# AR phone:
# Accessible Augmented Reality in the Intelligent Environment

Mark Assad, David J. Carmichael, Daniel Cutting, Adam Hudson
*University of Sydney*
{*massad, dcarmich, dcutting, ahudson*}*@it.usyd.edu.au*

## Abstract

*This paper introduces the concept of the AR phone, a mobile phone acting as an augmented reality interface into an intelligent environment. We present scenarios to illustrate this concept and describe the design and implementation of a prototype system whereby a phone can capture images, transfer them to an intelligent environment and receive some form of visually augmented result.*

## 1. Introduction

Intelligent environments (IEs) aim to complement users' daily experiences by providing them with computational resources and services that blend seamlessly into their surroundings. One of the most important aspects of such an environment is how users interact with it. This should be as intuitive and unobtrusive as possible, which makes traditional computing interfaces such as keyboards, mice and monitors undesirable. Techniques that allow a user to interact with an IE in a more natural manner are preferable.

Augmented reality (AR) is one technology that offers a lot of promise in this area. This is the concept of augmenting physical reality with additional information such as sound, text or three dimensional graphics.

Much work into AR is concerned with three dimensional graphics and current AR interfaces of this sort are typically head mounted displays (HMDs) or web cams attached to PCs. Both of these have significant drawbacks. HMDs are expensive and thus inaccessible to ordinary users. Web cams are non-intuitive since the input device is usually quite separate to the display device. The disparity of affect and effect in such an interface can make coordination quite difficult.

Additionally, much of the hardware currently used in AR either tethers users to workstations or is not generally portable enough to be carried or worn for long periods. It would be desirable to have an intuitive, accessible interface that users can carry at all times. Ideally, they should not even need a specialised device to interact with an IE: it should be something they already carry.

A mobile phone has the potential to perform this role. In particular, the emerging generation of phones with colour screens, cameras and wireless networking have all of the capabilities required to implement an AR interface and are therefore plausible platforms for deploying AR-based applications. Mobile phones are highly accessible, as seen by the vast number of people who own and carry them today. Even the latest high-end phones are affordable, given away cheaply or even for free as part of long term contracts.

This paper presents the concept of an AR phone, an accessible and intuitive system built around a mobile phone which can be used to present an AR interface into an intelligent environment. We discuss some similar projects, explore some scenarios that will be possible once such a system is ubiquitous and discuss the implementation of a prototype system.

## 2. Background

The metaphor of seeing through an object to modify or augment the view behind is familiar. For example, a magnifying glass appears to make things larger. It can be manipulated naturally by simply moving the lens over target objects. By designing a device that functions in a similar fashion, we can use this metaphor to provide an intuitive interface. The metaphor of a 'magic viewport' has been recognised in previous work such as the AR-PDA (2).

The AR-PDA project is in many ways similar to the AR phone. The ultimate goal is to help consumers in their daily lives with tasks such as inserting a memory card into a digital camera. All AR processing is offloaded to remote servers over wireless network links. The project appears to be designed to provide relatively static information such as training manuals and the client-server approach appears to have arisen due to technical limitations on the mobile devices. The client-server approach is also applicable to the AR phone, although for reasons of increasing interaction with the IE as well as reducing the processing load on the phone.

The AR toolkit (5) is commonly used to rapidly produce AR applications. Developed at the HIT Labs at the University of Washington, it is based on the concept of fiducial marker tracking and augmentation. This approach allows the determination of the relative transformation from the camera to the marker and as such requires no special infrastructure apart from a camera and a flat surface able to display a recognisable pattern. As we wish to use only commodity components for the user interface, and no current mobile phones have location or orientation sensors, it is appropriate to use the AR toolkit in our system.

The Handheld AR project (4) from Vienna University of Technology uses a PDA running the AR toolkit. Although the PDA can operate without external servers, it does use wireless technologies if available to offload the AR processing to a server. In this case however, thresholding of the image is applied prior to transmission to save bandwidth. Since the server does not receive a complete image, it cannot return an augmented image and all augmentation must occur on the client. This is achieved via a modeling language describing the placement, orientation and design of virtual objects that the mobile device can then use to build an augmented image. Due to the reduced capabilities of mobile devices, the augmented images are generally quite simplistic. By contrast, it is intended that the AR phone displays images fully generated within the intelligent environment which affords greater control over the augmentation, enables far more sophisticated augmented images, and requires less processing by the client. A major caveat is of course that far higher bandwidth is required between the phone and the intelligent environment.

## 3. Accessible AR Scenarios

One scenario sees the placement and organisation of virtual furniture in an empty room. This application of AR has been explored in the VOMAR project (as described in (1)), however the ubiquity of the mobile phone combined with access to an IE allows the organisation to take place in a real room. The user carries a series of fiducial markers which have been associated with pieces of furniture from a personalised furniture profile within the IE. These could be items they own or taken from a catalogue, with the associations managed through an interface similar to VOMAR. By first placing the markers on the floor and walls around a room and then viewing them through the AR phone, the user is able to see the items of furniture at full scale as they would really appear. Finding if and how they would all fit is simply a matter of reorganising the markers, in a similar fashion to the instrument panel application of the tiles project (6). This could be particularly useful for users on site, such as interior designers or people looking to rent or buy property. The interaction with an IE in this scenario means that the user only needs their phone and some markers to be able to see their furniture across a variety of different locales.

