# A Wearable System for Articulated Human Pose Tracking under Uncertainty of Sensor Placement

Xuesu Xiao and Shuayb Zarar

*Abstract*— To precisely track human motion, today's state-of-the-art employs either well-calibrated sensors tightly strapped to the body or high-speed cameras confined to a finite capture volume. These restrictions make such systems less mobile. In this paper, we aim to break this usability barrier around motion-capture technology through a wearable system that has sensors integrated directly into garments. We develop a pose-estimation approach based on classic kinematics and show that it is insufficient to analyze motion in such a system, leading to mean Euler angle errors of up to $\pm 60°$ and standard deviations of $120°$. Thus, we motivate the need for data-driven algorithms in this domain. Through a quantitative study, we attribute motion-estimation errors to the high-degree of sensor displacement (up to $118°$ standard deviation from the nominal value) with respect to the body segments that are present when human poses change. Based on controlled experiments, we develop a new dataset for such systems comprising over 3 hours of biomechanical motion recordings from 215 trials on 12 test subjects.

## I. INTRODUCTION

High-accuracy human pose tracking, also known as motion capture, is traditionally achieved with the use of external cameras placed in well-calibrated studio environments [1]–[3]. Emerging wearable sensors promise to take this technology outside of studio rooms into the real world [4]–[6]. However, such sensors still need to be positioned precisely and attached to the body with straps or tight-fitting clothes [1]–[9]. These restrictions hinder the use of wearable motion-capture technology in applications that rely on long-term tracking such as biomechanical gait assessment and performance measurement in the wild [10]–[12]. In this paper, we present a new kind of wearable system for articulated human-motion analysis that overcomes the existing sensor-mounting and positioning restrictions.

Our system comprises a dense network of sensors (inertial measurement units, infrared proximity and ultrasound) integrated into garments that can be worn over extended periods of time. The sensors are synchronized and connected over a high-bandwidth wireless channel. Within the network, multiple CPUs sample data from the sensors at rates of up to 760 Hz and stream to a base station at speeds of up to 27 Mbps ($\sim$1600 byte UDP broadcast payloads at 90% 802.11 PHY rate). At the base station, this data is processed to track the orientation of body-joints in free space. Simultaneously, the system records depth video from multiple calibrated Kinect sensors and fuses them to produce the ground-truth tracking information for comparison. Segmented and synchronized data from wearable and Kinect

Xuesu Xiao is with the CSE Dept., Texas A&M University, College Station, TX 77843, xiaoxuesu@tamu.edu

Shuayb Zarar is with Microsoft Research, Redmond, WA 98052, shuayb@microsoft.com



Fig. 1: Our system enables pose tracking over the long term with wearable sensors integrated into everyday clothing.

sensors, along with the corresponding RGB video are saved to local storage. We believe that these sensor-rich recordings are uniquely useful to study emerging challenges in human biomechanical analysis. Fig. 1 illustrates the application of our system for long-term, unconstrained motion capture. Our system thus opens up new applications in health monitoring, sports training and AR-VR. The following are the specific contributions we make in this paper:

- We propose a new wearable, garment-integrated system for pose tracking that is non-intrusive and mobile. It enables long-term analysis of human biomechanics.
- We develop a kinematics-based pose-tracking algorithm, apply it to our system, and demonstrate its limitations in offsetting artifacts from sensor displacement.
- We produce time-synchronized recordings from our dense network of sensors along with segmented ground-truth annotations for biomechanical motion analysis. We intend to make these recordings public.

The rest of the paper is organized as follows. In Sec. II, we review the state-of-the-art in wearable human pose tracking along with a background on fusing sensor data to accurately estimate orientation. In Sec. III, we describe our garment-integrated system followed by the data-collection process and a kinematics-based algorithm for pose tracking. In Sec. IV, we study performance of the algorithm and quantify the sensor displacement on body segments. Finally, we conclude in Sec. V.

## II. BACKGROUND AND RELATED WORK

In this section, we provide an overview of existing techniques for motion capture and their differences from ours. We also present the sensor-fusion algorithm that we use to accurately estimate 3D orientation.

## A. Existing Systems for Motion Capture

There are two main categories of pose-tracking systems in the literature: optical and non-optical. Optical systems rely on different kinds of cameras including vision, depth and infrared. This class can be further sub-divided into systems with or without markers attached to the body. An overview of reflective marker-based systems is provided in [13]. The use of Microsoft Kinect sensor for motion capture is a good example of markerless systems. It records depth information in addition to RGB wavelengths [14]. Multiple such sensors have been used for pose tracking [15]. Other markerless systems use computer vision techniques [16], [17]. Optical motion-capture systems tend to have good tracking accuracy. However, they are confined to a limited capture volume and are sensitive to occlusion and lighting variations.

