Environmental, User, and Social Context-Aware Augmented Reality for Supporting Personal Development and Change

Timothy Scargill* Ying Chen† Sangjun Eom‡ Jessilyn Dunn§ Maria Gorlatova¶

Duke University, Durham, NC, USA

Abstract

Robust pervasive context-aware augmented reality (AR) has the potential to enable a range of applications that support users in reaching their personal and professional goals. In such applications, AR can be used to deliver richer, more immersive, and more timely *just in time adaptive interventions* (JITAI) than conventional mobile solutions, leading to more effective support of the user. This position paper defines a research agenda centered on improving AR applications' environmental, user, and social context awareness. Specifically, we argue for two key architectural approaches that will allow pushing AR context awareness to the next level: use of wearable and Internet of Things (IoT) devices as additional data streams that complement the data captured by the AR devices, and the development of edge computing-based mechanisms for enriching existing scene understanding and simultaneous localization and mapping (SLAM) algorithms. The paper outlines a collection of specific research directions in the development of such architectures and in the design of next-generation environmental, user, and social context awareness algorithms.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 Introduction

Pervasive context-aware augmented reality (AR), in which immersive experiences generated for the user are influenced by the state of the environment and the state of the user [11,22], has already demonstrated notable potential in many applications aimed at improving users' quality of life and helping users reach their goals, such as in the treatment of phobias [6] and educational applications [11]. Enriched and more robust environmental, user, and social context awareness for AR applications is likely to lead to an explosion of context-aware AR applications that support users in reaching a variety of personal and professional goals.

In this position paper, we lay out a research agenda towards advancing context awareness for AR applications, with the ultimate goal of developing a class of context-aware AR applications that support users' personal and clinical development and change [36]. We believe that practical realization of this important class of applications hinges on both improvements in the performance of the existing context detection algorithms, and on the development of methods that enable operating over the types of context that current systems do not capture, such as the understanding of scene changes between subsequent AR sessions and the characterizations of users' encounters with others. The outlined research agenda spans multiple elements of *engineering the Metaverse*. It includes the development of integrated platforms where data streams of AR, wearable, and Internet of Things (IoT) devices are jointly used as sources of environmental, user, and social context information (see Figure 1), the design of enriched edge computing-supported semantic scene understanding and Simultaneous Localization and Mapping (SLAM) capabilities, and the development of new Deep Neural Network (DNN)-based context detection approaches that enable timely operation over multi-modal data streams with time-varying quality. Broadly, our research agenda contributes to realizing the vision of AR as a *positive technology*, that fosters positive growth of individuals [9].

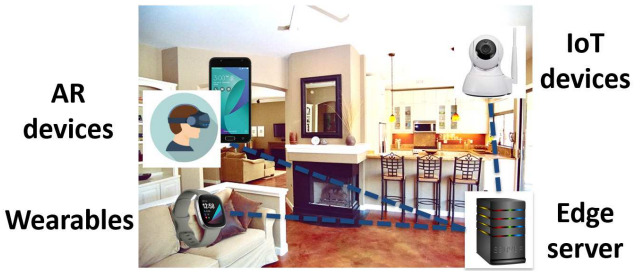


Figure 1: Multi-device platforms for reliable environmental, user, and social context awareness. AR devices' captures are complemented by the data captured by wearable and IoT devices. Edge computing offers a convenient option for efficient processing of the data collected by different devices.

First, in Section 2 we describe a class of applications that are likely to benefit from further development of robust pervasive context-aware AR: specifically, applications that support users' *personal and clinical development and change*. Such *eudaimonic* [21] AR applications will center on helping users achieve professional or educational goals, establish healthier habits, strengthen emotional regulation skills, or combat substance abuse. We note the relationship of the envisioned AR applications to the broader category of *just-in-time adaptive interventions* (JITAI) that have been developed in the context of mobile health technologies [30]. Compared to existing mobile health JITAI solutions, AR has the potential to offer richer, more immersive, and more timely interventions. The envisioned applications impel the development of context understanding techniques that are significantly more robust than the AR context detection approaches developed to date.

Next, in Section 3 we describe the core pillars for improving context awareness for the envisioned AR applications. We note the opportunities associated with *multi-device platforms*, that will bring together AR, wearable, and IoT devices, and will be supporting enhanced context awareness by offering rich complementary sources of contextual information. We also note the opportunities associated with the use of *edge computing* for improving the performance of state-of-the-art SLAM and scene understanding algorithms, and comment on the need for additional research in the area of enhanced context understanding as a whole. We elaborate on these 3 research

*Electrical and Computer Eng. E-mail: timothyjames.scargill@duke.edu
†Electrical and Computer Eng. E-mail: ying.chen151@duke.edu
‡Electrical and Computer Eng. E-mail: sangjun.eom@duke.edu
§Biomedical Eng. E-mail: jessilyn.dunn@duke.edu
¶Electrical and Computer Eng. E-mail: maria.gorlatova@duke.edu

areas in the subsequent 3 sections of the paper.

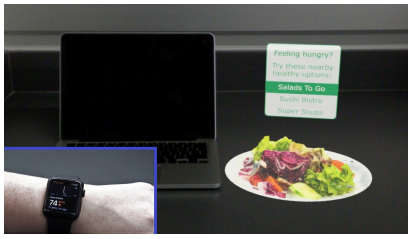
In Section 4 we outline several research directions in the area of multi-device – i.e., AR devices, wearable sensors, IoT devices – platforms for supporting context awareness in AR applications. The core connectivity between different devices we envision deploying can be readily established, via multiple existing communication technologies and platforms. However, as sources of context data, the 3 different device categories have so far been studied largely in isolation. To enable them to work together to achieve robust context awareness, we need to design solutions that operate over diverse, in many cases multi-modal, sources of data, where the quality and the relative importance of different data sources vary over time. We call for the development of domain-specific multi-modal DNNs optimized for real-time execution, and for the development of suitable online metrics of data quality, for different elements of environmental, user, and social context awareness.

