Mobile Augmented Reality Use Cases for Standards Development Uses

Abstract
The purpose of this position paper is to provide greater clarity for the AR Use Case categories and use cases proposed for discussion by the AR Standards Community. The three use case categories use specific language. We recommend that Standards Development Organizations that wish to focus on AR scenarios consider adoption of language/concepts made available in this AR use case document.

This document is published by the author for community discussion and is provided without restriction or condition as a contribution to the AR activities of the Web3D Consortium AR WG, the OGC ARML 2.0 SWG, OMA and other groups in ISO, ITU and elsewhere.

We conclude that work in standards organizations could easily result in misaligned outcomes due to each SDO beginning with and focusing on a different use case. Our recommendation is that activities begin with one or at most two use cases selected from those proposed below but consider possible implications of standards on the three use case categories.

What is a Use Case Category?
A category of use cases is composed of a “core use case” and many possible permutations. The category description includes the most simplified use case and short descriptions that show how the “core use case” can be adapted to multiple vertical markets and may be extended to support new features.

All Use Cases have several attributes in common
In the description of the AR use case, there does not need to be a detailed (or any) specification of the user’s device. The reason is that the standard that will result from the use case must be sufficiently broad to be used from any device with the sensors and processing ability to complete the tasks requested of it by the user. In other words, it doesn’t matter if the user has a tablet, a smartphone, a heads-up display with connection to a local processing unit. With the exception of standards that concern themselves directly with the connectivity and communications between the devices and information “sources,” the Internet where the processing is done in the cloud, or any other scenario for connectivity, there should not be any reference to where and how much data is stored for the completion of the use case.

Further, the initial step of opening a specific application in order to perform a task is superfluous. At this point, unless the SDO is defining application interfaces such as the SC29 MPEG-A group, the specification of an application (dedicated or general, such as a web browser), is a red herring and should not be included in the description of the use case.
The Guide Use Case Category
The simplest and most fundamental use case is one in which the real world provides a user interface for a person who is asking simple questions, most often a series of questions that lead to the successful learning, completion of a task or arrival at a destination.

The emphasis of a “guide” is on a point of interest and usually a process (step by step sequence). In the Guide, the user turns on the device, points it at a target (or in a direction) and queries the system to select the question that the user wants to answer or address.

Sample Guide Use Case-Geo-position
The user is of any age, gender or cultural background. The user is new to a particular location. In the scenario the purpose of the user’s arriving at the site is to explore cultural heritage. In advance, using the latest technologies, the site’s management has created interactive experiences for mobile AR platforms using historical information in one or more of these formats: text, sounds (speech), images, 3D models and animations.

The user’s geospatial position and the user’s orientation are determined (method is not relevant). Given the position and orientation, the system welcomes the user to the site. The system may prompt the user to determine some user preferences, but this is not desirable. The system can prompt the user to enter demographic data (age, gender, origin) that can be used for selection of preferred tour. In addition, based on preferences, the system can operate only with audio, with audio+video, with 3D models or only with text labels (augmentations). The system only proposes to the users those options that are supported by the user’s device.

The system also prompts the user to determine the level of interest in different domains (e.g., agriculture, architecture, lifestyle, commerce, entertainment, etc). This information can be organized in small chunks (atomic size?) for the user to explore to the level that they wish. Here the user could use the “more” or “next” button.

Once configured, the system proposes a starting point on the site. Assumption here is that there are unique "tours" that each has its own starting point. The user is guided to the starting point using arrows on the screen or audio instructions.

The user’s arrival at the starting point is detected. The user is prompted to start the tour. A context-sensitive menu should be available as the user move near points of interest, to prompt the user. The user then follows directions on the screen (defined by a story teller/architect of the tour) and at any point can interact with information assets provided as part of the experience (see above regarding the "next" and "more" note above).

