

VisBLE: BLE Tag Identification with Vision Enhancement

Abstract—IoT devices have penetrated every corner of our life. With a large number of IoT devices around, it becomes an interesting question how to identify them in an intuitive way. In this work, we take as an example a popular IoT technology, Bluetooth Low Energy (BLE), which is widely utilized in location ID and consumer promotion. Instead of the legacy method looking for a target BLE device in a long device list, we propose a novel vision-based interaction method on the camera screen, called VisBLE. VisBLE takes advantage of the new localization capability introduced in BLE 5.1 and advances in vision technologies for high accuracy, robust, and intuitive BLE device identification. There are two novel technical mechanisms: i) a rotation-based wireless localization mechanism that accurately and robustly locates the BLE transmitter in the camera coordinate and ii) a homography-based matching mechanism that identifies target BLE devices with high accuracy on the camera screen. We develop a prototype and deployed it on the Nexus 5x smartphone and the CC26X2 development board. Our results show that VisBLE identifies a BLE device with over 90% accuracy.

I. INTRODUCTION

With the advent of the Internet of Things (IoT) era, the number of IoT devices surges in recent years. By 2035, the number of IoT devices is estimated to reach a trillion [1]. Among them, Bluetooth Low Energy (BLE) has been widely adopted to pinpoint the location of interest (LOI) and for contact tracing. With a large number of BLE devices around, it becomes an interesting question how to identify them in an efficient and intuitive way.

Some exiting approaches exist in daily life and the research community coping with the BLE identification problem, but they have limitations in one kind or another. In daily life, users usually assign a human-friendly name to each BLE device [2] and leverage the device name to identify physical devices. Though commonly used, such an approach is not applicable when the device name is garbled or not understandable to the user. In the research community, the identification problem is scoped as a 3D wireless localization problem [3]. This line of studies adopts advanced radio technologies to seek centimeter-level accuracy [4]. Such an approach often relies on wide-bandwidth signal and a large-size antenna array. However, the which are not commonly supported on a mobile devices. In addition, outputting the target's 3D coordinate is not intuitive for ordinary users.

The opportunity comes from incorporating Augmented Reality (AR) in the BLE localization. BLE starts to support the angle-of-arrival (AoA) based localization technologies from the BLE 5.1 specification. It empowers the BLE devices with the capability of angular measurement. On the AR side, the

adoption of vision technologies such as image segmentation algorithms enables masking semantic objects in camera view and identifying visual objects. In addition, AR also provides an intuitive interface for interaction with human body to assist the localization task. These two lines of research, when working together properly, allow us to achieve high accuracy and intuitive BLE device identification. A conceptual use case is illustrated in Figure 1. Instead of looking for each in a long Bluetooth device list, a user opens the camera and scans the whole room. The connectable devices will appear on the screen as clickable AR objects. The user then operates on these devices directly from the screen. Besides in an office, other intuitive BLE identification application scenarios include malls, restaurants, and hospitals. Note that although we discuss specifically on BLE 5.1, the same idea works on any wireless radio with AoA capability.

Despite a simple idea, technology side is not straightforward. First, though AoA angle can be obtained from the commercial devices, it is a 2D-plane angle which cannot uniquely determine the position of the target device in the 3D world. Second, since a BLE device may come in various product forms, the vision technology alone cannot well identify the device of interest. Third, a straightforward fusing approach based on the pixel proximity [5] has a low accuracy since it does not make well use of the angular information.



Fig. 1: A conceptual scene of the VisBLE use case in the office to operate on various BLE devices from the screen.

In this paper, we present VisBLE, a system that enhances the BLE 5.1 device identification by combining wireless localization technology and AR. VisBLE is built with two novel technical mechanisms: (i) a rotation-based wireless localiza-

tion mechanism that accurately and robustly locates the BLE transmitter in the camera coordinate and (ii) a homography-based matching mechanism that identifies the target BLE device with high accuracy in the camera screen. The technical contributions of the work are four-fold;

- To the best of our knowledge, this work is the first to explore an intuitive vision-based Bluetooth 5.1 device identification. It paves the way for many AR applications as well as novel wireless interaction technologies.
- On the wireless side, we propose a novel azimuth and elevation estimation mechanism based on the Level meter (an electronic level) that converts AoA information obtained from the antenna to high accuracy and robust point cloud in the camera screen.
- On the vision side, we utilize the homography matrix to match the point cloud and the mask segments through both the pixel proximity and the depth of field. It well separates objects in the foreground and the background of the picture.
- We develop a prototype and evaluated the performance of VisBLE on the Nexus 5x smartphone and the CC26X2 development board. Our results show that VisBLE identifies a BLE device with over 90% accuracy.

The rest of the paper is organized as follows. Section II motivates the need for VisBLE and opportunity, while Section III overviews the design. Section IV and Section V describe the key designs: AoA point cloud extraction algorithms and Vision enhanced BLE identification, respectively. Section VI and Section VII presents the implementation and evaluation of VisBLE, followed by the related work and conclusion in Section VIII and Section IX respectively.

II. MOTIVATION

This work is motivated by the primitive device identification method in BLE devices. Under the current identification logic, a user needs to look for the target Bluetooth device in a long list of device names or universally unique identifiers (UUIDs) to launch a connection request. Such a text-based strategy is now challenged by the pervasive use of Bluetooth technology for location-based services and contact tracing. Its limitations are appearing in some practical scenarios: (i) when the device name is written in a foreign language or in an acronym which is not understandable to users; (ii) when the incompatible character encoding scheme, e.g., ASCII vs. GBK, makes the device name unreadable hex characters at the user's device; and (iii) when there are multiple identical BLE devices, e.g., multiple smart bulbs, which cannot be easily distinguished from either name or appearance. In the above scenarios, it calls for a more intuitive way of device identification than the text-based solution.

