A) and step width (B) extracted from video by the three raters. Comparison with results from an optoelectronic motion capture system as gold standard (C and D). Results were based on data from 30 steps for rater 1 and 3, and 27 steps for rater 2.

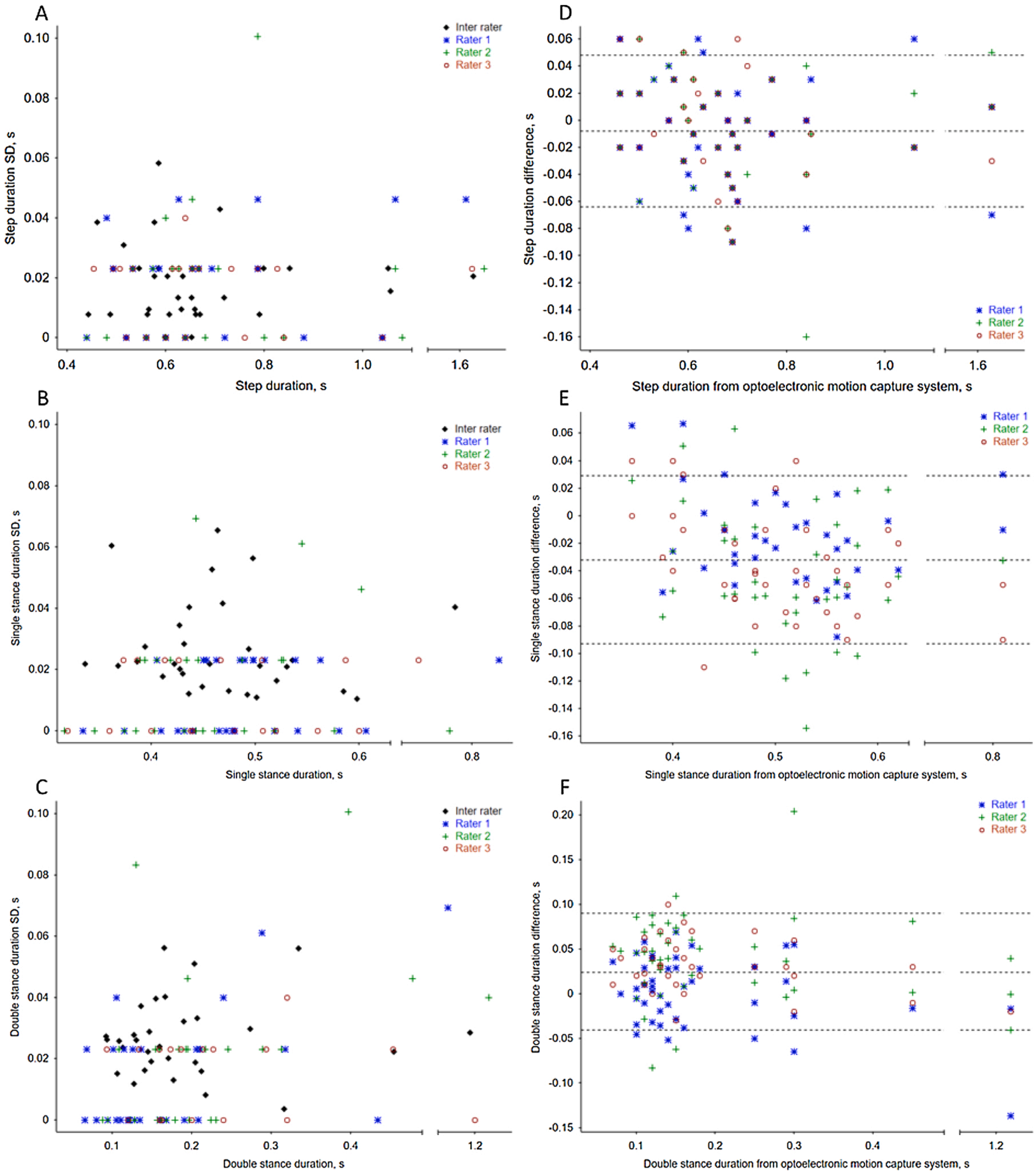

Fig. 3. Inter and intra-rater standard deviations versus mean values of durations of step (A), single stance (B) and double stance (C) extracded from video by three raters. Comparison with results from an optoelectronic motion capture system as gold standard (D, E and F). Results were based on data from 30 steps for rater 1, and 3 and 27 steps for rater 2.

of at least 24, which is a common cut-off for cognitive impairment [25], a walking speed of at least 0.8 m/s, and a hand-grip strength of above 53 lb for men and 36 lb for women, respectively.

