Telegeoinformatics: Location-Based Computing and Services. H Karimi and A. Hammad (eds.). Taylor & Francis Books Ltd., 01/2004

Chapter Nine

Mobile Augmented Reality

Tobias H. Höllerer
Steven K. Feiner

9.1 INTRODUCTION

As computers increase in power and decrease in size, new mobile, wearable, and pervasive computing applications are rapidly becoming feasible, providing people access to online resources always and everywhere. This new flexibility makes possible new kind of applications that exploit the person's surrounding context. Augmented reality (AR) presents a particularly powerful user interface (UI) to context-aware computing environments. AR systems integrate virtual information into a person's physical environment so that he or she will perceive that information as existing in their surroundings. Mobile augmented reality systems (MARS) provide this service without constraining the individual's whereabouts to a specially equipped area. Ideally, they work virtually anywhere, adding a palpable layer of information to any environment whenever desired. By doing so, they hold the potential to revolutionize the way in which information is presented to people. Computer-presented material is directly integrated with the real world surrounding the freely roaming person, who can interact with it to display related information, to pose and resolve queries, and to collaborate with other people. The world becomes the user interface.

This chapter provides a detailed introduction to mobile AR technology with in-depth reviews of important topics, such as wearable display and computing hardware, tracking, registration, user interaction, heterogeneous UIs, collaboration, and UI management for situated computing. As part of this introduction, we define what we mean by augmented reality, give a brief overview of the history of the field in general, and review some important mobile AR system considerations. Section 9.2 discusses the potential and possibilities of MARS technology, with a detailed overview of prototype application areas, and reviews the challenges that impede immediate widespread commercial adoption. In Section 9.3 we take a closer look at the requirements and specific components of MARS, before examining UI concepts in Section 9.4. We conclude the chapter with an outlook on research directions.

9.1.1 Definition

A*ugmented reality* is related to the concept of *virtual reality* (VR). VR attempts to create an artificial world that a person can experience and explore interactively, predominantly through his or her sense of vision, but also via audio, tactile, and other forms of feedback. AR also brings about an interactive experience, but aims to supplement the real world, rather than creating an entirely artificial environment. The physical objects in the individual's surroundings become the backdrop and target items for computer-generated annotations. Different researchers subscribe to narrower or wider definitions of exactly what constitutes AR. While the research community largely agrees on most of the elements of AR systems, helped along by the exchange and discussions at several international conferences in the field, there are still small differences in opinion and nomenclature. For the purpose of this chapter, we follow the definitions of Azuma (1997) and Azuma and colleagues (2001). We will define an AR system as one that combines real and computer-generated information in a real environment, interactively and in real time, and aligns virtual objects with physical ones. At the same time, AR is a subfield of the broader concept of *mixed reality* (MR) (Drascic and Milgram, 1996), which also includes simulations predominantly taking place in the virtual domain and not in the real world.

Mobile AR applies this concept in truly mobile settings; that is, away from the carefully conditioned environments of research laboratories and special-purpose work areas. Quite a few technologies, introduced in earlier chapters, must be combined to make this possible: global tracking technologies (Chapter 4), wireless communication (Chapter 5), location-based computing (LBC) and services (LBS) (Chapters 6 and 7), and wearable computing (Chapter 8).

After giving a brief historical overview of AR systems in the next subsection, we will take a look at the components needed to create a mobile AR experience. While AR can potentially supplement the physical environment with information perceptible by all human senses, visual and auditory overlays are currently the most commonly applied augmentations. In the case of visual AR, computer-generated graphics are spatially registered with, and overlaid on, real objects, using the display and tracking technologies that we describe in this chapter.

9.1.2 Historical Overview

While the term *augmented reality* was coined in the early 1990s, the first fully functional AR system dates back to the late 1960s, when Ivan Sutherland and colleagues (1968) built a mechanically tracked 3D see-through head-worn display, through which the wearer could see computer-generated information mixed with physical objects, such as signs on a laboratory wall. For the next few decades much research was done on getting computers to generate graphical information, and the emerging field of *interactive computer graphics* began to flourish. Photorealistic computer-generated images became an area of research in the late 1970s, and progress in tracking technology furthered the hopes to create the ultimate simulation machine. The field of VR began to emerge. Science fiction literature, in particular the early 1980s movement of *cyberpunk*, created visions of man-machine symbiosis. The entertainment industry jumped in with movies such as the *Terminator* series, which presented a specific rendition of what computer-annotated vision could look like. During the 1970s and 80s, AR was a research topic at some institutions, such as the U.S. Air Force's Armstrong Laboratory, the NASA Ames Research Center, the Massachusetts Institute of Technology, and the University of North Carolina at Chapel Hill. As part of the US Air Force *Super Cockpit* project, Tom Furness developed a high-resolution heads-up overlay display for fighter pilots, supported by 3D sound (Furness, 1986).

(a) **(b)**

Figure 9.1 Mobile AR restaurant guide. (a) User with MARS backpack, looking at a restaurant. (b) Annotated view of restaurant, imaged through the head-worn display.

It was not until the early 1990s, with research at the Boeing Corporation, that the notion of overlaying computer graphics on top of the real world received its current name. Caudell and Mizell (1992) worked at Boeing on simplifying the process of conveying wiring instructions for aircraft assembly to construction workers, and they referred to their proposed solution of overlaying computer-presented material on top of the real world as *augmented reality*. Even though this application was conceived with the goal of mobility in mind, true mobile graphical AR was out of reach for the available technology until a few years later. Also during the early 1990s, Loomis and colleagues (1993) at the University of California, Santa Barbara, developed a GPS-based outdoor system, presenting navigational assistance to the visually impaired with spatial audio overlays.

Since about the mid-1990s computing and tracking devices have become sufficiently powerful, and at the same time small enough, to support registered computer-generated graphical overlays in a dynamic mobile setting. The Columbia *Touring Machine* (Feiner *et al.*, 1997) is an early prototype of an outdoor MARS that presents 3D graphical tour guide information to campus visitors, registered with the buildings and artifacts the visitor sees. Figure 9.1 shows a more recent version of the Touring Machine, annotating restaurants in the Columbia University neighborhood.

In 1979, a mobile audio display changed the way people listened to music: the Sony Walkman. It was one of the three most successful consumer products of the 1980s, with the other two being roller skates and digital watches (another kind of mobile display). This commercial success paved the way for other mobile devices, among them personal digital organizers. The original Sony Walkman weighed in at 390g, not counting batteries and audio tape. Today, many MP3 players weigh less than 40 grams, including batteries.

Wearable computing (Mann, 1997; Starner *et al.*, 1997a), took off in the 1990s, when personal computers were becoming small enough to be carried or worn at all times. The earliest wearable system was a special purpose analog computer for predicting the outcome of gambling events, built in 1961 (Thorp, 1998). On the commercial front, palmtop computers embody the trend towards miniaturization. They date back to the Psion I organizer from 1984 and later became commonplace with the introduction of the Apple Newton MessagePad in 1993 and the Palm Pilot in 1996. Since the mid 1990s, wearable computing has received ever-increasing commercial backing, and the miniaturization and more cost-effective production of mobile computing equipment resulted in several companies now offering commercial wearable computing products (e.g., Xybernaut, Charmed Technology, ViA, Antelope Technologies).

In terms of the technologies necessary for a mobile AR experience, we will look briefly at the historical developments in the fields of *tracking and registration*, *wireless networking*, *display technology*, and *interaction technology* in Section 9.3. Now that the technological cornerstones of mobile AR have been placed, it might seem that it is purely a matter of improving the necessary components, putting it all together, and making the end result as reliable as possible. However, there are more challenges lying ahead, and, after giving an overview of the necessary components of a functional mobile AR system in the following subsection, we will come back to these challenges in Section 9.2.2.