A. Limitations of the SOA

In the literature, the identification problem is formulated as a 3D wireless localization problem [3]. It assumes that with high enough localization accuracy, say centimeter-level, the target device is identified [4], [6]. However, such a strategy confronts

a dilemma in practice. On the one hand, a narrowband BLE device by its nature is limited in localization accuracy. On the other hand, advanced radio technologies, though capable to increase localization accuracy, requires a large-size antenna array which is rarely supported in most mobile devices. To break the dilemma, we argue that the wireless localization problem is not equivalent to the device identification problem. First, localization algorithms output 3D coordinates which are not intuitive to users. Second, centimeter-level high-accuracy localization can be an over-kill for applications that only need to identify target BLE devices. These insights inspire us a new vision-based strategy to identify BLE devices in an accurate, robust, and intuitive way.

B. Opportunity

Our opportunities come from the advances in the latest Bluetooth specifications as well as the vision-based technologies. On the one hand, in the Bluetooth 5.1 specification released in 2019 [7], the “direction finding function” was introduced into the core specification for the first time. Two localization elements are proposed, namely the angle of arrival (AoA) and the angle of departure (AoD). AoA and AoD measure the angle of the target device through the differences of the signal arriving multiple antenna. Compared with existing BLE devices that only capable of proximity measurement through RSSI, the AoA and AoD are much more accurate due to the new capability of angular measurement.

On the other hand, advances in computer vision identify the visual objects in the camera screen and mark them with bounding boxes or masks [8]–[10]. They assign each bounding box and masked area a semantic label. From the perspective of localization, the bounding boxes and the masks provide help cross-validate the accuracy of the wireless approach. From the perspective of the identification, compared with the output of point cloud given by most wireless localization approaches, the bounding boxes and the masks are more intuitive to the users. They pave the way for developing AR applications and other complex sensing and recognition tasks [11]–[14].

C. Challenges

It is not straightforward to integrate the AoA information with the mask segments. First, the AoA information read from the Bluetooth device is 2D AoA information in the antenna coordinate. It cannot uniquely determine the position of the target device in a 3D world. Second, it is non-trivial to fuse the AoA angular information with the screen pixels information as in mask segments. Finally, proximity in the camera screen does not mean proximity in the real world due to the different depth of each object in the screen.

III. VISBLE OVERVIEW

This section presents the overview of VisBLE, a system that identifies BLE devices and masks their positions on the screen of the users' mobile devices. Figure 2 illustrates the workflow of VisBLE, which is divided into two components, i.e., (i) the AoA point cloud extraction and (ii) the vision enhanced BLE identification.

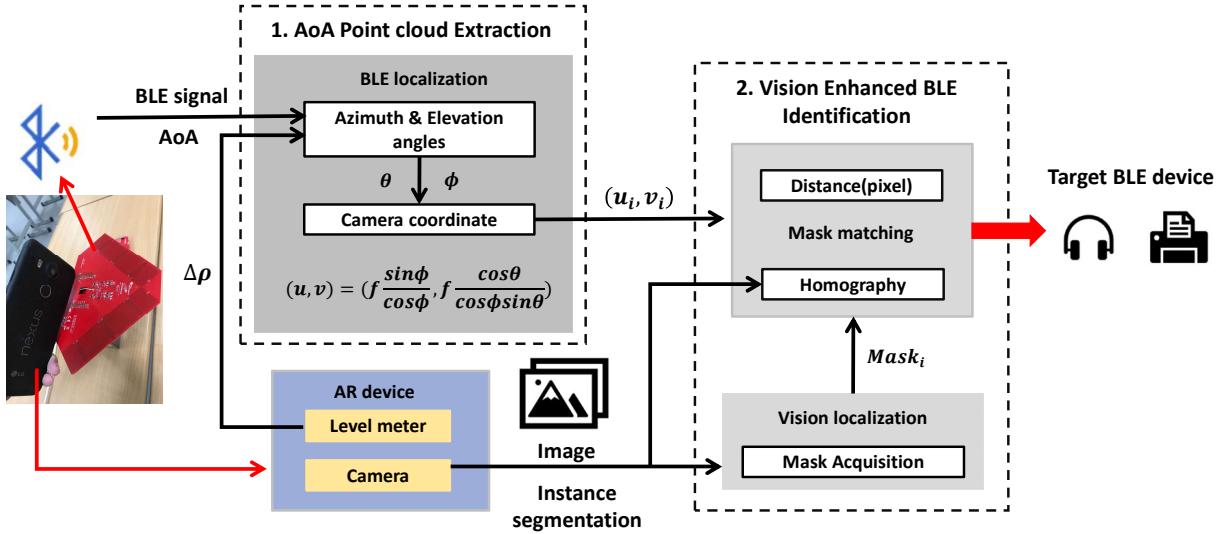


Fig. 2: VisBLE overview.

IV. AOA POINT CLOUD EXTRACTION

This section illustrates the VisBLE design in detail. We first give background about the AoA information obtained from BLE 5.1 devices. Then we introduce how to estimate the azimuth and elevation angles from the AoA.

From Bluetooth 5.1, AoA-based localization is introduced into the specification. It enables a BLE receiver to directly obtain the arrival angle of another BLE transmitter either in the connection mode or the connectionless-mode. We here review the AoA estimation approach adopted in the commercial development kit and the geometric relation between AoA and the imaging process of the camera. Under the far-field assumption, the angle of the arriving (AoA) signal Θ can be obtained by the phase difference Ψ of signals arriving at the antenna, as shown in Equation 1

$$\cos(\Theta) = \frac{\lambda\Psi}{2\pi d}, \quad (1)$$

where λ represents the wavelength of the incoming signal and d represents the interval between antennas. The AoA information, however, is not enough to localize a target object in the 3D world because (i) the distance information is lacking and (ii) the AoA angle obtained is only the azimuth angle in the antenna plane, without elevation information. To overcome these two challenges, we take the advantage of the camera and the Level meter available on mobile devices.

