BAIR: A Fine-Grained Real-Time Multi-Modal Ranging System on Smartphones

Xiao Li[1], Feiyu Han[1], Fei Shang[1], Shicheng Zheng[1], Chunyu He[1], Haohua Du[23], Xiang-Yang Li[1]
[1]University of Science and Technology of China, Hefei, Anhui, China
[2]Beihang University, Beijing, China
[3]Deqing Alpha Innovation Institute, Huzhou, Zhejiang, China
{xiao_li, fyhan, feishang, sc7543, hcyustc210404}@mail.ustc.edu.cn, duhaohua@buaa.edu.cn, xiangyangli@ustc.edu.cn

***Abstract*—Accurate and quick relative-distance measurement is crucial for supporting various intelligent transparent services, such as multi-device collaboration, screen rotation, and multi-device mirroring. Unfortunately, current methods often rely on single-modality sensing, resulting in various limitations: BLE-based and WiFi-based methods suffer from coarse-grained estimation, and ultrasound-based approaches suffer from limited sensing range.**

In this work, we aim at designing a distance-measurement system that enjoys long range, high accuracy, and small delay. Our designed system, named BAIR, relies on low-energy Bluetooth (BLE), acoustic sensors, and inertial measurement units (IMU) equipped on commercial smartphones for fine-grained and real-time relative distance estimation. BAIR effectively aligns multiple sensory signals with different sampling rates via the improved Kalman filter technology. To mitigate IMU's integration errors, BAIR calculates the average velocity over a preceding period and uses this, alongside accumulated velocity data from the IMU, significantly improving distance prediction accuracy.

We implemented our BAIR system on smartphones and conducted extensive experiments to evaluate its performance. Specifically, in static scenarios, BAIR achieves a mean average error (MAE) of 11 cm. In moving scenarios, the cumulative distribution function (CDF) values for 95%, 80%, and 50% are 31 cm, 13 cm, and 8 cm, respectively. The memory footprint of BAIR is 16.41 MB. We release an initial version of BAIR and a video demo on GitHub[1] and YouTube[2].

***Index Terms*—ranging, wireless sensing, multi-modality, mobile device, smartphone**

I. Introduction

As the level of intelligence and quantity of personal terminals gradually increase, the collaboration and interaction among multiple devices are attracting increasing attention. For instance, leading companies such as Apple, Samsung, and HUAWEI have continuously iterated related functions for several years, including iWatch-assisted Mac unlocking, multi-screen collaboration, screen rotation, iPhone mirroring, etc. *How to accurately estimate the relative distance between two devices* is a fundamental issue in achieving multi-device collaboration.

Although there have been many works on localization problems in the arts, including WiFi-based [1], [2], sound-based [3], and BLE-based [4], their systems rely on a device with a fixed position as an anchor and then obtain the relative positions of other devices to the anchor device. However, in the scenario of device interaction, the locations of all devices (e.g. laptops, smart watches, mobile phones, etc.) are often unfixed. [5]–[7] explore the possibility of self-localization for non-stationary devices, however, their sensing accuracy and range are insufficient. [5] develops a distributed FMCW system to achieve a tracking accuracy of 5 mm but it requires an elaborate calibration process to remove the unknown clock offset. [6] introduces a self-calibrated acoustic ranging system that achieves sub-millimeter accuracy on distributed asynchronous devices but its ranging distance is limited to 3 meters. [7] propose two indoor pedestrian localization methods based on contact information obtained from BLE installed in smartphones but its average positioning error is 0.74 m. In a flexible device interaction scenario, such as online conference transfer between devices, the system is not only required to provide a rough measurement over long distances but also to track movement and facilitate real-time interaction over short distances. Moreover, despite the high accuracy that many existing excellent algorithms can achieve, such as the EM algorithm, they often require iterations lasting dozens of seconds or even longer. This is unacceptable in device interaction scenarios.

In this work, we propose BAIR, a relative distance-ranging system for smart devices. Our system can be practically deployed and used on commercial smartphones. It enjoys several favorable characteristics, including:

- **Great Universality.** BAIR uses the existing COTS devices instead of installing other sophisticated hardware or making hardware changes on smartphones. Also, it does not rely on location-fixed anchor devices, and can still obtain their relative distance for device interaction when all the device locations are unknown or change.
- **High Availability.** On one hand, the working distance range of the system should be up to 10 meters to meet the needs of scenarios such as smart home wake-up in home settings and inter-device conference flow in office settings. On the other hand, the ranging error at short distances should be less than 10 centimeters to satisfy the requirements for sensitive device interaction.
- **Real-time Response**. To achieve a good user experience

Xiang-Yang Li is the corresponding author
[1]https://anonymous.4open.science/r/BAIR-B2A0/
[2]https://www.youtube.com/watch?v=Scrrz4CA2Cw

with device interaction, the ranging process must not take more than 0.1 s (humans' persistence of vision is 0.1 s). If the update frequency exceeds 10 Hz, the changes will not be noticeable to the user.

The variety of sensors that exist in the end devices gives us opportunities. For example, BLE delivers broader coverage and IMUs offer continuous tracking capabilities. However, two challenges need to be solved first:

(1) How to align multi-modal sensory signals? The measured entities and sampling rates of different sensors vary. To fuse multimodal data for ranging, we need to obtain a consistent representation of different modal data. And due to device differences, their information density is hugely different. For instance, sound-based methods operate at a frequency of 1 Hz and measure absolute distances between devices, while IMUs measure relative distances with a much higher sampling rate of 50 Hz.

