Techniques for Supporting the Author of Outdoor Mobile Multimodal Augmented Reality

Roger J. Chapman, Ph.D.
Collaborative Work Systems (CWS), Inc
Plant City, FL

Dawn L. Riddle, Ph.D.
Organizational Systems Design, Inc (OSDi)
Plant City, FL

LTC James Merlo, Ph.D.
United States Military Academy
West Point, NY

In augmented reality (AR) environments our senses with which we interact with the real world are selectively supplemented by graphics, sounds, haptics, and/or smell. Outdoor AR is a relatively new topic for research, but it is emerging quickly as technological obstacles are overcome and the utility of such systems is demonstrated in a wide range of contexts. The audiences for outdoor AR systems are those who can benefit from information and experiences presented relative to the surrounding outdoor real-world context, e.g., tourists, historians, accident investigators, builders, industrial processing plant operators, and military troops. In this paper, we describe the design of Geo-Docent, a multimodal AR system that augments a user's real world experience by presenting visual, auditory, and tactile information, relative to the user's geographical position, through an ultramobile computer paired with a Bluetooth headset and a tactile display. Geo-Docent supports content authoring, in a map-based GUI rather than a programmer's environment, in which the AR author geographically tags multimodal content. From designing, developing, and preliminary testing of this system we have learned there is utility in providing the AR author with: (1) a simple, but flexible, interface for defining geographical trigger points and regions, and for linking those triggers with a variety of interface actions; (2) an interface that optionally allows the author to utilize the author's current location and view of the natural world when creating AR content; and (3) support for authoring content in a variety of modalities.

INTRODUCTION

Augmented Reality

On the Virtuality Continuum (Milgram and Kishino, 1994), shown in Figure 1, where the mix of interaction with a virtual world and the real world varies, Augmented Reality (AR) lies much closer to the real world than the virtual one. AR systems typically superimpose graphics and audio sensory enhancements over a real world environment in real-time. Less often, tactile and/or olfactory information are included in AR systems.

Figure 1. The Virtuality Continuum

The ultimate goal is to make such systems light, mobile, and intuitive to use by, for instance, having the display look like a normal pair of glasses, where informative graphics appear in the user's field of view, synchronized with complementary audio, both based on the user's location and head movements. Such systems could be used indoors (where, for instance, a surgeon might be able to select an X-ray view mapped to the current field of view) or outdoors (where, for instance, military and emergency services could be presented with geo-triggered instructions, views of what lies hidden behind obstacles, maps showing friendly and enemy locations, or maneuver graphics).

Head-Mounted Displays

Head-Mounted Displays (HMDs) for viewing graphics and text created by AR systems are a core AR research topic, but they are not sufficiently developed so that the user can simply wear what looks like a normal pair of glasses and have a robust and effective display in a variety of outdoor conditions.

There are two basic types of HMDs: video-see-through and optical-see-through (Bonsor, 2001). Video-see-through displays block out the wearer's surrounding environment, using small video cameras attached to the

outside of the googles to capture images. On the inside of the display, the video image is played in real-time and the graphics are superimposed on the video. Unfortunately, these systems are susceptible to lag, where there is a delay in image-adjustment as the viewer's head moves. Optical-see-through involves projecting the computer generated image through a partially reflective mirror so that the real world view is seen directly. However, they cannot currently display mutual occlusion correctly and the synthetic objects appear as ghosts floating in front of the real image.

Smartphone and Ultramobile Displays

While we wait for HMDs to become reliable, affordable, and socially acceptable (by looking more like a regular pair of glasses than protective goggles), common mobile computing devices that contain cameras and visual displays offer a relatively inexpensive and reliable alternative, even if they do require more cognitive effort to coordinate their displayed view with the user's.

Today, Smartphone and ultramobile computers, with built-in cameras and GPS devices, are relatively common, and inexpensive. When these features are not built-in they can usually easily be added through a cable or Bluetooth interface. Accelerometers and electronic compasses are also more readily available and can be integrated in AR systems through serial port connections to help determine location and bearing of users moving in and out of satellite views and at pedestrian speeds (where the bearing information from a GPS device becomes particularly unreliable).

Wikitude AR

One example of a relatively mainstream commercial application that utilizes a mobile Smartphone with a GPS, compass, and accelerometer is the Wikitude AR travel guide (http://www.mobilizy.com/wikitude.php). The application, produced by Mobilizy, was developed using Google's phone platform Android, an open mobile phone software stack. Wikitude AR is a mobile travel guide that draws on articles from Wikitravel, a user-written guide modeled on Wikipedia. Figure 2 shows two screen captures from the program. In the first, the search interface is demonstrated, where it can be seen that the user can search for landmarks in the surrounding area and have them shown on a map, list, or in the camera view. The second screen capture in Figure 2 demonstrates the map view, and Figure 3 demonstrates the camera view, with the application running on T-Mobile's G1 Smartphone.

