The Need for an Open, Large Scale, Vision-Based, Mobile Augmented Reality Platform

Anthony Ashbrook\textsuperscript{1}, Geoff Ballinger\textsuperscript{2}

\textit{Mobile Acuity Limited}
Appleton Tower 6.15, 11 Crichton Street, Edinburgh, EH8 9LE, UK
\textsuperscript{1}anthony.ashbrook@mobileacuity.com
\textsuperscript{2}geoff.ballinger@mobileacuity.com

\textit{Abstract}—We describe the advantages of vision-based augmented reality over existing geolocation-based systems, the limitations of current vision-based systems and our vision and development of a truly large scale, vision-based augmented reality platform. Vision-based and geolocation-based approaches are actually complementary and we expect future systems to integrate these.

I. INTRODUCTION

Mobile augmented reality (AR) is a compelling and powerful new way for users to access location and object specific information and content. You simply point your phone’s camera at an object or place and instantly see a view of the real environment with an overlay of relevant digital content.

Mobile AR is part of a broader range of technologies that are enabling the physical world to become interactive and measurable, enabling new applications and business opportunities. This broader range of technologies includes active systems such as NFC and passive technologies such as barcodes and visual search.

Although still in its infancy today, the mobile AR industry is expected to grow rapidly and quickly evolve into a multi-billion dollar industry. Mobile AR technology is being adopted across a large number of application areas including navigation, games, tourism, advertising & marketing, product information, publishing and education.

II. GEOLOCATION-BASED AR

Despite a number of significant limitations, geolocation-based AR applications, which typically use your phones GPS and electronic compass to display geolocated information, have become very popular. The key limitations are:

1. \textbf{Poor Precision and Accuracy}: Poor precision and accuracy of the GPS and electronic compass sensors mean that applications work well for distant objects but are poor for objects that are close by. This limits the range of useful applications, working best for navigation or viewing objects at a distance and where position is not critical. High precision positioning systems are available but typically involve installing specialist infrastructure at a location and non-standard sensors on the mobile device.

2. \textbf{World-centric}: Content is locked to a world-centric, global coordinate system and lacks the flexibility to follow objects limiting the range and scope of applications.

3. \textbf{Non-Ubiquitous}: GPS is not ubiquitously available. It is frequently problematic in built up urban areas and completely unavailable indoors. Other positioning technologies based on data from WiFi access points and cell towers do work indoors and in dense urban environments but provide far less accuracy, typically in the range of 20 to hundreds of meters.

III. VISION-BASED AR

To address these limitations and to provide a much richer experience AR companies are exploring the use of a vision-based approach in which content is precisely attached to natural visual targets rather than being geo-located. Software on the mobile handset recognises and tracks these visual targets and estimates the relative camera position so that content can be rendered into the scene.

Notable vision-based platforms are Qualcomm’s AR SDK and Metaio’s Junaio Glue. Layar have also announced their plans to include image recognition in their platform and have announced a partnership with Kooaba. A number of standalone vision-based mobile AR applications have also been demonstrated over the last few years, typically linked to mobile marketing campaigns. Some of the earliest vision-based mobile AR used artificial visual markers to simplify the recognition and tracking problem.

Vision-based AR has four essential features that enable a much richer AR experience:

1. \textbf{Precision}: The location of the camera can be determined with very high precision (relative to the visual model of the scene). This allows for
precise interaction with objects and locations at close range and allows for a high density of adjacent content.

2. **Object-centric**: The location of the camera is determined with respect to an object, or even multiple objects simultaneously, rather than with respect to a global coordinate system. This allows content to be fixed to and to follow objects and enables a much wider and richer range of applications.

3. **Ubiquitous**: Vision-based systems can be used wherever there are sufficient natural visual targets including places where GPS systems cannot operate such as indoors or in closed urban environments.

4. **Visibility-aware**: Vision-based AR systems only display content if the visual target that the content is associated with is visible. Content therefore disappears if the visual target is obscured from view. This provides a much stronger illusion than a geolocation-based approach which has no knowledge of when objects are in view or not. This attribute can also be a limitation since partial occlusion can cause an object to disappear and destroy the AR experience.

IV. LIMITATIONS OF TODAY’S VISION-BASED AR PLATFORMS

In an ideal vision-based AR system the user would simply point at a visual target (an object or location) and instantly see the associated augmented reality content. Unfortunately today’s vision-based AR platforms fall short on delivering this experience. In the current systems the user either has to install a new application for each AR experience or they have to manually activate a particular AR experience within their existing AR application. This is because today’s vision-based solutions can only match the camera view to a small number of visual targets that are pre-loaded and stored locally on the mobile device.

