Mobile Augmented Reality: The Need for Visual Recognition

Version 1.0

Anthony Ashbrook
Geoff Ballinger

8th February 2010

Visual recognition algorithms from the computer vision community have a key part to play in unlocking the full potential of the new breed of mass market Mobile Augmented Reality applications that are now being produced.

Overview

Augmented reality (AR) is the provision of a digital environment that is spatially aligned with a real environment. The digital environment may be a visual one that is merged with the real one through a heads-up display or increasing on the screen of a mobile device overlaid onto the camera view. The digital environment may also be interactive such that the user can interact with the digital elements. The traditionalist view of AR is of a visual *and* interactive digital environment typically viewed through a heads-up display [1]. A more abstract but equally valid definition of AR includes a purely non-visual interactive-only environment where objects or places to be "clickable", a good example being mobile applications that can provide product information by reading a barcode, RFID tag or identifying the product through visual recognition.

The importance of AR is that by mapping the digital environment to the real one an intuitive and powerful new model of human-computer interaction is created. It is analogous to the change from a text based operating system like DOS to a graphical user interface such as Windows. Both of these systems could support similar applications but the new model of interaction provided by the interactive desktop transformed the way that people used computers. AR represents a similar quantum leap.

A key technical aspect of AR systems is how the digital content is spatially aligned with the world. In the early days of AR, and more recently in popular mobile AR applications, this has been done using the location and orientation of the viewer. This makes it very easy to present the relative location of digital objects that have been assigned a position relative to the real world. Historically the AR community quickly found that, although this was sufficient for points of interest that are fixed and some distance from the viewer, the approach is unsatisfactory for objects that can move or are close by. To address these limitations the AR community embraced computer vision technology as a complementary way to perform spatial alignment. Initially this was based upon markers – the best know example today is the AR toolkit – and more recently using more sophisticated object recognition technology.

The first breed of mobile AR applications are getting a lot of attention right now as they provide a very compelling use for a Smartphone. These applications are largely based upon position data and ultimately suffer from the limitations identified by the AR community. To unlock the full potential of mobile AR these new services must embrace computer vision technology. Position data and visual data are in fact very complimentary. Position data is well suited to objects and places in the distance that are not necessarily visible but where errors and jitter in the position measurements are not problematic. Visual recognition is well suited to objects in the viewer's immediate vicinity where precision is critical to support the illusion of spatial alignment.

AR Primer

There is a wealth of material available covering all aspects of AR so we present a small number of select references for the beginner.

One of the leading companies in the AR technology space is Total Immersion [2]. This is a great place to go for examples of AR. For the developers out there one of the standard SDK's used in AR is the AR Toolkit [3], versions of which have been ported to some mobile devices. For the latest research in this field the annual ISMAR conference is a great source [4].

Similar technology is also used extensively in film post-production to mix digital assets with live action footage. Traditionalist would not class this as AR since it is not a live, real-time application but there is a great deal of overlap in these problems. Boujou [5] is the leading post-production tool used for this purpose.

References

[1] http://en.wikipedia.org/wiki/Augmented_reality

[2] http://www.t-immersion.com/

[3] http://www.hitl.washington.edu/artoolkit/

[4] http://www.ismar09.org/

[5] http://www.vicon.com/boujou/