Subsequently, in Section 5 we describe a collection of research directions in the area of improved SLAM and scene understanding for context-aware AR. In particular, we describe the opportunities associated with the use of edge computing, a distributed computing paradigm that brings computing closer to the end-users [41], for improving AR device-captured environmental characterizations, semantic scene understanding, and traditional and semantic SLAM. We also describe the opportunities for using IoT-based cameras to improve the performance of semantic scene understanding algorithms and enable the detection of changes in a scene.

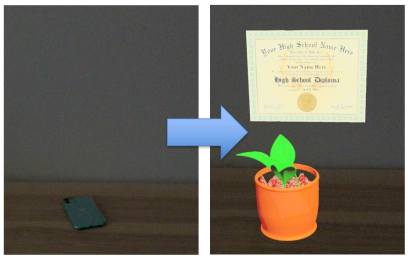
Finally, in Section 6 we outline specific context adaptation mechanisms enabled, in the envisioned applications, by environmental, user, and social context awareness, and describe a set of research directions specific to these 3 context types. In particular, we emphasize the need to develop environmental context awareness interfaces that allow users to correct errors made by context detection algorithms. We also describe the potential of *predicting*, rather than *detecting*, user context, given the richness of context data our applications will be able to collect. In addition, we note the challenge of ensuring *bystander privacy* in collecting social context for our applications.

2 Supporting Personal Development with Environmental, User, and Social Context-Aware AR

Pervasive *context-aware* AR, which generates experiences that are personalized for the user and adapted to the state of the environment around her [11], has recently been envisioned for many types of applications that *support personal development and change* [36], including reducing stress and improving emotion regulation [6], and treating different kinds of substance use disorders [42, 50]. In these applications, context-aware AR has enormous potential as a technology for delivering *just-in-time adaptive interventions* (JITAI) [30], defined as 'interventions aimed to provide the right type of support, at the right time, by adapting to an individual's changing internal and contextual state'. Current modes of delivery for JITAI using mobile health technologies may deliver messages tailored to the user through smartphone or smartwatch app notifications or text messaging. By contrast, AR provides the canvas on which the intervention can be brought to life, potentially allowing interventions to be seamless and immersive. For example, rather than sending a notification that buzzes on a phone as someone walks into a fast food restaurant to "Try the grilled chicken instead of the Big Mac", one can envision blurring out the menu options that are unhealthy, reducing the need for the user to exercise additional self-control after having seen the menu items available that they are more drawn to. AR also has the potential to reduce alarm fatigue associated with existing JITAI notification methods. Correspondingly, multiple practitioner communities are eagerly awaiting for AR to offer the robust environmental, user, and social context recognition that would enable such experiences. For instance, as emphasized in [42], *there is evidence that AR can offer real advances in the understanding and treatment of psychopathology.* Context-aware AR applications can be readily imagined for many different types of personal development and change, such as maintaining a healthy lifestyle and fulfilling personal and professional aspirations.



(a) Supporting healthy eating habits with *user context-aware* AR.



(b) Supporting professional and educational aspirations with *environmental context-aware* AR.

Figure 2: Magic Leap-based mock-ups of two context-aware AR applications that aim to support users' personal development and change. (a) Healthy food suggestions are presented to a user who is predicted to soon feel hunger; (b) To motivate a user to study, a distracting real-world object (phone) is covered with a hologram, and a motivational hologram (a diploma) is presented.

We demonstrate two Magic Leap-based mock-ups of context-aware AR applications we envision to support personal development and change in Figure 2. Both these applications fall in the category of *eudaimonic* technologies, that support human flourishing, growth, and the realization of one's full potential [21]. Figure 2(a) shows a mock-up of an AR application centered on supporting users' goals to maintain a healthy diet. Detecting, via a combination of AR device-based and wearable sensor-based signals, that the user will soon get hungry, the application proactively recommends healthy food options, accompanied by a vibrant, visually appealing depiction of one of them. Figure 2(b) shows a mock-up of a user productivity and motivation-focused application. This AR application is envisioned to be generating and placing, around the users, appropriate motivational holograms, while also hiding real-world distractions. For instance, a user who needs help motivating herself to study can be shown a hologram of a diploma (as depicted in Figure 2(b)) or a visual of herself in a cap and gown, while having distracting objects (TV, gaming consoles) obstructed by virtual objects or blurred.