Non-optical approaches do not typically require elaborate studio setups. Some examples in this category employ mechanical, magnetic or inertial sensors. Mechanical devices such as exoskeletons use built-in potentiometers on the joints to capture human motion [18]. Magnetic systems use sensors attached to the body within an artificially-created magnetic field [19]. Inertial sensing systems constitute the most common non-optical approach. Inertial measurement units (IMUs) are widely used to track the orientation of human body segments and reconstruct full pose. For instance, the authors in [7] and [20] utilize classical kinematics to recover articulated human motion from IMUs. There are also approaches that are based on optimization techniques [8]. To improve translational-tracking accuracy, rooted kinematic models are utilized [9]. Unfortunately, most of these approaches still rely on precise sensor positioning and studio calibration.

Some inertial motion-capture systems work outside of controlled laboratory environments, [4]–[6]. These are the closest to what we propose in this paper. However, despite superior mobility and minimality [21], such systems still require additional sensors to be rigidly attached to body segments. Although these approaches do not severely limit the capture volume, their tracking performance is affected by sensor drift. Moreover, the accuracy of these systems is poor in the presence of motion artifacts, which can be severe when the sensor and body segments are not tightly coupled [22]. These limitations thus preclude long-term and non-intrusive tracking. To overcome some of the accuracy limitations, hybrid trackers have been developed [1]–[3]. However, due to the involvement of complementary optical sensors, such techniques are not mobile.

## B. Existing Algorithms for Sensor Fusion

In this paper, we focus on motion capture with inertial sensors. This approach requires us to estimate orientation from IMU sensor readings corresponding to the tri-axis accelerometer, gyroscope and magnetometer. These readings can be converted to angular orientations (3 degrees of freedom) through a process called sensor fusion. A good overview of different sensor-fusion techniques is presented in [23]. A computationally efficient version of this algorithm
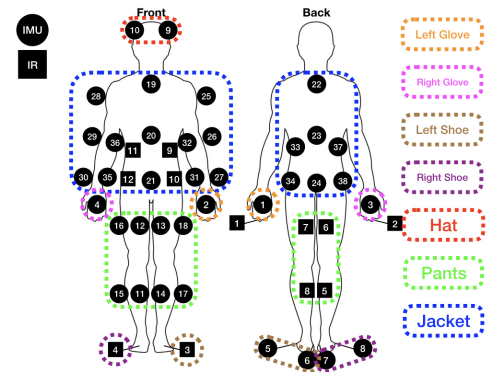


Fig. 2: A dense network of fabric-integrated sensors maintain an association with body parts even as they move around.

suitable for our purposes is the Madgwick filter [24]. This algorithm has comparable accuracy with traditional Kalman-filter based approaches. We thus utilize it extensively in our tracking system.

## III. SYSTEM OVERVIEW

In this section, we describe our wearable motion-capture system in detail. More information about the system is included in the video accompanying this paper. The entire sensor network is shown in Fig. 2. There are 38 IMUs distributed over the body. At least two IMUs are associated with each body segment, e.g., IMU numbers 27 and 31 are associated with the left lower arm. There are also IMUs that are placed close to joint locations, e.g., IMU number 29 is along the right elbow joint. These joint-location sensors are assumed to be associated with either of the abutting body segments. Infrared (IR) sensors are placed between arms and torso, and between two legs. Each hand and foot also has one IR sensor. These sensors complement the IMUs by detecting the distance between body parts based on time-of-flight proximity readings.

Note that there are multiple IMUs per body segment to allow future studies on the benefit of redundant sensing. This is also the reason for including IR sensors in the system. The latter would potentially provide a way of measuring distances between arms, torso and legs. For algorithms in this paper, we only use a single IMU per body segment. Although sensors are nominally assumed to be placed at the locations shown in Fig. 2, their position shifts over time during system operation. This is because these sensors are embedded into garments (and mounted loosely) as described next. Thus, the only usable geometric property maintained by these sensors is that they remain close to the originally associated body segments throughout the state of motion.

## A. Hardware Design with Fabric-integrated Sensors

The proposed wearable system comprises sensors that are embedded into multiple garments: a hat, pants, jacket, pair of shoes and gloves. These are shown in Fig. 3. Each piece of garment contains at least one CPU; to keep up data rate in the presence of high sensor density, we employ 3 CPUs in the jacket. A CPU in our case is defined as an integrated module comprising a processor and radio. The CPU specifications

TABLE I. Detailed specifications of the proposed system.

| CPU (ARM Cortex M0 + ESP8266) | | IR Sensor (ADI VCNL) | |
|---|---|---|---|
| Frequency | 80 MHz | Range | 1-200 mm |
| Flash | 4 MB | Rate | 250 Hz |
| $V_{dd}$ | 3.3 V | $V_{dd}$ | 2.5-3.6 V |
| Protocol | WiFI UDP RFC768 | Interface | I$^2$C |
| Total Power | 478 mW (16 dBm tx.) | Total Power | 50 mW |
| **IMU (ST Microelectronics LSM9DSO** | | | |
| Components 2.4 - 3.6 V SPI/I$^2$C | 3-axis Accelerometer | Range: $\pm 2g$, rate: 1600 Hz anti-aliasing filter: 773 Hz | |
| | 3-axis Gyroscope | Range: 245 DPS, rate: 760 Hz HPF cutoff frequency: 0.09 Hz | |
| | 3-axis Magnetometer | Range: 2 Gauss, rate: 100 Hz | |
| | Temperature | Range: -40 to 85°C | |
| Total Power | 16 - 25 mW (depending on duty cycle and data rate) | | |
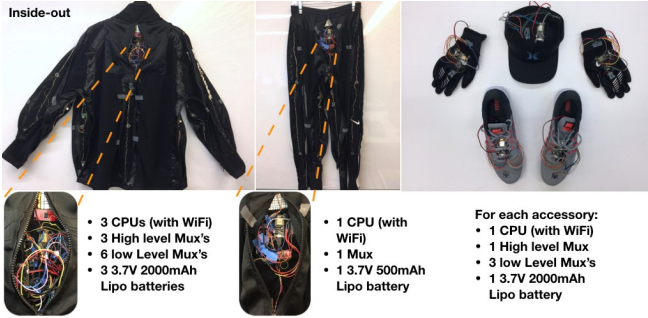


