MI-Poser: Human Body Pose Tracking Using Magnetic and Inertial Sensor Fusion with Metal Interference Mitigation

RIKU ARAKAWA*, Carnegie Mellon University, United States
BING ZHOU†, Snap Research, United States
GURUNANDAN KRISHNAN, Snap Research, United States
MAYANK GOEL, Carnegie Mellon University, United States
SHREE K. NAYAR, Snap Research, United States

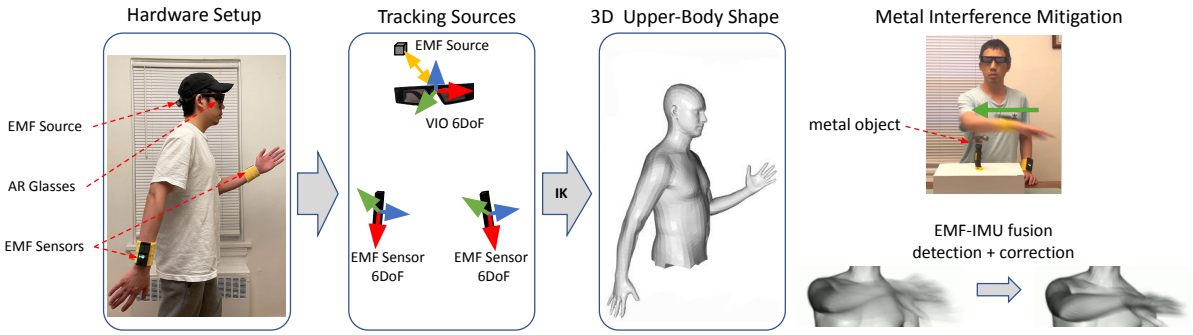

Fig. 1. MI-Poser takes the Visual-Inertial Odometry (VIO) tracking data from AR glasses and two EMF sensors on the wrists as input and generates 3D body shapes through a machine learning model for IK. To tackle the well-known magnetic metal interference issue, we propose Metal Interference Mitigation (MIM) which actively detects and corrects metal interference with EMF-IMU sensor fusion. As a result, the output body movements become steady and have more fidelity.

Inside-out tracking of human body poses using wearable sensors holds significant potential for AR/VR applications, such as remote communication through 3D avatars with expressive body language. Current inside-out systems often rely on vision-based methods utilizing handheld controllers or incorporating densely distributed body-worn IMU sensors. The former limits hands-free and occlusion-robust interactions, while the latter is plagued by inadequate accuracy and jittering. We introduce a novel body tracking system, *MI-Poser*, which employs AR glasses and two wrist-worn electromagnetic field (EMF) sensors to achieve high-fidelity upper-body pose estimation while mitigating metal interference. Our lightweight system demonstrates a minimal error (6.6 cm mean joint position error) with real-world data collected from 10 participants. It remains robust against various upper-body movements and operates efficiently at 60 Hz. Furthermore, by incorporating an IMU sensor co-located with the EMF sensor, MI-Poser presents solutions to counteract the effects of metal interference, which inherently

*The majority of work was done during an internship at Snap Research.
†Corresponding author.

Authors' addresses: Riku Arakawa, rarakawa@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, United States; Bing Zhou, bzhou@snapchat.com, Snap Research, New York, United States; Gurunandan Krishnan, gkrishnan@snap.com, Snap Research, New York, United States; Mayank Goel, mayankgoel@cmu.edu, Carnegie Mellon University, Pittsburgh, United States; Shree K. Nayar, snayar@snap.com, Snap Research, New York, United States.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2023 Copyright held by the owner/author(s).
2474-9567/2023/9-ART85
https://doi.org/10.1145/3610891

disrupts the EMF signal during tracking. Our evaluation effectively showcases the successful detection and correction of interference using our EMF-IMU fusion approach across environments with diverse metal profiles. Ultimately, MI-Poser offers a practical pose tracking system, particularly suited for body-centric AR applications.

CCS Concepts: • **Human-centered computing** ! **Mobile devices**; • **Computing methodologies** ! *Computer vision.*

Additional Key Words and Phrases: body pose tracking, inverse kinematics, sensor fusion

ACM Reference Format:
Riku Arakawa, Bing Zhou, Gurunandan Krishnan, Mayank Goel, and Shree K. Nayar. 2023. MI-Poser: Human Body Pose Tracking Using Magnetic and Inertial Sensor Fusion with Metal Interference Mitigation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 85 (September 2023), 24 pages. https://doi.org/10.1145/3610891

1 INTRODUCTION

Human motion tracking is integral to augmented reality (AR) and virtual reality (VR). Existing VR devices [20, 35, 54] primarily use cameras in head-mounted displays to track head pose and hand-held controllers for spatial input. However, cameras are power-intensive and impractical for AR glasses. Moreover, their limited field of view can lose track of controllers or hands, constraining user interaction. To achieve occlusion-robust, hands-free tracking, researchers have explored wearable-IMU-based solutions [21, 68–70]. However, these approaches typically require more sensors attached to the body (*e.g.*, waist to track torso movement), and the error can accumulate over time as the IMU sensor does not provide 3D position information directly. As a result, the solutions often suffer from imprecise tracking results.

