



SmartPoser: Arm Pose Estimation with a Smartphone and Smartwatch Using UWB and IMU Data

Nathan DeVrio*
Carnegie Mellon University
Pittsburgh, PA, USA
ndeivio@cmu.edu

Vimal Mollyn*
Carnegie Mellon University
Pittsburgh, PA, USA
vmollyn@andrew.cmu.edu

Chris Harrison
Carnegie Mellon University
Pittsburgh, PA, USA
chris.harrison@cs.cmu.edu

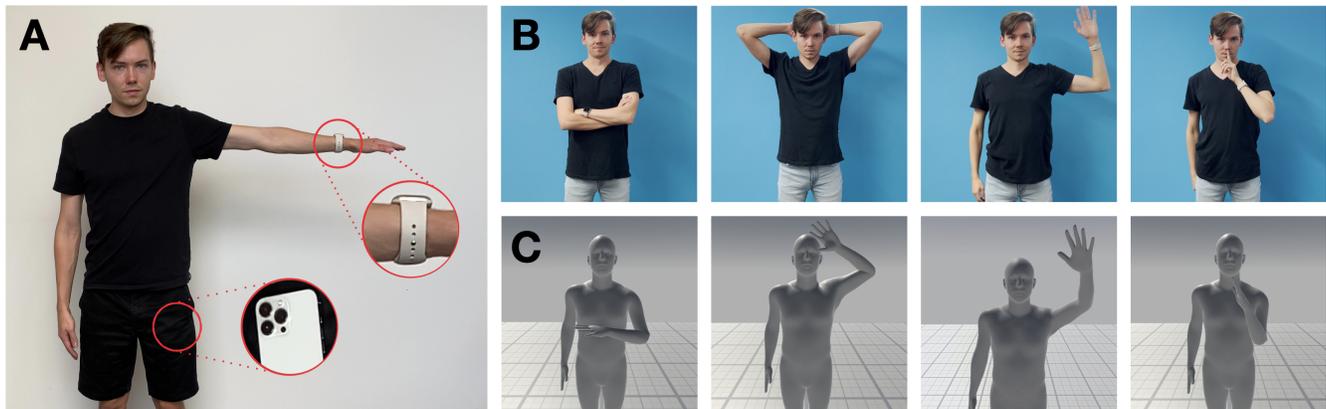


Figure 1: SmartPoser fuses UWB distance and IMU data from an off-the-shelf smartphone and smartwatch (A) to track a wearer’s real-time arm pose (B & C).

ABSTRACT

The ability to track a user’s arm pose could be valuable in a wide range of applications, including fitness, rehabilitation, augmented reality input, life logging, and context-aware assistants. Unfortunately, this capability is not readily available to consumers. Systems either require cameras, which carry privacy issues, or utilize multiple worn IMUs or markers. In this work, we describe how an off-the-shelf smartphone and smartwatch can work together to accurately estimate arm pose. Moving beyond prior work, we take advantage of more recent ultra-wideband (UWB) functionality on these devices to capture absolute distance between the two devices. This measurement is the perfect complement to inertial data, which is relative and suffers from drift. We quantify the performance of our software-only approach using off-the-shelf devices, showing it can estimate the wrist and elbow joints with a median positional error of 11.0 cm, without the user having to provide training data.

CCS CONCEPTS

• **Human-centered computing** → **Mobile devices.**

*Both authors contributed equally.



This work is licensed under a Creative Commons Attribution International 4.0 License.

UIST '23, October 29–November 01, 2023, San Francisco, CA, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0132-0/23/10.
<https://doi.org/10.1145/3586183.3606821>

KEYWORDS

Smartwatch, sensing, hand gestures, body pose, mobile devices, interaction techniques.

ACM Reference Format:

Nathan DeVrio, Vimal Mollyn, and Chris Harrison. 2023. SmartPoser: Arm Pose Estimation with a Smartphone and Smartwatch Using UWB and IMU Data. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, October 29–November 01, 2023, San Francisco, CA, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3586183.3606821>

1 INTRODUCTION

Despite being worn on the wrist, contemporary smartwatches know more about their users’ legs, lungs, and heart than their arms. For instance, smartwatches can predict locomotion mode, track total number of steps taken, and measure respiration and heart rate. Yet, beyond some very specific use cases (e.g., hand washing detection [3], the activities of the arms and hands are unknown, despite the fact that they are the chief appendage we use to manipulate the world around us. If we could better track the arms, we could power more sophisticated applications spanning fitness [10, 54], rehabilitation [5], life logging [30], occupational training [7], and context-aware assistants [34], to name just a few domains.

Researchers have long recognized the importance of tracking arm pose to support such use cases. As we will discuss in the Related Work section, a wide variety of methods have been brought to bear on this problem. Much prior work has utilized external, fixed equipment (most often cameras), but this immediately precludes mobile use. Next most common are worn sensor arrays or suits, which are comparatively more practical, but still face significant

consumer adoption headwinds. Ideally, we could track a user’s arm pose with devices they already own and routinely carry with them [17, 35], relying on no external infrastructure.

In this work, we present SmartPoser, a new arm pose tracking method that uses only an off-the-shelf smartwatch and smartphone (Figure 1). In other words, our technique is software only, and could be enabled on recent devices as an over-the-air update or app download. SmartPoser also allows users to be on the move while it performs its tracking, a capability not found in many prior approaches [42, 61]. Taken together, these advances could significantly improve the ease-of-use and practicality of mobile arm-tracking applications.

Uniquely, we take advantage of recent ultra-wideband (UWB) functionality to capture the distance between the smartwatch and smartphone. We fuse this with more conventional IMU-reported data, specifically watch and phone orientation and acceleration. Importantly, the UWB measurements are absolute, and less drift-prone as compared to relative inertial data used in most prior work. As a consequence, our tracking output is more stable and accurate. In our user study, our system achieved a median positional error of 11.0 cm for the wrist and elbow joints, which is comparable to prior work requiring uncommon hardware.

The contributions of this work are as follows: 1) A self-contained and software-only approach for arm pose tracking combining UWB and IMU data, which is unique in the literature. 2) A functional, real-time implementation using two common off-the-shelf devices. 3) User studies that show our system achieves the best tracking accuracy among systems that do not require special user instrumentation or external infrastructure. 4) Open-sourced, synchronized UWB/IMU/Kinect data, processing pipeline, and trained models. <https://github.com/FIGLAB/SmartPoser>.

2 RELATED WORK

In this section, we review past systems for arm pose tracking. In addition to arm tracking, we also discuss systems for whole-body pose tracking that have used comparable sensing technologies. We first cover systems that use UWB for pose tracking, before moving to IMU-based systems. We then conclude with systems most similar to our own — ones that combine inertial data with some form of distance-based measurement.

2.1 UWB-Based Pose Tracking

Our system takes advantage of ultra-wideband (UWB) ranging. This technique uses the time-of-flight of a challenge-response transmission to determine the distance between two devices. Unlike other more popular RF technologies, such as WiFi or Bluetooth, UWB operates at a very low energy level and instead of varying frequency or power, the technology transmits information at specific times using a large bandwidth. This results in multiple beneficial properties such as high immunity to multipath fading, high data throughput, and excellent time-domain resolution, allowing for more accurate localization and tracking, down to centimeters [12].

Due to the strengths just described, UWB has become a popular approach for pose tracking systems due to its ability to accurately measure the distance between devices in space. A common setup

is to place UWB beacons on the limbs and torso of a user and instrument an environment with one or more receivers to measure the time-of-flight of signals transmitted from the beacons [29, 57]. Researchers have also created a UWB radar with a large, MIMO antenna array that can capture a radio “image” of the user’s pose [44]. When compared to cameras, UWB can be more privacy-sensitive and can operate through walls and other occlusions [44].

