Exploring the Potential of a Robot-Assisted Frailty Assessment System for Elderly Care

Aniol Civit[1], Antonio Andriella[2], Maite Antonio[3], Casimiro Javierre[4], Concepción Boqué[3] and Guillem Alenyà[1]

***Abstract*— Frailty assessment plays a pivotal role in providing older adults care. However, the current process is time-consuming and only measures patients' completion time for each test. This paper introduces a set of algorithms to be used in robots to autonomously perform frailty assessments. In doing so we aim at reducing therapists' burden and provide additional frailty-related metrics that can enhance the effectiveness of diagnosis. We conducted a pilot study with 22 elderly participants and compared our system's performance with that of medical professionals to assess its precision. The results demonstrate that our approach achieved performances close to that of its human counterpart. This research represents an important step forward in the integration of social robotics in healthcare, offering potential benefits for patient care and clinical decision-making.**

I. INTRODUCTION

The continuous increase in lifespan over recent years has resulted in a growing proportion of older adults within the population [1]. This demographic shift presents unique challenges, particularly in healthcare, where older adults may face greater vulnerability and risk. One crucial aspect of older adults' health is frailty, a well-established indicator of their susceptibility to adverse health outcomes, particularly when coupled with other health issues [2]. Older adults with higher levels of frailty exhibit reduced tolerance to the stress induced by different treatments, making it imperative to tailor medical interventions to their specific needs [3].

The gold standard to assess frailty in older adults is geriatric screening followed by Comprehensive Geriatric Assessment (CGA) across essential GA-domains that include social status, nutrition, cognition, emotion, co-morbidity, polypharmacy, geriatric syndromes and physical condition [4]. Here, we focus on the physical component. A commonly used indicator of physical frailty is the performance in some standardized tests such as the Short Physical Performance Battery (SPPB) [5] and the Timed Up and Go (TUG) [6]. However, administering these tests demands valuable time from already busy healthcare professionals [7]. Moreover, the data collected through these tests is often limited in scope, e.g., completion time. Many additional metrics can be extracted from those tests that have been proven to be related to frailty [8]–[13], e.g., stride and step lengths, cadences, and speeds during walking tests, stability during the balance test, and others.

Fig. 1. Showcase of the frailty assessment. The robot monitors the patient (one of the authors) while he starts performing the TUG test.

Integrating robotics into healthcare is rapidly gaining recognition as a promising solution to bridge existing societal gaps in healthcare delivery. Robots can become powerful tools for therapists and doctors, enhancing their effectiveness and productivity [14], [15], especially in repetitive tasks such as frailty assessment. By automating these repetitive tasks, doctors can dedicate their time to activities that better utilize their expertise and gather valuable data that can enhance their ability to make accurate diagnoses.

In our previous work [16], we proposed a robotics framework to autonomously administer frailty assessments. The framework consists of two stages: in the first, the robot has to bring the patient to the evaluation room, and in the second, the robot has to make the patient perform the tests and

*This work was supported by the project ROB-IN PLEC2021-007859 funded by MCIN/ AEI /10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"; and the 23S06141-001 FRAILWATCH project funded by Barcelona Ciencia 2020-2023 Plan. A. Civit has been supported by AGAUR-FI ajuts (2023 FI-3 00065) Joan Oró of the Secretariat of Universities and Research of the Department of Research and Universities of the Generalitat of Catalonia and the European Social Plus Fund.

[1]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain; {acivit, galenya}@iri.upc.edu

[2]Artificial Intelligence Research Institute (IIIA-CSIC), Campus de la UAB, 08193 Bellaterra, Barcelona, Spain; aandriella@iri.upc.edu

[3]Oncohematogeriatrics Unit, Institut Català d'Oncologia, IDIBELL Hospitalet, 08908 Barcelona, Spain {marebollo, c.boque}@iconcologia.net

[4]Unitat de Fisiologia. Departament de Ciències Fisiològiques II Facultat de Medicina (Campus de Bellvitge) Ctra. Feixa Llarga, s/n. 08907 L'Hospitalet de Llobregat, Barcelona, Spain cjavierre@ub.edu

evaluate them. In this work, we focus on the second stage of the framework by developing and validating the robot's perceptual abilities, as well as the algorithms to extract the necessary metrics. To validate our system, we conducted a pilot study with 22 older adults who attended the hospital to assess their frailty. We compared the robot's ability to automatically measure the completion time of the tests with that of the doctor. The results of our study indicated that there was a strong correlation between our system's measurements and those taken by the doctor.

In summary, the contributions of this work are the following:

1) Designing a set of algorithms for automatically assessing frailty and extracting additional metrics according to the healthcare professionals' requirements.
2) Validating the tracking system and the metrics with an OptiTrack system as the gold standard.
3) Evaluating our system in a pilot study with N=22 older adults.

II. RELATED WORK

This section presents studies where the frailty assessment tests are performed and measured, and finally, it includes the use of social robots for that purpose. We will not be covering studies that have used wearable sensors such as accelerometers, gyroscopes, pressure sensors, and grip-strength sensors, as they can be intrusive and impractical in many assistive contexts [8], [10], [11], [12], [17].

