Current issues in handheld augmented reality

Conference Paper · June 2012

DOI: 10.1109/ICCITechnol.2012.6285844

4 authors, including:

Stan A Kurkovsky
Central Connecticut State University

Ranjana Koshy
University of Bridgeport

Vivian Novak
Central Connecticut State University

Current Issues in Handheld Augmented Reality

Stan Kurkovsky, Ranjana Koshy, Vivian Novak, Peter Szul
Central Connecticut State University
New Britain, CT, USA
kurkovskysta@ccsu.edu

Abstract—**Equipped with powerful processors, cameras for capturing still images and video, and a range of sensors capable of tracking location, orientation and motion of the user, modern smartphones offer a sophisticated platform for implementing handheld augmented reality (AR) applications. Despite the advances in research and development, implementing AR applications for the smartphone platform remains a challenge due to many open problems related to navigation, context-awareness, visualization, usability and interaction design, as well as content creation and sharing. This paper surveys a number of open challenges and issues, trade-offs and possible solutions in implementing handheld AR applications.**

Keywords-augmented reality, mobile computing, navigation, location tracking, usability, content management

I. Introduction

The concept of mixed reality describes systems that combine virtual and real worlds in order to create new visualized environments capable of real-time interaction where digital and physical objects co-exist. Augmented reality (AR), a special case of mixed reality, denotes a real-time representation of the real world that is digitally augmented by adding graphics, sound or video [9]. Handheld augmented reality systems often utilize smartphones that include powerful processors often supplemented with dedicated graphics co-processors, high-resolution cameras, and a range of sensors including Global Positioning System (GPS), accelerometers and magnetometers. Unlike other AR systems, handheld AR applications do not require users to carry or wear any special equipment and do not constrain the applications to any specialized (equipped or previously marked up) area. Handheld AR systems that utilize location and position information are often used to augment the view of the real world with relevant information about the currently visible points of interest (POI).

Using AR applications on smartphones equipped with a camera on the opposite side of the display encourages the use of the 'magic lens' metaphor describing the fact that the users have to point and look 'through' the device to view the augmented representation of the real world. This metaphor imposes a number of ergonomic constraints on the design of handheld AR applications. For example, the device must be held at a certain distance with the camera aimed in the direction of the real-world scene to be augmented. The field of view is limited by the size and resolution of the smartphone's screen and the camera's optical characteristics. Consequently, in order to obtain an augmented view of a broad area, the user should frequently move the device while properly aiming the camera. Although it is relatively easy to steadily move the device while standing, it is much more difficult to do so while walking, which would have a negative impact on the perceived quality of the AR imagery displayed on the screen.

Interaction design challenges described above present just a small sample of issues and open research problems related to navigation, context-awareness, visualization and content creation in handheld AR applications. This paper presents a survey of such problems and proposed solutions, which was prompted by the requirements analysis for a campus guide AR application for smartphones described in Section III. In the process of analyzing possible design solutions that would fit the project requirements we realized that not a single existing handheld AR platform would provide a satisfactory solution the full list of features demanded by our application. This paper outlines many of the open design questions that developers and researchers might encounter building handheld AR applications using currently available technologies and tools.

II. Handheld Augmented Reality Applications

The architecture of a typical handheld AR application generally consists of three components: the *mobile AR browser* for end-user interaction, the *AR server* responsible for identifying and querying one or more *POI provider* (Figure 1). AR browsers typically provide the user with a choice of information channels; upon selecting a channel, the browser sends a query requesting the relevant POIs, which are bounded by the channel selection, current location and a certain spatial range. The AR server acts as a broker and selects an appropriate POI provider to which the query is forwarded. Similarly, POI content is returned to the mobile AR browser via the server. Finally, the mobile AR browser overlays the POI-related content over a real-time view of the real world.

The task of scene identification to determine the correct location and orientation of the user is fundamental to any AR application and may be implemented on the mobile device, on the AR server, or distributed between the two. Marker-based scene identification techniques rely on previously placed artificial visual tags [2, 8, 25]. Non marker-based scene identification relies on computer vision [5], geopositioning [28] or a combination of these two techniques [13, 20]).

