EyeLoc: Towards Plug-and-play Indoor Localization on COTS Smartphone

Zhichao Cao*, Zhao Wang*, Jiliang Wang*
*School of Software, Tsinghua University, China
caozc@tsinghua.edu.com

***Abstract*—Indoor localization is becoming an emerging requirement for nowadays mobile applications. Existing indoor localization approaches, however, require exhausted system bootstrap and calibration phases. The huge sunk cost usually hinder practical deployment of indoor localization systems. Although crowdsourcing is widely adopted to reduce the burden, the concern of incentive and privacy usually limits the number of participants. In contrast, we observe that floor-plan images, which highlight the location of many landmarks, are widely available in many indoor environments. According to some observed landmarks, people can freely localize themselves. However, due to the requirements of direction-sense and space-imagination, not all people get used to this way, especially in complex or unfamiliar buildings.**

In this paper, we propose EyeLoc that uses smartphone to imitate the human vision based self-localization behavior. EyeLoc takes a floor-plan image as the input and adopts state-of-the-art text recognition algorithm to extract a set of available landmarks. Next, the key challenge is how to obtain the accurate geometric information of surrounding landmarks to infer the user's location with simple smartphone operations. When people hold the smartphone to shoot an image, camera facing direction can be measured by compass. The key observation is the landmark along the camera facing direction will appear in the middle of the camera image. The same text recognition algorithm is used to detect and localize a landmark in a camera image. After fixing a landmark, EyeLoc further calculates pixel distance between user location and the landmark with two or more camera images. Then, EyeLoc combines camera facing direction and landmark pixel distance to establish a geometric structure. Instead of manually searching landmarks and shooting images, EyeLoc requires a user to shoot a video of surrounding environments. Then, EyeLoc automatically detects several landmarks and extracts geometric structures of the detected landmarks. Furthermore, once three or more landmarks are observed, EyeLoc heuristically calculates the user's location to accurately match the extracted geometric structures on the floor-plan image. Finally, EyeLoc establishes a geometric matching function between floor-plan image coordinate system and landmark geometric structure. With the matching function, EyeLoc can accurately localize a user even only one landmark is observed. We evaluate EyeLoc in different scenarios with Android mobile phones. If only one landmark is observed, the 90-percentile localization precision can achieve ?? cm.

I. Introduction

Nowadays, the physical layout of many indoor spaces (e.g., shopping malls, airports, and hospitals) are becoming more and more complex. Indoor localization is therefore becoming an important service for people in those indoor spaces. Though outdoor localization (i.e., GPS) has been put in practice for many years, there is still no practical deployed indoor localization systems.

Most indoor localization systems rely on both pre-deployed infrastructure and pre-collected information (e.g., WiFi [13] [22] [27] [15] [19], lamp [9] [29] [12], images [4], magnetic field [20]) to construct a localizable map. The requirement of pre-effort, which incurs extensive bootstrap overhead that is difficult to realize in practice, hinders existing approaches from being widely used. As a result, there is no practical indoor localization system widely in use now. Crowdsourcing is a promising approach for collecting huge amount information. The incentive and privacy need to be addressed for using it in indoor navigation [31] [21]. Even when pre-collected information can be obtained, the information usually needs to be timely updated and calibrated [4] [10] which limits the applicability.

Therefore, the question is *Can a user setup a plug-and-play indoor localization system by himself*? We notice a possible way by leveraging the widely available floor-plan images, which can be obtained in various ways (e.g., physical guidance sign, publicity website). Those floor-plan images contain the location of many landmarks. These landmarks are often used as visual hints to help users manually localize themselves based on their observation. For example, in a large shopping mall, a customer watches the floor-plan image from a guidance sign. He looks around, searches for some landmarks of the floor-plan image, then tries to figure out his location. Although there is no pre-deployed infrastructure and pre-collected information, self-localization usually requires users have good direction-sense and space-imagination. To mitigate users' mental work, we imagine that can users automatically obtain their location on the floor-plan image from their mobile phone as traditional outdoor localization system such as Google Maps? In this sense, it is possible to achieve a plug-and-play indoor localization system by bridging this gap.