Camera-based systems have been widely used to monitor and assess user health status by extracting information through computer vision or deep learning algorithms. In [18], the OpenPose algorithm was applied to videos, followed by a CNN to obtain gait metrics, including walking speed, cadence, and knee flexion at maximum extension, they trained the algorithm with a dataset of 1026 patients with cerebral palsy, which means that the study is biased towards patients with that pathology. In [19] a multi-RGB camera system with Openpose and GrabCut was used to extract the skeleton, and the gait features were calculated based on the angle of the foot with the vertical image plane axis. A limitation of this approach is its dependence on the proper horizontal positioning of the camera. In [20] a multi-camera system was developed for a 10-meter walk gait analysis. The gait information was based on the distances between ankle joints. However, the case of irregular walking patterns was not considered. In [21], the authors employed the Kinect camera and a supervised learning approach to acquire gait information. Nonetheless, the system was trained with 23 subjects between the age of 26 and 56. Therefore, its application may be limited to older adults who might have different walking patterns than younger individuals.

Relating to frailty assessment, in [22], the Kinect camera was used to track the skeleton for assessing the Standing Balance test, they validated the measure of the movement of the center of mass of the patient, but the imbalance was not computed. In [23], they developed a frailty detection tool based on the TUG test using the RGBD Kinect camera. They determined a step when an ankle reached zero velocity. The time and distances between steps are calculated from the skeleton poses acquired using the camera. A limitation is the depth range of the Kinect camera, which complicated skeleton extraction at the chair position. As a result, the algorithm extracted information only during the walking phases, the time was computed manually. In [24], the SPPB and TUG tests are conducted. They employ Haar Cascades to locate the face of the patient and track it until it reaches the goal in all the tests. In the Standing Balance test, they also marked the feet' location on the floor. For the 5-Times Sit-Stand test, they used a virtual horizontal plane to determine if the face crossed it five times. A limitation of this study is the low flexibility in the experimental setup and test options, as a prior configuration is required, it doesn't ensure significant autonomy for healthcare environments with older adults using it.

There have been several studies in the literature that have investigated the use of social robotics to assist frail older adults. In [25], authors demonstrated the potential of social robots to provide information on how to perform frailty tests. The study was focused on the user experience of a robot application in which physical exercises were provided to frail older adults while monitored by a therapist. In [26], a robotic platform was employed to perform cognitive tests on the patients, using the Kinect camera for the TUG test, by means of which they computed the time, step length, number of steps, step frequency, velocity gait, arms swing amplitude, and upper-lower body synchrony, but the methodology was not explained or validated.

To address some of the limitations encountered in previous works, we propose a robotic system that makes use of a stereoscopic camera. This camera increases the distance working range thereby providing more flexibility in terms of setting up the environment. Additionally, we have equipped the robot with skeleton-based algorithms that are independent of the camera's orientation. This allows us to detect the beginning and end of each test in real time as well as monitor the user's performance during the tests.

III. METHOD

The system proposed in this work aims to enable robots to autonomously administer the SPPB and TUG tests. To achieve this goal, we develop a set of algorithms that identify the start and end of each test and also monitor the user while taking the tests. All the algorithms are based on the user's skeleton which is obtained and tracked from images captured by a camera. We show how the system is capable of (i) computing the user's completion time and (ii) extracting other frailty-related metrics deemed important by healthcare professionals to improve diagnosis.

In the following paragraphs, we describe each of the tests, how we detect when it starts and ends, how we compute the completion time, and the additional metrics.

A. SPPB - Standing Balance Test

In this test, the patient is requested to stand in three predetermined foot positions for 10 seconds while maintaining balance, the positions are: together, semi-tandem, and tandem. It is worth noting that is the robot that sets the start and the end of each position, which is fixed to 10 seconds. The following metrics apply to all the foot positions.

a) Completion time: To detect a loss of balance, we measure the distance between the heels and the tips of the toes on both feet. An imbalance is detected when the standard deviation σ of the difference in that distances between three consecutive frames is greater than a certain value λ (in the experiment, we use a value of $\lambda = 0.08$), where

$$\text{time} = \begin{cases} t & \text{if } \sigma(d_{heels}(t-1,t,t+1)) > \lambda \\ & \text{or } \sigma(d_{toes}(t-1,t,t+1)) > \lambda, \\ 10 & \text{otherwise.} \end{cases}$$

The time t corresponds to the total time that the patient has kept balance, or 10 if there were no issues.

b) Additional metric - Movement box: A metric that measures the movement of the shoulders and explains how balanced a patient is. It is the box containing the maximum and the minimum positions in each of the three coordinates for one of the patient's shoulders, it is calculated as follows:

$$\begin{aligned} M_{\text{box}}\ (m^3) = &|\max(\vec{p}_x) - \min(\vec{p}_x)| \\ &\cdot |\max(\vec{p}_y) - \min(\vec{p}_y)| \\ &\cdot |\max(\vec{p}_z) - \min(\vec{p}_z)| \end{aligned}$$

where p is the right or left shoulder 3D coordinates.

