# Insights on the Role of Depth Information for Human Motion Evaluation Using a Single RGB-D Camera

Beatrice Lagomarsino[1]
beatrice.lagomarsino@edu.unige.it
Giorgia Marchesi[2,4]
giorgia.marchesi@movendo.technology
Valeria Falzarano[2]
valeria.falzarano@movendo.technology
Tommaso Falchi Delitala[2,4]
tommaso.falchi@movendo.technology
Francesca Odone[1,3,4]
francesca.odone@unige.it
Maura Casadio[1,4]
maura.casadio@unige.it
Matteo Moro[1,3,4]
matteo.moro@unige.it

[1] DIBRIS University of Genoa
Genoa, Italy
[2] Movendo Technology s.r.l
Genoa, Italy
[3] Machine Learning Genoa Center
University of Genoa Genoa, Italy
[4] RAISE Ecosystem
16126, Genoa, Italy.

**Abstract**

Telerehabilitation is a promising approach for delivering remote therapy, particularly for patients with neurological disorders who need continuous monitoring and feedback. Single-camera markerless motion capture offers a convenient solution for motion assessment, especially suited to home environments, unlike multi-camera or marker-based systems. However, most studies focus on movements within the frontal plane, neglecting the accuracy of these systems for complex, multi-directional movements involving the sagittal plane. This study aims to characterise motion using a single RGB-D camera by comparing two different depth estimation methods: one based on 3D pose estimation from RGB video and the other using the camera's built-in depth sensor. Both methods are compared to a marker-based system, recognised as the gold standard. The findings indicate that while single-camera methods are accurate for frontal plane movements, they show significant differences with respect to the gold-standard in the sagittal plane due to depth estimation and joint occlusion issues. Nonetheless, the study highlights the potential of video-based markerless systems in telerehabilitation.

## 1 Introduction

Telerehabilitation is a method of providing rehabilitation services through telecommunication networks and internet, allowing people to access therapy from the comfort of their own homes or remote locations [18, 19, 39]. Telerehabilitation has demonstrated its effectiveness

© 2024. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

and benefits, particularly in the treatment of various neurological disorders, such as stroke, Parkinson's disease, multiple sclerosis, and cerebral palsy [25, 29, 37]. By enabling more frequent and accessible therapy sessions, telerehabilitation can improve patient outcomes, reduce the load on healthcare facilities and improve patient compliance with rehabilitation protocols. This approach includes a range of motor and cognitive exercises, some of which are more popular and commonly used. Motor exercises are designed to target muscle strength, endurance, balance, and aerobic capacity. Examples include knee flexion, trunk flexion and extension, sit-to-stand, step-ups, and walking [31]. While telerehabilitation shares similarities with traditional home training by enabling patients to exercise outside clinical settings, it differs significantly in its structured, interactive and technology-supported approach. While home training typically relies on patients autonomously following instructions contained in cards or videos, telerehabilitation aims to improve rehabilitation outcomes by providing real-time feedback [22]. This highlights a key challenge for telerehabilitation: ensuring accurate and reliable evaluation and feedback on users' movements and performance to motivate and guide them during therapy sessions [44, 45]. To address this challenge, markerless motion capture technology emerges as a promising solution, allowing the tracking and the analysis of human movement without the need for body markers, which are both time-consuming (for the setup) and user-dependent [6, 14, 33]. In particular, in recent years, techniques based on video analysis have been proposed for extracting quantitative parameters regarding movement (*e.g.*, joint angles, velocities, and trajectories [11, 13, 23, 24]). These systems use pose estimation algorithms based on deep learning and computer vision (*i.e.*, [7, 10, 43]) to detect points of interest on the human body. They have the advantage of being cheaper and less cumbersome than marker-based systems. However, from a clinical perspective, their accuracy has not been thoroughly explored and depends on various factors, including the quality and resolution of the images or videos captured, camera positioning, lighting conditions, occlusions of certain body parts, and the reliability of the pose estimation algorithms used to extract the position of the points of interest [6, 40]. In addition, such models are often constrained to two-dimensional (2D) motion analysis [32] or the use of multiple cameras [33] for the geometric reconstruction of three-dimensional (3D) information through stereo-vision techniques. Although the latter approach is essential for biomechanical analysis, single-camera-based motion analysis systems are emerging as alternatives to be explored. In this context, devices such as depth cameras (RGB-D) and recent development of 3D pose estimation algorithms on single images as input [7, 10, 17, 43] have the potential to further simplify motor task estimation by eliminating the need for two or more cameras. However, the accuracy and precision of such systems are still unclear, raising questions about their actual applicability in rehabilitation contexts. Although these systems have the ambition of making motion analysis possible in situations where it would otherwise be difficult or impossible through the use of markers (as, for example, in the case of home rehabilitation [21, 36, 38]), many studies using depth sensors (*i.e.*, Kinect or Leap Motion) have reported limited accuracy due to camera occlusion and difficulties in recognising movement characteristics [5, 9, 12, 20, 27, 35, 36, 41]. In addition, studies using a single camera often report exercises on the acquisition plane only, and depth information is generally not explored [8, 16, 26, 42]. Given the above-mentioned considerations, research on the accuracy of motion characterisation through video analysis tools, and tracking out-of-plane movements with respect to camera orientation or influenced by occlusion appears necessary. Given the current state of the art, this study aims to quantify motion using a single RGB-D camera by employing two approaches of depth estimation: (i) using the depth data from the sensor itself, and (ii) employing a monocular 3D pose estimation algorithm on the RGB video. The objective is to evaluate how these approaches affect the estimation of quantitative angle measures compared to those extracted with a gold-standard stereophotogrammetric system. This study has the potential to highlight the advantages and limitations of current video-based motion capture technologies in the clinical setting. In doing so, it sets the basis for bridging the gap between emerging technologies and clinical needs, with the ultimate goal of improving the accuracy and reliability of telerehabilitation interventions.

