Towards Seamless Mobile Learning with Mixed Reality on Head-Mounted Displays

Christian Sailer, David Rudi, Kuno Kurzhals, Martin Raubal

ETH Zurich

Institute of Cartography and Geoinformation Chair of Geoinformation Engineering
Stefano-Franscini-Platz 5 CH-8093 Zurich, Switzerland
{csailer, davidrudi, kunok, mraubal}@ethz.ch

ABSTRACT
This paper discusses the opportunities and challenges of utilizing optical (head-mounted) displays for mixed reality in the context of seamless mobile learning. Recent technological developments have significantly improved the capabilities and the mobility of head-mounted displays. Future displays, frequently referred to as data glasses, will be even more lightweight, have a larger naturally sized field of view, and will redefine the standing of mixed reality in economy and society. We envision that these immersive displays will introduce a new world of didactical opportunities and novel context-aware designs for seamless mobile learning, particularly in outdoor learning environments. Our discussion is based on an extensive literature review and first-hand experience with applying this type of novel technology in learning scenarios. Our considerations comprise new ways of collaborative learning and communication between peers, the role of human-computer interaction, the visualization of learning content, and exemplary learning scenarios with mixed reality. We further provide an overview of potential research directions to be pursued in the near future.

Author Keywords
Seamless Mobile Learning, Head-Mounted Display, Mixed Reality, Augmented Reality

INTRODUCTION
The educational advantages of seamless mobile learning with hand-held devices (e.g., mobile phones) have been extensively discussed and well established: As part of a technologically boundary-less environment, learners move in a continuum wherein they perform different tasks across time and locations, switch between formal and informal environments, interact in personal and social spaces, and have ubiquitous access to information and communication (Kukulska-Hulme et al. 2009; Wong and Looi 2011, p.9). One overarching goal is “doing the right thing at the right time and right place with the right contextual tools” (Sharplees 2015), i.e., to make full use of the contextual affordances available at every learning space to mediate varied learning tasks (e.g., geoinquiries, real-time face-to-face or online interactions with the teacher and peers, and sharing personal experiences to trigger collaborative knowledge building), and consciously bridge such learning tasks to form a recursive trajectory of learn-apply-reflect or knowledge synthesis (i.e., not just contextualization, but more importantly, continuous recontextualization of knowledge and skills).

Encompassing physical and digital worlds can be observed also in the field of mobile computing when users play location-based mixed reality games. Mixed reality (MR) describes the blend of real and virtual environments and is part of the reality-virtuality continuum (RVC) introduced by Milgram et al. (1994) (cf. Figure 1). The RVC describes the range of possible interactions between reality and virtuality, depending on their respective contributions to a given scenario. In the RVC, augmented reality (AR) is the depiction of virtual objects in a real-world context. Contemporary AR solutions on hand-held devices utilize a device’s real-world camera image and fuse it with virtual content. Similar solutions exist for see-through (head-mounted) displays (e.g., Microsoft HoloLens). Current AR applications also include tactile feedback and spatial sound (Radu and Schneider 2019; Xu et al. 2019).

Figure 1. Illustration of the reality-virtuality continuum, according to Milgram et al. (1994). We see mobile learning mainly in the area between the real environment and augmented reality due to mobility restrictions in augmented virtuality.

The remainder of this paper is structured as follows: first, we describe a situated learning scenario in an architectural AR context. Then, we identify important technological limitations that affect learning. Based on these we discuss how AR can be adapted to enhance the learning experience. We conclude with a discussion of future research directions.
SCENARIO: STUDENTS OF ARCHITECTURE USING MIXED REALITY ON THE MOVE

The following scenario is taken from one of our learning scenarios and based on one of the oldest AR examples by Feiner, Webster, Krueger, MacIntyre, & Keller (1995). It was designed to teach architecture students the concepts of illustrating, understanding, and modifying architectural anatomy. This mobile learning scenario included augmented reality features of architectural concepts. It was implemented as a city walk where the learning objectives consisted of acquiring the competence of classification and evaluation of structures of urban expansion in the last two centuries. In this example, the learning location is a city, which had previously been a riverine city and developed into a city where commercial and residential areas dominate. Amongst others, the learning content includes the shapes of buildings and their local arrangement in settlements, the main river’s network and its lateral flows, and the resulting placement of the roads and bridges (cf. Figure 2).

