The future directions of mobile augmented reality applications

Abstract

Mobile augmented reality is currently moving into the mainstream market. Both information retrieval, and crowd sourcing applications are rapidly being adopted. This paper describes the main mobile augmented reality applications on the current market and which techniques they use. This paper clusters new techniques suited for mobile augmented reality and evaluates which ones will be adopted by the market.

0.1 Introduction

Augmented Reality (AR) is the augmentation of the visual field of a user by enhancing the current field of vision with additional information Caudell and Mizell [1992].

A mobile AR systems (MARS) is defined as a system which Papagiannakis et al. [2008]:

- Combines real and virtual objects in a real environment
- Runs in real time and mobile mode
- Registers(aligns) real and virtual objects with each other
- The virtual augmentation is based on dynamic, 3-D objects (e.g. interative, deformable virtual characters)

These systems very in display type, reality, immersion and directness of interaction. These differences influence their place in the mixed reality continuum Milgram [1995]. All applications discussed in this paper will use monitor based displays. Most applications will use (video) images of the real world as their substrate and show the world from a egocentric perspective. The applications will differ in reproduction fidelity, extent of presence and directness of interaction.

0.2 Motivation

The recent rise of smartphones[1] has sparked the development of multiple MARS.

The market for MARS has been growing rapidly, with an estimated size of more then 400 million downloads in 2014[2].

The quick evolution of these applications warrants a overview of the current products in the market and an overview of recent technologies that have not yet made the transition from science to market.

In order to provide this overview a selection of very popular and innovative applications is provided. Some of these are not classical AR applications. However, all the applications present an augmented version of the real world. Google Goggles, Google Navigation, Layar and Sekai Camera have been chosen because their popularity in app stores on the 7th of July 2010. Acrossair, Junaio, Toozla, Waze and Wikitude Drive have been selected due to their innovativeness.

The techniques used in these applications will be evaluated. Afterwards, techniques from selected papers and the relevant papers from the proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR) 2006-2009 will be discussed.

[1] http://communities-dominate.blogs.com/brands/2010/07/an-8-segment-model-to-analyze-smartphone-market-consumers-and-html

[2] http://juniperresearch.com/viewpressrelease.php?pr=176

0.3 Research Question

With a focus on mobile augmented reality using a mobile phone, the research questions follow logically:

> *What are the future directions of mobile augmented reality applications?*

As discussed in the previous section, there are two specific subquestions that need to be answered:

Market What current MAR's are successfull;

Techniques What techniques are currently used in MAR's;

Future techniques Which techniques will appear in MAR's in the recent future?

0.4 Current Applications

A range of applications will be discussed. Afterwards, we will present a overview of the technologies used in these applications.

0.4.1 Acrossair

Acrossair is a company that produces iPhone applications. In July of 2009 they launched the first variation of their augmented reality browser, the new york nearest subway applications[3]. A mockup of this application can be seen in figure 1.

A second application is a (paid) browser to view the nearest wikipedia content, NearestWiki, which can be seen in 2. Creating a paid app that enables users to view (creative commons) wikipedia content is considered harmfull by some reviewers[4].

The Acrossair AR browser is also used as a frontend for third party data sources. In contrast to most other AR browsers, Acrossair does not allow third parties to add their own point of interest sets.

Acrossair is a iPhone only application, but it is very up to date. The latest release of their software includes support for the iPhone 4G's gyroscope and the iAd platform[5]. The gyroscope gives a better experience indoors then the compass which most other applications use.

Figure 1: Acrossair nearest subway

Figure 2: Nearest wikipedia entry

0.4.2 Google Goggles

Google Goggles[6] enables users to search the web by picture. At the moment it is most suited for recognizing landmarks and other inanimate objects. However, it also has the ability to recognize text and read miscellaneous tags like QR codes or barcodes.