In this work, we propose EyeLoc, a step towards plug-and-play indoor localization without pre-deployed infrastructure and pre-collected information. The key idea of EyeLoc is to imitate the human vision based self-localization behavior with smartphone. After obtaining the floor-plan image, EyeLoc uses state-of-the-art text recognition algorithm [14] to extract a set of available landmarks. The recognized text is used to identify a landmark. Moreover, the location of text bounding box provides geometric information of a landmark in the floor-plan image coordinate system.

A. Challenges, Innovations and Contributions

To further infer a user's location on the floor-plan image, EyeLoc needs to obtain accurate geometric information of surrounding landmarks. However, it faces three challenges.

- First, most smartphones have only monocular camera. Even with binocular camera, the optical properties is usually not known. Thus, EyeLoc cannot obtain landmark geometric information from single camera image.
- Second, considering the limited user patience, the smartphone operation must be kept as simple as possible. EyeLoc also should not require users have any special ability and prior knowledge.
- Third, in real indoor environment, the quantity of observed landmarks cannot be guaranteed. It is possible there is only one landmark can be observed. EyeLoc should support this scenario.

To address the first challenge (Section IV), EyeLoc uses monocular camera and compass to construct the similar function of a well calibrated binocular camera. Specifically, when people hold the smartphone to shoot an image, the camera facing direction can be measured by compass [32]. We notice that if a landmark is on the camera facing direction, the landmark will appear in the middle of the camera images. EyeLoc uses the same text recognition algorithm to detect and localize a landmark in a camera image. Hence, EyeLoc can obtain the direction of a landmark by requiring a user to shoot an image that contains the landmark in the middle. Then, after fixing a landmark, if the user shoots another image that contains the landmark with different camera facing direction, EyeLoc can approximately calculate the pixel distance between the smartphone camera and the landmark. As a result, EyeLoc establishes a geometric structure using both camera facing direction and landmark pixel distance. However, for common users, image shooting is non-trivial to guarantee a landmark exactly appears in the middle of the image. Moreover, users also need to follow the instruction to shoot an auxiliary image to calculate landmark pixel distance. The requests of user's special ability and prior knowledge are not acceptable. This is why the second challenge coms.

For the second challenge (Section **??**), EyeLoc uses video to replace labor-intensive image shooting and automatically detects landmarks and extracts their geometric structures. We know that users only need to press and hold a button to shoot a video with most modern smartphones. In comparison with intended image shooting, it is much easier to rotate the camera around human body and shoot a video of surround environment. We can treat the continuous video frames as a set of camera images that contain surrounding landmarks. EyeLoc further develops an efficient method to obtain the useful video frames and calculate the geometric structures of observed landmarks. Finally, if there are three or more observed landmarks, EyeLoc develops a triangle localization based heuristic algorithm to search user's location on floor-plan images. However, as the third challenge mentioned, in a lot of cases, it is hard to observe sufficient landmarks so that EyeLoc still cannot localize user.

To solve the last challenge (Section V), the basic idea is to initialize a geometric matching function between floor-plan image coordinate system and landmark geometric structure. Namely, if the direction offset and pixel distance scale between floor-plan image coordinate system and landmark geometric structure are known, EyeLoc can localize user by using one observed landmark. EyeLoc provides two alternative ways to initialize the geometric matching function. If there is the opportunity to observe three or more landmarks at a location, EyeLoc can establish the matching function after the user's location on the floor-plan image is determined. Otherwise, EyeLoc collects several landmark geometric structures at different locations. Then, combining with the walking direction [16] between two adjacent locations, EyeLoc can obtain the matching function. Now, users can freely localize themselves by EyeLoc.