In this environment, students explore various locations near the rivers and shores by activating their prior knowledge and following their own strategy. With AR, the city’s hidden structural changes can be made visible to the students when needed, e.g., parts of the river that were later filled. The students individually research phenomena (e.g., anthropogenic constructions) and reason about the urban expansion. An AR system can deliver augmented information as soon as the students enter a virtual perimeter. Meta information about learning objects can be uncovered upon request. Students compare local findings to related work by switching between the augmented and the real world. They debate and reflect on their results within their local group or across groups by means of current communication tools (e.g., chat or video conference). Moreover, they annotate the findings as small chunks of text, voice, or video memos.

Seamless mobile learning implies that students can always store, share, and recall contextualized knowledge autonomously despite a changing physical and social context (Sharples 2015). The extension of this principle would be that students proactively seek out information and use self-regulated learning strategies, such as self-evaluation, goal-setting and motivation, or seeking social assistance (Zimmerman and Pons 1986). This means that in such an educational environment, the students become producers, critical thinkers, team players and not only consumers of digital information.

DEFICIENCIES IN SELF-REGULATED LEARNING ON HAND-HELD DEVICES

In human-computer interaction (HCI) research, location-based AR applications (e.g., games) have been studied for over a decade (Colley et al. 2017; Magerkurth et al. 2005; Montola, Stenros, and Waern 2009). As a result, several issues have been identified related to immersion and situational awareness: Ambient Wood (Rogers et al., 2004), an educational AR game, enhances the woodland with experiments for children to explore the effect of light and moisture on habitats. For
some students, the task to research the woodland’s contextual factors were overstimulating and limitations of the game design distracted some students from different activities. Klopfer & Squire (2008) identified a cognitive overload in an inquiry-based task where students needed to organize, search and evaluate data and information in augmented environments. Attention and working memory are presented as critical factors limiting dynamic decision making and impairing the user’s situational awareness (Endsley 2018). Churchill & Hedberg (2008) consider the key limitation of mobile phones to be their the small screen, hindering the cognitively adequate distribution of learning objects. Specht, Ternier, & Greller (2011) have identified flexible display systems and AR browsers, tracking technologies with six degrees of freedom and computational power to be important. In summary, we identified the following aspects as the most limiting factors for AR on hand-held devices:

1. **Screen size:** The average diagonal screen size of contemporary mobile phones lies between 4-6 inches. This limits access to information during inquiries while reading text or scrolling through images compared to desktop computers (Churchill and Hedberg 2008). Working with small displays is particularly challenging during tasks that involve wayfinding or the collection and access of location data during field trips using maps (Bartoschek et al. 2013; Christian Sailer et al. 2015). Students from a mobile outdoor study by Santos, Hernández-Leo, & Blat (2014) even preferred to use a paper map over a mobile device to get an overview of the whole area. The mobile devices were only used to check their own location and the task.

2. **Field of view:** The image plane depends on the field of view (FOV) of the camera. The viewing frustum of mobile phones is by default very narrow (vertical to horizontal: 50 to 70 degrees) and it is therefore challenging to register the physical surrounding. A human being is naturally accustomed to a field of view of slightly over 210 degrees (Traquair 1927). As a result, users constantly move the hand-held device to explore the surroundings, impairing their immersion (cf. Table 1). To compensate this issue Newman, Ingram, & Hopper (2001) used a false perspective with a fish-eye lens effect to distort the surrounding. However, due to the distance between the device and the human eyes, the physical space around the display cannot be switched off, but can distract a learner’s situational awareness (Christian Sailer et al. 2015).

3. **Human-computer interaction:** Churchill & Hedberg (2008) highlight the lack of an external keyboard and mouse for text messaging with contemporary mobile phones. Today, this is mostly solved with virtual on-screen keyboards, which however occlude the already limited screen space available for most mobile devices. Therefore, services and applications need to reduce information and interaction complexity to fit the parameters of the mobile device.

At the same time, technological developments have significantly improved the capabilities and the mobility of head-mounted displays (HMD). These advancements have enabled a significantly better immersion, increased the devices’ popularity, and affected numerous fields of research and development (Pierdicca et al. 2017). The next section discusses how HMDs could impact the experience of self-regulated mobile learning, opening the door to a new world of didactics and novel context-aware designs to experience outdoor learning environments.