In this paper, we introduce *MI-Poser*, an upper-body pose-tracking system utilizing magnetic tracking in wristbands and AR glasses, as depicted in Figure 1. MI-Poser incorporates an electromagnetic field (EMF) source in AR glasses and two wrist-worn EMF sensors, enabling 6-DoF wrist tracking relative to the head. Combined with AR glasses' Visual-Inertial Odometry (VIO) tracking, MI-Poser can track 6-DoF poses for head and wrists. We trained deep neural networks for human pose inverse kinematics (IK) on a large dataset (AMASS [34]) to reconstruct upper-body pose from sparse signals. Importantly, EMF sensing often suffers from interference from nearby metallic objects. Thus, we propose *metal interference mitigation* (MIM) to enhance the input data of pose reconstruction by utilizing a collocated IMU sensor. MIM detects metal interference on EMF sensors and actively corrects the measured values. MI-Poser's IK and MIM incur little latency and the pipeline runs efficiently at 60 Hz.

We first evaluated MI-Poser's feasibility in body pose tracking by comparing its output with Microsoft Kinect [36] tracking. The study involved ten participants performing various upper-body movements, including out-of-sight hand motions. Results indicated that MI-Poser tracks upper-body pose with a mean joint position error of 6.6 cm. Additionally, we found that MI-Poser outperforms prior IMU-based work with the same sensor placement, thanks to the precise 6-DoF information EMF tracking provides.

Next, we assessed our approach to addressing the inherent metal interference issue by collecting a dataset of synchronized EMF sensor data, IMU sensor data, and ground truth pose data. Data were collected under three conditions with varying metal profiles: open-space, standard, and extreme cases. We first quantified errors between EMF and ground truth pose in each condition, finding significant tracking errors exist only briefly when the EMF sensor passes by a metal object. Our EMF-IMU fusion approach successfully detects such short-period metal interference (0.62 MCC in standard and 0.56 in extreme cases). Moreover, our correction approach reduces tracking error under interference by 11.6° max rotation error per session and 3.3 cm max position error per session in standard cases.

In sum, this work offers the following contributions:

(1) We devised a hands-free wearable upper-body pose tracking system with a natural form factor, using wrist-worn EMF sensors and AR glasses.

(2) We proposed metal interference mitigation (MIM) to address the inherent issue in utilizing EMF tracking in dynamic user environments.

(3) In User Study 1, we demonstrated our system's ability to reconstruct 3D body pose in various movements with a small error in real-time, even when hands are out of view from the AR glasses.

(4) In User Study 2, we showed that MIM significantly reduces error in EMF tracking in environments with different metal profiles. We also qualitatively demonstrated its efficacy in improving body pose tracking with smoother and more accurate output.

Although prior work used EMF sensors for tracking VR controllers [62], MI-Poser is the first hands-free, occlusion-robust body tracking system employing sparse on-body EMF sensors with metal interference solutions. The results suggest MI-Poser provides a practical body pose tracking system for everyday body-centered AR applications, such as remote communication through avatars with upper-body expression while walking outdoors.

2 RELATED WORK

To situate our work, we first review existing wearable systems for body pose tracking and discuss the need to address the limitations of current vision- and IMU-based systems. Then, we examine existing research employing magnetic tracking to provide a background for our EMF-based pose tracking system.

2.1 Wearable Systems for Body Pose Tracking

Researchers have explored wearable body pose estimation systems as a means to achieve portable and flexible interactions. Prior work has involved attaching an array of sensors on the body [51, 60] or using exoskeletons [71]. Recent developments in machine learning approaches have enabled systems with lower costs using sparse sensors, reducing the burden of wearing numerous sensors on the body. Vision-based approaches are the most popular in this context [2–4, 22, 39, 50, 58, 63, 66]. For example, Ahuja *et al.* [4] attached additional cameras to the Meta Quest 2 VR controllers to create an inside-out body capture system. Magic Leap 2 [33] employs a similar technique. Some researchers opted for wrist-worn vision sensors instead of VR controllers, using a spherical camera [8] or an array of small cameras [30]. Furthermore, several IK models have been developed to achieve high-fidelity body reconstruction from sparse sensor inputs. These models use the poses of the head and two hands as inputs to estimate the full body [5, 15, 22]. For instance, Jiang *et al.* [22] proposed a full-body pose tracking system using HTC VIVE (note it necessitates an additional base station in the environment). To train and evaluate these IK models, researchers utilized the extensive human motion database AMASS [34], which comprises a collection of high-precision MoCap datasets. Although these vision-based systems minimize error within the dataset, they unavoidably face challenges such as heavy computation (*e.g.*, model inference) and sensing costs[1], which can be critical in resource-constrained devices like AR glasses. Moreover, they depend on line of sight; therefore, trained IK models may not function correctly if a user's hand moves out of the camera's view, limiting the possible tracking range of human body movement. Unlike existing VR headsets (*e.g.*, Meta Quest Pro [35]) that use cameras for tracking hand-held controllers in 3D space, AR glasses lack sufficient space to accommodate multiple cameras, resulting in a limited field of view.

As alternatives, researchers have investigated IMU-based pose tracking [21, 37, 49, 56, 61, 64, 68–70]. Shen *et al.* [49] proposed a method for reconstructing an arm movement from a single smartwatch by using its embedded IMU sensors. Similarly, Tautges *et al.* [56] proposed an approach to reconstructing full-body animation from four acceleration sensors attached to wrists and ankles. In the context of IMU-based pose tracking, recent works

[1]For reference, the HTC Vive Lighthouse consumes approximately 5W for the source [62].

leverage more sophisticated machine/deep learning models using the AMASS dataset [34]. For example, Sparse Inertial Poser [61] allows 3D human pose estimation using six IMU sensors attached to wrists, lower legs, back, and head. Deep Inertial Poser [21] improves the approach by incorporating temporal pose priors through deep learning. TransPose [70], LoBSTr [68], and Physical Inertial Poser [69] build upon those works and further advance the performance of IMU-based body pose tracking. However, these approaches require not a small number of sensor-instrumented joints (typically six) and exhibit a certain degree of errors.