A. AoA on Camera Coordinate

With AoA alone, we are still not able to locate the target device, because the depth, i.e., distance r between the target object and the receiving device, is unknown. Traditional distance estimation relies on RSSI measurement or the Time of Flight (ToF). Their accuracy is low and unstable especially for low-end IoT devices [15], [16]. To address the issue of lacking distance measure, we use a coordinate transformation. In detail, instead of locating the target Bluetooth device in

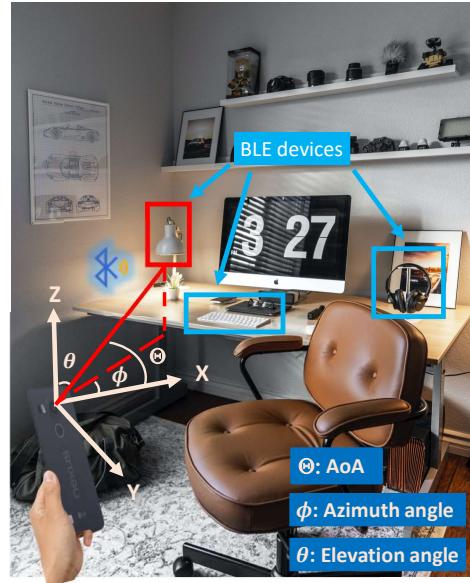


Fig. 3: Pinhole imaging.

the world coordinates, we find its location in the 2D camera coordinate of the AR device directly [5]. A special feature of such a transformation is that the distance r is eliminated. In such a manner, we bypass the localization accuracy loss from distance estimation.

Figure 3 illustrates the relationship between the AoA, azimuth angle, and the elevation angle.

$$(u, v) = \left(f \frac{\sin\phi}{\cos\phi}, f \frac{\cos\theta}{\cos\phi\sin\theta} \right) \quad (2)$$

where ϕ is the azimuth angle, and θ is the elevation angle. With the help of the spherical coordinates. From Equation 2, we find that the distance r is eliminated for the 2D position of the target device in the camera screen. Compared with the traditional localization estimation relying on distance

estimation, the angle of the signal incident estimation assisted by the antenna array is relatively more robust [15], [16].

B. Azimuth and Elevation Estimation

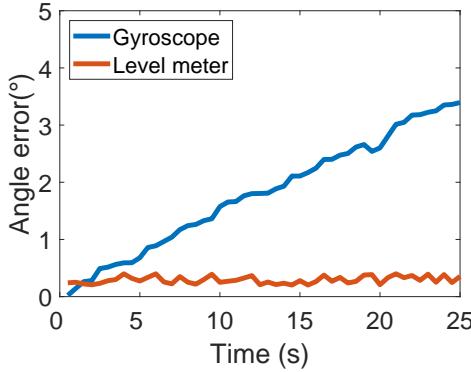


Fig. 4: Gyroscope vs., Level meter

From Section IV-A, we see that the position of the target Bluetooth device is determined in the camera screen if the azimuth and elevation angles are known. Then the question becomes how to estimate the azimuth and elevation angles. We will review several popular approaches in this subsection and propose a novel azimuth and elevation angle estimation approach based on the Level meter in the next subsection.

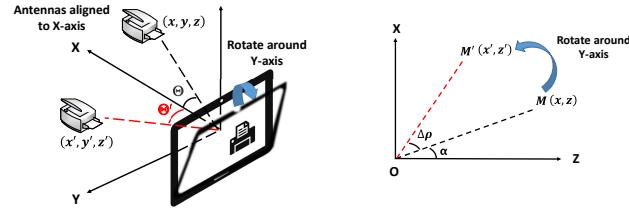
Direct Measurement: In terms of azimuth and elevation angles estimation, the direct method is to configure the AR device with the specialized antenna array for angle estimation, such as uniform circular array [17], [18], L-shaped array [1], [19], or parallel linear array [20], [21] with 8 or more antennas. However, on mobile devices, this approach is hardly applicable due to the limited space for the antenna.

With Motion Sensors: The azimuth and elevation angles are measured through device motion tracked by gyroscope or the inertial measurement unit (IMU) sensor (using gyroscope internally) [5]. These sensors commonly estimate the angular acceleration of the device's movement. They then calculate the angle of rotation of the device in real-time through integral calculation. The limitation of the integration is the accumulating drift as the measurement time increases. In Figure 4 we measured the accumulating error of the gyroscope over time. We find the gyroscope introduces 2.5° error after 18 seconds. In our study, we find it not necessary to track the motion of the device to estimate the azimuth and elevation angles. Instead, obtaining the start and final gestures of the mobile device is more than enough. Note that the Level meter derives data by processing the raw data from the accelerometer and the geomagnetic field sensor. The Level meter obtains the angle between the mobile device and the horizontal plane in a certain state, which is a state quantity. In such a manner, we eliminate the accumulating errors and get more robust azimuth and elevation angles estimation over time.

C. Level meter Based Angle Estimation

Inspired by this finding, we propose to utilize the Level meter available in the mobile devices to estimate the azimuth and

elevation angles. The Level meter gives the angle between the mobile device and the horizontal plane in a certain state, which is a state quantity. Compared with IMU sensors that require dynamic integral calculation, the state quantity provided by the Level meter is more reliable without accumulating errors in long-term measurements.



(a) Rotate the device along the Y- axis. (b) Projection of the target device on the XZ plane.

