[DC] Outdoor AR Tracking Evaluation and Tracking with Prior Map

Ziwen Lu*

University College London

ABSTRACT

We are very interested in addressing the problem of building city-scale AR systems where users can travel anywhere at any time and see the correct graphics registered in the world around them. One crucial requirement for this is accurate tracking and localisation.

In my work, I propose to tackle two themes. The first is to examine what good registration means in uncontrolled outdoor environments. The second is to explore how prior information can be used to support wide-area tracking efficiently and robustly.

1 INTRODUCTION

The potential now exists for Augmented Reality (AR) to be ubiquitous. Tracking enables AR, and AR tracking has been the research subject for over two decades, making them available on billions of mobile devices. However, state-of-the-art mobile AR systems can only run robustly in a small-scale environment. The lack of good tracking in large-scale outdoor environments is one of the key factors limiting AR usability in the outdoors.

The term *outdoor AR* in this paper mostly refers to AR in an urban outdoor environment, such as navigation or tour guide experience on streets or plazas. And ultimately, we want to expand the system to general outdoor scenarios. Outdoor environments are usually much larger and often unprepared or uncontrolled compared to indoor settings, leading to the higher complexity of the environment containing irregular structures such as vegetation and dynamic objects. Those factors pose challenges in occlusion and variation in appearance for visual-based tracking systems. Therefore, a more robust tracking system is necessary for outdoor AR.

Finding robust tracking systems requires good evaluation criteria, which could help us identify the requirements and rank algorithms. However, there is a lack of good benchmarks for AR tracking systems in the outdoor environment. More specifically, registration quality evaluation is more important than evaluating the trajectory error as it is crucial for AR tracking and affects the user experience. And there are only a few such evaluations for outdoor AR systems.

On the other hand, even though measuring the registration error will not directly mitigate the error, it could potentially provide real-time performance measures. Such measures give designers a better sense of users' situation for interaction, from which the impact on user experience from tracking error can be mitigated [7].

For tracking methods, in the current state-of-the-art AR tracking systems, there is a trade-off. Wide area and source-less systems are widely available (GNSS) but are very noisy. On the other hand, systems such as ARCore and ARKit are only mature for mobile AR in small or localised workspaces. To expand the usability of these local systems, Visual Position Services (VPS) are proposed for assistance in outdoor environments. VPS is commonly achieved through image-based localization or location retrieval, which compares the input image with a map database to localise the camera against the world, providing initialisation and correction for local tracking [10].

*e-mail: ziwen.lu@ucl.ac.uk

However, there are drawbacks associated with VPS. Most VPS systems rely on photometric maps, which can be time-consuming to create and require specialized equipment. Also, they are heavy to load and transfer, requiring high computation power and storage to operate. As a result, they are typically only invested in high-value locations and are often offloaded as a cloud service [10].

To mitigate the cost of the dense prior maps, other systems are proposed to leverage lightweight prior maps of lower fidelity [1, 13] for tracking and localization. Those prior maps could potentially provide larger coverage, given the constraint on mobile devices, than photometric maps, facilitating consistent tracking over a larger area. We wish to develop alternative systems based on those prior maps and achieve similar performance to dense photometric maps.

2 TRACKING EVALUATION FOR REGISTRATION QUALITY

Categorizing and understanding the sources of registration errors is crucial for improving the AR experience and estimating them provides a direct measure of the tracking system's performance. Holloway [3] provided anatomy on registration errors and pointed out that registration error is more sensitive to orientation errors. This finding was echoed by Macintyre et al. [6]. In outdoor AR scenarios, users may be more sensitive to orientation or lateral error due to the large distance between the user and the augmented object, as opposed to position or depth error. Livingston and Ai [5] provided a similar description of types of errors to [3] and agreed on the effect of latency on the error and the existence of noise that leads to jitter.