Realizing these and other context-aware AR applications that support personal development and change requires significant advances in AR's environmental, user, and social context awareness. For example, generating the experience shown in Figure 2(a) requires correctly predicting that the user will soon be hungry, and detecting that the user is likely to be able to process and use the information provided by the AR application (as opposed to, for instance, dismissing it while being engaged in an important task). Generating the experience shown in Figure 2(b) requires the mobile phone that is placed on the desk (shown on the left side of Figure 2(b)) to be reliably identified in a wide range of conditions, and requires the hologram covering it to consistently stay in place (i.e., not be subject to spatial drift, which may be significant in modern AR platforms

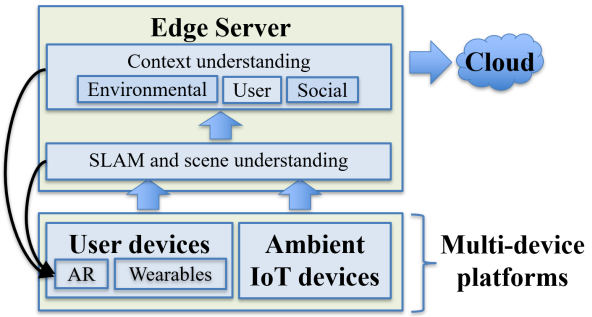


Figure 3: Envisioned pillars of next-generation environmental, user, and social context-aware AR: (1) multi-device platforms that integrate additional sources of context data with the data captured by the AR devices, (2) enhanced edge computing-supported SLAM and scene understanding, and (3) context understanding algorithms enriched by multi-device data, advanced SLAM and scene understanding, and computational capabilities of edge computing.

under many circumstances [43]).

3 Pillars of Next-Generation Context-Aware AR

Realizing AR applications described in Section 2 necessitates significant improvements in the quality and extent of context awareness available to the applications. We argue for the following pillars of next-generation context awareness. The pillars and their relationships to each other are shown schematically in Figure 3.

- **Multi-device integration**: We argue that supporting personal development and change with pervasive AR calls for the development of integrated platforms that bring together AR, wearables, and the IoT. Lightweight and ubiquitous wearable sensors, such as heart rate monitors, can provide important additional information about the physical, cognitive, and emotional state of the user. IoT devices, such as cameras and microphones, commonly deployed in smart homes and offices, can provide additional important signals for establishing environmental and social context. We elaborate on this in Section 4.

- **Edge computing-supported SLAM and scene understanding**: Modern SLAM and scene understanding algorithms may not be sufficiently powerful for the applications we have outlined in Section 2. We argue for the need to develop additional edge computing-supported algorithms, that will incorporate inputs from multiple devices, adapt to practical wireless communication environments, and both extend the range of SLAM and scene understanding features available to the AR devices (e.g., they will include scene change detection) and improve the robustness of the existing algorithms. We describe the specific associated research directions in Section 5.

- **Enhanced context understanding**: The already-impressive sensing capabilities of user-worn AR devices, coupled with multi-device support and advanced SLAM and scene understanding approaches, will allow for the development of the next generation of environmental, user, and social context understanding algorithms for the envisioned applications. We elaborate on this in Section 6.

These pillars will serve as the key enablers of the applications we outlined in Section 2. Additionally, practical realization of these applications, and their-long-term acceptance by the users, will certainly require enhancing *privacy*, which has long been recognized as one of the key concerns in both AR and other related technology areas (such as wearables, smart homes, edge computing, and SLAM) [26, 31, 44, 62]. Towards it, one may envision developing, for instance, edge computing-based *privacy mediators*, which have been proposed in multiple lines of work [41]. Many context-aware AR applications may also benefit from optimized approaches to *storing and delivering large context-specific virtual objects* (e.g., virtual objects that are particular to a given context may be stored on the edge or the cloud, and transmitted to the AR device in advance of being displayed to the user). Several platforms that enable such behavior have recently been proposed and developed [10, 19]. Due to page restrictions, in this position paper we focus on the three pillars highlighted above, leaving the discussion of other important topics for future work.

4 Multi-Device Integration

AR devices capture considerable amounts of information about the user and the environment through the egocentric video, inertial measurement units, eye tracking cameras, and other sensors. However, there are inherent limitations to having an AR headset as the only source of contextual data. To enable the applications we outlined in Section 2, we propose to develop platforms that combine AR devices with the following:

- **Wearable sensors**, such as smart watches, that can provide additional signals for capturing physical, cognitive, and emotional state of the user.

- **Ambient IoT devices**, such as cameras and microphones, that can provide signals for improving environmental, user, and social context detection.

- **Edge servers** that collect data from multiple devices and offer the computational capacity essential for the execution of complex context detection algorithms in real time.

We show the high-level view of the envisioned multi-device platforms in Figure 1. As shown in this figure, we envision using edge computing to collect data from multiple, user-worn and environmental, devices. While one user's AR device and wearable sensors can operate independently without relying on an edge server to coordinate them, edge computing offers a convenient option for running complex context detection algorithms on behalf of these devices. Edge computing also offers a natural point for collecting and analyzing the data captured by environmental IoT devices.

It is worth noting that the core connectivity between AR devices, wearable sensors, IoT devices, and edge servers can be readily established, via ubiquitous communication protocols (e.g., WiFi, Bluetooth Low Energy) and a wide range of available IoT gateways. Indeed, multiple lines of work have integrated wearable or IoT sensors with AR, with the intent of using AR to visualize the data generated by the other devices [16, 28, 32, 60]. However, in using multi-device architectures to enable next-generation context awareness, many research challenges remain, as we elaborate on below.

4.1 AR devices and wearable sensors

Integration of AR devices with wearable sensors that obtain additional biometrics is particularly helpful for understanding the state of the user. Convenient commonplace wearable devices have been used to aid in recognizing arousal, stress, illness, and more via measurements of movement, heart rate, skin conductance, blood oxygen saturation, and skin temperature, among others [8]. To date, AR devices and wearables, as sources of user context data, have largely been examined separately. However, jointly they offer diverse, complementary streams of data that can be used to improve upon existing single-source context detection algorithms. One can imagine exploring joint use of AR devices and wearables for detecting a wide range of different elements of physical, cognitive, and emotional states of the user. For example, how can fine-grained user activity recognition based on AR-provided egocentric camera feeds [52] benefit

from the information captured by wrist-worn inertial measurement units? Can heart rate variability measurements improve eye tracking-based mental and physical fatigue recognition algorithms [45, 58]? Which elements of AR-provided egocentric video captures and eye tracking-based recognition of cognitive and emotional states of the user [3, 20] could aid existing wearables-based stress recognition algorithms [40]? Translating multi-device signal collection to reliable and timely user context detection will likely require developing multi-modal DNNs, optimized for near-real-time execution; an edge server's computational capabilities are likely to be advantageous for ensuring near-real-time performance in these cases.