B. SPPB - Gait Speed Test

In this test, the patient is asked to walk six meters in a straight line. The first meter corresponds to acceleration and the last one to deceleration. Timing is taken during the constant speed phase, which corresponds to the central four meters.

a) Completion time: We detect the frames that correspond to the start and end of the 4-meter run. Then, we calculate the distance vector using the following method. Firstly, we create a vector that contains the central point between the landmarks of the head, nose, neck, left and right clavicle, chest, naval spine, and pelvis for all the frames, using the following equation:

$$\begin{aligned} \vec{p}_{\text{central}}(t) = \frac{1}{8} \cdot \Big(&\vec{p}_{\text{head}}(t) + \vec{p}_{\text{neck}}(t) + \vec{p}_{\text{nose}}(t) \\ &+ \vec{p}_{\text{right clavicle}}(t) + \vec{p}_{\text{left clavicle}}(t) \\ &+ \vec{p}_{\text{chest}}(t) + \vec{p}_{\text{naval spine}}(t) + \vec{p}_{\text{pelvis}}(t) \Big). \end{aligned}$$

Next, we set the origin to the initial frame point as:

$$\vec{d}_{central}(t_i) = \vec{p}_{central}(t_i) - \vec{p}_{central}(0), \quad \forall i.$$

It is important to note that the test includes two additional meters, which serve as acceleration (first meter) and deceleration (last meter) intervals. We need to eliminate these periods by identifying when the central point crosses the first meter:

$$t_{initial} = t_i \text{ when } ||\vec{d}_{central}(t_i)|| = 1 \text{ m}$$

and the fifth meter:

$$t_{end} = t_j \text{ when } ||\vec{d}_{central}(t_j)|| = 5 \text{ m}$$

to compute the total time as

$$t_{total} = t_{end} - t_{initial}\,.$$

b) Additional metrics - Gait information: We obtain gait metrics using Joint Relative Distance (JRD) during walking, similar to [27] where they use the Joint Relative Angle (JRA).

All the local maximum points in the JRD sequence are identified as steps since, physiologically, they correspond to the maximum leg aperture. We gather the vector $\vec{X}$ with all the steps from the sequence as follows:

$$\vec{X} = \{t \,|\, \text{Local maxima of } J\vec{R}D(t), \forall t\}.$$

Alternatively, the minimum points in the sequence are positions where the feet are the closest during walking, they are gathered in the vector

$$\vec{Y} = \{t \,|\, \text{Local minima of } J\vec{R}D(t), \forall t\}.$$

Next, we need to identify the right and left steps. To do so, we observe the maximum velocity of each foot between a step and its previous minimum. The foot with the higher maximum velocity is the one that is in motion, since during walking one foot swings while the other remains almost steady. This helps to differentiate between the two feet using

$$S(i) = \begin{cases} Left & \max \vec{v}_L(t) > \max \vec{v}_R(t), \\ & with \max(\vec{Y} | \vec{Y} < X_i) < t \leq X_i, \\ Right & otherwise. \end{cases}$$

Being S a vector containing the identity of the steps, marking if it is was taken with the right or left foot.

It is important to consider that older adults may have difficulty walking, resulting in irregular patterns. When a foot is moving but doesn't move forward the other foot, it implies that a minimum in JRD is actually a step. To detect these patterns, the method selects a local minimum between two consecutive steps of the same foot, that is when $S(i) = S(i-1)$. This local minimum is chosen as a step candidate $j = Y_i \mid X_{i-1} < Y_i \leq X_i$, and is then converted to the contrary classification $S(j) = \neg S(i)$.

Now, we need to check if the other foot has moved between that minimum and the previous step. The distances traveled by each foot between the previous step time and j are computed as

$$\begin{aligned} dist_L &= ||\vec{d}_L(j) - \vec{d}_L(i-1)||, \\ dist_R &= ||\vec{d}_R(j) - \vec{d}_R(i-1)||. \end{aligned}$$

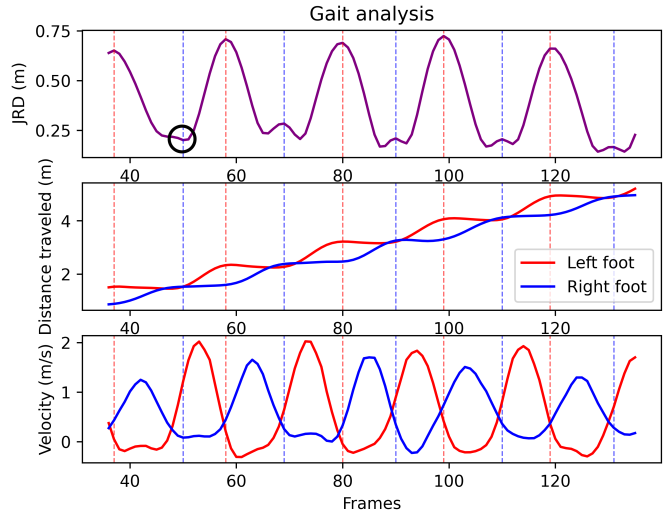

Fig. 2. Example of the gait analysis in the Gait Speed Test, the graphic shows the JRD, distance traveled, velocities of each foot during the performance, and which foot is stepping shown as vertical lines. It includes the detection of irregular walking patterns marked in a black circle in the first subplot.

