Lower Limbs Human Motion Estimation From Sparse Multi-Modal Measurements

Mohamed Adjel, Maxime Sabbah, Raphael Dumas, Marta Mirkov, Nicolas Mansard, Samer Mohammed, Vincent Bonnet

Lower Limbs Human Motion Estimation From Sparse Multi-Modal Measurements

Mohamed Adjel[1,2], Maxime Sabbah[3], Raphael Dumas[4],
Marta Mirkov[5], Nicolas Mansard[3], Samer Mohammed[1], Vincent Bonnet[3,6]

***Abstract*— This study aimed at the estimation of the 3D lower-limb joint kinematics during a sit-to-stand and a squat exercises using a new affordable motion capture system. Utilizing a reduced number of affordable visual inertial measurement units and markerless data, the study investigates the performance of these modalities in comparison to a reference stereophotogrammetric system. Indeed, markerless data are easily accessible from an RGB image, but few studies investigated their accuracy to perform inverse kinematics for rehabilitation exercise. Thus, ankle, knee, and hip joint center positions and joint angles were obtained through a novel sliding windows inverse kinematics algorithm. Joint angles were estimated with an average error of 8.1deg when inertial and visual data were used and 13.4deg when using solely markerless data. Joint center positions also displayed an estimation error reduced by 2.5 times when using the proposed approach over purely markerless data. These results, associated with the real affordability and ease of use of the proposed system open the door to future field applications in both rehabilitation and sport.**

I. INTRODUCTION

The challenge of creating minimally invasive instrumented clinical and sport protocols for accurately assessing an individual's motor capacity and performance was the target of numerous recent researches [1], [2], [3]. In the last decade, there has been a growing trend in the bioengineering community to maximize functional information extracted from low-cost and easy-to-use instruments and simplified experimental protocols. Beside walking, the most used exercises for balance recovery rehabilitation process are the squat [4] and the sit-to-stand [5] tasks. Squatting is highly effective for training lower-limb muscles during rehabilitation exercises or sport training, and is essential for daily tasks such as picking up items. Squat exercise is also used in assessments for pathologies that affect ankle joint range of motion, and age-related changes in lower limbs movements. However the efficiency of a squat is directly related to the execution of the task and of the underlying subject's motivation.

Difficulties in standing up, which is a fundamental movement and a prerequisite to standing or walking, pose a significant injury risk for individuals with limited mobility, including the elderly, and persons with musculoskeletal disorders, arthritis, or neurological conditions like stroke or Parkinson's disease. Sit-to-stand biomechanics, extensively studied [6], vary notably among individuals, especially those with disabilities. Analyzing squat and sit-to-stand biomechanics involves key metrics like ankle, knee, and hip 3D joint kinematics, crucial for evaluating body and activity levels [7].

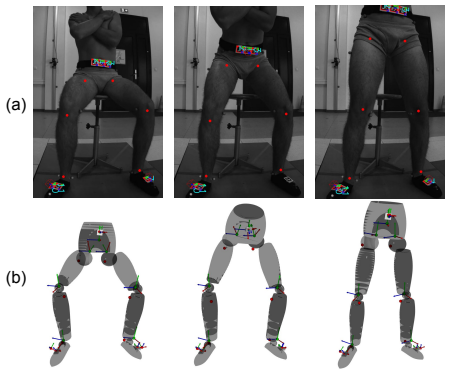

Fig. 1: (a) Multi-modal measurements shown in the camera frame, and (b) reconstructed 3D kinematics during sit-to-stand.

The standard for assessing reference 3D joint motion kinematics is Stereophotogrammetric Systems (SS), which are expensive and not easily portable. Consequently, many studies have turned to more affordable options like Inertial Measurement Units (IMU) [8], [2] or markerless pose estimation algorithms using RGB cameras [2]. IMUs, despite advanced filtering techniques, face issues with data drift and calibration inaccuracies, requiring one IMU per body segment [9], [10], [11]. Markerless pose estimation algorithms [12], [13] are user-friendly and calibration-free, but they lack the accuracy needed for rehabilitation assessments [3], despite their success in animation and gesture classification [14], [15].

Recent studies have suggested merging IMU and visual data to leverage the advantages of both approaches [16], [17]. Mallat et al. [16] introduced a cost-effective Visual Inertial Measurement Unit (VIMU) that combines IMUs with fiducial markers [18]. An extended Kalman Filter was proposed to fuse data from both sources, estimating upper limb kinematics during rehabilitation tasks [16]. The estimated joint angles were compared with those obtained from SS data and showed a Root Mean Square Error (RMSE) of 2.7 degrees, when the calibration offset was removed. However, this approach requires one VIMU per limb segment, and the

[1] LISSI, Univerité Paris-Est Créteil, Vitry-sur Seine, France
[2] NaturalPad, Montpellier, France
[3] LAAS-CNRS, Université Paul Sabatier, CNRS, Toulouse, France
[4] LBMC, Université Gustave Eiffel, IFFSTAR, Lyon, France
[5] School of Electrical Engineering, University of Belgrade, Belgrade, Serbia
[6] Image and Pervasive Access Laboratory (IPAL), CNRS-UMI, 2955, Singapore.