(2) How to achieve real-time response under the condition of limited computing power? In order to meet the demands of real-world scenarios, we need to obtain ranging results within 0.1 seconds. However, the computational power of terminals such as mobile phones is relatively weak. For instance, the computational power of the Qualcomm 865 chip is only 15 tops, which needs to be shared by dozens or even hundreds of system services and applications simultaneously.

To address these issues, we employ the Kalman filter to fuse multi-modal data. Firstly, the Kalman filter formulation incorporates both absolute and relative distances. Secondly, the parameters are updated exclusively upon receiving new absolute distance measurements. The dead reckoning algorithm is running during the interval. Thirdly, the Kalman filter utilizes a limited number of parameters and features a straightforward formula, thus conserving memory, power, and computational resources.

Summary of results: We implement BAIR using COTS BLEs, acoustic sensors, and IMUs on smartphones. We first test it in static scenarios. The results show that BAIR achieves a MAE of 11 cm, compared to 1.9 cm with single ultrasound signals and 99 cm with single BLE. BAI-BLE. BAIR has a working range of up to 10 meters, outperforming the 6-meter range of ultrasound. In short distances (< 4 m) and long distances (4 m $\sim$ 6 m), the MAEs are 2 cm and 18 cm, respectively. Next, we conduct experiments in moving scenarios. The CDF values for 95%, 80%, and 50% are 31 cm, 13 cm, and 8 cm, respectively. Compared with other ranging technologies, BAIR is the only one that satisfies all conditions: universality, long range (10 m), low error (11 cm), and real-time response (0.1 s). Additionally, we perform ablation studies that prove our fusion method is effective in reducing error. Furthermore, we explore whether the speed of movement and device type affect the performance of BAIR. We find that its influence is little and acceptable and the ranging error remains decentimeter level. Finally, we test BAIR's memory usage and power consumption. Its memory footprint is just 16.41 MB and it consumes 15% of battery life after one hour of operation.

Contributions: The main contributions of our work are in signal-fusion from various sensors, commercial system implementation on smartphones, and extensive evaluations, specifically 1) We propose a novel multi-modal fusion algorithm that leverages BLEs, acoustic sensors, and IMUs. By employing a dual-Kalman filter approach, we effectively fuse heterogeneous and asynchronous signals from these sensors; 2) We develop a ranging system called BAIR. It has been successfully deployed in smartphones with a minimal storage footprint of just 16.41 MB; 3) We systematically evaluate the performance of BAIR system in static and moving settings. Our results demonstrate that our BAIR system provides a real-time ranging with high precision in short distances and high availability in long distances.

The rest of this paper is organized as follows: Section II reviews related works about indirect and direct ranging methods; Section III demonstrates the implementation of three modules in our work; Section IV details our methodology; Section V first illustrates the parameter settings when deployed on smartphones and then presents various experiments and discusses the results; Section VI concludes our work.

II. Related Works

Ranging can be categorized as indirect ranging and direct ranging.

A. Indirect Ranging

Indirect ranging, such as localization-based ranging, determines the relative distance between two devices by first acquiring their positions and then deriving the target results.

Based on the type of signal and sensors being used, localization methods can be classified to Wi-Fi-based [8]–[10], BLE-based [11], [12], and sound-based [3], [13] approaches. However, these methods require fixed anchor devices such as Wi-Fi access points, beacons, or speakers in the environment. They either need to be conducted in a lab or require significant manpower to deploy due to authorization policies and the non-universality of these devices in daily life.

Some localization works utilize the Klaman filter to fuse signals from different modules and perform well, such as [14], [15]. However, they do not focus on relative distance problems based on smartphones.

B. Direct Ranging

Direct ranging determines the relative distance between two devices with a single measurement.

1) Single-modal Ranging Methods: The most common ranging methods are sound-based methods [5], [6], [16]. These methods calculate relative distance based on the time of flight (TOA) between devices. Although they can achieve high precision (*cm*- or even *mm*-level error), they face challenges such as susceptibility to disturbances and limited range (around 3 meters). Other researchers have explored the use of BLE RSSI for ranging [7], but it can only attain *m*-level accuracy because RSSI is easily interrupted by people's movements and environmental changes.

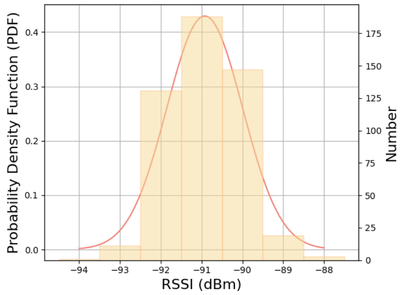


Fig. 1. **Distribution of BLE RSSI at 4 m.** To demonstrate the statistical properties of Gaussian distributions, we collect a total of 500 BLE packets. The orange histogram represents the number of BLE packets received at each RSSI value. The red curve is the fitting curve of the BLE RSSI distribution.