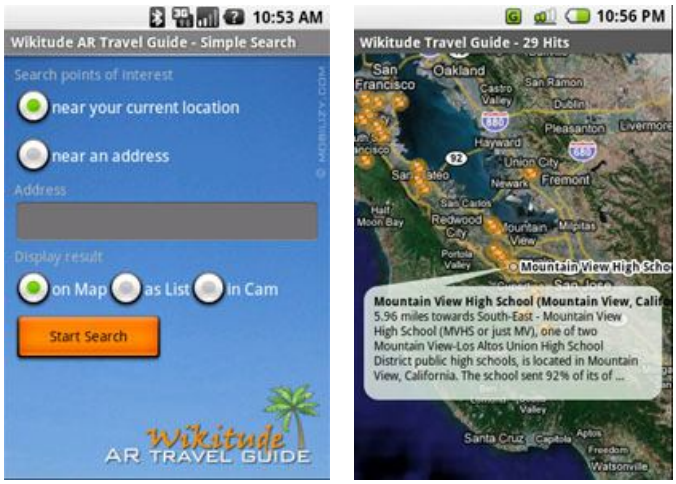

Figure 2. Wikitude's Search interface and "on Map" view

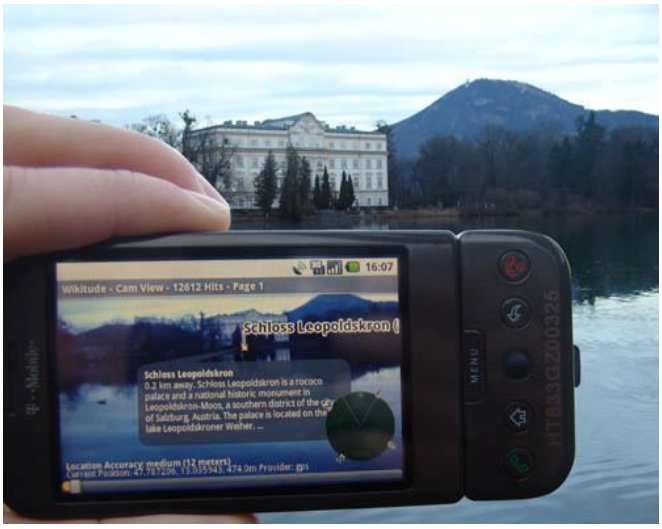

Figure 3. Wikitude's "in Cam" view

Wikitude demonstrates the practicality of a mainstream commercial outdoor AR application, but it does not include support for tailored content authoring by non-software developers. Further, the only modality used to present information to the user is visual. This is typical of outdoor AR applications and therefore there appears to be an opportunity to advance outdoor AR research and development by supporting these features.

PRACTICE INNOVATION

Geo-Docent

The primary goals in developing Geo-Docent were to create an outdoor multimodal AR application that would be operated by the user on an ultramobile computer and which would support (1) authoring of geographically triggered multimodal information by non-software developers, and (2) multiple communication modalities, including tactile. At the time of writing there has been less emphasis on actually integrating the information presented with the user's view through a camera.

Geo-Docent supports geographically triggered graphical, auditory, and tactile information presentation to the AR user through an ultramobile computer, with a built-in camera, paired with a Bluetooth headset and, optionally, a tactile display. At present, this information is preloaded on the ultramobile device rather than retrieving it from a network dynamically, although support for network utilization is currently under development (Chapman and Merlo, 2008). As the user moves through the real-world a log is generated capturing movements over time and interactions with the system to facilitate later review.

Trigger-Action Rules

Figure 4 demonstrates the AR author's map view in Geo-Docent. Direct annotation of a map to create "trigger regions" is illustrated where a map of the United States Military Academy at West Point, New York, has been annotated with elliptical and polygon regions. After the regions are drawn the user then completes the 'trigger' authoring by specifying the required relative location of the user for the trigger to execute an associated action. The available conditions are region type dependent, but for the closed areas shown in this example the conditions currently supported are "on proximity" – approaching the region, "on entry" – crossing into the region, and "on exit" – leaving the region.

For each geographically triggered condition one or more actions is specified, such as: show an image, play a video, play an audio file, synthesize speech, open a file using whatever is the appropriate operating system registered application, execute a program or even activate a tactile display with this set of parameters (tactor, frequency-gain-duration pattern) to, for instance, direct the wearer along a prescribed route or to a particular way point.

