Poster Abstract: Fusing Computer Vision and Wireless Signal for Accurate Sensor Localization in AR View

Md Fazlay Rabbi Masum Billah, Md Mofijul Islam, Nurani Saoda, Tariq Iqbal, Bradford Campbell
University of Virginia, Virginia, USA

Abstract

Recent years have seen increasing traction to enable new applications that can localize sensors on the screen of an Augmented Reality (AR) device (e.g. smartphone, tablet) so that sensors can be controlled more intuitively. Despite recent advances in this area, both wireless signal dependent and computer vision based localization solutions have seen a slow acceptance due to signal noise, multipath effect, and limited AR device-sensor interactivity. In this paper, we propose a novel solution to combine the complementary advantages of wireless signal based localization solution with the computer vision based solution to track IoT devices and sensors more accurately. Experimental result shows that our system can accurately track IoT devices with an average pixel error of 34 pixels in a 1024 × 768 pixels image, which is a 75.8% improvement from the state-of-the-art model.

CCS Concepts

• **Human-centered computing** → **Interaction design**.

Keywords

Sensor Localization, Augmented Reality, BLE, Computer Vision

ACM Reference Format:
Md Fazlay Rabbi Masum Billah, Md Mofijul Islam, Nurani Saoda, Tariq Iqbal, Bradford Campbell. 2022. Poster Abstract: Fusing Computer Vision and Wireless Signal for Accurate Sensor Localization in AR View. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3560905.3568095

1 Introduction

The massive growth of sensors and IoT devices promises a future where smart devices will work together to enhance the quality of human life. However, a growing concern with this continued surge is that an unscalable amount of user attention will be needed to interact with a large number of devices and current approach to rely on device-specific applications as the primary interface will fail to ensure a better user experience. Lately, new applications that integrate augmented reality, ubiquitous sensing, and the Internet of Things (IoT) have received considerable attention due to its promise in addressing these issues. For instance, using an AR device, a manager in a retail store can replace advertisements of Bluetooth Low Energy (BLE) beacons directly by holding the AR device over a beacon and dragging and dropping the new advertisements to it (Figure 1a). Whereas in homes, users can use the AR device to visualize an IoT device and tap on it to initiate control without looking for device-specific applications. (Figure 1b).

(a) Retail (b) Smart home

Figure 1: Visualizing and interacting with IoT devices in AR.

Projecting a sensor over the image plane can be done by processing the radio frequency (RF) signals coming from sensors or IoT devices to the AR device [1]. While the RF-dependent localization approaches have achieved a centimeter-level accuracy in controlled environments, it still tends to be error-prone in real-life settings due to the existence of the multipath effect, signal noise, and polarization effect. While vision based technique is more accurate in localizing sensors in the image plane, it alone cannot provide the means to tap on a sensor on the AR view and initiate control. In this paper, we argue that by fusing the RF-based localization features with visual features we can overcome these approaches drawbacks and accurately localize sensors on the 2D plane of an AR device's screen, while conserving the AR device-sensor interaction capability. We develop SpotBLE, a system that uses BLE RF signal parameters to detect the angle from which a sensor is transmitting and determines the sensors 2D coordinate on the AR screen. SpotBLE fuses this coordinate with the image feature and passes to a Faster-RCNN object detection model to precisely localize even adjacent sensors. While fusing, SpotBLE ensures that the metadata of the detected sensor (e.g. MAC address) is preserved so that the user can tap on the sensor and establish a connection.

2 System Design of SpotBLE

Figure 2 illustrates the high-level overview of the system, which consists of an AR device, transmitting BLE devices, a Wireless

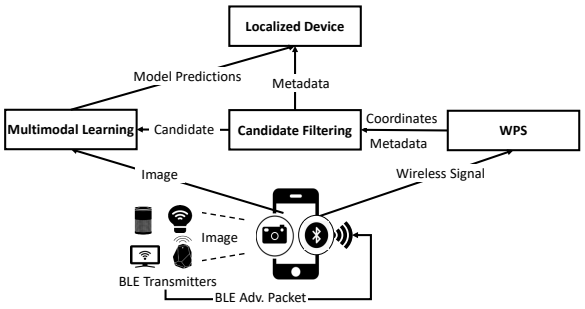


Figure 2: SpotBLE system overview.

SenSys '22, November 6–9, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9886-2/22/11…$15.00
https://doi.org/10.1145/3560905.3568095

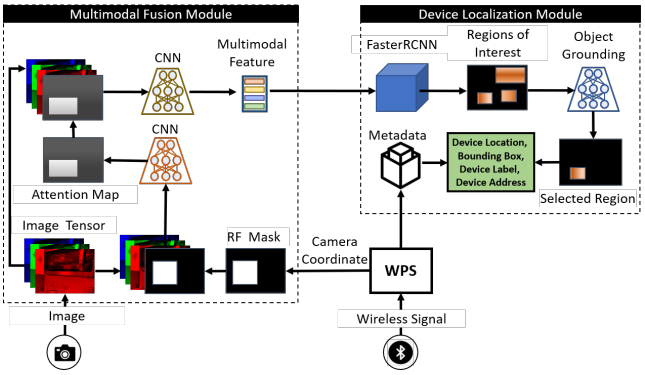


Figure 3: Multimodal learning model

Positioning Module (WPS), a filtering module, and a multimodal learning module. These devices and modules act together to predict the transmitting BLE device's location, create a bounding box around it, predict its label, and store its physical address.