9.1.3 Mobile AR Systems

Revisiting our definition of AR, we can identify the components needed for MARS. To begin with, one needs a *computational platform* that can generate and manage the virtual material to be layered on top of the physical environment, process the tracker information, and control the AR display(s).

Next, one needs *displays* to present the virtual material in the context of the physical world. In the case of augmenting the visual sense, these can be head-worn displays, mobile hand-held displays, or displays integrated into the physical world. Other senses (hearing, touch, or smell) can also be potentially augmented. Spatialized audio, in particular, is often used to convey localized information, either complementing or completely substituting for visual elements (Sawhney and Schmandt, 1998).

Registration must also be addressed: aligning the virtual elements with the physical objects they annotate. For visual and auditory registration, this can be done by *tracking* the position and orientation of the user’s head and relating that measurement to a model of the environment and/or by making the computer “see” and potentially interpret the environment by means of cameras and computer vision.

Wearable input and interaction technologies enable a mobile person to work with the augmented world (e.g., to make selections or access and visualize databases containing relevant material) and to further augment the world around them. They also make it possible for an individual to communicate and collaborate with other MARS users.

Wireless networking is needed to communicate with other people and computers while on the run. Dynamic and flexible mobile AR will rely on up-to-the-second information that cannot possibly be stored on the computing device before application run-time. For example, this would make it possible to report train or bus delays and traffic conditions to the busy commuter.

This brings us to another item in the list of requirements for MARS: *data storage and access technology*. If a MARS is to provide information about a roaming individual's current environment, it needs to get the data about that environment from somewhere. Data repositories must provide information suited for the roaming individual's current context. Data and service discovery, management, and access pose several research questions that are being examined by researchers in the database, middleware, and context-based services communities. From the user's point of view, the important question is how to get to the most relevant information with the least effort and how to minimize information overload.

AR is an immensely promising and increasingly feasible UI technology, but current systems are still mostly research prototypes. There is no obvious overall best solution to many of the challenging areas in the field of AR, such as tracking and display technologies. New research results constantly open up new avenues of exploration. A developer, planning to deploy AR technology for a specific task, has to make design decisions that optimize AR performance for a given application, based on careful task analysis. In some cases, it might be possible to come up with specialized AR solutions that will work well in constrained areas with special purpose hardware support. If, on the other hand, the task analysis for the AR scenario reveals that the end-user of the system is to be supported in a wide variety of locations and situations, possibly including outdoor activities, then we enter the realm of true mobile AR.

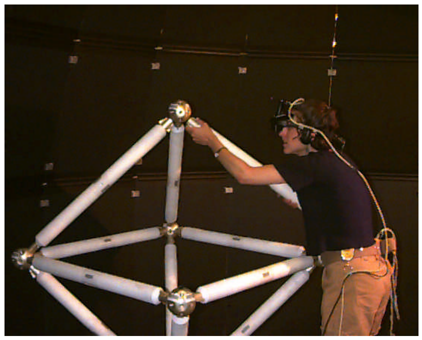
9.2 MARS: PROMISES, APPLICATIONS, AND CHALLENGES

The previous chapters on LBC and LBSs have shown how location-aware technology opens up new possibilities in the way we interact with computers, gather information, find our way in unfamiliar environments, and do business. AR can provide a powerful UI for this type of computing; one might even say, the ultimate interface: to interact directly with the world around us—the world becomes the user interface.

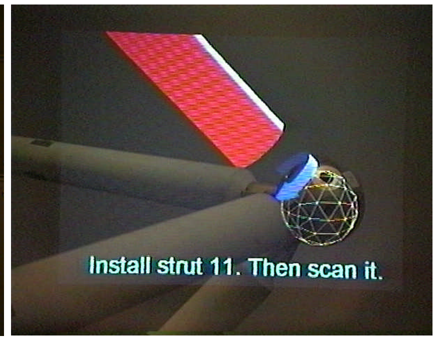
Mobile AR is particularly applicable whenever people require informational support for a task while needing to stay focused on that task. It has the potential to allow people to interact with computer-supported information (which might come from databases or as a live feed from a remote expert), without getting distracted from the real world around them. This is a very important feature for the mobile worker, or for anybody who needs or wants to use their hands, and some of their attention, for something other than controlling a computer. The next subsection gives a few examples of such occupations and summarizes application areas for which mobile AR prototypes have been tested.

9.2.1 Applications

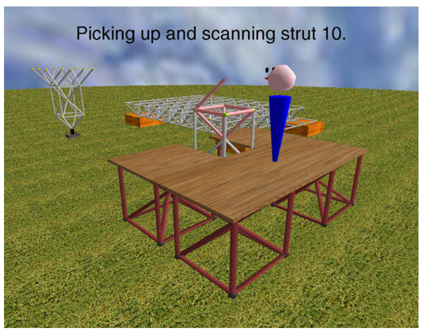
A high percentage of conference submissions on AR come from industrial research labs or are joint work between universities and industry. Many of the early AR publications are "application papers," describing applications of the new



(a) Installing a spaceframe strut



(b) View through the head-worn display



(c) Overview visualization of the tracked scene.

Figure 9.2 AR for construction.

technology in varied fields. In the rest of this section, we give an overview of potential uses for mobile AR systems.

Assembly and construction. Over a nine-year period, researchers at Boeing built several iterations of prototypes for AR-supported assembly of electrical wire bundles for aircraft (Mizell, 2001). AR overlaid schematic diagrams, as well as accompanying documentation, directly onto the wooden boards on which the cables are routed, bundled, and sleeved. The computer led (and could potentially talk) assembly workers through the wiring process. Since the resulting wire bundles were long enough to extend through considerable portions of an aircraft, stationary AR solutions were not sufficient, and the project became an exercise in making mobile AR work for a specific application scenario.

"Augmented Reality for Construction" (Feiner *et al.*, 1999), is an indoor prototype for the construction of spaceframe structures. As illustrated in Figure 9.2 (a–b), a construction worker would see and hear through their head-worn displays where the next structural element is to be installed. The construction worker scans the designated element with a tracked barcode reader before and after installation to verify that the right piece gets installed in the right place. The possibility of generating virtual overview renderings of the entire construction scene, as indicated in a small-scale example via live-updated networked graphics in one of the prototype's demonstrations (see Figure 9.2 (c)), is a side benefit of tracking each individual worker and following their actions, and could prove immensely useful

(a) (b)

Figure 9.3 Navigational AR Interfaces, imaged through head-worn displays. (a) Indoor guidance using overview visualization and arrows. (b) Virtual trails and flags outdoors (viewed from the roof of a building).

for the management of large complex construction projects. Such construction tasks would ultimately take place in the outdoors.

Maintenance and inspection. Apart from assembly and construction, inspection and maintenance are other areas in manufacturing that may benefit greatly from applying MARS technologies. Sato and colleagues (1999) propose a prototype AR system for inspection of electronic parts within the boundaries of a wide-area manufacturing plant. Their MARS backpack is tracked with a purely inertia-based (gyroscope orientation tracker plus acceleration sensor) tracking system, and calibration has to be frequently adjusted by hand. Zhang and colleagues (2001) suggest the use of visual coded markers for large industrial environments. Klinker and colleagues (2001) present the system architecture of a MARS prototype for use in maintenance for nuclear power plants. AR is well suited for situations that require “x-ray vision,” an ability to see through solid structures. Using direct overlays of hidden infrastructure, AR can assist maintenance workers who are trying to locate a broken cable connection within the walls of a building, or the location of a leaking pipe beneath a road’s surface. The exact position may have been detected automatically (e.g., by installed sensors), in which case direct visualization of the problem area via AR could be the fastest way to direct the worker’s attention to the right area. Alternatively, AR may be used as a supporting tool for determining the problem in the first place, in that it could instantaneously and directly visualize any data the worker might gather, probing the environment with various sensors.