If the distance traveled in that time interval of the foot that isn't detected as a step is greater than the foot detected as a step, it will be considered as a step and added to $\vec{X}$

$$\vec{X} = \begin{cases} \vec{X} + j & \text{if } dist_{S(j)} > dist_{S(i)} \\ \vec{X} & otherwise. \end{cases}$$

Fig. 2 shows an example of irregular walking. The top subplot shows the JRD (Joint Relative Distance) from the ankles, the middle one displays the distance covered by each foot, and the bottom one shows the velocities of each foot during the test. The vertical lines indicate steps detected for each foot, with blue for the right foot, and red for the left foot. The first right step (marked with a black circle in the first subplot) occurs in a local minimum from the JRD sequence, between two maxima of the left foot. By looking at the velocities (bottom subplot) and the foot displacements (middle subplot), we can see that the patient changed the moving foot between those two maxima of the left foot, indicating that the right foot was moved in between. The other steps are located in local maxima, even with the value being close to the local minima from JRD, meaning that the feet are close when stepping.

Now that we have the data, we extract various gait parameters, such as stride lengths and velocities, step lengths and velocities, step widths, and the number of steps. The stride length is the distance between two consecutive steps for the same foot, and the step length is the distance between two consecutive steps of different feet. The step width is the transversal distance. We also calculate the means for these values since there are multiple steps.

c) Additional metric - Gait balance: To measure the balance during walking, considering walking as a cyclic motion for most people, the initial positions of the pelvis and the shoulders are taken as a reference. In each frame, those initial positions are translated to the current pelvis position, the difference between the current 3D coordinates of the shoulders and the translated ones will be used to obtain the balance metric as:

$$balance(t) = \vec{p}_{shoulder}(t) - \vec{d}_{shoulder}(t),$$
$$\vec{d}_{shoulder}(t) = \vec{p}_{shoulder}(0) + \vec{d}_{pelvis}(t),$$
$$\vec{d}_{pelvis}(t) = \vec{p}_{pelvis}(t) - p_{pelvis}(0).$$

The vector $\vec{d}_{shoulder}$ is the translation of the right or left shoulder in the pelvis direction, and $\vec{d}_{pelvis}$ the translation of the pelvis, both in 3D coordinates. The acquired features for the balance are the ranges between the maximum and the minimum values of the difference, and its mean along the entire trajectory.

C. SPPB - 5 Times Sit-Stand Test

In this test, the patient is asked to stand up and sit down from a chair five times while measuring the total time taken.

a) Completion time: The total time is determined taking into account the angles between the torso and the thigh (hip angle), as well as the knee flexion angle. To identify whether the patient is sitting or standing, we rely on predetermined thresholds for these angles

$$Pose = \begin{cases} Stand\ up & \text{if } \alpha_{knee} > \lambda_{knee\_stand} \\ & \text{and } \alpha_{hip} > \lambda_{hip\_stand}, \\ Sit\ down & \text{if } \alpha_{knee} < \lambda_{knee\_sit} \\ & \text{and } \alpha_{hip} > \lambda_{hip\_sit}, \\ Transition & \text{otherwise.} \end{cases}$$

The total time is calculated by measuring the time it takes to complete five repetitions. For the experiment presented in Sec. V, when standing, we use a knee angle threshold of 130 degrees and a hip angle threshold of 120 degrees. When seated, the knee angle threshold is 114 degrees, and the hip angle threshold is 90 degrees.

b) Additional metric - Fatigue: This metric is the difference between the time of the first sit-stand and the last one, and it gives information on the patient's physical condition. We compute the time for each sit-stand action as $tss_i = tsu_i - tsd_i$, where tsu_i is the stand-up time and tsd_i is the sit-down time. The first and the last ones are selected to compute fatigue as

$$fatigue = t_{last} - t_{first}.$$

D. Timed Up and Go Test

In this test, the patient is instructed to stand up from a chair, walk three meters to an object, walk around it, walk back, and sit in the same chair.

a) Completion time: The test begins when the patient stands up and ends when they sit down, measuring the time taken between these two events. To detect these events online, the algorithm uses the same angles as the 5 Times Sit-Stand Test to identify when the patient transitions from sitting to standing (start) and from standing to sitting (end). After the online detection of the start and end of the

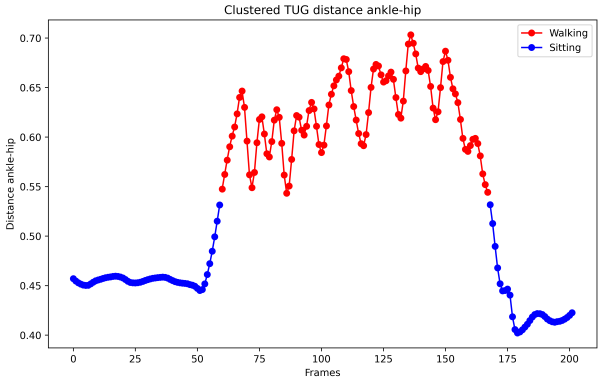

Fig. 3. Distance between the right ankle and hip and clustering for an example of the TUG test.

test, an offline method is used to increase the precision of the measurement. To detect these transitions, the distance between the patient's ankles and hips ($d = p_{hip} - p_{ankle}$) is used as a reference measure, which is higher while standing up than sitting. The K-means clustering algorithm with $k = 2$ is used on the distance d to segment the two actions. This assigns a cluster to each frame of the sequence, and the first cluster change in the sequence is considered as the patient standing up as follows:

$$t_{start} = t \text{ when } \min\{t \,|\, C_t \neq C_{t-1}\}$$

and when the cluster changes for the last time it is considered as sitting down

$$t_{end} = t \text{ when } \max\{t \,|\, C_t \neq C_{t-1}\}$$

being C_t the cluster sequence at time t. The total time is computed as

$$t_{total} = t_{end} - t_{start}\,.$$

An example of a real TUG test and the cluster segmentation is shown in Fig. 3.

b) Additional metrics - Gait information and balance: The metrics are computed as explained in Sec. III-B.

