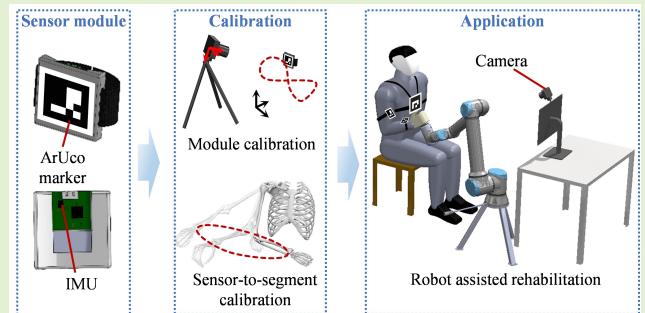


Upper Body Pose Estimation Using a Visual–Inertial Sensor System With Automatic Sensor-to-Segment Calibration

Tong Li^{ID}, Member, IEEE, and Haoyong Yu^{ID}, Senior Member, IEEE

Abstract—Upper body kinematics is essential for motor function assessment and robot-assisted rehabilitation training. Wearable sensor systems, such as inertial measurement units (IMUs), provide affordable solutions to replace laboratory-based motion capture systems for use in daily life. However, the sensor-to-segment calibration often relies on predefined posture or movements, which is hard to perform accurately, particularly for patients with a limited range of motion. A visual–inertial sensor system is presented, which includes three sensor modules attached to the trunk, upper arm, and forearm. Each module has an IMU and an ArUco marker, which can be captured by a camera and the driftless orientation of the modules is computed from visual–inertial fusion. The sensor-to-segment transformations are calibrated from a period of arbitrary arm movements in either a 2-D plane or 3-D space, simulating the training process assisted by end-effector robots. Experiments were conducted to validate the feasibility and evaluate the accuracy of the proposed method. The estimated shoulder and elbow joint angles correlated well (>0.986) with the ground truth from the optical motion capture (OMC) system. The joint angles presented low root-mean-square errors (RMSEs) ($<4^\circ$) except for the forearm pronation–supination angle (9.34°), which relied on manual alignment. The sensor system provides a simple and easy-to-use solution for movement assessment during robot-assisted training.



Index Terms—Kinematics, pose estimation, sensor-to-segment calibration, upper body, visual–inertial system.

I. INTRODUCTION

UPPER body kinematics measurement is essential to monitor human activities and is useful for many applications. For example, the measurement can be used to assess the motor function of stroke patients [1], [2] or as feedback information during robot-assisted rehabilitation to adjust the training process [3]. Virtual reality games employ the upper body pose as the user’s input to generate scenes in the game space [4], [5]. Human–robot collaboration and interaction may also use human movements as input for motion planning of

Manuscript received 8 January 2023; accepted 26 January 2023. Date of publication 3 February 2023; date of current version 14 March 2023. This work was supported by the Agency for Science, Technology and Research, Singapore, through the National Robotics Program under A*STAR SERC Grant 1922500045. The associate editor coordinating the review of this article and approving it for publication was Prof. Bin Gao. (Corresponding author: Haoyong Yu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of the National University of Singapore under Approval No. NUS-IRB-2022-216.

The authors are with the Department of Biomedical Engineering, National University of Singapore, Singapore 117583 (e-mail: tong.li@nus.edu.sg; biehyh@nus.edu.sg).

Digital Object Identifier 10.1109/JSEN.2023.3241084

robots to avoid unsafe collisions and achieve collaboration in industrial scenarios [6].

Unlike robotic systems where the kinematics of mechanical joints can be easily measured by installing sensors such as encoders at each joint, human movement measurement is much more challenging. Human joints usually include more than one degree of freedom (DoF) and noninvasive sensors can only be installed on the surface of the limb segments. The joint axis and joint center are difficult to identify from the surface [7], [8] and movements of the soft tissues (i.e., soft tissue artifact) further cause substantial errors in the skeletal pose estimation [9], [10], [11].

Vision-based systems are commonly used for human movement measurement. The optical motion capture (OMC) systems consisting of several high-speed cameras are widely used to capture the trajectories of passive or active markers that are attached to the human body for pose estimation. Such systems are highly expensive but can achieve submillimeter accuracy in 3-D space [12]. Pose estimation can also be achieved with multicamera systems to identify the human skeletal structure with machine learning-based approaches without markers [13], [14]. However, they usually also require

expensive high-speed and high-resolution cameras. Using a single red-green-blue (RGB) camera (e.g., webcam and phone camera) [15], [16] or RGB-depth (RGB-D) camera [17], [18] to detect body pose has been considered highly desirable for mobile and contactless capture. However, it suffers more significantly from occlusion and low accuracy issues, restraining the application to scenarios with high accuracy and robustness requirements.

Inertial measurement units (IMUs) have been extensively studied and widely applied in motion capture benefiting from their low cost and miniature size [19], [20]. Fusing data from accelerometers and gyroscopes can output precise inclination angles, and magnetometers are used to further estimate the 3-D orientation with driftless heading angles [21], [22]. However, the 3-D position cannot be directly computed from IMUs due to the rapid drifting in the double-integration process from accelerations. The magnetometers are also prone to be disturbed by surrounding magnetic or ferromagnetic materials [23], [24]. Therefore, applications of IMUs in close proximity to robotic devices usually can only rely on accelerometers and gyroscopes [25], [26]. Estimating human body pose and joint angles from inertial motion capture (IMC) systems attached to the segments also suffer from the aforementioned soft tissue artifact issue [27], [28]. This issue is more notable than markers attached to bony landmarks used in OMC systems since IMUs are usually attached to segments with muscles underneath.

A prerequisite of using IMUs to calculate body pose and joint angles is knowing the relative orientation of the sensor in the anatomical segment frame, i.e., the sensor-to-segment transformation [29], [30], [31]. IMUs can be carefully attached to the body segment to manually align the IMU frame or axis with the segment frame [29]. This manual alignment method causes large errors as segment frames are difficult to identify accurately with visual recognition [32]. Static poses have also been widely used for sensor-to-segment calibration [33], [34]. The accuracy of performing these poses is crucial for the calibration performance and using multiple predefined poses can improve the accuracy of alignment [35]. It is also possible to combine static posture with predetermined monoaxial joint rotations (i.e., functional movements) for calibration [36], [37]. Arbitrary movement-based calibration methods have also been explored to directly utilize the raw acceleration and angular velocity data from the IMUs [38], [39]. The joint axis and the joint center can be estimated from angular velocities and accelerations, respectively, using a period of arbitrary movements. Such methods are also used in OMC systems to identify the joint axis and joint center [40]. However, determining joint center locations from accelerations requires fast movement for calibration [41], which may not be suitable for the aged or patients. Calibrating anatomical landmarks with an additional IMU installed on a caliper-like device was proposed in [42] and only a static pose was required for calibration. However, this method requires assistance from a professional operator to conduct the calibration procedure.

In this article, we present a visual-inertial sensor system for estimation of the upper body pose to overcome the challenges in current purely visual or inertial systems. The system

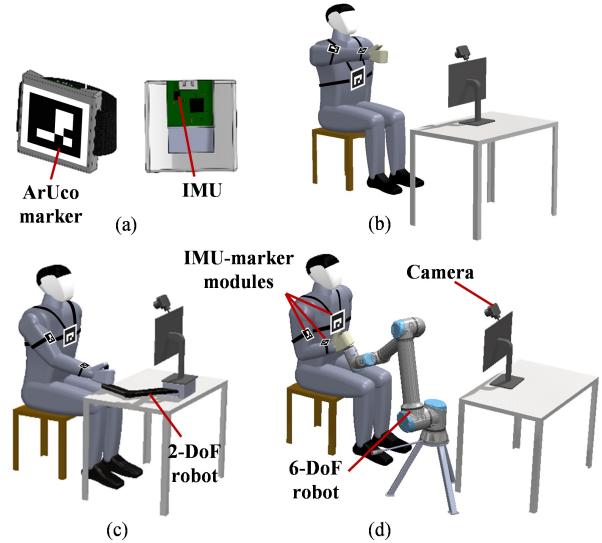


Fig. 1. Overview of the visual-inertial sensor system for upper body pose estimation. (a) Each sensor module consists of an ArUco marker and an IMU embedded underneath. (b)–(d) Application scenarios of free movements and movements led by end-effector robots in a 2-D plane or 3-D space.

includes a camera and three sensor modules each consisting of an IMU and an ArUco marker [43] [see Fig. 1(a)], similar to those proposed in [44] and [45]. The sensor modules are attached to the trunk, upper arm, and forearm, aiming at scenarios of either free movements or robot-assisted motions (see Fig. 1). The hand is not captured in the current system as it is usually fixed relative to the robot. The combination of visual and inertial sensors enables several important characteristics. The marker pose from vision can be fused with IMUs to remove the need for magnetometers, which is essential for applications close to assistive robots. The vision also provides additional position information compared to purely IMU-based systems, which can aid the calibration and estimation process. Anatomical segment frames may be calibrated by pointing to bony landmarks one by one with a wand as in [44], but this still requires assistance from extra operators. Instead, we propose an arbitrary movement-based calibration method to automatically align sensor frames to anatomical segment frames, which in practice can be performed voluntarily or led by robots during rehabilitation training. The feasibility and accuracy of the proposed method in upper body pose estimation are validated with experiments.

