Context-Aware Markerless Augmented Reality for Shared Educational Spaces

Tim Scargill*

Department of Electrical and Computer Engineering, Duke University

ABSTRACT

In order for markerless augmented reality (AR) to reach its full potential in educational settings it must be able to adapt to the wide range of possible environments, devices and user cognitive states that affect learning outcomes. Shared educational spaces, such as classrooms, art galleries, museums, teaching hospitals and wildlife centers present enticing opportunities in terms of specialized AR applications, existing infrastructure to leverage, and large numbers of AR sessions to gather data on. In this work we make feasible the challenging concept of context-aware AR through the use of a 'local expert', which learns the optimal configuration of AR algorithms and virtual content for the specific educational space and use case for which it is implemented. To compute insights from multiple AR devices, enable timely responses to fast-changing user cognitive states, and ensure the security of sensitive user data, we propose an edge-computing architecture, in which storage and computation related to our local expert is performed on a server on the same local area network as the mobile AR devices.

Keywords: Markerless augmented reality, context-aware augmented reality, edge computing.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Computing methodologies—Artificial intelligence—Computer vision—Scene understanding

1 INTRODUCTION

Education is an eminent use case for markerless augmented reality (AR). The ability to display the otherwise invisible wherever convenient for the user, to supplement a real-world view with rich visuals and additional information, and create engaging, interactive and memorable experiences carries huge potential for learning outcomes. There are obvious settings where this can be implemented to great benefit, from classrooms and museums to teaching hospitals and wildlife centers. However, while these scenarios share much in common, the context that the AR system is operating in can be very different. Are we in a large gallery with white walls, an operating room with reflective surfaces, or in a forest, amongst moving branches and shadows? Are the users focused medical students or excitable elementary school children? Even within the same scenario conditions are variable, as lighting, the number of concurrent users, or the cognitive states of users change. Understanding the context of an AR session is critical because it impacts both system performance and how a user perceives virtual content; if we want to optimize educational experiences it is essential that we understand the impact of different conditions and can adapt accordingly.

We model the impact of context on an AR session by considering how the *environment*, the *AR device*, and the *user* interact with one another, which we illustrate in Figure 1. We study these interactions through existing fields of research, shown in dashed boxes. For example how the properties of the environment affect the performance of AR device tracking requires experiments with visual-inertial SLAM (Simultaneous Localization and Mapping); how tracking performance impacts user experience requires studies rooted in visual perception. What is beneficial about this holistic view is that it allows us to appreciate that these relationships do not exist in isolation, that the properties of the environment and the current cognitive state of the user also play an important role in perception. The latter can be detected on an AR device by analyzing eye movements, facial expressions, or other physiological signals, which we term *cognitive context sensing*. Furthermore, it reveals previously unforeseen changes and feedback loops that may occur, e.g., a highly textured environment that generates a large number of map points may result in device heating, causing discomfort to the user and them to produce an unusual motion, in turn reducing the accuracy of device tracking. Only by considering the context of an AR system as a whole can we understand and improve how it functions in practice.

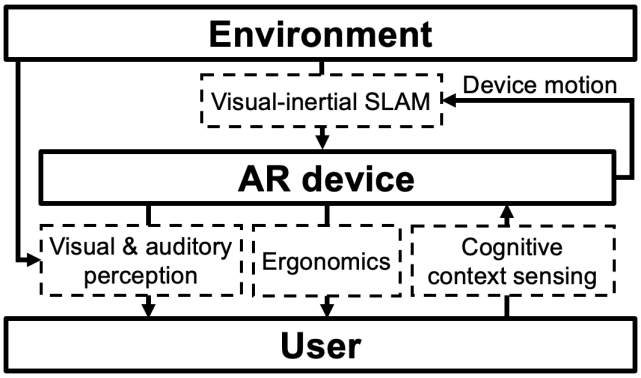

Figure 1: Our holistic model of an AR session. To produce high-quality educational experiences we must consider how both the environment and the user interact with the AR device and each other.