Fig. 3: System hardware: 9 CPUs aggregate data from 46 sensors over a high-bandwidth WiFi network.

are shown in Table I. Thus, a total of 9 CPUs poll data from sensors through an I$^2$C interface, compose UDP messages, and send them over a 2.4 GHz WiFi network to the base station, which is a PC. Each CPU and sensor assembly is powered by an independent lithium polymer (LiPo) battery. Furthermore, a total of 38 IMUs and 8 IR sensors are connected through a hierarchy of digital multiplexers that are able to disambiguate data from 8 sensors with the same I$^2$C address. Specifications of the IMU and IR sensors are also shown in Table I. Note that the current hardware is only a prototype, and thus bulky. However, it is still useful to demonstrate the impact of sensor displacement on motion tracking. The final system could be largely optimized by further engineering.

For complete mobility, the system utilizes multiple WiFi radios for communication. We use the UDP stack to achieve maximum throughput with a given number of radios. The messages composed by the CPU comprise relative time stamps at sampling intervals (that vary across different sensors). These packets are relayed to a base station (PC) over a high-bandwidth router that maintains an active local network connection. Packets are transmitted at a power level of +16 dBm allowing for sizes of approximately 1600 Bytes (400 samples of 4 Byte floats) to be transmitted at a time up to a distance of approximately one hundred meters. Thus, our system is able to achieve 80-90% of the PHY-layer throughput. The assembled UDP packets are processed at the base station to synchronize for time (up to the accuracy of the sensor CPU clock frequency) and
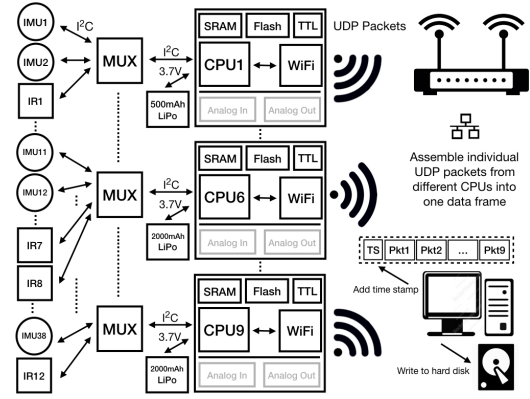


Fig. 4: Network-level connectivity of our system. Data is streamed over UDP to minimize latency.

superfluous packets are eliminated by detecting outliers. It worths to note that the base station (PC) can be replaced by any wireless device, which is able to log UDP packets into a large database. For example, a smart phone which receives data from WiFi/Bluetooth and streams data though LTE network to the cloud for storage could totally make data collection possible in the wild. Finally, the remaining clean packets are re-sampled to a uniform frequency, interpolated to account for measurement distortions, and archived on local storage for further processing necessary to track pose. The overall network-level connectivity is illustrated in Fig. 4. There are some analog (ultrasound) sensors along with the corresponding IO interfaces that are not connected in the current system but are provisioned to be used by the CPUs in future iterations. With this system, we collect approximately half a million synchronized sensor samples. Next we describe the data-collection process in detail.

### B. Data-collection Process

In this section, we describe our experimental framework including details on the subjects, motion patterns and data structures.

**Experimental studio.** We set up a laboratory space where test subjects performed predefined types of motion while wearing the sensor-integrated garments described in the previous section. The studio is shown in Fig. 5. It is a ventilated facility provisioned for charging the garment batteries, when not in use. There was also a monitor screen that played back subject motions as they were being performed and recorded *via* a video camera. Besides controlling the types of motion being performed, the goal of the studio was to also register ground-truth information for the joint angles as they changed during the course of movement. We utilized 2 Microsoft Kinect sensors to achieve this goal. These were calibrated using a white board so that the two streams of point-cloud data could be registered to the same local map. They were placed at a distance that was sufficient to capture the full range of motion. Multiple short video clips of different motion types were played on the monitor screen and the subjects followed the actor in the video. While Kinect recordings were made, data from the wearable sensors were
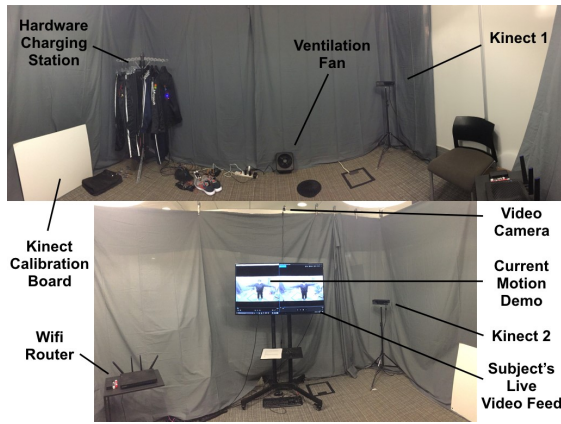
Fig. 5: Studio where sensor data and ground-truth joint angle information was obtained by fusing 2 Kinect outputs.

