Augmented Reality at IBM

Alex Phillips, Senior Managing Consultant, IBM Global Business Services.
Sinem Güven, IBM T.J Watson Research Center, Hawthorne, NY, USA

May 2010

IBM has recently been working with a number of partners to explore Augmented Reality (AR) technology, in order to understand how AR can enable new innovative and practical solutions for the future. Given their strong background in advanced graphics, image recognition and computer vision technologies, IBM aims to leverage their expertise in these areas to create compelling AR technology and solutions.

In this paper, we present a variety of applications developed at IBM, which include mobile as well as desktop-based solutions, and range from navigation guides for enhancing the way fans experience sports events to remote collaboration tools, shopping advisors and technical support systems.

ACME

With support from IBM Research and Nokia Research Center, the VTT Technical Research Centre of Finland created an experimental system that enables people in multiple locations to interact and collaborate with avatars and objects in a single, virtual meeting. Objects and avatars are located in a "virtual" space that mirrors the corresponding physical room. Sensors, cameras and microphones located on both ends of the conversation allow voices, head and hand gestures and movements to change in concert with the behavior of participants, enabling participants to sense the vital visual cues of body language. In this proof-of-concept, participants in physical rooms wear video see through glasses that depict three-dimensional images of their online counterparts as they stand, walk, talk or demonstrate and manipulate virtual objects shared between the spaces.

ACME [1], which stands for Augmented Collaboration in Mixed Environments, is a remote collaboration tool, which was assembled using an open source viewer from Linden Lab's Second Life virtual world, as well as from open source ARToolkit and OpenCV libraries. The use of open source components lowers the costs associated with the project as it matures, and encourages the participation of more computer programmers and developers. ACME technology provides a more affordable and eco-friendly alternative to physical meetings. It is also more interactive than telephone conferences, video conferences - and even on-screen meetings held exclusively in virtual spaces. "ACME is a compelling example of the kind of R&D now being conducted that will enable the business community to work more intelligently, in a more productive, efficient, convenient and immersive fashion," said Neil Katz, an IBM Distinguished Engineer in the company's CIO Office, and liaison with the ACME project. "It's easy to imagine that this technology, especially when it becomes somewhat more mature, will give people a promising new option for collaborating more interactively with colleagues in an increasingly decentralized world." The research towards this new level of meeting experience was supported by the Finnish Funding Agency for Technology and

Innovation (TEKES). IBM Research, together with Nokia Research Center, provided additional funding and contributed to the technical direction.

SEER

At Wimbledon 2009, the All England Lawn Tennis Club and IBM unveiled smart mobile applications designed to transform how fans access information and keep up with the action at Wimbledon 2009. Since 1990, IBM has worked with the Wimbledon team to send the captured score and statistical data around the world in an instant, keeping on-site broadcasters, media and tennis fans up to date with all the latest scores and statistics.

The Seer Android version features location-aware visualisation technology developed for the G1. The Augmented Reality application acts as a real-time guide and provides an interactive map of the 2009 tournament allowing selected users to see what others can not, as well as providing up to the second scores. The Seer Android is an innovative application that was trialed at Wimbledon 2009 that takes a live video feed from the handset's camera, and superimposes content and data associated with various points of interest into that video stream. From tennis to food courts, points of interest throughout the Wimbledon grounds have been plotted using GPS. By making use of the G1's digital compass and precise GPS coordinates, the application offers a 'heads up display' to show the user what they are looking at. It augments this with other live data from the scoring systems and IBM scouts reporting from around the grounds, to give the user a comprehensive and dynamic insight into their surroundings. For instance, pointing the camera lens towards a court will not only identify the court number, but also display details about the current and subsequent matches.

Seer Android users were also able to use their phones' Map view, which pinpoints their location on a detailed map of the grounds, and can be used as a way finder. The Timeline view is an aggregation of news feeds and updates from IBM scouts, and allows users to see in real-time what is happening around the site. And a handy 'Radar' function indicates the user's current position and nearby points of interest within range. As a result of the demonstration at Wimbledon, IBM has engaged with a number of clients who wanted to explore how it could be used in their business.

Figure 1. The AR Camera View for the Taxi stand shows the user the exact location behind the building in front of them.

Figure 2. The Map View of the Taxi Stand, indicating where the user is and where the nearest taxi stands are.

VIRA

IBM Research has also developed a number of AR prototypes that leverage advanced capabilities to combine various images and displays, real or virtual. For example, the VIRA (Virtual Agent) [2] system is an application for remote technical support, which allows agents to interact with and guide end-users to fix their computer problems. VIRA augments users' desktops or physical environments with the presence of the person they are communicating with such that they can use not only text or voice, but also hand gestures, such as pointing, to help communicate more effectively while experiencing an enhanced sense of co-location of the remote person. During the remote help session, the agent and the end-user can chat, and additionally, the agent points at icons, folders, executables, etc. to help the user fix the problem. This way, the user is in full control of their machine, but is guided by the agent as if they are physically co-located.