While a number of previous works have investigated context-aware AR, these generally either focus on one aspect of session context, for example the environment [19] or the user [15], or only adapt specific elements of the system, such as the user interface [7], tracking algorithm [17] or virtual content [21]. Our vision is closest to that proposed in [5], in which a wide variety of human, environmental and system factors inform changes to both device configuration and virtual content. Furthermore, the vast majority of current implementations are based on simple controllers, which limits their effectiveness; as Grubert et al. [5] note in a survey of existing context-aware AR approaches *"the context-controllers found currently in context-aware AR are too simplistic to model or infer detailed contexts"*. One element of novelty in our work is to bring greater intelligence to context-aware AR, by using machine learning to predict the impact of different conditions. Even then however, modeling and adapting to all possible scenarios and conditions remains an extremely challenging problem. Our key insight is that

*e-mail: ts352@duke.edu

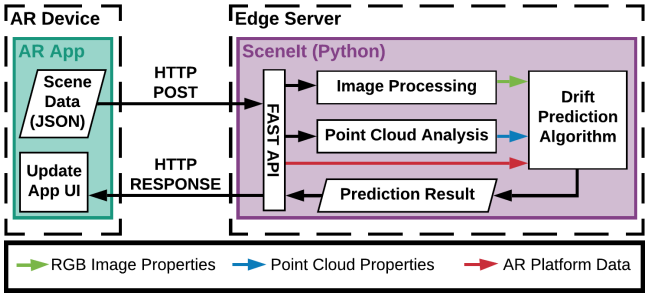

Figure 2: The edge computing-based architecture for our AR visual environment rating system [18].

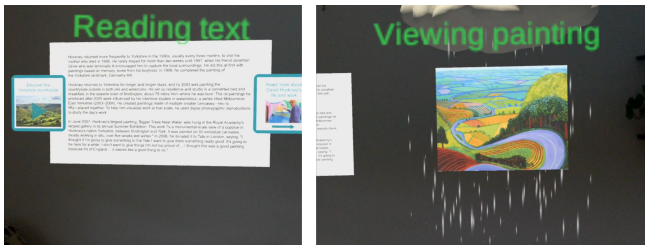

Figure 3: The prototype of our context-aware virtual art gallery experience on the Magic Leap One AR headset. Additional relevant virtual content is rendered when specific user activities are detected using eye tracking.

this is much more feasible for one specific location; a system which aggregates and analyzes the data from multiple users in that space to learn about the surrounding environment, how those users move around, and how they engage with virtual content. The shared educational spaces we described above (schools, museums, hospitals) are well suited to this specialized context-awareness because the AR users in these locations are interacting with specific applications and virtual content, and we can leverage data from large numbers of users over extended periods of time.

One requirement for this type of context-aware AR system is therefore the ability to store and process large amounts of data from multiple sources. This might lead us to consider a cloud computing-based solution. However, one issue with that is that the context of an AR session, in particular human cognitive states and gaze, can change rapidly, in the order of tens of milliseconds, which requires us to respond faster than the speed of communication with a distant server. Furthermore, there are serious privacy concerns with sending images of the environment that may contain other users' faces, sensitive eye movements or other biosignal data to an external location. It is clear that a fully context-aware AR system requires on-site storage and computation, much closer to the AR devices themselves.

Thankfully an emerging distributed computing architecture, edge computing, provides the answer. By placing a dedicated server on the same local area network as the AR devices we can achieve the low latency and security required for context-aware AR. While the combination of AR and edge computing has been proposed before to facilitate computation offloading (e.g., object recognition [11]), including systems which adapt to different network conditions [12], our work is the first to implement and test edge-supported AR which adapts to different environments and user states. Another benefit of working with shared educational spaces is that this architecture is usually relatively easy to implement, because institutions like schools, museums and hospitals often already have the necessary infrastructure. We envision the future of education in which AR plays a central role, and learning experiences in any given location are constantly being optimized for the current context by a 'local expert', enabled by edge computing. In our work we bring together the necessary expertise on both AR and edge computing to study the impact of context on educational AR experiences, and develop and test practical implementations of context-aware AR systems.