By pointing at the building (or where a building was), a reconstruction of the original site (or the building) appears. A photo overlay is one option to bring up the visuals. If the user has selected that they wish to see 3D models, they will see 3D. The 3D model remains registered to the site as the user position changes (tracking the building staying in view relative to the original building as long as the orientation indicates the user’s attention is on the same point of interest).
The distance of the user from the object of interest needs to be considered in how to design the interaction between the user and the data. Expressing the content as compact representations is useful here. This shows different types of information aggregation for AR authors.

The user’s orientation changes. Now it is focused on a monument (e.g. statue). This causes a label to appear with information about the date of the monument’s erection, the artist who made it and there is a short animation showing the artist at work (re-enactment of history).

The user wants to be in a photo with the statue. This is done using the user-facing camera. The composition results with the statue, the user and the fictional rendering of the artist leaning on the other side of the statue. The user sends this photo montage to friends via a social network. This portion of the use case can be extended to include capturing a video of the user, location and digital element.

Upon conclusion of the exploration of the site (but this can be available at any time), the user is prompted to produce a comment or to rate a particular component. The user’s rating becomes an augmentation that can then be viewed at another point (e.g., by the curator of the cultural heritage site). The important point is that the user can create and attach an augmentation.

**Sample Guide Use Case-Computer Vision**

The user can be of any age, any gender or cultural background, in any location at any time of day or year. The focus of the actor/user’s attention is a medium to large piece of machinery (a coffee maker, a washing machine for laundry, a office photocopier, a modern engine) that has continuous power supply, a panel or area on which there are LED indicators of status, an internal computing system and many complex moving parts. The user does not have a manual for configuration, use or repair. The machine’s manufacturer has developed an AR manual for diagnostic and repair. The AR-assisted manual has 3D and planar recognition algorithms and contains step-by-step instructions.

The device is not performing the task it should be and requires user intervention as indicated by a red light on the panel that is blinking at 1 second intervals. The user orients the device that has a front facing camera, computational resources, communications capabilities and a touch sensitive display, at the machine.

The user communicates with software on the device or in the cloud that they wish to search for an interactive, digital user manual about this machine. The user case does not describe whether the menu is multi-touch, speech controlled or any other details concerning the user interaction for starting, and communicating desires prior to the use of AR.

The user receives an indication from its device that it is searching for a manual but has not determined which manual since it does not know the model of machine the user wishes to receive/use. The software prompts the user to approach the object of interest.

When pointed at the machine, the device’s camera sends video of unknown resolution at an undetermined frame rate to a processing system. The system analyzes the video stream, recognizes the point of interest and highlights it with a
visual overlay. It can prompt the user by way of speech output, displaying messages on the screen (text or arrows) that the user must change positions with respect to the machine.

When the machine is recognized, the manual is prepared for viewing/interaction (could be local or in the cloud) and the user receives a signal that the first step of diagnostics is ready to begin. The user gives the device a signal (touch, speech, or a gesture) to begin.

A series of diagnostic tests are given to the user to perform. [The use case description can go into more detail here.] Each step is communicated by way of an on-screen (or optional auditory) indicator of successful completion of a step. The next step is then shown and the instructions given. The user gives a signal to the system by way of touching the screen, a gesture or an auditory input (speech).

At the end, the light (LED) turns green and is no longer blinking.

**Create Use Case Category**

The Create use case category involves the real world, a user's actions with a device. The user’s objective in the Create use case is to express an opinion, provide additional thoughts or questions, in other words to “author” something in the form of text, image, video or audio recording and to attach this original (user generated) information to a real world target.

In the Create use case category, the real world target (the Point of Interest of the user) can be anything that the device can sense repeatedly. An example of a Point of Interest that cannot be, in this category, considered a real world target is something ephemeral, such as a wave on the surface of water or a cloud.

In the Create use case category, all conditions with respect to viewing (“consuming”) the annotation are equally included. In other words, defining who can view the augmentation (only the author, only those designated by the author, anyone/public) can be part of a use case within the use case category.