**ADVANTAGES OF SELF-REGULATED LEARNING WITH HEAD-MOUNTED DISPLAYS**

HMDs are not subject to the previously mentioned limitations of hand-held devices, because of their theoretically infinite screen size, their large field of view, and novel interaction modalities. Although current generations of HMDs are still bulky and cumbersome to wear in pervasive scenarios such as mobile learning, it is safe to assume that future solutions will become more light-weight, have a larger field-of-view, and augment the real world with better rendering performance than the current hardware generation. Nonetheless, HMDs introduce new challenges regarding health and safety (Lin et al. 2017), ethics (Poretski, Lanir, et al. 2019), and privacy (Beresford and Stajano 2003; Duckham and Kulik 2006), which are not in the focus of this paper. Instead, we envision an AR device that covers the user’s FOV entirely, providing full immersion in a mixed reality experience for seamless mobile learning and discuss the educational consequences of such a device.

Suárez, Specht, Prinsen, Kalz, & Ternier (2018) conducted a literature review on self-regulated learning activities with hand-held devices and investigated the level of “learner agency” (the initiative or self-regulation of the learner) supported by mobile technology. The review shows that collaborative data collections and peer-to-peer interactions allow learners to have control over their inquiry process, while direct instruction, access to content, and the contextual support (augmented, immersive, and adaptive feedback experiences) are often regulated and superimposed with mobile technology. A head-mounted AR system could tackle the shortcomings of regulated access to content and contextual support but brings also new ideas for natural peer-to-peer interaction within the system. First, we discuss the enhancement of authenticity and affection through natural renderings of contextual support, which creates new teaching strategies, situated problems, experiments, and evaluations of higher complexity than current applications. Second, we explain multi-tasking ideas with peers (peer-to-peer communication), that will change the interaction between peers in the future. We describe the strengths and point towards threats based on the existing research and our own exploratory work.
Contextual Support

<table>
<thead>
<tr>
<th>Description</th>
<th>Hand-held</th>
<th>Head-mounted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextual support in hand-held computing is classified as augmented, immersive, and adaptive feedback experiences (Suárez et al. 2018). Augmented experiences regulate the delivery of information when the learner steps into the virtual perimeter of the real-world geographic area. Adaptive feedback is shown when exploration hints must be given based on the learners’ performance.</td>
<td><img src="image1" alt="Hand-held Contextual Support" /></td>
<td><img src="image2" alt="Head-mounted Contextual Support" /></td>
</tr>
</tbody>
</table>

Table 1. Contextual support with a hand-held device (left) and head-mounted display (right). The different levels of immersion related to the devices’ field-of-view become apparent.

Suárez et al. (2014) presented a prototype for inquiry-based learning with the Google Glass HMD. The authors projected a display in the FOV. The solution’s main advantage was the capability to annotate moments of curiosity with voice commands to take a picture, create a note, or record a video. On the other hand, we noticed that the small display in front of the eye (specifically for Google Glass) made focusing on virtual objects impossible at very close range, as is the case with hand-held devices (e.g., mobile phones). In our reference scenario, an HMD would show the augmented content with a wider FOV than current hand-held devices (cf. Table 1). A seemingly endless screen would maximize the learners’ immersion and remove the screen boundaries of current devices. Moreover, when comparing virtual reality (VR) with AR, recent research has shown that users develop an increased feeling of relatedness and psychological ownership to virtual artifacts in the latter (Poretski, Arazy, et al. 2019). This finding highlights the importance of “real” physical displays for shaping the learners’ perception and to support the motivation for self-regulated learning.

While most hand-held devices consist of built-in geolocation capabilities, current approaches based on simultaneous localization and mapping (SLAM) seem promising for the future to cover large areas that are necessary for contextual support in AR with HMDs (Engel, Schöps, & Cremers, 2014; Cadena et al., 2016). Assuming that future approaches will provide virtual meshes of larger real-world terrain, learners can ask for navigational hints (e.g., Where are my peers?) or context information (e.g., Labeling the city’s skyline) independent of real-world occlusions.