[3] http://www.acrossair.com/apps_newyorknearestsubway.htm
[4] http://itunes.apple.com/us/app/nearestwiki/id331305547?mt=8
[5] http://www.readwriteweb.com/archives/iphone_4_gyroscope_brings_silky_smooth_augmented_r.php
[6] http://www.google.com/mobile/goggles/#text

After the picture is taken, the picture is transmissed to Google's servers to be analysed. The results are shown on the mobile phone. Depending on the detected relevant options are shown. For example, the translation of text might be offered depending on your current location and home language.[7]

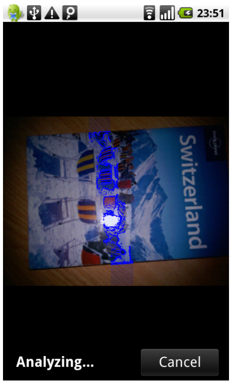

Figure 3: Photo processing

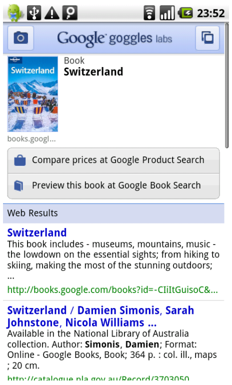

Figure 4: Goggles search results

0.4.3 Google Navigation

Google Navigation[8] is a navigation product that uses aerial photos (figure 7) and street view (figure 8) photo's to display a route. The information displayed is influeced by compass direction, location (GPS) and social traffic information. Destinations can be dictated and it includes voice guidance.

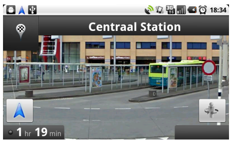

Figure 5: Arrival location in real world photo

Figure 6: Overview of the next turn

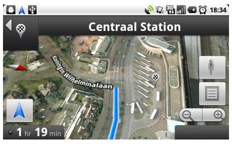

Figure 7: Arrival location in an aerial photo

Figure 8: Direction overlay

0.4.4 Junaio

Junaio is a mobile augmented reality browser that is focused on interactivity. It can provide live context-relevant information, such as traffic information and the location of the nearest train station and arrival times (figures 9 and 10). Users are able to give recommendations on shops and restaurants and are able to leave digital tags.

Junaio currently has a layer that gives live information on the Bay Area Rapid Transport (BART) network in San Francisco.

When used indoors it can recognise LLA tags/markers[9][10]. Junaio's database stores that tags position. These tags combined with the compass enable indoor AR (figure 12).

Junaio can also recognize other markers. Junaio Glue[11][12] is a technique that recognizes a 2d tag and it's orientation and adds a layer with a 3d model (figure 11).

[7] http://googlemobile.blogspot.com/2010/05/translate-real-world-with-google.html
[8] http://www.google.com/mobile/navigation/
[9] http://www.junaio.com/publisher/llamarker
[10] http://blog.converget.com/2010/03/14/new-junaio-2-0-worlds-first-indooroutdoor-ar-platform/
[11] http://www.junaio.com/publisher/junaioglue
[12] http://gamesalfresco.com/2010/07/07/5-things-to-do-with-junaio-glue-and-lla-markers/

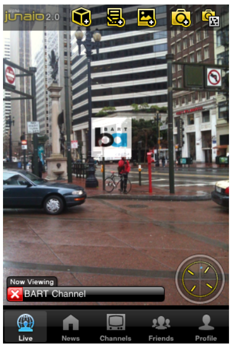

Figure 9: BART station location

Figure 10: BART arrival time

Figure 11: GLUE 3d model

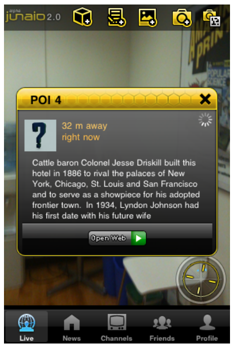

Figure 12: LLA tag used for positioning

0.4.5 Layar

Layar is an AR browser developed by a dutch company. The AR view in layer is based on the metaphore of an extra layer of data. Both user-provided as sponsored layers are available. Sponsored layers might appear higher in the list of suggested or popular layers. Layar also includes a store in which users can buy access to a layer with information.