Beyond UWB, there have also been other RF technologies used for pose tracking, many of which share similar advantages to those just described for UWB. In comparable setups, researchers have placed a transmitter in the environment and tracked the pose of a user (often uninstrumented) by measuring signal reflections. Examples of such systems include ones that use RFID scanners and tags [21, 49], mmWave radar [40], and even WiFi [36, 37]. One drawback of systems that use purely RF ranging is that although they are able to measure the distance to different parts of the body, it is difficult to distinguish fine limb orientations such as the rotation of the wrist. For this reason, RF systems are often combined with IMU sensors that can provide information on limb orientation [9, 15, 63].

2.2 IMU-Based Wearable Pose Tracking

The idea of tracking pose with a wearable system of inertial sensors did not surface until the miniaturization of magnetic, angular-rate, and gravity (MARG) sensors in the late 1990s [4]. Pioneering systems had to deal with the problem of fusing measurements from multiple instruments to track limbs, all the while minimizing drift. In addition, these systems also had difficulty detecting and correcting for erroneous measurements such as magnetic disturbances [4, 28]. The two dominant approaches that were developed to solve these problems include Complimentary filters by Bachman et al. [4] and Kalman filters by Marins et al. [28]. While these two seminal works would develop the theoretical foundations for tracking pose with inertial data, many others went on to take these ideas and develop them into real-world systems [25, 31, 60, 62]. The commercial Xsens system can be seen as a current peak of this classical filtering and optimization-based approach [38]. By using a suit embedded with 17 IMUs located all over the body, this system is able to track a wide range of motions at 120 FPS [38].

In recent years, a new class of systems has emerged which are able to use significantly fewer IMUs to track pose with a similar level of accuracy [8, 20, 41, 42, 50, 58, 59]. A breakthrough and common theme in these systems has been modeling inertial data over multiple frames in time, often with neural networks. A seminal system is Sparse Inertial Poser (SIP) by Marcard et al., which was able to drop the sensor burden from 17 IMUs down to 6 by using an optimization-based method [48]. Building upon this, Huang et al.’s Deep Inertial Poser (DIP) [16] removed the constraint in SIP of needing to access the entire sequence of data and instead only used a subset of nearby frames in a bidirectional RNN (Bi-RNN). More recently, TransPose [59] and Physical Inertial Poser (PIP) [58] advanced DIP’s Bi-RNN approach, improving fidelity and framerate (to 90 FPS). This was primarily achieved by decomposing the model into sub-tasks that are easier to solve and combining the outputs to get user pose and global translation. More recently, IMU-Poser [32] used IMUs already found in smartphones, smartwatches

and earbuds to produce a best-guess pose from one to three points of instrumentation.

Of the IMU-based systems, the most similar to ours is ArmTrak, developed by Shen et al. [42]. Similar to SmartPoser, the system focused on tracking arm pose (instead of full-body pose) and did this by using a single IMU on the wrist instead of multiple IMUs placed across the body. Unlike the later deep learning methods, they used a hidden Markov model to probabilistically estimate next states for arm joint positions. This approach produced good arm pose estimates using a single IMU. However, the method was slow (with reduced accuracy when run in real-time) and could not track users on the move (only evaluated with feet planted after calibration). We detail our improvements over ArmTrak, particularly tracking results, in subsection 5.5.

2.3 Fusing Inertial and Distance Measurements for Wearable Pose Tracking

One of the greatest challenges in working with IMUs, particularly consumer-grade components, is noise and drift. For this reason, it is often advantageous to have a secondary sensor stream that can provide absolute position or orientation data. In the context of body tracking, worn systems have combined IMU data with magnetic field sensors [39, 55], ultrasonic transmitters and microphones [19, 26, 47], infrared distance sensors [56], micro-flow sensors [27], acoustic ranging [61], and optical linear encoders [33]. As previously mentioned, UWB beacons have also been used for this purpose, albeit not in a wearable form factor [9, 15, 63]. For each of these sensing approaches, there are important trade-offs to consider such as cost, robustness to occlusion, and across-environment reliability. Yet despite their differences, these sensors all share the same operating principle of capturing absolute measurements to augment the relative measurements of worn IMUs. Most importantly, all of the aforementioned systems in this section require new and special devices to be worn by the user. A key contribution of this work is showing this ranging+IMU fusion approach can now be achieved on popular commodity devices that users already own.

3 IMPLEMENTATION

Our motivation in developing SmartPoser was to track arm pose using only a smartphone and smartwatch (Figure 1). Figure 2 provides a high-level system overview and data flow. We now describe the core components of our system.

3.1 Proof-of-Concept Devices

We selected an iPhone 13 Pro (iOS 16.4) and an Apple Watch Series 7 (watchOS 9.4) as a pair of example, off-the-shelf, and popular smart devices. This choice was also motivated by Apple’s mature UWB and inter-device communication APIs that facilitated our development. Android lags behind in this regard, with Google adding UWB support via a Jetpack library only in late 2022 [11]. This is to support applications like lost object tracking, unlocking cars, and AR localization. Fortunately, the number of UWB-capable smart devices has steadily grown in recent years [53], and the feature looks to become increasingly pervasive.

3.2 UWB Ranging & IMU Data

In our custom app, we use the iOS Nearby Interaction API [2] to set up a UWB ranging session between the watch and phone. Once the session has been established, we receive distance updates at approximately 5 Hz. We note that iPhones offer higher sampling rates and even azimuth and elevation readings when a UWB device is located within a narrow cone-shaped field-of-view projecting from the rear of the device. Unfortunately, the field of view is too narrow for us to make practical use of this capability.

To get data from the IMUs onboard the phone and watch, we used Apple’s Core Motion API [1], which is standard across all iOS devices. Using the API, we acquire data about the current orientation quaternions and acceleration measurements of each device at roughly 25 FPS. We use this as the native frame rate of our system, upsampling the 5 Hz UWB data to the more responsive 25 Hz IMU rate by duplicating the most recent UWB value. Finally, we note that although the two devices can estimate the gravity vector and magnetic north, we found it was more reliable to explicitly align the two devices’ frames of reference. For this, we require the devices to be placed side by side, which could happen e.g., in a charging dock overnight.

3.3 UWB Correction

In early piloting, we saw that UWB measurements exhibited systematic offsets and characteristic noise, especially for certain arm pose regions. This is chiefly due to loss of line-of-sight between the two devices due to body occlusion (e.g., putting the arms behind the head). The example data in Figure 3 illustrates this effect. To capture this data, the authors placed the iPhone 13 Pro smartphone in their left-front pants pocket, and wore the Apple Watch Series 7 smartwatch on their left wrist. The authors then performed a range of exemplary arm motions. During this process, the phone-to-watch distance was recorded (Figure 3, blue line). To provide a ground truth distance, we used an Azure Kinect and its native SDK to capture a 3D skeleton of the user in real-world units. Using this skeleton, we “attached” a virtual phone and watch, and compute the distance between the two (Figure 3, orange line).

Given the offset and noise is semi-structured, we can compensate for some error, especially if we take advantage of orientation data

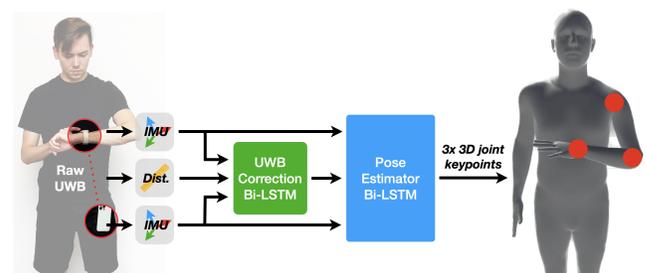


Figure 2: SmartPoser system overview. Our model has 46.4k parameters in total, with 3k parameters for the UWB Correction model and 43.4k parameters for the Pose Estimator model.