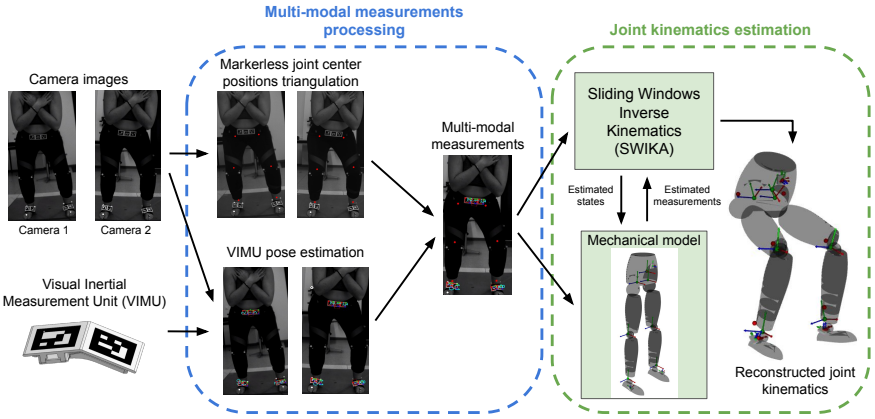

Fig. 2: Overview of the proposed affordable and multi-modal Sliding Windows Inverse Kinematics Algorithm (SWIKA).

use of markerless measurement modality was not investigated.

A study by Li et al. [19] introduced a VIMU and 2D markerless data-based sensor-to-segment calibration method for upper limb joint angle estimation. The method achieved less than 2deg and 10mm accuracy for VIMU sensor-to-segment transformation and a 6.6deg RMSE in joint angles without calibration offset removal compared to SS data. This approach, however, was limited to shoulder and elbow joints, required one VIMU per segment, and excluded markerless data in the Inverse Kinematics (IK) process.

Pearl et al. [2] and Cai et al. [20] also fused IMU and markerless data to estimate lower limb joint angles and positions, comparing results with SS data. Pearl et al. achieved less than 6deg RMSE in joint angles and below 50mm in joint center positions during gait, using anthropometric tables and SS data for model calibration, one IMU per segment, and 8 cameras for markerless triangulation, but limited their estimations to the sagittal plane. Cai et al. proposed a cost-effective system with only two cameras for triangulation in standing-up, turning, and walking tasks, achieving less than 3.5deg RMSE with calibration offset removal. However, their model calibration led to significant offsets, increasing RMSE by at least 12deg in joint angle estimation, and the study did not address joint center position accuracy. Both studies used one IMU per segment.

Previous studies typically employed one IMU per segment or simplified models. Only two studies [1], [8] explored joint kinematics estimation with fewer sensors. Mallat et al. [1] used three VIMUs on the pelvis and feet for lower limb kinematics during gait, achieving a 3.5 degrees RMSE with calibration offset removed. However, their study did not include markerless measurements, discuss RMSE without calibration, or address tasks involving significant knee flexions, crucial for accurate lower limb kinematics. Our previous study [8] employed two VIMUs on the trunk and hand, alongside two cameras, for upper limb joint angle estimation. They reported 3.4 degrees RMSE without calibration offset and 9.7 degrees with it. Their study, however, was limited to the right arm and low-amplitude tasks like picking and placing.

In this context, the present study explores the potential to estimate lower limbs kinematics during squat and sit-to-stand exercises using data gathered from a reduced number of VIMU and a markerless algorithm in conjunction with a 3D model of the locomotor system. Validated against SS measurements, the resulting RMSE on the joint angles and joint center positions estimated either from markerless data or from multiple modalities (i.e. markerless and VIMU data) are compared with respect to the literature.