Fig. 5: Spatial angle relation.

The operations we take to estimate the azimuth and elevation angles are as follows: We first take the angle between the device and the horizontal plane obtained by the Level meter in the AR device ρ and the AoA Θ obtained by the Bluetooth device as a known quantity. Then we rotate the device to change the position of the antenna and get another Level meter reading. The rotation angle is calculated from the Level meter reading difference before and after the rotation. After that, we calculate the azimuth and elevation angles from the rotation angle and AoA. The azimuth and elevation angles determine the pixel coordinates of the target Bluetooth device in the camera screen following Equation 2.

We hereby derive how to calculate the azimuth angle ϕ and elevation angle θ from the rotation angle and AoA. Without loss of generality, we assume the receiving antenna is located on the X-axis and the rotation is around the Y-axis as shown in Figure 5. When we rotate the device around the Y-axis, its Y-axis coordinate remains unchanged: $y = y'$. Based on this observation, the relation is as follows:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\Delta\rho & 0 & \sin\Delta\rho \\ 0 & 1 & 0 \\ -\sin\Delta\rho & 0 & \cos\Delta\rho \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (3)$$

To calculate the azimuth and elevation angles, we transform the coordinate of the target device to the spherical coordinates. The relation between AoA and the azimuth and elevation angles is:

$$\sin(\theta) \cos(\phi) = \cos\Theta \quad (4)$$

Since the horizontal angle $\Delta\rho$ and AoA Θ are known quantities, we calculate 1) elevation angle θ 2) azimuth angle ϕ

$$\theta = \arccos \frac{\cos(\Theta') - \cos(\Delta\rho)\cos(\Theta)}{\sin(\Delta\rho)} \quad (5)$$

$$\phi = \arccos \frac{\cos(\Theta)}{\sin(\theta)} \quad (6)$$

Through the calculation of the above Equations, we estimate the azimuth and elevation angles of the target Bluetooth device relative to the AR receiving device, and then estimate the

position of its signal source in the camera screen. It should be noted that if the sender is outside of the camera's viewing angle, (u, v) values from Equation 2 become infeasible (i.e., either smaller than zero or larger than the number of pixels of the video).

Compared with the state of the arts, the Level meter based angle estimation has several advantages:

- **Less accumulating errors.** We only need to know the difference between the angles before and after the rotation without tracking the angle changes during the rotation.
- **Easier operation.** We only need to rotate around a single axis once. Also, during the rotation of the device, we no longer perform threshold estimation on AoA [5], just collect the AoA and Level meter readings.
- **Low computational cost.** The proposed angle estimation algorithm causes low computing overhead and runs in real-time in mobile AR devices

V. VISION ENHANCED BLE IDENTIFICATION

Through the azimuth and elevation angle estimation mechanism, we estimate the position of the target device in the camera screen, but that is only halfway done. The pure wireless solution is limited in two aspects: (i) relative low accuracy and unstable localization due to the simple antenna and fast-changing environment and (ii) scattered point cloud that cannot directly mask the target object. In this section, we introduce how to borrow ideas from computer vision technology to locate the interactable BLE device in the camera screen.

A. Mask Acquisition

In contrast to the wireless localization approaches that generate point clouds, the computer vision algorithms segment the image into semantic objects. One state-of-the-art framework is the Mask R-CNN [9]. It clusters the semantic information of each feature point to generate a Mask of the target object. Compared with edge detection-based approaches, such as SIFT [22] or SURF [21] that are based on color gradients, segmentation masks in Mask R-CNN are more rational and interpretable with pre-trained models over large labeled datasets. So it is more suitable for the identification tasks in this work.

Although Mask R-CNN obtains accurate segmentation masks of the semantic objects in the image, the matching between the segmentation masks and the point cloud is not straightforward. Computer vision algorithms deal with pixels in the image, they are blind to the information obtained by the wireless technology through the signal or the wireless frames, such as the signal strength, AoA, and the device ID. In addition, the imaging information and the wireless information are of different models, they have to be transformed to the same modal for comparison and fusion. In the next subsections, we will introduce how we resolve such a multi-modal matching problem between the AoA point cloud and the segmentation masks.

B. Matching through Pixel Proximity

One straightforward approach is matching through the pixel proximity in the camera screen. Recall that in Section IV we have transformed the AoA point cloud to the camera coordinate so that it is fused with the masks. A naive approach is to simply match the AoA point cloud to the closest masks in the camera screen.

However, pixel proximity fails in distinguishing objects in the foreground from those in the background. In some cases, the point cloud may be scattered on two objects that are close to the camera screen but far away in the actual scene. For example, one object could be a bottle close to the user while the other is a clock hanging on the remote wall. The matching approach through pixel proximity may identify the wrong object. It is because the camera screen itself is a 2D plane and does not contain any depth information, it is impossible to distinguish objects of different depths from a static frame alone.

C. Matching through Homography

To distinguish objects at different depths of field, we employ the idea of homography from computer vision. Recall that the camera pose changes in the azimuth and elevation estimation algorithm. The camera viewing angle also changes accordingly during the process. In the field of computer vision, and two camera screens of the same planar surface are related by homography (assuming a pinhole camera model). From the homography matrix, we derive the camera's rotation and further figure out the depth of the target object.