IV. METRICS PRECISION VALIDATION

In this section, we compare our tracking system with other state-of-the-art solutions in assessing the tests' metrics (see Sec. III).

We use the ZED2i stereoscopic camera[1]. There are two interesting features of this device. Firstly, it can be used both indoors and outdoors, making it robust in varying daylight conditions. This is particularly useful in real-life daycare facilities and hospitals where lighting cannot be controlled. Secondly, it has a depth range of up to 8 meters for skeleton tracking, as reported by [28] which is higher than the 5 meters range from the most used camera in the literature, the Kinect camera [29]. This is a necessary requirement as the *Gait Speed Test* needs a distance of 6 meters.

[1]https://www.stereolabs.com/assets/datasheets/zed-2i-datasheet-feb2022.pdf

A. Tracking System Validation

Regarding tracking systems, there are several options available that have been validated in the literature. In [30], it is suggested that Mediapipe performs better than the commonly used Kinect's SDK for acquiring gait information. Therefore, we decided to compare our system with Mediapipe. Additionally, to ensure accuracy, we compared the metrics computed with our system with those extracted with the OptiTrack motion capture, considered the gold standard [31].

To conduct this comparison, 7 volunteers performed each of the tests three times wearing the OptiTrack suit with 41 markers. At the same time, the skeleton 3D points were recorded using both ZED SDK[2] and Mediapipe algorithms. The volunteers were asked to vary their behavior in each run of the tests to have more variance between the samples.

B. Results and Discussion

Mediapipe failed to provide accurate information beyond a working distance of 6 meters. The detection rate was frequently incorrect, and false positives began to appear as soon as the subject moved further than 3 meters away from the camera.

Consequently, the validation was made only for the ZED SDK algorithm in comparison to the gold standard OptiTrack. To do so, we used the Mean Absolute Percentage Error (MAPE) for each of the metrics defined in Sec. III as

$$MAPE(\%) = \frac{\sum_{i=1}^{n} \frac{|X_i - \hat{X}_i|}{|X_i|}}{n} \cdot 100$$

where X_i is the measurement with OptiTrack and $\hat{X}_i$ is the measurement with ZED2i. Given the low magnitude values from the metrics, any error under 15 percent was considered a good measurement.

Table I presents a summary of results. The mean error in the time measure, which is the only measure collected by the doctors, is less than 5 percent in all tests. This suggests that the time measures using ZED SDK are precise and can be used to assess the patients.

For the Gait Speed test, the right leg metrics show lower errors than the left leg's. This is because the camera is placed on the right side of the patient during the test, leading to occlusions in some frames for the left leg.

In the TUG test, the errors decrease in the left leg step and increase in the right leg step due to occlusions in both legs for each half of the test. A camera on each side of the patient could provide better precision for these metrics if necessary.

The method for both OptiTrack and ZED SDK correctly computed the times for the Standing Balance tests. However, the box in the Standing Balance and the balance in the Gait Speed and TUG tests present a high percentage of error due to camera measurement noise.

Finally, the fatigue metric in the Sit-Stand Test has a high MAPE value of 24.91%, but the difference between the mean

[2]https://www.stereolabs.com/docs/body-tracking/

Test	Metric	OptiTrack (M)	Ours (M)	MAPE(%)
Gait Speed	**Time (s)**	4.71	4.76	**4.73**
	Left Stride Length (m)	1.25	1.16	**8.41**
	Left Stride Velocity (m/s)	0.99	0.96	**5.12**
	Left Step Length (m)	0.67	0.52	22.26
	Left Step Velocity (m/s)	1.07	0.93	15.49
	Left Step Width (m)	0.22	0.13	38.79
	Right Stride Length (m)	1.24	1.13	**9.53**
	Right Stride Velocity (m/s)	0.99	0.97	**5.34**
	Right Step Length (m)	0.66	0.67	**12.26**
	Right Step Velocity (m/s)	1.05	1.09	**11.32**
	Right Step Width (m)	0.21	0.14	34.48
	Range Balance (m)	0.05	0.08	68.74
	Mean Balance (m)	0.03	0.04	66.02
	Number Steps	7.11	7.68	**11.11**
TUG	**Time (s)**	9.29	9.36	**2.03**
	Left Stride Length (m)	1.11	1	**12.41**
	Left Stride Velocity (m/s)	0.86	0.81	**9.5**
	Left Step Length (m)	0.59	0.56	**13.09**
	Left Step Velocity (m/s)	0.92	0.91	**9.73**
	Left Step Width (m)	0.19	0.14	33.28
	Right Stride Length (m)	1.06	1.02	**9.27**
	Right Stride Velocity (m/s)	0.82	0.79	**6.52**
	Right Step Length (m)	0.59	0.52	**13.54**
	Right Step Velocity (m/s)	0.94	0.8	15.41
	Right Step Width (m)	0.24	0.17	31.61
	Range Balance (m)	0.57	0.48	17.47
	Mean Balance (m)	0.42	0.33	20.55
	Number Steps	14	13.78	**9.52**
Standing Balance	**Time Together (s)**	10	10	**0**
	Time Semi-Tandem (s)	10	10	**0**
	Time Tandem (s)	10	10	**0**
	Box Together (m^3)	0.0021	0.0019	**14.98**
	Box Semi-Tandem (m^3)	0.0001	0.0002	69.51
	Box Tandem (m^3)	0.0011	0.0008	33.87
5 Times Sit-Stand	**Time (s)**	21.83	22.04	**1.2**
	Fatigue (s)	1.55	1.29	24.91
	Number Sequences	5	5	**0**

