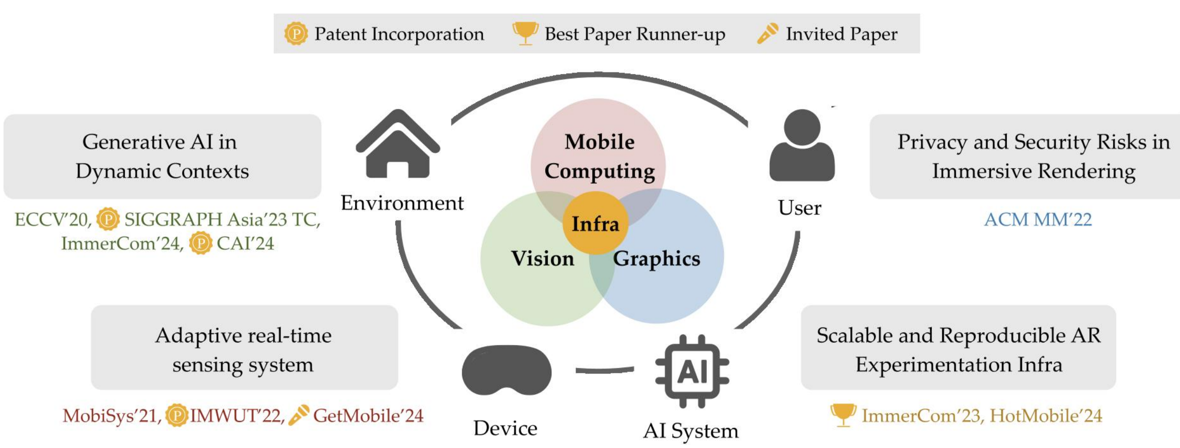
Dynamic Intelligent System for Immersive Mobile Computing

Yiqin Zhao (Worcester Polytechnic Institute)

The world is moving toward a future where *computing and information* can seamlessly blend into *physical spaces*. However, this transformation remains fundamentally challenging as mobile computing systems must now support accurate and robust real-world information understanding. Moreover, mobile computing systems must also deeply engage with dynamic physical environments and complex user activities. Towards addressing these fundamental challenges, my research objective focuses on empowering mobile computing systems to adapt to the complex *dynamics* from key mobile computing stakeholders–user, device, environment, and AI system. Centered around this objective, I have developed *AI models* and *mobile system supports* for high-impact dynamic application scenarios, with a particular focus on supporting *immersive user experiences*.

My research centers on mobile Augmented Reality (AR) as the primary application scanrio. A key aspect of my research is building *adaptive environment sensing system* to support unstable observations introduced by device movement and environmental dynamics. In particular, my Ph.D. thesis centers around lighting estimation [3, 4, 6, 7, 9], a key task that impacts user-perceived visual immersion. At the same time, my research also looks into prominent privacy issues arise in immersive applications. With sensed environmental information being used to create immersive experiences, user privacy in AR contents can also expose to maliciously or unintentionally usage. My work [12] protects out-of-screen user privacy leakage from reflective AR object renderings.

Beyond technical innovation, my work also has far-reaching societal and educational implications. In particular, ExShot [10] addresses the issue of *identity appearance shifting* by utilizing a context-aware neural rendering model with mobile photography information. I am also committed to democratizing mobile AR research and development by contributing to the foundational infrastructure research. I developed ARflow [8], an open-source data streaming and visualization framework for AR devices, and ExpAR [1], a controllable and programmable platform designed to evaluate various AR systems with standardized user mobility and physical environments.



Research overview. My research works have seen academic recognition, such as a best paper runner-up award for ExpAR at ImmerCom'23 [1], and industry product and patent incorporations at Google [10] and Adobe [11]. Throughout my Ph.D., I have fostered many connections with researchers from the mobile computing, computer vision, and computer graphics communities through collaborations. Looking forward, I plan to bring the same combination of cross-disciplinary collaboration and industry interaction to my work as a professor.