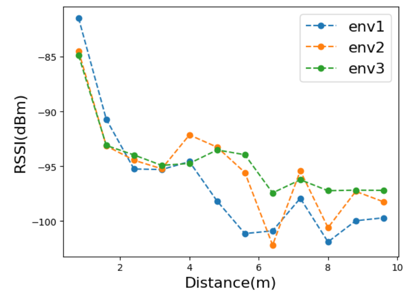


Fig. 2. **Relationship between BLE RSSI and distance in three different environments.** It is not ideally linear.

2) Multi-modal Ranging Methods: Due to the shortage of single-modal ranging systems in range or accuracy, more researchers become interested in multi-modal-fusion techniques [17]. Feishang et al. [18] demonstrate that multi-modal fusion can theoretically lower the information entropy's lower bound, which means better performance. [19], [20] use probability models, which are of high computational complexity and strong model dependence. Deep learning techniques [21], [22] perform well on this problem, but it demands a lot of time and effort to build a dataset and its memory footprint is huge, which hinders it from practical use.

In order to achieve real-time implementation on a resource-limited smartphone platform, we need to develop a simple and efficient fusion algorithm. As for the device's relative distance ranging problem, we can model the problem as a linear model. Therefore, a Kalman filter algorithm, which is computationally light, can be applied.

III. Ranging by Single Modal Sensors

In this section, we provide an outline of the implementation of BLE, IMU, and ultrasound-based methods in our work.

A. BLE-based Ranging

BLE is a wireless communication technology mainly used for short-distance (within a few dozen meters) data transmission. It offers low power consumption, high transmission rate, and swift connection speed, which meet the demand of IoT devices.

Nowadays, researchers have investigated BLE on a mobile phone and beacons fixed in the environment to estimate the user's location. Theoretically, the most used model of the relationship between BLE RSSI and transmission distance is $P = P_0 - 10n \lg(d/d_0) + \delta$ [23], as we only have access to BLE RSSI by the smartphone. In real-world sampling, nearby environments easily disturb the RF signal, leading to heavy fluctuations. We collect 500 BLE packets at 4 m and it shows a Gaussian distribution (shown in Fig. 1). To relieve the fluctuations, we adopt a Gaussian filtering model [11], which effectively filters out small probability, and short-term signal disturbances, thereby enhancing signal stability and accuracy.

In the real-world test, we discover a phenomenon that the relationship between RSSI and distance is not linear (shown in Fig. 2). There is a turning point around 4 meters, after which the RSSI becomes more volatile. We preliminarily judge that this may be caused by the system's adaptive frequency hopping (AFH) mechanism. That is, to improve the stability of data transmission and disturbance resistance, BLE will automatically choose its working frequency, leading to changes in transmitting power and receiver sensitivity. So we divide the task into two segments: short-distance (0 $\sim$ 4 m) and long-distance (4 $\sim$ 10 m) and use the Least Square (LS) method for curve fitting.

For real-time distance measurement, we use a weighted moving average to smooth real-time signals. However, extreme jumps due to environmental changes still occur occasionally even with Gaussian filtering. To mitigate these anomalies, we implement an outlier detection mechanism, retaining only data within the range $[\mu - 3\sigma, \mu + 3\sigma]$ both computed over a specified time window. Then, distance estimation is conducted using a pre-trained model based on filtered data.

B. Ultrasound-based ranging

The most common method is based on time of flight (ToF). By measuring the time difference between transmission and reception, the distance can be calculated by $d = \frac{c \cdot t}{2}$. However, the local clocks of different devices are not the same, which always causes estimation errors. To avoid this, we utilize the local elapsed time between two time-of-arrival (ETOA) [16]. The distance can be measured by

$$D = \frac{c}{2} \cdot ((t_{B1} - t_{A0}) + (t_{A3} - t_{B2}))$$

where t_{A0} and t_{B1} denote the time when the first sound wave was sent and received, t_{B3} and t_{A3} denote the time when the second sound wave was sent and received, and c is the speed of sound. Many experiments have demonstrated that temperature influences ranging errors because sound speed is temperature-dependent [24]. The current sound speed c is calculated using the formula $c = 331.4 + 0.61T$, where T is the current temperature in Celsius.

To find out the arrival time, we calculate the cross-correlation between recorded data and the reference signal. The surrounding noise causes a lot of messy peaks and the

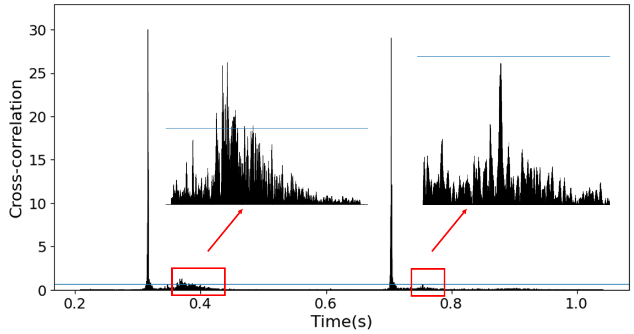


Fig. 3. **Example of cross-correlation in a time slot.** Two peaks are detected. The first is what we want and the second is caused by the multi-path effect. The cross-correlation of disturbance is zoomed in on the right.

multi-path effect may generate other peaks around the primary peak. To solve these issues, we employ two thresholds. The first one (blue line in Fig. 3) aims to filter the noise peaks. If the value of cross-correlation is smaller than the threshold, it must not be generated by the sound. The other is to judge the peak's received energy. If the energy of the received signal is smaller than the threshold, it can be considered not exit.