Navigation and path finding. Considering some other outdoor-oriented uses for mobile AR, an important application area for wearable systems is their use as navigational aids. Wearable computers can greatly assist blind users (Loomis *et al.*, 1993; Petrie *et al.*, 1996) via audio and tactile feedback. If auditory information relating to real-world waypoints and features of the environment is presented to the position-tracked wearer of the system via spatialized stereo audio, this clearly matches our definition of an AR system from Section 9.1.1. Visual AR can aid navigation by directly pointing out locations in the user’s field of view, by means of directional annotations, such as arrows and trails to follow along (see Figure 9.3), or by pointing out occluded infrastructure, either directly by visually encoded overlays (Furmanski *et al.*, 2002), or indirectly via 2D or 3D maps that are dynamically tailored to the situation's needs and presented to the user (Figures 9.3 (a), 9.10, and 9.11).

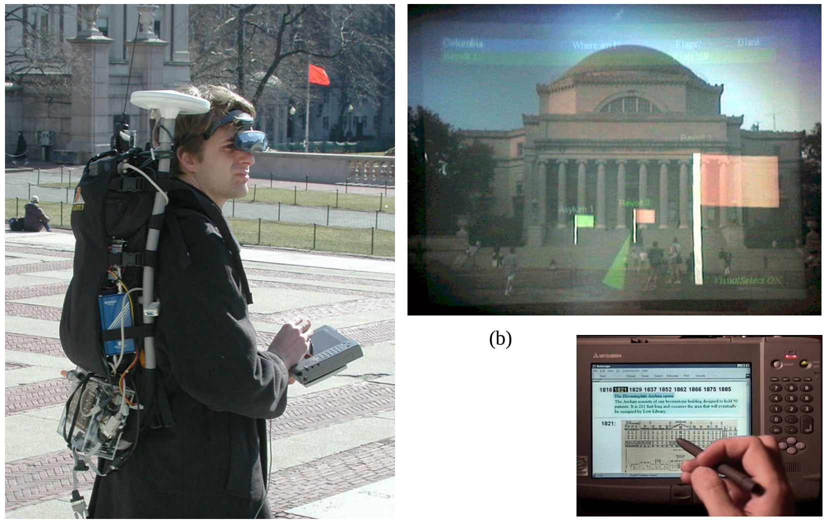
Tourism. Taking navigational UIs one step further by including information about objects in a mobile person's environment that might be of interest to a traveler, leads naturally to applications for tourism (Feiner *et al.*, 1997; Cheverst *et al.*, 2000). In this case, AR is not only used to find destinations, but also to display background information. For example, instead of looking up a description and historic account of a famous cathedral in a guide book, (or even on a wirelessly connected palm-sized computer in front of the site, AR can make the air around the church come alive with information: 3D models of related art or architecture, the life and work of the architect, or architectural changes over the centuries can be documented in situ with overlays. The possibilities are endless and only limited by the amount and type of information available to the AR-enabled individual and the capabilities of the AR device the individual is wearing.

Figure 9.1 shows an example of a mobile AR restaurant guide developed at Columbia University (Feiner, 2002). This prototype MARS provides an interface to a database of the restaurants in Morningside Heights, New York City. Information about restaurants is provided either via an overview 3D map, so that the user can be guided to a specific place of his or her choice, or as direct annotations of the actual restaurant locations themselves. Having selected an establishment, the user can bring up a popup window with further information on it: a brief description, address and phone number, an image of the interior and, accessible at a mouse click: the menu and, if available, reviews of the restaurant, and its external web page. We will discuss related information display issues in Section 9.4.1.

Geographical field work. Field workers in geography and regional sciences could use AR techniques to collect, compare, and update survey data and statistics in the field (Nusser *et al.*, 2003). By assisting data collection and display, an AR system could enable discovery of patterns in the field, not just the laboratory. Instant verification and comparison of information with data on file would be possible.

Journalism. Journalism is another area in which MARS could have a major impact. Pavlik (2001) discusses the use of wireless technology for the mobile journalist, who covers and documents a developing news story on the run. MARS could be used to leave notes in the scene for other collaborating journalists and photographers to view and act upon. The *Situated Documentaries* project at Columbia University (Höllerer *et al.*, 1999a), shown in Figures 9.4 and 9.5, is a collaboration between computer science and journalism, and uses MARS for storytelling and presentation of historical information.

Architecture and archaeology. AR is also especially useful to visualize the invisible: architects' designs of bridges or buildings that are about to be constructed on a particular site; historic buildings, long torn down, in their original location; or reconstructions of archaeological sites. Figure 9.5, which was shot through a see-through head-worn display presenting a situated documentary (Höllerer *et al.*, 1999a) of the history of Columbia, shows a model of part of the Bloomingdale Insane Asylum, which once occupied the main campus. The European sponsored project ARCHEOGUIDE (Vlahakis *et al.*, 2002) aims to reconstruct a cultural heritage site in AR and let visitors view and learn about the ancient architecture and



(a) (b) (c)

Figure 9.4 Situated documentaries. (a) User with backpack MARS. (b) View through head-worn display. Virtual flags representing points of interest. (c) Additional handheld interface for interacting with virtual material.

customs. The first place selected as a trial site is the ancient town of Olympia in Greece.

Urban modeling. AR will not be used solely for passive viewing or information retrieval via the occasional mouse-button (or should we say shirt-button) click. Many researchers are exploring how AR technologies could be used to enter information to the computer (Rekimoto *et al.*, 1998). One practical example is 3D modeling of outdoor scenes: using the mobile platform to create 3D renderings of buildings and other objects that model the very environment to be used later as a backdrop for AR presentations (Baillot *et al.*, 2001; Piekarski and Thomas, 2001).

Entertainment. The situated documentaries application also suggests the technology's potential for entertainment purposes. Instead of delivering 3D movie "rides", such as the popular *Terminator 2* presentation at Universal Studios, to audiences in special purpose theme park theatres, virtual actors in special effects scenes could one day populate the very streets of the theme parks, engaging AR outfitted guests in spectacular action. As an early start in this direction, several researchers have experimented with applying mobile AR technology to gaming (Thomas *et al.*, 2000; Starner *et al.*, 2000).

Medicine. Augmented reality has important application possibilities in medicine. Many of these, such as surgery support systems that assist surgeons in their operations via live overlays (Fuchs *et al.*, 1998), require very precise registration, but do not require that the surgeon be extremely mobile while supported by the AR system. There are, however, several possible applications of mobile AR in the medical field. In the hospital or nursing home, doctors or nurses on their rounds of visits to the patients could get important information about each

Figure 9.5 Situated documentaries: Historic building overlaid at its original location on Columbia's campus.

patient's status directly delivered to their glasses (Hasvold, 2002). Out in the field, emergency medicine personnel could assess a situation quicker with wearable sensing and AR technology. They could apply the wearable sensors to the patient and would, from then on, be able to check the patient's status through AR glasses, literally at one glance. Also, a remote expert at a distant hospital could be brought into the loop and communicate with the field worker via the AR system, seeing through camera feeds what the field worker is seeing, which could be important to prepare an imminent operation at the hospital.

Monitoring the health information of a group of people at the same time could be advantageous for trainers or coaches during athletic training or competition. The military also has potential medical uses for mobile AR technologies. The health status of soldiers on the battlefield could be monitored, and in case of any injuries, the commanding officer could get live overview visualizations of location and status of the wounded.