simultaneously streamed over to the base station PC.

**Offline processing.** At the base station, we synchronized data streams using time stamps and a pre-determined starting pose. We also utilized a third-party software to fuse Kinect sensor information and obtain the ground-truth joint-angle data [25]. We saved this data in a BioVision (BVH) format [26]. The BVH file is a widely-used standard for storing motion-capture data. It contains aspects of *hierarchy* (skeletal structure, dimension and connectivity relationship between joints) and *motion* (changing values of the joint angles over time for the skeletal model).

**Data structure.** The organization of our data is shown in Fig. 6. It comprises three parts: raw, BVH and motion data. The raw data corresponds to sensors (IMU+IR) readings on the garments. Each frame of sensor data is an ordered list of absolute time stamps and 9 UDP packets. Each packet includes a relative time stamp and sensor data in numeric order (as shown in Fig. 2). For each IMU, data is arranged in the order of accelerometer, magnetometer and gyroscope readings along the x, y and z axes. For the two packets that contain IR data (packets 6 and 9), readings are stored after IMU data. In the dataset structure of Fig. 6, the BVH data includes joint-angle information (hierarchy+motion) obtained by fusing the Kinect sensors. The motion data part is a stripped down version of the BVH data comprising just the motion information.

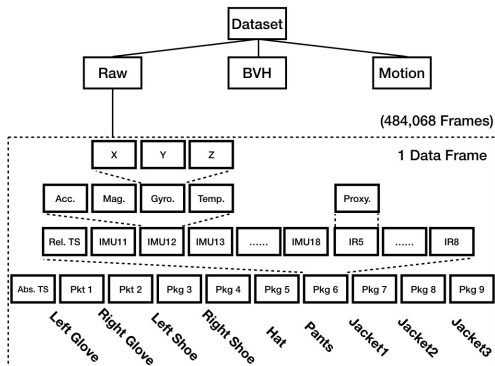**Motion patterns.** After an internal IRB approval process,



Fig. 6: Organization and structure of the collected data.

TABLE II. Summary of the collected dataset.

| Dataset | |
|---|---|
| No. motion types | 5 |
| No. repetitions per motion | 2 - 6 |
| No. subjects | 12 (8 male, 4 female) |
| Age of subjects | 26 - 36 yrs. |
| Total no. trials | 215 |
| Avg. length of one trial | 75 seconds |
| Data sampling frequency | 30 Hz |
| Total no. frames | ∼500k |
| Ground truth collection | 2 Kinect sensors + iPiSoft Motion Capture Studio |

we invited 12 subjects (8 male and 4 female) in the age group of 26-36 yrs. for data collection. We played actor recordings of different motion types and asked the subjects to mimic them over several repetitions. Since our objective was to only track the rotational degrees of freedom - and not the translational degrees - we restricted motion types to stimulate a wide variety of sensor and limb orientations. Furthermore, we avoided motion along the axis of the bones since Kinect is not capable of tracking this type of motion. We also did not track fingers and supinations/pronations of the feet and wrists. Eventually, we collected 215 clean data trials (averaging 75 seconds each with 30 Hz sampling rate), containing 2-6 repetitions of each motion per subject. The total number of BVH frames was close to 500k. The summary of the collected dataset is shown in Table II. The subjects performed the following types of motion:

(1). Upper Arms : move upper arm with elbow straight to reach multiple extremities
(2). Lower Arms : fix upper arm and move lower arm to reach multiple extremities
(3). Arm Swing : swing arms forward and backward
(4). Boxing : perform boxing motion
(5). Walking : perform walking motion

*C. Algorithm for Motion Capture*

In this section, we present our kinematics-based algorithm for motion capture, which assumes that the sensor locations do not change with respect to body segments; note that this assumption is not strictly true in our system. We provide theory for the algorithm in this section and study its performance with our dataset in the next.