# 2 Materials and Methods

## 2.1 Participants

Ten unimpaired volunteers (age: 26.5 mean ± 2.5 standard deviation (SD) years, two males) participated in this study. The inclusion criterion for this study was the absence of any history of neurological or orthopaedic disorders. This study conforms to the ethical principles for medical research involving human subjects of the Declaration of Helsinki (revision 2013) and was approved by the local ethical committee (Comitato etico per la ricerca di Ateneo CERA - Università degli Studi di Genova, protocol n. 2023/93, 14/12/2023). All participants signed an informed consent for the analysis and publication of their data for research purposes.

## 2.2 Experimental Setup

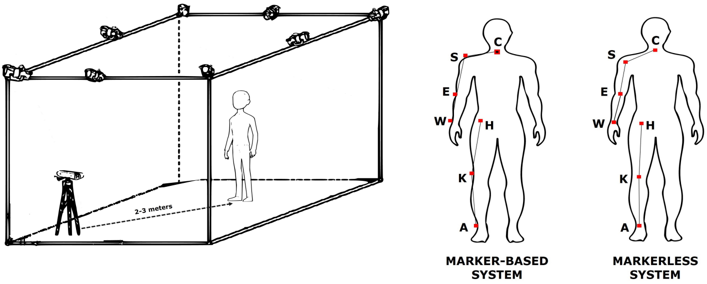
The experiment was conducted at the motion analysis laboratory of the University of Genoa. The set-up included a stereophotogrammetric system (Vicon, [2]) consisting of eight infrared cameras recording at 100$Hz$ calibrated to measure body movements (Figure 1a). We recorded 32 passive reflective spherical markers with a diameter of 19 mm placed in precise anatomical positions defined according to the DAVIS protocol [15] (partially shown in Figure 1b). A single RGB-D camera (Intel RealSense D435, [1]) recording at 30$Hz$ was included in the setup. This camera has an integrated depth sensor with stereo cameras and an infrared projector for high-resolution 3D depth data, along with an RGB camera. For markerless motion tracking, we recorded both RGB video and depth information. This was positioned in front of the participants (*i.e.*, who were facing the camera, Figure 1a) as we were interested in a possible scenario applicable to telerehabilitation. To overcome potential inaccuracies in depth measurements caused by clothing, all participants were required to wear tight-fitting attire during the experiments.

## 2.3 Protocol

The participants performed a series of exercises in random order, involving all planes of movement, at distances of two and three meters from the camera (Figure 1a). The exercises are represented in Figure 2 and could be categorized as follows:

**Trunk Exercises** Participants were instructed to perform trunk flexion (F) in the sagittal plane and lateral tilting (LT) in the frontal plane (Figure 2b). These movements were executed both in a standing position and while seated.

**Range of Motion (RoM) Exercises** Participants were instructed to perform shoulder and hip movements with both the upper and lower limbs. These included abduction (A) in the



(a) (b)

Figure 1: Setup. (a) The stereophotogrammetric system with eight infrared cameras and the RGB-D camera. (b) Markers and keypoints used for angle computation: C (C7 or midpoint between shoulders), S (shoulder), E (elbow), W (wrist), H (hip), K (knee), A (ankle). Left: Marker-based system; Right: Markerless system.

frontal plane, flexion (F), and extension (E) in the sagittal plane, as well as flexion-abduction (FA) and extension-abduction (EA) in both planes (Figure 2c). For shoulder movements, in particular, participants performed flexion, abduction, and flexion-abduction at three different amplitudes, aiming to create movement angles of 45, 90, and 180 degrees. For hip movements, participants were not required to achieve high amplitude, focusing instead on comfortable and controlled motion within a safe range.

**Body-weight Exercises** Participants were instructed to perform elbow flexions, squats, and skips (Figure 2a).

For all exercises, participants were required to return to the initial position after each movement. Each movement was repeated 10 times. Each exercise that required limb movements was performed with both the right and left limbs.

## 2.4 Data Analysis

For the gold standard marker-based approach (*Marker*), we recorded the (x, y, z) coordinates of the passive markers positioned on the body. The Vicon Nexus software [4], typically automates marker sorting and tracking using a human body model. However, manual intervention is often required when markers are occluded or affected by reflections, making the process time-consuming. This highlights a significant drawback of marker-based motion capture systems. At the end of this process, we obtained 10 matrices ($P_{\text{marker}_j}$ with $j = 1, ..., 10$ indicating the index for each participant), with shape $32 \times 3 \times M_j$ (32 markers; 3 for (x, y, z) markers' coordinates; $M_j$ for the number of samples for the acquisition of the $j-th$ participant).