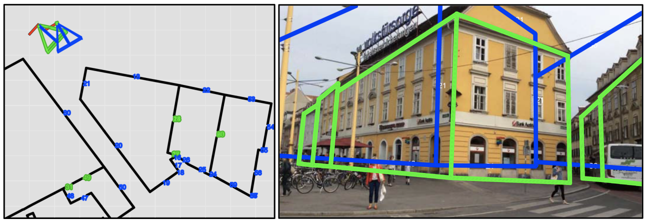
Assessing the accuracy of AR tracking systems should go beyond simply measuring metric accuracy, and also consider the system's ability to associate, disambiguate, and localize in relation to the prior model. AR tracking requirements are highly dependent on the specific application and user perspective, making it challenging to establish comprehensive evaluation criteria. For error estimation, Macintyre et al. [6] proposed a statistical method where convex hull vertices as measurement points for registration errors. Alternatively, tasks can be designed for users' participation to quantify their performance impact under different types of error [5]. Despite these efforts, there is still a need for a more comprehensive evaluation criterion for AR tracking systems.

Recently, Wilmott et al. [12] conducted a study using a psychophysical staircase procedure to measure the permeability of jitter in AR. Their findings indicate a positive correlation between jitter permeability and viewing distance, as well as a negative correlation with background illuminance. These findings could inspire further investigation in outdoor environments where background illuminance and depth vary greatly. For instance, compared to a plain background like a brick wall, it is much harder to spot registration errors about constantly moving clouds in the sky.

For model-based tracking, the relationship between prior map/model accuracy and tracking accuracy needs to be understood for effective AR, with the ultimate goal of establishing the minimum model and tracking accuracy required for a successful AR experience. Julier et al. [4] described the metrics for the required modelling error concerning the tracking error and augmentation error tolerance. They noted that as the viewing distance increases, the permissible error of the model decreases.

Adhering to the principles outlined in the relevant literature, our experiment will focus on the dynamic alignment of distant buildings

(a) Part of Figure 1 from Platinsky et al. [10]



(b) Figure 1(b,(c) from Arth et al. [1]

Figure 1: Comparison of a photometric map-based urban AR system with a line-based localisation system on a 2.5D wire-frame map.

in relation to prior models under varying tracking conditions [8], simulating challenging outdoor AR scenarios. Alternatively, other hardware, such as robotic arms, can be used for trajectory ground truth and repetitive testing. We also wish to investigate the impact of prior model accuracy on its usage in interaction. One simple test scenario could involve asking the user to distinguish between two windows on the same building based on the visual cues registered around the windows. Other factors, such as variations in the background and differences between video/optic see-through displays, can also be included for comparison.

3 Tracking with Prior Map

Although there are versatile SLAM methods leveraging various features and optimization methods, they still have issues such as accumulated drift and interference with dynamic objects in the scene, especially in city-scale environments. Photometric-based VPS are demonstrating good performance for city-scale AR [10]. We aim to develop a similar tracking system with a more compact map representation to reduce the reliance on connectivity and cloud servers.

Leveraging a prior map for tracking is not a novel concept. Model-based tracking [11] is common in early-stage AR and robotic systems and is still the preferred method in certain situations for its simplicity and robustness. Similar outdoor tracking problems have been studied in other fields, and variation in environmental constraints and tasks leads to divergent tracking requirements. For instance, high-definition (HD) maps [9] is used for autonomous driving scenarios, but these methods need adaptation for AR as mobile devices have much less payload and sensing capabilities.

The process of feature extraction in HD map production can simplify dense maps and images while preserving key information about the urban structure for robust and accurate tracking and association. The association can be achieved across different feature modalities, such as photometric features, geometry features, or semantic labels. A geometric-based SLAM system, StructSLAM [14], is one example that leverages line features and regularity in human-made structures. Outdoor model-based tracking [11] and localization [1] have also been demonstrated using common line and plane features from images and 2.5D maps. However, even with the aid of semantics [2,13], the association between observation and map is still challenging.