Previous and Current Work

Adaptive real-time sensing system. Sensing physical environments is a foundational supporting pillar of immersive mobile computing. A main chapter of my Ph.D. thesis focuses on designing *environment sensing systems that can adapt to dynamic environments*. With key goal of supporting visually immersive user experiences, my Ph.D. work primarily focuses on the lighting estimation task [3, 4, 6, 7, 9], and depth estimation [2, 5]. Below, I briefly described two of my recent works Xihe [7] and LitAR [9].

Immersive mobile computing devices, e.g., AR headsets, are resource-constrained computing platforms. Therefore, having real-time sensing workload scheduling is critical to ensure the temporal stability of immersive user experiences. Drawing insights from information theory and 3D geometry processing, I designed an *information entropy-based triggering mechanism* in Xihe [7]. This mechanism effectively triggers lighting estimation requests based on changes in device environment observationzsa data. Specifically, our evaluation shows this triggering strategy can reduce over *76.24%* of unnecessary lighting estimation requests.

Device movement plays a significant role in how environment observation changes. Therefore, understanding and modeling device mobility patterns are crucial for building system support for diverse user movements in mobile computing. In another work LitAR, I comprehensively studied device and user movement patterns in placement-based AR applications. The key challenge in this study is to generate controlled user mobility. To tackle this challenge, I developed a photorealistic simulation environment to study lighting estimation under controlled device usage patterns, including different device positions, heights, and movement patterns. Inspired by the findings, I characterized the user mobility patterns and developed a near real-time pipeline to generate environment maps from both *opportunistic- and guided-sensing data* from mobile AR devices.

Generative AI in dynamic contexts. While generative models can be powerful tools for immersive experience creations, they must be deeply integrated with the dynamic context to avoid unrealistic outputs. My past work explored effective and safe integrations of generative AI in immersive mobile computing, with a key focus on *context awareness*. In this aspect, my research includes generative lighting estimation [3] and neural rendering-based portrait editing [10].

The sensing capabilities of mobile computing devices are often limited by their physical form factors. As a representative example, lighting estimation is inherently challenging because mobile computing devices can not acquire spatially-variant omnidirectional environment observations. In my ongoing work, ClearAR [3], I leveraged AR context-conditioned image diffusion models to tackle this challenge. My approach leverages multimodal AR context data as control signals to guide a diffusion model-based pipeline for generating environment lighting. This generative estimation design effectively combines knowledge from both large generative models and dynamic runtime AR context to compensate for the limited environment sensing coverage.

Conforming to dynamic context is critical for sensitive content generation, such as portrait photo editing. The facial identity appearances in portrait photos include many details, like facial structures and skin features. These details can change based on factors like environmental lighting or skin conditions. In my previous work [10], I collaborated with experts from Google Research and addressed the *identity-shifting* issue during portrait expression editing with ExShot.My key insight in this work is to use photo sequences, key contextual data in mobile photography, to inform and condition expression editing models. My approach combines explicit expression and head motion modeling with neural rendering-based generative expression editing.

Privacy and security risks in immersive rendering. The technologies that enable immersive user experiences can also be used in malicious attacks. In my previous work [12], I studied a data privacy leakage issue in AR live streaming applications. The attack is made possible because the virtual object reflections can reveal the physical environment information of AR live streamers. Because AR live streamers seek visually coherent virtual content rendering, they often use advanced environment scanning or real-time 360° in their streaming. Under these setups, the attacking mechanism renders streamers unaware and unprepared for the attack. In defense, we proposed a privacy-preserving rendering technique that runs in parallel to the main stage setup pipeline and effectively filters out sensitive information without introducing time delay.