For markerless approaches, we used MediaPipe Pose [7]. MediaPipe Pose uses a deep learning pipeline for real-time human pose estimation from RGB video. The framework is based on a convolutional neural network (CNN) architecture consisting of two main stages. The network is trained on large-scale human pose datasets, such as COCO and MPII [28], which include a wide range of human poses in various environments. In the first stage, the BlazePose detector identifies the region of interest (ROI) corresponding to the human body within the frame. In the second stage, the pose tracker predicts the 33 2D keypoints of the human body from the detected ROI. Additionally, the framework can estimate the

3D coordinates of these keypoints through a regression model, using a coordinate system where the origin is defined as the midpoint between the hips. We used this pose estimation algorithm to estimate the $(x, y)$ coordinates of the keypoints of interest. For the z-coordinates, we employed two approaches: *Markerless1* used the depth information extracted from the sensor integrated into the RealSense, and *Markerless2* used the $z$ estimated by MediaPipe. In particular, for *Markerless1*, we applied MediaPipe Pose to the RGB video to obtain the pixel position of the keypoints of interest. Then, we converted the pixel into $(x, y, z)$ coordinates in the reference system centred in the camera using the depth sensor's point cloud of the Real Sense. The final outputs were 10 matrices $P_{\text{markerless1}_j}$ of shape 33 × 3 x $N_j$ ($N_j$ for the number of frames composing the video of the $j-th$ participant). For *Markerless2*, we applied MediaPipe Pose 3D to the RGB video that estimates the $(x, y, z)$ coordinates of the keypoints of interest. We obtained 10 matrices $P_{\text{markerless2}_j}$ of the same shape as matrices $P_{\text{markerless1}_j}$.

Once obtained the coordinates from all the systems, they were filtered using a fourth-order Butterworth low-pass filter with a cutoff frequency of 12 Hz [34]. Then, we calculated different angles of movement using the following formula:

$$\theta_{(\text{A,B})} = \cos^{-1}\left(\frac{A \cdot B}{|A|\,|B|}\right)$$

where A and B represented the body vectors considered for each exercise.
Specifically:

**Trunk Exercises** We calculated the trunk angle in 3D space as the angle between the trunk vector ($trunk$) and the initial trunk vector ($trunk_0$), which was measured at the start of the recording when the participant was in a comfortable position. The trunk vector was defined as connecting the midpoint of the shoulders' landmarks to the midpoint of the hips' landmarks. So we obtained $\theta_{(trunk, trunk_0)}$.

**Range of Motion Exercises** We computed the shoulder and hip angles in 3D space as the angle between the arm or the leg body segments at each time point and their initial position. Specifically, we compared the vectors connecting the shoulder and the wrist, or the hip and the knee landmarks (*arm*, *leg*, respectively) during the task with their corresponding vectors at the start of the acquisition (*arm*$_0$, *leg*$_0$, respectively). We obtained $\theta_{(arm, arm_0)}$ and $\theta_{(leg, leg_0)}$.

**Body-weight Exercises** We computed the elbow angle for the elbow flexion exercise considering the arm and forearm body segments. For the skip and squat exercises, we calculated the knee angle as the angle between the leg and the shank. We obtained $\theta_{(arm, forearm)}$ and $\theta_{(leg, shank)}$.

After obtaining the angles for each system, we extracted the peak values representing the maximum angular displacement for each movement. Then, we computed the mean absolute errors (MAEs) by comparing the maximum angular displacement obtained from the *Marker* approach and both *Markerless* approaches. Specifically, (*MAE#* represents the error between *Marker* and *Markerless#*, # referring to 1 and 2). Finally, the obtained values were averaged for repeated movements.