Many ongoing AR research projects focus on building proprietary systems to study different aspects of user interaction, as well as building AR interfaces. A few AR projects offer third-party developers an open set of application programming interfaces (API) to build their own systems or, at least, enable end users and developers populate these systems with their own content. Layar (http://layar.com) is a platform for creating and deploying handheld AR applications. Layar provides a customizable mobile browser along with a remote AR server. However, Layar's API requires developers to run their own web service to host application-specific POI providers and link them to Layar's AR server. Consequently, creating new content for Layar-enabled mobile AR applications is limited mainly to the developers who can run and maintain their own web services and related infrastructure.

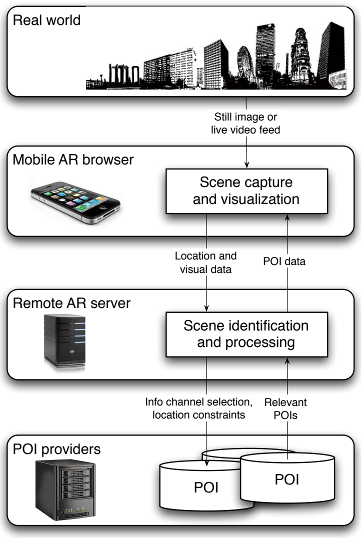

Figure 1. Architecture of a typical handheld AR application.

Wikitude (http://www.wikitude.com/en) provides developers for the Android, iOS and other mobile platforms with a set of client side services to create their own handheld AR applications. Wikitude pioneered the location-based approach in AR applications for smartphones. The platform leverages user-created content that can be provided using the Wikitude web interface, Keyhole Markup Language (KML) or ARML (Section III-B). Junaio (http://www.junaio.com) and Tagwhat (http://www.tagwhat.com) are good examples of user-centric AR applications allowing on-the-go content creation (tagging and uploading) as the users encounter objects of interest in the real world.

III. Issues And Challenges

Handheld tour guides and applications is one of the areas where AR has a strong potential to appeal to the broad user base. Handheld AR technologies, such as Layar and Wikitude, offer developers the technology to create relatively simple location-based AR applications that could be made available and would appeal to a broad spectrum of mass-market users. Handheld AR campus guides have been created by a number of universities in the US including Arizona State University, Indiana State University, Purdue University, University of California, Santa Barbara, University of Wisconsin, Madison, West Virginia University, and others. Such handheld AR applications could be particularly useful during the student orientation when new students are expected to make a transition to the life on an unfamiliar campus and be able to quickly navigate among different buildings that contain a multitude of academic departments and administrative offices.

Following a similar set of goals, we designed and developed an AR campus guide for the smartphone platform. During the requirements engineering phase of the project, we identified the following set of functional requirements. The application must be able to:

- Capture the position and orientation of the user;
- Identify buildings within the line of sight;
- Provide additional information about the POIs;
- Identify different entry points of campus buildings;
- Locate departments and offices in each building.

Additionally, the following non-functional requirements were identified:

- Work both indoors and outdoors;
- Operate on iOS and Android mobile devices;
- Usable while walking (from one building to another).

Once the set of requirements was identified and refined, we attempted to find a suitable existing handheld AR platform to use as the basis for implementing our application. Among the existing handheld AR platforms or ongoing research projects, regardless whether they were domain-specific or generic, we were unable to find any single implementation that would fit the compete set of our requirements.

Among the domain-specific AR systems, those used in the tourism industry are the closest in functionality to a campus navigation application. In many respects, handheld AR is positioned to make a substantial impact on the tourism industry. People visiting unfamiliar places could use handheld AR applications on their smartphones to learn about sightseeing locations, historical events, noteworthy buildings, dining options, etc. Travel guides currently available on mobile devices are often no more than just a specially formatted digital version of the paper guidebook. With the help of AR, these guides could be reimagined as dynamic real-time applications that could bring to life the history of any location [13]. Heritage sites attracted a special attention in the AR community as a fitting domain for deploying specialized handheld applications aimed to facilitate AR-based tours [17]. One of the issues specific to handheld AR deployment at a historical site is the diversity of the tracking targets, as well as the combination of indoor and outdoor environments, which makes it very difficult or even impossible to use a single method of location tracking. This often necessitates the fusion of sensor-based and computer vision-based tracking methods to achieve a robust level of the handheld AR application performance [20].

Careful examination of the existing non-domain-specific AR platforms and applications reveals that none of them provide a suitable set of features to match our set of requirements. The deficiencies of each platform typically represent one or several open research problems, for which there is usually a number alternative solutions, each offering a specific set of trade-offs. These problems can be broken down into three areas: navigation and tracking, content management, and usability (Figure 2). The remainder of this section outlines these open research problems and challenges in these areas, along with different promising ways to address them.