The contribution of this work includes several aspects. First, a visual-inertial sensor system for upper body pose estimation is presented and it can be integrated into current end-effector rehabilitation robots to acquire the body states during therapy as no magnetometers are used. Second, an automatic calibration method is proposed to compute sensor-to-segment transformation based on simple arbitrary joint movements, avoiding the use of predefined postures that are difficult to perform accurately, especially for patients. Third, the proposed method is validated in experimental tests and the sources of error are analyzed to guide future development and applications.

This article is organized as follows. Section II introduces the structure of the sensor system and methods of calibration and fusion. Section III presents the experiment setup and

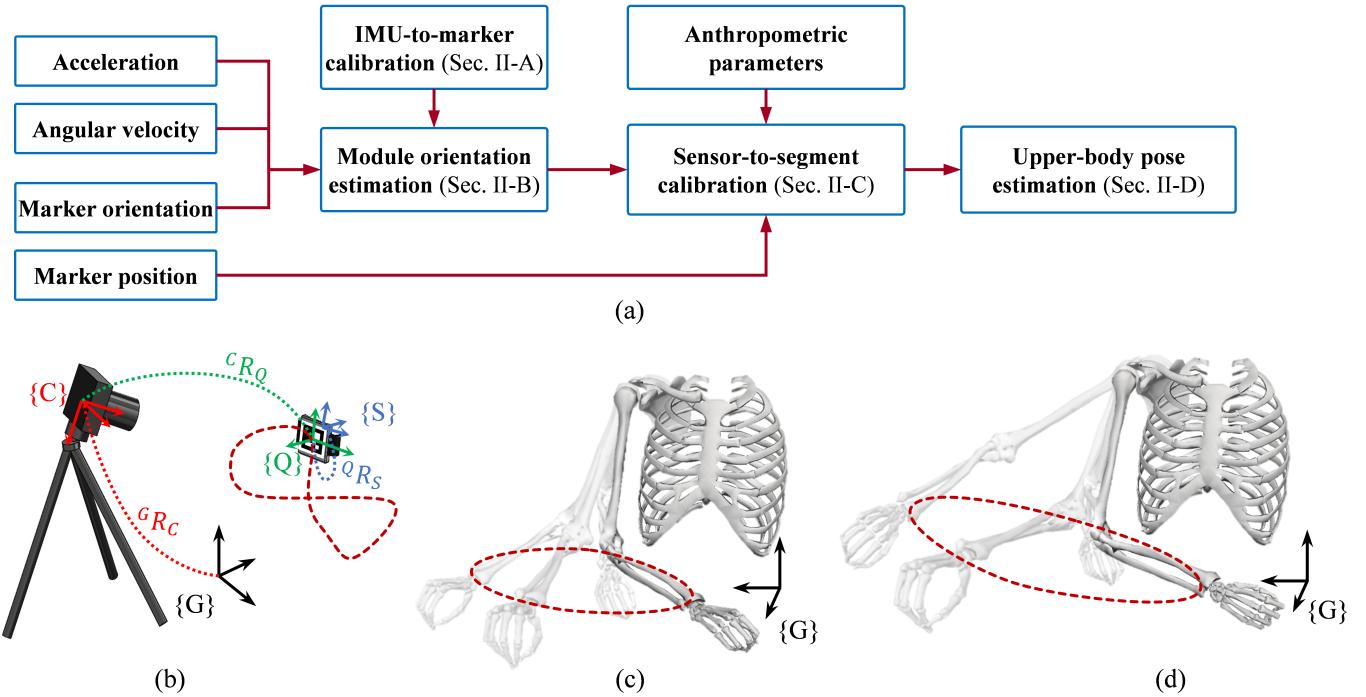


Fig. 2. Algorithm structure and illustration of the calibration process. (a) Data flow in the process of calibration and estimation. (b) IMU-to-marker transformation is calibrated with a single camera via dynamic movements. (c) and (d) Arm motion with the wrist/hand in a 2-D plane or 3-D space, respectively.

results. Discussions on the sensor system and sources of error are presented in Section IV. Finally, concluding remarks are conducted in Section V.

II. METHODS

The proposed framework for upper body pose estimation from visual-inertial sensors is shown in Fig. 2. The visual-inertial sensor modules are manually built with IMUs and printed markers and their relative orientation and position are calibrated first. The orientations of sensor modules are then estimated by fusing the inertial outputs and marker poses from the image processing. The module orientation and marker position are used to calibrate the sensor-to-segment transformation. The body pose and joint angles are then estimated via an extended Kalman filter (EKF) with the upper body skeletal model and poses of sensor modules.

A. IMU-to-Marker Calibration

The camera and IMUs can be obtained off the shelf and they are either precalibrated by manufacturers or manually calibrated with simple and mature methods available in [46] and [47]. The IMU-to-marker transformation (rotation matrix $\mathcal{Q}\mathbf{R}_S$ and position $\mathcal{Q}\mathbf{p}_S$) is calculated based on simple movements of the module in front of the camera [see Fig. 2(b)]. The marker location and orientation in the camera frame ($^C\mathbf{R}_Q$ and $^C\mathbf{p}_Q$) are detected from images and the velocity and acceleration of the marker ($^C\omega_Q$, $^C\dot{\mathbf{p}}_Q$, $^C\ddot{\mathbf{p}}_Q$, and $^C\ddot{\mathbf{p}}_Q$) are obtained via finite difference. The angular velocity and acceleration of the IMU are direct measurements ($^S\mathbf{a}$, $^S\omega$) expressed in the IMU local frame.

Consider a commonly used camera without built-in inertial sensors, and its orientation in the global frame (the vertical axis

coincides with the gravitational vector) is denoted as $^G\mathbf{R}_C$. The gravity vector ($^G\mathbf{g} = [0, 0, 1]^T$) expressed in the IMU frame should be

$$^S\mathbf{g} = \mathcal{Q}\mathbf{R}_S^{T,C}\mathbf{R}_Q^{T,G}\mathbf{R}_C^{T,G}\mathbf{g}. \quad (1)$$

For quasi-static movements, the acceleration of the IMU is close to gravitational acceleration ($^S\mathbf{a} \approx ^S\mathbf{g}$). Thus, a least-square optimization is used to find the transform between the marker and the IMU

$$\mathcal{Q}\mathbf{R}_S, ^G\mathbf{R}_C = \arg \min_{\mathcal{Q}\mathbf{R}_S, ^G\mathbf{R}_C} \int \left(\| ^S\mathbf{g} - ^S\mathbf{a} \|^2 \right). \quad (2)$$

Then, a dynamic trial is conducted to find the translational offset between the IMU and the marker. The acceleration of the IMU is expressed in the camera frame ($^C\ddot{\mathbf{p}}_S$) and derived from the detected marker motion

$$^C\ddot{\mathbf{p}}_S = ^C\dot{\omega}_Q \times (^C\mathbf{R}_Q \cdot \mathcal{Q}\mathbf{p}_S) + ^C\omega_Q \times (^C\omega_Q \times (^C\mathbf{R}_Q \cdot \mathcal{Q}\mathbf{p}_S)) + ^C\ddot{\mathbf{p}}_Q. \quad (3)$$