All gait parameters were normally distributed without transformation according to the Shapiro-Wilks test statistic (w>0.95).

Criterion validity, analysed by comparisons of gait parameter results extracted from video recordings and results generated by the motion capture system, are illustrated by plots of SL and SW (Fig. 2C and D) as well as SD, SSD, and DSD (Fig. 3D–F). These plots show the difference between gait parameter values from each of the three raters and corresponding values from the motions capture system versus the motions capture value itself, as well as 95 % limits of agreement. The mean differences of these comparisons (and the mean differences in percentage of the gold standard means values) including the 95 % CI for mean differences are shown in Table 2. Mean differences for SL (-0.1 %) and SD (-1.1 %) were not statistically significant while SW (-6.0 %), SSD (-6.4 %) and DSd (13.0 %) showed significant differences.

Inter- and intra-rater standard deviation for gait parameters extracted from video by the three raters are illustrated by plots of SL and SW (Fig. 2A and B) as well as SD, SSD, and DSD (Fig. 3A, B, C). Inter- and intra-rater agreements concerning these parameters are presented as ICC and SEM in Table 3, with the addition of SDC for inter-rater reliability. According to 95 % CIs all ICC estimates were good or excellent with estimated inter-rater ICCs ranging between 0.889 and 0.996 and intra-rater ICCs ranging between 0.898 and 0.999.

Table 2
Means of raters' measurements and means of motion capture system measurements (gold standard) and mean differences (and mean differences in percentage of the gold standard mean values) between them with 95 % confidence intervals and p-values. Results were based on data from 30 steps for motion capture, rater 1 and 3 and 27 steps for rater 2.

Parameter	Mean of raters' measurements	Mean of gold standard	Mean difference (% of gold standard)	95 % confidence interval	p-value
Step length (m)	0.574	0.575	−0.001 (−0.1 %)	−0.003; 0.002	0.699
Step width(m)	0.199	0.211	−0.013 (−6,0%)	−0.016; −0.009	<0.001
Step duration (s)	0.685	0.693	−0.008 (−1.1 %)	−0.016; 0.000	0.096
Single stance duration (s)	0.470	0.502	−0.032 (−6.4 %)	−0.040; −0.025	<0.001
Double stance duration (s)	0.215	0.190	0.025 (13.0 %)	0.017; 0.032	<0.001

Table 3
Intraclass correlation coefficients (ICC) with 95 % confidence intervals, standard error of measurements (SEM) for inter-rater and intra-rater variability, and smallest detectable change (SDC) for inter-rater variability. Results were based on data from 30 steps from rater 1 and 3, and 27 steps for rater 2.

Parameter	Variability	ICC	95 % confidence interval ICC	SEM	SDC
Step length (m)	Inter-rater	0.998	0.996;0.999	0.001	0.003
	Intra-rater, rater 1	0.998	0.995;0.999	0.004	
	Intra-rater, rater 2	0.993	0.985;0.997	0.008	
	Intra-rater, rater 3	0.996	0.992;0.998	0.006	
Step width (m)	Inter-rater	0.946	0.892;0.973	0.018	0.050
	Intra-rater, rater 1	0.992	0.984;0.996	0.005	
	Intra-rater, rater 2	0.978	0.953;0.989	0.009	
	Intra-rater, rater 3	0.987	0.974;0.994	0.007	
Step duration (s)	Inter-rater	0.995	0.989;0.997	0.006	0.016
	Intra-rater, rater 1	0.982	0.964;0.991	0.024	
	Intra-rater, rater 2	0.980	0.959;0.991	0.027	
	Intra-rater, rater 3	0.988	0.976;0.994	0.019	
Single stance duration (s)	Inter-rater	0.944	0.889;0.973	0.013	0.037
	Intra-rater, rater 1	0.950	0.899;0.975	0.016	
	Intra-rater, rater 2	0.898	0.793;0.951	0.024	
	Intra-rater, rater 3	0.953	0.905;0.977	0.015	
Double stance duration (s)	Inter-rater	0.985	0.97;0.993	0.020	0.055
	Intra-rater, rater 1	0.978	0.955;0.989	0.023	
	Intra-rater, rater 2	0.964	0.924;0.983	0.032	
	Intra-rater, rater 3	0.988	0.976;0.994	0.017	