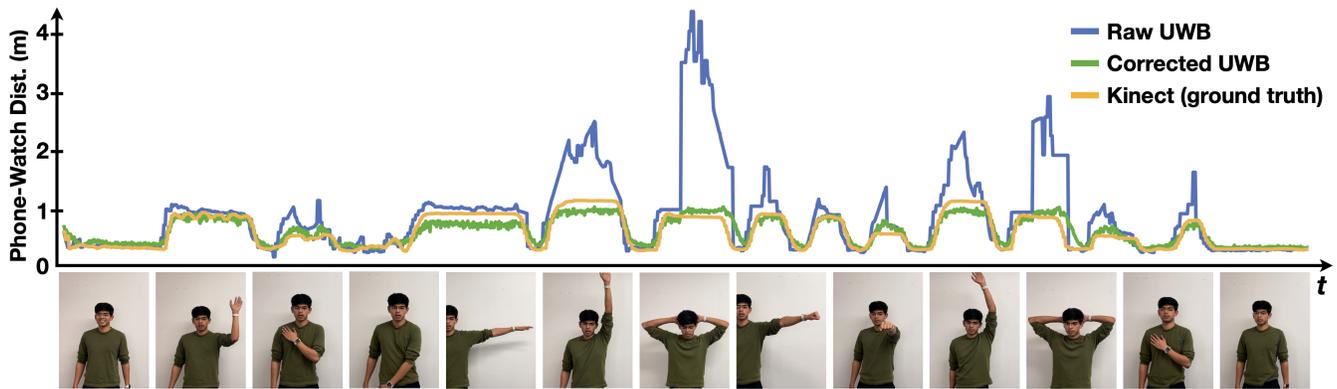


Figure 3: Example user motion sequence over one minute. Raw UWB distance is plotted in blue — note the significant errors, far exceeding possible arm lengths, when the watch loses line-of-sight to the phone. Our corrected UWB stream is plotted in green. Ground truth distance is provided by an external Azure Kinect is shown in orange.

from the IMUs, which already provide some cues as to the orientation and location of the arms. For this, we designed a UWB correction network, which we implemented as a 2-layer bi-directional LSTM with 8 hidden units each. The input to the RNN is a 125-long sequence (5 seconds of data at our system’s native framerate of 25 FPS) of raw UWB distance measurements, plus phone and watch orientations (3x3 rotation matrix format) and accelerations (X/Y/Z), for a total feature dimension of 25. We scale the UWB distance measurements by the wearer’s arm span. The scaling ensures that the model is able to generalize to various body sizes. We also scale the acceleration values by 30 in order to make them suitable for training, following prior work [58, 59]. The RNN outputs a sequence of corrected distance measurements. The correcting effect of this RNN is illustrated in Figure 3, green line. In this example sequence, mean UWB tracking error is reduced from 34.5 cm to 6.7 cm, and is considerably more stable as well.

3.4 Pose Estimation

Our pose estimation RNN ingests corrected UWB data along with raw IMU data (watch and phone orientation and acceleration). Taking inspiration from Deep Inertial Poser [16] and TransPose [59], this network is also implemented as a 2-layer bi-directional LSTM, with 32 hidden units each. Watch orientation and acceleration are normalized such that they are relative to the phone’s frame of reference. Along with having an absolute distance measurement, using relative orientations helps to make the system more robust against IMU drift. As with UWB correction, we scale the acceleration values by 30. We use the same input vector format as our UWB correction network – a 125-length sequence of UWB distances, phone and watch orientations and accelerations. This 125-length sequence is a rolling input buffer, and new predictions are made in real-time whenever a new sample arrives in the buffer (at 25 Hz).

This RNN outputs a 125-long arm pose sequence, containing 3D joint locations for the wrist, elbow, and shoulder. The predicted joint locations are 9 values normalized to the user’s arm span, which we can “reproject” into real-world units by the user’s arm span. Rather than taking the most recent pose prediction (i.e.,

the 125th frame), we use the 120th frame. Functionally, this means we are predicting 200 ms behind real-time. This approach to use the n-5th output frame was directly inspired by TransPose [59], and offers a compromise between latency vs. accuracy. Note, during testing we also used the n-5th frame of ground truth to match with our prediction frame.

3.5 Model Training

In summary, SmartPoser consists of two machine learning elements – UWB correction and then pose estimation. We train the full system end-to-end using the Adam optimizer, with a learning rate of $3e-4$. The UWB corrector is trained to regress to the Kinect-derived distances (Section 3.3), scaled by user arm span, and using mean squared error loss. The pose estimator network is trained to regress to arm joint positions (shoulder, elbow, and wrist) using a mean-per-joint-position-error (MPJPE) loss. Kinect-derived ground truth arm joint positions are preprocessed to have a shoulder frame of reference (i.e., ignoring world body rotation).

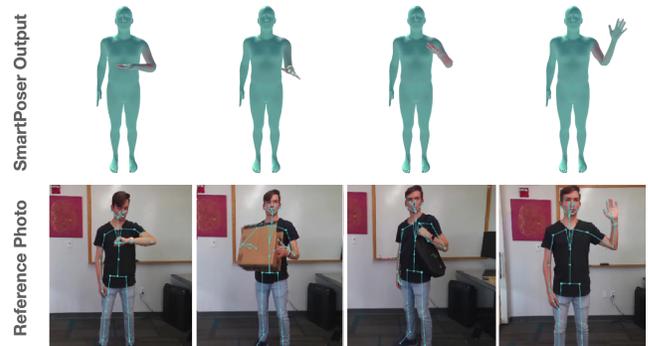


Figure 4: Example set of predicted poses. Vertex error (vs. Kinect ground truth) color coded.



Figure 5: The 20 terminal arm poses utilized in our user study. Note that participants were asked to transition between these poses and all intermediate data was captured for evaluation (i.e., not just the terminal poses).

The total loss of our system is the sum of the UWB corrector loss plus the pose estimator loss. The system is implemented using PyTorch and the PyTorch Lightning frameworks. For efficient training, we train the model to predict a full motion window of 125 samples (5 seconds) rather than just the last frame, following the method in [16, 59]. In other words, during training we use non-overlapping windows, however for testing we overlap windows by all but one new frame.

3.6 Model Inference and SmartPoser Output

To facilitate development and debugging, UWB and IMU data is streamed to a laptop (M1 MacBook Air 2021). On this machine, our trained model takes 8.1 ms per inference (on the CPU, using pytorch), meaning we output pose predictions as fast as data arrives (25 FPS). To verify SmartPoser could run entirely on-device, we also converted our trained model to CoreML with mixed precision compute and found it took only 0.99 ms per inference on an iPhone 12 Pro. Although we envision our arm pose data most often being used internally and unseen (discussion in Section 6), we did build a basic visualization tool, seen Figures 1, 2, and 4.

4 EVALUATION

To evaluate the performance of SmartPoser, we conducted a user study with 10 participants (mean age 25.4; 3 identified as female, 7 as male). The study took place in a typical indoor office environment, lasted approximately 45 minutes, and compensated \$20. The study was divided into two data collection phases. The first phase focused on arm motions a user might perform in the course of their daily activities. Phase two captured less ecologically valid, but more varied and challenging gross motor arm poses (Figure 5).

Before participants wore our test devices, the smartphone and smartwatch were placed side-by-side, as they might rest in a charging dock. We use this to align their independent reference frames into a common, global frame of reference. We then asked participants to wear the smartwatch on their left wrist and to place the smartphone in their left front pocket (see Figure 1), which is a common placement [17, 35] (though we note other arrangements are possible, which we discuss in Limitations). Before each study

phase began, we had users perform a T-pose with their arms (palms facing down, such that the watch face is parallel to the ground) in order to capture the relative orientations of the two devices on the body (and importantly, the same study procedure used in IMUPoser [32], Transpose [59], Deep Inertial Poser [16], and Physical Inertial Poser [58]).

