

Enhancing Smartphone-Based Multi-Modal Indoor Localization with Camera and WiFi Signal

Jing Xu[†], Yanchao Zhao[†], Jie Wu[‡] and Hongyan Qian[†]

[†]College of Computer Sciences and Technology, Nanjing University of Aeronautics and Astronautics, China

[‡]Center for Networked Computing, Department of Computer and Information Sciences, Temple University, USA

Abstract—One of the major challenges in indoor localization is the matching difficulty and prediction accuracy of anchor points. In this work, we innovate in proposing a camera-based, sensor- and WiFi-assisted, and easy-to-deploy system for localization. The proposed method is based on multi-modal sensing to enhancing localization measurement. We implement a prototype with smartphones and commercial WiFi devices and evaluate it in distinct indoor environments. Experimental results show that the 85-percentile error is within 0.21m for indoor POIs that sheds light on sub-meter level localization.

I. INTRODUCTION

Multi-modal sensing with Computer Vision (CV), sensors embedded in smartphones and wireless signal has been a fundamental solution to a wide range of applications and services such as customer navigation in shopping malls [1], tracking in airports [2] and routing robots in an automated factory [3]. The essence of these applications lies in the measurements of distance and azimuth for indoor localization. However, most state-of-art multimodal-based localization schemes, such as Argus [4] and ClickLoc [5] mainly have flaws in measurement accuracy or complexity.

In this context, we propose a solution for indoor localization, which uses Commercial Off-The-Shelf (COTS) WiFi devices, Internal Measurement Units (IMU) sensors and the monocular camera of smartphones.

II. METHODOLOGY

The methodology for multi-modal sensing indoor localization forecast used in this work consists of four steps:

1) Multi-Modal Data Collection: Our system instructs users to take photos of POI from two different angles. Both inertial sensors and CSI devices would be activated for collecting the associated signal in the process of moving.

2) Target Detection Scheme: Accelerometers and gyroscopes data are utilized to estimate the relative distance of anchor points. A multi-target detection framework based on YOLO V2 is adopted to identify objects in two photos.

3) Signal Processing: The acquired sensor data are processed to obtain the rough estimation of user's moving distance. Supplementary calculation is based on CSI extraction for motion features, including direction and distance.

This work was supported by the Natural Science Foundation of China under Grant 61602238, Natural Science Foundation of Jiangsu Province under Grant BK20160805, NSF grants CNS 1629746, CNS 1564128, CNS 1449860, CNS 1461932, CNS 1460971, and CNS 1439672.

4) Motion Information Extraction: The distance and orientation can be measured via cameras with smartphones and CSI of WiFi, referring to the distance and angle data in physical space, the coordinate information in image space, and the mapping relationship between image coordinate system and inertial coordinate system. The architecture of localization is depicted in Figure 1.

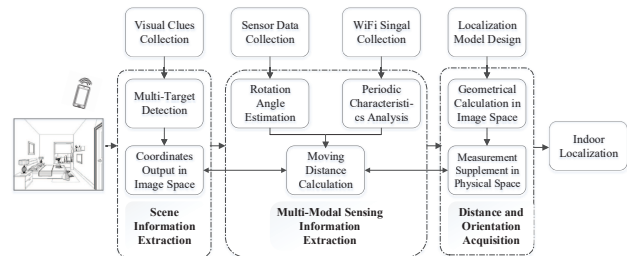


Fig. 1. System Architecture

Both permanent indoor environmental settings such as furniture and walls and temporal dynamics like wireless interference and movement by pedestrians have impact on transmission and reception of radio, which leads to fingerprint ambiguity resulting in errors for localization. In this work, we jointly leverage geometric relationship from ubiquitous images taken by monocular camera to solve the ambiguity problem of pure sensor or WiFi process described in step 3. The distance and angle ratio information can be acquired by the target recognition in the image, the rotation angle of the mobile devices, and the actual moving distance between two positions can be obtained through the inertial sensors and wireless CSI data. Due to the addition of wireless and sensor measurements, the proposed scheme does not require the user to add a new POI to the database and is valid for POIs that have unknown properties and do not have a floor plan.

III. RESULTS

In this work, the prototype consists of a Google Nexus 5X as the monocular camera device and sensor data acquisition device, an Intel NUC D54250WYKH laptop with an Intel 5300 NIC as receiver, and a mini R1C wireless router as transmitter. The implementation is conducted in the 5GHz frequency band with 20MHz band with channels.

We conduct experiments at 10 different POIs at the laboratory and the meeting hall. Fig. 2 summarizes the performance of