II. METHODS

A. Mechanical model

A mechanical model of the human locomotor apparatus composed of $N_l = 7$ rigid segments articulated by $N_j = 18$ joints was defined. A total of $N_v = 3$ VIMUs were rigidly linked to pelvis, right foot and left foot. The International Society of Biomechanics recommendations were considered to define the orientation of the segment frames as well as the order of successive joint rotations [21]. The thigh segment was linked to pelvis through three successive revolute joints, the shank was linked to thigh through a revolute joint and the foot was linked to shank through two revolute joints. The pelvis position and orientation with respect to the camera frame R_c was defined through three prismatic and three revolute joints (see Figure 3). A wand based anatomical calibration procedure was conducted to determine the local position and orientation (pose) of the VIMU with respect to the segment frames, as well as the successive segment frames' relative pose [22], [1] (see Section IV). The vector of joint angles $\boldsymbol{\theta}$ (and three translations/rotations for the pelvis, not further mentioned for better readability) was fed to the Forward Kinematics Model (FKM) to estimate the 3D positions of joint centers $\hat{\mathbf{p}}_j^c$, and the 3D orientation $\hat{\mathbf{R}}_v^c$ and position $\hat{\mathbf{p}}_v^c$ of VIMU relatively to the camera frame as follows:

$$\left[\hat{\mathbf{p}}_v^c,\ \hat{\mathbf{R}}_v^c,\ \hat{\mathbf{p}}_j^c\right] = FKM(\boldsymbol{\theta}, \mathbf{P}) \tag{1}$$

where $\mathbf{P}$ was the vector of model parameters containing the segment lengths and local VIMU poses. The vectors of joint velocities $\dot{\boldsymbol{\theta}}$ and accelerations $\ddot{\boldsymbol{\theta}}$ were used in the first and second order differential models to calculate the 3D angular velocities $\hat{\boldsymbol{\omega}}_v^v$ and linear accelerations $\hat{\mathbf{a}}_v^v$, respectively:

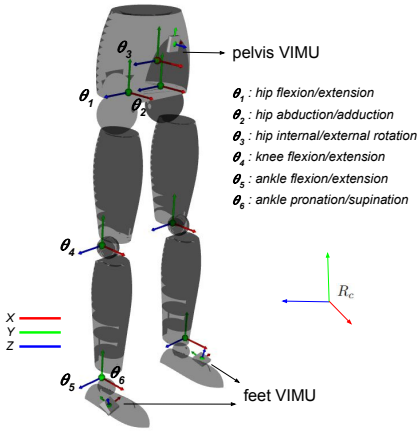

Fig. 3: Mechanical model of human locomotor apparatus and description of joint definition and technical frames.

$$\begin{aligned} \hat{\boldsymbol{\omega}}_v^v &= \mathbf{R}_v^{c^T} \mathbf{J}_r^c \dot{\boldsymbol{\theta}} + \mathbf{b}_\omega \\ \mathbf{a}_v^v &= \mathbf{R}_v^{c^T} (\mathbf{J}_p^c \ddot{\boldsymbol{\theta}} + \dot{\mathbf{J}}_p^c \dot{\boldsymbol{\theta}}) + \mathbf{b}_a \end{aligned} \tag{2}$$

where $\mathbf{b}_a$ and $\mathbf{b}_\omega$ were the acceleration and the gyroscope biases, respectively. $\mathbf{J}_p^c$ and $\mathbf{J}_r^c$ were the position and orientation Jacobian matrices expressed in the camera frame, respectively. The Pinocchio library was used to calculate the FKM and its derivatives [23]. The measurement function h was defined as follows:

$$h(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}, \ddot{\boldsymbol{\theta}}, \mathbf{b}_a, \mathbf{b}_\omega, \mathbf{P}) = \left[\hat{\mathbf{p}}_v^c,\ \hat{\mathbf{R}}_v^c,\ \hat{\mathbf{p}}_j^c,\ \hat{\mathbf{a}}_v^v,\ \hat{\boldsymbol{\omega}}_v^v\right] \tag{3}$$

B. Sensors measurements

The measured poses of the fiducial markers and joint center positions expressed in the camera frame as well as the linear accelerations and angular velocities of the IMUs were defined in a measurement matrix $\mathbf{y} = \left[\mathbf{p}_v^c,\ \mathbf{R}_v^c,\ \mathbf{p}_j^c,\ \mathbf{a}_v^v,\ \boldsymbol{\omega}_v^v\right]$. These measurements were collected by employing three VIMUs attached to the pelvis, left foot and right foot, each consisting of 0.036m fiducial markers affixed to affordable IMUs (*MPU6886, M5Stack MstickC-Plus*[1], *20€*), illustrated in Figures 4 and 5. A previously published self-calibration method was used to estimate the VIMU intrinsic parameters [16]. The 3D pose of each VIMU in the camera frame were determined utilizing the Aruco library [18], and the VIMUs linear accelerations and angular velocities were measured using their embedded IMUs frames. The BlazePose convolutional neural network architecture was used through the Mediapipe pipeline to estimate the 2D joint center positions of hips, knees and ankles, and finally a direct linear transformation triangulation[2] was employed to estimate their corresponding 3D coordinates in the camera frame [24].

C. Sliding Windows Inverse KInematics Algorithm (SWIKA)

To solve the IK, a Sliding Windows Inverse Kinematics Algorithm (SWIKA) was proposed. It had similar structure than the one presented in a previous work of our team [8].