2 Methodology

The goals of this work are to increase understanding of how the context of a markerless educational AR session impacts user experiences and learning outcomes, and to develop and test practical implementations of context-aware AR systems in real educational settings. We will achieve this through four phases of research, outlined below.

2.1 Effect of the environment on hologram stability

First, we have studied how the properties of the environment and device motion affect the performance of markerless AR systems, and in turn the appearance of virtual content. One of the key issues with markerless AR is that the spatial consistency of virtual objects relies on visual-inertial simultaneous localization and mapping (VI-SLAM), which is subject to errors if visual conditions or inertial data is challenging. These VI-SLAM errors manifest as hologram instability, which is distracting for users and reduces their level of engagement with the application. Indeed, we conducted an IRB-approved online Qualtrics survey of 39 AR users, and 74% of respondents noted that hologram instability had a significant impact on their satisfaction with markerless AR apps. In our ongoing work, selected elements of which we demonstrated in [18], we have developed the first prediction model for hologram stability based on visual conditions and device motion, and implemented it using an edge computing-based architecture, shown in Figure 2. In this way we demonstrated how the environmental context of an AR session can be assessed, and where appropriate used to guide the user towards better environments and experiences.

Another possibility is supplementing the environment data gathered by AR devices with data from external Internet of Things (IoT) devices, such as cameras or light sensors, enabling us to model the characteristics of a space over time. We have already implemented a proof of concept system for edge-based automatic environment improvements using IoT devices (by switching on a LIFX smart bulb when a low light level is detected by the AR device), and we will investigate combining these ideas into a system which automatically maintains the optimal light conditions for AR throughout the day. In planned work which also involves examining lighting changes, we will test how environmental conditions affect the accuracy and latency of spatial anchor-based localization for persistent virtual content. As well as performing quantitative measurements of spatial consistency errors, we will conduct user studies to better understand their subjective impact under different conditions, such as different screen sizes, viewing distances, virtual content properties and background textures. We have already obtained IRB approval for studies on user perception of AR content, for both smartphones and headsets.

2.2 Cognitive context-based content adaptation

Secondly we have explored how cognitive context sensing can be applied to detect the current state of the user and adapt virtual content accordingly. Specifically, in a recent paper [8] and work under review we developed eye tracking-based activity recognition for the scenario of an art gallery or museum, gathered training data for our activity classifier from multiple users using an IRB-approved process, and produced a working prototype on the Magic Leap One AR headset (shown in Figure 3). In future work we will investigate how cognitive attributes that affect learning outcomes, such as fatigue, mental workload and engagement can be best measured using AR

devices, informed by research from the fields of human-computer interaction, eye tracking and psychology. For example, a variety of eye movement measures such as blink rate have been used to predict fatigue [22], and studies indicate that both pupil dilation [20] and facial expressions [16] can be good indicators of cognitive workload. Recent work has also explored how other physiological signals such as electrodermal activity may be captured to aid in these tasks, while wearing a virtual reality headset [13]. We will then experiment with how these attributes can be used to inform virtual content adjustments, for example reducing the number of virtual objects when mental workload is high, or changing the type of content when engagement is low.