4.2 AR devices and IoT devices

IoT devices, such as cameras and microphones, have the advantage of providing persistent and continuous 3rd person captures of the environment, in addition to the transient 1st person captures obtained by an AR device. As such, they can be used to capture context that an AR device cannot obtain. For instance, IoT-based cameras can capture whether someone has recently been smoking in a given environment, which may be relevant to applications that aid in smoking cessation (since the environment is likely to have a lingering cigarette smell that may trigger a craving). IoT cameras can also be used to establish whether the environment has changed between different AR sessions, to trigger SLAM remapping as required to improve the quality of spatial scene understanding (and, conversely, to avoid unnecessarily time- and resource-consuming remapping if the environment did not change). We elaborate more on this *scene change detection* capability in Section 5.4.

Additionally, offering different vantage points of the same scene, simultaneous 1st person AR device-based and 3rd person IoT-based captures can be used to improve the accuracy of environmental, user, and social context detection algorithms. Captures from different vantage points have been shown to improve the performance of object detection algorithms [24]. Within the context of SLAM, use of multiple vantage points can be seen as a type of *collaborative SLAM*, where captures of different devices can be combined to obtain a higher-quality overall map that leads to higher-quality pose estimation by the AR device. For user context detection, multiple algorithms have been developed, disjointly, for either IoT camera-based or AR device camera-based operation [2, 52]. Combining IoT-based and AR-based approaches has the potential to offer improvements over either of the separate techniques. Finally, IoT-based monitoring of ambient acoustic signals [31] can be used to improve acoustic-based context detection for audio sources that are located far from the microphone of the AR device.

The key research direction in this space is the *development of robust approaches for combining signals from different devices*. Different vantage points' 'quality levels' will change over time: other noise sources will be interfering with signals captured by different microphones; users of AR devices will move closer and farther from different IoT-based cameras. It is thus important to design algorithms that will adaptively amplify 'good' signals and place less weight on signals that are less useful. In DNN-based context detection, this can be accomplished via the design of appropriate *attention mechanisms* [49]. In SLAM, assigning relative importance to different devices' inputs can be accomplished via measures of captured frames' *information gain* [13, 15].

5 SLAM and Scene Understanding

The applications we describe in Section 2 call for the development of environmental understanding algorithms that are significantly more robust than the state of the art. We highlight a set of related research directions below.

5.1 Edge computing for improving AR-captured environmental characterizations

Modern AR devices are increasingly relying on *depth sensors* to aid with both scene understanding and SLAM. However, depth captures remain sparse and imperfect even on higher-end AR devices. For example, modern LiDAR sensors only capture depth information for 200-500 pixels in an image [33]. In addition, many modern depth sensors struggle with obtaining reliable data when the observed scene contains materials with low reflectivity, strongly specular objects, and reflections from multiple objects [12, 48]. This leads to depth data captures that are missing valid depth estimates in large parts of a frame. For example, in our evaluation of depth estimates obtained by a Microsoft HoloLens 2 in the long throw mode, we found that *30% of depth pixels in a frame were missing, on average*, across a range of captures in representative indoor environments. These challenges can potentially be addressed with modern depth completion, super-resolution, and inpainting approaches such as [27, 33, 64]. Most such techniques are computationally complex; edge computing support is likely to be required to ensure that they are executed in real time.

Another area of improvement for AR device-based estimation of the state of the environment is lighting estimation. Generating realistic 3D rendering of virtual objects in AR requires matching virtual objects' lighting effects (shadows, reflections) to complex omnidirectional lighting conditions of the environment, i.e., *photometric registration*. Detecting lighting conditions correctly is a challenging task, for which multiple DNN-based solutions have been proposed. Recently, [65] demonstrated the use of edge computing to obtain high-quality lighting estimation in real time. Given the complexity of environmental lighting estimation required for realistic virtual object rendering, we believe that edge computing offers a natural fit for developing high-performing photometric registration solutions.

5.2 Edge-supported scene understanding

Multiple lines of work have recently demonstrated the use of edge computing for executing DNNs used for object detection in 2D images [23, 34]. A wide range of more computationally expensive scene understanding techniques, such as those that use both image and depth data, or those that use point clouds [14], are likely to benefit from edge computing support as well. Additional computational support offered by edge computing is particularly important for the types of applications we envision, since higher accuracy in environmental understanding is likely to directly influence user experience (e.g., scene understanding errors that lead to the generation of inappropriate interventions are likely to significantly reduce the effectiveness of the AR application). The tradeoffs associated with using more computationally efficient—but less accurate—models are not likely to be acceptable for these applications.

Envisioned applications' scene understanding is also likely to benefit from the inclusion of stationary IoT-based cameras into the envisioned platforms. For example, these cameras can be used to detect objects that are not in the field of view of the AR device, or provide an additional input for more accurate object classification.