Another possibility involves the augmentation of digital state and control information over physical devices. While many devices within an office such as printers and phones comprise physical interfaces, there are also many that do not. For example, electronic door locks can be configured via a central control system to only allow access to certain users, but viewing who may access them or changing these settings is rarely possible through the locks themselves. However by placing fiducial markers on such objects within a physical environment, they can be associated with their digital information. Viewing them through the AR phone allows the digital information to be extracted from the IE and augmented upon the device as different colours, floating menus and so on. The AR phone can also allow users to edit the information, write the changes back into the IE and thus affect the operation of the physical devices. The strength of the AR phone in this scenario is that it ties digital models of objects to their physical manifestations in an intuitive manner.

## 4. Architecture

The philosophy of the AR phone system is based around two principles. Firstly, the interface on the phone should be relatively ‘dumb’. It should simply act as a viewport of the AR application, with the bulk of the processing taking place in an IE. This keeps the applications that need to run on the phone simple, allowing them to be deployed across a range of phones with differing capabilities. Secondly, since the AR phone is intended to be unencumbered by tethers to workstations or particular locations, users should be able to roam throughout an IE and access the AR services regardless of their actual positions.

To explore the AR phone concept, we designed and implemented a simple, distributed prototype system. Following our design philosophy, we developed three separate modules. A user interface module runs on a mobile phone enabling it to communicate with one or more access point modules deployed around an IE using the low-power, low-range Bluetooth wireless communication standard. The access point modules are responsible for disseminating the received input to a centralised augmented reality service module. This processes the raw image input and returns the augmented results to the access point, which in turn transmits it back to the phone for display to the user. The images are sent through the system as uncompressed bitmaps.

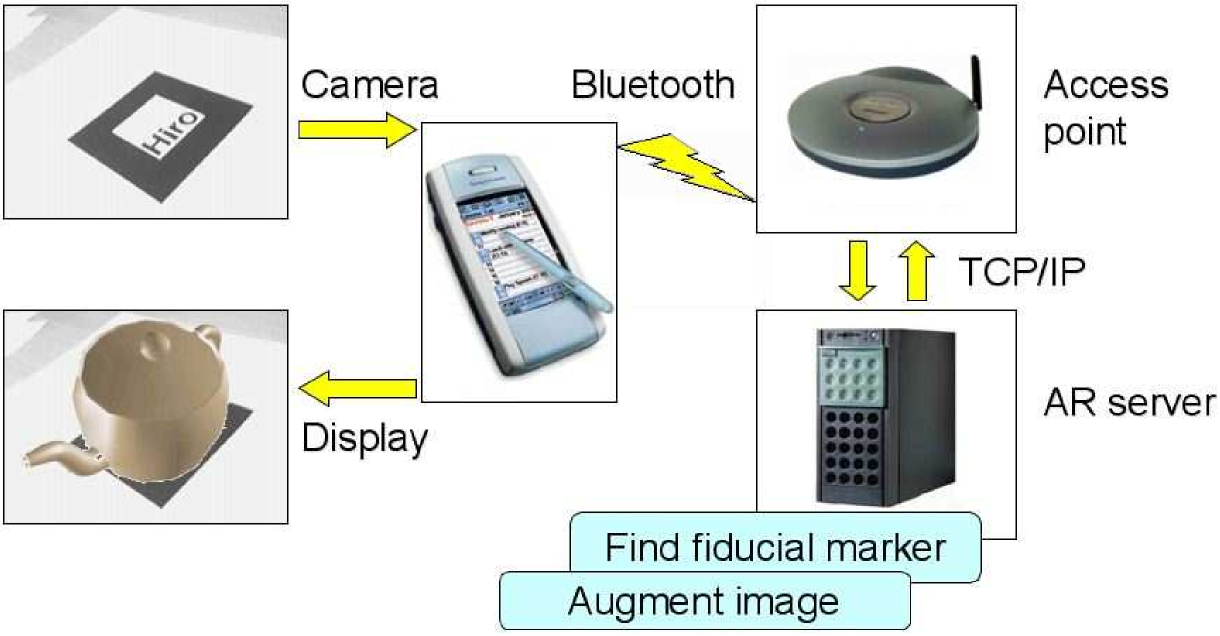

Figure 1: The AR phone pipeline.

### 4.1. User Interface

The user interface to the system has been developed on a Sony Ericsson P800 mobile phone running version 7.0 of the Symbian OS. The P800 is a good choice since it includes a built-in camera, a large colour screen and Bluetooth communication in the standard model.

The application allows the user to capture an image, send it to an access point via Bluetooth and display the augmented result on the phone's colour display. As of August 2003 there is no support for capturing video from the internal camera: it supports only the capture of snapshot images. These images can be taken at three resolutions (640x480, 320x240, and 160x120).