## 2.5 Statistical Analysis

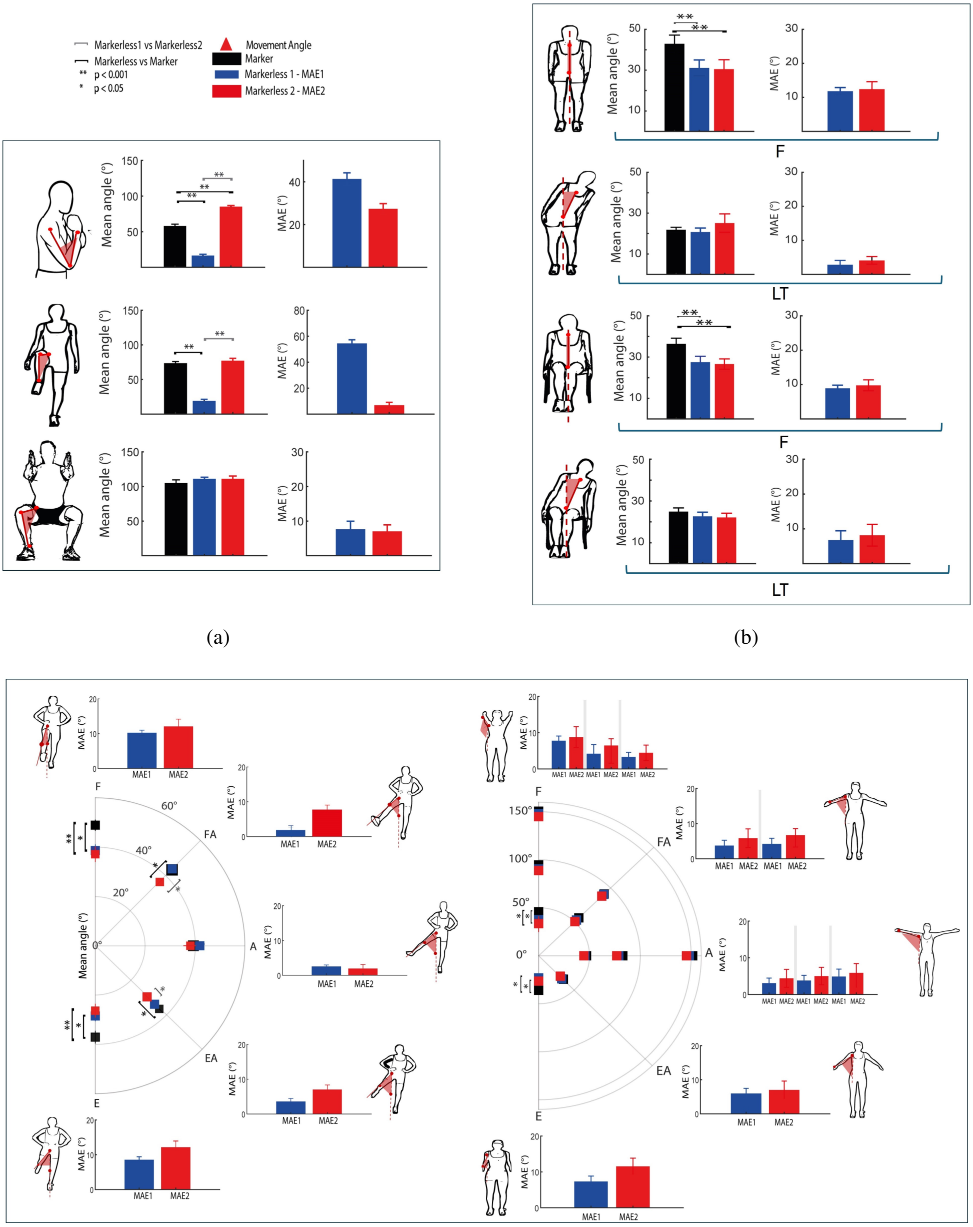
We aimed to evaluate if the performance of the markerless approaches was comparable to a marker-based method for assessing movement in different directions. Additionally, we examined potential differences by considering two distances from the camera and movements performed with the left and right limbs (when applicable). We conducted a repeated measures ANOVA (rANOVA) on angle values, considering four factors: approaches (Marker, Markerless), body side (if any), movement direction, and distance from the camera. The statistics were performed with the open statistical software Jamovi (jamovi.org [3]). The normality of the data was verified using the Shapiro-Wilk test. Sphericity was verified with the Mauchly test and corrected with the Greenhouse–Geisser method. The statistical significance was set at the error rate $\alpha = 0.05$. A Bonferroni-Holm correction for multiple comparisons was applied to the post-hoc rANOVA analysis performed following the significant factors.

# 3 Results

We conducted a quantitative analysis to evaluate the similarity between angle measurements obtained using a *Marker*-based technique and two *Markerless* approaches. For each exercise, we reported the movement angles in different directions obtained from each approach and the mean absolute error (MAE) of the *Markerless* approaches relative to the *Marker*-based one, which was used as ground truth. We reported the mean and standard deviation of these results across all participants (Figure 2).

## Trunk Exercises

We observed that all the systems performed similarly regardless of whether the exercises were performed while standing or seated (Figure 2b). As expected, the computed angle values showed no significant differences with respect to the distance from the camera (*p=0.112* standing, *p= 0.104* sitting). Therefore, we reported the averaged values across both distances. In contrast, a significant difference was found for the interaction factor between the different approaches and movement directions (approaches*direction of movement, *F(1.66,14.76) = 14.833* and *p<0.001* for standing exercises; *F(1.70, 13.6)= 14.27* and *p<0.001* for seated exercises), indicating that the performance of the systems varied depending on the direction of movement. Specifically, for movements performed in the frontal plane (*i.e.*, parallel to the acquisition plane, LT in Figure 2b, 2nd and 4th raws), where depth information is less critical, both *Markerless* approaches were comparable to the gold standard. There were no significant differences observed for either standing or seated exercises in both directions (so, we reported the averaged values). Interestingly, when standing, the *Markerless2* approach tended to overestimate compared to *Markerless1* approach (Figure 2b, LT 2nd raw) but this discrepancy was not observed when measurements were taken while seated (Figure 2, LT 4th raw). Despite these variations, the errors between the gold standard and each *Markerless* approaches were minimal and comparable across both exercises. Conversely, for flexion exercises performed purely in the sagittal plane (Figure 2b, F), both *Markerless* systems showed a significant difference compared to the gold standard (*t(9)=10.620* and *p<0.001* for *Markerless1* and *t(9)= 7.194* and *p<0.001* for *Markerless2* for standing exercises, *t(9)=8.499* and *p<0.001* for *Markerless1* and *t(9)= 9.366* and *p<0.001* for *Markerless2* for seated exercises), with error values increased but still within 10 degrees. Nevertheless, the two *Markerless* systems remained comparable to each other.



(a)

(b)

(c)

Figure 2: Angles and MAE (a) Range of Motion Exercises. (b) Trunk exercises: standing flexion (F) and lateral tilt (F) in the first two rows, seated flexion and lateral tilt in the third and fourth rows (c) Range of Motion Exercises: upper limb exercises on the left, lower limb exercises on the right; flexion (F), extension (E), abduction (A), and flexion-abduction (FA), extension-abduction (EA); for shoulder exercises, the vertical grey lines distinguish the error for different movement amplitudes.