TABLE I

COMPARISON OF THE PROPOSED METRICS BETWEEN ZED SDK AND OPTITRACK AS THE GOLD STANDARD REFERENCE.

values is only 0.26 seconds. Therefore, the metric is still considered a viable option.

It's important to note that there is a significant error in the step width measurement for both legs, with an error rate above 30 percent. Additionally, the means for the left and right legs are different. This is because the OptiTrack markers on the ankles are placed outside of the leg, while the ZED SDK provides an estimated midpoint in the ankle. As a result, there is an offset in the metric, which can be compensated for to obtain the correct step width measurement.

Generally, the metrics obtained using the camera and the gold standard have similar means in most cases, and the errors are mostly low. The results of the lab tests, including Mediapipe's, can be found at `http://www.iri.upc.edu/groups/perception/#FASE`. Given those promising results, we decided to conduct a pilot study with real end users.

V. EXPERIMENTAL DESIGN

The experiment was conducted in the Physiology of the Exercise laboratory at the University of Barcelona (UFEBELL) and it was approved by the Ethical Committee of the Hospital de Bellvitge. It was designed as a within-subject study, in which each participant underwent a standard

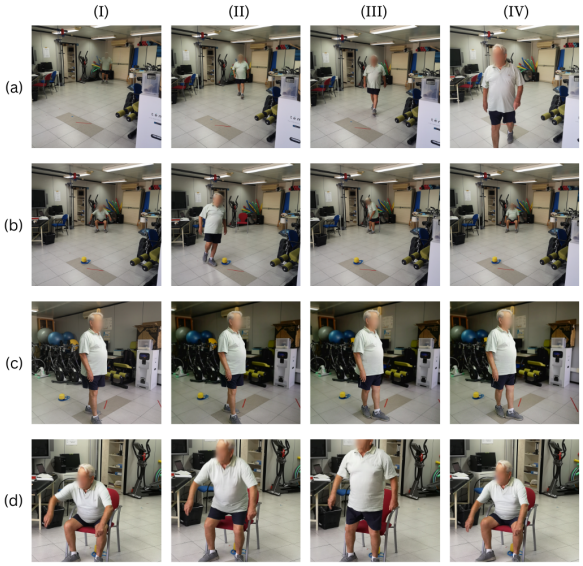

Fig. 4. Examples of sequences (I-IV) of the patient performing each of the four physical tests (a-d). (a) Gait Speed test, (b) TUG test, (c) Standing Balance test, in I and II the patient is standing in a semi-tandem position, in III and IV it is in a tandem position, and (d) 5 Times Sit-Stand test.

clinical check during which they were asked to perform the SPPB and the TUG tests while being assessed simultaneously by a human doctor and our robotic system (independent variable). In both cases, we measured the completion time (dependent variable). We hypothesized based on our previous experiment (see Sec. IV), that the robotic system's time measurement is highly correlated with that of the human doctor.

A. Participants

The inclusion criteria for the study required participants to be 65 years or older, and physically capable of performing the tests. The sample group consisted of 22 older adults ($M = 73.59$, $SD = 5.31$) with varying health and physical conditions to make the sample more heterogeneous.

B. Procedure

All participants were scheduled by time on different days. Random scheduling was employed to allocate participants across different days and hours to mitigate potential biases related to the time of day or day of the week. When they arrived at the lab, they followed the same procedure as in the geriatric clinic and began performing the tests. The assessment took place in a dedicated laboratory facility where geriatric clinicians used to conduct the frailty assessment. The laboratory was equipped with the necessary testing tools and robotics setup to facilitate the assessment process. Upon arrival at the laboratory, participants were provided with informed consent forms detailing the purpose and procedures of the study. They were briefed on the nature of the tests and any relevant instructions before commencing the assessment. The assessment comprised a series of standardized tests, following the established protocol used in geriatric clinics (see

Fig. 4). Each participant underwent the tests sequentially, with trained research personnel and doctors overseeing the process.

C. Measures

During the assessment, the doctor collected data using a stopwatch, while our system automatically captured skeletal data and completion time.

For the Gait Speed, TUG, and 5 Times Sit-Stand tests, the variable collected was time, while in the Standing Balance, a clinical measure was used. This number is based on the times of each foot position. For the together and semi-tandem positions, if a person kept the balance for 10 seconds, they scored one point. For the tandem position, if the balance was kept for 10 seconds, they scored two points, if it was kept between 3 and 10 seconds, they scored 1 point, otherwise, they got 0 points.

D. Results and Discussion

In this section, we analyze the results of the experiment comparing the completion time for each test computed using our system with those of the geriatrician. To do so, we conducted a correlation analysis using Pearson's correlation coefficient when possible.