C. IMU-based ranging

We utilize the inertial sensor, which includes the accelerometer and gyroscope, embedded in smartphones. We collect the acceleration and angular velocity data from the smartphone with a sampling frequency of 50 Hz. To calibrate the minor offset of the phone IMU system, we have analyzed the sensor data with the help of Allan variance modeling [25].

Data obtained from accelerometers and gyroscopes is in the device coordinate. We need to convert them to ground coordinates in order to know human motion. To avoid Gimbal Lock, we employ the quaternion instead of the Euler Angle. Quaternion is important in coordinate conversion, which can be acquired from the Android System. Mostly, a quaternion is represented as a four-vector $\boldsymbol{q} = [q_w, q_x, q_y, q_z]$. The conversion formula is

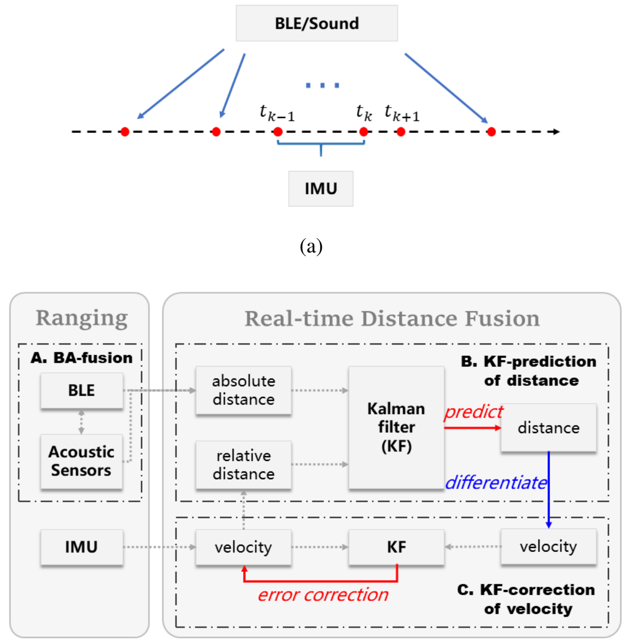
$$\boldsymbol{R}^g = \boldsymbol{q} \bigotimes \boldsymbol{R}^i \bigotimes \boldsymbol{q}^T$$

where $\boldsymbol{R}^g$ and $\boldsymbol{R}^i$ are denoted geocentric coordinate system and device coordinate system [26].

Subsequently, the acceleration data can be doubly integrated to determine the IMU's position relative to a known starting point. However, due to inherent noise in accelerometers and gyroscopes, directly integrating acceleration and angular velocity signals results in cumulative errors or "drift" that rapidly increase over time. Therefore, in practical applications, the IMU is used only for short-term distance estimation.

IV. System Design of BAIR

The schematic diagram of BAIR is shown in Fig. 4, which consists of two parts: 'Ranging' and 'Real-time Distance Fusion'. We have implemented the first part in Section III. We will give a detailed description of the second part in this section and it can further be divided into three blocks: BA-fusion, KF-prediction of distance, and KF-correction of velocity. The process is illustrated in Algorithm 1



(a)

(b)

Fig. 4. **Overview of the system.** (a) The red dots represent the absolute distance measuring times. During the intervals, the IMU works in pedestrian dead reckoning (PDR) mode. The update frequency of distance is 10 Hz. (b) The diagram illustrates the working flow of our proposed BAIR system. In the "Ranging" part (Section III), three modules provide their respective measurement results. In the "Real-time Distance Fusion" part (Section IV), the dual-Kalman filter is used to predict the relative distance and correct the velocity error of the IMU.

A. BA-fusion

If only BLE is available, the absolute distance is derived from BLE. When BLE and acoustic sensors are both capable of measuring distance, we choose acoustic sensors as our result provider because of its high accuracy. In short distances, environmental noises may cause sound detection failure, in which case BLE takes its role to prevent the system from collapsing or long-time drift. Failure is determined if we do not receive the ultrasound result in 3 seconds.

The interaction between BLE and acoustic sensors involves one device notifying the other via BLE before each sound measurement is initiated. This notification helps reduce false positive detections caused by other sounds in the environment. Additionally, the current distance information is useful for selecting the appropriate short-distance or long-distance fitting model for BLE.

B. KF-prediction of distance

The accelerometer data from the IMU is utilized to compute the current velocity and displacement for the last period using an inertial navigation algorithm. Once the absolute distance is obtained from acoustic sensors or BLE, the Kalman filter (KF)

Algorithm 1: Real-time Distance Fusion

Input: distance from BLE or ultrasound signals, velocity and acceleration from IMU
Output: predicted distance x_{now}

while *run* **do**
    if *velocity and acceleration* **then**
        integrate velocity $\Delta v = (a_{prev} + a_{now}) \cdot \Delta t_{IMU}/2$
        integrate distance $\Delta x = v \cdot \Delta t_{IMU} + \Delta v \cdot \Delta t_{IMU}/2$
        update velocity $v \leftarrow v + \Delta v$
        update distance $x_{rel} \leftarrow x_{rel} + \Delta x$
    end
    if *distance* **then**
        predict distance $x_{now} = KF(x_{abs}, x_{rel})$
        differentiate $v_{diff} = (x_{now} - x_{prev})/\Delta t$
        error correction $v \leftarrow KF(v_{IMU}, v_{diff})$
        reset $x_{ref} \leftarrow 0$
    end
end

prediction of distance executes, providing the current relative distance. The absolute distance acts as the observational input for the Kalman filter, while the relative distance represents the input information.