Military training and combat. Military research led to the development of satellite navigation systems and heads-up displays for combat pilots. Military research laboratories have also been exploring the potential of mobile AR technology for land warriors for some time now (Tappert *et al.*, 2001). In terms of the possible use of AR in military operations, there is considerable overlap with civilian applications on a general level. Navigational support, enhancement of communications, repair and maintenance, and emergency medicine, are important topics in civilian and military life. There are, however, specific benefits that AR technology could bring to the military user. Most missions take place in unfamiliar territories. Map views, projected naturally into a warrior's limited view of the battle scene, can provide additional information about terrain that cannot easily be overseen. Furthermore, reconnaissance data and mission planning information can be integrated into these information displays, clarifying the situation and outlining specific sub-missions for individual troops. Ongoing research at the Naval Research Laboratory is concerned with how such information displays can be delivered to the warriors most effectively (Julier *et al.*, 2000). Apart from its use in combat, mobile AR might also prove a valuable military tool for training and simulation purposes. For example, large-scale combat scenarios could be tested with simulated enemy action in real training environments.

Personal Information Management and Marketing. It is anybody's guess which endeavors in mobile AR might eventually lead to commercial success stories. Some small companies already offer specialized AR solutions (TriSense). Whatever might trigger widespread use, the biggest potential market for this technology could prove to be personal wearable computing. AR could serve as an advanced and immediate UI for wearable computing. In personal, daily use, AR could support and integrate common tasks, such as email and phone communication with location-aware overlays, provide navigational guidance, enable individuals to store personal information coupled with specific locations, and provide a unified control interface for all kinds of appliances in the home (Feiner, 2002). Of course, such a personal platform would be very attractive for direct marketing agencies. Stores could offer virtual discount coupons to passing pedestrians. Virtual billboards could advertise products based on the individual's profile. Virtual 3D product prototypes could pop up in the customer's eyewear (Zhang and Navab, 2000). To protect the individual from unwanted information, an AR platform would need to incorporate appropriate filtering and view management mechanisms (see Section 9.4.3 and Chapter 10).

9.2.2 Challenges

In spite of the great potential of mobile AR in many application areas, progress in the field has so far almost exclusively been demonstrated through research prototypes. The time is obviously not quite ripe yet for commercialization. When asking for the reasons why, one has to take a good look at the dimension of the task. While increasingly better solutions to the technical challenges of wearable computing are being introduced, a few problem areas remain, such as miniaturization of input/output technology, power sources, and thermal dissipation, especially in small high-performance systems. Ruggedness is also required. With some early wearable systems the phrase 'wear and tear' seemed to rather fittingly indicate the dire consequences of usage. In addition to these standard wearable computing requirements, mobile AR adds many more: reliable and ubiquitous wide area position tracking, accurate and self-calibrating (head-) orientation tracking; ultra-light, ultra-bright, ultra-transparent, display eyewear with wide field of view; fast 3D graphics capabilities, to name just the most important requirements. We will look at current technologies addressing these problem areas in the following section. If one were to select the best technologies available today for each of the necessary components, one could build a powerful (though large) MARS. Section 9.3.7 will take a closer look at such a hypothetical device.

AR in the outdoors is a particular challenge since there is a wide range of operating conditions to which the system could be exposed. Moreover, in contrast to controlled environments indoors, one has little influence over outdoor conditions; for example, lighting can range from direct sunlight, possibly exacerbated by a reflective environment (e.g., snow), to absolute darkness without artificial light sources during the night. Outdoor systems should withstand all possible weather conditions, including wind, rain, frost, and heat.

The list of challenges does not end with the technology on the user's side. Depending on the tracking technology, AR systems either need to have access to a model of the environment they are supposed to annotate, or require that such environments be *prepared* (e.g., equipped with visual markers or electronic tags). Vision-based tracking in unprepared environments is currently not a viable general solution, but research in this field is trying to create solutions for future systems. The data to be presented in AR overlays needs to be paired with locations in the environment. A standard access method needs to be in place for retrieving such data from databases responsible for the area the MARS user is currently passing through. This requires mechanisms such as automatic service detection and the definition of standard exchange formats that both the database servers and the MARS software support. It is clear from the history of protocol standards, that without big demand and money-making opportunities on the horizon, progress on these fronts can be expected to be slow. On the other hand, the World Wide Web, HTML, and HTTP evolved from similar starting conditions. Some researchers see location-aware computing on a global scale as a legitimate successor of the World Wide Web as we know it today (Spohrer, 1999).

In the movie industry, special effects seamlessly merge computer-generated worlds with real scenes. Currently these efforts take days and months of rendering time and very carefully handcrafted integration of the virtual material into real-world footage. In AR, not only does the rendering need to be performed in real time, but the decisions about what to display and the generation of this material must be triggered and controlled on the fly. Making the visuals as informative as possible, and, in some cases, also as realistic as possible, rendered with the correct lighting to provide a seamless experience, is an open-ended challenge for visual AR. Section 9.4 will present ongoing research in sorting out some UI challenges.

9.3 COMPONENTS AND REQUIREMENTS

In this section, we review in greater depth the basic components and infrastructure required for MARS, as outlined before in Section 9.1.3. We take a look at mobile computing platforms, displays for MARS, tracking and registration issues, environmental modeling, wearable input and interaction techniques, wireless communication, and distributed data storage and access. We give brief overviews of important historic developments in these areas, and point out technologies that have successfully been employed in MARS prototypes, or that have great potential to be employed in future systems. Finally, we summarize this material by describing a hypothetical top-of-the-line MARS, assembled from the most promising components currently available.

9.3.1 Mobile Computing Platforms

Mobile computing platforms have seen immense progress in miniaturization and performance over recent years, and are sold for increasingly less. Today, high-end notebook computers catch up in computing power with available desktop solutions very shortly after a new processor model hits the market. The trend towards more mobility is clearly visible. The wearable technology market, even though still in its infancy, has a growing customer base in industry, government, and military. Wearable computing solutions for personal use can now be purchased from various sources.

There are several decision factors when choosing a computing platform for mobile AR research, including the *computing power* needed, the *form factor* and *ruggedness* of the overall system, *power consumption*, the *graphics and multimedia capabilities*, availability of *expansion and interface ports*, available *memory and storage space*, *upgradeability* of components, the *operating system* and *software development environments*, availability of *technical support*, and last but not least, *price*. Quite clearly, many of these are interdependent. The smaller the computing device, the less likely it is to have the highest computing power and graphics and multimedia capabilities. Expansion and interface ports normally come at the price of increased size. So does upgradeability of components: if you have miniaturized functionality (e.g., graphics) by using special purpose integrated circuits, you no longer have the luxury of being able to easily replace that component with a newer model. Additionally, it takes a lot of effort and ingenuity to scale down any kind of technology to a size considerably smaller than what the competition is offering, so one can expect such equipment to be sold at a higher price.

The Columbia University MARS project (Höllerer *et al.*, 1999b) provides a concrete example of some of the tradeoffs involved in assembling a mobile AR platform. The hardware platform for this project, illustrated in Figures 1 (a) and 2 (a), was assembled from off-the-shelf components for maximum performance, upgradeability, ease of maintenance, and software development. These choices were made at the cost of the size and weight of the prototype system, whose parts were mounted on a backpack frame. From 1996 to 2002, every single component of the prototype was upgraded multiple times to a more powerful or otherwise more advantageous version, something that would not have been possible if a smaller, more integrated system had initially been chosen. The computing power tracked high-end mobile processing technologies, ranging from a 133MHz Pentium-based system in 1996 to a 2.2 GHz Pentium IV notebook in 2002. During the same time, the 3D graphics capabilities grew from a GLINT 500DTX chip with a claimed fill rate of about 16.5M pixels per second to an NVIDIA Quadro4 500 Go with announced 880M pixels per second.