The position of the IMU in the marker frame ($\mathcal{Q}\mathbf{p}_S$) is then solved by minimizing the acceleration difference between IMU measurement and camera measurement

$$\mathcal{Q}\mathbf{p}_S = \arg \min_{\mathcal{Q}\mathbf{p}_S} \int \left(\| ^S\mathbf{a} - ^S\mathbf{g} - \mathcal{Q}\mathbf{R}_S^T \cdot ^C\mathbf{R}_Q^T \cdot ^C\ddot{\mathbf{p}}_S \|^2 \right). \quad (4)$$

B. Module Orientation Estimation

An EKF is designed to estimate the orientation of sensor modules separately. Measurements of each IMU are fused with the orientation of the corresponding marker obtained from image processing. The state of the EKF is the orientation

of the sensor module in quaternions in the global frame ($\mathbf{x}^S = [{}^G\mathbf{q}^S]$). The orientation of the sensor is calculated via the integration of angular velocity from the gyroscope (${}^S\omega_k = [\omega_x, \omega_y, \omega_z]$)

$${}^G\mathbf{q}_{k|k-1}^S = {}^G\mathbf{q}_{k-1|k-1}^S + \frac{1}{2} {}^G\mathbf{q}_{k-1|k-1}^S \otimes [0 \ {}^S\omega_k] \Delta_t \quad (5)$$

where ${}^G\mathbf{q}_{k-1|k-1}^S$ is the sensor orientation at the last time sample ($k - 1$) and Δ_t is the sampling interval of IMUs. Thus, the predicted state estimate is

$$\hat{\mathbf{x}}_{k|k-1}^S = \mathbf{F}_k^S \hat{\mathbf{x}}_{k-1|k-1}^S + \mathbf{w}_k^S \quad (6)$$

where \mathbf{w}_k^S is the process noise and \mathbf{F}_k^S is the state transition model

$$\mathbf{F}_k^S = \mathbf{I}_{4 \times 4} + \frac{1}{2} \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix} \Delta_t. \quad (7)$$

The acceleration measurement (${}^S\mathbf{a}$) and sensor orientation (${}^G\mathbf{R}_k^{S|C}$) from the camera are then set as the observations for filtering. Since the sampling frequency of the camera is usually lower than IMUs, the visual data are only fused at the next sample time to avoid interpolation. The acceleration measurements of the IMU are considered to be only from gravitational acceleration

$$[0 \ {}^S\mathbf{g}_k] = {}^G\mathbf{q}_{k|k-1}^S * \otimes [0 \ {}^G\mathbf{g}] \otimes {}^G\mathbf{q}_{k|k-1}^S. \quad (8)$$

The measurement residual is calculated as

$$\tilde{\mathbf{y}}_k = [{}^S\mathbf{a}_k / \|{}^S\mathbf{a}_k\| - {}^S\mathbf{g}_k, {}^G\mathbf{R}_k^{S|C} - \mathbf{R}({}^G\mathbf{q}_{k|k-1}^S)] \quad (9)$$

where $\mathbf{R}({}^G\mathbf{q}_{k|k-1}^S)$ is the rotation matrix of the IMU transformed from the quaternion. The rotation matrixes are reshaped to vectors in implementation.

Then, the states can be updated with the Kalman gain (\mathbf{K}_k) and covariance ($\mathbf{P}_{k|k-1}$) calculated as

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k \quad (10)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k) \quad (11)$$

$$\hat{\mathbf{x}}_{k|k}^S = \hat{\mathbf{x}}_{k|k-1}^S + \mathbf{K}_k \tilde{\mathbf{y}}_k \quad (12)$$

where \mathbf{Q}_k and \mathbf{R}_k are the covariance of the process noise and measurement noise, which can be determined with sensor datasheets and manual calibration. We used $9e^{-6}$ for the process noise and 0.1 and 0.5 for the measurement noise of accelerations and orientations in implementations, respectively.

C. Sensor-to-Segment Calibration

The sensor system includes three modules attached to the trunk, upper arm, and forearm. The coordinate frames of these segments are denoted as BTK, BUA, and BFA, respectively, with STK, SUA, and SFA denoting the corresponding sensor frame. The shoulder joint is considered a spherical joint with three DoFs, including abduction-adduction, flexion-extension, and internal-external rotation. The elbow joint is treated as a 2-DoF joint with flexion-extension and pronation-supination

TABLE I
VARIATIONS OF THE SENSOR-TO-SEGMENT TRANSFORMATIONS

Angle (°)	X	Y	Z
STK	0.48 (0.18)	1.30 (0.34)	0.54 (0.15)
SUA	1.38 (0.36)	6.84 (2.26)	1.07 (0.27)
SFA	0.79 (0.17)	0.76 (0.24)	0.74 (0.30)
SUA (w/comp)	1.32 (0.38)	0.09 (0.05)	1.04 (0.26)
Position (mm)	X	Y	Z
STK	0.76 (0.26)	1.19 (0.38)	0.86 (0.27)
SUA	5.16 (2.04)	2.64 (1.34)	1.92 (0.45)
SFA	0.42 (0.06)	0.72 (0.18)	1.43 (0.35)
SUA (w/comp)	1.78 (0.59)	2.07 (0.89)	1.92 (0.45)

Results are presented in the form of mean (SD).

movements. To calibrate the sensor-to-segment transformation, the key is to identify segment frames. With the orientation and position of sensor modules, the shoulder and elbow joint center and the elbow joint axes can be estimated from dynamic movements as in previous studies [38], [40]. Thus, the segment frame of the upper arm can be built up but not enough for the trunk and forearm. For the trunk, we added an assumption that the trunk is upright (zero flexion and lateral bending) with the same heading angle as the STK at the initial static stage to determine the segment frame of the trunk. For the forearm, we assume that the wrist flexion axis (usually defined based on the markers on the ulna styloid and the radial styloid in the OMC system) is parallel to the marker plane of the SFA. We also assume that the position of the wrist joint center in the sensor frame is known since this module is worn close to the wrist joint. Therefore, the segment frames are theoretically able to be built up with these measurements and assumptions.

The sensor-to-segment transformation is usually assumed to be constant by ignoring the soft tissue artifact. However, variations are inevitable during dynamic movements and may affect the automatic calibration process. Previous studies revealed that notable soft tissue artifact exists during the humeral axial rotation captured by either the OMC system [48] or the IMC system [36]. During the experimental tests (see Section III-A), it is verified that the variations are relatively small for the module on the trunk and forearm but rather notable for the module on the upper arm (see Table I), where the mean orientation variation is up to 6.84° and the position variation is up to 5.16 mm. Since the variations in orientation and position mostly occur around the longitudinal axis of the upper arm (see Fig. 3), similar to the compensation method in [36] and [48], we introduce an additional compensation angle (θ_{zc}) to the shoulder internal-external rotation. The relationship between the sensor and segment frames on the upper arm can be approximated by

$${}^G\mathbf{R}_{\text{SUA}} = {}^G\mathbf{R}_{\text{BUA}} \mathbf{R}_z(\theta_{zc})^{\text{BUA}} \mathbf{R}_{\text{SUA}} \quad (13)$$

$${}^G\mathbf{p}_{\text{SUA}} = {}^G\mathbf{p}_{\text{BUA}} + {}^G\mathbf{R}_{\text{BUA}} \mathbf{R}_z(\theta_{zc})^{\text{BUA}} \mathbf{p}_{\text{SUA}} \quad (14)$$

where $\mathbf{R}_z(\theta_{zc})$ is the rotation matrix of a rotation around the z -axis by θ_{zc} . By introducing this compensation angle θ_{zc} , the orientation and location of the SUA in the virtual

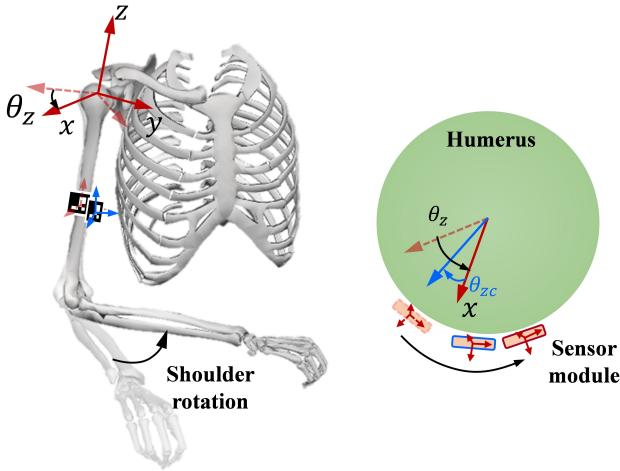


Fig. 3. Illustration of the variation of the transformation of the sensor module on the upper arm relative to the upper arm anatomical frame. When the shoulder has an internal rotation angle of θ_z , the sensor module will not follow the same motion but with an offset of θ_{zc} .