For the $i^{th}$ body segment and the associated sensor, we define two coordinate frames: body (segment) frame $B_i$ and the corresponding sensor frame $S_i$. We also define a global frame of reference $G$ (aligned with the Earth's magnetic field), which is the same for all body segments and sensors. Further, we denote a transformation between arbitrary frames $X$ and $Y$ *via* the rotation matrix $R_X^Y$. Before any motion begins, subjects hold a calibration pose (T-pose with hands lifted up on the side and head still) facing the global North direction pre-marked in the studio. In this pose, all joint angles are defined to be zero. Thus, frames $B_i$ and $G$ are aligned at the start, for all body segments. Since the

calibration process aligns coordinates with $G$, all rotation matrices below can be easily reproduced by other interested researchers. Further, the transformation $R_{S_i}^G$ is computed using the sensor-orientation information obtained from the Madgwick fusion algorithm described in Sec. II-B. Since frames $B_i$ and $G$ are aligned, $R_{S_i}^G$ is equal to $R_{S_i}^{B_i}$ in this initial pose.

**Kinematics with rotational transformations.** Suppose $B_i'$ and $S_i'$ denote the body and sensor frames, respectively, after motion occurs. The transformation matrices between all two connected body-segment frames $i$ and $j$ are the joint angles that we seek to track. We track a total of 12 joint segments shown in Fig. 2. For instance, $R_{B_j'}^{B_i'}$ is the joint angle of the left elbow after movement when $i$ and $j$ correspond to the left lower and upper arms, respectively. Given this framework, we can express any imaginary point $P$ in space *via* the following transformation of coordinate systems:

$$R_{S_i'}^G R_{B_i'}^{S_i'} R_{B_i'}^{B_i'} P = R_{S_j'}^G R_{B_j'}^{S_j'} P \tag{1}$$

By simplifying this expression with matrix inversion, we get:

$$\begin{aligned} R_{B_j'}^{B_i'} &= R_{B_i'}^{S_i'}{}^{-1} R_{S_i'}^{G}{}^{-1} R_{S_j'}^{G} R_{B_j'}^{S_j'} \\ &= R_{S_i'}^{B_i'} R_{S_i'}^{G}{}^{-1} R_{S_j'}^{G} R_{S_j'}^{B_j'}{}^{-1} \end{aligned} \tag{2}$$

$R_{S_i'}^G$ and $R_{S_j'}^G$ are the sensor orientations obtained through sensor fusion. Because of our initial assumption of invariant sensor displacement, $R_{S_i'}^{B_i'} \approx R_{S_i}^{B_i}$ and $R_{S_j'}^{B_j'} \approx R_{S_j}^{B_j}$. Furthermore, $R_{S_i}^{B_i}$ and $R_{S_j}^{B_j}$ are obtained from calibration, allowing us to determine the required joint angles $R_{B_j'}^{B_i'}$.

## IV. PERFORMANCE OF INERTIAL MOTION CAPTURE

In this section, we apply the previously-presented algorithm to the sensor data obtained from our wearable system. This algorithm provides a baseline performance level of an inertial motion-capture system. We analyze the different sources of error leading to inaccuracies of the kinematics algorithm. We also quantify the level of sensor displacement for different body segments and study its impact on the algorithm.

### A. Baseline Accuracy

We employ Eq. (2) to obtain all joint angles of interest, utilizing only one IMU sensor per segment (redundant sensors including IMUs and IR proximity sensors could potentially improve tracking accuracy in future research, but are ignored in the baseline tracker). The kinematic chain starts from the hip and reaches the extremities (hand, left and right feet, left and right forearms) during the estimation process that expresses joint angle transformations *via* Euler angles.

**Estimation error.** The average absolute joint-angle error and the associated standard deviation is shown in Fig. 7. From the figure, we observe that the algorithm has substantial error on upper body limbs: left and right shoulders and forearms (up to $\pm 60°$), which also exhibit large standard deviations (up to $120°$). We hypothesize that this error is
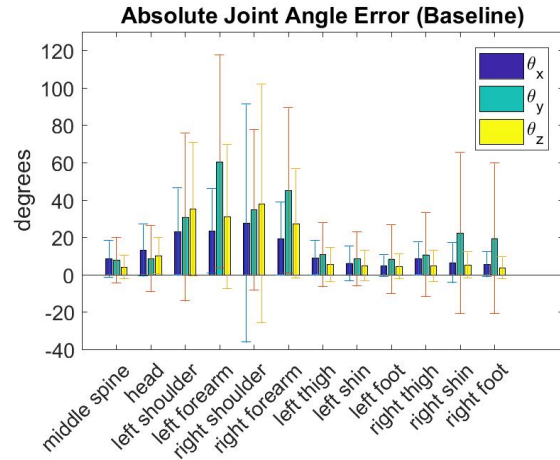


Fig. 7: Average absolute joint angle estimation error of the kinematics-based algorithm.

because the invariant sensor-displacement assumption made in the previous section is too strong for the upper limbs where the garments can shift dramatically leading to high error in joint-angle estimations. For other body parts, the error is within $\pm 10 - 20°$, which is reasonable for an inertial motion capture system. Details of the estimation error for different types of motion are shown in Table III.