5.3 Traditional and semantic SLAM

Remarkable progress has been made, over the last few years, in enabling reliable SLAM for AR devices. Yet, much room for improvement remains. Significant computational complexity of state-of-the art Visual-Inertial-SLAM (VI-SLAM) employed in modern AR platforms results in both high AR device resource consumption and reduced tracking and mapping performance. Within AR experiences, notable spatial artifacts continue to be encountered in a significant fraction of scenarios. These artifacts, some of which we quantified in our recent work [35, 43], include drift (unintended motion) of virtual objects, spatial inconsistency between the views of different AR devices on the same virtual objects, and long interruptions of

AR experiences following loss of tracking. Multiple research efforts have lately been focused on offloading computationally expensive parts of SLAM pipelines (such as global map optimization, place recognition, loop closing, and map fusion) to edge servers, to reduce mobile device resource consumption [1, 18, 55, 56]. In developing mobile offloading strategies, most of these efforts assume that the bandwidth available for offloading is sufficient and that wireless connectivity is stable, which is not likely to be the case in practice, particularly in environments with multiple AR devices. Realizing the benefits of edge offloading for improving VI-SLAM performance under computation and communication resource constraints in a diverse set of practical conditions is an important research direction.

An exciting and wide range of recent work is combining SLAM with machine learning-based perception (i.e., semantic) algorithms – the area referred to as *semantic SLAM* [4, 38]. SLAM and semantics can be combined in multiple ways [4]: SLAM helping semantics (e.g., [25]), semantics helping SLAM (e.g., [63]), and semantics and SLAM solved within a joint formulation (e.g., [59]). Edge-assisted semantic SLAM approaches have recently started to be developed [57]. Given the already-demonstrated fit of edge computing to improving semantic scene understanding, discussed in Section 5.2, edge computing holds notable promise for improving the performance of a wide range of semantic SLAM algorithms as well.

5.4 Scene change detection

Most environments where one is likely to use applications outlined in Section 2 are *semistatic*: in-between consecutive AR sessions, parts of the environment may be altered (e.g., in a study area, chairs and books may be moved, and blinds may be raised and lowered), which leads to reduced performance in SLAM systems that rely on the so-called *static world assumption* [38]. Stationary IoT-based cameras can be an important aid in determining whether the scene has changed (i.e., whether the AR device can rely on the map it has previously obtained, or whether it needs to map the environment again). Scene change detection based on stationary cameras' inputs is a long-examined, well-formulated problem, for which many solutions have been proposed [53]. Extending existing solutions to incorporate the specific constraints of heterogeneous multi-device platforms we envision (IoT and AR, stationary and mobile, devices), for the specific case of scene change detection in context of VI-SLAM, has the potential to significantly reduce the extent of mapping that would be required to achieve high-quality spatially aware AR experiences.

6 Context Awareness

This section outlines how environmental, user, and social context can be used to generate context-adaptive AR that supports personal development and change, and describes a number of important associated research directions.

6.1 Environmental Context Awareness

Applications: Reliable *semantic understanding* of an environment (i.e., accurate identification of objects and surfaces present in it) opens a wide range of exciting possibilities for AR applications that support personal development and change. One can imagine generating specific types of interventions when an environment includes a certain set of elements. For instance, the user could be reminded of her goals when the environment contains temptations for engaging in unhealthy behavior. It may also be important to ensure that the environment *does not* contain certain elements before an intervention is presented to the user. For instance, in AR-based extinction therapy for substance use disorders, where AR is envisioned to be generating craving-inducing virtual objects [50], the AR-based intervention should be delayed if physical craving-inducing objects are already present in the environment (e.g., it would not make sense to generate a virtual object representing a cigarette pack when a physical cigarette pack is already on the table in front of the user).

A combination of reliable *semantic and spatial understanding* of an environment (i.e., accurate identification of objects and surfaces, coupled with accurate real-time captures of their positions with respect to the user) can enable further important application capabilities. One can imagine highlighting, altering, or blocking certain features in the environment: for instance, in a supermarket, drawing a user's attention to healthy food options and blurring unhealthy ones. The foundations for such *spatially and semantically aware* AR have already been laid, and many exciting futuristic applications that incorporate them have been demonstrated. For instance, TransforMR recently showed how to compose AR scenes so that virtual objects assume behavioral and environment-contextual properties of the real-world objects they replace [17]. Further improvements in semantic and spatial awareness will make an expanding range of environmental context-aware AR-based interventions a reality.

Challenges and research directions: Additional research directions in environmental context awareness, not covered in Section 5, include the need to develop multi-device, edge-supported algorithms for *acoustic environmental context understanding* (e.g., detecting music, loud noise, conversations and their properties), which can impact intervention generation strategies. Another important research direction is in determining the best approaches to representing relevant properties of the environment. Recent examinations of *3D scene graphs*, which are used to capture the semantics and the relationships between objects in an environment [39, 47, 51], may offer a path to compact and easily modifiable representations that capture the properties of the environment important for the envisioned applications.

Another important research direction is developing AR application interfaces that deal with errors in semantic and spatial awareness. The multi-device edge computing-supported mechanisms we have described in Sections 4 and 5 will improve the reliability of existing environmental context detection algorithms. However, it is unrealistic to expect these techniques to fully eliminate all sources of error. At its core, environmental context awareness for AR will continue relying on DNNs and SLAM, both of which are known to fail unpredictably, and for both of which mathematical foundations of robustness have not yet been fully established [38, 46]. We envision borrowing the Internet designers' philosophy of *building reliable applications on top of best-effort lower-layer mechanisms*: AR applications will need to be engineered to take the likelihood of environmental context detection errors into account. For example, one can develop AR applications that allow users to correct errors made by scene understanding algorithms described in Section 5.2: e.g., by correcting incorrectly generated labels, or removing incorrectly generated holograms.