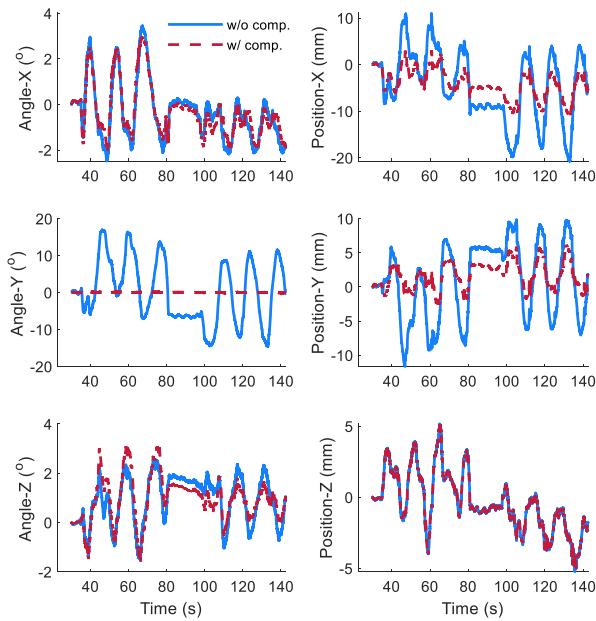


Fig. 4. Orientation and position variations of SUA in the upper arm anatomical frame with and without compensation.

segment frame (after an additional rotation of θ_{zc}) shows a much smaller variation (see Fig. 4).

Unlike the sensor-to-segment calibration based on predefined body poses, joint angles are unknown for arbitrary dynamic calibration movements. Therefore, a bilevel nonlinear optimization process is used to identify the sensor-to-segment transformation. These parameters are set as the decision variables of the upper level optimization. Given the guess of these parameters, a lower level optimization process (i.e., an inverse kinematics process) is performed to find the optimal joint angles to minimize measurement errors (sensor orientation and position). The computation methods of the calibration process are detailed as follows.

With the assumptions on the initial trunk pose, the relative orientation of STK in BTK (${}^{BTK}R_{STK}$) can be computed with

the sensor orientation at the static phase. Thus, the remaining parameters to identify (x_{ID}) include the position of three sensor modules in the corresponding segment frame (${}^{BTK}p_{STK}$, ${}^{BUA}p_{SUA}$, and ${}^{BFA}p_{SFA}$), and the orientation of SUA and SFA in the corresponding segment frame (${}^{BUA}R_{SUA}$ and ${}^{BFA}R_{SFA}$ expressed with Euler angles ${}^{BUA}\theta_{SUA}$ and ${}^{BFA}\theta_{SFA}$)

$$x_{ID} = [{}^{BTK}p_{STK}, {}^{BUA}p_{SUA}, {}^{BFA}p_{SFA}, {}^{BTK}\theta_{STK}, {}^{BFA}\theta_{SFA}]. \quad (15)$$

To reduce the number of decision variables and the computation cost of the lower level optimization, we consider the measurement error of the sensor module on the trunk to be negligible. Thus the trunk position and orientation are directly computed from the sensor module and guess of relative orientation and position in each iteration. The compensation angle (θ_{zc}) for the upper arm is also set as an additional variable to calculate the pose of the SUA using (13) and (14). Therefore, the decision variable of the inverse kinematics process is set as

$$x_{IK} = [\theta_{SD}, \theta_{EB}, \theta_{zc}] \quad (16)$$

where θ_{SD} and θ_{EB} are the shoulder and elbow joint angles, respectively. Given the guess of x_{ID} and x_{IK} , the orientation and position of SUA and SFA can be calculated via forward kinematics. The difference between the values from forward kinematics and measurements are set as the objective function to minimize

$$x_{IK} = \arg \min_{x_{IK}} (e_{R_{QUA}} + e_{R_{QFA}} + e_{p_{QUA}} + e_{p_{QFA}}) \quad (17)$$

where $e_{R_{QUA}}$ and $e_{R_{QFA}}$ are the orientation error calculated from the quaternion angle and $e_{p_{QUA}}$ and $e_{p_{QFA}}$ are the position error calculated as the distance.

For the upper level optimization, with the joint angles from the inverse kinematics process, the objective function is computed as the integral of the error in the calibration stage

$$\begin{aligned} x_{ID} = \arg \min_{x_{ID}} & \int (e_{R_{QUA}} + e_{R_{QFA}} + e_{p_{QUA}} \\ & + e_{p_{QFA}} + e_{p_{BFA}} + \theta_w) \end{aligned} \quad (18)$$

where $e_{p_{BFA}} = \|{}^G\hat{P}_{BFA} - {}^Gp_{BFA}\|^2$ is the error computed as the difference between the position of the wrist joint center from forward kinematics (${}^Gp_{BFA}$) and the value estimated from the SFA (${}^G\hat{P}_{BFA}$) via manual measurement. θ_w is the angle difference between the wrist flexion axis and the marker plane of SFA, which is assumed to be zero with the proper donning process.

D. Upper Body Pose Estimation

Once the sensor-to-segment transformation is calibrated, the model states can be estimated from the measurements at each time frame. The joint angles are usually obtained via inverse kinematics at each data frame when using a single modal of sensors. However, it is hard to directly apply as the sampling rates from the IMUs and the camera are different and the large noise from visual data may result in unsMOOTH curves.

Therefore, we use an EKF to estimate joint angles by fusing the sensor orientation and position data. The states of the EKF are set as $\mathbf{x}^C = [p_{BTK}, \theta_{BTK}, \theta_{SD}, \theta_{EB}, \theta_{zC}]$, where p_{BTK} and θ_{BTK} are the position and orientation (in Euler angles) of the trunk segment frame, respectively. The process model to predict the prior states ($\mathbf{x}_{k|k-1}^C$) from the states of the last time sample ($\mathbf{x}_{k-1|k-1}^C$) is set as

$$\mathbf{x}_{k|k-1}^C = \mathbf{I}^C \mathbf{x}_{k-1|k-1}^C + \mathbf{w}_k^C \quad (19)$$

where \mathbf{I}^C is the identity matrix and \mathbf{w}_k^C is the process noise.

The observations include all the fused orientations of the sensor modules and the position of the sensor modules. Similarly, the position data are only fused whenever available

$$= \mathbf{y}_k = \left[{}^G \mathbf{R}_{STK}, {}^G \mathbf{R}_{SUA}, {}^G \mathbf{R}_{SFA}, {}^G \mathbf{p}_{STK}, {}^G \mathbf{p}_{SUA}, {}^G \mathbf{p}_{SFA} \right]. \quad (20)$$

The states are predicted and updated similar to the fusion process of a single module.

III. EXPERIMENTS AND RESULTS

A. Experiment Setup

The sensor system designed for experiments includes three sensor modules with IMUs (WIT Motion, China) all connected to a Teensy board (version 4.0) for data collection. The Teensy board and a ZED2 camera (Stereolabs Inc., San Francisco, CA, USA) were both connected to a laptop via USB cables, and a customized Python program was used to acquire and log the inertial and visual data simultaneously. The acceleration and angular velocities were collected at 100 Hz, while the camera was recorded at 30 Hz, 1080 P. The accelerometers were calibrated following the methods in [47] and the raw data were corrected with the scale and misalignment factors from the calibration process.

To evaluate the accuracy of the sensor system, an OMC system (VICON, Oxford, U.K.) was set around the subjects to capture and compute the body motion as the ground truth [see Fig. 5(a)]. A total of 17 reflective spherical markers were attached to the upper body and sensor modules. There were four markers on the upper trunk at the left and right acromion, sternum jugular notch, and cervical vertebrae 7; two markers on the elbow joint at the medial epicondyle and lateral epicondyle; and two markers on the forearm at the ulna styloid and radius styloid. Besides, three markers were attached to each sensor module. To align the coordinate frames of the OMC system and sensor system, an additional ArUco marker plate with three reflective markers was set in the field of view of the camera [ground marker plate in Fig. 5(a)], which was only used for testing and is not needed in real applications of the system. The location of the reflective markers and ground marker plate and sensor modules was precalibrated before experiments on subjects.

