Multi-modal Upper Limbs Human Motion Estimation From a Reduced Set of Affordable Sensors

Mohamed Adjel, Maxime Sabbah, Raphael Dumas, Nicolas Mansard, Samer Mohammed, Bruno Watier, Vincent Bonnet

Multi-modal Upper Limbs Human Motion Estimation From a Reduced Set of Affordable Sensors

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

***Abstract*— This study aims at developing a new affordable motion capture system for human upper limbs' joint kinematics estimation based on a reduced set of visual inertial measurement units coupled with a markerless skeleton tracking algorithm. The markerless skeleton tracking algorithm allows to alleviate the kinematic redundancy that is observed if only a single visual inertial measurement unit is used at the hand level but it introduces undesired outliers. A Sliding Window Inverse Kinematics Algoritm based on a biomechanical model is proposed to filter out outliers. It has the advantage to constrain the evolution of joint kinematics while being able to handle multi-modalities. The proposed system was validated with five healthy volunteers performing a popular rehabilitation pick and place task. Joint angles estimated using our method were compared with the ones obtained using a reference stereophotogrammetric system. The results showed an average root mean square error of 9.7deg along with an average correlation of 0.8. These results compare favorably with literature results obtained with more numerous and relatively costly sensors or more elaborated and expensive markerless systems.**

I. INTRODUCTION

Accurate estimation of human movement is necessary in numerous robotics and rehabilitation applications. For physical rehabilitation, a physiotherapist usually assess the patient achievement of the prescribed physical exercises. To do so, several clinical indexes such as the Frenchay arm test are used [1]. However, giving a quantitative assessment of a multi-segment motion from a visual assessment is not an easy task. Moroever, the inter/intra-clinician variability in the motion evaluation can be up to 9deg on average [2]. This assessment can be done automatically if a reliable joint angles estimate is available. When a human motion analysis system is to be used outside of the laboratory, its usability is as important as the minimum metrological performance required to ensure the validity of the results obtained. Nowadays, Stereophotogrammetric Systems (SS) are considered as the reference in human motion analysis. However, these systems are relatively expensive and non portable, limiting *de-facto* the natural observation of humans in real life situations.

The emergence of wearable sensors such as Inertial Measurement Unit (IMU), depth cameras and markerless skeleton tracking algorithm for RGB camera(s) has recently revived the research for personal monitoring outside the laboratory. Numerous affordable and easy-to-use systems based on IMU [3], RGB cameras [4], and RGB-D sensors [5] have been developed. Each of these systems has some drawbacks, but it is often possible to improve their accuracy by fusing these different modalities. In this context, the next section of this paper will focus on the literature related to IMU and markerless systems for human motion estimation.

[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] CNRS-AIST JRL (Joint Robotics Laboratory), IRL, Tsukuba, Japon
[6] Image and Pervasive Access Laboratory (IPAL), CNRS-UMI, 2955, Singapore.

II. RELATED WORK

Markerless approaches are not yet adopted by the clinical community as there is still a debate about the level of accuracy that can be achieved using only markerless and RGB camera(s) [6]. Indeed, several studies use only the Joint Center Positions (JCP) to assess the quality of human motion [7], [8], [6]. Nakano et al. [6] used 5 high-speed cameras and OpenPose [9] to compare JCP estimation with respect to SS. They showed that in the best condition the JCP can be estimated with a RMSE of 30mm but also that the accuracy dramatically decreases if the scene has object interacting with the subject or in case of subject partial occlusion. Beside the fact that JCP are not a clinical metric, it is difficult to understand how the joint angles, required to build clinical indexes, are impacted.

Better results are expected if markerless algorithms and model-based approaches were to be used together [8]. This was done by several studies using RGB-D skeleton tracking algorithms. In this case, the JCP estimates were shown to be not enough accurate for rehabilitation assessment due to segment length variations [10], [11]. Thus, our group, among others, proposed to use a multi-body inverse kinematics approach with data measured using a RGB-D camera to improve the joint angle estimates. It drastically reduced outliers thanks to fixed segment lengths and joint limits [12]. Nevertheless, even with a complex and cumbersome optimal tuning of the IK process, the Root Mean Square Error (RMSE) was often superior to 20deg for many tasks and joints. In a recent study, Lahkar et al. [13] used an expensive commercial markerless system with 10 high speed cameras together with a multi-body IK optimization process for studying the motions of a single boxer on a ring. Even in this ideal setup, they reported large RMSE going from 6.3deg for the shoulder up to 20deg for the wrist. They claimed that these errors were largely due to differences in the calibration of the models used for the markerless and the SS. Indeed, it is not possible to properly calibrate the model used in any markerless algorithms. We believe that part of