4. Discussion

The findings show that the presented method for extraction of gait parameters possesses good to excellent validity and inter-, and intra-rater reliability. There was a span in the agreement between the proposed method and the gold-standard as expressed in the % absolute mean difference, ranging from less than 1.1 % (SL and SD) to around 6% (SW and SSD) to 13.6 % (DSD). The validity must therefore be judged individually for the different step parameters. The limits of agreement for SL are better by a factor three, but for SW worse by a factor 2, compared to a recent validation of a marker-less video-based system [26]. For SD and DSD the results are similar. A recent study uses OpenPose keypoints directly to estimate gait parameters [27]. This method does not use a human rater to refine gait events and to manually mark the contact points between the heel and surface, but instead relies on manually cleaning the trajectories of OpenPose keypoints. In comparisons with a gold standard similar to our study, they report a % absolute mean difference greater than 7.8 % for SL, 3.3 % for SD and 14.2 % for DSD [27]. This is substantially poorer accuracy compared to our results (except for DSD which is on par) and it shows the potential gain of including human raters in the process.

Concerning reliability, almost all 95 % CI of the ICC estimates exceeded 0.90, indicating excellent reliability [24]. Only inter-rater reliability for SW (95 % CI = 0.892;0.973) and one rater's intra-rater reliability for SSD (95 % CI = 0.793;0.951) were lower, but still on a level of good to excellent reliability. The SEM, which allows for differentiating between real differences and random errors [28] showed overall small values, with ranges between 0.001−0.018 m for spatial, and 0.006−0.18 s for temporal measurements. The SDC, presented for inter-rater variability and regarded as the smallest change between any two steps/ratings that cannot be attributed to measurement errors [28] were also small. Hence, the small SEM and SDC values in this study indicate that small real differences between steps could be detected by our method [28].

Some limitations should be considered when interpreting the results. We used no markers, which may have improved measurement quality even more [17], but were judged as clinically inconvenient particularly for the target population of UDDGait™ i.e. people undergoing memory assessment. The accuracy may be different for steps at different locations in space due to differences in distance to the cameras. The reported results represent average accuracy. Furthermore, we did not perform a power calculation to determine the sample size. However, it exceeded what is commonly required for reliability studies [24], since three raters processed thirty samples/steps (from 30 individuals), three times each. This study design resulted in a high precision of the estimated ICCs, which was shown in the narrow CIs in the reliability analyses. Additionally, the validity analyses resulted in precise estimates of differences between video recordings and the motion capture system. Nevertheless, more assessor pre-training might have improved the inter-rater reliability, since manually determining events and keypoints in videos implies decision making by individuals, and is therefore susceptible to subjective bias. For a team working together with data processing for a laboratory, it is advised to train together on the identification of gait events and keypoints to improve the inter-rater reliability. The results showed that reliability is better within the same rater, which makes it advantageous, but not critical, that the same individual processes the data for any repeated visits/trials of the same participant.

The current results are promising for future applications of our method that include the deep learning model OpenPose, to speed up and guide analysis of video data. Such methods can successfully classify different types of gait [29] and novel attempts to automatically extract quantitative information on gait from video are encouraging, but require large data sets [13]. Our method involves manual tasks (labelling events and keypoints) which we register in order to accumulate a database of training data that will enable training of deep learning networks for fully automatic analysis in the future.

5. Conclusions and significance

The presented method for extraction of gait parameters from video appear suitable for valid and reliable quantification of gait, which is more available and affordable than laboratory-based motion capture systems and easier to accomplish in a clinical context compared to marker-based video recordings [17,30].

Extraction and quantification of gait parameters from video facilitate analyses that may contribute to the knowledge of cognitive-motor interference in dual-task testing. We are further developing our methodology to make it more automatic, faster and hence, useful for analysis of our extensive UDDGait™ video-data. For these purposes, we use deep neural networks trained specifically for detecting the keypoints on the heel. The performance of such networks trained on image data labelled by humans depends on the quality of the training data, i.e. the validity and reliability. The results of this study establish the quality of such training data.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This work was supported by the Promobila foundation (grant numbers 18098 and 20098) and the Dementia Foundation (grant number 2019-04-09). The authors thank MSc Isabell Ahlström for carrying out some of the data processing.