**Time-domain analysis.** The performance (mean norm of all joint angles) of one sample trial over the entire course of movement is shown in Fig. 8. In this trial, the same motion is repeated three times after initial calibration. The ground-truth information from the Kinect sensors is also shown alongside the inertial estimation. We recognize multiple error sources in this profile. During calibration, inertial motion capture assumes that the subject is performing a perfect T-pose and thus all angles are 0. However, errors occur when the T-pose is not performed correctly. This shows up as a gap between the ground truth and estimated angles during the initial 700 frames. When the subject starts to move, the ground truth visual tracker registers the movement immediately. However, thanks to system inertia, the sensors in the garment may not move right away until the fabric is stretched, which puts the sensors into motion. Another source is the delay of UDP packets in the network. UDP packets sent out by the CPUs may be received later by the listener. Sensor fusion algorithm may also lack responsiveness and introduce further delay. We denote the combination of all these errors as the sensor movement latency. This latency is not apparent when the body segment stops moving since when the body segment reaches the desired location and stops, the garments and thus the sensors usually stop immediately as well. The final source of error that we can identify from Fig. 8 is the sensor-displacement error. This error is exaggerated in two cases:

(1). At the extreme positions of movement, the estimation is far from the true value because of error accumulated over the course of movement. This explains the error at the three profile peaks in Fig. 8.

(2). Even when the body segments go back to a previous position during repetition, the tracker may yield a differ-

TABLE III. Mean estimation error for different motion types. Expressed as the norm of Euler angle in degrees.

| | Middle spine | Head | Left shoulder | Left forearm | Right shoulder | Right forearm | Left thigh | Left shin | Left foot | Right thigh | Right shin | Right foot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper Arms | 8.7655 | 11.544 | 45.611 | 47.165 | 55.509 | 37.146 | 8.7738 | 4.792 | 5.6409 | 8.5375 | 12.792 | 11.433 |
| Lower Arms | 5.9076 | 9.3416 | 24.241 | 53.602 | 21.485 | 45.103 | 6.7996 | 4.2821 | 3.8721 | 8.0608 | 14.192 | 11.61 |
| Arm Swing | 7.1241 | 13.171 | 35.044 | 45.901 | 32.43 | 31.499 | 7.3654 | 5.4035 | 5.8525 | 8.4775 | 15.635 | 13.137 |
| Boxing | 9.9247 | 15.929 | 33.679 | 50.91 | 43.939 | 42.817 | 11.162 | 6.1338 | 5.5107 | 11.533 | 12.635 | 11.122 |
| Walking | 8.5368 | 13.814 | 27.774 | 38.49 | 32.475 | 31.875 | 16.76 | 18.161 | 14.535 | 13.827 | 22.365 | 18.115 |

ent estimate. This behavior is observed at the beginning and between extremities of motion where the inertial estimations are different even though they are supposed to be the same since the body segment returns to the very same position.

**Dynamic response.** Fig. 9 shows the mean and standard deviation of the estimation error at different movement speeds. The algorithm shows a relatively good performance when the motion is slow *e.g.*, at 0-72 degree per second (dps) the error is within $\pm 20°$. The error and standard deviation increase with higher movement speeds. This is expected because intense motion introduces drastic garment offsets, and therefore sensor displacement on the body segment. These set of experiments further validate the hypotheses about the different error sources that we presented above.

**Error correlations.** Fig. 10 shows how *errors* of different joint angles are correlated. We observe that most significant errors (highlighted in the figure) occur within one piece of clothing (jacket or pants) or between the bridges that connect two garment pieces *i.e.*, hat to jacket, jacket to pants and pants to shoes. Within each piece of fabric, the closer the sensors are, the more correlated the errors are between them. This is because all sensors are not separately mounted on to body segments, as in conventional strap-based inertial motion-capture approaches, but rather are inter-related through a common piece of fabric. The displacement of one sensor, which leads to tracking errors for the associated body segment, may easily propagate to the neighboring sensors by stretching of fabric. This effect diminishes when the sensors are far apart, even through they are still on the same piece of fabric. For instance, errors on the right shoulder and left forearm have low correlation values. The two pairs of limb extremities, left/right forearms and shins are also highly correlated. This is because the motion within these
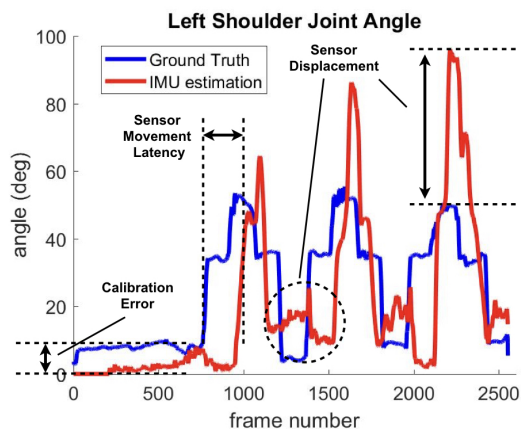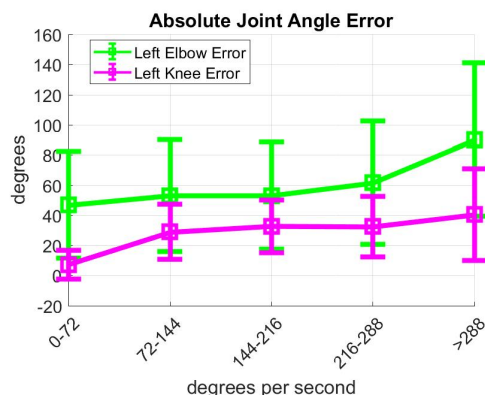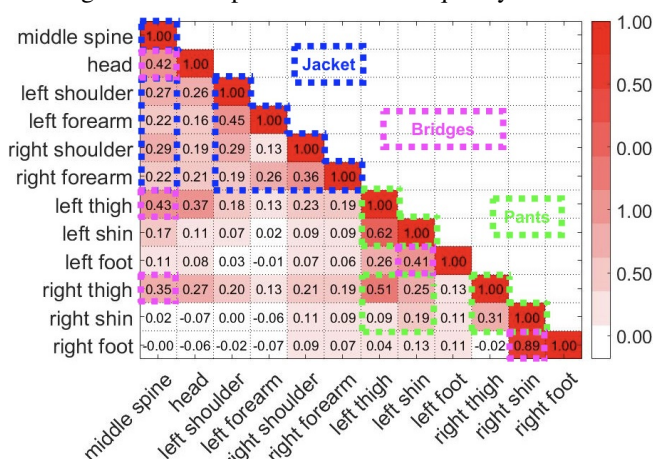
Fig. 9: Tracker performance in frequency domain