4.1 Daily Activities Procedure

The first phase of data collection focused on arm poses that users might assume during daily routines. To capture such data, we had participants complete an "obstacle course" style study (see e.g., [6, 18]). We selected 20 brief mock activities, setup at stations within our study environment. These activities were divided into seven categories: *Domestic* (ironing, open jar, vacuum floor, wipe table), *Grooming* (comb hair, wipe forehead, wash with loofa), *Food & Beverage* (stir pot, pour teapot, drink from cup), *Moving Items* (move grocery bag, lift box above head, wear bag on shoulder), *Exercise* (jumping jacks, biceps curls), *Home DIY* (use screwdriver, use paint roller on wall), and *Miscellaneous* (check time on watch, read magazine, flick light switch). Although this set of activities is not comprehensive, we have confidence our system did not overfit. For example, the Demo Activities segment of our Video Figure shows predictions of unseen poses resulting from naturalistic activities. This is a good signal of expected performance for poses and activities not used in training and matches our anecdotal observations throughout the course of development.

A large TV screen was positioned in the study area with a list of the activities. In addition to this visual prompt, the experimenter also spoke aloud the next activity to perform, and controlled the visual display. Although some activities were bimanual, in general we asked participants to use their left arm, as it was the one being tracked. All activities were performed one after another while data was continuously recorded, thus beyond the activities themselves, we also captured the transitions between them and miscellaneous organic gestures such as participants adjusting their clothes. Participants completed this procedure twice, taking roughly 10-15 minutes in total.

4.2 Gross Motor Arm Pose Procedure

Real-world activities tend to be performed in front of the body, or at least close to the torso. As such, they tend to have less variance and less extreme arm positions. Thus, as a compliment to our phase one data collection, we also captured a more gross-motor-oriented arm pose dataset. For this, we use a set of 20 terminal arm poses, seen in Figure 5. These exemplify a variety of spatial endpoints, joint positions, and joint orientations.

As with phase one, we used a TV screen to prompt participants with a random sequence of terminal poses (20 poses, 10 repeats, random order). Importantly, data was continuously captured as participants transitioned between these poses. Thus, instead of capturing 20 poses, our study captures transitions between $\sim 20C_2$ pose pairs, so the effective variety and volume of motions we captured was large. In addition, the amount of time spent in the terminal poses is very low, avoiding bias. Transitioning through the 200-pose sequence took approximately 10-15 minutes.

4.3 Data Capture & Ground Truth

Throughout both phases of the study, we used an Azure Kinect [46] to capture ground truth data on joint positions and orientations of participants' arms. The Kinect ran continuously, and whenever we got a new frame of UWB+IMU sensor data (at ~ 25 Hz), we paired it with the most recent Kinect frame. Across all participants, our activities dataset contained 151k frames of data (representing roughly 105 minutes of data), and our gross motor dataset contained 129K frames of data (~ 90 minutes).

5 RESULTS & DISCUSSION

5.1 Train/Test Procedure

We evaluated our model using a leave-one-participant-out cross validation scheme (i.e., ten-fold cross validation with whole participants' data as folds). This simulates "out-of-the-box" accuracy that requires only a single calibratory pose and no training data from the wearer (whereas much prior work requires per-user training data, or even per-worn-session training data). While we train on real-world activities and gross motor arm motions datasets combined, we break these results out where appropriate to give further insight into real-world accuracy.

5.2 Pose Tracking Accuracy

Across all participants, we found a median error (averaging elbow and wrist joints) of 11.0 cm (SD=0.9 cm). As we move along the kinematic chain, error unsurprisingly increases — median error by joint: shoulder 0.17 cm (SD=0.02 cm), elbow 8.7 cm (SD=0.9 cm), and wrist 13.3 cm (SD=0.98 cm). Looking at error broken out by gross motor arm poses vs. real-world activities, we find median error (elbow and wrist) of 11.2 cm (SD=2.3 cm) and 11.1 cm (SD=1.3 cm) respectively. Figure 6 provides a summary of these results for median error. Other results metrics for the combined data set of gross motions and real-world activities can be found in Table 1, and the calculated CDFs in Figure 8. Our results already used a leave-one-user-out cross validation and had a low STD of 0.9 cm across users.

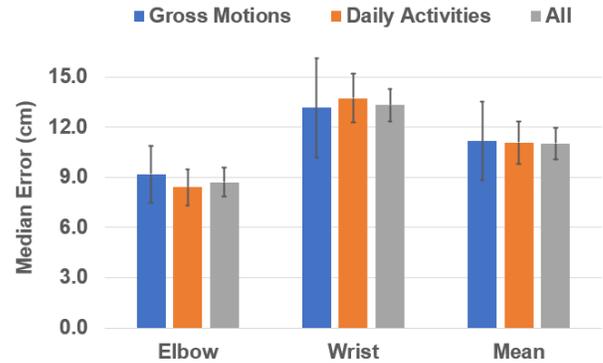


Figure 6: Median per-joint positional error was 11.0 cm averaged across all study participants for the elbow and wrist joints. For the gross motions phase, median per-joint positional error was 11.2 cm, 9.2 cm for just the elbow, and 13.2 cm for the wrist. For the real-world activities phase, median per-joint position error was 11.1 cm, 8.4 cm for just the elbow, and 13.7 for the wrist. Error bars are standard deviation (cm).

Joint	Elbow		Wrist		Mean	
	Mean	STD	Mean	STD	Mean	STD
Cross-user metric (cm)						
Median error	8.7	0.9	13.3	1.0	11.0	0.9
Mean per-joint positional error	10.0	0.9	15.1	1.2	12.6	1.1
Mean absolute error	15.0	1.4	22.4	1.7	18.7	1.5

Table 1: Calculated metrics for all arm joints for the combined data set of gross motions and daily activities. All metrics are reported in cm.

However, to further verify we were not overfitting, we also computed metrics testing on the training set specifically and observed a median wrist error of 11.9 cm (vs. 13.3 cm from cross validation).

5.3 Spatial Error

As discussed earlier, UWB ranging is sensitive to occlusion (in our case, body occlusion). Our UWB Correction RNN (Section 3.3) compensates for some of this error, but not entirely. Figure 7 compares mean euclidean error of the raw UWB distances (21.0 cm) vs. our corrected UWB values (8.6 cm), broken out by study dataset. We also generated a scatter plot (Figure 7) plotting our corrected UWB distance vs. the Kinect-derived ground truth distance for all 131k frames of study data from our gross arm pose procedure (representing ~ 1.5 hours of data). We include a line of best fit, which has an $R^2 = 0.72$.

Although the correlation is strong, there are other error modes and behaviors visible. To further investigate this, we created a series of heatmaps visualizing the maximum UWB error in space around the body using our study data. We look at arm pose data from all

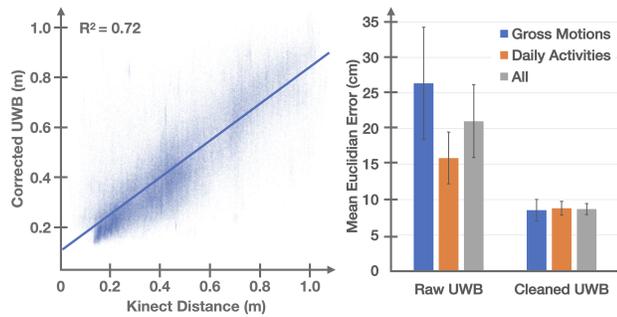


Figure 7: Left: Plot of corrected UWB distance vs. Kinect-derived distance (280k datapoints). Right: UWB Correction RNN performance across study datasets. Mean Euclidean error reduces from 21.0 cm to 8.6 cm on average. Error bars are standard deviation.