[1]https://shop.m5stack.com/

[2]https://github.com/TemugeB/bodypose3d

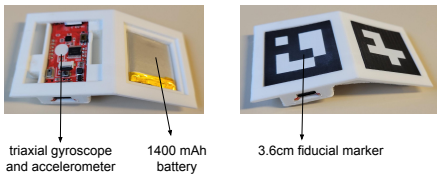

Fig. 4: Visual Inertial Measurement Unit (VIMU)

The algorithm takes into account the temporal evolution of data, allowing inverse kinematics to be solved over a windowed time horizon while effectively smoothing out outliers observed in the markerless data. Furthermore, continuity constraints on joint angles and velocities are considered by this method, setting it apart from classical multi-body inverse kinematics approaches that handle each time sample independently. For a given window, the state vector is composed by joint angles $\underline{\boldsymbol{\theta}}$ ($((N_j \times N_s) \times 1)$, velocities $\underline{\dot{\boldsymbol{\theta}}}$ ($(N_j \times N_s) \times 1$), and accelerations $\underline{\ddot{\boldsymbol{\theta}}}$ ($((N_j \times N_s) \times 1$). Additionally, accelerometer biases $\mathbf{b}_a$ ($((N_v \times 3) \times 1$) and gyroscope biases $\mathbf{b}_\omega$ ($((N_v \times 3) \times 1$) are included in the state vector as follows:

$$\mathbf{X} = [\underline{\boldsymbol{\theta}}, \underline{\dot{\boldsymbol{\theta}}}, \underline{\ddot{\boldsymbol{\theta}}}, \mathbf{b}_\omega, \mathbf{b}_a] \tag{4}$$

The window was set for a time horizon of $N_s = 10$ samples, and the following optimization problem was solved:

$$\min_{\hat{\mathbf{X}}} J(\hat{\mathbf{X}}) = \sum_{i=1}^{N_s} ||\mathbf{w}_i(h(\hat{\mathbf{X}}_i) \ominus \mathbf{y}_i)||^2 \tag{5a}$$

$$\text{s.t} \quad \theta_j^- \leq \theta_j \leq \theta_j^+,\ \forall j = 1, .., N_j \tag{5b}$$

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + T_s \dot{\boldsymbol{\theta}}_k,\ \forall k = 1, .., N_s \tag{5c}$$

$$\dot{\boldsymbol{\theta}}_{k+1} = \dot{\boldsymbol{\theta}}_k + T_s \ddot{\boldsymbol{\theta}}_k,\ \forall k = 1..N_s \tag{5d}$$

where $h(\hat{\mathbf{X}}_i)$, $\mathbf{y}_i$, $\mathbf{w}_i$ are the estimated measurements, sensor measurements, and positive definite diagonal weighting matrix, respectively, at i-th sample of the window. $\ominus$ is a retraction operator subtracting 3D cartesian measurements such as positions, angular velocities, linear accelerations, and more especially rotations using log operator [25]. $T_s = \frac{1}{50} s$ is the sampling time considered. $\boldsymbol{\theta}_j^-$ and $\boldsymbol{\theta}_j^+$ are the physiological lower and upper joint limits, represented by bound constraints (5b). Equations (5c) and (5d) correspond to the joint angles and velocities continuity constraint functions, respectively. The optimization problem was solved using an interior point method with the C++ Ipopt solver [26]. The cost function's gradient was calculated with automatic differentiation using CppAD library [27].

III. EXPERIMENTAL SETUP

Three healthy female and two healthy male participants (68 ± 9kg, 24 ± 2years, 1.71 ± 0.05m) were instructed to carry out three repetitions of squat and three repetitions of sit-to-stand tasks. Participants gave written informed consent prior to engaging in the experimental procedures. A reference SS was used to experimentally validate the proposed approach. Reflective markers were placed on 16 anatomical landmarks on both left and right side of the lower

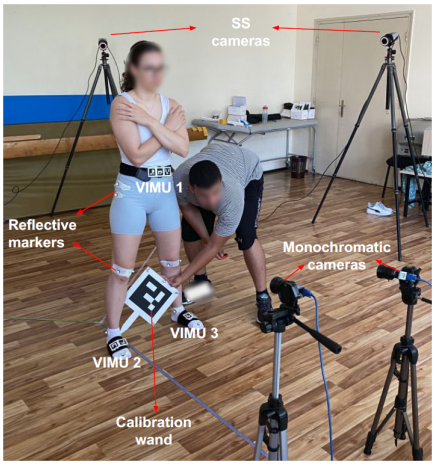

Fig. 5: Experimental setup and anatomical calibration.

limbs: anterior superior iliac spine, posterior superior iliac spine, medial/lateral femoral epicondyle, medial/lateral tibial malleolous, second metatarsal head, and heel. This marker template, the mechanical model described in section II-A and a classical multibody kinematics optimization were used to calculate the reference joint angles [28].