Fig. 10: Correlation of estimation error for the joint angles.

pairs are usually similar. To not confuse with correlations of joint angle estimates with skeletal structure of human body, the values presented in Fig. 10 are correlations of errors, which are caused by sensor motion artifacts. These artifacts are interrelated due to a continuum of fabric.

### B. Quantification of Sensor Displacement

By re-arranging terms in Eq. (2), we obtain the following relationship:

$$R_{S_j'}^{B_j'} = R_{B_j'}^{B_i'}{}^{-1} R_{S_i'}^{B_i'} R_{S_i'}^{G}{}^{-1} R_{S_j'}^{G} \qquad (3)$$

$R_{S_i'}^{G}$ and $R_{S_j'}^{G}$ are results of the sensor-fusion algorithm. Suppose, we plug in the ground-truth joint-angle value (from the optical tracker) for $R_{B_j'}^{B_i'}$ and assume that the hat is tightly fixed to the head. We obtain the following approximation:

$$R_{B_{hat}'}^{S_{hat}'} \approx R_{B_{hat}}^{S_{hat}}. \qquad (4)$$

This is not an unreasonable assumption since the hat does not move a lot on the subjects' head and the orientations of the IMUs on the hat are fixed relative to the head. Thus, utilizing
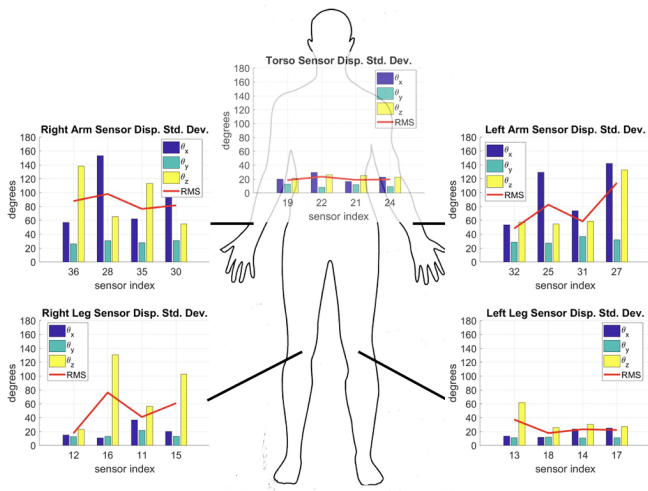
Fig. 8: Errors in inertial pose tracking due to various sources.

Fig. 11: Sensor displacement for different body segments.

the hat as an anchor point, if we plug $R_{S'_{hat}}^{B'_{hat}}$ in place of $R_{S'_i}^{B'_i}$ in Eq. (3), we solve for $R_{S'_j}^{B'_j}$, which is the body segment connected to the head *i.e.*, middle spine. We continue to utilize this concept along the kinematic chain on the body up to the 4 limb extremities. In this way, we quantify the sensor displacements with respect to the body segments $R_{S'_j}^{B'_j}$. The results after transforming the rotation matrices to Euler angles are presented in Fig. 11. We observe from the figure that the sensors on the upper limbs are displaced to the largest degree, while those on the torso are relatively stable.

## V. CONCLUSIONS

By integrating sensors into everyday garments, we proposed a system that overcomes the barriers of low mobility and intrusiveness posed by current inertial motion-capture technologies. No special setup of any kind is necessary to use our pose-tracking system, thus enabling long-term biomechanical motion analysis in the wild. Furthermore, we also showed that traditional approaches based on kinematics-based modeling are insufficient for tracking motion when sensors are integrated into garments. This is because of the high degree of sensor displacement present due to unconstrained mounting. Thus, we motivated the need for better data-driven algorithms for motion capture. Through a carefully designed experimental framework, we produced a synchronized dataset containing measurements from IMUs and IR sensors along with annotated ground truth information about joint angles from Kinect sensors. We will make this dataset public to encourage future research on inertial motion capture that is insensitive to sensor placement. In the future, we intend to augment this dataset with more sensors and more accurate visual baselines like the Vicon and OptiTrack.