We implement EyeLoc in Android. EyeLoc does not have special requirements on the smartphone hardware, and thus can be used for most modern smartphones. We evaluate EyeLoc on three COTS Android smartphones (i.e., Huawei Mate7, XiaoMi MI4 and Nexus 6P). The evaluation results show that a user can set up EyeLoc within ?? minutes in large shopping malls. With three or more observed landmarks, the 90-percentile localization precision can achieve ?? cm. Moreover, if only one landmark is observed, the 90-percentile localization precision can achieve ?? cm.

The contribution of this paper is as follows.

- We propose EyeLoc, a plug-and-play indoor localization system on COTS smartphones without pre-deployed infrastructure and pre-collected fingerprints.
- Our key principle is to imitate human vision based self-localization with smartphone. EyeLoc proposes the novel and effective countermeasures to address several practical challenges.
- We implement EyeLoc in Android and evaluate its performance in various environments. The evaluation results show that EyeLoc is effective on both execution time and localization precision.

The rest of this paper is organized as follows. Section II introduces the related work. Section III shows the overview of EyeLoc. Section IV illustrates . Section **??** gives . Section V illustrates . Section VI and Section VII show the implementation and evaluation details, respectively. Finally, we conclude this work in Section VIII.

II. Related Work

In recent years, many works target on developing efficient indoor localization and positioning systems. However, most of them need either pre-deployed infrastructure or pre-collected information. Some of them even need custom hardwares. According to the different types of indoor localization sources, we divide existing works into four categories.

WiFi Signal A large portion of indoor localization methods are proposed based on WiFi signal. One approach is fingerprinting based. The basic observation is WiFi signal patten can serve as the fingerprint that represents every location. System manager builds a fingerprint database in the target areas to initialize localization service. The user's location is estimated by matching the measured fingerprint to database records. RADAR [1], Horus [28], Place Lab [2], Active Campus [6] and PinLoc [19] use site survey to construct the fingerprint database. LIFS [27] and Zee [15] further utilize crowdsourcing to alleviate the burden of labor-intensive site survey.

The other approach is model-based. The basic principle is the relationship between the geometric structure from WiFi access point to user's location and the physical features of the received WiFi signal can be modeled. If the location of WiFi access point (AP) is pre-known, the user's location can be further inferred. Based on the log-distance path loss (LDPL) model, EZ [3] uses the received signal strength (RSS) to estimate the signal propagation distance and combines several estimations of different APs to find the user's location. SpinLoc [17] and Borealis [30] observe if a user faces an AP, the RSS is usually higher than the user turns his back on the AP. After making a full 360° turn, SpinLoc and Borealis extract the angle-of-arrival (AoA) of several APs to determine the user's location. ArrayTrack [26] uses antenna array and WiFi signal phase to obtain accurate AoA spectrum to calculate a user's location. CUPID [18], SAIL [13], Chronos [23] and Ubicarse [8] further refine the distance and AoA measurement methods to achieve high localization precision or adapt to COST WiFi AP.

Visible Light In a typical visible light positioning (VLP) system, lamps (fluorescent and LED) are served as landmarks. After a light receiver (smartphone camera or photodiode) obtains several lamps' location, the light receiver further measures the geometric structure from the observed lamps to find the user's location. Luxapose [9] takes an image which contains several LEDs as input. The LEDs send their location to smartphone through visible light communication technique. According to the optical geometric translation function of camera biconvex lens, Luxapose constructs several constrains of LEDs' AoA to calculate the user's location. According to inherent and common optical emission features of both LED and fluorescent, iLAMP [33] and Pulsar [29] identify a lamp's location from a pre-configured database by feature matching. iLAMP further combines camera image and inertial sensors to infer user's location. Pulsar utilizes a custom device to measure the lamp's AoA, then determine user's location.

The other approach is to customize the lamp to establish a mapping function between the location of light receiver and the corresponding received light physical features. CELLI [25] develops a custom LED bulb which projects a large number of fine-grained light beams toward the service area. CELLI adopts a modulation method to encode the coordinate of a fine-grained cell into the corresponding light beam. Thus, the light receiver can obtain its location by visible light communication. SmartLight [12] uses LED array and a lens to form the light transmitter. On LED array, different LED lamps use different PWM frequencies. According to the frequencies of the received light, the coordinate of the observed LEDs circle can be inferred on LED array. The location of light receiver can be further calculated by optical geometric translation function of the lens.

