The Need for AI: Intuitive User Interfaces for Mobile Augmented Reality Systems

S. Julier, D. Brown and Y. Baillot*
Advanced Information Technology
Naval Research Laboratory
Washington DC
{*julier, dbrown, baillot*}@*ait.nrl.navy.mil*

April, 2001

Abstract

This position paper discusses some of the difficulties posed by developing user interfaces for mobile augmented reality systems. We argue that these are a superset of the challenges which are encountered by mobile computers which use 2D textual displays and we discuss the potential role which could be played by artificial intelligence methods.

1 Introduction

Mobile computers will dramatically change the way in which information is delivered to individuals. Through the use of laptop computers, personal digital assistants and mobile telephones, it is possible to read email and even surf the web. However, the power of mobile computing is that the information which is being displayed can be tailored to the user's current tasks and contexts. Relatively unobtrusive information delivery systems, such as the Wearable Remembrance Agent [10], have been developed.

Recent development in portable computing hardware, position and orientation trackers, and see-through displays have begun to make mobile augmented reality systems feasible. Augmented reality (AR) integrates virtual information with the user's physical environment. Graphics-based AR can provide a user with a "heads up display" in which computer graphics is spatially registered with, and overlaid on, geographic locations and real objects.

Mobile AR becomes a superset of conventional wearable computers. However, AR offers a fundamentally different way of displaying the information and providing the users with tools to interact with that information. Rather than just provide 2D textual displays, a user sees information and is able to directly interact with it within the user's own 3D space. The extra dimension means that augmented reality has the potential to be much more valuable and much more complicated than traditional wearable systems. However, with this extra freedom comes increased complexities. A display which is heavily cluttered is unreadable. Poorly positioned labels or annotations can be confusing or highly misleading.

In this position paper we argue that Artificial Intelligence (AI) techniques have the potential to play a significant role in developing intuitive interfaces which will help to mitigate or overcome some of the difficulties involved. We survey some of the technology and approaches which have been used, and identify key areas where we feel developments are needed. The structure of this paper is as follows. The next section discusses the application scenario in more detail. Section 3 discusses the problem of generating

*S. Julier, D. Brown and Y. Baillot are with ITT Advanced Engineering Systems, Alexandria, VA.

and managing a graphical display to minimize the problems of clutter and information overload. Section 4 discusses the issues of how a user can issue instructions and commands to a system. We summarize and conclude in Section 5.

2 Application Scenario

Our goal is to develop software systems and interaction techniques to support multiple, mobile, collaborating users with wearable AR systems [6]. These users would interact with other users of stationary VR, AR, and desktop systems. To this end, several systems have been developed. One such system is shown in Figure 1 shows the Battlefield Augmented Reality System (BARS) which is currently under development at NRL in collaboration with Columbia University. Mostly built from common off the shelf (COTS) products, the system is composed of 6DOF trackers (an Ashtech GG Surveyor real-time–kinematic GPS for position, an InterSense IS300Pro for orientation), a see-through head-worn display (Sony LDI-D100B Glasstron), a wireless network and a wearable computer with 3D hardware graphics acceleration. The purpose of the system is to provide a user with situation awareness — given a set of tasks to complete in an urban environment, the system must provide the user with information which is pertinent to that task. The types of information which can be displayed include the names of buildings, routes which have to be followed, objectives which have to be achieved and the locations of other users. As an example, Figure 2 shows an actual image taken through the see through head mounted display.

The display in Figure 2 appears to be relatively straightforward. However, if the display becomes much more complicated, significant improvements in systems technology will be required. We focus on two such issues — display management through information filtering and intuitive 3D user interaction paradigms.

Figure 1: Prototype mobile augmented reality system. This system is constructed from COTS products.

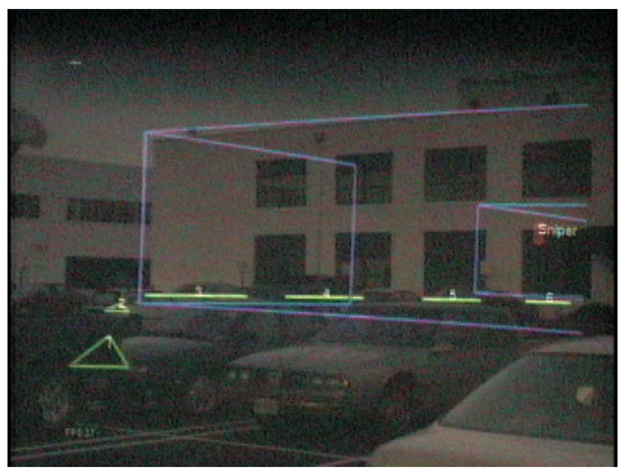

Figure 2: Sample output from the prototype augmented reality system. In this scenario, a user follows a route (triangles) around the edge of a building. Sniper is visible beyond.