A. Navigation and Tracking

Continuous localization of the user is a key component of any AR system. The vast majority of outdoor handheld AR systems utilize GPS for navigation because of its wide availability and relatively high accuracy. But even outdoors, GPS reception and accuracy can substantially deteriorate in urban environments, where the GPS satellite signal may be shadowed and reflected by the surrounding buildings. Additionally, magnetometers can be affected by local magnetic fields that are not uncommon in urban environments. Indoor navigation for AR systems does not have a similar commonly accepted solution. Satellite signals used by the GPS are too weak or unavailable indoors unless special High Sensitivity GPS (HSGPS) or Ultra-wide Band (UWB) location sensors are used. Furthermore, it has long been recognized that no single sensor technology is currently capable of providing robust tracking with high enough precision both indoors and outdoors [26]. Hightower and Borriello [7] survey a wide range of hardware technologies that could be used for indoor user localization. Deploying a specialized hardware infrastructure could be costly and unfeasible, in which case developers of handheld AR systems may resort to using the sensors already available on the mobile device. Images or video captured with the built-in camera can be processed to recognize the features of indoor environment or previously placed QR (or similar) codes (e.g., [8]). Multiple WiFi signal triangulation could be used for approximate localization [1]. Finally, localization can be achieved by combining sparsely placed 'info points' whose precise location is known, accelerometer and compass data, with activity-based instructions, such as "walk five steps and turn right" [15]. Similarly, indoor navigation in museum tour guide applications can be achieved by asking the user to follow a path formed by a predefined itinerary, while the user orientation and movement are tracked by compass and gyroscope [14]. Gee et al describe an approach where GPS and UWB-based location sensing is combined with vision-based tracking that offers a reliable platform for both indoor and outdoor handheld AR applications [6].

Although a tracking solution based on computer vision could offer the best precision, real-time object recognition from a live video feed may be too taxing for a smartphone CPU. Wither et al propose a compromise solution, Indirect AR, which replaces a true AR based on the live camera feed with a previously captured panoramic view of the environment [27]. A solution suggested by Gammeter at al [5] suggests using a remote server to split the tasks of object tracking and recognition: tracking is performed on the mobile device that periodically sends still images to the server, which is responsible for object recognition. Such an approach could have several advantages: instead of keeping a database on the device, objects can be retrieved from large server-side databases in close to real-time; the bandwidth usage is reasonable since only still images are transmitted to the server instead of a constant video feed. An approach suggested by Takacs et al [23] performs on-device object recognition using a local database of previously captured location-tagged images, which helps to limit the search only to the objects in the close proximity to the user. In case if no match is found, the system offers an option to send the image to the server along with a label describing the relevant POI. It is possible to extend this approach with equipping the server with a larger image database and/or a more robust content-based image retrieval algorithm that would be impractical to implement on a mobile device. Unlike GPS-based tracking, computer vision could offer accurate information about the user location, as well as the pose of the user, with a refresh rate exceeding that of a GPS-based solution. Langlotz et al [11] propose a computer vision-based solution that enables high-precision tracking and object registration without the need to construct a 3D object database. Instead, this approach takes advantage of natural-feature mapping performed on the device that enables tracking with three degrees of freedom. Natural features of the surrounding environment are mapped to the panoramic view captured by the device in real time.

B. Content Management

Many existing handheld AR systems are limited in the way how the new content can be added to them. In most cases, these systems are constrained to a few specialized domains. Adding new content is limited to the application developers and is accessible mainly from the application backend because this often requires programming skills for linking existing systems and data sources together. For an AR system that is truly mobile, it is imperative that regular users, such as tourists and small business owners, are able to add their own content on the go and with a minimal technical effort. Another highly attractive feature would be to provide an easy way for the users to mash up user-created content from multiple sources into a uniform handheld AR view. Belimpasakis et al [3] describe a handheld AR system that addresses these concerns by creating a generic Mixed Reality Web Service Platform enabling users to geo-register new content in this open system without a substantial expertise in AR systems.