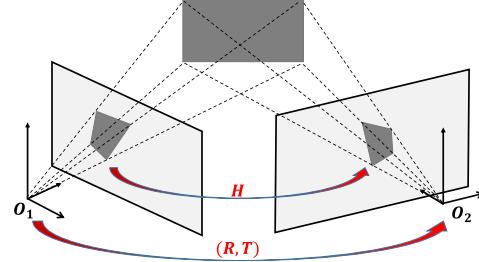


Fig. 6: Homography transformation

Homography matrix: As illustrated in Figure 6, a homography matrix describes the relation between the position projection of the feature point on the two frames of the camera on the same plane [23]:

$$\begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = H \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix}, \quad (7)$$

where $(u_1, v_1, 1)^T$ represents the image point in image 1, $(u_2, v_2, 1)^T$ is the image point in image 2. Image 2 is transformed into image 1 through the Homography matrix H. During the azimuth and elevation estimation algorithm, we change the pose of the camera. Such a movement generates a unique homography matrix in the camera screen for each mask at a certain depth in the real scene, assuming each Mask approximates a plane. In other words, by calculating

the homography matrix of each mask in the camera screen, we figure out the depth of the object.

Homography matrix and depth: The formula between the homography matrix and depth is as follows [23]:

$$H = K(R + T \frac{1}{d} N^T)K^{-1}, \quad (8)$$

where K is the camera internal parameter, R is the camera rotation matrix, T is the translation matrix, d is the depth, and N is the normal vector of the plane in the frame.

From Equation 8 and Figure 6, we observe that the change of camera pose (rotation and translation) produces two frames of images. The Homography matrix of the plane is calculated by the matching feature points in the two frames. Since each mask contains thousands of pixels, we apply edge detection approaches to extract feature points from the masks [21], [22]. From Equation 8, we find that the homography matrix H is inversely proportional to the depths d . Based on it, we leverage the homography matrix as a characteristic feature to distinguish masks of different depths to enhance the matching algorithm.

A more robust matching mechanism: Note that we have projected the AoA point cloud to the camera screen. The AoA point cloud is supposed to be located on the same plane as the target device as the signal is sent from a target device. According to the principle of homography, the homography matrix of the AoA point cloud and the target device should always be similar no matter how the camera pose changes. Based on that, by calculating the similarity between each Mask's homography matrix and the AoA point cloud's Homography matrix, we distinguish Masks at the same depth as the AoA point cloud.

To put everything together, our matching algorithm works as follows. We design the instance segmentation method: Mask R-CNN processes the original image and generates several Masks. Each mask represents an object plane. We utilize the AoA point cloud of the wireless positioning result as the observations, and filter the Masks from two aspects: i) From the camera screen distance, select the mask with the closest distance to the AoA point cloud. ii) Using the camera pose changes as a guide, filter masks of varying depths through the Homography matrices, and select the mask which is most similar to the AoA point cloud Homography matrix. We take the intersection of the two and finally determine the mask corresponding to the target device.

VI. IMPLEMENTATION

Platform setting. We build VisBLE, which locates the position of the Bluetooth device signal source in the video and provide interactive Bluetooth device objects. Figure 7 illustrates our implementation setting platform of VisBLE. It consists of three modules: i) Bluetooth signal receiving module: CC26X2 development board and attached BOOSTXL-AOA antenna array as the Master terminal. ii) We utilize Nexus5x as the main AR device, which is equipped with a Level meter and a rear camera. The Level meter reads the angle between the phone and the horizontal plane, and the camera captures

images; iii) The entire device is placed on a tripod and rotated around the Y-axis. We place the mobile phone and the Master module together and align the center of the antenna with the center of the camera. on the X-axis. Note that it is difficult to align the two coordinates strictly from the hardware, we first perform coordinate calibration before proceeding with the experiment. Make sure that the antenna coordinate and the camera coordinate are in the same coordinate.

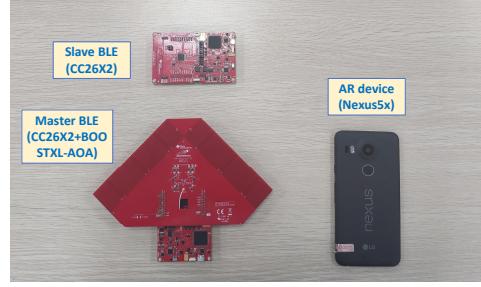


Fig. 7: Experiment Setting for VisBLE.

VII. PERFORMANCE AND EVALUATION

We conducted our experiments in a conference room, surrounded by tables, chairs, and other wireless devices. As mentioned in the previous implementation, the CC26X2 development board with BOOSTXL-AOA together with the Nexus5x device as the receiving device, and the separate CC26X2 development board as the target Bluetooth device. 2.4GHz channel is used for Bluetooth communication. By default, connectionless Bluetooth devices broadcast and send data packets with 100ms intervals. The resolution of the image is 4032×3024 pixels.

For the performance comparison, we also implemented VisIoT [5], which uses the phase difference of the received signals from the two antennas, combined with the motion of the receiver device tracked by IMU sensors to estimate the azimuth and elevation angles. From Equation 1 in Section IV-A, the phase difference and the AoA are transformed. We realize the visualization method of VisIoT by observing the AoA threshold. Note that VisIoT is a solution that completely relies on wireless signals and IMU sensor data. By itself, it locates the position of the target device in the camera coordinate, but it is difficult to provide interactive and reliable semantic objects. In the following, we will compare the performance of angular accuracy and localization accuracy as well as the BLE identification accuracy with the vision technology enhancement.