Scene Image A scene image contains many landmark details and architectural features of the indoor environment. With a large collection of scene image, SfM (Structure from Motion) [4] [5] and SLAM (Simultaneous Localization and Mapping) [24] can construct a 3D indoor model. Based on the 3D model, we can know the camera pose and shooting location given another image. iMoon [4] and Jigsaw [5] adopt SfM to construct indoor 3D model and enable localization services.

Others There are several other methods to build localization and positioning system. Magicol [20] and FollowMe [21] combine the geomagnetic field and user trajectory as the fingerprint to localize user's location. With acoustic speakers as the landmarks, Swadloon [7] and Guoguo [11] use acoustic signal based geometric model and inertial sensors to localize user's location.

To conclude, comparing with these methods, EyeLoc depends on neither pre-deployed infrastructure (i.e., WiFi APs, custom lamps, acoustic speaker) nor pre-collected information (i.e., WiFi fingerprint, lamp optical emission features, scene images, geomagnetic field). Shenlong Wang, etc. [24] utilize the floor-plan image and a scene image to localize a user in large shopping mall. Based on edge, text and layout features of the scene image, they use Markov random field model to infer the camera pose on the floor-plan image. However, considering the computation complexity, it needs to deploy the service on cloud instead of COTS smartphone. In contrast, EyeLoc allows user to set up localization system by oneself on COTS smartphone in any indoor environment.

III. EyeLoc Overview

A. Design Goals

Overall, EyeLoc has the following design goals.

- *user-initialized*. EyeLoc should not assume any pre-deployed infrastructure, pre-collected information and powerful cloud service. A user can initialize the localization system by himself.
- *plug-and-play*. EyeLoc should be able to establish localization service in various of indoor environments (e.g., shopping mall, office building) with different types of floor-plan images (e.g., color image, CAD image).
- *light-weight and accurate*. EyeLoc should be able to run on COTS smartphones with low operation overhead while providing accurate indoor localization services.

B. System Architecture

The main architecture of MI-Navi is shown in Figure 1. Towards the design goals, MI-Navi mainly consists of the following components. The first module is the map translation module (Section **??**), which translates raw image map to structured navigation map. This module takes an image map as input. Then, geographical ingredients such as roads and POIs

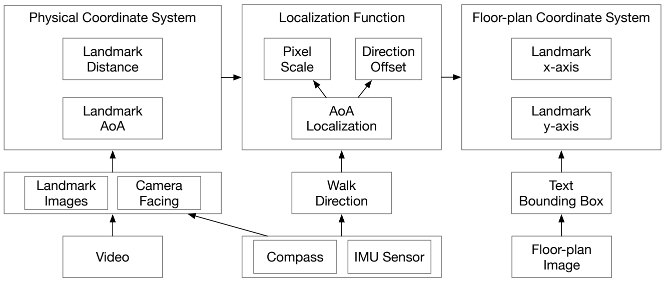

Fig. 1: Illustration of system architecture of MI-Navi.

are extracted from the map by image map parsing methods. Finally, navigation map uses a connected undirected graph to represent the road skeleton and associates the location of POI on the corresponding edges. Hence, giving the start location and destination POI on navigation map, MI-Navi can plan the shortest path by Dijkstra algorithm as the navigation path.

The second module is the localization module (Section **??**), which enables MI-Navi to locate user on navigation map. User uses the mobile phone to take a video of 360° surrounding view. A series of monocular images are then extracted as the input of this module. Several features are, then, extracted from the monocular images by feature extraction methods. The features mainly include the name of surrounding POIs and rotation angle between two adjacent POIs. Taking the information of navigation map as input, with the name of surrounding POIs, MI-Navi can find the possible roads that the user stands. Further, along the chosen roads, MI-Navi searches the MI-location, which has the minimum error between extracted rotation angles and calculated rotation angles on navigation map. Thus, MI-Navi obtains localizable navigation map.