Wikitude and Layar allow developers create AR browser applications accessible to a wide spectrum of end users solving the problems of user location tracking and visualization. However, AR applications will not be able to gain much traction with the end users without a broad availability of diverse sources of content. Active user participation in content authoring is leading the evolution of the World Wide Web. A similar trend could be applied to the AR applications. Schmalstieg et al [19] introduce the concept of Social AR, in which regular users can actively participate and create their own content instead of only consuming the content authored by a select group of professional AR modelers and developers. Langlotz et al [10], describe a system for on-the-go, on-device content authoring and sharing. Using this system, end users can create 2D and 3D content on a mobile device and publish it to their private library on a remote server that supports ARML (described below). Users are then free to share this content with others or reuse the objects they created for marking other real world locations.

In an AR application utilizing multiple POI providers, the AR server acts as the only point of interaction between the POIs from different sources. For example, the only connection between a bus stop and a nearby restaurant will be their close proximity, which will only become apparent when the AR server processes both sets of POIs. There is no logical or symbolic relationship between such two POIs, although it could be of a great benefit. A possible solution to this problem could be to utilize the Linked Open Data (LOD) principles [4], which suggest using URIs as names for all data elements, including POIs, as well as cross-referencing among the data elements. Augmented Reality Markup Language (ARML) used by Wikitude provides a native LOD support and it is gaining traction among AR system developers as the Open Geospatial Consortium uniting over 440 international industry, government and academic organizations has established the ARML 2.0 Standards Working Group in September 2011.

One of the key features of AR applications is the ability to present a subset of available information in the current geospatial context. Research in context awareness focuses on creating intelligent systems that can adapt to the surrounding environment and user behavior, thereby reducing information overload and providing the user with relevant services and information. Although AR systems take advantage of the user location and orientation context, it should be possible to offer users a more personalized experience by utilizing other contextual dimensions, such as user intention based on the past behavioral profile ([12], [21]). In addition to improving the level of personalization, context-awareness in handheld AR applications could facilitate sharing of personalized content and social collaboration among the users [22].

C. Usability

Current applications address only the most obvious and simplest challenges that could be solved by handheld AR systems. Typically, the user of a handheld AR application can receive information about an unfamiliar location or surrounding objects, such as nearby buildings or businesses that may or may not be within the line of sight. Nack [16] notes that currently available handheld AR applications take advantage mainly of the user position and orientation, which is contextualized, provided that the correct information channel is available. GPS sensors on current smartphones have the accuracy of about 20 meters, while the magnetometers offer the compass orientation within about 20 degrees. This could lead to problems with calculating the correct field of view of the camera making real and digital objects not perfectly aligned against each other. Consequently, current mobile AR systems may not offer the precision necessary to identify the specific location of the entrance door or even distinguish between different entrances to a building.

Although the manufacturers equip smartphones with high-resolution cameras, they provide a limited field of view that is significantly smaller than that of the human eye. Consequently, current handheld AR applications can only augment a small portion of the mobile user's field of view. The 'magic lens' design of the current handheld AR applications requires the user to stretch out their hand while holding the device and pointing it in various directions. Tokusho and Feiner [24] point out that this is the mobile AR user's equivalent of the texting stance of the Blackberry user. This problem could possibly be resolved by "freezing" the augmented view to allow the user to see it in a more comfortable position.

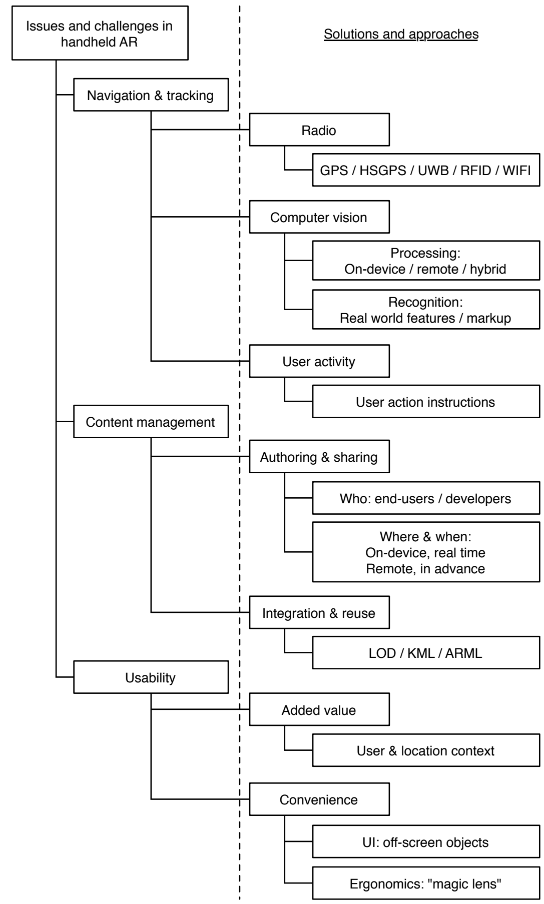

Figure 2. Challenges and their solutions in handheld AR.