A. Angular and Position Accuracy

Angular accuracy: We first evaluate the angular accuracy of VisBLE, say the azimuth and elevation angles. For the performance evaluation, we conducted experiments while varying the position of the target BLE device. The distance between the target device and the receiving device is between 1m and 6m. Figure 8(a) and 8(b) show the CDF of the azimuth and elevation errors with 300 experiments. The median errors of the azimuth and elevation angles of VisBLE are 2.2° and

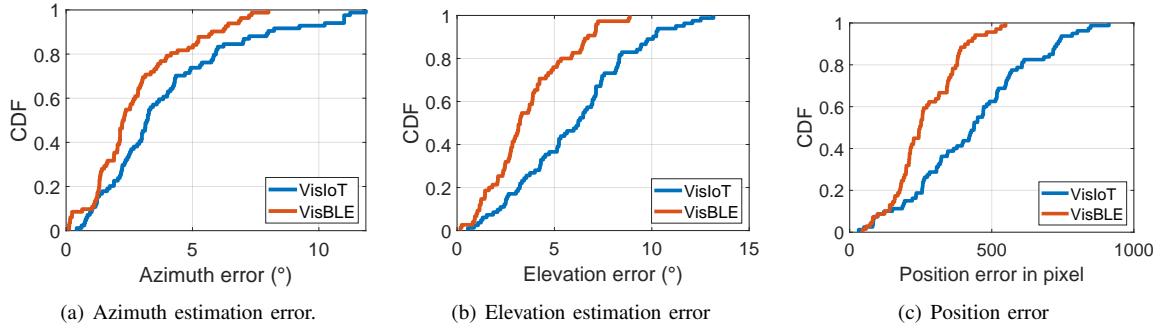


Fig. 8: CDFs for estimation angles and position error comparison between VisBLE and VisIoT

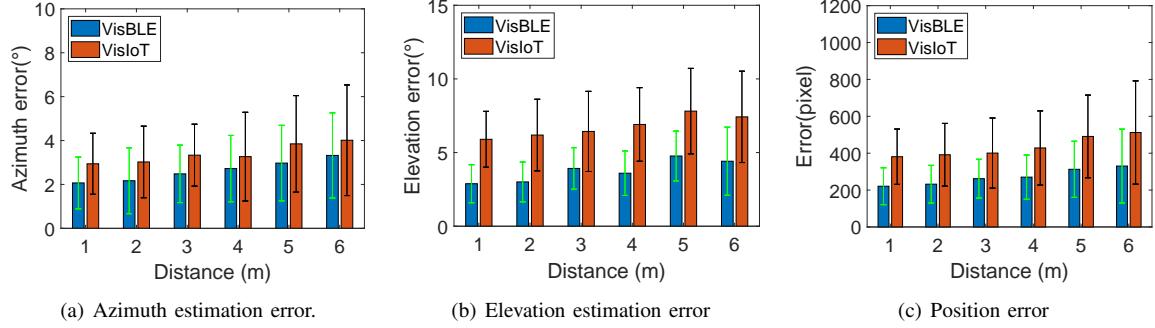


Fig. 9: Evaluation in various distances

3.1° , respectively. On the other hand, VisIoT yields much higher azimuth and elevation angle estimation errors, with median errors of 3.4° and 6.4° , respectively. Compared with VisIoT, our azimuth and elevation estimation algorithms have increased by 33% and 46.8%, respectively. It is observed that VisBLE and VisIoT are comparable in azimuth estimation, and have a huge difference in elevation estimation. The reason is that VisBLE uses the state quantities which brings much less cumulative errors during the rotation.

Position accuracy: We then evaluate the errors between the estimated target and the actual target in the camera coordinate. Figure 8(c) shows the corresponding CDF of the positioning pixel error. The median and the 95-percentile pixel errors of VisBLE of 246 and 481 pixels, respectively. When normalized by the number of pixels in the diagonal line (i.e., 5040 pixels in 4032×3024 pixels image), they correspond to 4.8% and 9.5% of errors, respectively. On the other hand, VisIoT yields higher position errors, with a median error and 95-percentile errors of 437 and 783 pixels, respectively. Normalized according to the number of pixels on the diagonal, the positioning errors of VisIoT are 8.6% and 15.5% errors, respectively. Note that when the receiving and sending devices are one meter apart, the median error of 243 pixels according to VisBLE is equivalent to the actual positioning median accuracy of 8.3cm . VisBLE achieves not only accurate tracking in the camera coordinate but also centimeter-level localization accuracy.

Impact of distance: Since Bluetooth is suitable for short-distance recognition and transmission, the scenario in this paper focuses on the indoor environment. In this section, we discuss the impact of distance on VisBLE positioning in an indoor environment. Figure 9(a) and Figure 9(b) show the me-

dian angles estimation errors within 95% confidence interval in various distances. And, Figure 9(c) shows the position error. We observe that VisIoT has slightly lower azimuth estimation accuracy than VisBLE, and VisBLE is much better than VisIoT in elevation estimation. The conclusion is consistent with the previous results. It once again proves that the calculation of the two-state angles by Level meter is more reliable than the calculation of rotation by Gyroscope integration.

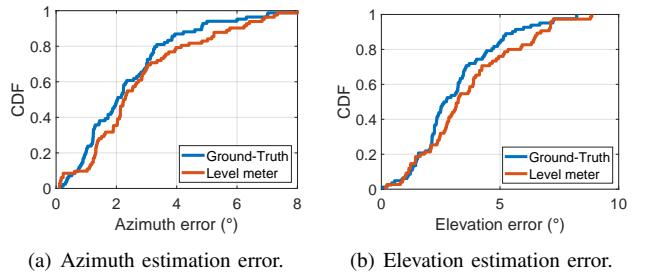


Fig. 10: Angles estimation error with ground-truth rotation.

Impact of Level meter: In our azimuth and elevation estimation algorithm, the level tracks the motion of the device and records the angle of rotation. Here, we analyze how much estimation error is caused by the imperfect Level meter. We replace the rotation angle obtained based on the Level meter with the ground-truth rotation angle. Figure 10 shows the performance comparison. The results show that the Level meter is reliable, where VisBLE does not seriously suffer from the drift problem. When the ground-truth rotation angle is used, the median azimuth estimation error reduces from 2.2° to 2.0° , and the median elevation angle estimation error reduces from 3.2° to 2.6° . Regardless of the CDF estimation of azimuth

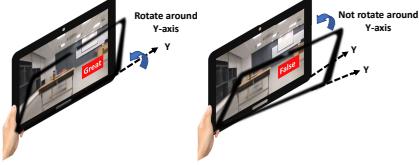


Fig. 11: Rely on prompt instead of tripods. Fig. 12: The performance of prompt on localization.