Considering previous research, there is a demand for hands-free, occlusion-robust pose tracking systems with minimal errors, particularly in AR contexts. As a result, we developed a body pose tracking system utilizing a practically sparse sensor input (AR glasses and two wrist-worn sensors) based on a different sensing modality: EMF sensing.

2.2 Magnetic Tracking for Interaction

Magnetic field sensing-based tracking has a long history [46, 47] with a comprehensive review available in [44]. Several HCI applications have been proposed [9, 11, 18, 43, 52], leveraging occlusion-free tracking. For example, Abracadabra [18] tracks finger radial position relative to a watch using an attached magnet, and Nenya [9] measures magnetic field changes with a magnetometer-equipped smartwatch and a ring with two permanent magnets. Chen *et al.* [11] proposed uTrack, which tracks finger movements using a pair of magnetometers on the back of the fingers and a permanent magnet to the back of the thumb. While precise in short-range tracking like centimeters, these approaches can not be extended to long-range (*e.g.*, body-scale) since the Earth's geomagnetic field easily influences the tracking. Razer Hydra [52] extends the sensing range of the controller by using a base station that generates a weak magnetic field.

Electromagnetic field (EMF) tracking involves oscillating magnetic fields [25, 29] and has gained attention for its precision in medium-range 6-DoF tracking. For a detailed review, see [17]. Several HCI applications have been proposed using EMF tracking, such as Finexus [12], which advances uTrack [11] by tracking multiple fingertips in real time. AuraRing [42] offers 5-DoF finger tracking for VR/AR applications with low power consumption (*i.e.*, around 2.3mW for a sensor in a ring and 73.3mW for a transmitter in a wristband). Whitmire *et al.* [62] extended the tracking range to body scale, enabling VR controller tracking by embedding three coils in HMD and a set of orthogonal receivers in hand-held devices with reasonable power consumption (*i.e.*, around 45mW for a sensor in a controller and 224mW for a transmitter in HMD without wireless communication module). Similar work to ours is EM-Pose [23], a wearable EMF-based body tracking system that uses 6 or 12 on-body sensors, which requires users to wear an additional EMF source on their back.

A known drawback of EMF tracking is its susceptibility to magnetic field distortion by environmental metals [42, 62], particularly in dynamic environments like VR/AR. Interference becomes more significant as the tracking range increase es, such as body-scale [62]. Some work has attempted offline calibration to account for magnetic interference [26, 27], but online calibration is desirable. Although previous work [23, 62] recognized the issue, no work has been proposed to date to quantify the interference effect in different user environments and to devise solutions to it. Therefore, we propose approaches to addressing the metal interference issue in our EMF-based body pose tracking system.

3 PROPOSED METHOD

We design a hands-free, wearable upper-body pose tracking system with an extensive tracking range, MI-Poser. The system pipeline is illustrated in Figure 2. Unlike existing VR tracking systems that employ cameras and handheld controllers, MI-Poser utilizes wrist-worn EMF sensors for tracking. As depicted in Figure 1, users wear an EMF receiver on each wrist while an EMF source is mounted to the AR glasses. The EMF tracking captures the wrist poses relative to the EMF source. Simultaneously, the AR glasses track the user's head position in the

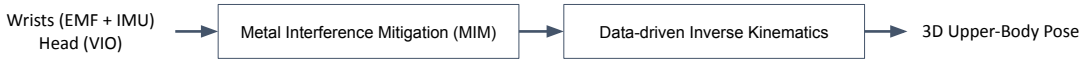

Fig. 2. The overview of MI-Poser's pipeline.

world coordinate using Visual-Inertial Odometry [16]. These sparse measurements are input into our IK model to reconstruct a high-fidelity upper-body pose. We prioritized upper-body pose tracking due to its applicability to various AR applications. While our sensor setup could estimate full-body pose in a data-driven manner for specific motions through hallucination, like walking as shown in [37], we focus on the fidelity of the reconstructed pose in general conditions.

Metal interference is inevitable when using EMF sensors for body tracking in dynamic user environments. Therefore, we propose Metal Interference Mitigation (MIM) methods that operate online with minimal latency. While previous work [23, 62] acknowledged the issue, no concrete solutions have been proposed, as discussed in Section 2.2. Our solution incorporates an IMU sensor embedded with the EMF receiver. The fusion of EMF and IMU sensors has been studied to enhance EMF tracking performance in ideal, metal-free environments using a static filter like Kalman Filter [47]. However, in practical situations, the tracking algorithm should actively detect interference presence in real time and dynamically correct the trajectory.

3.1 MIM Overivew

Initially, we examined the behavior of our EMF tracking under metal interference. Within a 1.5 m range from the EMF source, metal effects are seldom present in open-space environments (*e.g.*, outside), leading to accurate tracking performance. However, in typical spaces with some metal objects (*e.g.*, a desk with a laptop), interference emerges when the EMF receiver approaches a metal object, confirming previous observations [62]. Even over a short period, interference significantly impacts the position and rotation of the EMF sensor, potentially degrading IK performance. We identified two types of metal interference based on how the sensor moves around metal objects. On one hand, when the sensor passes by a static metal object, a spike-like short-period error occurs in the tracking. On the other hand, when a metal object and the EMF receiver move together, an error persists as long as they remain in close range. Both cases can occur in end-user scenarios, such as users moving their arm near metal objects or holding a metal can while interacting with AR content.