Specific use cases within the broader Create use case category can also embellish aspects of security, curation/review, and distribution of the content. A use case within the Create use case category case specify where the platform for acknowledging the user's actions is based (local, in the cloud, or both) and the parameters related to the transmission, storage and processing of the user-generated augmentation.

**Play Use Case Category**

The Play use case category encompasses all use cases in which there is the real world, devices, digital content (augmentations) and two or more users interacting with one another in real time. In the Play use case category, there is no apriori limit to where the users are in the world with respect to one another.

A specific Play use case can specify the distance between the users (proximity) in meters. Other Play user cases can specify the categories of objects that constitute the focus of attention. For example, there are use cases in Play category involving
manufacturing, repair, maintenance of machinery, infrastructure or some stationary, man-made object.

Play use cases are not limited to games. There are use cases in the Play category that can describe a military scenario in which two or more people target a common person, place or object and then see what the other sees.

**Summary of the proposal**

The first and most generic use case category for AR is called “The Guide.” It covers all scenarios in which the user is consuming content through an AR view with the real world as trigger, with and without network connectivity.

The second level of use case category for AR involves the user being a consumer of AR experiences as well as publishing content that is “attached” or anchored to the real world where the user designates, and can be viewed by the user at a later time or by others to whom the user grants permission, in AR view.

The third use case category is the most complex, involving the real world, digital data synchronized with triggers in the real world, and two or more people.
Appendix A

Comments on OMA Touristic Use Case (see next page)

The OMA Tourist Use Case is a very specific scenario that is well contained in the “Guide” use case category. The Guide use case simply covers all scenarios in which the user receives step-by-step information as the situation around them changes.

This use case is detailed, however, suffers from a number of important weaknesses.

1. The use case specifies that the local tourism office has published the “AR application.” Does this mean that the application only provides information about local cultural heritage sights in AR view? Or does it also provide these data in map view? If, as would be logical and reach a larger audience, the tourism office has published an application that includes Points of Interest and these can be viewed in AR “mode,” then the scenario is that of a person running a tourism application in AR mode, which together with the real world, produces an AR “experience.”

2. Why does the user begin the application only upon entering the center of the city (third paragraph)? A user has reached the city center by some means (pedestrian, public transport, etc) and could have used AR all along.

3. In the scenario, the user begins by permitting the application to obtain/use positioning information. Then, the user “sets preferences”. This can also be called “filtering” the information that’s available in AR view.

4. The user seeks to store data. Why not take a photo or a video for this purpose (memory)? The purpose of this section may be to suggest that applications that foster a connection between the user and a place have the data capture (use of camera for capturing physical environment) as part of the same application. If the camera is being used for AR view, then using it also in capture mode requires using both a camera capture API and a camera “view reality as background” mode.

5. The user points the device at a building and receives content but it is not clear from the description if the AR content the user sees is ever tied to an object that is recognized by the camera, or if it is entirely the result of the device determining the user’s location and orientation.

6. Use of AR markers as a term for something shown on the screen is not consistent with industry terminology

7. How do we define a tourist? Is it a question of how far the person has traveled or if they have never been to a destination of interest? The tourist can as easily be a person who is in their home town or city but has never visited a particular neighborhood. A person who seeks help to replace the cartridge on their printer.

8. The use case does not distinguish between the content that is shown in the user’s field of view as a result of a public
B.1 Touristic Use Case
This use case describes the usage of the AR technology in a touristic scenario where the user is walking around
the city and enjoying its beauty and cultural aspects.