this error is also due to the fact that the temporal relationship between samples are not taken into account as most of the markerless skeleton tracking algorithm use only a single image to estimate the subject pose. Li et al. [8] proposed to use an optimisation approach to determine the kinematics but also the dynamics of several parkour tasks. Their original approach involved monitoring the JCP estimates obtained from the OpenPose algorithm. To reduce the outliers impact, they simultaneously minimized the dynamic biomechanical cost functions while incorporating temporal relations between the components of the state vector as constraints. Their approach resulted in a more reliable outcome.. Unfortunately, they reported solely JCP estimate compared to other markerless methods. In a survey [14], it is suggested that merging RGB-D and IMU data can improve joint angle estimates accuracy. Unfortunately, Feng et al. [15], who proposed to fuse JCP estimated from RGB-D camera with IMU measurement in a Kalman filter, did not assess the accuracy of upper limb joint kinematics. They only showed that fusing data reduces the uncertainty of hand position and acceleration estimates. Later on, for video gaming application it was proposed to use one IMU per segment and RGB-D data to reduce markerless outliers but again joint angle estimates were not reported [16].

Low-cost IMU, measuring 3D angular velocity and linear acceleration, are subjects to a large non-linear and low frequency drift that is due to manufacturing inaccuracies, temperature change or ageing of electronic components. This drift jeopardizes time integration of raw IMU signals. Adaptive filters and magnetometers can be used to reduce the drift effects but the latter are sensitive to ferromagnetic disturbances and thus the community in biomechanics tends to limit their use. Nevertheless, following several recent studies, it is possible to estimate joint angles of upper limbs with a multi-body kinematics model and one affordable IMU per investigated segments with an accuracy lower of 5deg [17], [18], [19]. Obviously, equipping a patient with one IMU per segment is very cumbersome and limit the usability of the system while increasing its costs.

To allow a more natural motion of the wearer, while solving the multiple possible IK solutions problem, this study proposes to use a method based on a reduced set of Visual Inertial Measurement Units (VIMUs) coupled with a markerless skeleton tracking algorithm. The use of VIMUs also offers the opportunity to perform an anatomical sensor to segment calibration for a better clinical interpretation of the joint angles.

III. METHODS

We proposed to fuse IMU data, from a reduced sensor set, with markerless data. Markerless data was used to solve the redundant IK problem but introduced rather noisy data. To reduce markerless outliers, a kinematic model of human upper limbs including anthropomorphic constraints was used and temporal relations between the state variables were taken into account with a Sliding Windows IK Algorithm (SWIKA). The overall method is described in Fig. 1.

A. Mechanical model

The mechanical model was composed of $N_L = 4$ rigid links articulated with $N_J = 13$ joints. The relative position of successive segment Coordinates Systems (CS), as well as the local VIMU position and orientation (pose) w.r.t to segment CS, were determined through a wand based anatomical calibration method [20], [21] (details about wand anatomical pointing accuracy are discussed in Section V). The successive segment CS of the kinematic chain as well as the order of successive rotations were defined following the International Society of Biomechanics recommendations [22]. The upper-arm segment was linked to trunk through three successive hinge joints, the lower-arm was linked to upper-arm through two successive hinge joints and the hand was linked to the lower-arm through two successive hinge joints. The trunk floating base with respect to the camera coordinate system R_c is defined through three prismatic and three hinge joints. The Forward Kinematics Model is used to calculate the position of joint centers $\hat{\mathbf{p}}_{\mathbf{j}}^{\mathbf{c}}$, the orientation $\hat{\mathbf{R}}_{\mathbf{v}}^{\mathbf{c}}$ and position $\hat{\mathbf{p}}_{\mathbf{v}}^{\mathbf{c}}$ of VIMU with respect to the camera coordinate system R_c as follows:

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

where $\boldsymbol{\theta}$ is the vector of joint angles and $\mathbf{P}$ is the vector containing local VIMU pose and segment lengths. The first and second differential models were used to estimate the 3D angular velocities $\hat{\boldsymbol{\omega}}_{\mathbf{v}}^{\mathbf{v}}$ and linear accelerations $\hat{\mathbf{a}}_{\mathbf{v}}^{\mathbf{v}}$ measured by the IMU, respectively:

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

where $\dot{\boldsymbol{\theta}}$ and $\ddot{\boldsymbol{\theta}}$ are the joint velocity and acceleration vectors, respectively. $\mathbf{J}_P$, $\mathbf{J}_R$, $\mathbf{b}_a$ and $\mathbf{b}_\omega$ are the positions and orientation Jacobian matrices and the acceleration and gyroscope bias, respectively.
The measurement function $\mathbf{h}$ was then defined as follows:

$$\mathbf{h} = \left[\hat{\mathbf{p}}_{\mathbf{v}}^{\mathbf{c}},\ \hat{\mathbf{q}}_{\mathbf{v}}^{\mathbf{c}},\ \hat{\mathbf{p}}_{\mathbf{j}}^{\mathbf{c}},\ \hat{\mathbf{a}}_{\mathbf{v}}^{\mathbf{v}},\ \hat{\boldsymbol{\omega}}_{\mathbf{v}}^{\mathbf{v}}\right] \tag{3}$$

The FKM and its derivatives were calculated with the Pinocchio library [23] that efficiently implements state-of-the-art rigid body algorithms for poly-articulated systems.