References

[1] Y. Cedervall, A.M. Stenberg, H.B. Åhman, V. Giedraitis, F. Tinmark, B. Berglund, et al., Timed Up-and-Go Dual-Task testing in the assessment of cognitive function: a mixed methods observational study for development of the UDDGait Protocol, J. Environ. Res. Public Health 17 (5) (2020). https://www.mdpi.com/1660-4601/17/5/1715.

[2] D. Podsiadlo, S. Richardson, The timed "Up & Go": a test of basic functional mobility for frail elderly persons, J. Am. Geriatr. Soc. 39 (2) (1991) 142–148, https://doi.org/10.1111/j.1532-5415.1991.tb01616.x.

[3] H.B. Åhman, L. Berglund, Y. Cedervall, L. Kilander, V. Giedraitis, et al., Dual-task tests predict conversion to dementia-a prospective memory-clinic-based cohort study, Int. J. Environ. Res. Public Health 17 (21) (2020). https://www.ncbi.nlm.nih.gov/pubmed/33153203.

[4] H.B. Åhman, Y. Cedervall, L. Kilander, V. Giedraitis, L. Berglund, K.J. McKee, et al., Dual-task tests discriminate between dementia, mild cognitive impairment, subjective cognitive impairment, and healthy controls - a cross-sectional cohort study, BMC Geriatr. 20 (1) (2020) 258. https://www.ncbi.nlm.nih.gov/pubmed/32727472.

[5] H.B. Åhman, V. Giedraitis, Y. Cedervall, B. Lennhed, L. Berglund, K. McKee, et al., Dual-task performance and neurodegeneration: correlations between Timed Up-and-Go Dual-Task test outcomes and Alzheimer's disease cerebrospinal fluid biomarkers, J. Alzheimers Dis. 71 (s1) (2019) S75–S83. https://www.ncbi.nlm.nih.gov/pubmed/31104024.

[6] C. Buckley, L. Alcock, R. McArdle, R.Z.U. Rehman, S. Del Din, C. Mazza, et al., The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: insights from gait and postural control, Brain Sci. 9 (2) (2019). https://www.ncbi.nlm.nih.gov/pubmed/30736374.

[7] Y. Cedervall, K. Halvorsen, A.C. Aberg, A longitudinal study of gait function and characteristics of gait disturbance in individuals with Alzheimer's disease, Gait Posture 39 (4) (2014) 1022–1027. https://www.ncbi.nlm.nih.gov/pubmed/24491520.

[8] R. Chiaramonte, M. Cioni, Critical spatiotemporal gait parameters for individuals with dementia: a systematic review and meta-analysis, Hong Kong Physiother. J. 41 (1) (2021) 1–14. https://www.ncbi.nlm.nih.gov/pubmed/34054252.

[9] S. Cullen, M. Montero-Odasso, L. Bherer, Q. Almeida, S. Fraser, S. Muir-Hunter, et al., Guidelines for gait assessments in the canadian consortium on neurodegeneration in aging (CCNA), Can. Geriatr. J. 21 (2) (2018) 157–165. https://www.ncbi.nlm.nih.gov/pubmed/29977431.

[10] Z. Cao, G.H. Martinez, T. Simon, S.E. Wei, Y.A. Sheikh, OpenPose: realtime multi-person 2D pose estimation using part affinity fields, IEEE Trans. Pattern Anal. Mach. Intell. (2019). https://ieeexplore.ieee.org/document/8765346.

[11] E. D'Antonio, J. Taborri, E. Palermo, S. Rossi, F. Patanè, A markerless system for gait analysis based on OpenPose library. 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2020, pp. 1–6. https://ieeexplore.ieee.org/document/9128918.

[12] W. Chen, Z. Jiang, H. Guo, X. Ni, Fall detection based on key points of human-skeleton using OpenPose, Symetri 12 (5) (2020) 744. https://www.mdpi.com/2073-8994/12/5/744/htm.

[13] L. Kidzinski, B. Yang, J.L. Hicks, A. Rajagopal, S.L. Delp, M.H. Schwartz, Deep neural networks enable quantitative movement analysis using single-camera videos, Nat. Commun. 11 (1) (2020) 4054. https://www.ncbi.nlm.nih.gov/pubmed/32792511.