B.1.1 Short Description
Isabel, a tourist, is visiting an ancient city. The Tourism Office, in accordance with the Service Provider, offers a
new AR application to find touristic information (and much more) in an easy and entertaining new way. In
particular, a 3D reconstruction of the city ruins is available.
Isabel loves to know everything about cultural places, so she used to carry several guides and make use of audio-
guides. This time Isabel has downloaded on her mobile the AR application, so she doesn’t have to carry heavy
guides, as all cultural content and information available is easily accessible from her mobile.
When walking in the city centre, she starts the AR application to discover monuments and Points Of Interest
(POI) in her proximity, as she doesn’t want to miss any of them. The application makes use of Isabel’s position in
order to retrieve geolocalized content and information.
The application allows Isabel to set preferences (e.g. search radius, ...) and select the type of content she wants to
consume (e.g. category of POIs, ...). Based on her preferences, or the choice made on the move, geolocalized
multimedia contents are shown on the screen of her mobile.
Excited by the charm and beauty of the ancient city augmented by cultural content and information that AR
application provides, Isabel wants to store these valuable views to review later. With a simple touch, Isabel
successfully records what she is watching on her mobile screen.
She is in front of a very original building and she frames it with her mobile. The scene is augmented with a mash-
up of information and content about that building (e.g. POI name, its description, audio guide, related UGC
content...); indeed her mobile screen displays AR Markers to highlight the availability of AR Content. AR
Markers provide links to additional information sources and content related to that place. Isabel finds out that,
among such content, the application also offers premium content: a 3D reconstruction of the ancient building. She
decides to get it and starts interacting and virtually navigating the 3D model.
After that, Isabel decides to enjoy the beautiful scenery of this original building without outside interference, so
she switches her mobile from AR view mode to camera-view mode to watch the building directly. For the
convenience of following operations, this change does not need Isabel to shut down AR application.

B.1.2 Market benefits
This use case shows some benefits for User and Content Provider. The User can consume easily touristic
information and entertaining content in an attractive way and in a single point of access: maps, basic information,
multimedia content are all available through the AR application. Then the User is enabled to see their surrounding
environment overlapped by enriched cultural multimedia content through the use of their mobile and intensify the
user involvement by providing an interactive service.
The Content Providers (e.g. POI owners, ...) are enabled to offer their content and information at the right
moment, then their information is easily retrieved and consumed, allowing new business model.
Appendix B

Comments on OMA Gaming Use Case

The Gaming use case proposed is very similar to the guide use case (the user follows a trajectory while playing with a digital object) except that the user’s motivations are different.

1. It is not necessary to specify for the purpose of the AR Standardization, that the park has partnered with another company, or for that matter, to specify the provider of the experiences.
2. There is no need and it is limiting to have a use case in which the user must (to begin) download a piece of software. The use case should be sufficiently defined that the same could be implemented with a Web browser in the future.
3. The use case refers to solving puzzles and riddles but does not say the exact role of AR in the scenarios. It appears that computer vision is used because there is a reference to a camera view and AR markers.
4. The purpose of the user’s earnings (redeemed for a discount on the purchase of a park item) is not necessary or relevant to the use case for AR standardization. And, redeeming points could be performed using many technologies.
5. Including the game characters in the photos taken while in the park involves insertion of digital objects into a still image. This is included in the proposed “guide” use cases.

B.2 Gaming Use Case

The use case described here focus on the rich experience offered by AR technology to the user actively involved in a Gaming scenario.

B.2.1 Short Description

Harry, an ardent gamer, is on visit to an adventurous and popular park Disney Park with its special effects and features. The Park, in conjunction with the local Service Provider, sets up an AR gaming application version that offers amazing entertainment experience to the user. Harry downloads this AR app on his mobile to have fun in the park.

Through this application, Harry navigates his way through the park by following a map or trajectory that lets him know where the magical creatures live. Of course, this app is magical – as he moves past landmarks in the park the app tells him where to go next by showing on his mobile screen. Harry must solve puzzles and riddles on his way to the next destination. Clues to the answers can be found on the AR markers in the park on his camera view.

At each stage, Harry is shown the relevant content to proceed further in the game to accumulate points or coupons on his way out. These coupons can be redeemed for a discount on the purchase of any item in the park.