Marker pointing in the direction of the nearest points of interest are overlaid on the camera images. Markers can include extra information, like pictures or videos.

Examples of the usage of layar are sponsored layers with the location of casino's (figure 15), properties that are for sale or the location of nearby ATM's. Sponsored layers might include objects that allow the users to create funny snapshots. This has been done to promote avatar and for the "ronald gone mad" campaigne (figures 13 and 14)

Other uses for layar include (paid) city tours [13][14] and a layar with user-provided information on the rolling stones. The information on the rolling stones label is provided by fans that are using a facebook application [15].

Figure 13: Layar commercial for Avatar

Figure 14: Ronald gone mad commercial

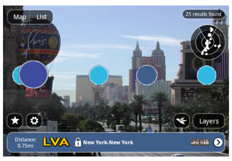

Figure 15: Layer with Las Vegas casino's

0.4.6 Sekai Camera

Sekai Camera is a augmented reality from Tonchidot, which was introduced at TechCrunch50 in 2008[16]. After the initial buzz, Tonchidot received $4 million in a first round of venture capital [17].

[13] http://site.layar.com/paid-layer-launching-partners/
[14] http://www.nai.nl/uar
[15] http://site.layar.com/company/blog/layer-of-the-week-rolling-stones/
[16] http://techcrunch.com/2008/09/17/tonchidot-madness-the-video/
[17] http://blogs.wsj.com/venturecapital/2009/12/08/augmented-reality-start-up-tonchidot-moves-beyond-buzz/

Initially Sekai Camera was the most popular iphone application the Japanse app store. Later on they deployed their application worldwide and released an android application.

Tonchidot allows users to place "AirTags" in the real world. Additional information can be added to a tag, which can be viewed later on (figure 18). The client also views and posts geolocated tweets (see figure 16). Sekai Camera has two models for partners to advertise. Recently Tonchidot has been experimenting with indoor markers (figure 17).

Sekai Camera eX enables publishers to display static information to visitors. The target group for this application are visitors at shops/malls/museums or visitors at outdoor entertainment parks.

The second solution, the "Open Air API" is most suited to publishers that want to display up to date, numerous, geotagged data that is often updated [18].

Another usage for the Sekai Camera platform are augmented reality games. The first (Japanse only game) that was launched is Sekai Hero[19][20]. Sekai hero is a online augmented reality role playing game in the world of Sekai. A screenshot of Sekai Hero is in figure 19.

The combination of information from partners and tags from users, Sekai camera strives to deliver a very interactive environment for their users.

Figure 16: Cloud of Sekai tags on a street

Figure 17: Indoor Sekai tags

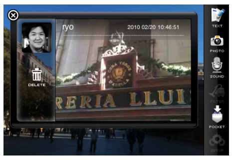

Figure 18: Viewing an AirTag

Figure 19: Sekai Hero

0.4.7 Toozla

Toozla is global audio guide[21]. Toozla provides multiple location-based streams of information. These streams include local chat, tourism information, the weather and stories on nearby objects from wikipedia.

Users can also record their own messages and share these with others.

0.4.8 Waze

Waze is a social mobile application developed by Waze Mobile providing free turn-by-turn navigation using crowdsourced traffic information[22]. The data in Waze is provided by it's users.

Waze extracts a roadmap from the GPS logs of from travel patterns. These maps are then annotated with extra data (for example names) by it's users. Waze's source is publically available[23].

While driving Waze shows a map with additional crowdsourced information. This information can be traffic information but also events like accidents, speed cameras or police patrols (see figure 20).

0.4.9 Wikitude Drive

Mobilizy is a company that released one of the earlier AR browsers. Because of it's lack of popularity, it was excluded from this overview. However, their second product, Wikitude Drive is more interesting. After announcing it in september 2009, recently wikitude opened the beta of their navigation product.

Wikitude Drive is a navigation product that overlays the route on live camera images, providing a more natural driving experience. The user is able to switch between a 3d view of the route (figure 22) and an augmented road view (figure 23).