For the Standing Balance test, the geriatrician reported that all patients except one performed the tests correctly ($M = 3.95$, $SD = 0.21$), while our system results showed that all of them performed correctly ($M = 4$, $SD = 0$), this suggests that the thresholds can be slightly adjusted considering the geriatrician criteria. In this case, as both sets of data are identical except for one entry, it means that there is no variability to measure the degree of linear association between them. As a result, we did not compute the correlation coefficient.

For the Gait speed test, we found a significant positive correlation between the completion times obtained from our system ($M = 3.06$, $SD = 0.56$) and those assessed by the geriatrician ($M = 2.95$, $SD = 0.59$) ($r(22) = 0.92$, $p < 0.05$), with an R^2 value of 0.85.

Similarly, for the Sit-Stand test, we observed a highly significant positive correlation between the completion times derived from our system ($M = 10.53$, $SD = 1.72$) and those measured by the geriatrician ($M = 10.82$, $SD = 1.8$) ($r(22) = 0.98$, $p < 0.01$), yielding an R^2 value of 0.96.

Finally, for the TUG test, we found a strong positive correlation between the completion times obtained from our system ($M = 7.01$, $SD = 1.47$) and those recorded by the geriatrician ($M = 7.68$, $SD = 1.49$) ($r(22) = 0.93$, $p < 0.01$), with an associated coefficient of determination R^2 of 0.88. It is interesting to note that the difference is the highest among the 4 tests due to the clusterization of the data that accounts for detecting whether the patient stands up or sits down. As can be observed in Fig. 3, the body is already in motion when the person is considered standing or sitting. This means that the proposed method might measure less time than the geriatrician, which starts counting at the moment the person starts moving.

Overall, these results indicate a strong agreement between the completion times obtained from our system and those recorded by the geriatrician across all tested tasks, suggesting the reliability and validity of our approach in assessing frailty in older adults.

VI. CONCLUSION AND FUTURE WORK

In this article, we propose a set of algorithms that enable robots to assess frailty automatically in the same way as geriatricians currently do. Moreover, we demonstrate that our system can extract complementary and significant frailty-related metrics requested by clinicians from those tests.

To evidence the effectiveness of our robotic system, we conducted two evaluation experiments. In the first experiment, we evaluated the accuracy of the skeleton tracking of our system and that of the OptiTrack, which is considered the gold standard, in measuring the frailty metrics. The results showed satisfactory performance of our system.

In the second experiment, we conducted a pilot study with 22 older adults in a hospital and evaluated the correlation between the completion time measured by our system and that measured by the geriatrician. Our findings indicated a high correlation between our system and that of the clinician.

The next phase of development will focus on developing new human-robot interaction techniques to enable robots to personalize the interaction with respect to the different patient's unique needs and adjust it to different situations. There may be instances where the patient performs a test incorrectly because they cannot understand the instructions, decide voluntarily to cheat, or are unable to complete it. The robot will need to handle these challenges. At the same time, we plan to conduct more experiments in real-world scenarios to (i) validate user acceptance and (ii) compare the additional metrics obtained with our system with other sensors used by clinicians (e.g., accelerometers and gait platform).

REFERENCES

[1] "Global Aging." [Online]. Available: https://www.nia.nih.gov/research/dbsr/global-aging

[2] L. P. Fried, C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, M. A. McBurnie, and Cardiovascular Health Study Collaborative Research Group, "Frailty in older adults: evidence for a phenotype," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 56, no. 3, pp. M146–156, Mar. 2001.

[3] C. G. Ethun, M. A. Bilen, A. B. Jani, S. K. Maithel, K. Ogan, and V. A. Master, "Frailty and cancer: Implications for oncology surgery, medical oncology, and radiation oncology," *CA: a cancer journal for clinicians*, vol. 67, no. 5, pp. 362–377, Sep. 2017.

[4] H. Lee, E. Lee, and I.-Y. Jang, "Frailty and comprehensive geriatric assessment," *Journal of Korean medical science*, vol. 35, no. 3, 2020.

[5] J. Guralnik, E. Simonsick, L. Ferrucci, R. Glynn, L. Berkman, D. Blazer, P. Scherr, and R. Wallace, "A Short Physical Performance Battery Assessing Lower Extremity Function: Association With Self-Reported Disability and Prediction of Mortality and Nursing Home Admission," *Journal of gerontology*, vol. 49, pp. M85–94, Apr. 1994.

[6] 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.

[7] O. Bruyère, F. Buckinx, C. Beaudart, J.-Y. Reginster, J. Bauer, T. Cederholm, A. Cherubini, C. Cooper, A. J. Cruz-Jentoft, F. Landi *et al.*, "How clinical practitioners assess frailty in their daily practice: an international survey," *Aging clinical and experimental research*, vol. 29, pp. 905–912, 2017.

[8] A. Martínez-Ramírez, P. Lecumberri, M. Gómez, L. Rodriguez-Mañas, F. García, and M. Izquierdo, "Frailty assessment based on wavelet analysis during quiet standing balance test," *Journal of biomechanics*, vol. 44, no. 12, pp. 2213–2220, 2011.

[9] M. Schwenk, C. Howe, A. Saleh, J. Mohler, G. Grewal, D. Armstrong, and B. Najafi, "Frailty and Technology: A Systematic Review of Gait Analysis in Those with Frailty," *Gerontology*, vol. 60, no. 1, pp. 79–89, 08 2013.