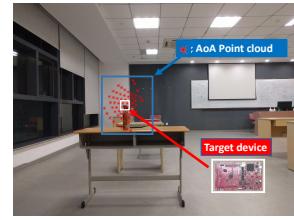
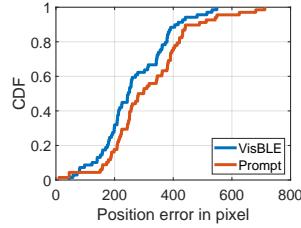
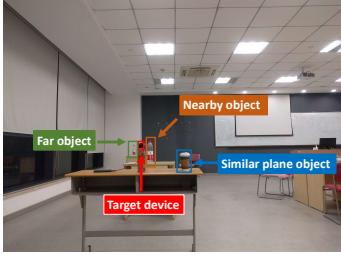
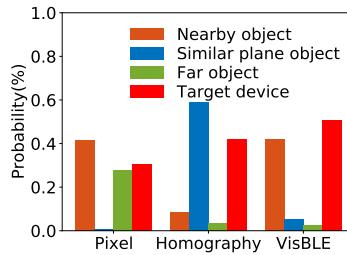


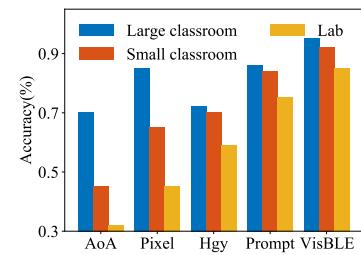
Fig. 13: Enhancement through the vision technology.



(a) Example of the experiment scene.



(b) Probability of the target device.



(c) Accuracy of 5 methods in different scenarios (Hgy is homography's abbreviation).

Fig. 14: The performance of matching algorithms.

or elevation, the overall trend of the curve is very similar to Ground truth. Note that VisBLE only needs to count the angles of the two-state quantities before and after the rotation to estimate the angle of the device rotation. Reduced rotation time significantly reduces the accumulation of errors due to drift. The error is more caused by the equipment itself.

Impact of tripod: Figure 11 shows that if the user rotates strictly according to the Y-axis, the system gives the prompt as "Great", otherwise "False". Figure 12 shows the CDFs of the position error when the rotation mode is changed. As shown in the figure, the median positioning error increased from 246 pixels to 280 pixels when we rely on the prompt provided by the Level meter instead of tripods. Our results demonstrate that we rely on the Level meter to constrain user operations for high-precision positioning even if mobile devices are not placed on tripods.

The performance of matching algorithms: Figure 13 shows the scene of VisBLE. We utilize the AoA point cloud obtained by Bluetooth signal positioning as the observation value. In order to evaluate the matching effect of visual positioning and Bluetooth positioning, in the visual matching method, we design three different experimental schemes: 1) Pixel: utilize pixel distance to filter masks. 2) Homography: leverage the homography matrix to filter the masks with different depths. 3) VisBLE: Combine pixel distance and homography matrix to filter masks. Each scheme calculates the probability that the mask in the scene is the target device. Our experimental scene is shown in Figure 14(a), where there are interfering objects around the target device.

Figure 14(b) shows the predicted probability of masks in the camera screen for the above three schemes. We can see that the Nearby object and Far object surround the target device in the pixel image, but is on a different plane from the target device.

A similar object is far away from the target device in pixels but on the same plane. VisBLE combines the two methods, we provide users with accurate interactive objects.

Performance in different scenarios: To confirm the reliability of VisBLE in more general, we deploy VisBLE, the above two methods, and the no-vision-based AoA point cloud method in different scenarios, including a large classroom, a small classroom, and a lab. In a large classroom, there is a large distance between objects, while in a laboratory environment, the interference object is closer to the target device. As shown in Figure 14(c), AoA point cloud alone obtains higher recognition accuracy in large classrooms, but in small classrooms and laboratories, it is significantly reduced. After enhancing by visual method, regional segmentation enhances the accuracy of AoA point cloud positioning objects, and the homography roughly distinguishes objects in front and back.

The reasons of identify failure: There are two main reasons for VisBLE identification failure: First, due to the limitations of indoor positioning, we obtained the AoA with the median error of 3° , the resulting point cloud area with the median error of 246 pixels. Second, as described in section V-C, the homography makes a rough distinction between front and rear objects when the position changes in a small range ($\leq 1m$), and cannot make a clear distinction between objects with close depth.

In summary, our extensive experiment results have shown the effectiveness of the Level meter and the vision technology in providing an accurate and robust BLE identification. It paves the way for a lot of AR applications in the future.

VIII. RELATED WORK

Vision-based localization: With the rapid development of computer vision technology, it might not be difficult to identify and locate known IoT devices in the video. The

concepts of AR visualization with wireless devices have been proposed. However, there is no feasible solution to achieve it at the existing visual inspection. For example, researchers deliberately construct marks in visual images, such as 2D barcodes [24], retroreflective or luminous points [25] or other patterns [26]. To this extent, previous works primarily focused on local interactions with a device with prior knowledge of where it is located in the environment, which limits the scalability. In addition, IoT devices do not have specific appearance conditions. Therefore, it is challenging to identify and distinguish wireless devices with the same appearance based on visual inspection methods.