[18] http://openair.sekaicamera.com/about-api
[19] http://sekaiyuusya.jp/pc/
[20] http://support.sekaicamera.com/
[21] http://www.toozla.com/
[22] http://world.waze.com/
[23] http://www.waze.com/wiki/index.php/Source_code

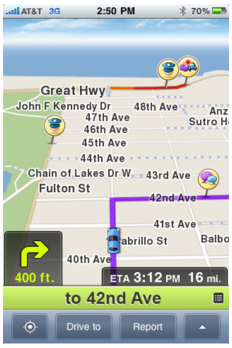

Figure 20: Waze navigation screen

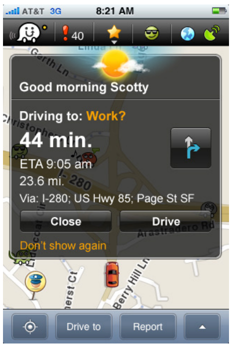

Figure 21: Welcome screen

Figure 22: 3d map view

Figure 23: Lane overlay

0.5 Current Techniques

In the previous section a selection of MAR's was discussed. Table 1 section will give an overview of the techniques used in these applications. All applications share the usage of GPS and a data connection, thus this is not noted in the table.

	Camera	Compass	Accelerometer	Push data	Crowd sourced data	Markers	OCR	Gyroscope	Microphone	Speaker
Acrossair	✓	✓	✓					✓		
Google Goggles	✓	✓				✓[24]	✓			
Google Navigation		✓		✓	✓[25]				✓	✓
Junaio	✓	✓	✓		✓[26]	✓[27]				
Layar	✓	✓	✓							
Sekai Camera	✓	✓	✓	✓	✓[28]	✓				
Toozla					✓				✓	✓
Waze	✓	✓	✓	✓	✓					
Wikitude Drive	✓	✓	✓							

Table 1: Techniques used in the MAR's

As you can see, there is a large spread of techniques used in these applications. Google Goggles (image recognition), Waze (social sourced traffic information) and Toozla (augmented reality audio) stand out.

0.6 New Techniques

In order to give a overview of feasible new techniques, a selection has been made among the papers from the ISMAR 2006-2009 and selected papers have been added. The techniques that are discussed are clustered per similar subject.

0.6.1 Augmented Reality Audio

At the moment most MAR applications do not use audio. High quality audio adds to the immersion of a augmented reality application. [Hrm, 2004] researched the requirements that augmented reality audio needs to fulfill in order to have the wanted effect.

0.6.2 Seethrough

With the usage of additional cameras, augmented reality techniques can give the illusion of seeing moving objects through occluding surfaces. Both a source (near the user) and reference camera are needed. [Barnum et al., 2009] proposes a technique that is suited for ad hoc sharing of images in a dynamic network of camera nodes.

0.6.3 Object recognition, tracking and pose estimation

In order to recognize and track shapes and poses, high contrast and precise shapes are needed. [Hagbi et al., 2009] describe a object recognition and pose estimation system that is feasible on current generation mobile devices. The library size causes a problem, both for confusion rate and matching speed.

[Waechter et al., 2009] investigate the possibility of using camera's to track indoor positions of users. They suggest that indoor cameras could be used to provide a localisation interface, but warn about the privacy issues it might introduce.

[Simon, 2007] proposes a new method to automatically detect and reconstruct walls in AR tasks. Walls can be detected and the performance is feasible on current mobile telephones. As a side product the position of walls is detected.

0.6.4 Positioning

Ubiquitous Tracking uses multiple sensors to increase the robustness and accuracy of position tracking. The dynamic combination of mobile and stationary trackers creates new wide-area tracking setups.

[Pustka and Klinker, 2008] introduces the concept of Spatial Relationship Graphs. The additional sensors could increase the precision of indoor position tracking and increase the relative accuracy overall.

0.6.5 Mobile applications

[Gunnarsson et al., 2006] use a MAR to visualize wireless sensor node data by overlaying the measurements on camera images. The orientation of the camera is detected with the help of markers.