## Range of Motion Exercises

Also in this case, neither the body side (left and right limb) nor the distance from the camera (2 and 3 meters) affected the quantification of performance (body side factors: *p= 0.233* for hip movements, *p= 0.121* for shoulder movements; distance factor: p= 0.326 for hip movements, *p= 0.145* for shoulder movements). Therefore, we averaged data across both distances and limbs. A significant difference was found among the measurement approaches depending on the movement direction (approaches*direction, *F(1.77,15.84)= 4.52* and *p=0.034* for hip movements, *F(7.2,64.8)= 7.19* and *p<0.001* for shoulder movements). Figure 2c shows the mean angle values for each direction of movement for both *Marker* and *Markerless* approaches. For the movements performed purely in the frontal plane (*i.e.*, parallel to the acquisition plane, A in Figure 2c), the angles computed using the three approaches were not significantly different. The MAE was similar for both markerless approaches, regardless of the body district or movement amplitude. Conversely, in the movements where the depth information is crucial (*i.e.* with a sagittal component: F, E, FA, and EA, Figure 2c), the *Markerless* approaches underestimated the angle especially in hip movements. Specifically, for the lower limb (Figure 2c on the left) the difference between the gold standard and *Markerless2* was significant in the combined planes (*p= 0.032* for FA, *p= 0.021* for FE), resulting in a significantly higher error compared to *Markerless1* (*Markerless1* vs *Markerless2*, *p= 0.004* for FA, *p= 0.015* for FE). For movements performed purely in the sagittal plane (*i.e.*, perpendicular to the camera acquisition plane: flexion (F) and extension (E)) both markerless approaches showed highly significant but comparable differences with respect to the gold standard (*p< 0.001* for *Markerless2*, *p= 0.05* and *p= 0.006* for *Markerless1* respectively in F and E directions). On the other hand, for the upper limb (Figure 2c on the right), the three systems reported comparable values for combined planes (FA and EA). However, similar to the observations with the hip, for upper limb movements performed purely in the sagittal plane (F and E) but limited to movement amplitudes less than 50 degrees, the *Markerless* systems significantly but comparably underestimated the angle values compared to the gold standard (*p= 0.009* and *p= 0.014* for *Markerless1*, *p= 0.006* and *p= 0.021* for *Markerless2* respectively in F and E direction). In contrast, when observing shoulder movements with larger amplitudes, the performance of both camera-based systems aligned more closely with the gold standard with no significant differences.

## Body-Weight Exercises

Also in this case, the angles did not show significant differences between the body side (*p=0.302* for elbow, *p=0.262* for skip, and *p=0.298* for squat) and the distance from the camera (*p=0.199* for elbow, *p=0.447* for skip, and *p=0.312* for squat). Therefore, we averaged the data across both distances and limbs for each exercise type. Interestingly, the results of the body-weight exercises contrasted with the previous observations (Figure 2a). Specifically, for elbow flexion performed purely in the sagittal plane, *Markerless1*, which relied on depth information provided by the RealSense camera, significantly miscalculated the elbow angle (*p<0.001*) due to shoulder joint occlusion, with an error of approximately 40 degrees compared to the gold standard (Figure 2a, 1st raw). Conversely, the *Markerless2* overestimated the elbow flexion angle (*p<0.001*) with an error of about 25 degrees (Figure 2a, 1st raw). A similar pattern was observed in the skip exercise (Figure 2a, 2nd raw), performed mainly in the sagittal plane. Here, *Markerless1*, miscalculated the knee angle with a significant error exceeding 50 degrees compared to the gold standard (*t(9)=5.25 p<0.001*)

(Figure 2a, 2nd raw). In contrast, *Markerless2* did not differ significantly from the gold standard and reported a minimal error (Figure 2a, 2nd raw). Finally, in the squat exercise that involved both sagittal and frontal planes, both camera-based systems were comparable to the gold standard, providing a good estimation of the knee angle (Figure 2a, 3rd raw).

## 4 Discussion and Conclusions

This paper discussed two single-camera, markerless approaches for motion analysis. The objective was to evaluate their performance for potential use in a telerehabilitation context. The single-camera approach offers several advantages: it requires less expertise, avoids operator biases, and prevents distortion from improper marker placement on the subject's skin. It also maintains natural movement without cumbersome markers or sensors. Additionally, it is more cost-effective, easier to set up, and user-friendly outside laboratory environments, requiring only one RGB-D camera [6, 33]. Despite these advantages, the camera-based approaches showed results that differed from the gold standard. For exercises performed purely in the frontal plane, where depth information is not required, both camera-based approaches showed comparable results to the gold standard, with non-significant errors consistent with the literature ( [16, 26, 32]). On the other side, for exercises involving depth information due to sagittal or combined-plane movements, the two systems reported significant errors compared to the gold standard. In particular, although for large movements, such as those involving the shoulder, both camera-based systems achieved results comparable to the gold standard, for movements performed in the sagittal plane, such as trunk, hip, and shoulder flexion with a limited range of motion, both markerless approaches reported significant errors compared to the gold standard. This indicates a difficulty in accurately capturing small movements in this plane due to poor depth estimation. This issue can be attributed to two main factors: the low resolution of the depth sensor [20, 41] and incorrect estimation by MediaPipe Pose. Specifically, with the Real-Sense based method, inaccuracies in pixel positioning by MediaPipe Pose can lead to incorrect projection within the point clouds of the camera. For the MediaPipe Pose-based depth estimation, inaccuracies are hypothesized to derive from its method of estimating 3D coordinates centred on the subject's hips [7]. Indeed, this can lead to underestimation and significant errors, particularly in movements involving compensatory hip motion like trunk and hip flexion. In the first case, participants move their hips backwards to perform the exercise, while in the second, they maintain balance by moving and rotating the hips. The MediaPipe Pose-based depth estimation method adjusts its reference system based on hip movement, resulting in incorrect quantification of trunk or leg displacement. This results in significant errors compared to the gold standard, especially for hip and trunk exercises in the sagittal plane. It also differs significantly from the Real-Sense method, particularly during hip movements involving combined planes. The major limitations of camera-based systems emerged in body-weight exercises that involve the evaluation of relative angles and potential occlusions. In both the skip and elbow flexion exercises, occlusion led to significant errors with the RealSense-based method. These exercises, performed primarily in the sagittal plane, involved the wrist and knee joints occluding the shoulder and hip joints, respectively. As a result, the coordinates of these joints coincided in the depth map, causing the body vectors describing movement to be misinterpreted as overlapping. This leads to a significant underestimation of joint positions compared to the gold standard. In contrast, the MediaPipe Pose-based depth estimation method significantly overestimated the elbow flexion angle compared to the gold standard, reporting an angle value of approximately 90 degrees. This highlights the algorithm's difficulty in recognizing the subtle retraction of the wrist towards the body after its initial range of motion. Indeed, during the elbow flexion exercise, the wrist initially moves away from the body to allow the elbow to reach an angle of approximately 90 degrees, then moves back towards the body to complete the motion. MediaPipe Pose-based depth estimation method had difficulty detecting this wrist retraction, leading to significant errors in measuring the elbow angle. Conversely, during the skip exercise, this issue did not occur because the movement is limited to an angle of about 90 degrees, and the knee joint does not move closer to the hips.