Wireless Signal-based localization:

In the latest wireless signal localization research [4], [6], researchers have achieved centimeter-level 3D localization accuracy. They both track the device in 3D space by estimating the distance from 3 or more anchor points, and finding the intersections of the circles whose radii are the estimated distances. Although they provide a reliable location estimation method, it cannot meet the requirement of visualizing the target device to the AR device because it cannot estimate the direction of arrival of the received signal. In order to estimate the direction of arrival of the received signal, one possible approach of azimuth and elevation estimation would be emulating circular antenna array using the rotation of the device(i.e., Synthetic Aperture Radar) [27]. Swarun Kumar et al. [28] propose Ubicarse, emulates Synthetic Aperture Radar (SAR) design a virtual circular antenna array to estimate the direction of arrival of the received signal.

IX. CONCLUSION

This work presents VisBLE, which provides users with a novel way to interact with Bluetooth devices. VisBLE is the first work to achieve visual identification with Bluetooth devices under the Bluetooth 5.1 protocol. In contrast with previous wireless signal localization visualization works, VisBLE proposes a novel azimuth and elevation angle estimation mechanism to simplify the user's operation and combines it with computer vision technology to improve the reliability of the system. Our results are divided into the following two aspects: i) Position accuracy: VisBLE positions a BLE device with the median position error of 4.8%. ii) Accuracy of identification: VisBLE identifies a BLE device with a success rate of over 90%. Looking to the future, we intend to develop a new visualization framework for the wireless source that requires no user movement.

REFERENCES

- [1] H. C. Chao, "Internet of things and cloud computing for future internet," 2011.
- [2] S. Seneviratne, F. Jiang, M. Cunche, and A. Seneviratne, "Ssids in the wild: Extracting semantic information from wifi ssids," in *2015 IEEE 40th Conference on Local Computer Networks (LCN 2015)*, 2015.
- [3] G. White, C. Cabrera, and A. Palade, "Augmented reality in iot," in *The 8th International Workshop on Context-Aware and IoT Services (CIoTS 2018)*, 2018.
- [4] R. Nandakumar, V. Iyer, and S. Gollakota, "3d localization for sub-centimeter sized devices," *ACM*, 2018.
- [5] Y. Park, S. Yun, and K. H. Kim, "When iot met augmented reality: Visualizing the source of the wireless signal in ar view," in *the 17th Annual International Conference*, 2019.
- [6] W. Mao, H. Jian, H. Zheng, Z. Zhang, and L. Qiu, "High-precision acoustic motion tracking: demo," in *International Conference on Mobile Computing & Networking*, 2016.
- [7] M. Woolley, "Bluetooth direction finding," *A technical Overview*, 2019.
- [8] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128 837–128 868, 2019.
- [9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *IEEE*, 2017.
- [10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [11] H. Xue, A. Zhang, L. Su, W. Jiang, and W. Xu, "Deepfusion: A deep learning framework for the fusion of heterogeneous sensory data," in *the Twentieth ACM International Symposium*, 2019.
- [12] B. Korany, C. R. Karanam, H. Cai, and Y. Mostofi, "Xmodal-id: Using wifi for through-wall person identification from candidate video footage," in *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom 19)*, 2019.
- [13] H. Ke, Y. Cao, H. Y. Sang, Z. Xu, and K. Ramani, "Scenariot: Spatially mapping smart things within augmented reality scenes," in *the 2018 CHI Conference*, 2018.
- [14] T. V. Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, "Recovering accurate 3d human pose in the wild using imus and a moving camera: 15th european conference, munich, germany, september 8-14, 2018, proceedings, part x," *Springer, Cham*, 2018.
- [15] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," *AcM Sigcomm Computer Communication Review*, vol. 45, no. 4, pp. 269–282, 2015.
- [16] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Proceedings of the 10th USENIX conference on Networked Systems Design and Implementation*, 2013.
- [17] Mathews, P. C., Zoltowski, and M. D., "Eigenstructure techniques for 2-d angle estimation with uniform circular arrays," *Signal Processing, IEEE Transactions on*, vol. 42, no. 9, pp. 2395–2407, 1994.
- [18] Y. Wu and H. C. So, "Simple and accurate two-dimensional angle estimation for a single source with uniform circular array," *IEEE Antennas & Wireless Propagation Letters*, vol. 7, pp. 78–80, 2008.
- [19] N. Xi and L. Li, "A computationally efficient subspace algorithm for 2-d doa estimation with l-shaped array," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 971–974, 2014.
- [20] T. Xia, Z. Yi, Q. Wan, and X. Wang, "Decoupled estimation of 2-d angles of arrival using two parallel uniform linear arrays," *IEEE Transactions on Antennas & Propagation*, vol. 55, no. 9, pp. 2627–2632, 2007.
- [21] L. Gan, J. F. Gu, and P. Wei, "Estimation of 2-d doa for noncircular sources using simultaneous svd technique," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, pp. 385–388, 2008.
- [22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [23] A. Agarwal, C. Jawahar, and P. Narayanan, "A survey of planar homography estimation techniques," *Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12*, 2005.
- [24] R. R. A, M. S. A. B, and M. C. A. B, "Speeded up detection of squared fiducial markers," *Image and Vision Computing*, vol. 76, pp. 38–47, 2018.
- [25] K. Dorfmuller, "Robust tracking for augmented reality using retroreflective markers," *Computers & Graphics*, vol. 23, no. 6, pp. 795–800, 1999.
- [26] V. Heun, S. Kasahara, and P. Maes, "Smarter objects: Using ar technology to program physical objects and their interactions," in *Chi 13 Extended Abstracts on Human Factors in Computing Systems*, 2013.
- [27] I. Pefkianakis and K. H. Kim, "Accurate 3d localization for 60 ghz networks," in *the 16th ACM Conference*, 2018.
- [28] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 483–494.