A. Cameras calibration

Two global shutter monochromatic cameras (*TheImagingSource DMK33UP2000*, 1920 x 1200 MJPEG, 50fps) were placed statically facing the subject (see Fig. 5). The intrinsic and extrinsic parameters of each camera were determined before the experiment through the use of 30 static chessboard poses and the calibration functions provided by OpenCV[3]. The resulting reprojection error of intrinsic and extrinsic calibration was below 0.3 pixels.

B. Biomechanical models calibration and alignment

A calibration wand, featuring four reflective markers and a 0.182m wide fiducial marker (Fig. 5), was utilized for anatomical calibration. Sphere fitting algorithms, fed by wand tip motions, estimated the wand tip's position relative to both fiducial and reflective marker frames. This yielded a residual error of 8.0×10^{-5}m for reflective markers and 1.0×10^{-3}m for the fiducial marker. The wand facilitated anatomical calibration by sequentially pointing to 16 anatomical landmarks [22], [1], determining segment lengths and constructing anatomical frames [21]. Reflective markers and VIMU measurements then ascertained their positions/poses relative to these anatomical frames.

IV. EXPERIMENTAL VALIDATION

A. Cartesian space comparison

SS data-derived joint center positions served as a reference to evaluate the accuracy of positions from raw markerless measurements, markerless SWIKA, and multi-modal SWIKA. The comparison involved estimating a transformation matrix between SS and camera frames using 3D positions of 16 anatomical points ([29], section III-B), yielding a RMSE of 4.0 ± 2.0mm for wand tip positions.

[3]https://opencv.org/

Table I compares joint center positions during sit-to-stand movements. RMSE and Pearson Correlation Coefficients (CC) were calculated for raw markerless, markerless SWIKA, and multi-modal SWIKA, with average RMSEs of 7.4 ± 1.5cm, 5.6 ± 1.2cm, and 3.0 ± 0.7cm, respectively, and corresponding CCs of 0.56, 0.59, and 0.71. Table II shows similar comparisons for squat movements, with average RMSEs of 6.4 ± 1.7cm, 5.0 ± 1.3cm, and 2.7 ± 0.5cm, and CCs of 0.55, 0.60, and 0.70, respectively.

Both markerless and multi-modal SWIKA demonstrate superior accuracy over raw markerless measurements in RMSE and CC, emphasizing SWIKA's effectiveness in joint center position estimation. Additionally, multi-modal SWIKA's consistent outperformance of markerless SWIKA underscores the benefits of integrating various data sources.

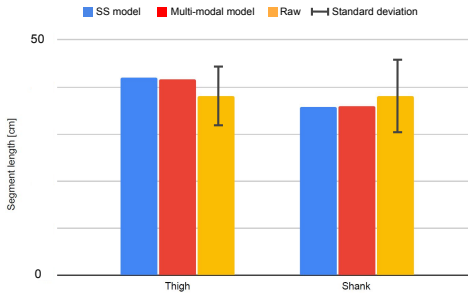

Fig. 6: Thigh and shank lengths from SS model calibration (blue), multi-modal calibration (red), and raw markerless data (yellow). Segment lengths from SS and multi-modal data were as constant due to the model calibration. Segment lengths estimated from raw markerless data were subject to variations, which were represented by a standard deviation bar.

B. Joint space comparison

Joint angles from SS data were used as a reference to compare with those derived from SWIKA using markerless and multi-modal data. Table III shows joint angle comparisons during sit-to-stand movements, calculating RMSE, RMSE without offset (correcting for orientation misalignments during subject-specific calibration), and CC for both markerless and multi-modal SWIKA. Markerless SWIKA showed an average RMSE of 13.1 ± 4.3deg and CC of 0.8, while multi-modal SWIKA had an average RMSE of 7.2 ± 2.3deg and CC of 0.72. Table IV compares squat movements, with markerless SWIKA having an average RMSE of 13.4 ± 6.3deg and CC of 0.87, and multi-modal SWIKA an average RMSE of 8.1 ± 2.9deg and CC of 0.63. These results highlight the multi-modal approach's superior accuracy and strong correlation (CC) with SS reference data. Note, markerless SWIKA could only measure sagittal plane joint angles, a topic discussed further in the next section.

V. DISCUSSION

This study aims to assess the accuracy of lower limbs joint angles and Cartesian position estimation, based on a markerless tracking algorithm, multi-modal measurements and a novel SWIKA approach, considering widely used

TABLE I: Comparison of joint center positions obtained from SWIKA based on markerless data, SWIKA based on multi-modal data, and raw markerless data, with respect to joint center positions obtained from SS data, during sit-to-stand.