[10] N. Millor, P. Lecumberri, M. Gómez, A. Martínez-Ramírez, and M. Izquierdo, "An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, pp. 1–9, 2013.

[11] A. Martínez-Ramírez, I. Martinikorena, M. Gómez, P. Lecumberri, N. Millor, L. Rodríguez-Mañas, F. J. García García, and M. Izquierdo, "Frailty assessment based on trunk kinematic parameters during walking," *Journal of Neuroengineering and Rehabilitation*, vol. 12, no. 1, pp. 1–10, 2015.

[12] L. Ruiz-Ruiz, A. R. Jimenez, G. Garcia-Villamil, and F. Seco, "Detecting Fall Risk and Frailty in Elders with Inertial Motion Sensors: A Survey of Significant Gait Parameters," *Sensors*, vol. 21, no. 20, p. 6918, Jan. 2021.

[13] M. Ando, N. Kamide, M. Sakamoto, and Y. Shiba, "Step length is associated with comprehensive frailty status in community-dwelling older people," *Geriatrics & Gerontology International*, vol. 24, no. 1, pp. 18–24, 2024.

[14] A. Andriella, C. Torras, C. Abdelnour, and G. Alenyà, "Introducing caresser: A framework for in situ learning robot social assistance from expert knowledge and demonstrations," *User Modeling and User-Adapted Interaction*, vol. 33, pp. 441–496, 2023.

[15] J. Holland, L. Kingston, C. McCarthy, E. Armstrong, P. O'Dwyer, F. Merz, and M. McConnell, "Service Robots in the Healthcare Sector," *Robotics*, vol. 10, no. 1, p. 47, Mar. 2021.

[16] A. Civit, A. Andriella, C. Barrué, M. Antonio, C. Boqué, and G. Alenyà, "Introducing social robots to assess frailty in older adults," in *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, Mar. 2024, pp. 342–346.

[17] E. R. Vieira, R. A. Da Silva, M. T. Severi, A. C. Barbosa, B. C. Amick III, J. C. Zevallos, I. L. Martinez, and P. H. Chaves, "Balance and gait of frail, pre-frail, and robust older hispanics," *Geriatrics*, vol. 3, no. 3, p. 42, 2018.

[18] Kidziński, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Communications*, vol. 11, no. 1, p. 4054, Aug. 2020.

[19] X. Gu, F. Deligianni, B. Lo, W. Chen, and G. Yang, "Markerless gait analysis based on a single RGB camera," in *IEEE International Conference on Wearable and Implantable Body Sensor Networks*, Mar. 2018, pp. 42–45.

[20] 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.

[21] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with Kinect," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2012, pp. 1964–1967.

[22] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li, "Reliability and Validity of Kinect RGB-D Sensor for Assessing Standing Balance," *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1633–1638, May 2014.

[23] E. Gianaria, M. Grangetto, M. Roppolo, A. Mulasso, and E. Rabaglietti, "Kinect-based gait analysis for automatic frailty syndrome assessment," in *IEEE International Conference on Image Processing*, Sep. 2016, pp. 1314–1318.

[24] L. Duncan, S. Zhu, M. Pergolotti, S. Giri, H. Salsabili, M. Faezipour, S. Ostadabbas, and S. A. Mirbozorgi, "Camera-Based Short Physical Performance Battery and Timed Up and Go Assessment for Older Adults with Cancer," *IEEE Transactions on Biomedical Engineering*, pp. 1–10, 2023.

[25] R. A. C. M. Olde Keizer, L. van Velsen, M. Moncharmont, B. Riche, N. Ammour, S. Del Signore, G. Zia, H. Hermens, and A. N'Dja, "Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges," *Health and Technology*, vol. 9, no. 4, pp. 595–605, Aug. 2019.

[26] G. Palestra, C. Granata, I. Hupont, and M. Chetouani, *Technology for Assisting During the Comprehensive Geriatric Assessment Process: The ASSESSTRONIC Project*. Cham: Springer International Publishing, 2020, pp. 229–247.

[27] F. Ahmed, P. P. Paul, and M. Gavrilova, "DTW-based kernel and rank level fusion for 3D gait recognition using Kinect," *The Visual Computer*, vol. 31, pp. 915–924, Apr. 2015.

[28] L. Ortiz, E. Cabrera, and L. Gonçalves, "Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs," *Electronic Letters on Computer Vision and Image Analysis*, vol. 17, Apr. 2018.

[29] K. Khoshelham, "Accuracy analysis of Kinect depth data," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-5/W12, May 2012.

[30] T. B. d. G. Lafayette, V. H. d. L. Kunst, P. V. d. S. Melo, P. d. O. Guedes, J. M. X. N. Teixeira, C. R. d. Vasconcelos, V. Teichrieb, and A. E. F. da Gama, "Validation of Angle Estimation Based on Body Tracking Data from RGB-D and RGB Cameras for Biomechanical Assessment," *Sensors*, vol. 23, no. 1, p. 3, Jan. 2023.

[31] G. Nagymáté and R. M. Kiss, "Application of optitrack motion capture systems in human movement analysis: A systematic literature review," *Recent Innovations in Mechatronics*, vol. 5, no. 1., pp. 1–9, 2018.