We recruited nine subjects (age = 27.6 ± 2.6 years, weight = 69.1 ± 10.4 kg, height = 1.72 ± 0.06 m, males = 8, and females = 1) for the experiments and the study was approved by the Institutional Review Board (IRB) of the National University of Singapore (No. NUS-IRB-2022-216).

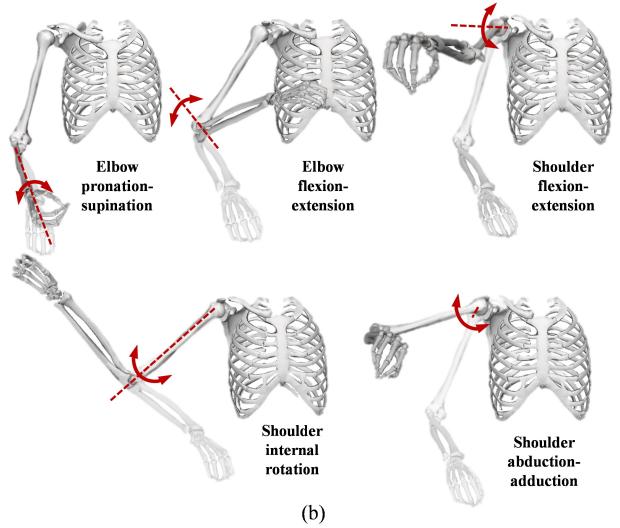
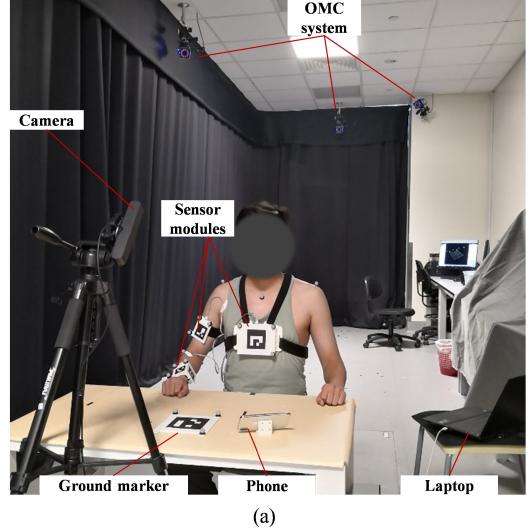


Fig. 5. Experimental setup and movements during the testing. (a) Experiment setup. (b) Testing movements by exciting each DoF of the shoulder and elbow joints.

The subjects were shown the detailed procedure of the experiments and gave written consent before the experiment. The experiment included a static stage, calibration stage, and testing stage. After donning the sensor system and marker placement, the location of reflective markers on the radius styloid and ulna styloid in the marker plate of SFA was manually measured and then averaged as the location of the wrist joint center. The subject started at a neutral upright trunk pose with the arm placed on the table [see Fig. 5(a)] for a static phase of 10 s. Then, in the calibration stage, the subject first moved the arm with the hand going through a planar 2-D elliptical trajectory [see Fig. 2(c)] for three cycles with the forearm approximately horizontal. Then, the subject moved the arm with the hand going through a 3-D elliptical trajectory [see Fig. 2(d)] for three cycles. The elliptical trajectory had a length of around 45 and 25 cm for the major and minor axes, respectively. To control the speed of the movement, a video where a red point is moving along an elliptical trajectory at a cycle period of 15 s was shown to the subject to follow via the

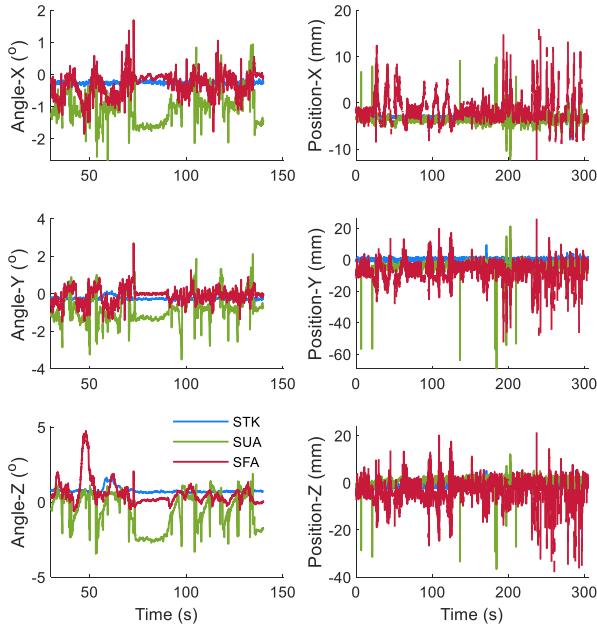


Fig. 6. Orientation and position errors of the sensor modules in a representative trial.

phone on the table [see Fig. 5(a)]. Then, in the testing stage, the subjects executed joint movements to excite each DoF of the elbow and shoulder joints in turn as shown in Fig. 5(b), and each movement was repeated three times as well.

B. Motion Capture Data Process

The trajectory data from the OMC system were used to compute the ground truth kinematics of the sensor modules and upper body. The orientation and location of the sensor modules were obtained from three reflective markers on each plate as ground truth for comparison. For the upper body, the joint center is required to build up the anatomical frame and the segment parameters will also be used in further calibration and fusion process. Although joint center position can be estimated from markers at a static frame, functional movement-based methods using least-square optimization can give better accuracy [49]. We used trajectories of the reflective markers on the sensor modules during the 3-D calibration movements to identify the location of the shoulder joint and elbow joint center similar to the method in [40]. The wrist joint center was determined as the midpoint between the markers on the ulna styloid and radius styloid. The anatomical segment frames were built at the static pose and the location of the reflective markers in the segment frames was computed. The joint angles were then obtained by solving an inverse kinematics problem using all the marker trajectories [50].

C. Single Module Accuracy

The orientation of each sensor module was first obtained by fusing the inertial data and marker orientation from images. The estimated orientation and position of the sensor module coincide well with the ground truth from the OMC system showing small errors (see Fig. 6). As shown in Table II,

TABLE II
ESTIMATION ERROR OF THE ORIENTATION AND POSITION
OF SENSOR MODULES

Error	STK	SUA	SFA
Roll angle (°)	0.26 (0.12)	1.03 (0.44)	0.49 (0.19)
Pitch angle (°)	0.23 (0.22)	0.79 (0.51)	0.53 (0.14)
Yaw angle (°)	0.35 (0.26)	1.45 (0.64)	1.01 (0.32)
Orientation (°)	0.55 (0.27)	1.81 (0.76)	1.22 (0.26)
Position (mm)	5.81 (2.03)	7.90 (3.65)	9.97 (1.10)

Results are presented in the form of mean (standard deviation).

the root-mean-square error (RMSE) of orientation (calculated from quaternion angle) is 0.55° – 1.81° . The orientation error expressed in Euler angles shows a relatively larger error in the yaw angle (0.35° – 1.45°) than the roll and pitch angle (0.23° – 0.53°). The position is estimated from the camera and the error is in the range of 5.81–9.97 mm. The orientation and position errors of the sensor module on the trunk show the smallest error compared to the other two modules, possibly due to its larger marker size.

D. Upper Body Pose Estimation

The sensor-to-segment transformation was identified with the orientation and position of sensor modules during the calibration stage, which included three cycles of arm motion lasting 45 s in total. The orientation data have a sampling frequency of 100 Hz, the same as the IMU raw data, and we interpolated the Euler angles of the sensor modules to the timing of the position data that have a frequency of 30 Hz, the same as video recording. The orientation and position data were further resampled to a frequency of 0.5 Hz to reduce the number of samples used in the lower level optimization. The samples with incomplete detection of sensor positions (i.e., at least one marker on the sensors not detected) were also discarded. As a result, the identification process used around 18–22 samples in the trials. The calibration process and estimation process were performed separately using the 3-D and 2-D calibration movements. For comparison, we also conducted the calibration process without compensating for the variation of sensor-to-segment transformation of SUA by setting θ_{zc} in (16) to be zero.

The upper body pose was then estimated from the EKF after the sensor-to-segment transformation had been determined. The estimation process was repeated with five sets of sensor-to-segment transformations, including the transformations computed at the initial static stage (SS-true) and the transformations solved using the 3-D and 2-D calibration movements, with (SS-3-D and SS-2-D) and without (SS-3-D-NC and SS-2-D-NC) compensation (θ_{zc}).