To address this, we divided the problem into two parts: *interference detection* and *interference correction*. The overview of our MIM approach is presented in Figure 3. The detection part aims to identify moments when tracking errors arise due to metal interference, while the correction part seeks to mitigate errors by adjusting the interfered EMF sensor values. In this paper, we focus on correcting the first type of interference, the short-period error that occurs when metal objects are placed statically in an environment and the sensor encounters them occasionally (*e.g.*, users swinging their arms). This is due to the challenge of tracking pose over a long period (more than a few seconds) using EMF and IMU sensors under metal interference. However, our detection method can also address the second type of interference, informing users of the reasons for degraded body tracking performance and improving user experience [6]. For instance, if a user holds a smartphone that causes interference to the tracking of the corresponding wrist, MI-Poser can notify the user via AR glasses that the tracking performance is low because the metal object is close to the hand.

3.2 Interference Detection

The first part is interference detection. There are two values regarding the rotation of the sensor based on different principles: angular momentum from the gyro sensor and orientation from the EMF sensor. Let's assume a rotation in axis-angle representation $_B$ " $^1\mathcal{C}^o$ 2 R^1 ^{3} given time $\mathcal{C}$ when there is no metal interference. We use $^1\mathcal{C}^o$ as

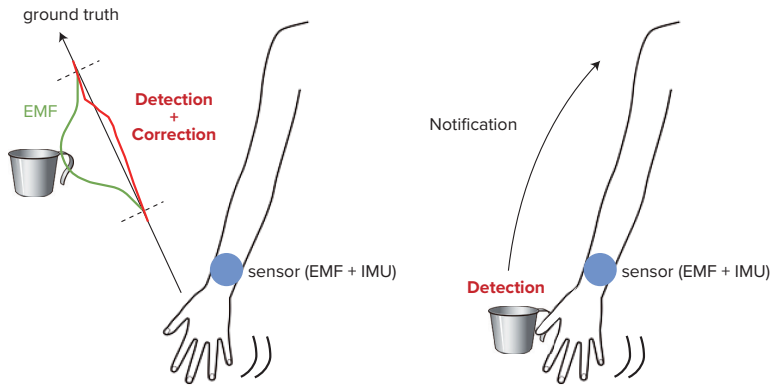

Fig. 3. Overview of the proposed metal interference mitigation (MIM) according to the two types of errors. When a user encounters a metal object for a short period, it is detected and the trajectory is corrected (left). When the interference is longer, *e.g.*, when the user holds a metal object, it is detected for notifying the user (right).

a binary index to represent the presence or absence of interference; $^{1}C^{o} = 0$ in this case. Simultaneously, an angular momentum $_{B}{}^{''*}{}^{1}C^{o}$ is measured in the same coordinate as $_{B}{}^{''}{}^{1}C^{o}$. At time C , C, there is rotation information from the EMF sensor as $_{B}{}^{''}{}^{1}C$, C^{o}. If no interference occurs at C , C, an approximation holds: $_{B}{}^{''}{}^{1}C$, C^{o} $_{B}{}^{''}{}^{1}C^{o}$, $_{B}{}^{''*}{}^{1}C^{o}$ C.

We can then introduce an error threshold 4_{C} to estimate the interference state ^{1}C , C^{o} by comparing 4_{C} with the distance between $_{B}{}^{''}{}^{1}C$, C^{o} and $_{B}{}^{''}{}^{1}C^{o}$, $_{B}{}^{''*}{}^{1}C^{o}$ C. The distance is the intrinsic geodesic distance between two angles. If the distance is larger than the threshold, the system predicts $\hat{}^{1}C$, $C^{o} = 1$. Otherwise, it predicts $\hat{}^{1}C$, $C^{o} = 0$.

3.3 Interference Correction

The second part is interference correction, which dynamically adjusts the measured value from the EMF sensor based on the detection result. It is essential to perform this correction online, as MI-Poser aims to be a real-time body pose tracking system. This means that if $\hat{}^{1}C^{o} = 1$, we need to correct the current position $\%_{B}{}^{''}{}^{1}C^{o}$ and rotation $_{B}{}^{''}{}^{1}C^{o}$ using past tracking and sensor data up to time C. If interference persists in $\hat{}^{1}C$, $C^{o} = 1$, we must correct them using the past data up to C , C. As noted, we apply this correction as long as the detected interference is of short duration to avoid drift error.

For the rotation correction, we use $_{B}{}^{''}{}^{1}C^{o}$, $_{B}{}^{''*}{}^{1}C^{o}$ C instead of $_{B}{}^{''}{}^{1}C$, C^{o}. Given that IMU-based rotation tracking is fairly feasible, we expected this simple solution to work well. Meanwhile, IMU-based position tracking is known to be a challenging problem, and we have prepared three methods.