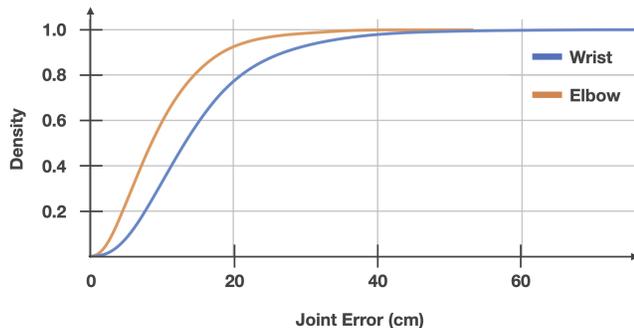


Figure 8: Cumulative distribution functions (CDF) of the elbow and wrist joint errors.

participants combined, and then only the wrist joint, which had the largest addressable volume and largest median error.

Figure 9 provides three heatmaps – a frontal, side, and top-down view of a wearer. Heatmaps are shown for the error of raw UWB distance, corrected UWB distance, and joint predictions versus Kinect-derived ground truth. All the heatmaps have some common error hot spots (red-orange regions in Figure 9), while the volume in the front and to the left side of the body is reasonably accurate. Most of the regions show up in blue, corresponding to an error of between 0-15 cm. Wrist locations that could lead to potential self-occlusion, such as the right thigh and the space above the left shoulder, intuitively show up as highly erroneous regions in the raw UWB distance heatmaps. Note how our UWB correction module is able to account for most of these errors. The error-prone areas for the wrist joint error map well with the error-prone regions for corrected UWB distance and are altogether much much lower than the original errors for raw UWB distance.

5.4 UWB Ablation Study

The primary contribution of SmartPoser is the introduction of UWB distance data to improve arm tracking results. To quantify how

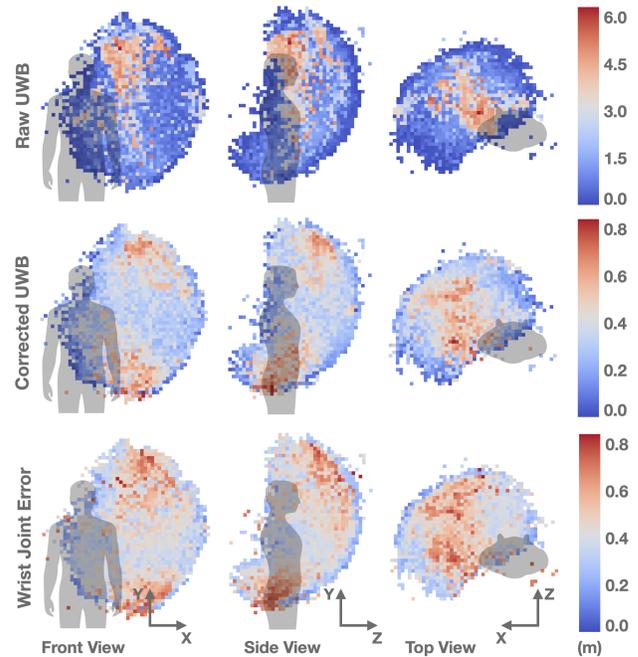


Figure 9: Three rows of heatmaps visualizing maximum absolute Euclidean distance error for raw UWB distances (i.e., not corrected), corrected UWB distances, and predicted wrist joint position. For all measurements, error (in meters) is calculated with respect to Kinect-derived ground truth. Note the scale for the raw UWB distances (top row) is different than the bottom two rows.

much the improvement was, we ran an ablation study of our complete UWB+IMU system versus one that used only IMU data. The results of this study can be seen in Figure 10.

Averaging the results from both the gross motions and daily activities phases, we observed a median error of 11.7 cm for the elbow and 16.7 cm for the wrist for the model with only IMU data. Compared to our standard results from the model with UWB+IMU, this meant error was 3.0 cm (25.6%) greater for the elbow and 3.4 cm (20.4%) greater for the wrist when using only IMU. Additionally, when looking at pose outputs, we find UWB+IMU is much more stable. For example, when holding static poses with IMUs alone they tend to "droop" slowly to a mean pose. These results show that the contribution of UWB distance measurements is significant to SmartPoser and a key to helping enable higher-fidelity pose tracking.

5.5 Comparison to Prior Work

To contextualize our results, we compare SmartPoser accuracy to other arm tracking systems that also do not require special instrumentation (i.e., no new or exotic sensors worn on the body) or external infrastructure (i.e., entirely self-reliant and work on-the-go). While these systems were evaluated on different datasets, we believe it is applicable to compare them to SmartPoser.

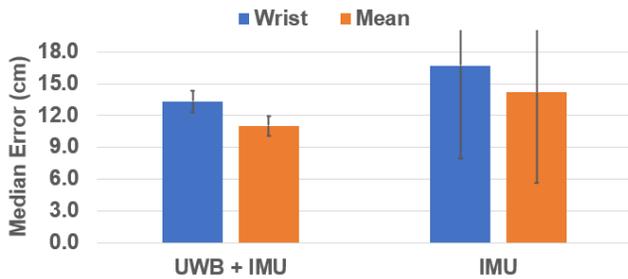


Figure 10: Ablation results showing the effectiveness using both UWB and IMU data for pose estimation. Results are calculated on data from both phases of the user study. When adding UWB data to the model for IMU alone, median error for both joints decreases from 14.2 cm to 11.0 cm.

ArmTrak [42] achieved a real-time median error of 12.0 cm for elbow and 13.3 cm for wrist while performing free motions with the arm (accuracy was higher when using an offline model). Of note, ArmTrak required participants to stand still (i.e., feet planted without the ability to change body orientation), unlike SmartPoser. IMUPoser [32] did not report median error, but achieved a combined real-time mean error of 21.6 cm for all arm joints and 26.9 cm for just the wrist. MUSE [41] only estimated position at the wrist for 5 participants and achieved a real-time (10 Hz on a desktop computer) median error of 13.0 cm for motions when the body moves around. Compared to these works, our system achieved a real-time (25 Hz) median error of 11.0 cm for the shoulder and wrist, and 13.3 cm for just the wrist.

6 EXAMPLE USES

Arm pose tracking (and body pose more broadly) has been well-motivated in prior work. Our system does not enable any new applications per se, but rather makes them more practical – able to run on off-the-shelf consumer hardware many people already own. For this reason, we do not re-implement any application demos, and instead point to the literature for motivating use cases. Please see our Video Figure for real-time demos.

Fitness - Arms are involved in many exercises, and thus tracking the arms has obvious and immediate utility for fitness-oriented applications. The most straightforward are applications that automatically track rep counts for lifting free weights, jumping jacks, jump rope, and similar activities. One could also imagine a smartwatch assistant verbally guiding a user through a tai chi routine, automatically advancing steps. It may even be possible to track exercise form for personalized feedback; for example, proper extension of the arm when hitting a punching bag or form when shooting a basketball.

Rehabilitation - Tracking of the arms could also prove useful in rehabilitation (e.g., rotor cuff injuries). For example, continuous background tracking could supplement periodic sessions with physical and occupational therapists by monitoring arm metrics such as total motion, duration lifted, maximum extension, and maximum height reached. It could even be that appointments are automatically scheduled if there is limited progress or performance loss in a

given metric. Alternatively, instead of passive, background monitoring, arm tracking could be employed in interactive rehabilitation applications. For example, an app could guide users through a series of balance exercises or reaching tasks. Such apps could not only automatically advance steps, but provide instructive feedback and encouragement.

Life Logging - Arm tracking could allow for new types of activities to be automatically tracked for life logging purposes. For example, it might be possible to track cigarette consumption (repeatedly raising one’s hands to the mouth with a particular cadence) as part of a smoking cessation effort. Likewise, detecting periods of food consumption may also be possible for dieting apps. It may also be possible to log domestic activities, such as cooking, cleaning, vacuuming, folding clothes, and putting away dishes.