B. Affordable measurements

The vector of measurements $\mathbf{y} = \left[\mathbf{p}_{\mathbf{v}}^{\mathbf{c}},\ \mathbf{q}_{\mathbf{v}}^{\mathbf{c}},\ \mathbf{p}_{\mathbf{j}}^{\mathbf{c}},\ \mathbf{a}_{\mathbf{v}}^{\mathbf{v}},\ \boldsymbol{\omega}_{\mathbf{v}}^{\mathbf{v}}\right]$ was composed of the poses of the fiducal markers in the camera frame, the JCP from markerless algorithm in the camera frame and the acceleration and angular velocity of the IMUs in their own frames. These data were gathered using two VIMU attached to the trunk and the hand and composed of 3.6cm fiducial markers located onto affordable IMU (*MPU6886, M5Stack MstickC-Plus*[1], *20€*), see Fig.4. The intrinsic parameters of the VIMU were estimated using an already published self-calibration method [17]. The 3D

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

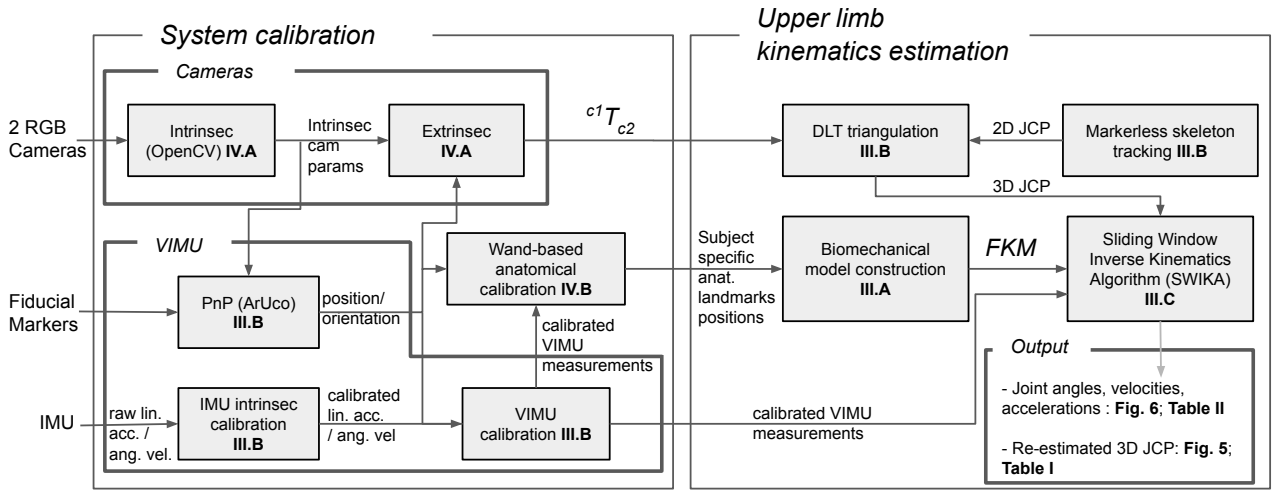

Fig. 1: Overview of the proposed affordable and multi-modal Sliding Windows IK Algorithm (SWIKA).

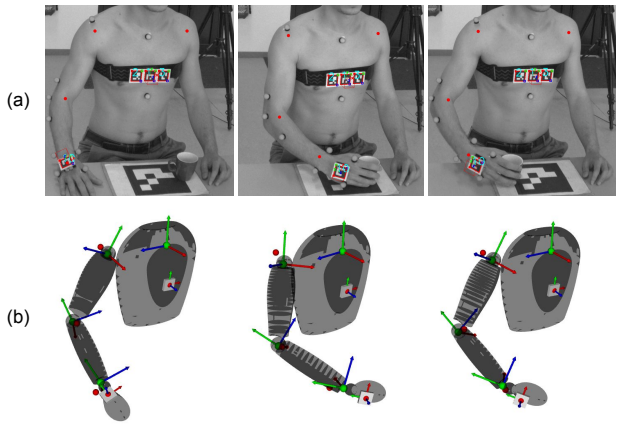

Fig. 2: VIMU and markerless measurements in the camera frame (a). Representation of the joint configuration using the proposed SWIKA during a pick and place task (b).