The Kalman filter method used in the fusion algorithm consists of two parts: state prediction and parameter correction:

1) State Prediction

$$\begin{cases} \hat{x}_k^- = \hat{x}_{k-1} + u_k \\ P_k^- = P_{k-1} + Q \end{cases} \tag{1}$$

where u is the input variable, x is the predicted variable, P is the covariance error, and Q is the observation noise error.

2) Parameter Correction

$$\begin{cases} K_k = \frac{P_k^-}{P_k^- + R} \\ \hat{x}_k = \hat{x}_k^- + K_k \cdot (z_k - \hat{x}_k^-) \\ P_k = (1 - K_k) \cdot P_k^- \end{cases} \tag{2}$$

where z is the observed variable, K is the Kalman gain, and R is the observation error.

C. KF-correction of velocity

The average velocity over the last period is calculated using $\frac{\Delta d}{\Delta t}$, where Δd denotes the distance difference and Δt is the time difference from the last period. This average velocity functions as observational data for the Kalman filter, while the accumulated IMU velocity acts as the input data. The predicted velocity from the Kalman filter for the current period is then used as the initial velocity for the subsequent period. After each cycle, the IMU's accumulated distance and velocity are reset to zero, preventing long-term error accumulation due to IMU integration.

The IMU operates at a frequency of up to 50 Hz, whereas BLE and acoustic sensors work at a much lower frequency, approximately 3 Hz and 1 Hz. So their updates do not coincide. To address this discrepancy, the dual-Kalman filter executes only upon receiving measurement results from BLE or sound. During the intervals, dead reckoning is performed using the IMU data (shown in Fig. 4(a)).

The accuracy and effective range of BLE and acoustic sensors differ significantly. In scenarios where one device is mobile, and the other remains stationary, distance sensing is categorized into short-distance and long-distance phases. BLE offers meter-level accuracy over distances up to 10 meters, while sound provides centimeter-level precision within a range of less than 5 meters. Consequently, the sound is employed during the short-distance phase and BLE during the long-distance phase. This approach also conserves smartphone energy since generating high-frequency sound signals is more power-intensive than BLE.

This approach leverages the IMU's high sampling rate and responsiveness while mitigating the drift issues. By integrating IMU measurements with data from other sensors such as Bluetooth and acoustic sensors, we enhance ranging accuracy and stability. The absolute distance data from these additional sensors calibrates and corrects the IMU's output, while the displacement calculated by the IMU provides constraint information for BLE and ultrasound ranging.

V. Implementation and Experimental Results

In this section, we first introduce the implementation of BAIR on Android system and the experiment settings. Then, we demonstrate the performance of our method in static scenarios and moving scenarios. Next, we explore the impact of some variables on performance. Finally, we point out the memory occupation and power consumption of the system on a smartphone.

A. Implementation

We have implemented BAIR on Android systems. For the BLE-based ranging method, we set the lower threshold of the Gaussian filter modal as 0.6. The transmitting frequency of BLE is 3 Hz. For the ultrasound-based ranging method, the sampling rate of sound is set to 48000 Hz. The duration of each chirp is 60 ms. To ensure that most people are insensible to the modulated chirp signals, we use a 3 kHz bandwidth, sweeping from 17.5 kHz to 20.5 kHz, which is nearly ultrasonic. Although a larger frequency can not be heard either, the signal is distorted due to hardware limitations. For IMU, the sampling rate is 50 Hz. Kalman filter parameters P and Q are chosen according to [27].

B. Experimental Setup

Fig. 5(a) gives an outline of the evaluation environment. It is the main passage of our office building. The length and width of the floor tiles are both 0.8 m. In order to get a realistic assessment, we follow daily scenarios without controlling the movement and whisper of non-staff members. We use two

(a) (b)

Fig. 5. **Experiment setups.** (a) The evaluation environment. The red stars mark the position of the server and client. The blue dashed line represents the moving track. (b) The evaluation setting. The server and Client are marked with red boxes and indicated nearby.

smartphones to evaluate the ranging performance. HUAWEI P40 Pro is the server and HUAWEI Mate40 Pro is the client. The server is fixed on the tripod and the client is held by a volunteer (as shown in Fig. 5(b)).