Occupational Safety & Training - Many forms of manual labor employ the hands, such as carrying boxes, stacking groceries, picking vegetables, driving a truck, laying bricks, mowing lawns, painting a house, cleaning windows, etc., which could be tracked to ensure workers are not fatigued, exploited or injured, e.g., Repetitive Strain Injury (RSI) [14], Hand-Arm Vibration Syndrome (HAVS) [43]. It may even be possible to detect improper handling of power tools or heavy equipment to improve site safety. Training on how to properly handle a tool could also happen via a companion smartwatch app.

Gaming - The commercial success of titles like Nintendo Wii Sports (the fourth best-selling video game of all time [52]) and Xbox’s Dance Central franchise ushered in a new era of more physically-oriented games. While some mobile games have taken advantage of inertial or spatial data, none have leveraged arm pose to date. In particular, this could be an interesting opportunity to make smartwatch games larger and more interesting than their diminutive screens might otherwise allow. For instance, users could practice their golf swings or frisbee throws, with the watch reporting on achieved range and form quality. While we only considered arm pose tracking with the phone stored in the pocket, our technique should also extend to a phone held in the hand (with a different machine learning model). This would allow for gaming experiences with coordinated graphics. More specifically, the user would look at a game running on the phone held in one hand, with the other hand free to perform game actions, such as throwing a ball or fighting with a sword.

Context-Aware Interfaces & Assistants - Devices that are context-aware can automatically summon relevant interfaces and information to augment a user’s activity. Arm pose is a useful piece of information that could improve the accuracy and breadth of prior context sensing work (using e.g., environmental sounds [22], bioacoustics [24], movement data [23]). For instance, arm pose tracking could help disambiguate if a user is driving an automobile vs. being a passenger. As discussed in the above subsections on Life Logging and Occupational Safety and Training, there is a long tail of arm-centric activities that could be automatically detected, which could launch context-aware functionality.

7 OPEN SOURCE

To enable others to build upon our system, we have made our dataset, architecture, and trained model freely available at

<https://github.com/FIGLAB/SmartPoser>, with the gracious permission of our participants.

8 LIMITATIONS & FUTURE WORK

We believe our proof-of-concept implementation demonstrates the feasibility of our method, but there are nevertheless limitations of note. First among these limitations is that our current work only considers the case when the phone is stored in a front pocket, a common but not ubiquitous placement [17, 35]. In general, our pipeline should scale to other pocket locations, including rear pants pockets, and potentially even cycling jersey pockets and fanny packs. However, when the phone is taken out and held in the hand, and can exist in many possible positions with respect to the body, a third spatial reference would be required (potentially head position captured via the phone’s user-facing camera or earbuds with UWB). There already exists work that can automatically detect pocket vs. hand phone placement [13, 51], which could toggle between different models.

Another piece of information we need is which pocket a user typically stores their phone and which wrist the user wears a watch, which could be captured during a one-time setup wizard. Even with all this data, the user still needs to perform a calibration procedure, which is a limitation of our current implementation. This process is described in the Evaluation section, but briefly here, participants in our study had to hold the devices near each other, then assume a T-pose with their arm to orient the IMU reference frames (e.g. to offset how the phone is placed rotationally in the pocket). While inconvenient to perform once, we found the system was able to maintain its calibration for an extended period of time. We note the T-pose used is not special and could be replaced by a pose users more naturally perform, such as arm by side (potentially auto-detected using UWB distance). If the calibration process could be condensed to a single gesture, such as holding the arms by the side of the body, we imagine there are many contexts in which it would be acceptable to do this once at the beginning of an activity, such as starting a workout or a rehab session. This is one aspect where the enhanced tracking mode for the iOS UWB API mentioned in the Implementation section could be particularly useful. This mode provides both distance and 3D angles between the phone and watch. Thus, if the watch was held in the phone’s FoV cone, the phone would know its absolute position relative to the watch and could use this to use this data to calibrate the IMU reference frames automatically.

One additional issue is that many people do not carry their phone in their pocket at all [17, 35]. Feminine clothing lacking functional pockets [45] is a prominent example. People may opt to carry their phone in a bag or purse. Tracking watch position with a phone in a bag is a more challenging problem than a phone in a pocket because, while there is variation in pocket design, the phone is located in the same area on the body in the majority of cases. This is not true of bags where the space of possible phone locations is much larger. Since the phone could move inside bags, we do not expect SmartPoser to work well in such situations. However, as a solution, instead of using a phone, our system could generalize to other devices that are worn in static locations and may become UWB-enabled in the future, such as smart glasses or earbuds.

9 CONCLUSION

We have presented SmartPoser, a system that enables 3D arm pose tracking using an off-the-shelf smartphone and smartwatch in their normally-worn positions. In our approach, we build upon a state-of-the-art bidirectional LSTM model for arm pose prediction and fused more traditional inertial data with new UWB time-of-flight range measurements between the two devices. Independent of using commodity devices, this technique is novel as UWB ranging has never been used to augment IMU data in a wearable pose-tracking system before. Through a user study involving real-world activities and gross arm motions, our approach yielded a median positional error of 11.0 cm for the wrist and elbow joints, with no user data in the training set. This is a level of tracking fidelity that has only been achieved in the past through full-body instrumentation with IMUs or UWB-instrumented environments. Overall, by achieving this result in this consumer form factor, we believe SmartPoser provides the most realistic pathway to date for arm-based tracking and interactions to be adopted for widespread use.

ACKNOWLEDGMENTS

We would like to thank Riku Arakawa for his early help with application development. We are also grateful to Dr. David Lindlbauer and the Augmented Perception Lab at Carnegie Mellon University for sharing technical equipment and support.

REFERENCES

- [1] Apple. 2022. Core Motion | Apple Developer Documentation. <https://developer.apple.com/documentation/coremotion/>
- [2] Apple. 2022. Nearby Interaction | Apple Developer Documentation. <https://developer.apple.com/documentation/nearbyinteraction>
- [3] Apple. 2023. Set up Handwashing on Apple Watch. <https://support.apple.com/guide/watch/set-up-handwashing-apdc9b9f04a8/watchos>
- [4] Eric R. Bachmann, Robert B. McGhee, Xiaoping Yun, and Michael J. Zyda. 2001. Inertial and magnetic posture tracking for inserting humans into networked virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology (VRST '01)*. Association for Computing Machinery, New York, NY, USA, 9–16. <https://doi.org/10.1145/505008.505011>
- [5] Alejandro Baldominos, Yago Saez, and Cristina García del Pozo. 2015. An Approach to Physical Rehabilitation Using State-of-the-art Virtual Reality and Motion Tracking Technologies. *Procedia Computer Science* 64 (2015), 10–16. <https://doi.org/10.1016/j.procs.2015.08.457>
- [6] Ling Bao and Stephen S. Intille. 2004. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive Computing (Lecture Notes in Computer Science)*, Alois Ferscha and Friedemann Mattern (Eds.). Springer, Berlin, Heidelberg, 1–17. https://doi.org/10.1007/978-3-540-24646-6_1
- [7] Frédéric Bosché, Mohamed Abdel-Wahab, and Ludovico Carozza. 2016. Towards a Mixed Reality System for Construction Trade Training. *Journal of Computing in Civil Engineering* 30, 2 (March 2016), 04015016. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000479](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000479) Publisher: American Society of Civil Engineers.
- [8] Hammad Tanveer Butt, Bertram Taetz, Mathias Musahl, Maria A. Sanchez, Pramod Murthy, and Didier Stricker. 2021. Magnetometer Robust Deep Human Pose Regression With Uncertainty Prediction Using Sparse Body Worn Magnetic Inertial Measurement Units. *IEEE Access* 9 (2021), 36657–36673. <https://doi.org/10.1109/ACCESS.2021.3062545> Conference Name: IEEE Access.
- [9] J. A. Corrales, F. A. Candelas, and F. Torres. 2008. Hybrid tracking of human operators using IMU/UWB data fusion by a Kalman filter. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 193–200. <https://doi.org/10.1145/1349822.1349848> ISSN: 2167-2148.
- [10] Han Ding, Jinsong Han, Longfei Shanguan, Wei Xi, Zhiping Jiang, Zheng Yang, Zimu Zhou, Panlong Yang, and Jizhong Zhao. 2017. A Platform for Free-Weight Exercise Monitoring with Passive Tags. *IEEE Transactions on Mobile Computing* 16, 12 (Dec. 2017), 3279–3293. <https://doi.org/10.1109/TMC.2017.2691705> Conference Name: IEEE Transactions on Mobile Computing.
- [11] Google. 2022. Android Developers, Ultra-wideband (UWB) communication. <https://developer.android.com/guide/topics/connectivity/uwb>
- [12] Bernhard Großwindhager, Carlo Alberto Boano, Michael Rath, and Kay Römer. 2018. Enabling Runtime Adaptation of Physical Layer Settings for Dependable