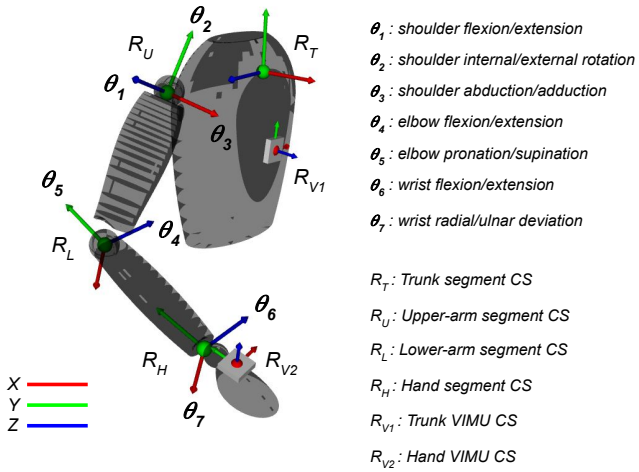

Fig. 3: Mechanical model of human upper limbs and description of joint definition and technical CS.

pose of each VIMU with respect to the camera frame was estimated using Aruco library [24] as shown in Fig. 2. VIMU angular velocities and linear accelerations were measured from the embedded IMU. The *so-called* markerless data, ie. the shoulder, elbow and wrist 3D JCP are estimated using Mediapipe pipeline [25]. The 2D estimate of JCP in both camera images were obtained using BlazePose convolutionnal neural network architecture. Then, the corresponding 3D JCP expressed in camera CS were obtained thanks to a triangulation[2] calculated with a direct linear transform [26]. Fig. 2.a represents, as red dots, the 2D JCP estimates from markerless data.

C. Sliding Windows Inverse Kinematics Algorithm (SWIKA)

The use of markerless skeletal tracker allows for alleviating the redundancy in solving the IK problem for the given mechanical model. However, this comes at the price of introducing noisy and discontinuous data in the IK process. To solve IK within such context, the SWIKA is introduced to take into account the whole time history of data evolution. Doing so, the frequent outliers observed in markerless data can be smoothed on a whole given trajectory. Moreover, a sliding window has the advantage of imposing continuity constraints for velocity and acceleration estimate unlike classical multi-body kinematics optimisation approach that are solving IK at a single sample. In addition, it seems relevant to consider velocities and accelerations due to use of IMU data. A sliding window was set for 1s of data of $N_S = 30$ samples with a sampling time interval $T_s = \frac{1}{30}$s. The state vector of the SWIKA was defined as follows:

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

where $\underline{\boldsymbol{\theta}}$, $\underline{\dot{\boldsymbol{\theta}}}$, $\underline{\ddot{\boldsymbol{\theta}}} \in \mathbb{R}^{N_J \times N_S}$ are the respective trajectories for joint configurations, velocities and accelerations on a whole window. $\mathbf{b}_\omega \in \mathbb{R}^{6\times 1}$ and $\mathbf{b}_\mathbf{a} \in \mathbb{R}^{6\times 1}$ are the bias associated to each axis of gyroscope and accelerometer sensors, respectively.

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

Hence, the formulation of the total optimisation problem boils down to determine $\mathbf{X}^*$ that allows to track the measurement vector $\mathbf{y}$ as follows:

$$\min_{\mathbf{X}^*} F(\mathbf{X}^*) = \|\mathbf{r}(\mathbf{X}^*)\|^2 \tag{5a}$$

$$\text{subject to} \quad \boldsymbol{\theta}_\mathbf{j}^- \leqslant \boldsymbol{\theta}_\mathbf{j} \leqslant \boldsymbol{\theta}_\mathbf{j}^+, \ \forall j = 1..N_J, \tag{5b}$$

$$\boldsymbol{\theta}_{\mathbf{t+1}} = \boldsymbol{\theta}_\mathbf{t} + T_s \dot{\boldsymbol{\theta}}_\mathbf{t}, \ \forall t = 1..N_S \tag{5c}$$

where $\mathbf{r}(\mathbf{X})$ is the function mapping the residuals between the measurement vector $\mathbf{y}$ and its estimate by the model $\mathbf{h}$ on the whole window as follows:

$$\mathbf{r}(\mathbf{X}) = \Omega(\mathbf{y} \ominus \mathbf{h}(\mathbf{X})) \tag{6}$$

Ω is the weight associated to the measurement (i.e. diagonal covariance matrix). $\ominus$ is a retraction operator that allowed to subtract 3D elements such as position, angular velocities and linear accelerations but also elements from the Lie algebra such as orientations using the log function [27], defined as :

$$\mathbf{y} \ominus \mathbf{h} = \begin{bmatrix} (\mathbf{p_v^c} - \hat{\mathbf{p}}_\mathbf{v}^\mathbf{c})^T \\ (\mathbf{p_j^c} - \hat{\mathbf{p}}_\mathbf{j}^\mathbf{c})^T \\ (log_3(\mathbf{R_v^c}^T \hat{\mathbf{R}}_\mathbf{v}^\mathbf{c}))^T \\ (\mathbf{a_v^v} - \hat{\mathbf{a}}_\mathbf{v}^\mathbf{v})^T \\ (\boldsymbol{\omega}_v^v - \hat{\boldsymbol{\omega}}_v^v)^T \end{bmatrix} \tag{7}$$

The physiological lower and upper joint limits $\boldsymbol{\theta}_j^-$ and $\boldsymbol{\theta}_j^+$ are enforced by constraints (5b). (5c) embody an Euler representation between some elements of the state vector and their derivatives. They ensure that the solution is dynamically consistent.