3.3.1 IMU Odometry Model. This physics-based method uses initial velocity and a time series of acceleration from the IMU sensor to calculate position through dual integration.

$$G^{1}C_{0} \text{ , } C^{o} = G^{1}C_{0}{}^{o} \text{ , } E^{1}C_{0}{}^{o} \quad C \text{ , } \int_{C_0}^{\int_{C_0}^{C} {}^{1} g} O^{1}g^{0o}3g^{0}3g \tag{1}$$

, where $G^{1}C^{o}$, $E^{1}C^{o}$, and $O^{1}C^{o}$ represent position, velocity, and acceleration at time C, and C_0 represents the initial reference time. While straightforward, this method performs poorly when there is noise in $O^{1}C^{o}$ and $E^{1}C_{0}{}^{o}$ [57]. Since we use this approach around the interference moments, the EMF position tracking can contain noises

during the moments, resulting in noisy $E^{1}C_0^{o}$. Another limitation is that successful short-time arm tracking based on an IMU sensor [32] requires a high sampling frequency like 2000 Hz, indicating that lower frequency leads to larger errors.

3.3.2 Trajectory Forecasting Model. The IMU odometry method may not account for trajectory trends and seasonality, which are often used in practical time-series forecasting methods [14]. Human body movements, particularly arm movements, include short-duration trends and can be forecasted based on past trajectories [19, 55, 65]. We anticipated that a short-period future trajectory could be forecasted using previous tracking history, which could then be used to correct EMF position data under metal interference. We adopted the N-BEATS method [41], a state-of-the-art deep learning approach using backward and forward residual links and a deep stack of fully-connected layers. The model's input and output can be written as:

$$fG^{1}C_0^{o}\bullet^{\prime\prime\prime\prime}\bullet G^{1}C_0\ ,\quad C_{>DC?DC}^{o}g = \text{N-BEATS}^{1}fG^{1}C_0 \qquad C_{8=?DC}^{o}\bullet^{\prime\prime\prime\prime}\bullet G^{1}C_0^{o}g^{o} \tag{2}$$

, where $C_{>DC?DC}$ and $C_{8=?DC}$ correspond to the amount of future data the model outputs and the amount of previous data the model takes as inputs, respectively.

In testing this model, we found that significant prediction errors tended to occur when there were large position changes right after the moment the model predicted. This can be understood as the time-series forecasting model estimating the trajectory based on past data but not reflecting future acceleration information. This observation led us to introduce the following model.

3.3.3 Fusion Model. To consider future acceleration while avoiding error due to noisy $E^{1}C_0^{o}$, we approximate the trajectory as follows:

$$G^{1}C_0\ ,\quad C^{o} = \text{N-BEATS}^{1}fG^{1}C_0 \qquad C_{8=?DC}^{o}\bullet^{\prime\prime\prime\prime}\bullet G^{1}C_0^{o}g^{o}j_{C=C_0\ ,\ C\ ,\ \substack{1\ C_0\ ,\ C\ 1\ g \\ C_0 \qquad C_0}}\ O^{1}g^{0o}3g^{0}3g \tag{3}$$

, where C is small enough (at least, $C < C_{>DC?DC}$). We iteratively use the same N-BEATS model. In detail, while the N-BEATS model outputs estimation for $C_{>DC?DC}$ seconds, we use the single prediction value corresponding to time $C_0\ ,\ C$. After adding the acceleration component through integration, we use this value as the input for the next N-BEATS inference for the next frame ($C_0\ ,\ 2\ C$) if metal interference still exists (*i.e.*, $\hat{}^{1}C_0\ ,\ 2\ C^{o} = 1$). In this way, we can adjust the N-BEATS prediction by adding the acceleration component, which further influences the subsequent trajectory forecasting.

4 IMPLEMENTATION AND SYSTEM PERFORMANCE

4.1 Hardware

Our EMF tracking system has an EMF transmitter (source) and two EMF receivers (sensors) using off-the-shelf 3D coils (See Figure 4). These components communicate with AR glasses through Bluetooth Low Energy (BLE) using the Enhanced ShockBurst protocol [40] to minimize latency. Designing an EMF tracking system for human body tracking involves multiple trade-offs, such as source coil size, electric current, sensor coil size, and tracking range/error requirements. Our final sensor configuration meets our 1.5 meter range requirement (typical arm reach) with a position RMS error of 0.9 mm and angle RMS error of 0.5° within a 1 meter range in an *ideal metal interference-free* lab environment. Additionally, an IMU sensor is integrated into the sensor, which we leverage for a sensor-fusion approach to MIM. The EMF and IMU data streams are synchronized and accessible via a BLE connection. We run the tracking at 120 fps with a latency of around 15 ms, and the tracking algorithm runs locally on the sensor. The algorithm incorporates Kalman Filter to stabilize long-range tracking and reduce jitter. We use Spectacles [53], commercial AR glasses, featuring a precise VIO algorithm running at 60 Hz, which is accessible via a BLE connection.

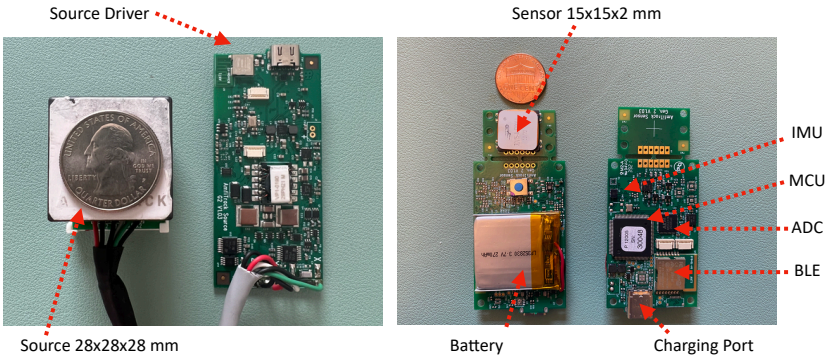

Fig. 4. EMF tracking hardware. The source (left) is integrated into AR glasses and the sensors (right) are attached to the user's wrists.