The joint angles and estimation errors in a representative trial are shown in Fig. 7. The estimated joint angles using true locations (SS-true) coincide well with the joint angles from the inverse kinematics process and the results using identified transformations (SS-3-D) are also close to the ground truth, with an offset mainly observed in forearm

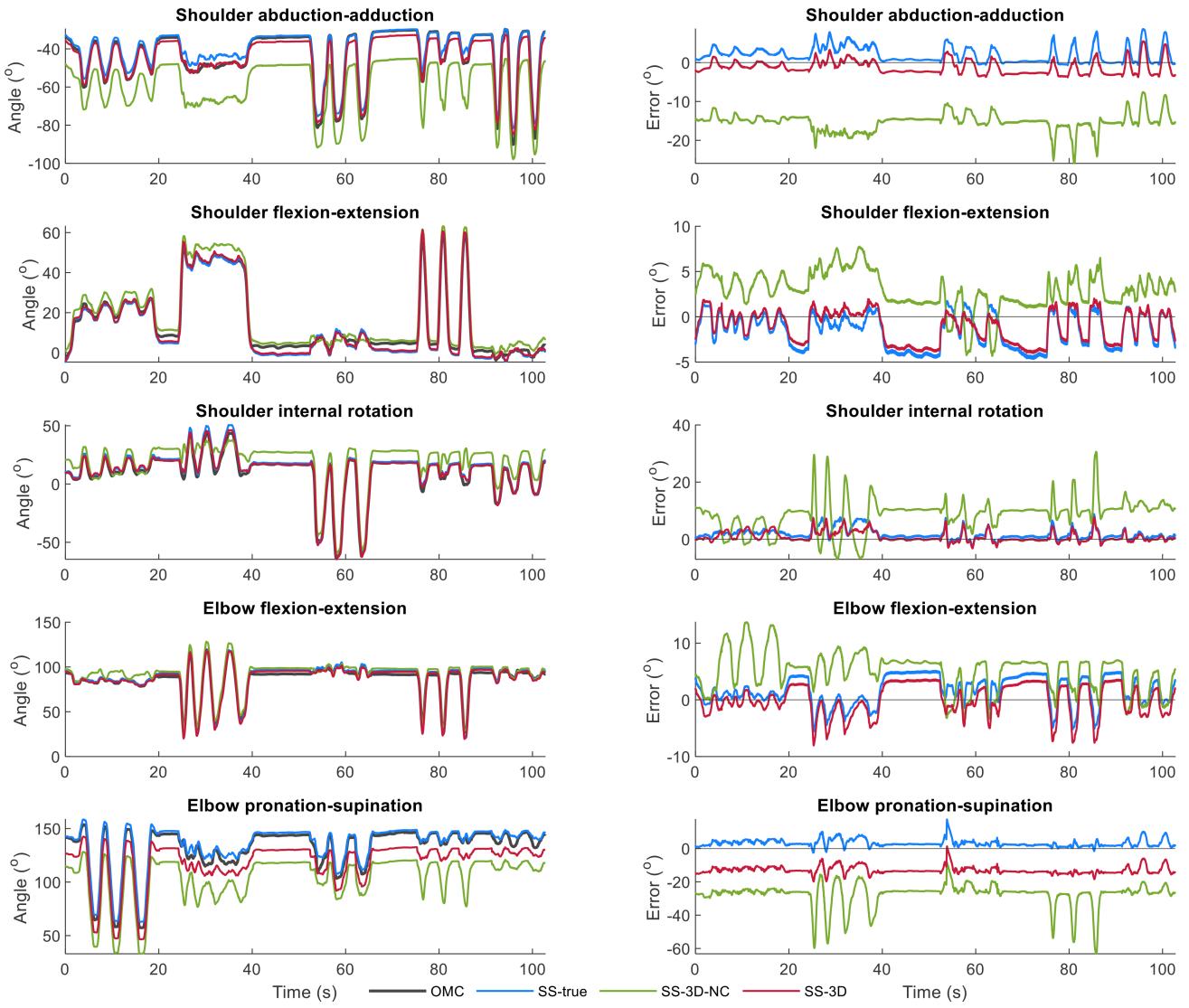


Fig. 7. Joint angle estimation results from one trial using different sensor-to-segment transformations. The left column shows the estimated joints and the right column shows the estimation errors compared with the ground truth (OMC). The sensor-to-segment transformations are calculated from static motion capture data (SS-true), and identification results using 3-D movements with and without compensation (SS-3-D and SS-3-D-NC) for estimation.

pronation/supination angle. The RMSEs and correlation coefficients between the estimated pose and ground truth of all trials are listed in Table III. The trunk angles are small and range from 0.75° – 2.09° , indicating that the assumption of neutral upright pose during calibration is valid. The estimation error using 3-D movements shows a relatively large error in forearm pronation–supination (9.34°) and shoulder internal–external rotation (3.91°), while errors of other DoFs are much lower ($\leq 3.23^\circ$). The error using 2-D movements is slightly larger for most DoFs of shoulder and elbow joints (0.75° – 5.44°), while the forearm pronation/supination presents a smaller error (7.44°). The estimation results from the sensor-to-segment transformation without compensation show a much larger error in shoulder abduction/adduction, shoulder internal rotation, and forearm pronation/supination angles for both 3-D and 2-D calibration movements (see Table III).

The estimated joint angles also show high correlation coefficients with the ground truth for shoulder and elbow

joints (≥ 0.986 and ≥ 0.984 for results using 3-D and 2-D movements). The correlation coefficients are similar to the corresponding values using true sensor-to-segment transformation, implying that the angle errors mainly come from the offsets due to the errors in identified sensor-to-segment transformation.

IV. DISCUSSION

A. Sensor System

Upper body kinematics measurement has wide application prospects in human-related scenarios and various sensor systems based on cameras or IMU have been proposed in the literature. We combine the camera and IMUs in this work and present a visual–inertial sensor system for upper body pose estimation. Previous work mainly explored various fusion methods for visual and inertial data [44], [51]. Different from those studies, we focus on the sensor-to-segment calibration method based on the abundant information

TABLE III
RMSEs AND CORRELATION COEFFICIENTS BETWEEN ESTIMATED UPPER BODY POSE AND GROUND TRUTH

	RMSE					Correlation coefficients				
	SS-true	SS-2D-NC	SS-3D-NC	SS-2D	SS-3D	SS-true	SS-2D-NC	SS-3D-NC	SS-2D	SS-3D
Trunk flexion-extension	1.19 (0.56)	1.21 (0.57)	1.21 (0.57)	1.21 (0.57)	1.21 (0.57)	0.810 (0.121)	0.810 (0.122)	0.809 (0.121)	0.810 (0.122)	0.810 (0.121)
Trunk lateral bending	0.75 (0.29)	0.76 (0.30)	0.76 (0.30)	0.75 (0.30)	0.76 (0.30)	0.882 (0.089)	0.882 (0.088)	0.880 (0.091)	0.883 (0.087)	0.881 (0.090)
Trunk axial rotation	2.08 (1.07)	2.09 (1.07)	2.09 (1.07)	2.09 (1.07)	2.09 (1.07)	0.736 (0.151)	0.733 (0.155)	0.735 (0.155)	0.731 (0.154)	0.733 (0.155)
Shoulder abduction-adduction	2.71 (0.94)	6.94 (2.43)	7.96 (3.62)	3.76 (1.16)	3.23 (0.96)	0.988 (0.014)	0.985 (0.023)	0.992 (0.004)	0.988 (0.011)	0.991 (0.008)
Shoulder flexion-extension	2.26 (1.27)	2.76 (0.87)	2.68 (0.78)	3.50 (1.20)	2.59 (1.22)	0.998 (0.001)	0.996 (0.002)	0.995 (0.004)	0.998 (0.001)	0.998 (0.001)
Shoulder internal-external rotation	4.18 (1.56)	12.09 (5.21)	7.04 (2.98)	5.44 (3.08)	3.91 (1.53)	0.986 (0.010)	0.953 (0.037)	0.974 (0.017)	0.984 (0.011)	0.986 (0.007)
Elbow flexion-extension	2.60 (1.16)	3.93 (1.76)	3.58 (1.27)	2.45 (1.02)	2.31 (0.59)	0.998 (0.002)	0.989 (0.011)	0.994 (0.006)	0.996 (0.004)	0.997 (0.002)
Forearm pronation-supination	3.16 (1.00)	16.10 (7.04)	15.99 (6.78)	7.44 (3.52)	9.34 (5.10)	0.989 (0.005)	0.917 (0.064)	0.956 (0.025)	0.985 (0.008)	0.986 (0.009)

Results are presented in the form of mean (standard deviation).

from cameras and IMUs. Although it is possible to calibrate the bony landmarks as used in [42] and [44] to calculate the sensor-to-segment transformation, we propose a dynamic movement-based method to automatically identify the transformations, which enables the system to be used without assistance from other persons.