IV. EXPERIMENTAL SETUP

Three healthy male and two healthy female participants (72 ± 13kg, 25 ± 1years, 1.78 ± 0.02m) were asked to perform three repetitions of a pick and place task that is the core of several rehabilitation exercises for upper arm. As shown in Fig.5.a, the proposed approach was experimentally validated using a SS as a gold standard reference system. Reflective markers of the SS were placed on 10 anatomical landmarks : processus xiphoideus, acromion, incisura jugularis, C7 cervical vertebra, lateral epicondyle, medial epicondyle, radial styloid, ulnar styloid, the second metacarpal head and the fifth Metacarpal Head. Reference SS joint angles were calculated using this marker template, the mechanical model described in Section III.A and a classical extended Kalman filter [28].

A. Affordable visual system setup and alignment

Two rolling shutter cameras (*ELP-usbfhd08s*, 1920 x 1080 MJPEG, 30fps, 70€) were placed in a static pose in front of the subject. The intrinsic parameters, i.e. the projection matrix and the distortion parameters, of each camera were estimated prior the experiment using 50 chess-board static poses and OpenCV[3] camera calibration function. A reprojection error inferior of 0.8pixel for each camera calibration was obtained. The homogeneous transformation matrix from camera 1 to camera 2 CS ${}^{c1}T_{c2}$ was estimated using the fiducial markers positions, estimated from both camera images during the participant's pick and place motion, in an overdetermined system of equations [29] solved with a SVD. The resulting residual was of 6.0×10^{-3}m.

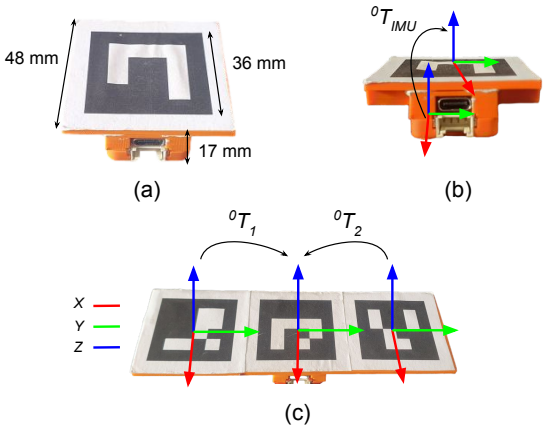

Fig. 4: Figure representing (a) the affordable Visual Inertial Measurement Unit (VIMU), (b) the transformation matrix from IMU CS to fiducial marker CS and (c) the transformation matrices from lateral fiducial markers to central one.

B. Biomechanical models calibration and alignment

A calibration wand was designed with a fiducial marker of 0.18m width, and three reflective markers as represented in Fig. 5.b. The position of the wand tip with respect to wand fiducial marker CS and reflective markers CS were estimated using a spherical motion of the wand, centered on the wand tip, and performing a sphere fitting. The resulting residual was of 1.0×10^{-4}m with the reflective markers and of 3.0×10^{-3}m with the fiducial marker. The 10 anatomical landmarks were pointed with the wand tip during an anatomical calibration phase [20], [17], constructing thus the segment anatomical CS [22] as well as their successive local positions. Doing so, the segments lengths were estimated. The local position of reflective markers as well as the local pose of VIMU with respect to the constructed segment anatomical CS were calculated using their respective measurements during the wand pointing phase.

V. EXPERIMENTAL VALIDATION

A. Cartesian space comparison

The raw 3D JCP estimated from the markerless skeleton tracking algorithm and the ones when using the SWIKA were compared to those obtained from the gold standard SS. This comparison required the determination of the homogeneous transformation matrix ${}^{SS}T_{cam}$ expressing the position and orientation of camera CS with respect to the SS one. We propose to use directly the 3D positions of the 10 pin-pointed

[3]https://opencv.org/

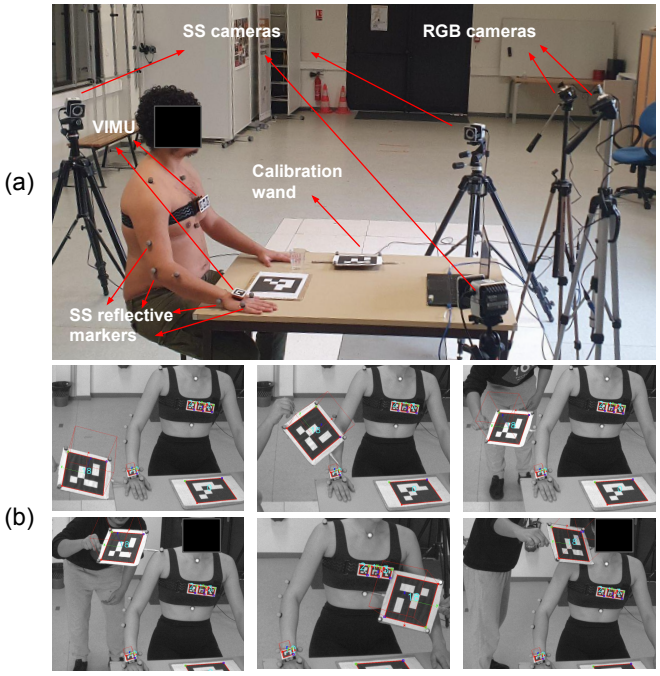

Fig. 5: Experimental setup (a) and anatomical calibration process performed with the wand (b).