4.2 IK Model for Pose Estimation from Sparse Sensor Inputs

We used the SMPL model [31] to represent and animate the human body pose. We trained our IK model to reconstruct the upper body from a sparse sensor set using the AMASS dataset [34], similarly to prior work discussed in Section 2.1. For the IK model, we adopted the state-of-the-art model from AvatarPoser [22]. The key difference is that our proposed system has EMF sensors on the wrists, while AvatarPoser assumes hand-held controllers. We expected that this different sensor placement would improve upper-body tracking by avoiding rotation noises from hand movements and helps the model infer more plausible arm poses.

In AMASS, we used a subset combination of the CMU, Eyes_Japan, KIT, MPI_HDM05, and TotalCapture datasets as the training set, and MPI_Limits as the validation set. We down-sampled the MoCap dataset from 120 Hz to 60 Hz and generated windowed segments of 40 frames (*i.e.*, 2/3 second window) with a stride length of 0.1 seconds to match with the original work [22]. We used the Adam optimizer [28] with a batch size of 32, and a starting learning rate was 0.001, which decays by a factor of 0.8 every 20 epochs. We performed the training with PyTorch on Google Cloud Platforms with NVIDIA Tesla V100 GPU.

To account for variations in body size and sensor-wearing positions, we calibrated the sensor outputs before inputting them into the IK model. For sensor-position calibration, the user simply maintained the default T-pose (Figure 13 in Appendix) for a few seconds. We used sensor measurements taken during this period for calibration, similar to prior work [21, 70]. We first estimated a scaling factor by comparing the arm span between actual sensor measurements and the SMPL model definition. To compensate for minor sensor offsets, we applied spatial transformations to the sensor output, ensuring alignment with the SMPL model definitions in the default pose. We applied the scaling factor and transformations to each frame throughout the entire body pose tracking session.

4.3 Trajectory Forecasting Model for MIM

To train the N-BEATS model for MIM (interference correction), we first collected data by moving the EMF sensor freely in open space without metal objects. Approximately 40 minutes of data were used for training the N-BEATS model with two stacks where there are two blocks per stack with 512 hidden layer units. The model takes 120 samples (corresponding to 1 second of $C_{8=?DC}$) as input and outputs 60 samples (corresponding to 0.5 seconds of $C_{>DC?DC}$) position data (Recall the EMF tracking runs at 120 Hz). We used the Adam optimizer [28] with batch size 16 up to 20 epochs. Within the dataset, the best model performed a 1.26 cm mean absolute position error in the validation dataset. This means that the model can forecast a position of 0.5 seconds ahead with small errors.

4.4 Real-Time System Performance

Currently, the IK model and MIM process run on a laptop (MacBook Pro with a 2.6 GHz 6-Core CPU and 16 GB memory) written in Python while streaming data in real time. Future work will involve transferring the process to AR glasses using JavaScript. Spectacles are equipped with an Octa-core CPU (2 2.52 GHz + 6 1.7 GHz). Given the limited computational resources, it is crucial to consider power consumption and inference speed, and we report them in our current prototype in this section.

4.4.1 Power Consumption. The power consumption at the source and the sensor is 1.4W and 0.68W, involving the communication modules, respectively. Further optimization, as demonstrated in [62], is advisable. For instance, replacing the currently used microcontroller (F7), which has more capabilities than necessary, with a lower power consumption alternative (H7) could reduce power consumption. Nonetheless, the sensor-level power consumption is significantly lower than existing camera-based research work. For example, ControllerPose [4] attaches a camera to each controller to capture upper-body movements, and a single camera's power consumption is approximately 3.3W, which does not involve the hand tracking algorithm. While we must consider the power required to run the model on the device, our prototype is suggested to operate with reasonable power consumption.

4.4.2 Inference Speed. The IK model's current latency is 4.2 ms on the laptop, from captured EMF values to the output. Likewise, MIM's detection and correction models incur average latencies of 0.09 ms and 0.50 ms, respectively. These latencies do not significantly impact the body pose tracking pipeline, making MIM a suitable complement to the EMF-based upper-body pose tracking system. Together with the MI-Poser's IK model for reconstructing body pose, our pipeline takes approximately 5 ms to process one frame on a laptop. The current MI-Poser pipeline operates efficiently at 60 Hz (recall the EMF sensor runs at 120 Hz and the VIO tracking runs at 60 Hz). Notably, commercial on-device tracking speeds, such as Meta Quest 2 [35], are also 60 Hz. For further comparison, we ran ControllerPose [4] and IMUPoser [37] systems using the same laptop, with pipeline speeds of approximately 4 Hz and 48 Hz, respectively. Please refer to the Video Figure for a real-time demonstration.

5 USER STUDY 1: UPPER-BODY POSE TRACKING IN THE ABSENCE OF METAL INTERFERENCE

Since MI-Poser is the first setup for an upper-body tracking system with two wrist-worn EMF sensors and AR glasses, we first examined its tracking performance in an open space (without visible metal objects) and compared it with similar setups using IMU sensors. We trained an IK model using the existing AMASS dataset [34] and tested it with real sensor data.

5.1 Data Collection

We collected sensor data from AR glasses and EMF sensors to demonstrate system performance using the setup shown in Figure 1. Ground truth data were obtained using Microsoft Kinect [36], following prior work [4]. To ensure the reliability of the ground truth data from Kinect, we filtered out unreliable inferred tracking frames in post-processing, based on the tracking state Kinect logged, which constituted approximately 10% of the data.