The smallest wearable computing platforms currently available (Windows CE-based Xybernaut Poma, or the soon-to-be-available higher performance OQO Ultra-Personal Computer, Tiqit eightythree, and Antelope Technologies Mobile Computer Core) provide only modest graphics performance and their main processors do not have enough power for software renderings of complex 3D scenes at interactive speeds. Decision factors in choosing a 3D graphics platform for mobile AR include the *graphics performance* required, *video and texture memory*, *graphics library* support (OpenGL or Direct-X), availability of *stereo* *drivers*, *power consumption*, and *price*. The most practical solution for a mobile AR system that can support complex 3D interactive graphics comes in the form of small notebook computers with integrated 3D graphics chip. The display could be removed if the computer is exclusively used with a near-eye display. However, in our experience with the Touring Machine, it can be put to good use in prototype systems for debugging purposes and for providing a view for onlookers during technology demonstrations.

Specific application requirements can drastically limit the choices for a MARS computing platform. For example, in late 2002 there are no integrated wearable computing solutions available that support rendering and display of complex graphical scenes in stereo. A system designer targeting such applications either has to assemble their own hardware to create a small form-factor solution, or resort to the smallest available power notebook that has sufficient graphics performance and a graphics chip supporting stereo.

Mobile systems do not necessarily have to follow the pattern of one standalone device, carried or worn, that generates and presents all the information to the user. Instead, there can be varying degrees of "environment participation," making use of resources that are not necessarily located on the user's body. In the most device-centric case, all information is generated and displayed on one single device that the user wears or carries. Examples include portable audio players and hand-held organizers without wireless communication option.

Departing one step from the device-centric approach, functionality can be distributed over multiple devices. Wireless connectivity technologies, such as IEEE 802.11b or Bluetooth come in handy for data exchange between different devices. For example, a personal organizer or wearable computer can send data over a wireless connection to an audio/video headset. With the addition of a Bluetooth-enabled cell phone for global communication purposes, such a combination would constitute a complete wireless mobile computing solution. Not all devices need to be carried by the user at all times. Suppose that we want to minimize the wearable computing equipment's size and power consumption. Lacking the computational power to generate and process the information that is to be displayed, we can turn the wearable device into a so-called *thin client*, relying on outside servers to collect and process information and feed it to the portable device as a data stream that can be comfortably presented with the limited resources available.

In the extreme case, a mobile user would not need to wear or carry *any* equipment and still be able to experience mobile AR. All the computation and sensing could occur in the environment. A grid of cameras could be set up so that multiple cameras would cover any possible location the person could occupy. Information could be stored, collected, processed, and generated on a network of computing servers that would not need to be in view, or even nearby. Displays, such as loudspeakers, video-walls, and projected video could bring personalized information to the people that are standing or moving nearby. However, the infrastructure needed for such a scenario is quite high. The task of making mobile AR work in unprepared environments requires solutions closer to the "one-device" end of the spectrum.

9.3.2 Displays for Mobile AR

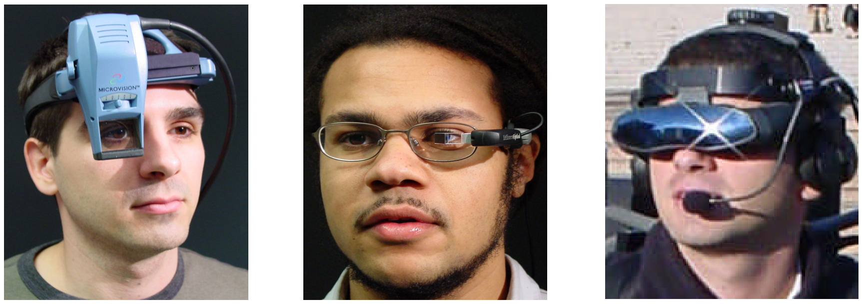
There are various approaches to display information to a mobile person and a variety of different types of displays can be employed for this purpose: personal hand-held, wrist-worn, or head-worn displays; screens and directed loudspeakers embedded in the environment; image projection on arbitrary surfaces, to name but a few. Several of these display possibilities may also be used in combination.

In general, one can distinguish between displays that the person carries on the body and displays that make use of resources in the environment. Wearable audio players and personal digital organizers use displays that fall in the first category, as do wearable computers with head-worn displays (see Figure 9.6). An example of the second category would be personalized advertisements that are displayed on video screens that a person passes. For such a scenario, one would need a fairly sophisticated environmental infrastructure. Displays would need to be embedded in walls and other physical objects, and they would either have to be equipped with sensors that can detect a particular individual's presence, or they could receive the tracking information of passersby via a computer network. Such environments do not yet exist outside of research laboratories, but several research groups have begun exploring ubiquitous display environments as part of *Smart Home* or *Collaborative Tele-Immersion* setups, such as the Microsoft Research EasyLiving project (Brumitt *et al.*, 2000), or the University of North Carolina Office of the Future (Raskar *et al.*, 1998).

Another display type that is being explored in AR research is the *head-worn projective display* (Hua *et al.*, 2001). This type of head-worn display consists of a pair of micro displays, beam splitters, and miniature projection lenses. It requires that retroreflective sheeting material be placed strategically in the environment. The head-worn display projects images out into the world, and only when users look at a patch of retroreflective material, they see the image that was sent out from their display. This approach aims to combine the concept of physical display surfaces (in this case: patches of retroreflective material) with the flexibility of personalized overlays with AR eyewear. A unique personalized image can be generated for each person in a set of people looking at the same object with retroreflective coating, as long as their viewing angles are not too close to each other.

One promising approach for mobile AR might be to combine different display technologies. Head-worn displays provide one of the most immediate means of accessing graphical information. The viewer does not need to divert his or her eyes away from their object of focus in the real world. The immediateness and privacy of a personal head-worn display is complemented well by the high text readability of hand-held plasma or LCD displays, and by the collaboration possibilities of wall-sized displays. For example, mobile AR research at Columbia University experimented early on with head-worn and hand-held displays (cf. Figure 9.2) used in synergistic combination (Feiner *et al.*, 1997; Höllerer *et al.*, 1999a).

As mentioned in the historical overview in Section 9.1.2, the concept of *see-through head-worn computer graphics displays* dates back to Ivan Sutherland's work on a head-worn 3D display (Sutherland, 1968). Some time before that, in 1957, Morton Heilig had filed a patent for a head-worn display fitted with two color TV units. In later years, several head-worn displays were developed for research in computer simulations and the military, including Tom Furness's work on heads-up display systems for fighter pilots. VPL Research and Autodesk



(a) (b) (c)

Figure 9.6 Monocular and binocular optical see-through head-worn displays. (a) Microvision Nomad. (b) MicroOptical Clip-on. (c) Sony LDI-D100B (retrofit onto a customer-provided mount).

introduced a commercial head-worn display for VR in 1989. In the same year, Reflection Technology introduced a small personal near-eye display, the P4 Private Eye. This display is noteworthy, because it gave rise to a number of wearable computing and AR and VR efforts in the early 1990s (Pausch, 1991, Feiner *et al.*, 1991). It sported a resolution of 720x280 pixels, using a dense column of 280 red LEDs and a vibrating mirror. The display was well suited for showing text and simple line drawings.