The orientations of the sensor modules can be obtained by fusing the marker pose and inertial data and removing the need for magnetometers for heading angle estimation. Therefore, the sensor system will not be affected by magnetic or ferromagnetic materials that are commonly used in assistive robots. However, the camera may be affected by occlusion caused by robots or objects. Since the IMUs can provide relatively accurate orientations in short-term integration, occlusion for a short period, which also occurred in our testing stage, has little influence in real applications. Only a single camera is used and its location can be selected based on the application scenario with little calibration effort. The camera-marker distance can be further increased when a higher resolution is used. The camera can also be replaced with affordable webcams or phone cameras in home-based applications.

B. Estimation Accuracy

With the proposed calibration and estimation methods, most of the joint angles show high accuracy and correlation with ground truth, in the range of 0.76° – 3.91° and 0.75° – 5.44° when using 3-D and 2-D calibration movements, respectively. The forearm pronation-supination, however, presented a larger error (9.34° and 7.44°). A major reason accounting for this error is that this angle is mainly determined by manual alignment of the marker plate of SFA with the wrist flexion axis, which may introduce an offset between the estimated angle and the ground truth. Adding a sensor module on the hand in the future to estimate the wrist joint axis would also be helpful to improve the accuracy of the pronation-supination angle.

The automatic calibration method proposed in [38] to estimate elbow angles resulted in a similar accuracy in elbow flexion/extension angle (2.7°) and higher accuracy in pronation/supination angle (3.8°). However, it should be noted that this error is calculated by removing the angle offset at the initial static phase (the “zero pose”). Manually calibrating the bony landmarks with a wand in [44] achieved an average RMSE of 2.7° and a correlation of 0.87. In comparison, the proposed fusion method using the identified sensor-to-segment transformation achieves a similar RMSE of 2.3° and a higher correlation of 0.93. Bouvier et al. [29] examined several sensor-to-segment calibration methods for IMC systems based on the static pose or dynamic movements and the estimation error of upper limb joint angles was in the range of 5° – 10° , larger than this work. Picerno et al. [42] calibrated the sensor-to-segment transformation with an extra IMU on a caliper-like device and the error of the estimated joint range of motion is around 1.8° – 4.4° , which is also comparable to the proposed automatic method without extra assistance.

The errors in the estimated joint angles come from multiple sources. First, at the module level, the orientation errors of single marker modules with the proposed visual–inertial fusion method are around 0.55° – 1.81° . Although this error is small relative to the errors in joint angles and similar to the results in studies fusing IMU and magnetometers, it may introduce errors to the identification process together with the position error from marker detection. The error in the identified sensor-to-segment transformation will result in a constant offset to the estimated joint angles. Second, the soft tissue artifact issue causes variations in the sensor-to-segment transformations up to 1.38° and 2.07 mm, even after the proposed compensation (see Table I). This error affects both the calibration process and the fusion process. Third, the inaccuracy of the model used in the calibration process leads to errors in estimating sensor-to-segment transformation and estimated joint angles. For example, the ideal spherical shoulder joint in the model

introduces consistency between the measured and calculated positions of the more distal sensor modules and segment frames. More accurate skeletal models as those defined in the OpenSim software may be adopted.

C. Limitations

There are still several limitations in the current study that need to be addressed in future work. First, we only tested two typical arm movements for calibration to validate the feasibility of the proposed methods. Further tests using end-effector robots with precisely controlled movements would be helpful to analyze the impacts of different movements on the estimation accuracy. Second, the computation cost of the identification process with bilevel optimization is quite high and real-time optimization will be developed to reduce the time of the calibration stage. Third, the trunk pose is assumed to be neutral upright at the static phase since the trunk presents little motion during the calibration. Estimating the trunk pose from multiple markers or using a markerless detection method would be desirable in the future.

V. CONCLUSION

In this work, we presented a visual–inertial system for upper body pose estimation. The inertial data from IMUs are fused with marker poses to obtain driftless 3-D orientations of the sensor modules. With the orientations and position of sensor modules, an automatic sensor-to-segment calibration method is proposed. Simple arbitrary movements are used for the calibration process, enabling easy use for users at home or community. The variation of sensor-to-segment transformation is also compensated to achieve higher identification accuracy. Upper body poses are then estimated from the orientation and position of sensor modules via an EKF. Experimental results validated the feasibility of the proposed methods and high accuracy in the shoulder and elbow joint angles is achieved. The possible sources of errors are analyzed to discuss future directions to improve accuracy. The proposed sensor system and estimation methods do not need predefined poses for calibration and thus are suitable for patients with movement issues and applications such as robot-assisted rehabilitation.

ACKNOWLEDGMENT

The authors would like to thank the subjects who volunteered to participate in this study.

REFERENCES

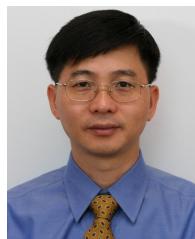
- [1] O. Pila, C. Duret, J.-M. Gracies, G. E. Francisco, N. Bayle, and É. Hutin, “Evolution of upper limb kinematics four years after subacute robot-assisted rehabilitation in stroke patients,” *Int. J. Neurosci.*, vol. 128, no. 11, pp. 1030–1039, Nov. 2018.
- [2] A. Schwarz, C. M. Kanzler, O. Lamberty, A. R. Luft, and J. M. Veerbeek, “Systematic review on kinematic assessments of upper limb movements after stroke,” *Stroke*, vol. 50, no. 3, pp. 718–727, Mar. 2019.
- [3] Y. Hwang, S. Lee, J. Hong, and J. Kim, “A novel end-effector robot system enabling to monitor upper-extremity posture during robot-aided planar reaching movements,” *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3035–3041, Apr. 2020.
- [4] P. Domínguez-Téllez, J. A. Moral-Muñoz, A. Salazar, E. Casado-Fernández, and D. Lucena-Antón, “Game-based virtual reality interventions to improve upper limb motor function and quality of life after stroke: Systematic review and meta-analysis,” *Games Health J.*, vol. 9, no. 1, pp. 1–10, Feb. 2020.
- [5] K. H. Kong et al., “Efficacy of a virtual reality commercial gaming device in upper limb recovery after stroke: A randomized, controlled study,” *Topics Stroke Rehabil.*, vol. 23, no. 5, pp. 333–340, Oct. 2016.
- [6] Y. Liu, Z. Li, H. Liu, and Z. Kan, “Skill transfer learning for autonomous robots and human–robot cooperation: A survey,” *Robot. Auto. Syst.*, vol. 128, Jun. 2020, Art. no. 103515.
- [7] C. Cael, *Functional Anatomy: Musculoskeletal Anatomy, Kinesiology, and Palpation for Manual Therapists*. Burlington, MA, USA: Jones & Bartlett Learning, 2022.
- [8] G. Rab, K. Petuskey, and A. Bagley, “A method for determination of upper extremity kinematics,” *Gait Posture*, vol. 15, no. 2, pp. 113–119, Apr. 2002.
- [9] A. Leardini, L. Chiari, U. Della Croce, and A. Cappozzo, “Human movement analysis using stereophotogrammetry. Part 3. Soft tissue artifact assessment and compensation,” *Gait Posture*, vol. 21, no. 2, pp. 212–225, Feb. 2005.
- [10] N. M. Fiorentino, P. R. Atkins, M. J. Kutschke, J. M. Goebel, K. B. Foreman, and A. E. Anderson, “Soft tissue artifact causes significant errors in the calculation of joint angles and range of motion at the hip,” *Gait Posture*, vol. 55, pp. 184–190, Jun. 2017.
- [11] A. Barre, J. P. Thiran, B. M. Jolles, N. Theumann, and K. Aminian, “Soft tissue artifact assessment during treadmill walking in subjects with total knee arthroplasty,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3131–3140, Nov. 2013.
- [12] P. Eichelberger et al., “Analysis of accuracy in optical motion capture—A protocol for laboratory setup evaluation,” *J. Biomech.*, vol. 49, no. 10, pp. 2085–2088, Jul. 2016.
- [13] M. Hofmann and D. M. Gavrila, “Multi-view 3D human pose estimation in complex environment,” *Int. J. Comput. Vis.*, vol. 96, no. 1, pp. 103–124, Jan. 2012.
- [14] L. Needham et al., “The accuracy of several pose estimation methods for 3D joint centre localisation,” *Sci. Rep.*, vol. 11, no. 1, p. 20673, Oct. 2021.
- [15] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.
- [16] N. U. Khan and W. Wan, “A review of human pose estimation from single image,” in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Jul. 2018, pp. 230–236.
- [17] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with Microsoft Kinect sensor: A review,” *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Oct. 2013.
- [18] Y. Guo, F. Deligianni, X. Gu, and G.-Z. Yang, “3-D canonical pose estimation and abnormal gait recognition with a single RGB-D camera,” *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3617–3624, Oct. 2019.
- [19] P. Picerno, “25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches,” *Gait Posture*, vol. 51, pp. 239–246, Jan. 2017.
- [20] D. Roetenberg, H. Luinge, and P. Slycke, “Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors,” Xsens Motion Technol. BV, Enschede, The Netherlands, Tech. Rep., 2009, pp. 1–7.
- [21] D. Laidig, I. Weygers, S. Bachhuber, and T. Seel, “VQF: A milestone in accuracy and versatility of 6D and 9D inertial orientation estimation,” in *Proc. 25th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2022, pp. 1–6.
- [22] M. Nazarhari 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*, vol. 68, pp. 67–84, Apr. 2021.
- [23] B. Fan, Q. Li, and T. Liu, “How magnetic disturbance influences the attitude and heading in magnetic and inertial sensor-based orientation estimation,” *Sensors*, vol. 18, no. 1, p. 76, 2018.
- [24] S. O. H. Madgwick, A. J. L. Harrison, and A. Vaidyanathan, “Estimation of IMU and MARG orientation using a gradient descent algorithm,” in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun./Jul. 2011, pp. 1–7.
- [25] F. Gao, G. Liu, F. Liang, and W.-H. Liao, “IMU-based locomotion mode identification for transtibial prostheses, orthoses, and exoskeletons,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1334–1343, Jun. 2020.