To inspect the fine-grained performance of MI-Poser across various upper-body movements, we designed an obstacle course-style setup inspired by previous works [3, 21]. We selected motions that encompassed a diverse range of upper-body movements:

Punch: the participant alternately punches with both arms in front of their body.
Wave: the participant randomly raises their arms and waves in the air.
Swing: the participant alternately swings their arms from side to side.
Rotate: the participant rotates both arms against each other in front of their chest.
Walk: the participant walks randomly with their arms swinging naturally.
Basketball: the participant jumps and performs basketball shooting gestures over their head.

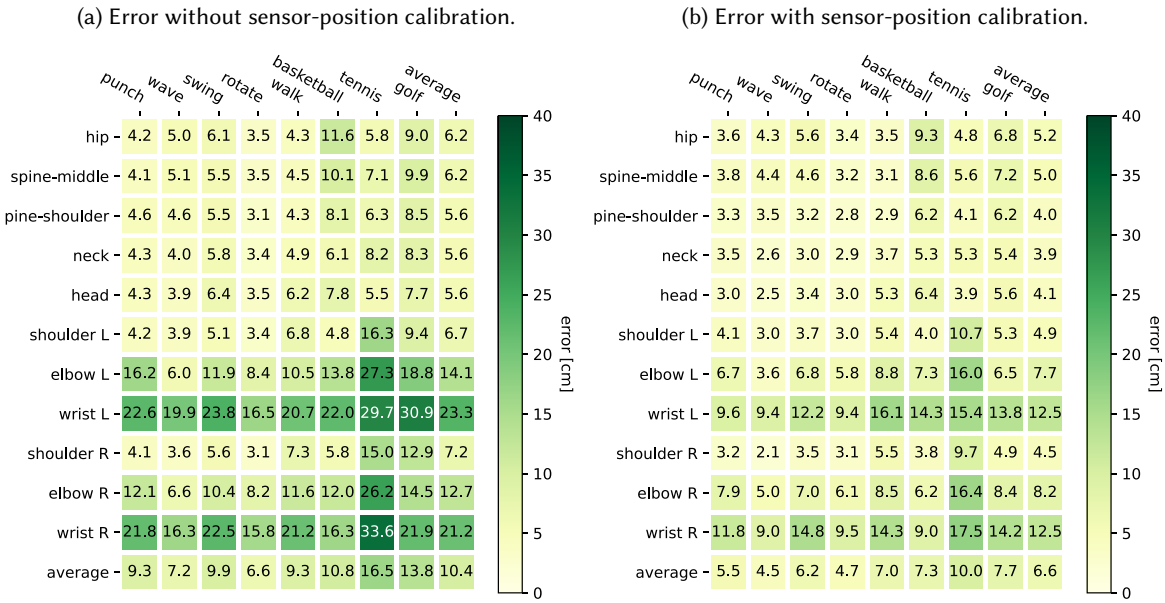

Fig. 5. MI-Poser's performance (joint position error) on real sensor data across body region and motion. The error is overall small thanks to the precise EMF tracking. The sensor-position calibration significantly reduces the error. Error accumulates from the hip (alignment root) to the end-effectors. The largest error is observed during the tennis motion, which involves rapid and extensive torso movements.

Tennis: the participant swings their arms from behind their body to the front, with torso rotation.
Golf: the participant swings both arms together, with torso rotation.

Several of these motions, such as swinging, walking, basketball, tennis, and golf, include moments when hands move outside the field of view of the cameras on AR glasses. These moments are often challenging to track in conventional camera-based AR systems.

We recruited 10 participants from our institution with diverse genders, ages, weights, and body shapes for data collection. Participants performed each motion for 50 seconds, with 10-second rest periods in between. The entire data collection process, including the initial calibration, took approximately 10 minutes per participant. We obtained approval from our institution to conduct the study.

5.2 Results

5.2.1 Fine-Grained Error Metric. The results across different joints and motions are presented in Figure 5. The overall error (a) without and (b) with sensor-position calibration is 10.4 cm and 6.6 cm, respectively, demonstrating a significant improvement due to the sensor-position calibration. As we align the root (hip) in calculating the error metric following prior work such as [21, 37], the hip generally has the smallest error, and the error propagates to the end-effectors like wrists, leading to the largest error. Still, the overall error is reasonably small after the sensor-position calibration, including when hands are out of view from the AR glasses.

However, there are several performance limitations to consider. First, by examining the error by motions in Figure 5, larger errors arise from those involving fast and extensive torso movements, such as tennis. Next, the EMF tracking method has a minimum working distance of about 10 cm to prevent signal saturation. As a result,

gestures involving close proximity to the source, like both hands being close in golf, can cause significant signal jitters. Lastly, while our sensor-position calibration accounts for body skeleton scale and minor sensor position shifts, it does not adjust for variations in body shape. This can lead to inaccurate hand-body contact, such as when the hand penetrates or hovers over the body mesh despite physical contact.

Additionally, we tested a scenario where a user wears a single EMF receiver on their wrist, assuming the sensor is embedded in a smartwatch. We trained a different model with this configuration using the AMASS dataset and evaluated its performance with our dataset, using data corresponding to one EMF receiver and AR glasses. The overall error without and with sensor-position calibration is 22.0 cm and 15.9 cm. The largest error comes from the hand without the EMF receiver, while similar performance is maintained for the hand with the receiver. Anecdotally, the error decreases in some movements (*e.g.*, rotating, walking), which can be attributed to hallucination from the training dataset, as indicated by [37]. Thus, while depending on applications, utilizing a pervasive configuration of on-body sensors could be valuable when users employ tracking in their everyday lives, such as in AR applications while walking.