A common question concerning the rendering of virtual objects is: How realistic should the rendering quality be for didactic purposes? Given the constant developments with regard to photorealistic graphics, it is very likely that in the near future virtual objects will no longer be distinguishable from those in the real world. This, however, redefines data collection activities in the real world. In our reference scenario, buildings from previous epochs could be embedded into the real-world scenery and students could collect data in a simulated ancient town and directly compare virtual objects with real ones.

Having virtual objects that are indistinguishable from real ones can result in a false sensation of situational awareness, or the underestimation of hazards. For example, learners in a flow could be thinking that they are moving on structures, while actually being about to fall down a slope. Further, the additional number of contextual factors could risk overstimulation as described above in Ambient Wood (Rogers et al., 2004).

Peer-to-Peer Communication

<table>
<thead>
<tr>
<th>Description</th>
<th>Hand-held</th>
<th>Head-mounted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners interact socially to achieve certain learning goals (Dewey and Boydston 1938). Communication can be asynchronous (e.g., forums) or synchronous where messages are exchanged in real-time (e.g., chats) (Suárez et al. 2018). A head-mounted device additionally provides the possibility to connect remote peers via holographic telepresence.</td>
<td><img src="image3" alt="Hand-held Peer Communication" /></td>
<td><img src="image4" alt="Head-mounted Peer Communication" /></td>
</tr>
</tbody>
</table>

Table 2. Peer-to-peer communication with a hand-held device (left) and head-mounted display (right). Bridging distances (as one of many physical boundaries) is more difficult with hand-held devices than with HMDs.

Scenarios such as our reference example are typically conducted in exploration teams (Christian Sailer et al. 2015), for which we consider new ways of communication and collaboration that arise when employing AR on HMDs:

When an individual learner makes a discovery and wants to share it with peers, HMDs with eye tracking support could project the learner’s gaze in the display of the other peers (Göbel, Kwok, & Rudi, 2019). Peers could easily follow an
individual discovery by simply highlighting an important real or virtual object which could also be enriched with additional information.

AR on HMDs allows us to bring distant peers closer using projections. One option may be the use of an avatar, as it is common on hand-held devices, but remains impersonal. Transmitting a natural projection of the real person (cf. Table 2) via holographic telepresence may be more personal but requires further research from a technical perspective.

An endless screen would further allow learners to research in the augmented reality using shared workspaces between them. Findings can be collected as shared concept maps visualized in the AR space (e.g., (Rudi et al. 2016)). This would allow learners to easily compare considerations, help debate, reflect the peers’ findings, and foster social knowledge construction. Such visualizations are hardly feasible on mobile phones due to the limited screen space.

While exploring cities in different decoupled groups, synchronous communication can become overwhelming. This can lead to cognitive overload and limit self-regulated strategies such as goal-setting, keeping records, self-evaluation, and it can impair the users’ situational awareness (Endsley 2018). The teaching design of the learning scenario plays an important role here and is comparable to learning with hand-held devices (Christian Sailer et al. 2015). Asynchronous communication allows learners to have more breaks to reflect and elaborate (Suárez et al. 2018), which is preferable if the goal is to achieve higher-order cognitive skill such as problem-solving (Garrison, Anderson, and Archer 1999).

**DISCUSSION**

Based on the previously introduced aspects of contextual support and communication between learners, we summarize a set of research directions to improve seamless mobile learning in the future with AR on HMDs. From a technical perspective, many aspects of AR technology will progress independently from didactic purposes. However, we see a large potential for interdisciplinary research between computer scientists, engineers, and pedagogues to tailor AR experiences on HMDs for purposes of mobile learning.

**Visualization of Learning Content**

Following the discussion on rendering quality and situational awareness, we conclude that a separation between two types of scenarios will be necessary: (1) Fully immersive experiences that represent photorealistic content that is seamlessly integrated into the real world. Examples are the representation of former urban landscapes, individual objects of urban space, or (cf. Figure 2) visualizations of traffic routes near and on the river. (2) Scenarios that require abstracted information, for example, statistical and textual information about objects. Here, representations of information on panels that emulate traditional displays, or the spatio-temporal visualization of learners’ trajectories could be applied using the endless screen space (C. Sailer et al. 2016).