C. Static Scenario

In static scenarios, we select a set of test points ranging from 0.8 m to 9.6 m, with intervals of 0.8 m. At each point, we collect 10 results three times.

a) Overall performance: The average of 30 results for each test point is depicted in Fig. 6. The ultrasound-based ranging method can measure distances within 6 m and its MAE is 1.9 cm. Although it can sometimes successfully measure when the distance exceeds 6 m, the high probability of failure prevents us from considering this case. The BLE-based ranging method can measure distances up to 10 m but its accuracy is much lower than the ultrasound-based method. As we analyzed earlier, BLE RSSI's turning point is around 4 m, which also causes the ranging error near 4 m. The two-stage fit trick performs well except for the boundary point. The MAE of the BLE-based ranging method is 99 cm. Our proposed BAIR achieves 11 cm MAE and increases the ranging distance to 10 m. It demonstrates that our fusion algorithm has a practical effect, broadening sound's ranging distance and improving BLE's ranging accuracy. Considering two different use cases - interaction between devices in short distances and arrival detection in long distances, we calculate MAE below 4 m and MAE from 4 m to 10 m. They are 2 cm and 18 cm, respectively. Our goal of achieving high accuracy at short distances and maintaining feasibility at long distances is realized in the static scenario.

b) Comparison with other technologies: Here, we compare BAIR with other wireless ranging technologies across five different aspects (as shown in TABLE I). CAT [5], SwordFight [28] and SCALAR [6] utilize ultrasound to measure the distance between two devices achieving high accuracy and real-time response. However, CAT requires a series of speakers in addition to the smartphones. The SwordFight and SCALAR systems suffer from limited range, which may not suffice for many indoor interaction purposes. Shino Shiraki et al. [7] use BLE RSSI to determine the relative distance but still need beacons to assist, and its error margin in static scenarios is quite large.

Our proposed BAIR system integrates BLE RSSI, ultrasound signals, and IMU, only requiring two COTS smartphones without any hardware modifications. With a range of 10 meters, it surpasses other mentioned technologies and is well-suited for most indoor scenarios. Additionally, its dual-Kalman algorithm ensures real-time response.

c) Results under different environments: To explore the impact of different environments on our BAIR system, we conducted experiments in four different environments: meeting room, office, passage, and playground (as shown in Fig. 9). The MAE results are presented in TABLE II. The data shows that BAIR performs better in spacious environments with few obstacles, such as passages and playgrounds, compared to more complex indoor environments like meeting room and office. The primary reasons for this disparity are that noisy environments lead to the early failure of ultrasound signals, and complex environments cause more multi-path effects.

D. Moving Scenario

In the moving scenarios, volunteers are required to move towards and backwards to the server three times. Suppose that our track is a straight line. We shoot a video for each term and manually align ground truth.

a) Different fusion combinations: Fig. 7 presents the Cumulative Distribution Function (CDF) curves for our proposed system alongside other variants from ablation studies. Notably, our proposed system outperforms the other four configurations at both the CDF 0.5 and CDF 0.8 marks. These results compellingly illustrate the efficacy of our multi-modal fusion approach. Interestingly, while the Ultrasound+BLE+IMU configuration demonstrates slightly inferior performance compared to the Ultrasound+IMU setup at the CDF 0.95 point, it compensates with a considerably extended ranging distance. Specifically, Ultrasound+BLE+IMU remains effective up to 10 m range, whereas the Ultrasound+IMU combination begins to fail beyond 5 meters. This highlights the trade-offs between accuracy and range. The results underscore the superior overall capability of our proposed multi-modal fusion system, which attains high accuracy in short distances and feasibility in long distances.

b) Impact of the moving speed: The volunteers are required to walk at three different speeds: slow (0.1 m/s), middle (0.2 m/s), and fast (0.4 m/s). The results are shown in Fig. 8. The algorithm error increases with the increase of speed because a larger speed means a larger fluctuation of the received signal leading to the ranging error of BLE and ultrasound. The ranging error of three different speeds is less than 0.2 m, 0.3 m, and 0.4 m when the value of CDF is equal to 0.5, 0.8, and 0.95, respectively. The difference between them is not obvious. It turns out that our KF-correction of velocity plays a role.

c) Impact of device type: To show that our system is applicable to other devices, three smartphones of HUAWEI are used in the test. The types of their models, operation

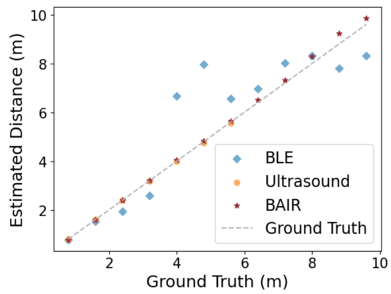


Fig. 6. **Distance ranging comparison in static scenario.**

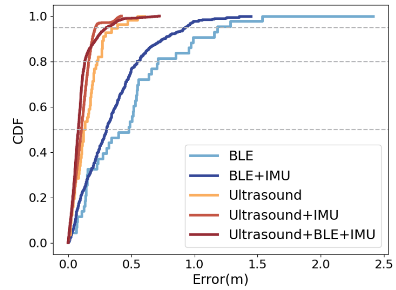


Fig. 7. **CDF comparison for different sensor combinations.**

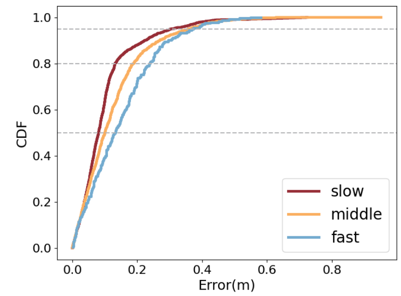
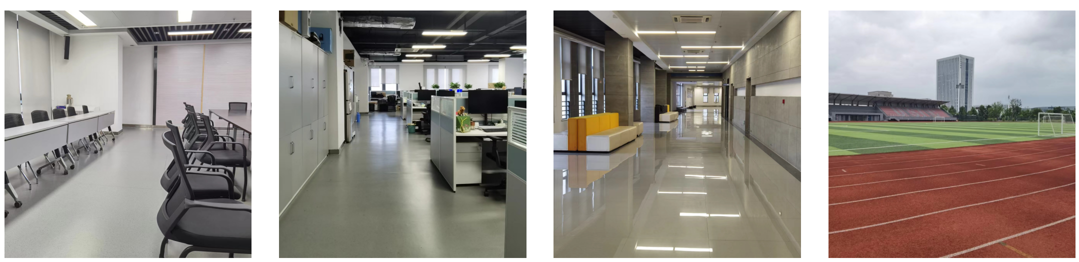