-	RMSE [cm]			CC		
	Markerless SWIKA	Multi-modal SWIKA	Markerless raw	Markerless SWIKA	Multi-modal SWIKA	Markerless raw
Hip	3.6 ± 0.3	3.0 ± 0.3	7.8 ± 1.6	0.95	0.94	0.94
Knee	6.3 ± 1.4	3.9 ± 1.2	6.6 ± 1.0	0.51	0.64	0.67
Ankle	7.1 ± 1.9	2.2 ± 0.5	7.6 ± 2.0	0.32	0.53	0.09
Average	$\mathbf{5.6 \pm 1.2}$	$\mathbf{3.0 \pm 0.7}$	$\mathbf{7.4 \pm 1.5}$	**0.59**	**0.71**	**0.56**

TABLE II: Comparison of joint center positions obtained from SWIKA based on markerless data, SWIKA based on multi-modal data, and raw markerless data, with respect to joint center positions obtained from SS data, during squat.

-	RMSE [cm]			CC		
	Markerless SWIKA	Multi-modal SWIKA	Markerless raw	Markerless SWIKA	Multi-modal SWIKA	Markerless raw
Hip	3.4 ± 0.2	2.7 ± 0.4	6.7 ± 0.6	0.85	0.85	0.83
Knee	5.2 ± 1.6	3.1 ± 0.4	6.5 ± 1.9	0.73	0.79	0.79
Ankle	6.6 ± 2.0	2.4 ± 0.7	6.2 ± 2.6	0.18	0.46	0.05
Average	$\mathbf{5.0 \pm 1.3}$	$\mathbf{2.7 \pm 0.5}$	$\mathbf{6.4 \pm 1.7}$	**0.60**	**0.70**	**0.55**

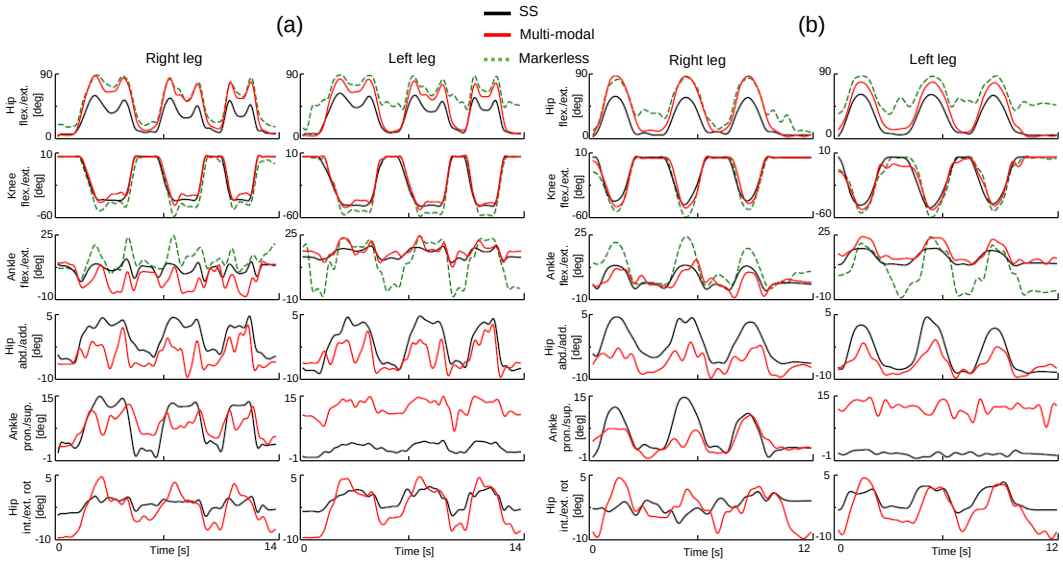

Fig. 7: Lower limbs joint angles estimated from the SS (black), from markerless SWIKA (dashed green), and from the multi-modal SWIKA (red), during a sit-to-stand trial (a) and a squat trial (b).

movements for lower limbs rehabilitation. The goals of the comparative analysis were twofold : investigate the benefits of using a SWIKA for joint angles and joint centers positions estimation, and investigate the advantages of using a multi-modal approach over a markerless based approach for the SWIKA.

The results presented in Tables I and II emphasize the superiority of joint center positions estimated with markerless-based SWIKA over those calculated from raw markerless measurements. This improvement can be attributed to the use of an accurately calibrated biomechanical model within SWIKA, enforcing constant segment lengths and removing poses that would exceed unfeasible joint angles. In contrast, raw markerless measurements often suffer from significant variations in segment lengths due to noise, as illustrated in Fig. 6.

Despite both markerless SWIKA and multi-modal SWIKA employing the same biomechanical model, the joint center positions estimated by multi-modal SWIKA demonstrate a significant increase in accuracy. This result emphasizes the effectiveness of integrating measurements from multiple sources, effectively mitigating the noise inherent to markerless measurements.