AR experimentation infrastructure. Conceptually simple, it can be extremely difficult to evaluate an AR system fairly and in scale to understand its performance in dynamic real-world environments. To understand how well an AR system works, existing research often conducted tailored and isolated evaluations for specific AR tasks, e.g., depth or lighting estimation, and compared them to easy-to-setup baselines, either using datasets or resorting to time-consuming data capturing. In our previous work ExpAR [1], *we pinpointed the limitations of existing AR evaluation methodology by characterizing recent papers and reflecting on our evaluation practices.* Leveraging these insights, we designed an AR researcher-centered evaluation platform prototype framework with controllable and programmable mobile devices. Continuing this research vision, I developed and released ARFlow [8]. As a key component in our infrastructure project, ARFlow enables high-performance real-time data streaming and visualization with major AR software and hardware platforms. I am currently mentoring two undergraduate students to improve ARFlow and expand its hardware compatibility and data type support.

Future Directions

My future work will continue my current research momentum and expand to the broad areas of *dynamic intelligent systems* and *immersive mobile computing*, including the following thrusts.

Short-term: sense to immerse. My short-term research objective is to develop *sensing* system support for *immersive* user experiences. I plan to explore this objective from two directions: *(i)* foundational spatial sensing system and *(ii)* personalized long-term context sensing system.

Immersive user experiences require a *foundational spatial sensing system* that understands the complex dynamics introduced by all mobile computing stakeholders. Building on my Ph.D. thesis work on lighting estimation, I plan to expand my research to broad types of sensing objectives. In addition, I will explore user-centered evaluation methods for how sensing systems impact immersive experiences. My previous work ExpAR, which is supported by an NSF planning grant, takes the first step towards building a controllable evaluation platform for mobile AR. My future research will take another step forward by scientifically studying the gap between physical sensors and user perceptions. On the theory side, I aim to leverage psychophysics theories to develop a model of user perception of physical sensing errors. On the application side, I plan to develop tools and benchmarks for quantitative measurement of sensing systems' accuracy and sensational measurement of user experiences from large-scale user studies. With these theories and applications coming together, I plan to achieve a standardized evaluation platform that lays the foundation for evaluating user-centered sensing system support for immersive experiences.

Immersive user experiences also demand personalized content generation to cater to per-user's needs. Sensing long-term context from user activity data can help mobile systems understand critical user information, just like the user identity appearances in portrait editing. However, as the system collects and processes user-related context, the risk of privacy leakage also increases. My previous research in privacy-preserving reflection rendering well positions me to explore new challenges, particularly those involving broader spectrums of attacking entities. In particular, I plan to research how to protect face authentication systems from advanced deep fake technologies and maliciously manipulated environment light sources. This exciting direction is currently supported by NSF and will continue to evolve.

Long-term: continuous learning and protection. My long-term objective is to further adapt computing systems to dynamic environments by enabling continous system learning and protection. Building upon my short-term objective, I aim to explore these topics through two approaches. First, I plan to develop in-context learning system supports for context-aware AI models. Specifically, I will explore efficient and safe on-device in-context learning methods. Second, I will explore theories and tools to enable dynamic updating of neural network weights for in applications that use long-term context. This approach will provide key support in building AI-powered personalized application experiences in the future.

Information privacy protection is critical and crucial for systems that continuously learn from highly sensitive user data. Therefore, I plan to design new data privacy frameworks, with a particular focus on safeguarding physical data privacy. Specifically, I plan to develop advanced data obfuscation techniques to prevent sensitive information from being leaked during continuous learning. These efforts will enable intelligent systems to adapt safely in privacy-critical environments, such as enterprise workplaces, ensuring flexible and secure system usage.

Broader vision. Dynamic intelligent systems for immersive mobile computing have the potential to be a key enabler across multiple domains. My broader vision is to empower humans to transcend the limitations of distance and physical presence through these systems. This could be realized through applications such as robot teleoperation via augmented reality (AR), advanced remote sensing, and telepresence technologies. I will collaborate with experts from these domains to explore foundational techniques to enable public adoption of these applications.