Fig. 8. **CDF comparison for different moving speed.**

TABLE I
COMPARISON WITH OTHER TECHNOLOGIES

Technology	Sensing Modality	Universality	Range	Error (Static Scenarios)	Real-Time
CAT [5]	Ultrasound	✗	7 m	9 mm	✓
SwordFight [28]	Ultrasound	✓	3 m	2 cm	✓
SCALAR [6]	Ultrasound	✓	3 m	0.39 mm	✓
Shino Shiraki et al. [7]	RSSI	✗	15 m $\times$ 30 m	0.74 m	-
BAIR	**RSSI, Ultrasound, IMU**	✓	**10 m**	**11 cm**	✓



(a) Meeting room (b) Office (c) Passage (d) Playground

Fig. 9. **Four different test environments.**

TABLE II
MAE UNDER DIFFERENT ENVIRONMENTS

	meeting room	office	passage	playground
MAE	15 cm	17 cm	11 cm	8 cm

TABLE III
DIFFERENT TYPES OF SMARTPHONES

Smartphone Type	Operation System	Processor
P40 Pro	Harmony 4.0.0	HUAWEI Kirin 990 5G
Mate40 Pro	Harmony 4.0.0	HUAWEI Kirin 9000
Nova10 SE	Harmony 4.0.0	Qualcomm Snapdragon 680

TABLE IV
ERROR CORRESPONDING TO DIFFERENT CDF VALUES FOR DIFFERENT SMARTPHONE PAIRS

	CDF 0.95	CDF 0.8	CDF 0.5
P+M	29.11 cm	17.95 cm	10.39 cm
P+N	41.19 cm	23.18 cm	15.13 cm
M+P	34.28 cm	18.63 cm	12.21 cm
M+N	51.19 cm	27.56 cm	18.23 cm
N+P	32.12 cm	21.94 cm	11.51 cm
N+M	32.29 cm	22.38 cm	12.47 cm

systems, and processors are shown in TABLE III. For each pair of test devices, the BLE offline model should be trained again. The results of the combination of different servers and clients are displayed in TABLE IV. In the first column, P, M, and N represent the P40 Pro, Mate40 Pro, and Nova10 SE, respectively. The standard configuration is “Server+Client”.

When the CDF equals 0.5, the errors are approximately 10 cm, which is comparable to the performance of the Ultrasound+IMU setup at short distances. This indicates that ultrasound plays the dominant role in short-distance measurements. However, from CDF 0.5 to CDF 0.95, there is a significant increase in ranging errors. This surge occurs because, when ultrasound fails to measure accurately, BLE takes over, and BLE’s accuracy is substantially lower than that of ultrasound. The performance is acceptable because the ratio of error to distance is not large, which meets our requirements of high accuracy at short distances and reliability at long distances.

Due to hardware differences, the performance of different types of smartphones varies. We found that the performance is more influenced by the client device, particularly its IMU. For instance, when the Nova10 SE serves as the client (in P+N and M+N configurations), the ranging error is larger compared to other pairs. However, when the Nova10 SE acts as the server (in N+P and N+M configurations), its performance is comparable to smartphone pairs that do not include the Nova10 SE.

E. Memory Usage and Power Consumption

In order to be practically deployed on mobile devices without affecting the use of other applications or functions, our proposed system needs to be small in both memory usage and power consumption. For memory occupation, we check the app management in the settings of the smartphone and find that it is 16.41 MB, which is much smaller than most existing applications. For power consumption, we let our system run 1 hour. In BLE+IMU mode, the smartphone battery dropped 5%. In Ultrasound+BLE+IMU mode, the smartphone battery dropped 15%. It shows that ultrasound generating consumes much more power than BLE. So we can do a trade-off between long-distance ranging accuracy and power consumption. The strategy is ultrasound ranging is enabled only when the device enters the close-range scene.

VI. Conclusion

In this work, we present BAIR for real-time relative distance measurement between two devices using Commercial Off-The-Shelf (COTS) BLEs, acoustic sensors, and IMUs on smartphones. A key innovation is employing a lightweight improved Kalman filter to fuse the heteroid and asynchronous signals from three modules. Its memory footprint is 16.41 MB. BAIR has been tested with extensive experiments. It shows great universality when the environment and devices change, and keeps the error at decentimeter level both in static and moving scenarios. Our work holds promise for the future, as it can be utilized in a ubiquitous real-time location system. Our goal is to develop a 2-D localization system based on more devices.

References

[1] S. Eleftherakis, G. Santaromita, M. Rea, X. Costa-Pérez, and D. Giustiniano, "Spring+: Smartphone positioning from a single wi-fi access point," *IEEE TMC*, 2024.