6.2 User Context Awareness

Applications: Within the AR applications we envision, user context detection will allow tailoring virtual objects that are rendered by the AR device to the state of the user. One can imagine, for instance, generating different virtual objects depending on whether the user is sitting still or moving, focused or distracted, tired or energetic. Formally, accurate and timely understanding of the state of the user will translate to the correct identification of the user's *states of vulnerability* (i.e., periods of heightened susceptibility for an undesirable outcome) [30] or *states of receptivity* (i.e., periods when the user is able to process and use interventions provided) [29]. That is, user context detection will allow delivering interventions when they are most likely to be needed by the user, and most likely to be useful to her.

For applications that support personal development and change, the holy grail is *predicting* the future state of the user (rather than detecting the state the user has already entered), to perform an early

intervention. One can imagine, for instance, displaying a calming virtual object before the user gets stressed, or helping the user take a break in advance of cognitive fatigue setting in. The richness of environmental and user context the envisioned systems are expected to be able to gather holds promise to enable such predictions, which in turn will translate to the increased effectiveness of the systems in supporting personal development and change.

Challenges and research directions: In addition to the research directions outlined in Sections 4.1 and 4.2, an important challenge is the expected need to *personalize*, i.e., tailor to the individual users, the algorithms we will be developing for user context detection and prediction. The advantages of model personalization have been established in a wide range of recent work on physical, cognitive, emotional, and other human state recognition and prediction [5, 7]. Due to the difficulties of collecting large amounts of data from individual users, personalization generally involves the development of *few-shot learning* techniques [54], potentially coupled with inventive application-specific approaches to data labeling and augmentation. The challenges associated with collecting, labeling, and training the model with the data of an individual user may be exacerbated in the heterogeneous multi-device systems we envision in this work; innovative approaches are likely to be required in these and other elements of algorithm personalization.

6.3 Social Context Awareness

Applications: Finally, for supporting personal development and change, *social context*, i.e., the state of other humans in the environment, and the nature of a user's interactions with them, plays a very important role as well. For instance, [31] emphasizes the role of positive social contacts for a user's stress response, and [42] notes the importance of social context in the treatment of addiction. For social context detection, the combination of AR devices' and IoT devices' data streams offers a wide range of possibilities that go beyond the 'encounter profiling' possible with traditional mobile technologies (such as users' proximity estimates based on Bluetooth Low Energy, Near-Field Communications, or acoustic signals [61]). One can imagine using AR device-captured 1st person egocentric views to obtain accurate characterizations of the length and the nature of users' encounters with others around her, and using 3rd person IoT-based captures to obtain additional insights into the state of other humans in the environment and the interactions between them. Additional context information can be gathered in cases where multiple users are wearing AR devices or wearable sensors. In this case we would be able to gain further insights into their state, thus obtaining richer context for quantifying the interactions between users. We can also envision developing methods for *joint* adaptation of AR-based interventions. For example, when one of the users is concentrating on her task, others can be shown an augmentation that warns against disturbing her; when two users have been recorded as having had a heated conversation, both can be shown calming, stress-reducing, virtual objects.

Challenges and research directions: Social context recognition methods developed for the applications outlined in Section 2 need to be cognizant of the need to ensure *bystander privacy* [37], which is paramount for the societal acceptance of AR as a whole. Because of this, social context recognition detection may be limited, at least at first, to the environments where all humans support the goals of the AR user (and consent to having their behavior analyzed for the benefit of the user), as could be the case within user's home or, potentially, workplace. In such scenarios, we imagine early experiments and pilot deployments to build on a set of well-developed frameworks for privacy in the context of smart homes, and include explicit consent from all inhabitants and visitors. We envision using edge computing for further privacy mediation: one can imagine, for instance, storing all information gathered about the other humans on a local edge server only, without transmitting it to the cloud. Wider use of social context detection methods would require the development of a range of bystander privacy preservation techniques.

7 Conclusion

This position paper outlines a research agenda towards improving AR applications' ability to capture and adapt to environmental, user, and social context, with the overarching goal of enabling an emerging class of context-aware AR applications that will support users' personal and clinical development and change. The outlined research will enable improvements in the performance of current context detection algorithms. It will also result in the development of context detection techniques that are not feasible with state of the art AR architectures. We define how this enhanced context awareness will be used to enable the envisioned AR applications, and point out several research directions associated with exploiting different types of context that we will be able to collect.

Acknowledgments

We thank Dr. Guohao Lan for helpful discussions. This work was supported in part by NSF grants CSR-1903136 and CNS-1908051, NSF CAREER Award IIS-2046072, an IBM Faculty Award, and grant 2020-218599 from the Chan Zuckerberg Initiative Doner-Advised Fund, an advised fund of Silicon Valley Community Foundation.

References

[1] A. J. B. Ali, Z. S. Hashemifar, and K. Dantu. Edge-SLAM: Edge-assisted visual simultaneous localization and mapping. In *Proc. ACM MobiSys'20*, 2020.

[2] A. Benmansour, A. Bouchachia, and M. Feham. Multioccupant activity recognition in pervasive smart home environments. *ACM Computing Surveys (CSUR)*, 48(3):1–36, 2015.

[3] A. Bulling and T. O. Zander. Cognition-aware computing. *IEEE Pervasive Computing*, 13(3):80–83, 2014.

[4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

[5] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4):1–40, 2021.