2.3 Virtual content perception in educational scenarios

Next we will research how user perception of virtual content in educational scenarios is impacted by environment properties and user cognitive state, and in turn how that impacts learning. This includes both the perception of information embedded within virtual objects, and the perception of virtual content errors such as hologram instability. For example, we should know if the characteristics of a virtual animal might not be seen by the user when it appears in certain real-world spaces, or when the user is not paying attention to the AR application, but that type of context might be useful for hiding temporary virtual content errors. Alternatively we could adapt the color or dynamics of a virtual body part to make it stand out against the current background or lighting conditions, or direct the user's focus towards it. We will draw from existing research on text legibility in AR with different backgrounds, including the automatic adaptation of text color [1, 2, 14], as well as work on quantifying and adjusting for the color blending effects in optical see-through AR [3, 6, 10]. Our goal is to expand the types of virtual content that is adapted, focus on educational scenarios, and consider how optimal content configurations can be learned and stored for a specific environment.

To this end we will conduct user studies that test context-aware content for education-related tasks. For example we will measure how the speed and accuracy of reading comprehension (rather than only text legibility) is affected when text color is altered according to the surfaces in the environment. We will test the ability to recognize and recall shapes, anatomical or botanical features when the original information was presented without AR, using AR content with fixed properties, or with adaptive AR content. We will also develop a virtual museum experience for which we can measure user engagement, information retention and overall subjective experience. For this scenario we will test optimizing content placement to minimize user perception of hologram stability errors or latency, and changing types of content according to the user's current cognitive state, for example presenting different subject matter if we detect the user is bored or distracted.

2.4 Edge-supported testbeds in educational settings

Finally, we will implement and test these context-aware AR systems in real educational scenarios and settings; we have and will continue to establish partnerships with local wildlife centers, art galleries and museums and other institutions to aid in this. These projects will enable us to uncover and develop creative solutions for challenges related to environmental conditions, resource-constrained mobile devices, multiple users and limited network bandwidth. For example, our work with Duke Lemur Center and discussions about the difficulty seeing and identifying some species motivated aspects of our collaborative image recognition system for AR [9]. We also tested the smartphone resource consumption of highly complex 3D models such as sculptures that might appear in art gallery or museum settings, and found that the scale and complexity of models and the type of shadows rendered had a significant impact on CPU and memory usage, as well as the frame rate achieved. Possible ways to support the demands of context-aware AR, in which multiple different complex models may need to be provisioned, include loading holographic content at runtime from an edge server as we presented in [4], or offloading some or part of the rendering to the edge server. In future we will implement full edge-supported testbeds that allow us to capture environment, device and user data over extended periods, and conduct in-the-wild studies of context-aware educational AR.

3 Results and Future Work

Our work will make the following contributions:

- We have designed and implemented edge-based machine learning classifiers that receive data from mobile AR devices (e.g., camera images, eye tracking data) and return results describing the current environment or user context, for example whether a space is likely to support stable holograms, or the activity a user is performing.
- We have developed the first predictive models for visual-inertial SLAM performance (achieving 94% accuracy in a binary classification of error magnitude) and hologram stability from environment characteristics and device motion.
- We are currently designing an experiment to test how environment characteristics impact the accuracy and latency of spatial anchor-based localization for persistent virtual content scenarios.
- We will study how levels of user attention, engagement and fatigue can be detected using device motion, face tracking and eye tracking, and virtual content adapted accordingly.
- We will conduct user studies on the perception of virtual content during education-related tasks, and measure how learning and information retention is affected by virtual content configurations.
- We will implement edge computing context-aware AR systems in at least two real-world educational scenarios, including a virtual art gallery on campus and a wildlife center educational app, and collect user feedback on their experiences.

Acknowledgments

This work was done in collaboration with Maria Gorlatova, Guohao Lan, Jiasi Chen, Shreya Hurli and Luca Ferranti. A number of students have contributed through undergraduate research projects: Bogyung Kim, Priya Rathinavelu, Orion Hsu, Megan Mott and Alex Xu. Some work has also been done with the help of Erin Ehmke and Megan McGrath at Duke Lemur Center. This work was supported in part by the Lord Foundation of North Carolina and by NSF awards CSR 1903136, CNS 1908051 and CAREER 2046072.

References

[1] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva. Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Transactions on Visualization and Computer Graphics*, 20(1):125–139, 2013.