In the first step, an AR device performs a BLE scan to discover all the transmitters sending BLE advertisement packets. The WPS module analyzes these advertisement packets and calculates the 2D coordinate of these devices over the AR device's screen. After that, it forwards these device coordinates and their associated metadata (e.g. device ID, MAC address) to a candidate filtering module. The filtering module discards those coordinates which are out of the bound of the image plane and sends the valid coordinates to the multimodal learning module. The learning module takes coordinates of each candidate device sequentially, fuses them with a captured image, and predicts a more precise location, device label, and a bounding box for the located device. Once located, the filtering module forwards the device metadata which can be used to establish a connection or engage with the device if required.

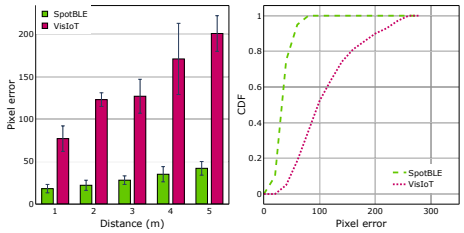
3 Multimodal Learning Model

The multimodal learning model fuses RF signals from BLE devices and camera images to localize the BLE devices more accurately. This model creates an RF mask using the BLE device position, which is calculated from the wireless signal coming from that BLE device. This RF and the captured image from the mobile device are passed through a CNN to produce a saliency attention map, which guides the multimodal model to fuse positional information from RF signal and image features. The fused multimodal features are passed through a Faster-RCNN model to produce a set of candidate locations of the BLE device. Finally, these candidate locations are passed through a device grounding model to determine the location of the BLE device on the camera scene.

4 Evaluation

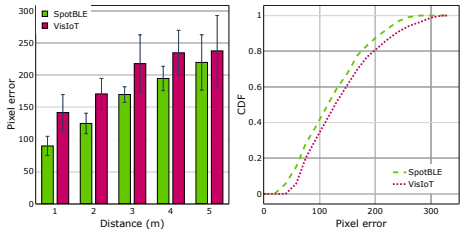
We evaluated SpotBLE in a 5m × 10m office area equipped with regular furniture. We placed up to five IoT transmitters in marked locations and walked with the receiver AR module toward the transmitters from five different directions.

We compared SpotBLE with the benchmark model VisIoT [1]. Figure 4 (a) illustrates the finding of our 500 experiments where the AR device is in the line-of-sight (LOS) of BLE transmitters. It indicates, SpotBLE outperforms VisIoT regardless of the distance



(a) Error with distance **(b) Error distribution**

Figure 4: Pixel error in LOS.



(a) Error with distance **(b) Error distribution**

Figure 5: Pixel error in NLOS.

between transmitters and receivers in terms of pixel error. Where pixel error is the diagonal distance between the center pixel of the predicted box and the center pixel of the ground truth box. Results show that on average SpotBLE achieves projecting device on an image plane with 34 pixel error, whereas VisIoT yields a much higher 134 pixel error. This means SpotBLE outperforms VisIoT with 75.8% improvement. Figure 4(b) represents the CDF of the pixel error. It indicates, SpotBLE achieves 95 percentile with 62 pixel errors, whereas VisIoT achieves 95 percentile with 232 pixels.

To understand if the RF feature is assisting the learning model or not, we evaluated SpotBLE in a non-line-of-sight (NLOS) scenario. We covered the transmitting device with a metal sheet and recorded samples from different directions and distances. Figure 5 (a) shows that SpotBLE can locate transmitting devices even if they are in NLOS with 168 average pixel error. On the other hand, VisIoT gets a much higher 201 average pixel error, meaning SpotBLE achieves 16.4% improvement in this scenario. Figure 5(b) illustrates the CDF of the pixel error in NLOS case. It shows, SpotBLE is able to accurately identify devices with 95 percentile with 232 pixel error. On the other hand, VisIot achieves this with 271 pixel errors.

5 Conclusion

Localizing sensors in augmented reality has proven to be a difficult challenge and a system that is accurate and intuitive remains elusive. In this paper, we revisit the challenges associated with RF-based and vision-based localization solutions and propose SpotBLE, which uses a novel fusion technique to overcome the drawbacks. We show that our multimodal fusion approach to merge information from BLE signals with the visual data can enhance the system capability with a 75.8% improvement while preserving the metadata of transmitting sensors.

References

[1] Yongtae Park, Sangki Yun, and Kyu-Han Kim. 2019. When IoT met augmented reality: Visualizing the source of the wireless signal in AR view. In *MobiSys 2019.*