RCA made the first experimental liquid crystal display (LCD) in 1968, a non-emissive technology (requiring a separate light source) that steadily developed and later enabled a whole generation of small computing devices to display information. Today, many different technologies, especially emissive ones, are being explored for displays of a wide variety of sizes and shapes. *Plasma displays* provide bright images and wide viewing angles for medium-sized to large flat panels. *Organic light emitting diodes* (OLED) can be used to produce ultra-thin displays. Certain types of OLED technology, such as *light emitting polymers*, might one day lead to display products that can be bent and shaped as required. Of high interest for the development of personal displays are display technologies that are so small, that optical magnification is needed to view the images. These are collectively referred to as *microdisplays.* OLED on silicon is one of the most promising approaches to produce such miniature displays. Non-emissive technologies for microdisplays include transmissive *poly-silicon LCDs*, and several reflective technologies, such as *liquid crystal on silicon* (LCoS) and *digital micromirror devices* (DMD).

One technology that is particularly interesting for mobile AR purposes is the one employed in Microvision’s monochromatic, single-eye, Nomad *retinal scanning display*, shown in Figure 9.6 (a). It is one of the few displays that can produce good results in direct sunlight outdoors. It works by pointing a red laser diode towards an electromagnetically controlled pivoting micromirror and diverting the beam via an optical combiner through the viewer’s pupil into the eye, where it sweeps across the retina to recreate the digital image. This technology produces a very crisp and bright image, and exhibits the highest transparency any optical see-through display offers today. Microvision has also prototyped a much larger, full-color and optionally stereoscopic display.

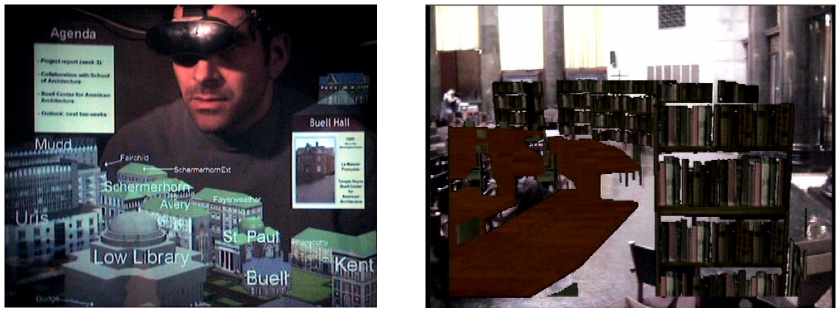


Figure 9.7 (a) Optical see-through and (b) Video see through indoor AR.

When choosing a head-worn display for mobile AR, several decision factors have to be considered. One of the more controversial issues within the AR research community is the choice between optical see-through and video see-through displays. Optical see-through displays are transparent, the way prescription glasses or sunglasses are. They use optical combiners, such as mirror beam-splitters, to layer the computer generated image on top of the user's view of the environment. Figure 9.7 (a) shows an image shot through such glasses. In contrast, video see-through displays present a more indirect, mediated view of the environment. One or two small video cameras, mounted on the head-worn display, capture video streams of the environment in front of the user, which are displayed on non-transparent screens with suitable optics, right in front of the user's eyes. The computer can modify the video image before it is sent to the glasses to create AR overlays. An example is shown in Figure 9.7 (b). More details and a discussion of the advantages and disadvantages of both approaches are given in Azuma (1997) and Feiner (2002). Non-AR wearable computing applications often use monocular displays. Even if the display is non-transparent, the user is able to see the real world with the non-occluded eye. However, perceiving a true mixture of computer overlay and real world can be somewhat of a challenge in that case.

For mobile AR work, the authors of this chapter prefer optical see-through displays. We believe that a person walking around in the environment should be able to rely on their full natural sense of vision. While AR can enhance their vision, it should not unduly lessen it. In our opinion, several drawbacks of current video see-through technology stand in the way of their adoption in truly mobile applications: Seeing the real world at video resolution and at the same small field of view angle used for the graphical overlays, having to compensate for image distortions introduced by the cameras, the risk of latency in the video feed to the display, and safety concerns about seeing the world solely through cameras.

In our experience, monocular displays can yield acceptable results for AR if the display is see-through to make it easier for the user to fuse the augmented view with the other eye's view of the real world, as is the case with the MicroVision Nomad. A larger field of view is also helpful. It is hard to discuss such display properties in isolation, however, since quite a few display factors influence the quality of mobile AR presentations, among them monocular vs. biocular (two-eye) vs. binocular (stereo), resolution, color depth, luminance, contrast, field of view, focus depth, degree of transparency, weight, ergonomics, and appearance. Power consumption is an additional factor with extended mobile use.

Stereo displays can greatly enhance the AR experience, since virtual objects can then be better perceived at the same distance as the real world objects they annotate. Note though, that even though stereo allows objects to be displayed with the correct left/right eye disparity, all currently available displays display graphics at the same apparent depth, and hence require the viewer's eyes to accommodate at that particular distance, which leads to an accommodation-vergence conflict. For example, when overlaying a virtual object on a real one in an optical see-through display, unless the real object is located at that particular fixed distance, the viewer needs to adjust accommodation in order to see either the real object or the virtual one in focus.

Currently, the options for optical see-through head-worn displays are quite limited. If stereo is a necessity, the options are even more restricted. The Columbia University MARS prototypes employed several stereo capable optical see-through displays over the years, none of which are on the market anymore. Figure 9.6 (c) shows the Sony LDI-D100B, a display that was discontinued in June 2000.

Displays are for the most part still bulky and awkward in appearance today. Smaller monocular displays, such as the MicroOptical CO-1, pictured in Figure 9.6 (b), or the Minolta 'Forgettable Display' prototype (Kasai *et al.*, 2000), are much more inconspicuous, but do not afford the high field-of-view angles necessary for true immersion nor the brightness of, for example, the Microvision Nomad. Meanwhile, manufacturers are working hard on improving and further miniaturizing display optics. Microdisplays can today be found in a diverse set of products including viewfinders for cameras, displays for cell phones and other mobile devices, and portable video projectors. Near-eye displays constitute a growing application segment in the microdisplay market. The attractiveness of mobile AR relies on further progress in this area.

9.3.3 Tracking and Registration

Apart from the display technology, the single most important technological challenge to general mobile AR is tracking and registration. AR requires extremely accurate position and orientation tracking to align, or register, virtual information with the physical objects that are to be annotated. It is difficult to convince people that computer-generated virtual objects actually live in the same physical space as the real world objects around us. In controlled environments of constrained size in indoor computing laboratories, researchers have succeeded in creating environments in which a person's head and hands can be motion-tracked with sufficiently high spatial accuracy and resolution, low latency, and high update rates, to create fairly realistic interactive computer graphics environments that seemingly coexist with the physical environment. Doing the same in a general mobile setting is much more challenging. In the general mobile case, one cannot expect to rely on any kind of tracking infrastructure in the environment. Tracking equipment needs to be light enough to wear, fairly resistant to shock and abuse, and functional across a wide spectrum of environmental conditions, including lighting, temperature, and weather. Under these circumstances, there does not currently exist a perfect tracking solution, nor can we expect to find one in the near future. Compromises in tracking performance have to be made, and applications will have to adjust.

Tracking technology has improved steadily since the early days of head-tracked computer graphics. Sutherland's original head-worn display was tracked mechanically through ceiling-mounted hardware, and, because of all the equipment suspended from the ceiling, was humorously referred to as the "Sword of Damocles." Sutherland also explored the use of an ultrasonic head-tracker (Sutherland, 1968). The introduction of the Polhemus magnetic tracker in the late 1970s (Raab *et al*., 1979) had a big impact on VR and AR research, and the same technology, in improved form, is still in use today. During the 1990s, commercial hybrid tracking systems became available, based on different technologies, all explored separately in experimental tracking systems over the previous decades, such as ultrasonic, magnetic, and optical position tracking, and inertial and magnetometer-based orientation tracking. With respect to global positioning systems, the idea for today's *NAVSTAR* GPS (Getting, 1993) was born in 1973, the first operational GPS satellite was launched in 1978, and the 24-satellite constellation was completed in 1993. Satellites for the Russian counterpart constellation, *Glonass*, were launched from 1982 onwards. The European Union has plans underway to launch a separate 30-satellite GPS, called *Galileo*. Chapter 4 gives a detailed introduction to GPS technology.