We aim to improve model-based tracking in AR by exploring the relationship between model accuracy and tracking accuracy. We started with a comprehensive literature review on related topics, including SLAM, localization, 3D reconstruction, and the topological model of outdoor AR coverage. As a starting point, we captured a high-accuracy point cloud of a building, which we will use as a prior map for evaluating registration errors and for tracking.

Following the work of [13], we wish to expand the base system to support line features [1, 14] for the outdoor environment with other prior maps, for which investigation on cross-modality data association through geometry and semantic features will be conducted. The tracking system will be evaluated using a standard dataset or our own captured on campus and eventually integrated with AR displays for user studies to assess registration errors and user experiences.

Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation programme under RISE, Grant Agreement No 739578. The author thanks Prof Simon Julier and Prof Anthony Steed for their supervision and valuable comment.

References

[1] C. Arth, C. Pirchheim, J. Ventura, D. Schmalstieg, and V. Lepetit. Instant Outdoor Localization and SLAM Initialization from 2.5D Maps. *IEEE Transactions on Visualization and Computer Graphics*, 21(11), Nov. 2015.

[2] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas. Probabilistic data association for semantic SLAM. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017.

[3] R. L. Holloway. Registration Error Analysis for Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), Aug. 1997.

[4] S. Julier, Y. Baillot, M. Lanzagorta, L. Rosenblum, and D. Brown. Urban Terrain Modeling For Augmented Reality Applications. In M. Abdelguerfi, editor, *3D Synthetic Environment Reconstruction*, The Springer International Series in Engineering and Computer Science, pages 119–136. Springer US, Boston, MA, 2001.

[5] M. A. Livingston and Zhuming Ai. The effect of registration error on tracking distant augmented objects. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 77–86, Cambridge, UK, Sept. 2008. IEEE.

[6] B. Macintyre, E. M. Coelho, and S. J. Julier. Estimating and Adapting to Registration Errors in Augmented Reality Systems. In *IEEE VR 2002*, pages 73–80. IEEE Press, 2002.

[7] B. MacIntyre and E. Machado Coelho. Adapting to dynamic registration errors using level of error (LOE) filtering. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 85–88, Munich, Germany, 2000. IEEE.

[8] N. Numan, Z. Lu, B. Congdon, D. Giunchi, A. Rotsidis, A. Lernis, K. Larmos, T. Kourra, P. Charalambous, Y. Chrysanthou, S. Julier, and A. Steed. Towards Outdoor Collaborative Mixed Reality: Lessons Learnt from a Prototype System. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Shanghai, China, 2023.

[9] J.-H. Pauls, K. Petek, F. Poggenhans, and C. Stiller. Monocular Localization in HD Maps by Combining Semantic Segmentation and Distance Transform. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4595–4601, Oct. 2020.

[10] L. Platinsky, M. Szabados, F. Hlasek, R. Hemsley, L. D. Pero, A. Pancik, B. Baum, H. Grimmett, and P. Ondruska. Collaborative Augmented Reality on Smartphones via Life-long City-scale Maps. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 533–541, Nov. 2020. ISSN: 1554-7868.

[11] G. Reitmayr and T. W. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, Oct. 2006.

[12] J. P. Wilmott, I. M. Erkelens, T. S. Murdison, and K. W. Rio. Perceptibility of Jitter in Augmented Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 470–478, Oct. 2022.

[13] C. Ye, Y. Wang, Z. Lu, I. Gilitschenski, M. Parsley, and S. J. Julier. Exploiting Semantic and Public Prior Information in MonoSLAM. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4936–4941, Oct. 2020.

[14] H. Zhou, D. Zou, L. Pei, R. Ying, P. Liu, and W. Yu. StructSLAM: Visual SLAM with building structure lines. *IEEE Transactions on Vehicular Technology*, 64(4):1364–1375, 2015.