- UWB Communications. In *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 01–11. <https://doi.org/10.1109/WoWMoM.2018.8449776>
- [13] Chris Harrison and Scott E. Hudson. 2008. Lightweight material detection for placement-aware mobile computing. In *Proceedings of the 21st annual ACM symposium on User interface software and technology (UIST '08)*. Association for Computing Machinery, New York, NY, USA, 279–282. <https://doi.org/10.1145/1449715.1449761>
- [14] P Helliwell and W Taylor. 2004. Repetitive strain injury. *Postgraduate Medical Journal* 80, 946 (Aug. 2004), 438–443. <https://doi.org/10.1136/pgmj.2003.012591>
- [15] Jeroen D. Hol, Fred Dijkstra, Henk Luinge, and Thomas B. Schon. 2009. Tightly coupled UWB/IMU pose estimation. In *2009 IEEE International Conference on Ultra-Wideband*. 688–692. <https://doi.org/10.1109/ICUWB.2009.5288724>
- [16] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J. Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 1–15. <https://doi.org/10.1145/3272127.3275108>
- [17] Fumiko Ichikawa, Jan Chipchase, and R. Grignani. 2005. Where's the phone? A study of mobile phone location in public spaces. *Mobile Technology, Applications and Systems, 2005 2nd International Conference on* (Jan. 2005), 1–8.
- [18] Stephen S. Intille, Ling Bao, Emmanuel Munguia Tapia, and John Rondoni. 2004. Acquiring in situ training data for context-aware ubiquitous computing applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/985692.985693>
- [19] Silje Ekroll Jahren, Niels Aakvaag, Frode Strisland, Andreas Vogl, Alessandro Liberale, and Anders E. Liverud. 2021. Towards Human Motion Tracking Enhanced by Semi-Continuous Ultrasonic Time-of-Flight Measurements. *Sensors* 21, 7 (Jan. 2021), 2259. <https://doi.org/10.3390/s21072259>
- [20] Jiayi Jiang, Paul Strelci, Huajian Qiu, Andreas Fender, Larissa Laich, Patrick Snape, and Christian Holz. 2022. AvatarPoser: Articulated Full-Body Pose Tracking from Sparse Motion Sensing. <http://arxiv.org/abs/2207.13784> arXiv:2207.13784 [cs].
- [21] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I. Hong. 2018. Towards Wearable Everyday Body-Frame Tracking using Passive RFIDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (Jan. 2018), 145:1–145:23. <https://doi.org/10.1145/3161199>
- [22] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. Association for Computing Machinery, New York, NY, USA, 213–224. <https://doi.org/10.1145/3242587.3242609>
- [23] Gierad Laput and Chris Harrison. 2019. Sensing Fine-Grained Hand Activity with Smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300568>
- [24] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 321–333. <https://doi.org/10.1145/2984511.2984582>
- [25] Tommaso Lisini Baldi, Francesco Farina, Andrea Garulli, Antonio Giannitrapani, and Domenico Prattichizzo. 2020. Upper Body Pose Estimation Using Wearable Inertial Sensors and Multiplicative Kalman Filter. *IEEE Sensors Journal* 20, 1 (Jan. 2020), 492–500. <https://doi.org/10.1109/JSEN.2019.2940612> Conference Name: IEEE Sensors Journal.
- [26] Huajun Liu, Xiaolin Wei, Jinxiang Chai, Inwoo Ha, and Taehyun Rhee. 2011. Realtime human motion control with a small number of inertial sensors. In *Symposium on Interactive 3D Graphics and Games (I3D '11)*. Association for Computing Machinery, New York, NY, USA, 133–140. <https://doi.org/10.1145/1944745.1944768>
- [27] Shi Qiang Liu, Jun Chang Zhang, and Rong Zhu. 2020. A Wearable Human Motion Tracking Device Using Micro Flow Sensor Incorporating a Micro Accelerometer. *IEEE Transactions on Biomedical Engineering* 67, 4 (April 2020), 940–948. <https://doi.org/10.1109/TBME.2019.2924689> Conference Name: IEEE Transactions on Biomedical Engineering.
- [28] J.L. Marins, Xiaoping Yun, E.R. Bachmann, R.B. McGhee, and M.J. Zyda. 2001. An extended Kalman filter for quaternion-based orientation estimation using MARG sensors. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, Vol. 4. 2003–2011 vol.4. <https://doi.org/10.1109/IROS.2001.976367>
- [29] Zemene W. Mekonnen, Eric Slotkce, Heinrich Luecken, Christoph Steiner, and Armin Witteben. 2010. Constrained maximum likelihood positioning for UWB based human motion tracking. In *2010 International Conference on Indoor Positioning and Indoor Navigation*. 1–10. <https://doi.org/10.1109/IPIN.2010.5647912>
- [30] Jochen Meyer, Jutta Fortmann, Merlin Wasmann, and Wilko Heuten. 2015. Making Lifelogging Usable: Design Guidelines for Activity Trackers. In *MultiMedia Modeling (Lecture Notes in Computer Science)*, Xianguan He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and Muhammad Abul Hasan (Eds.). Springer International Publishing, Cham, 323–334. https://doi.org/10.1007/978-3-319-14442-9_39
- [31] Markus Miezal, Gabriele Bleser, Norbert Schmitz, and Didier Stricker. 2013. A generic approach to inertial tracking of arbitrary kinematic chains. In *Proceedings of the 8th International Conference on Body Area Networks (BodyNets '13)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 189–192. <https://doi.org/10.4108/icst.bodynets.2013.253608>
- [32] Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 2023. IMUPoser: Full-Body Pose Estimation using IMUs in Phones, Watches, and Earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3544548.3581392>
- [33] Kim Doang Nguyen, I-Ming Chen, Zhiqiang Luo, Song Huat Yeo, and Henry Been-Lirn Duh. 2011. A Wearable Sensing System for Tracking and Monitoring of Functional Arm Movement. *IEEE/ASME Transactions on Mechatronics* 16, 2 (April 2011), 213–220. <https://doi.org/10.1109/TMECH.2009.2039222> Conference Name: IEEE/ASME Transactions on Mechatronics.
- [34] Vijay Rajanna, Raniero Lara-Garduno, Dev Jyoti Behera, Karthic Madanagopal, Daniel Goldberg, and Tracy Hammond. 2014. Step up life: a context aware health assistant. In *Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health (HealthGIS '14)*. Association for Computing Machinery, New York, NY, USA, 21–30. <https://doi.org/10.1145/2676629.2676636>
- [35] Mary Redmayne. 2017. Where's Your Phone? A Survey of Where Women Aged 15–40 Carry Their Smartphone and Related Risk Perception: A Survey and Pilot Study. *PLoS ONE* 12, 1 (Jan. 2017), e0167996. <https://doi.org/10.1371/journal.pone.0167996>
- [36] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Winect: 3D Human Pose Tracking for Free-form Activity Using Commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2022), 176:1–176:29. <https://doi.org/10.1145/3494973>
- [37] Yili Ren, Zi Wang, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. GoPose: 3D Human Pose Estimation Using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (July 2022), 69:1–69:25. <https://doi.org/10.1145/3534605>
- [38] Daniel Roetenberg, Henk Luinge, and Per Slycke. 2013. Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors. (2013), 9.
- [39] Daniel Roetenberg, Per J. Slycke, and Peter H. Veltink. 2007. Ambulatory Position and Orientation Tracking Fusing Magnetic and Inertial Sensing. *IEEE Transactions on Biomedical Engineering* 54, 5 (May 2007), 883–890. <https://doi.org/10.1109/TBME.2006.889184> Conference Name: IEEE Transactions on Biomedical Engineering.
- [40] Arindam Sengupta, Feng Jin, Renyuan Zhang, and Siyang Cao. 2020. mm-Pose: Real-Time Human Skeletal Posture Estimation Using mmWave Radars and CNNs. *IEEE Sensors Journal* 20, 17 (Sept. 2020), 10032–10044. <https://doi.org/10.1109/JSEN.2020.2991741> Conference Name: IEEE Sensors Journal.
- [41] Sheng Shen, Mahanth Gowda, and Romit Roy Choudhury. 2018. Closing the Gaps in Inertial Motion Tracking. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (New Delhi, India) (MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 429–444. <https://doi.org/10.1145/3241539.3241582>
- [42] Sheng Shen, He Wang, and Romit Roy Choudhury. 2016. I am a Smartwatch and I can Track my User's Arm. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '16)*. Association for Computing Machinery, New York, NY, USA, 85–96. <https://doi.org/10.1145/2906388.2906407>
- [43] Shixin (Cindy) Shen and Ronald A. House. 2017. Hand-arm vibration syndrome. *Canadian Family Physician* 63, 3 (March 2017), 206–210. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5349719/>
- [44] Yongkun Song, Tian Jin, Yongpeng Dai, Yongping Song, and Xiaolong Zhou. 2021. Through-Wall Human Pose Reconstruction via UWB MIMO Radar and 3D CNN. *Remote Sensing* 13, 2 (Jan. 2021), 241. <https://doi.org/10.3390/rs13020241>
- [45] Elizabeth Spencer. 2020. The Pocket: A Hidden History of Women's Lives, 1660–1900. *Cultural and Social History* 17, 4 (Aug. 2020), 570–572. <https://doi.org/10.1080/14780038.2020.1810955> Publisher: Routledge _eprint: <https://doi.org/10.1080/14780038.2020.1810955>
- [46] tesych. 2020. About Azure Kinect DK. <https://docs.microsoft.com/en-us/azure/kinect-dk/about-azure-kinect-dk>
- [47] Daniel Vlasic, Rolf Adelsberger, Giovanni Vannucci, John Barnwell, Markus Gross, Wojciech Matusik, and Jovan Popović. 2007. Practical motion capture in everyday surroundings. *ACM Transactions on Graphics* 26, 3 (July 2007), 35. <https://doi.org/10.1145/1276377.1276421>
- [48] Timo von Marcard, Bodo Rosenhahn, Michael J. Black, and Gerard Pons-Moll. 2017. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. <http://arxiv.org/abs/1703.08014> arXiv:1703.08014 [cs].
- [49] Chuyu Wang, Jian Liu, Yingying Chen, Lei Xie, Hong Bo Liu, and Sanclu Lu. 2018. RF-Kinect: A Wearable RFID-based Approach Towards 3D Body Movement Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous*