The third module is the real-time navigation module (Section **??**), which embeds user's physical walking on navigation path and gives the motion hints. User's walking is depicted by its distance and direction. The walking distance is measured as the number of steps. The mobile phone is horizontally held in hand of users MI-Navi utilizes multiple sensors (i.e., compass, accelerometer, gyroscope) of mobile phone to count the accumulated steps and measure the instant direction. With step counting, direction measurement and navigation path as input, MI-Navi can trace user's real-time location with an assumption of map scale and direction. However, since the error of direction measurement and the change of stride frequency, the accumulative error of navigation trace matching quickly increases when the map scale and direction is fixed. Hence, during user's walking, navigation engine asks user to fetch several MI-locations in a on-demand manner and automatically detects several turns. Both MI-location and turn serve as anchor points to identify user's real-time location on navigation map. Further, navigation engine continuously adjusts map scale and direction to minimize the error of navigation trace matching. Considering the error of MI-localization and turn detection, MI-Navi utilize particle filter to alleviate the influence of the error.

C. *MI-Navi Example*

IV. Landmark Geometric Structure

V. Geometric Matching Function

VI. Implementation

VII. Evaluation

VIII. Conclusion

In this paper, we propose EyeLoc, a plug-and-play indoor localization system, which does not need the burden of system bootstrap and calibration. The observation is that people can combine floor-plan images and observed landmarks to find their location. Although there is no pre-deployed infrastructure and pre-collected information, not all people like this way because the requirement of direction-sense and space-imagination. EyeLoc enables smartphones can imitate the human vision based self-localization behavior. The inputs of EyeLoc are a floor-plan image and a video of surrounding environments. For a landmark on the floor-plan image and a video frame, to obtain the text content and the coordinate of the text bounding box, EyeLoc adopts state-of-the-art scene text recognition tools. In floor-plan image, the coordinate of the text bounding box can directly be used to describe the geometric relationship of all landmarks, but it does not work in a video frame. We notice that if the text bounding box of a landmark appears in the middle of a video frame, the landmark is on the camera facing direction. Furthermore, combining two or more video frames that contains an identical landmark, the pixel distance between the landmark and smartphone camera can be calculated. EyeLoc develops an efficient method to automatically calculate the geometric structures of observed landmarks. If three or more landmarks are extracted from a video, EyeLoc can heuristically the user's location to accurately match extracted geometric strtuctures on the floor-plan image. To localize a user with two or less observed landmarks, EyeLoc further develops several lightweight methods to establish a geometric matching between floor-plan image coordinate system and landmark geometric structure. We implement EyeLoc as an Android application and evaluate its performance in several environments. The evaluation results show that if three or more vision clues can be observed, the 90-percentile localization precision is ?? cm. When only one vision clue is observed, the 90-percentile localization precision is ?? cm. The performance is comparable with previous WiFi, visible light and image based indoor localization systems.

References

[1] P. Bahl and V. N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *Proceedings of INFOCOM*, 2000.

[2] Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale wi-fi localization. In *Proceedings of MobiSys*, 2005.

[3] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan. Indoor localization without the pain. In *Proceedings of MobiCom*, 2010.

[4] J. Dong, Y. Xiao, M. Noreikis, Z. Ou, and A. Ylä-Jääski. imoon: Using smartphones for image-based indoor navigation. In *Proceedings of Sensys*, 2015.

[5] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li. Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing. In *Proceedings of MobiCom*, 2014.

[6] W. G. Griswold, P. Shanahan, S. W. Brown, R. Boyer, M. Ratto, R. B. Shapiro, and T. M. Truong. Activecampus: experiments in community-oriented ubiquitous computing. *Computer*, 37(10):73–81, 2004.

[7] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, and Y. Liu. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *Proceedings of INFOCOM*, 2014.

[8] S. Kumar, S. Gil, D. Katabi, and D. Rus. Accurate indoor localization with zero start-up cost. In *Proceedings of MobiCom*, 2014.