## 5 Limitations and Future Works

This study highlighted the strengths and limitations of two single-camera methods for motion assessment. Specifically, the accuracy of depth estimation using the camera's built-in depth sensor was compared with that of a pre-trained deep learning framework applied to RGB images, as well as with a marker-based system, which served as the gold standard for motion analysis. As a preliminary study, the goal was to explore simple, accessible solutions without the need for training or customizing neural networks. The focus was on developing a practical, straightforward system using existing technologies, such as pose estimation algorithms and RGB-D camera, without relying on complex training processes that would require larger, more diverse datasets. However, several limitations were identified. One of the main limitations of this type of study is the impact of environmental variability on the accuracy of depth estimation. Indeed, changes in lighting, background objects, or room layout can negatively affect both depth sensors and pose estimation algorithms, leading to a reduction in motion capture accuracy. However, since all our experiments were conducted in a controlled environment with consistent lighting conditions, we did not consider lighting variations as a factor that could affect the accuracy of the two methods. Future work will explore model fine-tuning techniques and the combination of the presented methods to overcome these limitations. Future research will also compare the current methods with more advanced pose estimation and SMPL algorithms [30], which are known for improved performance and better handling of pose plausibility. Such comparisons could offer valuable insights and potentially enhance the accuracy and robustness of the system. Despite these challenges, the markerless pipeline has shown promise as an alternative for evaluating kinematic parameters across various exercises and planes of motion, with the potential to be effective even in less controlled environments outside traditional lab settings. In both clinical and home rehabilitation contexts, this approach could serve as a valuable tool for quantitative movement assessment and comprehensive exercise monitoring. The system could be adapted to automatically count repetitions, assess movement quality, and provide real-time feedback, which is particularly beneficial for individuals with neurological disorders. For these patients, continuous motor activity and precise monitoring, especially at home, are crucial for effective rehabilitation. By facilitating remote supervision, this approach could allow patients to regularly engage in rehabilitation exercises with confidence, knowing their movements are accurately assessed, even in the absence of a therapist. Overall, the integration of a markerless system into telerehabilitation could revolutionize rehabilitation practices, offering a user-friendly solution that improves patient outcomes through continuous monitoring and feedback.

# 6 Acknowledgement

This work is funded by the European Union—NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project “RAISE—Robotics and AI for Socio-economic Empowerment” (ECS00000035). G.M., T.F.D., F.O., M.C., and M.M., are part of RAISE Innovation Ecosystem.

# References

[1] https://www.intelrealsense.com/.

[2] https://www.vicon.com/.

[3] https://www.jamovi.org/.

[4] https://www.vicon.com/software/nexus/.

[5] Lars Adde, Jorunn L. Helbostad, Alexander R. Jensenius, Gunnar Taraldsen, Kristine H. Grunewaldt, and Ragnhild StØen. Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Developmental Medicine and Child Neurology*, 52, 2010. ISSN 00121622. doi: 10.1111/j.1469-8749.2010.03629.x.

[6] Andrea Avogaro, Federico Cunico, Bodo Rosenhahn, and Francesco Setti. Markerless human pose estimation for biomedical applications: a survey. *Frontiers in Computer Science*, 5, 7 2023. ISSN 2624-9898. doi: 10.3389/fcomp.2023.1153160.

[7] Valentin Bazarevsky and Ivan Grishchenko. Google ai blog: On-device, real-time body pose tracking with mediapipe blazepose. *Google AI Blog*, 2020.

[8] Peter Beshara, David B. Anderson, Matthew Pelletier, and William R. Walsh. The reliability of the microsoft kinect and ambulatory sensor-based motion tracking devices to measure shoulder range-of-motion: A systematic review and meta-analysis, 2021. ISSN 14248220.