TABLE I: Comparison of the JCP estimated from the markerless algorithm and from the proposed SWIKA approach with the JCP estimated with the reference SS.

-	RMSE [m]		PC	
	Markerless JCP	SWIKA JCP	Markerless JCP	SWIKA JCP
Shoulder	0.017 ± 0.01	0.013 ± 0.005	0.60 ± 0.23	0.50 ± 0.20
Elbow	0.024 ± 0.009	0.016 ± 0.007	0.83 ± 0.11	0.89 ± 0.05
Wrist	0.023 ± 0.007	0.019 ± 0.01	0.83 ± 0.11	0.93 ± 0.04
Mean	0.021 ± 0.009	0.016 ± 0.007	0.76 ± 0.15	0.77 ± 0.1

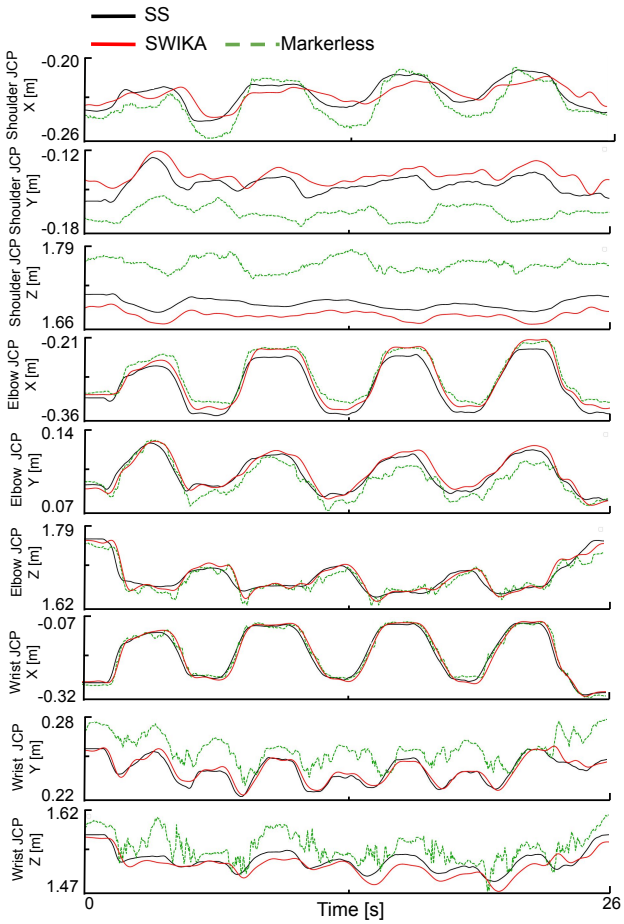

Fig. 6: Shoulder, Elbow and Wrist JCP estimated from the SS (black), from the markerless data (dashed green) and from the SWIKA with VIMU and markerless data (red)

anatomical landmarks expressed in both camera and SS CS to solve an overdetermined system of equations [29]. The estimated matrix resulted in a RMSE of 0.006 ± 0.004m between the SS wand tip positions and the fiducial marker wand tip positions expressed in SS CS. Using the identified ${}^{SS}T_{cam}$ matrix, the comparison of the markerless and SWIKA obtained JCP is presented in Table I. Surprisingly, the markerless JCP displayed a relatively low RMSE of 0.021 ± 0.009m and a Pearson Correlation coefficient (PC) of 0.76 ± 0.15. When using the SWIKA, the RMSE was 23% inferior but with a similar PC. Fig. 6 shows a typical comparison between JCP estimated with the SS, with the markerless algorithm and with the SWIKA.

B. Joint space comparison

The reference joint angles obtained from the SS were used as a reference to compare the joint angles obtained with the SWIKA based on different modalities. Modality 1 refer to the use of markerless data solely. Modality 2 refers to the use of VIMU measurements only and Modality 3 refers to the use of the SWIKA with both markerless and VIMU data. Table II shows that the joint angles estimated with VIMU only have the larger RMSE. This is most likely due to the fact that the IK on the 7 DoF arm is redundant when using only a VIMU at the end effector. Indeed, even within joint limits, there are several possible shoulder/elbow/wrist configurations for a given pose of the hand. Shoulder internal/external rotation and elbow pronation/supination are also impossible to compute when the arm is fully extended. The markerless algorithm shows a better RMSE but was not able to estimate elbow pronation/supination and wrist joint angles. Thus the corresponding average RMSE can not be fully compared with the Modality 3. Modality 3, as expected, displays the lowest RMSE and the significantly higher PC of 0.8 compared to other modalities. This is due to the fact that the markerless algorithm provides a relatively correct measurement of the elbow JCP, with an accuracy of 0.024 ± 0.009m. Thus, it can can be used to overcome the redundancy of the IK.