**Human-Computer/Human-Human Interaction**

Building on the extensive body of research on HCI in augmented reality, the development and application of concepts to interact with real and virtual objects in such as environment is necessary. Existing techniques based on gestures, voice commands, or tangible controllers might be extendable to improve the learning experience. Furthermore, alternative input modalities, such as from eye tracking, seem promising to identify parts of the environment that could be adjusted to tailor the AR experience to the interests of the user (Kwok and Kiefer 2019). One important question is how future communication between humans can be supported in AR. This needs to be extended to improve collaborative scenarios in mobile learning. If all users perceive the same real and virtual content, shared attention could be guided by augmented content, for example, to point out specific findings in a spatial context. Regarding the urban scenario, one student could show the others where to find a specific building that is highlighted by interacting with the AR environment.

**New Research Directions for Mobile Learning Scenarios**

In the case of a non-technology learning task, technology changes the task at different levels such as enhancement (Substitute with no functional change, or Augment with functional improvement) or transformation (Modify to significant redesigns, or Redefine new unimaginable tasks) of interaction and communication based on the SAMR model by Puenteuda (2010). Mobile learning takes place in the geographic space and links surroundings with digital content. This content is sometimes structured as spatial data (e.g., map features). That’s why our research directions are inspired by the broad field of Geographic Information Science (GIS) (Gold 2006; Longley et al. 2011). We identified six mobile learning research directions based on our reference example where a perfect augmented environment redefines the creation of situated and mobile assisted seamless learning tasks, which are still inconceivable today with hand-held devices:

(1) **Immersive simulations:** One scenario could be the situated inquiry in generated 3D models and simulations of past centuries (e.g., the river city 200 years ago) or in futuristic environments (e.g., to teach urban planning) as real-world renderings. Such inquiries modify or redefine data collections by hand-held devices because the student can choose mixed contextual filters based on individual discoveries and intentions. These allow for discussion and co-constructions that were previously unthinkable. In addition, to collaboratively assess the learning progress, a non-photorealistic rendering can show the data collection or features of interest, for example, in-place representations of
building blueprints (Nienhaus and Döllner 2004). In such an augmented experience, the architecture student should be able to choose the right amount of information needed to avoid missing findings and information overload.

2) **Geospatial analysis:** Situated learning promotes the goal of understanding the learning content in the context of space. Urban space, for example, is characterized by size, population and building density, as well as functional-spatial specialization and socio-spatial differentiation. These peculiarities can be analyzed in a self-regulated way with the help of spatial procedures and visualization in AR displays. In the urban expansion scenario students could perform shadow calculations and experience the results authentically on-site. Such a scenario *redefines* geospatial analysis, which is usually carried out by experts on desktop computers (Turner et al. 2001).

3) **Geospatial modeling:** Another scenario of urban development is the generation of virtual, future neighborhoods. It is based on urban planning rules for predicting population growth, zoning, and for determining transport needs, allocation of social services, etc. (Litman 1995). The modeling of physical effects of congested roads or abstract concepts such as land price disparities or gentrification patterns can be displayed on AR hand-held devices as digital maps. Since visual communication is a crucial factor for the professional handling of important construction projects, future AR displays can *redefine* learning tasks in order to model geographical features at authentic scales.

4) **Gaze-based interaction:** In future scenarios, gaze-based interaction techniques could be used to, e.g., examine an exterior façade of a building for certain historical features using a combination of classical paper plans and utilize one’s gaze for hands-free interaction with audio information objects associated with the building. The system could then make the interaction results available to peers as needed. In this example, eye tracking *augments* mobile data collection and gestures techniques and supports seamless mobile learning by *substituting* the sending and receiving of audio messages and enabling hands-free interaction to use the hands for other tasks.

5) **Close collaboration:** After the students have mastered a more complex task alone or in pairs during an urban expansion scenario, an exchange of experiences and results in larger groups is often necessary. The oral face-to-face exchange of experiences is *redefined* by a screen extension, where the recorded exploration process of the peer groups can be played back. The additional video will possibly enrich the discussion and trigger new questions and new tasks.