In summary, today’s vision-based AR suffers from these two competing requirements:

1. To provide a compelling and interactive user experience visual tracking is implemented in real-time on the mobile device using a model of the visual target stored locally.

2. So that a user can instantly interact with any of potentially billions of augmented objects and locations the system must access a vast visual model database. It is clearly not possible or sensible to store such a database on a mobile handset.

V. THE FUTURE OF VISION-BASED AR

Future vision-based augmented reality platforms should be able to match the camera view to potentially billions of visual targets in near real-time and the system should adopt the web model of a standard browser and distributed content.

To create this truly scalable vision-based AR platform there are a number of technical problems still to be addressed but Mobile Acuity and others specialists are developing solutions to these problems.

Our recommendation for a vision-based AR platform has the following attributes:

1. **The Scene Manager**: We use the term “scene” to mean a visual target with which AR content is associated and relative to which the content is displayed. A scene may be an extended visual model such as a room or city or it may be a localised object such as a book. In this system the mobile device communicates with the remote scene manger to determine which scene(s) are visible and where the camera is relative to those visible scene(s). Multiple simultaneous scenes are allowed – for example the local environment may be one scene and some objects in the environment may be separate local scenes – each with their own associated AR content. The scene manager must be able to perform image-based recognition very efficiently and allow content providers to register the scenes they wish to populate with AR content.

2. **ARML (Augmented Reality Mark-up Language)**: To enable mass content creation it must be possible for non-programmers to be able to create content. This will require a standardized mark-up language for building interactive AR content allowing for the specification of audio/visual content, dynamic behaviours, modes of interaction and so on. An important feature of ARML is that is must allow for object-centric as well as geolocated content.

3. **Content Servers**: Content should be distributed and not hosted on central servers to provide flexibility and scale for the creation and management of content. In other words, this should work like the web.

4. **Modularised Technology**: One of the objectives of standardising the architecture of the AR platform is to allow anyone to create their own AR browsers and applications. The core technology should be modularised so that developers can incorporate this into their own applications. This will allow more rapid adoption across platforms.
encourage innovation and allow the technology to be incorporated into existing browsers.

VI. MOBILE ACUITY

Mobile Acuity was incorporated in 2006 by computer vision specialists from the University of Edinburgh in the UK who invented the concept of visual tagging is a system they call “Spellbinder”. Visual tagging allows you to take a picture of an object or place and then leave digital content at that location for other people to discover and access.

The team at Mobile Acuity are now developing a demonstrator for the world’s first large-scale, vision-based, mobile augmented reality platform. This builds upon the company’s existing mobile visual search technology. The key components of the system are:

1. The system utilises Mobile Acuity’s large scale image recognition platform that can match a query image to a database of millions of candidate images in real-time (For example, the figures below present our visual music discovery service provided through our partnership with leading digital media delivery company 7digital).

2. The system uses Mobile Acuity’s technology to extend image-based recognition to fixed shape 3D objects such as buildings.

3. Mobile Acuity is developing a very efficient mechanism for the mobile client and image recognition server to communicate and interoperate over a mobile network.

VII. CONCLUSIONS AND RECOMMENDATIONS

Geolocation-based mobile AR has rapidly grown in popularity due to the mass market availability of position and compass sensors on smartphones, the simplicity of the approach and the availability of existing and newly created geolocated content. Vision-based AR, however, provides a number of significant benefits over geolocation-based AR offering the possibility of ubiquitous, high precision AR experiences with the flexibly for both fixed locations and for following objects.

A particular problem that needs to be addressed to maximise the benefit of this approach is to be able to recognise and track any of potential billions of objects or locations (“scenes”) in real-time.

We have three primary recommendations:

1. **Technology**: Vision-based AR technology has reached a level of maturity where real applications are now viable but further progress is required from the technologists:

   a. Solutions are needed for the real-time recognition and tracking of billions of potential objects

   b. Techniques are needed to deal with sudden occlusions and very distant visual targets. Fusion of vision-based data with location and inertial data may provide good solutions.

2. **Standards**: Developing AR standards must include provision for a vision-based as well as geolocation-based approach.

3. **Applications**: Applications designers and developers should establish the best practice for combining both geolocation-based and vision-based approaches to create the ideal user experience.

Mobile Acuity is developing a real-time system capable of recognizing and tracking potentially billions of visual targets. This client-server technology will be available for licensing so that the AR community can incorporate this capability into their own applications.