Unlike existing remote take over systems, VIRA provides a visual representation of the agent, i.e. their presence, to help clarify targets, thereby easing the interaction and enriching the end-user's experience. VIRA works for augmenting desktop computers (Figure 3) as well as real objects (Figure 4) situated in the user's real world environment. When the user enables the presence-enhanced world application, currently, a window appears on their desktop, which displays their camera feed as well as the overlaid avatar video being streamed out of Second Life (SL). As a result, the user sees the SL avatar, superimposed onto their real world environment, to guide the user through the support session.

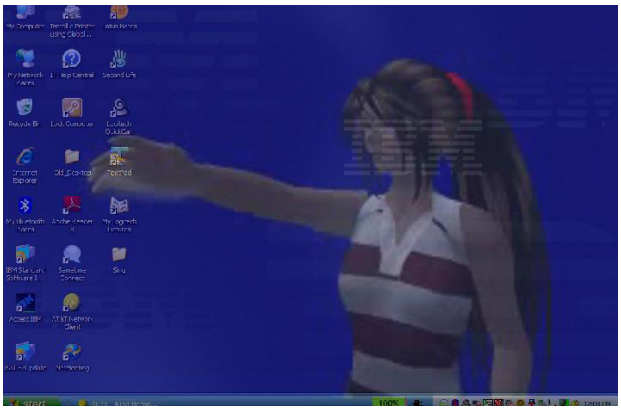

Figure 3. Although the user's desktop is 2D, the avatar can interact in 3D, moving closer or retracting when necessary.

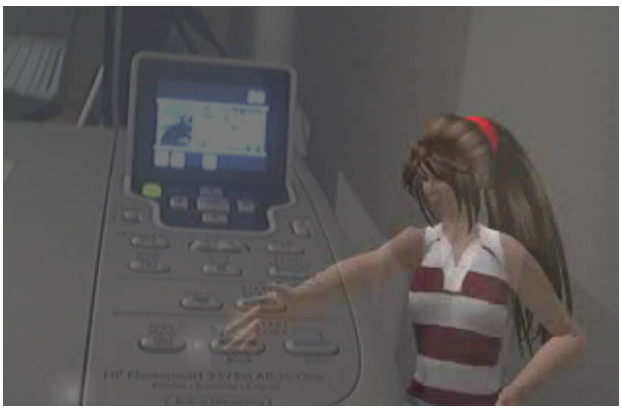

Figure 4. 3D avatar representing the agent augments the real world, in this case, the user's printer.

MAR

MAR (Mobile Augmented Reality for Retail) [3] is an AR application, which runs on a mobile phone, and depicts the social content and networks as well as product information associated with various consumer items within the visual context of these items. By aiming a mobile phone's camera at an item's 2D bar code, a consumer can visualize this information as a swarm of animated 3D avatars that appear to be hovering above the item in the physical world (Figures 5 and 6). MAR avatars typically have different roles such as online reviewers, designers of the system, sales representatives, and even other consumers who have physically been in the store and have shared their opinion on an item.

Figure 5. Shopper browsing through different reviewers for a book of interest.

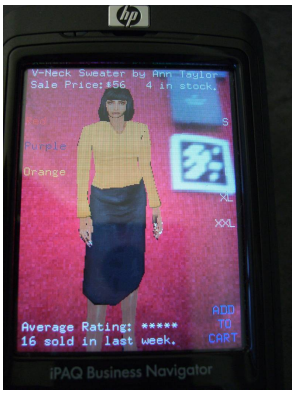

Figure 6. Shopper visualizing different colors of a sweater of interest in AR.

Interaction with the avatars is multi-modal. Users can browse through the avatars and engage in a limited, structured conversation with an avatar to elicit content previously generated by its owner – such content can be created either by physically leaving an on-site (text or voice) review using the VIRA system, or by publishing it online. Alternatively, if an avatar's owner is online and available, users could approach the avatar and engage in a live chat like those experienced in other virtual worlds, such as Second Life. Avatars can deliver not only voice notes, but also text-based content (such as reviews previously entered online) using text-to-speech (TTS) delivered via the mobile phone's speaker or a Bluetooth headset.

MAR also has applications in technical support and maintenance where maintenance personnel can access a social network of experts from their portable device. The experts are selected based on the content that the device is being pointed at. Concept mock-ups are shown in Figure 7.

Figure 7. User consulting to experts who left an opinion on this motherboard as FAQs.

REFERENCES

[1] ACME: http://www.youtube.com/watch?v=DNB0_c-5TSk

[2] VIRA: http://domino.research.ibm.com/comm/research_projects.nsf/pages/projects.presence.html

[3] MAR: http://domino.research.ibm.com/comm/research_projects.nsf/pages/projects.mar.html