[6] D. Colombo, J. Fernández-Álvarez, A. Garcia Palacios, P. Cipresso, C. Botella, and G. Riva. New technologies for the understanding, assessment, and intervention of emotion regulation. *Frontiers in Psychology*, 10:1261, 2019.

[7] J. Dunn, L. Kidzinski, R. Runge, D. Witt, J. L. Hicks, S. M. S.-F. Rose, X. Li, A. Bahmani, S. L. Delp, T. Hastie, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. *Nature Medicine*, 27(6):1105–1112, 2021.

[8] J. Dunn, R. Runge, and M. Snyder. Wearables and the medical revolution. *Personalized Medicine*, 15(5):429–448, 2018.

[9] A. Gaggioli, D. Villani, S. Serino, R. Banos, and C. Botella. Editorial: Positive technology: Designing e-experiences for positive change. *Frontiers in Psychology*, 10:1571, 2019.

[10] M. Glushakov, Y. Zhang, Y. Han, T. J. Scargill, G. Lan, and M. Gorlatova. Edge-based provisioning of holographic content for contextual and personalized augmented reality. In *Proc. IEEE PerCom'20 Workshops*, 2020.

[11] 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, 2017.

[12] M. Hansard, S. Lee, O. Choi, and R. Horaud. *Time-of-flight Cameras: Principles, Methods and Applications*. Springer, 2012.

[13] B. Hepp, M. Nießner, and O. Hilliges. Plan3D: Viewpoint and trajectory optimization for aerial multi-view stereo reconstruction. *ACM Trans. Graph.*, 38(1), 2018.

[14] S. Herrmann. *Object detection with Microsoft HoloLens 2. A comparison between image and point cloud based algorithms*. PhD thesis, TU Wien, 2021.

[15] B.-J. Ho, P. Sodhi, P. Teixeira, M. Hsiao, T. Kusnur, and M. Kaess. Virtual occupancy grid map for submap-based pose graph SLAM and planning in 3D environments. In *Proc. IEEE IROS'18*, 2018.

[16] D. Jo and G. J. Kim. ARIoT: Scalable augmented reality framework for interacting with Internet of Things appliances everywhere. *IEEE Transactions on Consumer Electronics*, 62(3):334–340, 2016.

[17] M. Kari, T. Grosse-Puppendahl, L. F. Coelho, A. R. Fender, D. Bethge, R. Schütte, and C. Holz. TransforMR: Pose-aware object substitution for composing alternate mixed realities. In *Proc. IEEE ISMAR'21*, 2021.

[18] M. Karrer, P. Schmuck, and M. Chli. CVI-SLAM—collaborative visual-inertial SLAM. *IEEE Robotics and Automation Letters*, 3(4):2762–2769, 2018.

[19] M. Khan and A. Nandi. Dreamstore: A data platform for enabling shared augmented reality. In *Proc. IEEE VR'21*, 2021.

[20] G. Lan, B. Heit, T. Scargill, and M. Gorlatova. GazeGraph: Graph-based few-shot cognitive context sensing from human visual behavior. In *Proc. ACM SenSys'20*, 2020.

[21] J. J. Lee and E. Hu-Au. E3XR: An analytical framework for ethical, educational and eudaimonic XR design. *Frontiers in Virtual Reality*, 2:135, 2021.

[22] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proc. ACM UIST'19*, 2019.

[23] L. Liu, H. Li, and M. Gruteser. Edge assisted real-time object detection for mobile augmented reality. In *Proc. ACM MobiCom'19*, 2019.

[24] Z. Liu, G. Lan, J. Stojkovic, Y. Zhang, C. Joe-Wong, and M. Gorlatova. CollabAR: Edge-assisted collaborative image recognition for mobile augmented reality. In *Proc. IEEE IPSN'20*, 2020.

[25] G. Marchesi, C. Eichhorn, D. A. Plecher, Y. Itoh, and G. Klinker. EnvSLAM: Combining SLAM systems and neural networks to improve the environment fusion in AR applications. *ISPRS International Journal of Geo-Information*, 10(11):772, 2021.

[26] B. Martínez-Pérez, I. De La Torre-Díez, and M. López-Coronado. Privacy and security in mobile health apps: a review and recommendations. *Journal of Medical Systems*, 39(1):1–8, 2015.

[27] N. Merrill, P. Geneva, and G. Huang. Robust monocular visual-inertial depth completion for embedded systems. In *Proc. IEEE ICRA'21*, 2021.

[28] K. Michalakis, J. Aliprantis, and G. Caridakis. Visualizing the Internet of Things: Naturalizing human-computer interaction by incorporating AR features. *IEEE Consumer Electronics Magazine*, 7(3):64–72, 2018.

[29] V. Mishra, F. Künzler, J.-N. Kramer, E. Fleisch, T. Kowatsch, and D. Kotz. Detecting receptivity for mHealth interventions in the natural environment. *Proc. ACM IMWUT'21*, 5(2):1–24, 2021.

[30] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, and S. A. Murphy. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462, 2018.

[31] B. W. Nelson and N. B. Allen. Extending the passive-sensing toolbox: Using smart-home technology in psychological science. *Perspectives on Psychological Science*, 13(6):718–733, 2018.

[32] Y. Park, S. Yun, and K.-H. Kim. When IoT met augmented reality: Visualizing the source of the wireless signal in AR view. In *Proc. ACM MobiSys'19*, 2019.