Tables III and IV demonstrate the superior performance of multi-modal SWIKA over markerless SWIKA in estimating joint angles. Markerless SWIKA is limited to flexion/extension angles in the sagittal plane due to insufficient information. Despite this, multi-modal SWIKA outperforms markerless SWIKA even in sagittal degrees of freedom, highlighting its effectiveness. Its ability to estimate angles in multiple degrees of freedom broadens its utility. Figures 6 and 7 further illustrate the enhanced accuracy of multi-modal SWIKA in various joint movements compared to raw markerless data, and markerless SWIKA.

VI. CONCLUSION

This study assessed the accuracy of estimating lower limb joint centers and angles during rehabilitation exercises using markerless, multi-modal measurements and SWIKA. It focused on two goals: evaluating the effectiveness of a model-based SWIKA and comparing multi-modal versus markerless methods. Results showed that markerless SWIKA, enhanced by a calibrated biomechanical model, more accurately determined joint centers than raw markerless

TABLE III: Comparison of joint angles obtained from SWIKA based on multi-modal and markerless data, with respect to joint angles obtained from SS data, during sit-to-stand. The RMSE without offset consists in removing the subject specific joint angles offset caused by calibration misalignment between SS and multi-modal system.

	RMSE [deg]		RMSE without offset [deg]		CC	
-	Markerless	Multi-modal	Markerless	Multi-modal	Markerless	Multi-modal
Hip flex./ext.	22.9 ± 7.4	13.5 ± 3.9	8.1 ± 1.8	9.3 ± 1.8	0.94	0.97
Hip abd./add.	-	4.4 ± 1.5	-	3.0 ± 0.9	-	0.6
Hip int./ext. rot.	-	7.5 ± 2.8	-	4.3 ± 0.9	-	0.8
Knee flex./ext.	6.8 ± 1.2	6.3 ± 1.4	6.0 ± 1.2	5.1 ± 1.1	0.97	0.97
Ankle flex./ext.	9.6 ± 4.0	6.0 ± 1.3	6.8 ± 3.9	4.1 ± 1.0	0.64	0.62
Ankle pron./sup.	-	11.7 ± 6.6	-	4.3 ± 2.6	-	0.54
Average	$\mathbf{13.1 \pm 4.3}$	$\mathbf{7.2 \pm 2.3}$	$\mathbf{6.9 \pm 2.3}$	$\mathbf{4.2 \pm 1.3}$	**0.8**	**0.72**

TABLE IV: Comparison of joint angles obtained from SWIKA based on multi-modal and markerless data, with respect to joint angles obtained from SS data, during squat.

	RMSE [deg]		RMSE without offset [deg]		CC	
-	Markerless	Multi-modal	Markerless	Multi-modal	Markerless	Multi-modal
Hip flex./ext.	20.1 ± 7.7	11.9 ± 3.9	10.1 ± 2.0	9.2 ± 2.4	0.93	0.97
Hip abd./add.	-	4.8 ± 1.6	-	3.4 ± 0.8	-	0.63
Hip int./ext. rot.	-	8.1 ± 2.1	-	4.8 ± 0.8	-	0.48
Knee flex./ext.	9.2 ± 1.7	7.2 ± 1.4	7.9 ± 1.4	6.6 ± 1.3	0.96	0.96
Ankle flex./ext.	10.9 ± 3.3	6.0 ± 1.3	6.8 ± 2.6	4.6 ± 0.8	0.77	0.67
Ankle pron./sup.	-	14.77 ± 7.9	-	6.6 ± 5.6	-	0.44
Average	$\mathbf{13.4 \pm 6.3}$	$\mathbf{8.1 \pm 2.9}$	$\mathbf{8.2 \pm 2.0}$	$\mathbf{5.2 \pm 1.8}$	**0.87**	**0.63**

methods. Additionally, multi-modal SWIKA was superior to markerless in both joint angle and center position accuracy, highlighting the value of integrating diverse data sources to reduce markerless measurement errors. Although this increased accuracy requires additional embedded sensors, using SWIKA reduces the sensor count to one VIMU per kinematic chain.

VII. Acknowledgements

We acknowledge the WILLOW team at ENS Paris for their support with the Pinocchio library and Jelena Aleksic, Kosta Jovanovic, and Dragan Mirkov from the University of Belgrade for their assistance in experimental validation.

References

[1] R. Mallat, V. Bonnet, R. Dumas, M. Adjel, G. Venture, M. Khalil, and S. Mohammed, "Sparse visual-inertial measurement units placement for gait kinematics assessment," *IEEE TNSRE*, 2021.

[2] O. Pearl, S. Shin, A. Godura, S. Bergbreiter, and E. Halilaj, "Fusion of video and inertial sensing data via dynamic optimization of a biomechanical model," *J. Biomechs*, 2023.