[9] A. H. Butt, E. Rovini, C. Dolciotti, P. Bongioanni, G. De Petris, and F. Cavallo. Leap motion evaluation for assessment of upper limb motor skills in parkinson’s disease. In *IEEE International Conference on Rehabilitation Robotics*, 2017. doi: 10.1109/ICORR.2017.8009232.

[10] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 2021. ISSN 19393539. doi: 10.1109/TPAMI.2019.2929257.

[11] Andrea Castelli, Gabriele Paolini, Andrea Cereatti, and Ugo Della Croce. A 2d markerless gait analysis methodology: Validation on healthy subjects. *Computational and Mathematical Methods in Medicine*, 2015, 2015. ISSN 17486718. doi: 10.1155/2015/186780.

[12] Kai Hsiang Chen, Po Chieh Lin, Yu Jung Chen, Bing Shiang Yang, and Chin Hsien Lin. Development of method for quantifying essential tremor using a small optical device. *Journal of Neuroscience Methods*, 266, 2016. ISSN 1872678X. doi: 10.1016/j.jneumeth.2016.03.014.

[13] Ross A. Clark, Stephanie Vernon, Benjamin F. Mentiplay, Kimberly J. Miller, Jennifer L. McGinley, Yong Hao Pua, Kade Paterson, and Kelly J. Bower. Instrumenting gait assessment using the kinect in people living with stroke: reliability and association with balance tests. *Journal of neuroengineering and rehabilitation*, 12:15, 2 2015. ISSN 17430003. doi: 10.1186/s12984-015-0006-8.

[14] Steffi L. Colyer, Murray Evans, Darren P. Cosker, and Aki I.T. Salo. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system, 2018. ISSN 21989761.

[15] R.B. Davis. Clinical gait analysis. *IEEE Engineering in Medicine and Biology Magazine*, 7(3):35–40, 1988. doi: 10.1109/51.7933.

[16] Thiago Buarque de Gusmao Lafayette, Victor Hugo de Lima Kunst, Pedro Vanderlei de Sousa Melo, Paulo de Oliveira Guedes, Joao Marcelo Xavier Natario Teixeira, Cinthia Rodrigues de Vasconcelos, Veronica Teichrieb, and Alana Elza Fontes da Gama. Validation of angle estimation based on body tracking data from rgb-d and rgb cameras for biomechanical assessment. *Sensors*, 23, 2023. ISSN 14248220. doi: 10.3390/s23010003.

[17] Hao Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, 2023. ISSN 19393539. doi: 10.1109/TPAMI.2022.3222784.

[18] Nada Fares, R. Simon Sherratt, and Imad H. Elhajj. Directing and orienting ict healthcare solutions to address the needs of the aging population, 2021. ISSN 22279032.

[19] Alessia Gallucci, Pietro Davide Trimarchi, Carlo Abbate, Cosimo Tuena, Elisa Pedroli, Fabrizia Lattanzio, Marco Stramba-Badiale, Matteo Cesari, and Fabrizio Giunco. Ict technologies as new promising tools for the managing of frailty: a systematic review, 2021. ISSN 17208319.

[20] Brook Galna, Gillian Barry, Dan Jackson, Dadirayi Mhiripiri, Patrick Olivier, and Lynn Rochester. Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait and Posture*, 39, 2014. ISSN 18792219. doi: 10.1016/j.gaitpost.2014.01.008.

[21] Ronny Grunert, Andre Krause, Silvio Feig, Juergen Meixensberger, Christian Rotsch, Welf Guntram Drossel, Peter Themann, and Dirk Winkler. A technical concept of a computer game for patients with parkinson's disease - a new form of pc-based physiotherapy. *International Journal of Neuroscience*, 129, 2019. ISSN 15635279. doi: 10.1080/00207454.2019.1567510.

[22] Tian Huang, Wei Zhang, Bing Yan, Haoyang Liu, and Olivier Girard. Comparing telerehabilitation and home-based exercise for shoulder disorders: A systematic review

and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, 2024. ISSN 0003-9993. doi: https://doi.org/10.1016/j.apmr.2024.02.723. URL https://www.sciencedirect.com/science/article/pii/S0003999324008360.

[23] Lukasz Kidzinski, Bryan Yang, Jennifer L. Hicks, Apoorva Rajagopal, Scott L. Delp, and Michael H. Schwartz. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications*, 11, 12 2020. ISSN 20411723. doi: 10.1038/s41467-020-17807-z.

[24] Bogdan Kwolek, Agnieszka Michalczuk, Tomasz Krzeszowski, Adam Switonski, Henryk Josinski, and Konrad Wojciechowski. Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition. *Multimedia Tools and Applications*, 78, 2019. ISSN 15737721. doi: 10.1007/s11042-019-07945-y.

[25] Winnie W.T. Lam, Yuk Ming Tang, and Kenneth N.K. Fong. A systematic review of the applications of markerless motion capture (mmc) technology for clinical measurement in rehabilitation, 2023. ISSN 17430003.

[26] Ameur Latreche, Ridha Kelaiaia, Ahmed Chemori, and Adlen Kerboua. Reliability and validity analysis of mediapipe-based measurement system for some human rehabilitation motions. *Measurement: Journal of the International Measurement Confederation*, 214, 2023. ISSN 02632241. doi: 10.1016/j.measurement.2023.112826.

[27] Wee Lih Lee, Nicholas C. Sinclair, Mary Jones, Joy L. Tan, Elizabeth L. Proud, Richard Peppard, Hugh J. McDermott, and Thushara Perera. Objective evaluation of bradykinesia in parkinson's disease using an inexpensive marker-less motion tracking system. *Physiological Measurement*, 40, 2019. ISSN 13616579. doi: 10.1088/1361-6579/aafef2.