- [26] S. Ding, X. Ouyang, T. Liu, Z. Li, and H. Yang, "Gait event detection of a lower extremity exoskeleton robot by an intelligent IMU," *IEEE Sensors J.*, vol. 18, no. 23, pp. 9728–9735, Dec. 2018.
- [27] S. Garcia-de-Villa, A. Jimenez-Martin, and J. J. Garcia-Dominguez, "Novel IMU-based adaptive estimator of the center of rotation of joints for movement analysis," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [28] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon, "Accuracy and repeatability of single-pose calibration of inertial measurement units for whole-body motion analysis," *Gait Posture*, vol. 54, pp. 80–86, May 2017.
- [29] 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*, vol. 15, no. 8, pp. 18813–18833, Jul. 2015.
- [30] L. Pacher, C. Chatellier, R. Vauzelle, and L. Fradet, "Sensor-to-segment calibration methodologies for lower-body kinematic analysis with inertial sensors: A systematic review," *Sensors*, vol. 20, no. 11, p. 3322, Jun. 2020.
- [31] R. V. Vitali and N. C. Perkins, "Determining anatomical frames via inertial motion capture: A survey of methods," *J. Biomech.*, vol. 106, Jun. 2020, Art. no. 109832.
- [32] I. Weygers et al., "In-vitro validation of inertial-sensor-to-bone alignment," *J. Biomech.*, vol. 128, Jan. 2021, Art. no. 110781.
- [33] B. Fan, Q. Li, T. Tan, P. Kang, and P. B. Shull, "Effects of IMU sensor-to-segment misalignment and orientation error on 3-D knee joint angle estimation," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2543–2552, Feb. 2022.
- [34] Z.-Q. Zhang and J.-K. Wu, "A novel hierarchical information fusion method for three-dimensional upper limb motion estimation," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3709–3719, Nov. 2011.
- [35] M. Zabat, A. Ababou, N. Ababou, and R. Dumas, "IMU-based sensor-to-segment multiple calibration for upper limb joint angle measurement—A proof of concept," *Med. Biol. Eng. Comput.*, vol. 57, no. 11, pp. 2449–2460, Nov. 2019.
- [36] A. G. Cutti, A. Giovanardi, L. Rocchi, A. Davalli, and R. Sacchetti, "Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors," *Med. Biol. Eng. Comput.*, vol. 46, no. 2, pp. 169–178, 2008.
- [37] J. Z. Nie, J. W. Nie, N.-T. Hung, R. J. Cotton, and M. W. Slutsky, "Portable, open-source solutions for estimating wrist position during reaching in people with stroke," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, Nov. 2021.
- [38] P. Müller, M.-A. S. T. Bégin, and T. Seel, "Alignment-free, self-calibrating elbow angles measurement using inertial sensors," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 2, pp. 312–319, Dec. 2017.
- [39] T. Seel, T. Schauer, and J. Raisch, "Joint axis and position estimation from inertial measurement data by exploiting kinematic constraints," in *Proc. IEEE Int. Conf. Control Appl.*, Oct. 2012, pp. 45–49.
- [40] M. H. Schwartz and A. Rozumalski, "A new method for estimating joint parameters from motion data," *J. Biomech.*, vol. 38, no. 1, pp. 16–107, Jan. 2005.
- [41] M. Crabolu, D. Pani, L. Raffo, M. Conti, P. Crivelli, and A. Cereatti, "In vivo estimation of the shoulder joint center of rotation using magneto-inertial sensors: MRI-based accuracy and repeatability assessment," *Biomed. Eng. OnLine*, vol. 16, no. 1, p. 34, Mar. 2017.
- [42] P. Picerno et al., "Upper limb joint kinematics using wearable magnetic and inertial measurement units: An anatomical calibration procedure based on bony landmark identification," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019.
- [43] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [44] R. Mallat, V. Bonnet, M. A. Khalil, and S. Mohammed, "Upper limbs kinematics estimation using affordable visual-inertial sensors," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 1–11, Jan. 2022.
- [45] T. K. Chan, Y. K. Yu, H. C. Kam, and K. H. Wong, "Robust hand gesture input using computer vision, inertial measurement unit (IMU) and flex sensors," in *Proc. IEEE Int. Conf. Mechatronics, Robot. Autom. (ICMRA)*, May 2018, pp. 95–99.
- [46] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [47] D. Tedaldi, A. Pretto, and E. Menegatti, "A robust and easy to implement method for IMU calibration without external equipments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 3042–3049.
- [48] A. G. Cutti, A. Cappello, and A. Davalli, "In vivo validation of a new technique that compensates for soft tissue artefact in the upper-arm: Preliminary results," *Clin. Biomech.*, vol. 21, pp. S13–S19, Jan. 2006.
- [49] H. J. Bennett, K. Fleenor, and J. T. Weinhandl, "A normative database of hip and knee joint biomechanics during dynamic tasks using anatomical regression prediction methods," *J. Biomech.*, vol. 81, pp. 122–131, Nov. 2018.
- [50] T.-W. Lu and J. J. O'Connor, "Bone position estimation from skin marker co-ordinates using global optimisation with joint constraints," *J. Biomech.*, vol. 32, no. 2, pp. 129–134, 1999.
- [51] X. Lu, K. Yu, Y. Zhang, J. Yi, and J. Liu, "Whole-body pose estimation in physical rider-bicycle interactions with a monocular camera and a set of wearable gyroscopes," *J. Dyn. Syst., Meas., Control*, vol. 139, no. 7, May 2017, Art. no. 071005.



Tong Li (Member, IEEE) received the B.S. and Ph.D. degrees in mechatronic engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively.

He is currently a Research Fellow with the Department of Biomedical Engineering, National University of Singapore, Singapore. His current research interests include wearable sensors, assistive devices, and human modeling.



Haoyong Yu (Senior Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1988 and 1991, respectively, and the Ph.D. degree in mechanical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002.

He is currently an Associate Professor with the Department of Biomedical Engineering, National University of Singapore, Singapore. His areas of research include medical robotics, rehabilitation engineering and assistive technologies, and system dynamics and control.