[2] X. Tong, H. Wang, X. Liu, and W. Qu, "Mapfi: Autonomous mapping of wi-fi infrastructure for indoor localization," *IEEE TMC*, vol. 22, no. 3, pp. 1566–1580, 2023.

[3] G. Guo, R. Chen, K. Yan, Z. Li, L. Qian, S. Xu, X. Niu, and L. Chen, "Large-scale indoor localization solution for pervasive smartphones using corrected acoustic signals and data-driven pdr," *IEEE Internet of Things Journal*, 2023.

[4] M. Murata, D. Ahmetovic, D. Sato, H. Takagi, K. M. Kitani, and C. Asakawa, "Smartphone-based indoor localization for blind navigation across building complexes," in *IEEE PerCom*, 2018.

[5] W. Mao, J. He, and L. Qiu, "Cat: high-precision acoustic motion tracking," in *MobiCom '16*. ACM, 2016, p. 69–81.

[6] L. Wang, H. Wan, T. Zhao, K. Sun, S. Shi, H. Dai, G. Chen, H. Liu, and W. Wang, "Scalar: Self-calibrated acoustic ranging for distributed mobile devices," *IEEE TMC*, vol. 23, no. 2, pp. 1701–1716, 2024.

[7] S. Shiraki, A. Suzuki, T. Uehara, Y. Ohashi, and S. Shioda, "Indoor pedestrian localization methods using contact information from bluetooth low energy beacons between smartphones," in *IEEE 95th VTC-Spring*, 2022, pp. 1–7.

[8] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *SIGCOMM '15*. ACM, 2015, p. 269–282. [Online]. Available: https://doi.org/10.1145/2785956.2787487

[9] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "Indotrack: Device-free indoor human tracking with commodity wi-fi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, 2017. [Online]. Available: https://doi.org/10.1145/3130940

[10] M. Rea, T. E. Abrudan, D. Giustiniano, H. Claussen, and V.-M. Kolmonen, "Smartphone positioning with radio measurements from a single wifi access point," in *CoNEXT '19*. ACM, 2019, p. 200–206.

[11] Z. Jianyong, L. Haiyong, C. Zili, and L. Zhaohui, "Rssi based bluetooth low energy indoor positioning," in *IPIN*. IEEE, 2014, pp. 526–533.

[12] A. Awad, T. Frunzke, and F. Dressler, "Adaptive distance estimation and localization in wsn using rssi measures," in *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools*, 2007, pp. 471–478.

[13] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *MobiSys '15*. ACM, 2015, p. 15–29. [Online]. Available: https://doi.org/10.1145/2742647.2742662

[14] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of wifi, smartphone sensors and landmarks using the kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, 2015.

[15] R. M. B. Reza Zekavat, *An Introduction to Kalman Filtering Implementation for Localization and Tracking Applications*. Wiley-IEEE Press, 2019, pp. 143–195.

[16] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *SenSys '07*. ACM, 2007, p. 1–14. [Online]. Available: https://doi.org/10.1145/1322263.1322265

[17] X. Niu, K. Zou, D. Shen, S. Drew, S. Wu, G. Guo, and R. Chen, "Ultramotion: High-precision ultrasonic arm tracking for real-world exercises," *IEEE TMC*, vol. 23, no. 2, pp. 1846–1862, 2024.

[18] F. Shang, H. Du, P. Yang, X. He, W. Ma, and X.-Y. Li, "Towards the limits: Sensing capability measurement for isac through channel encoder," *arXiv preprint arXiv:2405.09497*, 2024.

[19] S. He, S.-H. G. Chan, L. Yu, and N. Liu, "Maxlifd: Joint maximum likelihood localization fusing fingerprints and mutual distances," *IEEE TMC*, vol. 18, no. 3, pp. 602–617, 2019.

[20] S. Sun, S. Li, Y. Li, B. Moran, and W. S. Rowe, "Smartphone user tracking by incorporating user orientation using a double-layer hmm," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7780–7790, 2022.

[21] G. Guo, R. Chen, K. Yan, P. Li, L. Yuan, and L. Chen, "Multichannel and multi-rss based ble range estimation for indoor tracking of commercial smartphones," *IEEE Sensors Journal*, 2023.

[22] X. Kong, C. Wu, Y. You, Z. Lv, and Z. Zhao, "Hybrid indoor positioning method of ble and monocular vins based smartphone," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[23] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, 1980.

[24] N. A. Singh and M. Borschbach, "Effect of external factors on accuracy of distance measurement using ultrasonic sensors," in *ICSigSys*, 2017, pp. 266–271.

[25] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart, "Maximum likelihood identification of inertial sensor noise model parameters," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 163–176, 2016.

[26] H. Wang, C. Xue, Z. Wang, L. Zhang, X. Luo, and X. Wang, "Smartphone-based pedestrian nlos positioning based on acoustics and imu parameter estimation," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23 095–23 108, 2022.

[27] S. Akhlaghi, N. Zhou, and Z. Huang, "Adaptive adjustment of noise covariance in kalman filter for dynamic state estimation," in *IEEE Power & Energy Society General Meeting*, 2017, pp. 1–5.

[28] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, "Swordfight: enabling a new class of phone-to-phone action games on commodity phones," in *MobiSys '12*. ACM, 2012, p. 1–14. [Online]. Available: https://doi.org/10.1145/2307636.2307638