[3] J. Colombel, V. Bonnet, D. Daney, R. Dumas, A. Seilles, and F. Charpillet, "Physically consistent whole-body kinematics assessment based on an rgb-d sensor. application to simple rehabilitation exercises," *Sensors*, 2020.

[4] Q. Muaidi, "Rehabilitation of patellar tendinopathy," *JMNI*, 2020.

[5] A. Kerr, J. Dawson, C. Robertson, P. Rowe, and T. Quinn, "Sit to stand activity during stroke rehabilitation," *Top Stroke Rehabil*, 2017.

[6] N. Shimizu, H. Hashidate, T. Ota, and Y. Kawai, "Physical activity according to sit-to-stand, standing, and stand-to-sit abilities in subacute stroke with walking difficulty: a cross-sectional study," *Physiother Theory Pract*, 2022.

[7] J. Wall, C. Bell, S. Campbell, and J. Davis, "The timed get-up-and-go test revisited: Measurement of the component tasks," *JRRD*, 2000.

[8] M. Adjel, M. Sabbah, R. Dumas, N. Mansard, S. Mohammed, W. Bruno, and V. Bonnet, "Multi-modal upper limbs human motion estimation from a reduced set of affordable sensors," *IEEE IROS*, 2023.

[9] P. Picerno, P. Caliandro, C. Iacovelli, C. Simbolotti, M. Crabolu, D. Pani, G. Vannozzi, G. Reale, P. Rossini, L. Padua, and A. Cereatti, "Upper limb joint kinematics using wearable magnetic and inertial measurement units: an anatomical calibration procedure based on bony landmark identification," *Sci. Rep.*, 2019.

[10] L. Meng, M. Chen, B. Li, F. He, R. Xu, and D. Ming, "An inertial-based upper-limb motion assessment model: Performance validation across various motion tasks," *IEEE Sens. J.*, 2023.

[11] M. Nazarahari and H. Rouhani, "40 years of sensor fusion for orientation tracking via magnetic and inertial measurement units: Methods, lessons learned, and future challenges," *Inf Fusion*, 2021.

[12] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields," *arXiv:1812.08008 [cs]*, 2018.

[13] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for perceiving and processing reality," *IEEE CVPR*, 2019.

[14] Z. Li, J. Sedlar, J. Carpentier, I. Laptev, N. Mansard, and J. Sivic, "Estimating 3D motion and forces of person-object interactions from monocular video," *CVPR*, 2019.

[15] A. Singh, M. Adjel, V. Bonnet, R. Passama, and A. Cherubini, "A framework for recognizing industrial actions via joint angles," 2022.

[16] R. Mallat, V. Bonnet, M. Khalil, and S. Mohammed, "Upper limbs kinematics estimation using affordable visual-inertial sensors," *IEEE Trans. Autom. Sci.*, 2020.

[17] T. Li and H. Yu, "Visual–inertial fusion-based human pose estimation: A review," *IEEE T INSTRUM MEAS*, 2023.

[18] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Tracking fiducial markers with discriminative correlation filters," *Image Vis Comput*, 2021.

[19] T. Li and T. Dong, "Monocular camera-based online sensor-to-segment calibration for upper body pose estimation," *Sens Actuator A Phys*, 2023.

[20] S. Cai, M. Shao, M. Du, G. Bao, and B. Fan, "A binocular-camera-assisted sensor-to-segment alignment method for inertial sensor-based human gait analysis," *IEEE Sens. J.*, 2022.

[21] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D. D'Lima, L. Cristofolini, H. Witte, O. A. Schmid, and I. Stokes, "Isb recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion–part i: ankle, hip, and spine. international society of biomechanics," *J. Biomechs*, 2002.

[22] M. Bisi, R. Stagni, A. Caroselli, and A. Cappello, "Anatomical calibration for wearable motion capture systems: Video calibrated anatomical system technique," *Med. Eng. Phys.*, 2015.

[23] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, and N. Mansard, "The pinocchio c++ library – a fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives," *IEEE/SICE*, 2019.

[24] A. Andrew, "Multiple view geometry in computer vision . cambridge: Cambridge university press." 2001.

[25] J. Sola, "Course on slam," 2017.

[26] A. Wächter and L. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, 2006.

[27] B. Bell, "CppAD: a package for C++ algorithmic differentiation," *http://www.coin-or.org/CppAD*, 2021.

[28] T.-W. Lu and J. O'Connor, "Bone position estimation from skin marker co-ordinates using global optimisation with joint constraints," *J. Biomechs*, 1999.

[29] I. Söderkvist and P. Åke Wedin, "Determining the movements of the skeleton using well-configured markers," *J. Biomechs*, 1993.