[33] J. Qiu, Z. Cui, Y. Zhang, X. Zhang, S. Liu, B. Zeng, and M. Pollefeys. DeepLiDAR: Deep surface normal guided depth prediction for outdoor scene from sparse LiDAR data and single color image. In *Proc. IEEE CVPR'19*, 2019.

[34] X. Ran, H. Chen, X. Zhu, Z. Liu, and J. Chen. DeepDecision: A mobile deep learning framework for edge video analytics. In *Proc. IEEE INFOCOM'18*, 2018.

[35] X. Ran, C. Slocum, Y.-Z. Tsai, K. Apicharttrisorn, M. Gorlatova, and J. Chen. Multi-user augmented reality with communication efficient and spatially consistent virtual objects. In *Proc. ACM CoNEXT'20*, 2020.

[36] G. Riva, R. M. Baños, C. Botella, F. Mantovani, and A. Gaggioli. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in Psychiatry*, 7:164, 2016.

[37] F. Roesner, T. Kohno, and D. Molnar. Security and privacy for augmented reality systems. *Communications of the ACM*, 57(4):88–96, 2014.

[38] D. M. Rosen, K. J. Doherty, A. Terán Espinoza, and J. J. Leonard. Advances in inference and representation for simultaneous localization and mapping. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):215–242, 2021.

[39] A. Rosinol, A. Violette, M. Abate, N. Hughes, Y. Chang, J. Shi, A. Gupta, and L. Carlone. Kimera: From SLAM to spatial perception with 3D dynamic scene graphs. *arXiv preprint arXiv:2101.06894*, 2021.

[40] C. Samson and A. Koh. Stress monitoring and recent advancements in wearable biosensors. *Frontiers in Bioengineering and Biotechnology*, 8:1037, 2020.

[41] M. Satyanarayanan. The emergence of edge computing. *Computer*, 50(1):30–39, 2017.

[42] M. A. Sayette and M. E. Goodwin. Augmented reality in addiction: Promises and challenges. *Clinical Psychology: Science and Practice*, 27(3), 2020.

[43] T. Scargill, J. Chen, and M. Gorlatova. Here to stay: Measuring hologram stability in markerless smartphone augmented reality. *arXiv preprint arXiv:2109.14757*, 2021.

[44] M. Shibuya, S. Sumikura, and K. Sakurada. Privacy preserving visual SLAM. In *Proc. ECCV'20*, 2020.

[45] G. Sikander and S. Anwar. Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6):2339–2352, 2018.

[46] S. H. Silva and P. Najafirad. Opportunities and challenges in deep learning adversarial robustness: A survey. *arXiv preprint arXiv:2007.00753*, 2020.

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

[48] Texas Instruments. Introduction to time-of-flight long range proximity and distance sensor system design (Rev. B). `https://www.ti.com/lit/pdf/sbau305`, 2019.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Proc. of NIPS'17*, 2017.

[50] C. Vinci, K. O. Brandon, M. Kleinjan, and T. H. Brandon. The clinical potential of augmented reality. *Clinical Psychology: Science and Practice*, 27(3):e12357, 2020.

[51] J. Wald, H. Dhamo, N. Navab, and F. Tombari. Learning 3D semantic scene graphs from 3D indoor reconstructions. In *Proc. IEEE CVPR'20*, 2020.

[52] T. Wang, X. Qian, F. He, X. Hu, K. Huo, Y. Cao, and K. Ramani. CAPturAR: An augmented reality tool for authoring human-involved context-aware applications. In *Proc. ACM UIST'20*, 2020.

[53] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. CDnet 2014: An expanded change detection benchmark dataset. In *Proc. IEEE CVPR'14 Workshops*, 2014.

[54] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.

[55] S. Wen, J. Chen, F. R. Yu, F. Sun, Z. Wang, and S. Fan. Edge computing-based collaborative vehicles 3D mapping in real time. *IEEE Transactions on Vehicular Technology*, 69(11):12470–12481, 2020.

[56] K.-L. Wright, A. Sivakumar, P. Steenkiste, B. Yu, and F. Bai. CloudSLAM: Edge offloading of stateful vehicular applications. In *Proc. IEEE/ACM SEC'20*, 2020.

[57] J. Xu, H. Cao, D. Li, K. Huang, C. Qian, L. Shangguan, and Z. Yang. Edge assisted mobile semantic visual SLAM. In *Proc. IEEE INFOCOM'20*, 2020.

[58] 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.

[59] Z. Yang and C. Liu. TUPPer-Map: Temporal and unified panoptic

perception for 3D metric-semantic mapping. In *Proc. IEEE IROS'21*, 2021.

[60] T. Zachariah and P. Dutta. Browsing the web of things in mobile augmented reality. In *Proc. ACM HotMobile'19*, 2019.

[61] H. Zhang, W. Du, P. Zhou, M. Li, and P. Mohapatra. DopEnc: Acoustic-based encounter profiling using smartphones. In *Proc. ACM MobiCom'16*, 2016.

[62] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu. Data security and privacy-preserving in edge computing paradigm: Survey and open issues. *IEEE Access*, 6:18209–18237, 2018.

[63] T. Zhang, H. Zhang, Y. Li, Y. Nakamura, and L. Zhang. FlowFusion: Dynamic dense RGB-D SLAM based on optical flow. In *Proc. IEEE ICRA'20*, 2020.

[64] Y. Zhang and T. Funkhouser. Deep depth completion of a single RGB-D image. In *Proc. IEEE CVPR'18*, 2018.

[65] Y. Zhao and T. Guo. Xihe: A 3D vision-based lighting estimation framework for mobile augmented reality. In *Proc. ACM MobiSys'21*, 2021.