VI. CONCLUSIONS

This study introduces a multi-modal IK method based on an affordable and reduced set of sensors for a rehabilitation pick and place task. Using an anatomical calibration method and the proposed SWIKA, it was possible to estimate joint angles of the upper limbs with an average accuracy of

TABLE II: Joint angles comparison between the ones obtained from the reference SS and the ones calculated with the SWIKA based on different modalities.

-	RMSE [deg]			RMSE without offset [deg]			PC		
	Mod3	Mod2	Mod1	Mod3	Mod2	Mod1	Mod3	Mod2	Mod1
Shoulder flex./ext.	10.0	35.7	20.2	2.8	22.3	5.7	0.85	0.33	0.56
Shoulder int./ext. rot.	11.7	16.9	7.5	4.4	14.0	7.1	0.87	0.50	0.54
Shoulder abd./add.	4.9	37.9	10.9	2.0	28.2	3.8	0.89	0.31	0.78
elbow flex./ext.	6.0	8.9	6.7	3.7	4.4	5.9	0.80	0.76	0.62
elbow pro./sup.	9.4	29.3	-	3.4	26.5	-	0.98	0.58	-
wrist flex./ext.	18.3	20.7	-	4.2	16.6	-	0.98	0.63	-
wrist radial/ulnar dev.	7.3	14.58	-	3.4	10.7	-	0.57	0.50	-
Average	9.7	23.4	11.3	3.4	17.5	5.7	0.8	0.5	0.63

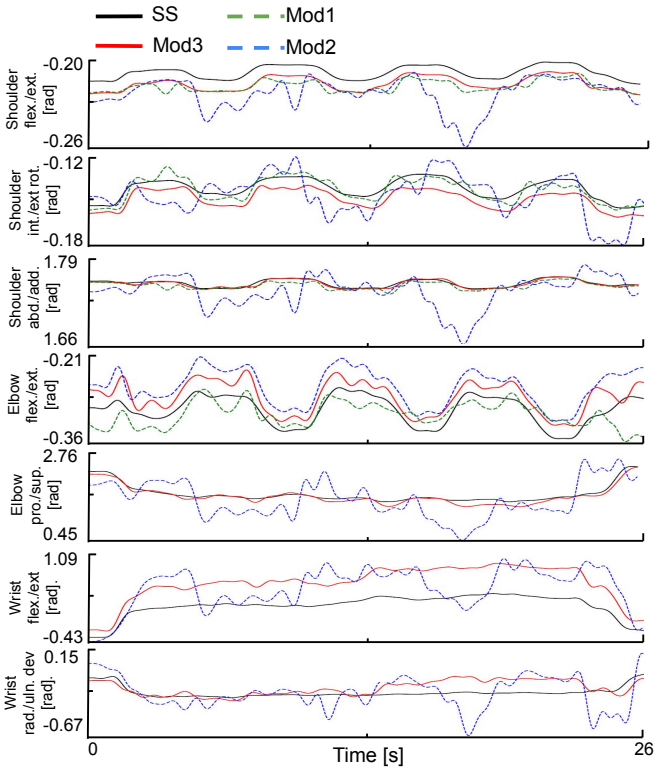

Fig. 7: Upper limb joint angles estimated from the SS (black), and the SWIKA using solely the markerless data (dashed green), the VIMU (dashed blue) and the VIMU and markerless data (red).

9.7deg including the calibration offset and 3.4deg when the calibration offset at the joint level was removed. This result is comparable to state-of-the art studies that were using one IMU per segments [17], [18] or very expensive and cumbersome multi-camera markerless algorithm [6]. The DoFs with the largest errors were the shoulder internal/external rotation and the wrist flexion/extension. In the literature reporting results from IMUs or markerless system, shoulder internal/external rotation is generally the joint angle with the least estimation accuracy. This is due to complexity of the shoulder joint composed of three bones and 2 joints leading to large shoulder instantaneous joint centre misplacement, indeterminacy when the arm is fully extended, and soft tissue artefact affecting both the orientation of sensors and position of reflective markers. Wrist flexion/extension may be specifically affected, in the present study, by errors in the pose of the fiducial marker placed on the hand. As represented in Fig.5 our approach preserves the natural motion of the participants and is easy to setup as only a single VIMU was located on the hand. In general it is commonly accepted that below 5deg of RMSE without the calibration offset the joint angle estimates are considered accurate enough for functional rehabilitation purpose [17]. The sensor to segment calibration method has a major impact on the results [30] and this step, of the proposed method, could be further improved including functional axes or the use of the JCP obtained by fusing VIMU and markerless data in the calibration process. Future works will typically focus on using functional calibration for improving the self-usability of patient for in home applications. We also would like to assess the proposed approach in the context of human-robot interaction and for more rehabilitation tasks as described in clinical context [1], [17].