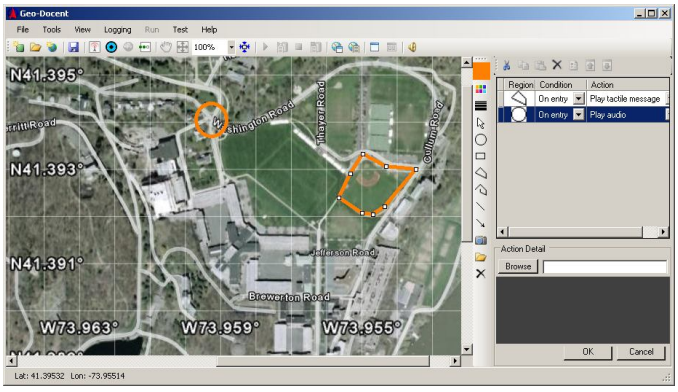

Figure 4 Authoring Geo-Docent content in map view mode

Authoring While in the Real-World

Figure 5 demonstrates the AR author's camera view in Geo-Docent. While standing at a location with a view of the world for which content is to be created the user can take a picture and then utilize the stylus of the ultramobile to annotate the image in a very natural way, pointing to various parts of the image, drawing, and speaking at the same time to efficiently and unambiguously create the geo-tagged multimodal content. Geo-Docent does not currently support creation of real-time voice over video through the camera, but this will be relatively simple to implement and provide the author with additional flexibility.

Figure 5 Authoring Geo-Docent content while at the respective geographic location

Multimodal Augmentations to Reality

Geo-Docent supports geo-triggered presentation of (1) visual information that is either static (text, pictures, and graphics) or dynamic (video and dynamic playback of author recorded annotation over a static image as shown in Figure 5), although neither is yet directly integrated with the camera view; (2) auditory information (through speech synthesis or playing of a previously recorded audio file); and (3) tactile information through a Tactile Display Unit. The non-visual modalities are particularly useful when a HMD is not being used in an AR system because they do not require coordinating multiple visual views related to the current location. They further facilitate hands free movement in the real world. Users are more familiar with auditory communications from technologies than tactile, and natural languages are richer than tactile, but tactile displays can provide a useful alternative when the auditory modality is occupied. Tactile displays have also been proven effective for navigational tasks (Veen et al., 2005) and effective navigation is clearly a requirement for outdoor AR systems. Further, tactile displays have been shown to be complementary with auditory and visual communications for some tasks so that when a communication is presented in both modalities it is perceived more quickly and accurately than if a single modality were used alone (Ernst and Banks, 2002). Tactile displays have also been shown to be effective in "cueing" users to attend to visual or auditory information (Ferris & Sarter, 2008) and have shown to be successful under high physiological stress (Merlo, Stafford, Gilson, & Hancock, 2006)..

Figure 6 demonstrates how eight tactors placed around the waist (the four cardinal directions (North, West, etc.) and the four intra-cardinal (Northeast, Southwest, etc.) can be used to provide directional guidance when a tactor display is combined with a GPS device, accelerometer, and an electronic compass. The figure also presents an example notational representation for how other information could be encoded in a tactile display. In this case it is assumed the augmented reality system can be updated with real-time information about the location of other entities (e.g. soldiers in Figure 6) and this is then translated to as an intuitive tactile display (i.e. patterns of tactor-frequency-gain-duration) as possible for the intended user.

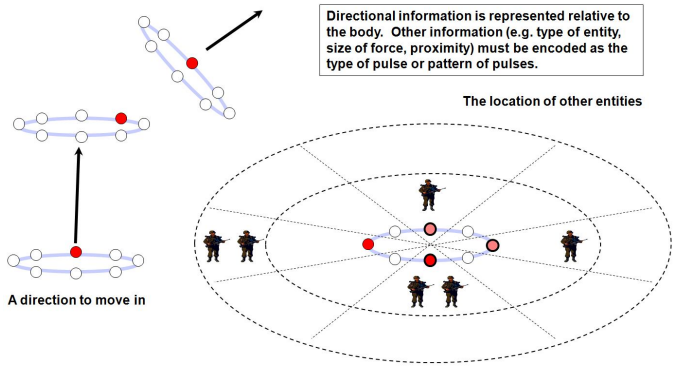

Figure 6 Augmenting reality with spatial information through a tactile display

Geo-Docent supports tactile display design through a GUI that allows the author to define each tactile message as a particular output to a belt containing eight tactors. Once defined the message is available in the drop-down list of actions for the trigger-action pair shown in Figure 4. At present Geo-Docent data is not updated dynamically over a network, so tactile actions are limited to navigational and other preplanned messages, but when the networking features are implemented Geo-Docent will be a multiuser application optionally supporting user location awareness.