5.2.2 Comparison to Prior Work. Our system is the first of its kind to use two wrist-worn EMF sensors for upper-body pose tracking. To the best of our knowledge, no prior research has reported the performance of upper-body pose tracking using such a sparse set of real sensors (head and two wrists). Although most IMU-based approaches were tested using the AMASS dataset, several works also reported performance on real sensor data. However, these studies used different datasets and involved more sensors (including leg-worn IMUs) in reporting full-body performance, making direct comparisons difficult. Moreover, the current Spectacles' SDK[2] does not provide raw IMU signals, preventing us from obtaining the IMU signal for the head in our setup, which makes a fair comparison with IMU-based solutions using our dataset impossible. This remains a limitation of the current study.

As a remedy, we adopted IMUPoser's model [37] (two-layer bi-directional LSTM) as a baseline for the IMU-based upper-body pose tracking and tested its error on the DIP-IMU dataset and IMUPoser dataset Both datasets contain human body poses across different motions similar to ours, such as arm raise, arm swing, and walking, involving 10 participants. The DIP-IMU dataset [21] includes 17 IMUs (X-Sense sensors), while IMUPoser uses common wearables as sensors, such as smartwatches and earbuds. We trained their model with the AMASS dataset (the same subsets we used for MI-Poser) using the configuration of three IMU sensors corresponding to our setup (*i.e.*, the head and two wrists) and evaluated its performance on the datasets. As a result, the joint position error for the upper body is 8.3 cm and 10.4 cm with sensor-position calibration for the DIP-IMU and IMUPoser datasets, respectively. Although not a direct comparison in terms of evaluating with different datasets, it is suggested that MI-Poser has a better performance compared to IMU-based systems, thanks to the high precision EMF/VIO tracking.

6 USER STUDY 2: MEASURING AND MITIGATING METAL INTERFERENCE

User Study 1 demonstrated that our EMF-based tracking setup achieves accurate upper-body pose tracking. In this section, we quantify the metal interference on EMF tracking in various environments and examine the effectiveness of MIM in enhancing the input data for our pose-tracking pipeline.

6.1 Data Collection

6.1.1 Apparatus. We used the same sensor described in Section 4.1, which streams EMF tracking and IMU tracking data synchronously. To obtain ground truth tracking data, we used Apple ARKit 4 [7], which enables a state-of-the-art self-localization in world coordinates based on VIO tracking [24]. We developed a custom

[2]https://docs.spectacles.dev/app/reference/api

Fig. 6. Setup for data collection in User Study 2. (a) An EMF sensor and iPhone 13 Pro are tightly attached to a non-metal object. The iPhone tracks its pose based on ARKit's world tracking function. (b) Open-space with few metal objects. We intentionally chose locations with visually striking objects to aid ARKit's VIO tracking. (c) Standard condition with common metal objects such as a monitor. (d) Extreme condition where we intentionally move the sensor close to common metal objects such as a can. Note that the apparatus is not visible in (c) and (d).

iOS application using Xcode 13.3 that records the position, orientation, and UNIX timestamp. We attach the EMF sensor and iPhone 13 Pro, which can utilize built-in LiDAR for enhanced ARKit tracking performance, to a non-metal rigid body with a space of 20 cm between them. This decision was based on the observation by Whitmire *et al.* [62] that significant interference occurs if an EMF receiver and a smartphone are closer than approximately 5 cm. The apparatus used for data collection is shown in Figure 6 (a). When the apparatus is moved, the sensor streams its EMF pose tracking data along with the IMU data at 120 Hz relative to the EMF source, while the iPhone captures its pose at 60 Hz relative to the reference world coordinate. Since these two sensors are attached to a rigid body and their transformation is constant, we can align their coordinates using a calibration process, which we elaborate on in Appendix B.

We did not use high-end MoCap systems such as OptiTrack [38] because we wanted to conduct data collection in multiple and actual environments. Our apparatus offers a convenient method for data collection. We tested whether ARKit could work properly when we moved the apparatus naturally to represent arm movements in VR/AR scenarios. ARKit reports its tracking state, and based on that, we observed that the tracking performance worsens if we rotate the apparatus too quickly. Consequently, we could not include such movements in the data collection.

6.1.2 Condition. We selected three representative cases for data collection concerning the level of metal interference: open-space, standard, and extreme conditions. The open-space condition represents locations with minimal metal interference (See Figure 6 (b)). In contrast, the standard condition includes typical places like desks and rooms with a few metal objects (*e.g.*, a desk with metal support, Figure 6 (c)). For the open-space condition, we chose two locations in a building where no visible metal objects were present within a 2-meter range, except for the floor[3]. For the standard condition, we selected three locations in the same building: a desk with a few everyday metal objects, such as a laptop and monitor, a meeting space with a metal door, and a crafting room with some metal objects like a hammer. In addition to these two conditions, we added an extreme condition, where we intentionally moved the apparatus closer to metal objects for an extended period, such as touching a laptop or holding a metal can (Figure 6 (d)). This condition was introduced to examine the extent of errors that could occur in potential end-user environments.

6.1.3 Procedure. We collected four session data for each environment, resulting in 8 sessions for the open-space condition, 12 sessions for the standard condition, and 4 sessions for the extreme condition. Five people from our

[3]As a typical building, the floor has steel support.