- Technologies* 2, 1 (March 2018), 1–28. <https://doi.org/10.1145/3191773>
- [50] Wenchuan Wei, Keiko Kurita, Jilong Kuang, and Alex Gao. 2021. Real-Time 3D Arm Motion Tracking Using the 6-axis IMU Sensor of a Smartwatch. In *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–4. <https://doi.org/10.1109/BSN51625.2021.9507012>
- [51] Jason Wiese, T. Scott Saponas, and A.J. Bernheim Brush. 2013. Phoneprioception: enabling mobile phones to infer where they are kept. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 2157–2166. <https://doi.org/10.1145/2470654.2481296>
- [52] Wikipedia. 2022. List of best-selling video games. https://en.wikipedia.org/w/index.php?title=List_of_best-selling_video_games&oldid=1104319901 Page Version ID: 1104319901.
- [53] Wikipedia. 2022. List of UWB-enabled mobile devices. https://en.wikipedia.org/w/index.php?title=List_of_UWB-enabled_mobile_devices&oldid=1109918576 Page Version ID: 1109918576.
- [54] Mariusz P. Wilk, Michael Walsh, and Brendan O'Flynn. 2021. Multimodal Sensor Fusion for Low-Power Wearable Human Motion Tracking Systems in Sports Applications. *IEEE Sensors Journal* 21, 4 (Feb. 2021), 5195–5212. <https://doi.org/10.1109/JSEN.2020.3030779> Conference Name: IEEE Sensors Journal.
- [55] Frieder Wittmann, Olivier Lambercy, and Roger Gassert. 2019. Magnetometer-Based Drift Correction During Rest in IMU Arm Motion Tracking. *Sensors* 19, 6 (Jan. 2019), 1312. <https://doi.org/10.3390/s19061312>
- [56] Xuesu Xiao and Shuayb Zarar. 2018. A Wearable System for Articulated Human Pose Tracking Under Uncertainty of Sensor Placement. In *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 1144–1150. <https://doi.org/10.1109/BIOROB.2018.8487858> ISSN: 2155-1782.
- [57] Chao Yang, Lingxiao Wang, Xuyu Wang, and Shiwen Mao. 2022. Environment Adaptive RFID-Based 3D Human Pose Tracking With a Meta-Learning Approach. *IEEE Journal of Radio Frequency Identification* 6 (2022), 413–425. <https://doi.org/10.1109/JRFID.2022.3140256> Conference Name: IEEE Journal of Radio Frequency Identification.
- [58] Xinyu Yi, Yuxiao Zhou, Vladislav Golyanik, Marc Habermann, Soshi Shimada, Christian Theobalt, and Feng Xu. 2022. Physical Inertial Poser (PIP): Physics-aware Real-time Human Motion Tracking from Sparse Inertial Sensors. (June 2022), 15.
- [59] Xinyu Yi, Yuxiao Zhou, and Feng Xu. 2021. TransPose: real-time 3D human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics* 40, 4 (July 2021), 86:1–86:13. <https://doi.org/10.1145/3450626.3459786>
- [60] Alexander David Young. 2010. Use of Body Model Constraints to Improve Accuracy of Inertial Motion Capture. In *2010 International Conference on Body Sensor Networks*. 180–186. <https://doi.org/10.1109/BSN.2010.30> ISSN: 2376-8894.
- [61] Han Zhou, Yi Gao, Xinyi Song, Wenxin Liu, and Wei Dong. 2020. LimbMotion: Decimeter-Level Limb Tracking for Wearable-Based Human-Computer Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 161 (sep 2020), 24 pages. <https://doi.org/10.1145/3369836>
- [62] Huiyu Zhou, Thomas Stone, Huosheng Hu, and Nigel Harris. 2008. Use of multiple wearable inertial sensors in upper limb motion tracking. *Medical Engineering & Physics* 30, 1 (Jan. 2008), 123–133. <https://doi.org/10.1016/j.medengphy.2006.11.010>
- [63] Shaghayegh Zihajehzadeh, Paul K. Yoon, Bong-Soo Kang, and Edward J. Park. 2015. UWB-Aided Inertial Motion Capture for Lower Body 3-D Dynamic Activity and Trajectory Tracking. *IEEE Transactions on Instrumentation and Measurement* 64, 12 (Dec. 2015), 3577–3587. <https://doi.org/10.1109/TIM.2015.2459532>