FINDINGS

From designing, developing, and conducting preliminary testing of this system, we have learned there is utility in providing the AR author with capabilities we have not seen in other AR systems:

(1) a simple, but flexible, interface for defining geographical trigger points and regions, and for linking those triggers with a variety of interface actions.

Such an interface provides flexibility in specifying the trigger regions so they can not only be based on what is within a certain range of a landmark, but can be tied to any shaped region. If the author has access to terrain or other detailed data for a region, this may be useful to take into account what will actually be viewable and/or accessible. Flexibility in the action types facilitates user adaption of Geo-Docent to a variety of applications.

(2) an interface that optionally allows the author to utilize the author's current location and view of the natural world when creating AR content.

When it is important to support spontaneous development of the content in the real-world, or the real-world context is needed to cue the author, it is particularly valuable to provide a means for authoring in the real-world itself, and in doing so there is an opportunity to automate the geo-tagging of that content. Further, while the full context of the real-world can help provide useful referents, the author may have time constraints, meaning that authoring must either be done quickly while in the full context or it must be postponed. Research has shown that creating multimodal communications by synthesizing speech with pointing and drawing over an image is more effective and efficient than text based annotation of an image or only speech annotation of an image, suggesting that support for multimodal annotation of real-world captured content for an AR system is valuable (Daly-Jones et al., 1997; Faraday & Sutcliffe, 1997; Wasinger, 2006).

(3) the ability to author flexibility in the modality used to present information to the user.

As previously mentioned the author may be in a situation where one modality is preferred over another, due to the user's current activities, or the environment. Further, appropriate orchestration of multiple modalities for communicating information has repeatedly been shown to be more effective than one modality alone when the context will accommodate this (Merlo, Gilson, Hancock, 2008) .

DISCUSSION

The focus of this research and development effort has been on the interface for AR authors and supporting the creation of multimodal augmented reality content. To further develop Geo-Docent itself as a complete and viable AR system more work is needed on the user interface. In particular, user preferences need to be supported, facilitating modality selection just as a regular cell phone may be put in a quiet, vibrate, or normal mode, and more integration between the augmenting content and the camera view is needed. Also, while the trend is for ultramobile computers to become less expensive, they are still more expensive than a Smartphone and therefore a Smartphone version would be more affordable. Given Geo-Docent was developed using Microsoft Visual Studio 2008, the first version of such a system would most easily be developed for Smartphones running Microsoft Pocket PC.

Linking multiple Geo-Docent users over a network will facilitate providing awareness of where the other users are, but it will also provide an opportunity to share authored AR content among users. Thus, for instance, a user may be able to view and/or listen to the comments another user has made at a particular location upon visiting it. While the content may range from nothing more than virtual graffiti to very informative information, it will nevertheless open the door to explore collaborative work and collaborative learning among distributed users in an AR world.

REFERENCES

Bonsor, K. (2001). *How Augmented Reality Will Work*. 19 February 2001. HowStuffWorks.com. http://computer.howstuffworks.com/augmented-reality.htm. Retrieved 4th February 2009.

Chapman, R. and Merlo, J.L. (2008). Geo-Docent: A system for authoring and utilizing geographically triggered multimodal information. 16th Annual ARL/USMA Technical Symposium in Atlantic City, NJ.

Daly-Jones, O., Monk, A., Frohlich, D.M., Geelhoed E. & Loughran S. (1997). Multimodal messages: The pen and voice opportunity. *Interacting with Computers*, 9, 1-25.

Ernst, M.O. & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415* (6870), 429-433.

Faraday, P. M. & Sutcliffe, A. G. (1997). Designing effective multimedia presentations. *Proceedings of CHI '97*, 272-279.

Ferris, T. K., & Sarter, N. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors.* Vol 50(1) 17-26.

Merlo, J., Gilson, R., Hancock, P.A. (2008) Cross-modal congruency benefits for tactile and visual signaling. *Proceedings of the Human Factors and Ergonomics Society*, **52**, 1297-1301.

Merlo, J.L., Stafford, S.C., Gilson, R.D., & Hancock, P.A. (2006). The effects of physiological stress on tactile communication. Paper presented at the 50th Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, CA, October.

Milgram, P., & Kishino, A. F. (1994). *Taxonomy of Mixed Reality Visual Displays.* IEICE Transactions on Information and Systems, E77-D(12), pp. 1321-1329.

Veen, E. J., Jansen, C., & Dobbin, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception* 2(2): 106-117.

Wasinger, R. (2006). *Multimodal interaction with mobile devices: fusing a broad spectrum of modality combinations.* IOS Press.