VII. Acknowledgements

We thank the WILLOW team at ENS Paris for their help and their work for the community in integrating Casadi with Pinocchio.

References

[1] A. Heller, D. Wade, V. Wood, A. Sunderland, R. Hewer, and E. Ward, "Arm function after stroke - measurement and recovery over the 1st 3 months," *Journal of neurology, neurosurgery, and psychiatry*, pp. 714–9, 1987.

[2] C. Lavernia, M. D'Apuzzo, M. D. Rossi, and D. Lee, "Accuracy of knee range of motion assessment after total knee arthroplasty," *The Journal of Arthroplasty*, no. 6, Supplement, 2008.

[3] I. H. López-Nava and A. Muñoz-Meléndez, "Wearable inertial sensors for human motion analysis: A review," *IEEE Sensors Journal*, 2016.

[4] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, pp. 231–268, 2001.

[5] L. Chen, H. Wei, and J. Ferryman, "A survey of human motion analysis using depth imagery," *Pattern Recognition Letters*, 2013.

[6] N. Nakano, T. Sakura, K. Ueda, L. Omura, K. Arata, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3d markerless motion capture accuracy using openpose with multiple video cameras," *Frontiers in Sports and Active Living*, 2020.

[7] M. Naeemabadi, B. Dinesen, O. K. Andersen, S. Najafi, and J. Hansen, "Evaluating accuracy and usability of Microsoft Kinect sensors and wearable sensor for tele knee rehabilitation after knee operation," pp. 128–135, 2018.

[8] 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," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8632–8641, 2019.

[9] 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.

[10] A. Da Gama, P. Fallavollita, V. Teichrieb, and N. Navab, "Motor rehabilitation using Kinect: A systematic review," *Games for Health Journal*, pp. 123–135, 2015.

[11] K. Otte, B. Kayser, S. Mansow-Model, J. Verrel, F. Paul, A. U. Brandt, and T. Schmitz-Hübsch, "Accuracy and reliability of the Kinect version 2 for clinical measurement of motor function," *PLOS ONE*, 2016.

[12] 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*, p. 2848, 2020.

[13] B. Lahkar, A. Muller, R. Dumas, L. Reveret, and T. Robert, "Accuracy of a markerless motion capture system in estimating upper extremity kinematics during boxing," *Frontiers in Sports and Active Living*, p. 939980, 2022.

[14] C. Chen, R. Jafari, and N. Kehtarnavaz, "A survey of depth and inertial sensor fusion for human action recognition," *Multimedia Tools and Applications*, 2017.

[15] S. Feng and R. Murray-Smith, "Fusing Kinect sensor and inertial sensors with multi-rate Kalman filter," pp. 1–8, 2014.

[16] Y.-C. Du, C.-B. Shih, S.-C. Fan, H.-T. Lin, and P.-J. Chen, "An IMU-compensated skeletal tracking system using Kinect for the upper limb," *Microsystem Technologies*, pp. 4317–4327, 2018.

[17] R. Mallat, V. Bonnet, M. Khalil, and S. Mohammed, "Upper limbs kinematics estimation using affordable visual-inertial sensors," *IEEE Transactions on Automation Science and Engineering*, 2020.

[18] P. Slade, A. Habib, J. Hicks, and S. Delp, "An open-source and wearable system for measuring 3d human motion in real-time," 2021.

[19] T. Li and H. Yu, "Upper-body pose estimation using a visual-inertial sensor system with automatic sensorto-segment calibration," *IEEE Sensors Journal*, pp. 1–1, 2023.

[20] M. Bisi, R. Stagni, A. Caroselli, and A. Cappello, "Anatomical calibration for wearable motion capture systems: Video calibrated anatomical system technique," *Medical engineering physics*, 2015.

[21] 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 Transactions on Neural Systems and Rehabilitation Engineering*, 2021.

[22] G. Wu, F. van der Helm, D. Veeger, M. Makhsous, P. Roy, C. Anglin, J. Nagels, A. Karduna, K. McQuade, X. Wang, F. Werner, and B. Buchholz, "Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion - part ii: Shoulder, elbow, wrist and hand," *J. Biomechs*, 2005.

[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," 2019.

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

[25] 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," 2019.

[26] A. Andrew, "Multiple view geometry in computer vision . cambridge: Cambridge university press 2000." pp. 1333–1341, 2001.

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

[28] V. Fohanno, M. Begon, P. Lacouture, and F. Colloud, "Estimating joint kinematics of a whole body chain model with closed-loop constraints," *Multibody System Dynamics*, 2014.

[29] I. Söderkvist and P. Åke Wedin, "Determining the movements of the skeleton using well-configured markers," *J. Biomechs*, pp. 1473–1477, 1993.

[30] B. Bouvier, S. Duprey, L. Claudon, R. Dumas, and A. Savescu, "Upper limb kinematics using inertial and magnetic sensors: Comparison of sensor-to-segment calibrations," *Sensors (Basel, Switzerland)*, pp. 18 813–33, 2015.