Representing off-screen objects offers a set of related challenges to the developers of handheld AR applications. In order to see the augmented view of real-world objects that are currently to the sides or behind the user, the user needs to either change their orientation or use a mini-map showing all nearby POIs that is typically displayed on the screen by many current handheld AR applications. Having to rotate around while holding the phone in an outstretched hand may be rather awkward, while interpreting the POIs on the mini-map and matching them to the augmented view and the rest of the unfamiliar real-world surroundings might require a substantial mental effort. Schinke et al [18] suggests using arrows embedded in the AR view to point at the surrounding off-screen POIs, which can make the task of interpreting such information much less demanding for the user.

IV. CONCLUSION

Although there is a large number of ongoing research projects and off-the-shelf solutions for taking advantage of smartphones as a platform for handheld AR, many of them limit themselves to leveraging only the user geographic location and orientation information. Today, smartphones are already quite capable of providing tracking services using computer vision algorithms, fusing different methods of location tracking for robust indoor and outdoor navigation, providing tools for easy on-device content creation and sharing, leveraging user and location context, taking advantage of heterogeneous sources of POI and other data, and providing more unobtrusive user interaction than what is currently offered by the existing handheld AR applications.

Each of the three areas of handheld AR solutions identified in Figure 2 presents its own set of challenges. Based on the body of surveyed work, it also appears that there is no single 'magic bullet' solution or approach in any of these areas. Achieving a reasonable navigation performance both indoors and outdoors seems to be especially difficult due to current technological limitations. Existing and proposed radio-based solutions achieve a good performance in either of the environments, but not in both due to their inherent technological constraints. Using computer vision may seem a reasonable alternative for user tracking and navigation by recognizing existing POIs and the features of the environment. However, such a solution may not be as appealing as radio-based navigation because it requires having pre-existing location data either on device or elsewhere. Depending on the location of this data, such a solution would require either a substantial computational effort on device or a constant availability of a low latency broadband connection to a remote server running the computer vision algorithms.

The issues related to content management seem closely related to the problem of where POI data is processed. The key to achieving an AR solution that would allow its users easily share and reuse POI information is using an open standard such as ARML and supporting the LOD principles. The issues of usability seem to be the least explored in the current body of work, possibly because it is imperative to first achieve a functional technological AR solution, which then can be made more practical by studying the usage patterns, user preferences, and interaction modalities. It also appears that the 'magic lens' metaphor that influences the ergonomics of most handheld AR applications is here to stay due to the inherent hardware characteristics of existing mobile devices.

REFERENCES

[1] C. Arth, D. Wagner, M. Klopschitz, A. Irschara, D. Schmalstieg, "Wide area localization on mobile phones," International Symp. on Mixed and Augmented Reality, 2009.

[2] B. Avery, B. Thomas, W. Piekarski, "User evaluation of see-through vision for mobile outdoor augmented reality," 7th IEEE/ACM Int'l Symp. on Mixed and Augmented Reality, pp. 69-72, 2008.

[3] P. Belimpasakis, Y. You, P. Selonen, "Enabling rapid creation of content for consumption in mobile augmented reality," In 4th Int'l Conf. on Next Generation Mobile Applications, Services and Technologies, 2010.

[4] T. Berners-Lee, "Linked data," Retrieved from http://www.w3.org/DesignIssues/LinkedData.

[5] S. Gammeter, A. Gassmann, L. Bossard, T. Quack, L. Van Gool, "Server-side object recognition and client-side object tracking for mobile augmented reality," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[6] A. Gee, M. Webb, J. Escamilla-Ambrosio, W. Mayol-Cuevas, A. Calway, "A topometric system for wide area augmented reality." Computers & Graphics, vol. 35, no. 4, pp. 854–868, 2011.

[7] J. Hightower, G. Borriello, "Location systems for ubiquitous computing," Computer, vol. 34, no. 8, pp. 57-66, 2001.

[8] T. Kan, C. Teng, W. Chou, "Applying QR code in augmented reality applications," 8th Int'l Conf. on Virtual Reality Continuum and its Applications in Industry, 2008.