6) **Remote collaboration:** If the students can’t meet in the same place, the usual text-based exchange can be *redefined* (cf. Table 2) via holographic telesence of remote peers. In addition, individual points of view could be exchanged between peers. This enables learning activities as in close collaboration and new collaborative learning activities such as locally shared data collection, comparison of one’s own and transferred urban environment. Another possible scenario could be that a student remotely changes the shapes of some buildings and locally retrieves, analyses, and evaluates the effects, such as the shadow calculations.

In summary, with the help of AR displays and a variety of communication, visualization, simulation, and interaction techniques inspired by GIS, learners can dive into completely different contexts and profit from better user experience of the system to their needs. Learners could begin to reconstruct the past or create a future vision on-site, observe ancient phenomena and evaluate geographic models, and make valuable connections with peers and experts. These activities go far beyond architectural and geographical education and also support many other subjects.

**Current Shortcomings**

We postulate that there will be technological improvements that resolve the issues of contemporary HMDs. However, there still exist shortcomings that have to be overcome by future research, which we briefly address in the following.

**Hardware Limitations:** Although contemporary systems can be used outdoors, they still have issues with the lighting conditions (Butz et al., 1999). If the surrounding world is much brighter than the screens or vice-versa, the virtual objects become invisible. Lighting issues have also been observed in the context of outdoor mobile learning with hand-held devices (Christian Sailer et al. 2015). Further limitations that constitute future challenges are the massive amount of data that has to be processed by the SLAM-based algorithms and the data traffic between devices for online data processing.

**Educational Limitations:** Suárez et al. (2018) mentioned the lack of support and guidance in self-regulated learning scenarios and stated: “Too much freedom can generate less desirable learning outcomes and make learners struggling to select, organize and integrate relevant information” (p.41). Suárez et al. (2018) concluded the use of mobile devices for inquiry-based learning is able to foster different types of mobile activities combining a mix of instruction and agency for the learners. This review has distillated suggestions for such a combination to be successful” (p.49).

- Continuous notifications can be annoying and disruptive. The reduction of situational awareness is highly dangerous while moving in public space. Asynchronous communication allows learners to have more breaks to reflect and elaborate, and not be urged to keep up the flow of the conversation (Suárez et al. 2018). This type of communication might be preferable when the goal is to achieve higher-order cognitive skills (Garrison, Anderson, and Archer 1999).

- The learning gain by adding augmented visualizations was not significantly different across conditions on several concepts, even though participant engagement was high (Radu & Schneider, 2019). Some participants had difficulties in understanding the augmented environment, which caused misconceptions. However, these problems occur in mobile learning with hand-held devices as well, and the approach to tackle this problem can be formative assessment on-site or de-briefings, e.g., back in the classroom.

- Related work often considers seams as something negative that needs to be avoided (Dilger, Gommer, and Rapp 2019). However, education should be the preparation for the real world, where seams are everywhere, occur all the
time, and must be learned to cope with. The experience of tuckling seams is important and can also be conceptualized as a trigger for learning (Bronkhorst and Akkerman 2016).

CONCLUSION
HMDs are evolving toward endless screens and natural FOVs. Given these developments, we identify several opportunities for seamless mobile learning activities concerning contextual support and peer-to-peer interactions presented in Suárez et al. (2018). Moreover, we state that the challenges (e.g., hardware limitations, cognitive overload, situational awareness distractions) of integrating augmented environments in education are worthwhile investigating in future research. We expect that current technological developments will provide us advanced AR technology built into normal glasses, or even contact lenses. GIS has many proven methods and applications that enable 2D or 3D representation of the earth and provide geospatial tools to solve geospatial and locational problems across disciplines. Using this field in our mobile AR scenarios, we extracted six research directions such as real-world sized simulations, analysis, and modeling as well as new interactions with gaze and gestures, and collaboration in close (face-to-face) or remote (online) ways. To evaluate our educational scenarios, especially which learning domains and goals can be respected and if recontextualization of knowledge and skills in general works, further empirical research will be required. In particular, we need to identify which principles hold up in AR displays and where new concepts will be necessary to foster seamless mobile learning.

REFERENCES


Suarez, Angel, Stefaan Ternier, Marco Kalz, and Marcus Specht. 2014. “GPIM: Google Glassware for Inquiry-Based Learning.” In Open Learning and Teaching in Educational Communities, Cham: Springer International Publishing, 530–33.