Figure 3: Sample output from the prototype augmented reality system when all information about the system is displayed.

3 Managing the Displays

Arguably one of the most fundamental problems with an AR system is the potential for information overload. In a dense environment, such as a city, there can be a substantial amount of information which can be available. However, näively displaying all available information can be highly confusing. This is illustrated in Figure 3 which shows the output from the system, rendered in the same view as Figure 2, with all data from the database is displayed. As can be seen, the result is extremely confusing and is difficult to interpret. Initial thinking suggests that a simple line-of-sight analysis might be sufficient to overcome this difficulty. One pragmatic approach, which was utilized in the ARQuake system, is to build a 3D model of the environment and make it black so that occluded objects are not visible [12]. Although this approach is suitable for games it is not suitable for information systems where the ability for the system to provide "X-ray vision" and let the user see through obstructions is extremely important. Figure 2 shows that, if a user follows a route, they will become visible to a sniper as they walk around a building.

To overcome these difficulties, it is necessary to apply some form of autonomous information filtering. This technique attempts, from a user's current context and task set, to choose the most appropriate subset of available information. However, the method of choosing the information is a function of the information requirements for any particular task. In [7] and [11] we presented a two-step framework for information filtering. The first stage employs the spatial model of interaction [1]: a user is surrounded by a *focus* region and all objects are surrounded by a *nimbus* region. If an object's nimbus intersects with the user's focus, the object is a potential candidate to be shown. The second step uses task-dependent logic to cull the list of objects to a critical subset to be shown. An approach based on dot products to score the relevance of an object with respect to a task was developed.

Although this algorithm dramatically reduces the clutter in a display, (compares Figure 2 and 3), it has a number of important shortcomings. The most important of these is that the current implementation is very limited in its capability to encode both the context (according to a user's task vector) and domain knowledge (a set of parameterized functions). Better context management can be achieved through the use of more sophisticated sensors and processing algorithms to detect context. For example, Golding [4] applied simple machine learning techniques (such as Bayesian nets) to deduce the context of a user from a range of sensors.

4 User Interfaces

The mobile outdoor system is designed to aid a user in completing a task. It must provide information to the user, and vice-versa, without distracting the user from that task. We feel the system can provide the best interface by monitoring many sources of data about the user and using intelligent heuristics to combine that data with information about the environment and task to produce a highly usable interface.

The system contains a detailed physical model of objects in the real environment that is used to generate the registered graphical overlay. This model is stored in a shared database that also contains information about the objects such as a general description, threat classification, and so on. Using knowledge representation and reasoning techniques, we can also store in this database information about the objects' relevance to each other and to the user's task.

We believe that early uses of BARS for situational awareness will mainly consist of users picking objects in the environment, either to find out more about them ("Where is the electrical cut off switch?") or to add information about them ("I saw a sniper on the third floor of that building"). Thus, we need to find a way to let the user easily pick items in the environment.

Using tracking sensors, we can measure the user's position and head orientation in the environment, and use that information to determine a reasonable approximation of the user's gaze direction–though we are looking at eye trackers to make this measurement more accurate, such as described in [13]. We can also track the user's hands relative to the body position. This information gives us rays of gaze and gesture that beam through the 3D model.

Once we have the rays of gaze and gesture, we have a set of possible objects at which the user may be looking and pointing that sit along and near these rays. This set will usually be larger than a single object for each mode because the our models are of dense urban terrain and many objects will sit along or near these rays. Also, because this is an AR system, even if an object is occluded in the real world, it may be visible on the AR display, so that a user may pick an object whose physical counterpart is not visible at that time.

We will use an inductive heuristic, such as the Inductive Dichotomizer Algorithm ID3 [9], to determine the single object at which a user is looking or pointing out of the many possible. If the heuristic picks the object the user really wants, it is considered a successful determination. The user will have the option of rejecting the selection and choosing another object, creating a feedback mechanism. Thus, the heuristic will analyze data about objects in the set of possible selections, such as their task relevance values and whether or not they are occluded, as well as its own previous successes, to pick the selected object.