## REFERENCES

[1] E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A wide-range wireless motion-tracking system for augmented reality and virtual set applications," in *Proc. Annual Conf. Computer Graphics and Interactive Techniques*. ACM, Jul. 1998, pp. 371–378.

[2] T. von Marcard, G. Pons-Moll, and B. Rosenhahn, "Human pose estimation from video and IMUs," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1533–1547, Aug. 2016.

[3] S. Andrews, I. H. Casado, T. Komura, L. Sigal, and K. Mitchell, "Real-time physics-based motion capture with sparse sensors." in *European Conf. Visual Media Production*, Dec. 2016, pp. 1–5.

[4] D. Vlasic *et. al.*, "Practical motion capture in everyday surroundings," in *ACM Trans. Graphics*, vol. 26, no. 3. ACM, Jul. 2007, p. 35.

[5] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors," *Xsens Motion Technologies, Tech. Rep*, Jan. 2009.

[6] T. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll, "Sparse inertial poser: Automatic 3D human pose estimation from sparse IMUs," in *J. Computer Graphics Forum*, vol. 36, no. 2. Wiley Online Library, May 2017, pp. 349–360.

[7] D. H. Kang, Y. J. Jung, A. J. Park, and J. W. Kim, "Human body motion capture system using magnetic and inertial sensor modules," in *Proc. Int. Universal Communication Symp.*, Oct. 2011, pp. 1–6.

[8] M. Kok, J. D. Hol, and T. B. Schön, "An optimization-based approach to human body motion capture using inertial sensors," *Proc. Int. Federation of Automatic Control*, vol. 47, no. 3, pp. 79–85, Aug. 2014.

[9] Y. Zheng, K.-C. Chan, and C. C. Wang, "Pedalvatar: An IMU-based real-time body motion capture system using foot rooted kinematic model," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*. IEEE, Sep. 2014, pp. 4130–4135.

[10] M. J. Mathie, A. C. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *J. Physio. Measurement*, vol. 25, no. 2, pp. 1–20, Feb. 2004.

[11] M. Lapinski, E. Berkson, T. Gill, M. Reinold, and J. A. Paradiso, "A distributed wearable, wireless sensor system for evaluating professional baseball pitchers and batters," in *Int. Symp. Wearable Computers*. IEEE, Sep. 2009, pp. 131–138.

[12] S. T. Moore, H. G. MacDougall, J.-M. Gracies, H. S. Cohen, and W. G. Ondo, "Long-term monitoring of gait in Parkinson's disease," *J. Gait & Posture*, vol. 26, no. 2, pp. 200–207, Jul. 2007.

[13] M. Windolf, N. Götzen, and M. Morlock, "Systematic accuracy and precision analysis of video motion capturing systems-exemplified on the Vicon-460 system," *J. Biomechanics*, vol. 41, no. 12, pp. 2776–2780, Jan. 2008.

[14] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Apr. 2012.

[15] K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, A. Scholz, and M. A. Magnor, "Markerless motion capture using multiple color-depth sensors." in *Proc. Vision, Modeling and Visualization Wkshp.*, Oct. 2011, pp. 317–324.

[16] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *J. Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, Mar. 2001.

[17] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *J. Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 90–126, Nov. 2006.

[18] X. Gu, Y. Zhang, W. Sun, Y. Bian, D. Zhou, and P. O. Kristensson, "Dexmo: An inexpensive and lightweight mechanical exoskeleton for motion capture and force feedback in VR," in *Proc. Conf. Human Factors in Computing Systems*. ACM, May 2016, pp. 1991–1995.

[19] J. F. O'Brien, R. E. Bodenheimer Jr, G. J. Brostow, and J. K. Hodgins, "Automatic joint parameter estimation from magnetic motion capture data," in *Proc. Graphics Interface*, May 2000, pp. 53–60.

[20] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 12, no. 2, pp. 295–302, Jun. 2004.

[21] Notch Interfaces Inc. Notch your movements reconstructed on your smartphone in 3d. [Online]. Available: https://wearnotch.com/

[22] S. Mohammed and I. Tashev, "Unsupervised deep representation learning to remove motion artifacts in free-mode body sensor networks," in *Int. Conf. Body Sensor Networks*. IEEE, May 2017, pp. 183–188.

[23] D. Novak and R. Riener, "A survey of sensor fusion methods in wearable robotics," *Robotics and Autonomous Systems*, vol. 73, pp. 155–170, Nov. 2015.

[24] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Int. Conf. Rehabilitation Robotics*. IEEE, Jul. 2011, pp. 1–7.

[25] iPi Soft LLC. iPi Soft motion capture for the masses. [Online]. Available: http://ipisoft.com/

[26] J. Thingvold, "Biovision BVH format. [online] available at http://www.cs. wisc. edu/graphics/courses/cs-838-1999/jeff," *(accessed 03 March, 2005)*, 1999.