Following these directions, Harry may take photos of various landmarks. To his wonder, App’s fantastical friends may appear in the photos – sometimes right next to him.

These photos are stored in a gallery to make his adventure unforgettable. The application typically may use all the features of the mobile to lead the user into a fantasy world of trolls, fairies and tree genies - right in their local park.

B.2.2 Market benefits

There is a win-win situation for both user and content provider in terms of usage and revenues. The user enjoy the gaming experience by downloading the application and consumes the resources (i.e. markers) overlaid on the real-time park environment. He also can utilize the discount coupons accumulated during the game play. The content provider is benefited by inviting more and more users by offering value-added services in their play-zone.
Appendix C

Comments on OMA Authenticated Scenario Use Case

Personalized services are desirable in all technologies and services. Adapting experiences to a user’s preferences can be achieved in many ways.

Likewise, authentication of user identity can be performed in many ways. New systems for authentication will continue to be introduced and evolve as devices are endowed with new sensors and other components.

The authentication of user does not need to (should not) be part of an open AR standard. Nor should explicit descriptions of the authentication interaction be part of a reference model, architecture but it can be part of any solution provider’s approach to ensuring log in by the authorized user.

B.3 Authenticated Scenario Use Case

This use case describes the support to the authentication in order to enable the user to access special or personalized content.

B.3.1 Short Description
Alice is using is preferred AR application while visiting a touristic city. She discovers that special contents are associated to a monument nearby. To access this content she needs to authenticate herself on the application.
By clicking on the AR marker, a window pops up allowing Alice to enter her credentials. After that she can enjoy the multimedia contents that show what the nearby area will look like after its renovation. She discovers also some other special personalized contents that are provided based on her preferences and User Context.

B.3.2 Market benefits
This use case shows some benefits for the actors involved. The user through the authentication can consume a personalized service and access special contents reserved to her.
The Content Providers (e.g. POI owners, ...) can offer contents that reach the right consumers. The Service Provider can tailor the offered service based on its customers and differentiate its service offer.
Appendix D

Comment on OMA Network and Client Use Case

It is not clear where the proposed Network and Client API use case differs from that which is currently available from many proprietary providers of AR platforms.

Using today’s platforms a brand can commission the development of an application that gives the user (with a brand-specified client application on their smartphone) access to brand-specific content in specific locations. The brand probably does not seek or desire authentication or any other obstacles to engagement with its customers.

B.4 Network and Client APIs Use Case
This use case describes how a MobAR Enabler implementation exposes its information and capabilities to third-party applications via (client and network) API.

B.4.1 Short Description
A big fashion brand company wants to provide its customers with an AR application that allows discovering all the shops selling its clothes.

The company makes a deal with the service provider that has deployed the MobAR Enabler and hires a developer to develop an application able to access the exposed MobAR APIs accordingly to the appropriate authentication and authorization mechanisms.

There are several customization scenarios:

1) (Basic) Use of MobAR Network API: the application running in the content provider environment can customize AR contents based on the brands graphics; for example, the Nike company may provide AR contents for advertising Nike shops;

2) (Intermediate) Use of MobAR Client API: the AR application running on the device adds new functionalities to the ones provided by the MobAR Client; no customization is applied to the AR Content; for example, the Nike company may provide a dedicated MobAR application running on the device that customize the user interface (but not providing contents as in the previous case).

3) (Advanced) Use of MobAR Client API and MobAR Network API; the two cases described above apply. The developer can exploit both sets of APIs to implement a stunning AR application.

B.4.2 Market benefits
This use case shows some benefits for User, Service and Content Providers. The User can consume easily AR Content from her preferred clothes brand.

The Content Provider (e.g. brand company) is enabled to offer its own application and contents to its customers. The Service Provider can monetize its deployment of the MobAR Enabler diversifying the users’ portfolio (end users, enterprises, ...).