So far we have addressed how AI can help the user pick objects in the environment in a natural way. Now that we can select items, how do we decide what to do with them without resorting to cumbersome or non-intuitive input devices? For this purpose we will look at multi-modal inputs.

4.1 Multi-modal Inputs

We have determined a way to allow the user to select one or more objects on which to perform an operation. However, how will the user specify the operation to be performed? One way is to superimpose a traditional 2D WIMP interface on the display [3]. For example, a context menu may appear next to the projection of the selected object allowing the user to pick an operation. However, in a wearable outdoor system, the user will not have a mouse and keyboard in which to interact with the WIMP interface, so we will look to more natural interfaces.

Gesturing and speaking are natural human interaction techniques. Multi-modal interfaces involving gesturing and speaking have shown promising results [8]. Additionally, there are mature systems using these techniques to determine the intent of a user, such as QuickSet [2] which utilizes a set of agents which communicate through a centralized blackboard or facilitator.

As with object selection, we propose that a heuristic can be developed to use results of multi-modal integration algorithms along with user feedback to create a highly usable interface for performing actions on selected objects.

5 Conclusions

Mobile augmented reality has the potential to be an extremely powerful paradigm for presenting information to a mobile user. However, the problems which are faced in developing an effective user interface are a superset of those faced with mobile computers with 2D textual displays. We have discussed two problems — the control of what information will be displayed and the way in which a user can interact with the system. In both cases, we believe that many AI principles will greatly contribute towards the design of these interfaces. In a related paper, our colleagues at Columbia University discuss how limitations in the resolution of a tracking system provide even more constraints on the design and operation of a user interface [5].

References

[1] S. Benford and L. Fahlén. A Spatial Model of Interaction in Large Virtual Environments. In *Proceedings of ECSCW '93*, Milan, Italy, September 1993.

[2] P. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pitman, I. Smtih, L. Chen and J. Clow. QuickSet: Multimodal interaction for distributed applications. In *Proceedings of the Fifth ACM International Multimedia Conference, New York, NY*, pages 31–40. ACM Press, 1997.

[3] S. Feiner, B. MacIntyre M. Haupt and E. Solomon. Windows on The World: 2D Windows for 3D Augmented Reality. In *Proceedings of UIST '93, Atlanta, Georgia*, pages 145–155, November 1993.

[4] A. R. Golding and N. Leash. Indoor Navigation Using a Diverse Set of Cheap, Wearable Sensors. In *Proceedings of the Fourth International Symposium on Wearable Computers, San Francisco, California*, 1999.

[5] T. Höllerer, D. Hallaway, N. Tinna and S. Feiner. Steps toward accommodating variable position tracking accuracy in a mobile augmented reality system. In *submitted to the 2nd Int. Worksh. on Artificial Intelligence in Mobile Systems (AIMS '01)*, 2001. currently under review.

[6] T. Höllerer, S., Feiner, T. Terauchi, G. Rashid and D. Hallaway. Exploring MARS: Developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers and Graphics*, 23(6):779–785, December 1999.

[7] S. Julier, M. Lanzagorta, S. Sestito, L. Rosenblum, T. Höllerer and S. Feiner. Information Filtering for Mobile Augmented Reality. In *Proceedings of the IEEE 2000 International Symposium on Augmented Reality, Germany*. IEEE, October 2000.

[8] C. Mignot, C. Valot and N. Carbonell. Multi-Modal User Interfaces. In *Proceedings of Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*, pages 67–68, 1993.

[9] J. R. Quinlan. Discovering Rules from Large Collections of Examples. In D. Mitchie, editor, *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, Edinburgh, 1979.

[10] B. J. Rhodes. The wearable remembrance agent: A system for augmented memory. In *Proc. ISWC '97 (First Int. Symp. on Wearable Computers)*, pages 123–128, Cambridge, Mass., USA, October 1997.

[11] S. Sestito, S. Julier, M. Lanzagorta and L. Rosenblum. Intelligent Filtering for Augmented Reality. In *Proceedings of SimTecT 2000, Sydney, Australia*, 2000.

[12] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris and W. Piekarski. ARQuake: An Outdoor/Indoor Augmented Reality First Person Application. In *Proceedings of the Fourth International Symposium on Wearable Computers, Atlanta, Georgia*, 2000.

[13] S. Zhai, C. Morimoto and S. Ihde. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of CHI '99*, pages 246–253, 1999.