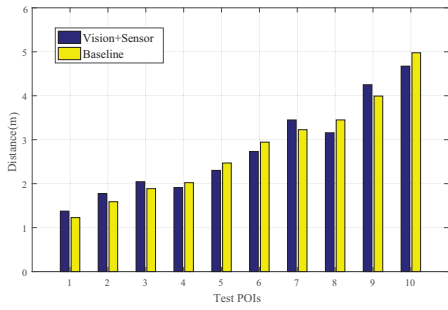


Fig. 2. Overall distance measurement performance

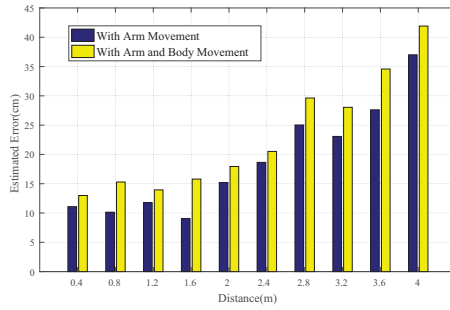


Fig. 3. Impact of activity diversity

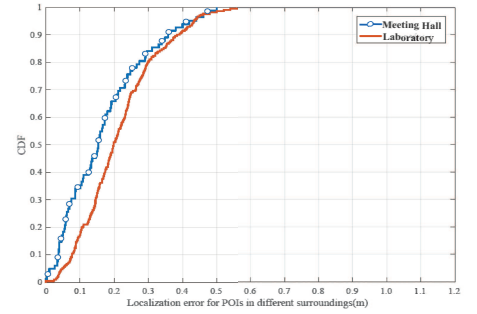


Fig. 4. Impact of moving device diversity

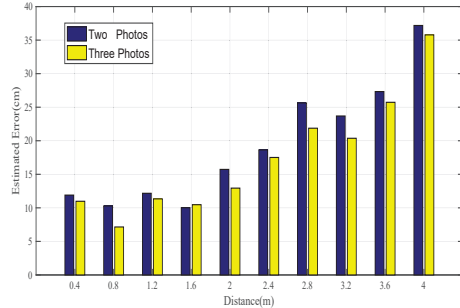


Fig. 5. Performance with different image numbers

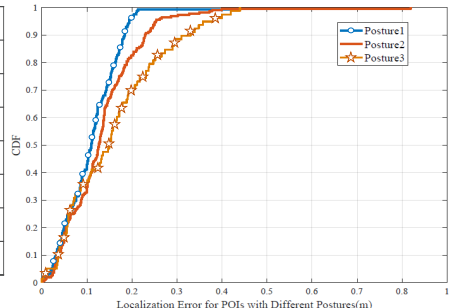


Fig. 6. Performance with different environments

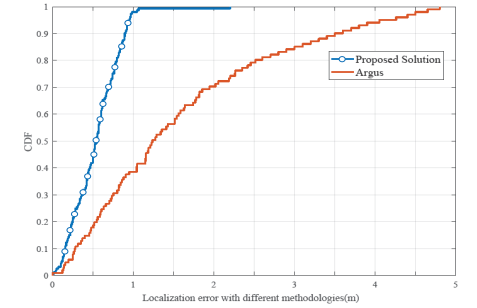


Fig. 7. Proposed solution v.s. Argus

proposed multi-modal strategy among the test locations. We evaluate the performance of our prototype through five sets of experiments. The 50-percentile and 85-percentile estimated errors for scenarios are about 0.13m and 0.21m.

Localization Case Study: Different Movement Methods

We divide the movement into translations, rotations, and actions that can be decomposed into the above two categories. As shown in Fig. 4, the 85-percentile estimated errors of the above postures are 0.18m, 0.2m and 0.23m, respectively. The reason for this result is that the first two movement methods do not need to consider the measurements of rotation angles.

Localization Case Study: Two Photos v.s. Three Photos

Fig. 5 illustrates the localization accuracy of taking two photos and three photos in the same scene. As shown in the figure, collecting three images reduces the estimated error, and the reason for decrease in the error is that the third image outputs an additional distance and direction from the user to target objects in images, correcting the previous localization results.

Localization Case Study: Distinct Environments

We conduct experiments in the laboratory and the meeting hall. The 85-percentile estimated error in the meeting hall is about 0.21m, and that in the laboratory is 0.25m. Fig. 6 shows that the localization system performs well in both environments and the error is smaller compared with that in the complex indoor environment.

Localization Case Study: Comparison with Existing Methodology

Argus is an indoor localization system that estimates user's distance and direction via combining WiFi with visual clues,

which extracts geometric constraints in the image space and takes joint methods to map the constraints to the fingerprint space. For a fair comparison, we evaluate Argus and our proposed solution. We further provide the same condition of sampling and measurement in the same dataset. Fig. 7 demonstrates comparison results over two different approaches.

IV. CONCLUSION

This work proposes a multi-modal approach to enhance indoor localization. The core techniques are rooted in mapping model from image space to physical space, the algorithm of distance, and orientation measurements. We conduct comprehensive theoretical studies and the experimental results show that the 85-percentile error is within 0.21m for indoor POIs within 5m. Our estimation could localize the user with only one POI in two pictures. Furthermore, our method is robust in case of insufficient data, so it can also be applied to indoor localization systems with sparse information.

REFERENCES

- [1] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi:decimeter level localization using wifi," *Acm Sigcomm Computer Communication Review*, pp. 269–282, 2015.
- [2] L. Yang, Y. Chen, X. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proc. of ACM MobiCom*, pp. 237–248, 2014.
- [3] R. K. D. K. Jue Wang, Fadel Adib and D. Rus, "Rf-compass: Robot object manipulation using rfids," in *Proc. of ACM MobiCom*, 2013.
- [4] H. Xu, Z. Yang, Z. Zhou, L. Shangguan, K. Yi, and Y. Liu, "Enhancing wifi-based localization with visual clues," in *Proc. of ACM Ubicomp*, pp. 963–974, 2015.
- [5] H. Xu, Z. Yang, Z. Zhou, L. Shangguan, K. Yi, and Y. Liu, "Indoor localization via multi-modal sensing on smartphones," in *Proc. of ACM Ubicomp*, pp. 208–219, 2016.