[14] L. Needham, M. Evans, D.P. Cosker, S.L. Colyer, Can markerless pose estimation algorithms estimate 3D mass centre positions and velocities during linear sprinting activities? Sensors Basel (Basel) 21 (8) (2021). https://www.ncbi.nlm.nih.gov/pubmed/33924266.

[15] K. Sato, Y. Nagashima, T. Mano, A. Iwata, T. Toda, Quantifying normal and parkinsonian gait features from home movies: practical application of a deep learning-based 2D pose estimator, PLoS One 14 (11) (2019) e0223549. https://www.ncbi.nlm.nih.gov/pubmed/31725754.

[16] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, et al., Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras, Front. Sports Act Living 2 (2020) 50. https://www.ncbi.nlm.nih.gov/pubmed/33345042.

[17] A. Michelini, A. Eshraghi, J. Andrysek, Two-dimensional video gait analysis: a systematic review of reliability, validity, and best practice considerations, Prosthet. Orthot. Int. 44 (4) (2020) 245–262. https://www.ncbi.nlm.nih.gov/pubmed/32507049.

[18] L.B. Mokkink, C.B. Terwee, D.L. Patrick, J. Alonso, P.W. Stratford, D.L. Knol, et al., The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study, Qual. Life Res. 19 (4) (2010) 539–549. https://www.ncbi.nlm.nih.gov/pubmed/20169472.

[19] D.M. Peters, S.L. Fritz, D.E. Krotish, Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults, J. Geriatr. Phys. Ther. 36 (1) (2013) 24–30. https://www.ncbi.nlm.nih.gov/pubmed/22415358.

[20] M.F. Folstein, S.E. Folstein, P.R. McHugh, Mini-mental state". A practical method for grading the cognitive state of patients for the clinician, J. Psychiatr. Res. 12 (3) (1975) 189–198. https://www.sciencedirect.com/science/article/abs/pii/0022395675900266?via%3Dihub.

[21] R.T. Manning, The serial sevens test, Arch. Intern. Med. 142 (6) (1982) 1192. https://www.ncbi.nlm.nih.gov/pubmed/7092427.

[22] I.T. Jolliffe, Principal Component Analysis, second edition ed., Springer, New York, 2002.

[23] L.B. Mokkink, C.A. Prinsen, D.L. Patrick, J. Alonso, L.M.Lex M. Bouter, H.C. de Vet, et al., COSMIN Study Design Checklist for Patient-reported Outcome Measurement Instruments, 2019 (accessed 2021).

[24] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J. Chiropr. Med. 15 (2) (2016) 155–163. https://www.ncbi.nlm.nih.gov/pubmed/27330520.

[25] S.T. Creavin, S. Wisniewski, A.H. Noel-Storr, C.M. Trevelyan, T. Hampton, D. Rayment, et al., Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations, Cochrane Database Syst. Rev. 1 (2016) CD011145. https://www.ncbi.nlm.nih.gov/pubmed/26760674.

[26] R.M. Kanko, E.K. Laende, G. Strutzenberger, M. Brown, W.S. Selbie, V. DePaul, et al., Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system, J. Biomech. 122 (2021) 110414. https://www.ncbi.nlm.nih.gov/pubmed/33915475.

[27] J. Stenum, C. Rossi, R.T. Roemmich, Two-dimensional video-based analysis of human gait using pose estimation, PLoS Comput. Biol. 17 (4) (2021) e1008935. https://www.ncbi.nlm.nih.gov/pubmed/33891585.

[28] L. Furlan, A. Sterr, The applicability of standard error of measurement and minimal detectable change to motor learning Research-A behavioral study, Front. Hum. Neurosci. 12 (2018) 95. https://www.ncbi.nlm.nih.gov/pubmed/29623034.

[29] A.S. Alharthi, S.U. Yunas, K.B. Ozanyan, Deep learning for monitoring of human gait: a review, IEEE Sens. J. 19 (21) (2019) 9575–9591. https://ieeexplore.ieee.org/document/8762100.

[30] F. de Oliveira Silva, J.V. Ferreira, J. Placido, D. Chagas, J. Praxedes, C. Guimaraes, et al., Gait analysis with videogrammetry can differentiate healthy elderly, mild cognitive impairment, and Alzheimer's disease: a cross-sectional study, Exp. Gerontol. 131 (2020) 110816. https://www.ncbi.nlm.nih.gov/pubmed/31862421.