An analysis of the delay and client-side processing required for server-side processing of camera images has been made by [Chen et al., 2009]. They show that the latency caused by the combination of local processing, network latency and server-side processing is under one second and should still provide a good user experience.

Marker tracking is a heavy operation. [Wagner et al., 2009] investigates the performance of various algorithms. In the end it is shown that multiple markers can be tracked on current generation telephones, while archieving sufficient performance.

0.6.6 3D/Displays

3D headsets have been used for augmented- and virtual reality since the 1990's. A better 3D headset increases the fidelity of reproduction and immersiveness, moving the application on the mixed reality continuum. Until recently, 3D headsets were very invasive and covered a large piece of the field of vision and face. The size of these headsets has been descreasing and the resolution has been increasing. [Rolland, 2005] suggest that 3D headsets are almost ready for primetime deployment because they have become a lot better and less invasive.

Most current MAR's their overlays provide inconsistent lighting. [Yeoh and Zhou, 2009] present a technique for rendering realisting shadows of virtual objects in a mixed reality environment by recovering the lighting positions by analysing the shadows of a known object.

0.6.7 Surface projection

[Gupta, 2007] explore the integration of projected imagery with a physical book. The book itself is markerless. In the end they create a "universal" book.

0.6.8 Interaction

A 3d object that is positioned near a marker can not always be directly controlled. [Seichter et al., 2009] propose additional markers near the main marker that control the 3d object.

[Benford, 2005] describe the mismatch between the movements that users perform, those that can be measured by a computer and the interactions needed to control an application. They provide an approach to evaluate potential problems with an gesture interface.

The combination of mobile-phone based input methods with a displayed image is explored by Hong et al. [2006]. Even though the usage of the mobile telephone input methods is feasible, they suggest the usage of other input methods when interacting with 3d objects.

[Thomas, 2007] evaluate three different modes of cursor manipulation as input methods for a user operated MAR. Their results show that a combination of a head cursor with a vision based gesture interface would provide the best interface.

0.6.9 Interacting with the world

[Suh, 2007] describes a method to control nearby devices in a ubiquitous computing environment with a augmented reality device.

0.7 Conclusions

At the moment two groups of MAR's are successfull. One group is focused on retrieving needed information easily. The second group has a focus on interactions and towards crowd sourcing.

Layar is very good at retrieving data, and while not technically excellent it has very broad support across platforms.

Sekai is a good example of the second group. The platform allows for interactivity and is one of the first augmented reality games.

Another group of applications is upcoming. These are the applications that use crowdsourcing for data acquisition and use the power of the crowd. Route planning applications are a good example of this.

Most applications share a common set of functionality. At the moment only acrossair uses a gyroscope. We assume this is because of hardware availability, since at the moment only the iPhone 4G has a gyroscope. A other exception is push data, which is relatively hard to implement at scale, but which is needed for real time interactions. Furthermore voice recognition and OCR are relatively rare. We assume this is because of the marginal gain of it's applications, and the server-side processing power needed to process this data at scale.

The main technique to appear in future MAR's will be object recognition. Some applications have already hit the market and proven to be feasible. The ability to reuse this technique to provide absolute indoor positioning is a nice bonus.

The second improvement will be among the display techniques. 3D headsets or image projection are ready for the main market and applications using them will be more immersive and allow more natural (gesture) interfaces to be used.

In the end augmented reality applications will become easier to use and more immersive. Augmented reality will become a easy, widely used information retrieval and interaction technique.

0.8 References

Barnum, P. et al. *Dynamic seethroughs: Synthesizing hidden views of moving objects.* In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 111 –114. 2009.