References

[1] Ashkan Ganj, **Yiqin Zhao**, Federico Galbiati, and Tian Guo. 2023. Toward Scalable and Controllable AR Experimentation. In *Proceedings of the 1st ACM Workshop on Mobile Immersive Computing, Networking, and Systems (ImmerCom'23)* (Madrid, Spain). Association for Computing Machinery, 237–246. **[PDF]. Best paper runner-up award**.

[2] Ashkan Ganj, **Yiqin Zhao**, Hang Su, and Tian Guo. 2024. Mobile AR Depth Estimation: Challenges & Prospects. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications* (San Diego, CA, USA) *(HOTMOBILE '24)*. Association for Computing Machinery, New York, NY, USA, 21–26. **[PDF]**.

[3] **Yiqin Zhao**, Mallesham Dasari, and Tian Guo. 2024. CleAR: Robust Context-Guided Generative Lighting Estimation for Mobile Augmented Reality. In *Submission*. **[arXiv]**.

[4] **Yiqin Zhao**, Sean Fanello, and Tian Guo. 2024. Multi-Camera Lighting Estimation for Mobile Augmented Reality. In *GetMobile: Mobile Computing and Communications (GetMobile Vol.28, No.3)*. ACM New York, NY, USA. **[PDF]. Invited paper**.

[5] **Yiqin Zhao**, Ashkan Ganj, and Tian Guo. 2024. Toward In-Context Environment Sensing for Mobile Augmented Reality. In *Proceedings of the 2st ACM Workshop on Mobile Immersive Computing, Networking, and Systems (ImmerCom'24)*. Association for Computing Machinery, Washington, D.C., USA. **[PDF]**..

[6] **Yiqin Zhao** and Tian Guo. 2020. PointAR: Efficient Lighting Estimation for Mobile Augmented Reality. In *Computer Vision – ECCV 2020 (ECCV'20)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 678–693. **[PDF]. Top-tier computer vision conference**.

[7] **Yiqin Zhao** and Tian Guo. 2021. Xihe: a 3D vision-based lighting estimation framework for mobile augmented reality. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services* (Virtual Event, Wisconsin) *(MobiSys '21)*. Association for Computing Machinery, New York, NY, USA, 28–40. **[PDF]. Top-tier mobile system conference**.

[8] **Yiqin Zhao** and Tian Guo. 2024. Demo: ARFlow: A Framework for Simplifying AR Experimentation Workflow. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications (HOTMOBILE '24)*. Association for Computing Machinery. **[PDF]. Project repository: https://github.com/cake-lab/ARFlow**.

[9] **Yiqin Zhao**, Chongyang Ma, Haibin Huang, and Tian Guo. 2022. LitAR: Visually Coherent Lighting for Mobile Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 153, 29 pages. **[PDF]. Top-tier mobile system conference. Incorporated in US patent application: 18/237,095, 2024**.

[10] **Yiqin Zhao**, Rohit Pandey, Yinda Zhang, Ruofei Du, Feitong Tan, Chetan Ramaiah, Tian Guo, and Sean Fanello. 2023. Portrait Expression Editing With Mobile Photo Sequence. In *SIGGRAPH Asia 2023 Technical Communications* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 17, 4 pages. **[PDF]. Top-tier computer graphics conference. Incorporated in a Google classified patent application**.

[11] **Yiqin Zhao**, Yu Shen, Stefano Petrangeli, Matheus Gadelha, Cuong Nguyen, and Gang Wu. 2024. Copiloting Creative 3D Scene Modeling and Visualization with Generative Agents. In *Creativity & Generative AI: A dialogue between machine learning researchers and creative professionals (CAI'24)*. **Incorporated in an Adobe classified patent application**.

[12] **Yiqin Zhao**, Sheng Wei, and Tian Guo. 2022. Privacy-preserving Reflection Rendering for Augmented Reality. In *Proceedings of the 30th ACM International Conference on Multimedia* (Lisboa, Portugal) *(MM '22)*. Association for Computing Machinery, New York, NY, USA, 2909–2918. **[PDF]. Top-tier AI conference**.