In the remainder of this section, we will review the tracking technologies most suited for mobile AR. For a more comprehensive overview of tracking technologies for AR and VR, we refer the reader to existing surveys of motion tracking technologies and techniques, such as Rolland *et al.* (2001), or a recent journal special issue on tracking (Julier and Bishop, 2002).

Visual registration of virtual and physical objects can be achieved in several ways. One common approach is to determine the person's head pose in some global coordinate system, and relate it to a computer model of the current environment. Note that in this case a computer model of the environment has to be created in a step called *environmental modeling* (see Section 9.3.4). This model should use the same global coordinate system as the tracking system, or the necessary conversion transformation has to be known. Determining position and orientation of an object is often referred to as six-degree-of-freedom (6DOF) tracking, for the six parameters sensed: position in x, y, and z, and orientation in yaw, pitch, and roll angles.

Absolute position and orientation of the user's head and the physical objects to be annotated do not necessarily need to be known. In one of the most direct approaches to visual registration, cameras observe specific unique landmarks (e.g., artificial markers) in the environment. If the camera's viewing parameters (position, orientation, field of view) coincide with the display's viewing parameters (e.g., because the display is showing the camera image, as in the case of video see-through displays), and stereo graphics are not employed, the virtual annotations can be inserted directly in pixel coordinates without having to establish the exact geometric relationship between the marker and the camera (Rekimoto and Nagao, 1995). On the other hand, if the precise locations of the landmarks in the environment are known, computer vision techniques can be used to estimate the camera pose. The use of cameras mounted on the display together with landmark recognition is sometimes referred to as *closed-loop tracking,* in which tracking accuracy can be corrected to the nearest pixel, if camera image and graphics display coincide. This is in contrast to *open-loop tracking*, which tries to align the virtual annotations with the physical objects in the real world by relying solely on the sensed 6DOF pose of the person and the computer model of the environment. Any inaccuracies in the tracking devices or the geometrical model will cause the annotation to be slightly off from its intended position in relation to the physical world.

An important criterion for mobile AR tracking is how much tracking equipment is needed on the user's body and in the environment. The obvious goal is to wear as little equipment as possible, and to not be required to prepare the environment in any way. Note that a system such as GPS meets this requirement for all intended purposes, even though a "prepared environment" on a global scale is needed in the form of a satellite constellation. Several tracking approaches require some knowledge about the environment. To create any useful AR annotations, either the objects to be annotated have to be modeled or *geo-referenced* in absolute terms, or their location must be able to be inferred by a known relationship to pre-selected and identifiable landmarks.

The tracking accuracies required for mobile AR depend very much on the application and the distance to the objects to be annotated. If we are annotating the rough outlines of buildings, we can afford some registration error. When trying to pinpoint down the exact location of a particular window, we have to be more accurate. When registration errors are measured as the screen distance between the projected physical target point and the point where the annotation gets drawn, the following observation holds: The further away the object that is to be annotated, the less errors in position tracking impact registration accuracy. The opposite is true for errors in orientation tracking. Since most targets in outdoor mobile AR tend to be some distance away from the viewer, one can assume that errors in orientation tracking contribute much more to overall misregistration than do errors in position tracking (Azuma, 1999). Since there are no standalone sensors that afford general reliable 6DOF tracking in unprepared outdoor environments, mobile AR systems normally resort to hybrid approaches, often employing separate mechanisms for position and orientation tracking.

Position tracking via GPS is a natural candidate for outdoor environments, since it is functional on a global scale, as long as signals from at least four satellites can be received. While the use of GPS navigation has long been restricted to areas that afford direct visibility to the satellites, so-called *assisted GPS* (A-GPS) manages to sidestep that restriction in many cases. A-GPS makes use of a world-wide reference network of servers and base stations for terrestrial signal broadcast. In combination with a large number of parallel correlation reception circuits in the mobile GPS receiver, the area of tracking can be extended to many previously uncovered areas, such as urban canyons and indoor environments in which the signal is sufficiently strong (GlobalLocate, 2002).

Plain GPS without selective availability is accurate to about 10–15 meters. GPS using the wide area augmentation system (WAAS) is typically accurate to 3–4 meters in the US and other countries that adopt this technology. Differential GPS typically yields a position estimate that is accurate to about 1–3 meters with a local base station. Real-time–kinematic GPS (RTK GPS) with carrier-phase ambiguity resolution can produce centimeter-accurate position estimates. The latter two options require the existence of a nearby base station from which a differential error-correction signal can be sent to the roaming unit. Therefore, one cannot really speak of an unprepared environment anymore in that case. Commercial differential services are available, however, with base stations covering most of North America. For a long time, commercial differential GPS receivers provided update rates of up to 5Hz, which is suboptimal for tracking fast motion of people or objects. Newer products provide update rates of up to 20 Hz. (Trimble, 2002). More details on GPS and other global tracking systems can be found in Chapter 4.

Another position tracking system applicable for wide-area mobile AR involves calculating a person's location from signal quality measures of IEEE 802.11b (WiFi) wireless networking. This obviously also requires the deployment of equipment in the environment, in this case the WiFi access points, but if such a wireless network is the communication technology of choice for the mobile AR system, the positioning system can serve as an added benefit. Several research projects, and at least one commercial product, are exploring this concept. The RADAR system uses multilateration and precomputed signal strength maps for this purpose (Bahl and Padmanabhan, 2000), while Castro and colleagues (2001) employ a Bayesian networks approach. The achievable resolution depends on the density of access points deployed to form the wireless network. Ekahau (Ekahau, 2002) offer a software product that allows position tracking of WiFi enabled devices after a manual data collection/calibration step.

Two additional means of determining position are often employed in MARS, mostly as part of hybrid tracking systems: Inertial sensors and vision-based approaches. Accelerometers and gyroscopes are self-contained or *sourceless* inertial sensors. Their main problem is drift. The output of accelerometers needs to be integrated once with respect to time, in order to recover velocity, and twice to recover position. Hence, any performance degradations in the raw data lead to rapidly increasing errors in the resulting position estimate. In practice, this approach to position estimation can only be employed for very small time periods between updates gathered from a more reliable source. Inertial sensors can also be used to detect the act of a pedestrian taking a step. This is the functional principle of *pedometers*, which, when combined with accurate heading information, can provide a practical *dead-reckoning* method (Point Research, 2002; Höllerer *et al.*, 2001b).

Vision-based approaches are a promising option for 6DOF pose estimation in a general mobile setting. One or two tiny cameras are mounted on the glasses, so that the computer can approximately see what the user sees. Model-based vision techniques require an accurate model of the environment with known landmarks that can be recognized in the image feeds. In contrast, move-matching algorithms track dynamically chosen key points along the image sequence, leading to relative, rather than absolute, tracking solutions, which means that further registration of the image sequence coordinate system with the physical world needs to be established to enable 3D graphical overlays. Simultaneous reconstruction of the camera motion and scene geometry is possible, but such computations are highly computationally expensive, and existing algorithms require a "batch bundle adjustment," a global offline computation over the entire image sequence. Finally, 2D image-based feature tracking techniques measure so-called "optical flow" between subsequent video images. Such techniques are comparatively fast, but by themselves cannot estimate 3D camera motion. Combinations of all these approaches are possible. Recent research reports promising results for some test scenarios (Julier and Bishop, 2002). However, in general, computer vision algorithms still lack robustness and require such high amounts of computation that pure vision solutions for general-case real-time tracking are still out of reach. For the time being, hybrids of vision based tracking and other sensing technologies show the biggest promise.