[9] D. van Krevenlen, R. Poelman, "A survey of augmented reality technology, applications and limitations," The Int'l Journal of Virtual Reality, vol. 9, no 2, pp. 1-20, 2010.

[10] T. Langlotz, S. Mooslechner, S., Zollmann, C. Degendorfer, G. Reitmayr, D. Schmalstieg, "Sketching up the world: in situ authoring for mobile augmented reality," Personal and Ubiquitous Computing, , 2011.

[11] T. Langlotz, D. Wagner, A. Mulloni, D. Schmalstieg, "Online creation of panoramic augmented reality annotations on mobile phones," Pervasive Computing, 2010.

[12] W. Lee, W., Woo, "Exploiting context-awareness in augmented reality applications," Int'l Symp. on Ubiquitous Virtual Reality, pp. 51-54, 2008.

[13] D. Marimon, C. Sarasua, P. Carrasco, R. Álvarez, J. Montesa, T. Adamek, I. Romero, M. Ortega, P. Gascó, "MobiAR: tourist experiences through mobile augmented reality," Networked and Electronic Media Summit, 2010.

[14] F. Mata, C. Claramunt, A. Juarez, "An experimental virtual museum based on augmented reality and navigation," 19th ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems, 2011.

[15] A. Mulloni, H. Seichter, D. Schmalstieg, "Handheld augmented reality indoor navigation with activity-based instructions." 13th Int'l Conf. on Human-Computer Interaction with Mobile Devices and Services, 2011.

[16] F. Nack, "Add to the real," IEEE Multimedia, vol. 17, no. 1, pp. 4-7, 2010.

[17] Z. Noh, M. Sunar, Z Pan, "A review on augmented reality for virtual heritage system," 4th Int'l Conf. on E-Learning and Games: Learning by Playing. Game-based Education System Design and Development, pp. 50-61, 2009.

[18] T. Schinke, N. Henze, S. Boll, "Visualization of off-screen objects in mobile augmented reality," 12th Int'l Conf. on Human-Computer Interaction with Mobile Devices and Services, 2010.

[19] D. Schmalstieg, T. Langlotz, M. Billinghurst, "Augmented reality 2.0," In: Coquillart. S., Brunnett. G., Welch. G. (eds), Virtual Realities, pp. 13–37, 2010.

[20] D. Seo, K. Kim, J. Park, "A tracking framework for augmented reality tours on cultural heritage sites," 9th ACM SIGGRAPH Conf. on Virtual-Reality Continuum and its Applications in Industry, pp. 169-174, 2010.

[21] C. Shin, W. Lee, Y. Suh, H. Yoon, Y. Lee, W. Woo, "CAMAR 2.0: future direction of context-aware mobile augmented reality," Int'l Symp. on Ubiquitous Virtual Reality, pp.21-24, 2009.

[22] Y. Suh, Y., Park, H. Yoon, Y. Chang, W. Woo, "Context-aware mobile AR system for personalization, selective sharing, and interaction of contents in ubiquitous computing environments," 12th Int'l Conf. on Human-Computer Interaction, 2007.

[23] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, B. Girod, "Outdoors augmented reality on mobile phone using loxel-based visual feature organization," 1st ACM Int'l Conf. on Multimedia Information Retrieval, pp. 427-434, 2008.

[24] Y. Tokusho, S. Feiner, "Prototyping an outdoor mobile augmented reality street view application," 8th IEEE Int'l Symp. on Mixed and Augmented Reality, 2009.

[25] D. Wagner, T. Langlotz, D. Schmalstieg, "Robust and unobtrusive marker tracking on mobile phones," 7th IEEE and ACM Int'l Symp. on Mixed and Augmented Reality, 2008.

[26] G. Welch, E. Foxlin, "Motion tracking: no silver bullet, but a respectable arsenal," IEEE Computer Graphics and Applications, vol. 22, no. 6, pp. 24–38, 2002.

[27] J. Wither, Y. Tsai, R. Azuma, "Mobile augmented reality: indirect augmented reality," Computers & Graphics, vol. 35, no. 4, pp. 810-822, 2011.

[28] T. You, T. Chin, J. Lim, J., Chevallet, C. Coutrix, L. Nigay, "Deploying and evaluating a mixed reality mobile treasure hunt: Snap2Play," 10th Int'l Conf. on Human Computer Interaction with Mobile Devices and Services, pp. 335-338, 2008.