[2] J. L. Gabbard, J. E. Swan, D. Hix, S. Kim, and G. Fitch. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In *Proceedings of IEEE VR 2007*.

[3] J. L. Gabbard, J. E. Swan, and A. Zarger. Color blending in outdoor optical see-through AR: The effect of real-world backgrounds on user interface color. In *Proceedings of IEEE VR 2013*.

[4] M. Glushakov, Y. Zhang, Y. Han, T. Scargill, G. Lan, and M. Gorlatova. Edge-based provisioning of holographic content for contextual and personalized augmented reality. In *Proceedings of IEEE SmartEdge 2020*.

[5] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724, 2016.

[6] N. Hassani and M. J. Murdoch. Investigating color appearance in optical see-through augmented reality. *Color Research & Application*, 44(4):492–507, 2019.

[7] S. J. Henderson and S. Feiner. Opportunistic controls: Leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of ACM VRST 2008*.

[8] G. Lan, B. Heit, T. Scargill, and M. Gorlatova. GazeGraph: Graph-based few-shot cognitive context sensing from human visual behavior. In *Proceedings of ACM SenSys 2020*.

[9] G. Lan, Z. Liu, Y. Zhang, T. Scargill, J. Stojkovic, C. Joe-Wong, and M. Gorlatova. CollabAR: Edge-assisted collaborative image recognition for augmented reality. *To appear in the ACM Transactions on Sensor Networks*, 2021.

[10] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2385–2394, 2016.

[11] L. Liu, H. Li, and M. Gruteser. Edge assisted real-time object detection for mobile augmented reality. In *Proceedings of ACM MobiCom 2019*.

[12] Q. Liu and T. Han. DARE: Dynamic adaptive mobile augmented reality with edge computing. In *Proceedings of IEEE ICNP 2018*.

[13] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez, and A. Lécuyer. Towards real-time recognition of users mental workload using integrated physiological sensors into a VR HMD. In *Proceedings of IEEE ISMAR 2020*.

[14] V. M. Manghisi, M. Gattullo, M. Fiorentino, A. E. Uva, F. Marino, V. Bevilacqua, and G. Monno. Predicting text legibility over textured digital backgrounds for a monocular optical see-through display. *Presence*, 26(1):1–15, 2017.

[15] S. Oh, W. Woo, et al. CAMAR: Context-aware mobile augmented reality in smart space. In *Proceedings of IWUVR 2009*.

[16] A. Pecchinenda and M. Petrucci. Emotion unchained: Facial expression modulates gaze cueing under cognitive load. *PLoS One*, 11(12):e0168111, 2016.

[17] J. Piao and S. Kim. Adaptive monocular visual–inertial SLAM for real-time augmented reality applications in mobile devices. *Sensors*, 17(11):2567, 2017.

[18] T. Scargill, S. Hurli, J. Chen, and M. Gorlatova. Demo: Will it move? Indoor scene characterization for hologram stability in mobile AR. In *Proceedings of ACM HotMobile 2021*. Demo video available at `https://sites.duke.edu/timscargill/sceneit-prototype/`.

[19] T. Tahara, T. Seno, G. Narita, and T. Ishikawa. Retargetable AR: Context-aware augmented reality in indoor scenes based on 3D scene graph. In *Proceedings of IEEE ISMAR 2020*.

[20] W. Wang, Z. Li, Y. Wang, and F. Chen. Indexing cognitive workload based on pupillary response under luminance and emotional changes. In *Proceedings of ACM IUI 2013*.

[21] Y. Xu, N. Stojanovic, L. Stojanovic, A. Cabrera, and T. Schuchert. An approach for using complex event processing for adaptive augmented reality in cultural heritage domain: Experience report. In *Proceedings of ACM DEBS 2012*.

[22] Y. Yamada and M. Kobayashi. Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults. *Artificial Intelligence in Medicine*, 91:39–48, 2018.