[9] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta. Luxapose: Indoor positioning with mobile phones and visible light. In *Proceedings of MobiCom*, 2014.

[10] L. Li, G. Shen, C. Zhao, T. Moscibroda, J.-H. Lin, and F. Zhao. Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service. In *Proceedings of MobiCom*, 2014.

[11] K. Liu, X. Liu, and X. Li. Guoguo: Enabling fine-grained indoor localization via smartphone. In *Proceeding of MobiSys*, 2013.

[12] S. Liu and T. He. Smartlight: Light-weight 3d indoor localization using a single led lamp. In *Proceedings of SenSys*, 2017.

[13] A. T. Mariakakis, S. Sen, J. Lee, and K.-H. Kim. Sail: Single access point-based indoor localization. In *Proceedings of MobiSys*, 2014.

[14] L. Neumann and J. Matas. Real-time scene text localization and recognition. In *Proceedings of CVPR*, 2012.

[15] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen. Zee: zero-effort crowdsourcing for indoor localization. In *Proceedings of MobiCom*, 2012.

[16] N. Roy, H. Wang, and R. Roy Choudhury. I am a smartphone and i can tell my user's walking direction. In *Proceedings of MobiSys*, 2014.

[17] S. Sen, R. R. Choudhury, and S. Nelakuditi. Spinloc: Spin once to know your location. In *Proceedings of HotMobile Workshop*, 2012.

[18] S. Sen, J. Lee, K.-H. Kim, and P. Congdon. Avoiding multipath to revive inbuilding wifi localization. In *Proceeding of MobiSys*, 2013.

[19] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. You are facing the mona lisa: spot localization using phy layer information. In *Proceedings of MobiSys*, 2012.

[20] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li, and F. Zhao. Magicol: Indoor localization using pervasive magnetic field and opportunistic wifi sensing. *IEEE Journal on Selected Areas in Communications*, 33(7):1443–1457, 2015.

[21] Y. Shu, K. G. Shin, T. He, and J. Chen. Last-mile navigation using smartphones. In *Proceedings of MobiCom*, 2015.

[22] D. Vasisht, S. Kumar, and D. Katabi. Decimeter-level localization with a single wifi access point. In *Proceedings of NSDI*, 2016.

[23] D. Vasisht, S. Kumar, and D. Katabi. Decimeter-level localization with a single wifi access point. In *Proceedings of NSDI*, 2016.

[24] S. Wang, S. Fidler, and R. Urtasun. Lost shopping! monocular localization in large indoor spaces. In *Proceedings of ICCV*, 2015.

[25] Y.-L. Wei, C.-J. Huang, H.-M. Tsai, and K. C.-J. Lin. Celli: Indoor positioning using polarized sweeping light beams. In *Proceedings of MobiSys*, 2017.

[26] J. Xiong and K. Jamieson. Arraytrack: A fine-grained indoor location system. In *Proceedings of NSDI*, 2013.

[27] Z. Yang, C. Wu, and Y. Liu. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of MobiCom*, 2012.

[28] M. Youssef and A. Agrawala. The horus wlan location determination system. In *Proceedings of MobiSys*, 2005.

[29] C. Zhang and X. Zhang. Pulsar: Towards ubiquitous visible light localization. In *Proceedings of MobiCom*, 2017.

[30] Z. Zhang, X. Zhou, W. Zhang, Y. Zhang, G. Wang, B. Y. Zhao, and H. Zheng. I am the antenna: accurate outdoor ap location using smartphones. In *Proceedings of MobiCom*, 2011.

[31] Y. Zheng, G. Shen, L. Li, C. Zhao, M. Li, and F. Zhao. Travi-navi: Self-deployable indoor navigation system. In *Proceedings of MobiCom*, 2014.

[32] P. Zhou, M. Li, and G. Shen. Use it free: instantly knowing your phone attitude. In *Proceedings of MobiCom*, 2014.

[33] S. Zhu and X. Zhang. Enabling high-precision visible light localization in today's buildings. In *Proceedings of MobiSys*, 2017.