Benford, S., S.H.K.B.A.R.G.C.R.T.G.J.G.A.P.T.G.B.B.A.W.B.P.S.S.A.G.H.S.A. *Expected, sensed, and desired: A framework for designing sensing-based interaction. ACM Transactions on Computer-Human Interaction*, 12(1):3–30, 2005.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-21044437342&partnerID=40&md5=f90f966e93df3231e04c3037a94dc660

Caudell, T. and Mizell, D. *Augmented reality: An application of heads-up display technology to manual manufacturing processes.* In *Proceedings of the Hawaii International Conference on System Sciences*, volume 25, pages 659–659. IEEE INSTITUTE OF ELECTRICAL AND ELECTRONICS, 1992.

Chen, D. et al. *Streaming mobile augmented reality on mobile phones.* In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 181–182. Citeseer, 2009.

Gunnarsson, A. et al. *Visualization of sensor data using mobile phone augmented reality.* In *Proceedings of the 2006 Fifth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'06)-Volume 00*, pages 233–234. IEEE Computer Society, 2006.

Gupta, S., J.C. *The universal media book: Tracking and augmenting moving surfaces with projected information.* pages 177–180. 2007.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-45149098843&partnerID=40&md5=b7902aa5c7dcf07c730f288c32030f05

Hagbi, N. et al. *Shape recognition and pose estimation for mobile augmented reality.* In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 65 –71. 2009.

Hong, Y. et al. *Mobile pointing and input system using active marker.* In *IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006*, pages 237–238. 2006.

Hrm, A., J.J.T.M.K.M.L.T.H.J.L.G. *Augmented reality audio for mobile and wearable appliances. AES: Journal of the Audio Engineering Society*, 52(6):618–639, 2004.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-4344645826&partnerID=40&md5=0669099b83cc2cd049231f578de827ec

Milgram, Paul, T.H.U.A.K.F. *Augmented reality: a class of displays on the reality-virtuality continuum.* volume 2351, pages 282–292. 1995.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-0029211386&partnerID=40&md5=5b462c4a3302c15337dc75ecf522b105

Papagiannakis, G., Singh, G. and Magnenat-Thalmann, N. *A survey of mobile and wireless technologies for augmented reality systems. Computer Animation and Virtual Worlds*, 19(1):3, 2008.

Pustka, D. and Klinker, G. *Dynamic gyroscope fusion in Ubiquitous Tracking environments.* In *Proceedings of the 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality-Volume 00*, pages 13–20. IEEE Computer Society, 2008.

Rolland, J.P., B.F.H.L.F.H.Y.M.R. *Development of head-mounted projection displays for distributed, collaborative, augmented reality applications. Presence: Teleoperators and Virtual Environments*, 14(5):528–549, 2005.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-33645750066&partnerID=40&md5=68eea0186bd7cea397b17ea404928e9d

Seichter, H. et al. *Multitouch interaction for Tangible User Interfaces.* In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 213 –214. 2009.

Simon, G. *Automatic online walls detection for immediate use in AR tasks.* pages 39–42. 2007.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-45149127927&partnerID=40&md5=30d5bee7987478cc1a575bb670ca6010

Suh, Y., P.Y.Y.H.C.Y.W.N. *Context-aware mobile AR system for personalization, selective sharing, and interaction of contents in ubiquitous computing environments. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4551 LNCS(PART 2):966–974, 2007.
URL http://www.scopus.com/inward/record.url?eid=2-s2.0-38149082253&partnerID=40&md5=aa9610e2753fc440a54cb128bf1fa511

Thomas, B. *Evaluation of three input techniques for selection and annotation of physical objects through an augmented reality view.* pages 33–36. 2007.
URL `http://www.scopus.com/inward/record.url?eid=2-s2.0-45149101369&partnerID=40&md5=f92972c26cbaa6e7ce20902975adddfd`

Waechter, C., Pustka, D. and Klinker, G. *Vision based people tracking for ubiquitous Augmented Reality applications.* In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 221–222. IEEE Computer Society, 2009.

Wagner, D., Schmalstieg, D. and Bischof, H. *Multiple target detection and tracking with guaranteed framerates on mobile phones. ISMAR09*, 2009.

Yeoh, R. and Zhou, S. *Consistent real-time lighting for virtual objects in augmented reality.* In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 223–224. IEEE Computer Society, 2009.