The application is able to scan its local area for any available Bluetooth access points which are then displayed to the user for selection. Once an access point is selected a L2CAP connection is established between it and the phone. When a user captures an image, the phone sends it to the access point and waits for an augmented result whereupon it is displayed to the user. At this point a new image can be captured and the process can be repeated as often as desired.

### 4.2. Access Points

The user interface running on the phone connects to the AR service though the access points. These access points have been developed to run under Linux with the BlueZ Bluetooth stack. They accept raw image data sent from the AR phones, convert the input to a common format, and forward the result to the AR service module. After processing, the results are converted back into the phone's native format and sent back to the phone over the Bluetooth connection. Due to the short range of the Bluetooth protocol, numerous access points need to be placed around an environment to enable easy access from any location.

### 4.3. AR Service

The AR service is a generic module that is able to receive image input from the access points and perform processing and manipulation of the data. The specifics of the processing are dependent on the service provided. The decoupling of the service processing allows for multiple AR applications to be developed for the phone using common access points and user interfaces. After processing of the image, the result is returned to the access point module requesting the service, which then passes it onto the phone.

### 4.4. Prototype Application

We implemented an AR service for our prototype which simply augments a snapshot image with a three dimensional graphic. The image augmentation service examines the input image for a particular fiducial marker which, if found, is overlaid in three dimensional space with a virtual teapot (Figure 1). This is achieved with the AR toolkit (5), and the OpenGL GLUT library. Depending on the chosen input resolution, the round trip time varies from two to 15 seconds. The largest bottleneck in this system is the transmission of the uncompressed bitmap image over Bluetooth.

## 5. Future Directions

The AR phone is a work in progress. The underlying concepts and our prototype implementation leave much room for improvement and extension.

For our prototype, we have focused on a ‘viewport interface’. That is, the phone is nothing but a viewport into an intelligent environment; we do not make use of any input modalities apart from the camera. This is appealing as it allows the application on the phone to be extremely simple, but it raises the problem of how to actually interact with an environment as opposed to merely view it. An obvious solution is to use ancillary markers as tangible control elements (such as virtual dials, sliders, levers, etc.) as in (1).

An alternative approach could be a ‘control interface’ where other input modalities on the phone such as a touch screen, speech recognition or jog dials are used to affect the environment. This could make for more efficient interfaces but would require more complicated applications on the mobile phone.

In the near future we hope to create user interface modules for other mobile devices including the Nokia 7650 mobile phone and the Toshiba e740 PDA with an attached camera. In the case of the latter, we will be able to make use of a higher bandwidth 802.11b wireless connection instead of Bluetooth which should lead to a great performance improvement in our prototype application.

It should also be possible to improve performance by applying compression algorithms to the data transmitted over Bluetooth. Since the data forms an image, we are able to use lossy compression techniques like JPEG which should significantly reduce the amount of traffic.

It may also be possible to move some parts of the AR toolkit into the user interface module such as the thresholding of images or detection of markers. This would mean far less data would need to be transmitted over Bluetooth and should increase performance. It should be noted that this approach strays from our design philosophy however.

The current prototype requires fiducial markers to be placed within the physical environment. It would be less intrusive if the system could recognise natural features. For example, the IE visualisation scenario mentioned in section 3 might be able to support the recognition of physical devices like printers in the environment without the need for them to be tagged with fiducial markers. One particularly interesting possibility is that of capturing images from a real street directory, determining the location in a city based on the streets visible in the image, and overlaying real-time traffic information (similar to the TrafficGuage project (7)) or photographs of the location.

## Acknowledgements

We would like to thank: Dr. Mark Billinghurst and the other members of HIT Lab New Zealand for their hospitality and inspiration during our recent visit; The School of Information Technologies at the University of Sydney for providing us with the opportunity to travel to New Zealand; and Dr. Aaron Quigley for his guidance on this paper.

## References

[1] Billinghurst, M., Kato, H. and Poupyrev, I., Collaboration with Tangible Augmented Reality Interfaces. *Presented at HCI International 2001, August 5-10, 2001, New Orleans, LA, USA.*

[2] Geiger, C., et al. Mobile AR4ALL. *http://citeseer.nj.nec.com/kleinjohann01mobile.html*.

[3] Hightower, J. and Borriello. G. (August 2001). Location systems for ubiquitous computing. *Computer*, 34(8), 57-66.

[4] Interactive Media Systems Group, Institute for Software Technology and Interactive Systems, Vienna University of Technology *http://www.ims.tuwien.ac.at/research/handheld_ar*.

[5] Kato, H. and Billinghurst, M. (October 1999). Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. *Proceedings International Workshop on Augmented Reality*. (IWAR99).

[6] Poupyrev, I., Tan, D.S., Billinghurst, M., Kato, H., Regenbrecht, H., Tetsutani, N., Tiles: A Mixed Reality Authoring Interface. *INTERACT 2001.*

[7] TrafficGauge, Inc., TrafficGauge. *http://www.trafficgauge.com/*.