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015. URL https://arxiv.org/abs/1405.0312.

[29] Roberto Llorens, Enrique Noe, Carolina Colomer, and Mariano Alcaniz. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation*, 96:418–425.e2, 3 2015. ISSN 1532821X. doi: 10.1016/j.apmr.2014.10.019.

[30] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), October 2015. ISSN 0730-0301. doi: 10.1145/2816795.2818013. URL https://doi.org/10.1145/2816795.2818013.

[31] Giuseppa Maresca, Maria Grazia Maggio, Rosaria De Luca, Alfredo Manuli, Paolo Tonin, Loris Pignolo, and Rocco Salvatore Calabrò. Tele-neuro-rehabilitation in italy: State of the art and future perspectives, 2020. ISSN 16642295.

[32] Matteo Moro, Giorgia Marchesi, Francesca Odone, and Maura Casadio. Markerless gait analysis in stroke survivors based on computer vision and deep learning: A pilot

study. In *Proceedings of the ACM Symposium on Applied Computing*, pages 2097–2104. Association for Computing Machinery, 3 2020. ISBN 9781450368667. doi: 10.1145/3341105.3373963.

[33] Matteo Moro, Giorgia Marchesi, Filip Hesse, Francesca Odone, and Maura Casadio. Markerless vs. marker-based gait analysis: A proof of concept study. *Sensors*, 22, 3 2022. ISSN 14248220. doi: 10.3390/s22052011.

[34] Matteo Moro, Giorgia Marchesi, Filip Hesse, Francesca Odone, and Maura Casadio. Markerless vs. marker-based gait analysis: A proof of concept study. *Sensors*, 22(5), 2022. ISSN 1424-8220. doi: 10.3390/s22052011. URL https://www.mdpi.com/1424-8220/22/5/2011.

[35] Takuto Nakamura, Satoko Sekimoto, Genko Oyama, Yasushi Shimo, Nobutaka Hattori, and Hiroyuki Kajimoto. Pilot feasibility study of a semi-automated three-dimensional scoring system for cervical dystonia. *PLoS ONE*, 14, 2019. ISSN 19326203. doi: 10.1371/journal.pone.0219758.

[36] Edwin Daniel Ona, Carlos Balaguer, Roberto Cano-De La Cuerda, Susana Collado-Vazquez, and Alberto Jardón. Effectiveness of serious games for leap motion on the functionality of the upper limb in parkinson's disease: A feasibility study. *Computational Intelligence and Neuroscience*, 2018, 2018. ISSN 16875273. doi: 10.1155/2018/7148427.

[37] Shintaro Oyama, Masaomi Saeki, Satoshi Kaneta, Shingo Shimoda, Hidemasa Yoneda, and Hitoshi Hirata. Telerehabilitation based on markerless motion capture and imt-2020 (5g) networks. In *Studies in Health Technology and Informatics*, volume 290, 2022. doi: 10.3233/SHTI220291.

[38] Guillermo Palacios-Navarro, Ivan García-Magarino, and Pedro Ramos-Lorente. A kinect-based system for lower limb rehabilitation in parkinson's disease patients: a pilot study. *Journal of Medical Systems*, 39, 2015. ISSN 1573689X. doi: 10.1007/s10916-015-0289-0.

[39] Alessandro Peretti, Francesco Amenta, Seyed Khosrow Tayebati, Giulio Nittari, and Syed Sarosh Mahdi. Telerehabilitation: Review of the state-of-the-art and areas of application, 2017. ISSN 23692529.

[40] Bradley Scott, Martin Seyres, Fraser Philp, Edward K. Chadwick, and Dimitra Blana. Healthcare applications of single camera markerless motion capture: a scoping review, 2022. ISSN 21678359.

[41] Toshio Tsuji, Shota Nakashima, Hideaki Hayashi, Zu Soh, Akira Furui, Taro Shibanoki, Keisuke Shima, and Koji Shimatani. Markerless measurement and evaluation of general movements in infants. *Scientific Reports*, 10, 2020. ISSN 20452322. doi: 10.1038/s41598-020-57580-z.

[42] Zun Rong Wang, Ping Wang, Liang Xing, Li Ping Mei, Jun Zhao, and Tong Zhang. Leap motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients. *Neural Regeneration Research*, 12:1823–1831, 11 2017. ISSN 18767958. doi: 10.4103/1673-5374.219043.

[43] Xu Xu, Raymond W. McGorry, Li Shan Chou, Jia hua Lin, and Chien chi Chang. Accuracy of the microsoft kinect™ for measuring gait parameters during treadmill walking. *Gait and Posture*, 42, 2015. ISSN 18792219. doi: 10.1016/j.gaitpost.2015.05.002.

[44] Hang Yan, Beichen Hu, Gang Chen, and E. Zhengyuan. Real-time continuous human rehabilitation action recognition using openpose and fcn. In *Proceedings - 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering, AEMCSE 2020*, 2020. doi: 10.1109/AEMCSE50948.2020.00058.

[45] Fan Yang, Yang Wu, Sakriani Sakti, and Satoshi Nakamura. Make skeleton-based action recognition model smaller, faster and better. In *1st ACM International Conference on Multimedia in Asia, MMAsia 2019*, 2019. doi: 10.1145/3338533.3366569.