Orientation tracking also benefits greatly from hybrid approaches. The basic technologies available for orientation sensing are electromagnetic compasses (magnetometers), gravitational tilt sensors (inclinometers), and gyroscopes (mechanical and optical). Hybrid solutions have been developed, both as commercial products and research prototypes. The IS300 and InertiaCube2 orientation sensors by InterSense (2002) combine three micro-electromechanical gyroscopes, three accelerometers (for motion prediction), and an electromagnetic compass in one small integrated sensor. Azuma and colleagues (1999) presented a hybrid tracker that combines a carefully calibrated compass and tilt sensor with three rate gyroscopes. You and colleagues (1999) extended that system by a move-matching vision algorithm, which did not, however, run in real time. Behringer (1999) presented a vision-based correction method based on comparing the silhouette of the horizon line with a model of local geography. Satoh and colleagues (2001) employed a template-matching technique on manually selected landmarks in a real-time algorithm that corrects for the orientation drift of a highly accurate fiber optic gyroscope (Sawada *et al.*, 2001).

In summary, the problem of tracking a person's pose for general mobile AR purposes is a hard problem with no single best solution. Hybrid tracking approaches are currently the most promising way to deal with the difficulties posed by general indoor and outdoor mobile AR environments.

9.3.4 Environmental Modeling

For AR purposes, it is often useful to have access to geometrical models of objects in the physical environment. As mentioned above, one use of such models is in registration. If you want to annotate a window in a building, the computer has to know where that window is located with regard to the user's current position and field of view. Having a detailed hierarchical 3D model of the building, including elements such as floors, rooms, doors, and windows, gives the computer flexibility in answering such questions. Some tracking techniques, such as the model-based computer vision approaches mentioned in Section 9.3.3, rely explicitly on features represented in more or less detailed models of the tracking environment. Geometrical computer models are also used for figuring out *occlusion* with respect to the observer's current view. For example, if portions of a building in front of the observer are occluded by other objects, only the non-occluded building parts should be annotated with the building's name to avoid confusing the observer as to which object is annotated (Bell *et al.*, 2001).

For the purposes mentioned so far, an environment model does not need to be photorealistic. One can disregard materials, textures, and possibly even geometric detail. In fact, in model-based tracking, often only a "cloud" of unconnected 3D sample points is used. More realistic geometric models of real-world structures, such as the ones depicted in Figure 9.8, are often used for annotation purposes, or for giving an overview of the real environment. For example, a building that is occluded from the user's view can be displayed in its exact hidden location via AR, enabling the user, in a sense, to see through walls.

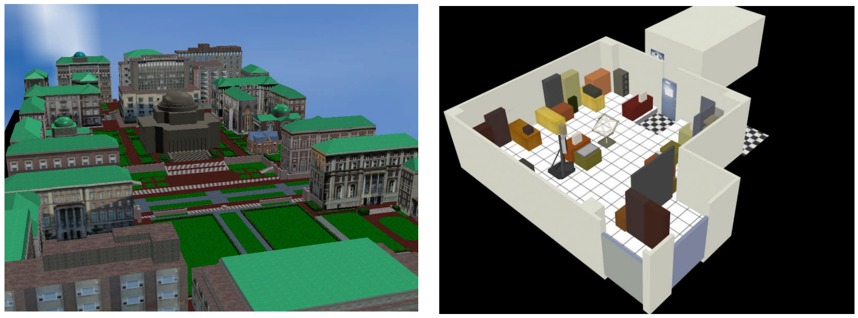


Figure 9.8 Environmental modelling: (a) Model of a campus. (b) Model of a laboratory.

Somebody looking for a specific building can be shown a virtual version of it on the AR screen. Having gotten an idea of the building's shape and texture, the person might watch the model move off in the correct direction, until it coincides with the real building in physical space. A three-dimensional map of the environment can be presented to the user to give a bird's-eye overview of the surroundings (Stoakley *et al.*, 1995). Figures 9.9 and 9.10 show examples of such *worlds in miniature* (WIM) used in AR.

Creating 3D models of large environments is a research challenge in its own right. Automatic, semiautomatic, and manual techniques can be employed, among them 3D reconstruction from satellite imagery and aerial photographs, 3D imaging with laser range finders, reconstruction from a set of partly overlapping photographs, surveying with total stations and other telemetry tools, and manual reconstruction using 3D modeling software. Even AR itself can be employed for modeling purposes, as mentioned in Section 9.2.1. Abdelguerfi (2001) provides an overview of 3D synthetic environment reconstruction. The models in Figure 9.8 were reconstructed by extruding 2D map outlines of Columbia University's campus and our research laboratory, refining the resulting models by hand, and texture mapping them selectively with photographs taken from various strategic positions. The WIM of Figure 9.9 shows a 3D model of the currently selected building (Dodge Hall) in the context of an aerial photograph of the current environment.

3D spatial models can be arbitrarily complex. Consider, for example, the task of completely modeling a large urban area, down to the level of water pipes and electric circuits in walls of buildings. There are significant research problems involved in the modeling, as well as the organization and storage of such data in spatial databases and data structures optimized for specific queries. Finally, environmental modeling does not end with a static model of the geometry. Most environments are dynamic: changes in the geometric models (due to moving objects, construction, or destruction) need to be tracked, and reflected in the environmental model. The databases may need to change quite rapidly, depending on the level of detail considered.

Figure 9.9 Context-overview: World-in-Miniature map.

9.3.5 Wearable Input and Interaction Technologies

How to interact with wearable computers effectively and efficiently is another open research question. The desktop UI metaphor, often referred to as WIMP (windows, icons, menus, and pointing), is not a good match for mobile and wearable computing, mostly because it places unreasonable motor skill and attention demands on mobile users interacting with the real world.

As a general UI principle, AR can provide a user with immediate access to the physical world. Visual attention does not need to be divided between the task in the physical world and a separate computer screen. However, interacting seamlessly with such a computing paradigm is a challenge. In this section, we review interaction technologies that have been tried for MARS.

Basic interaction tasks that graphical UIs handle, include *selecting, positioning,* and *rotating* virtual objects, *drawing paths* or *trajectories,* assigning quantitative values, referred to as *quantification,* and *text input.* AR UIs deal as much with the physical world as with virtual objects. Therefore, selection, annotation, and, possibly, direct manipulation of physical objects also play an important role in these kind of UIs.

We already mentioned one class of input devices, namely the sensors that afford tracking and registration. Position tracking determines the user's locale, and head orientation tracking assists in figuring out the user's focus. Establishing the user's context in this fashion can effectively support user interaction. The UI can adapt to such input by limiting the choices for possible courses of action to a context-relevant subset. Both, position and head orientation tracking can also be employed for object selection. Suppose that the task is to select a building in an urban neighborhood. With position tracking only, the closest building, or a list of the *n* closest ones, might be listed on the display for direct selection via a button or scrolling input device. With head orientation, the user can point their head in the direction of the object to be selected. Selection can take place by dwelling for a certain time period on the object in view, or by active selection via button-like devices. Höllerer *et al.* (1999a) discuss several tracking-prompted selection mechanisms for a mobile AR system. Additional orientation trackers can provide hand tracking, which can be used to control pointers or manipulate virtual objects on the AR screen. Tracking hand or finger position for full 6DOF hand tracking, as is common in indoor virtual or augmented environments, would be a great plus for MARS, but is hard to achieve with mobile hardware in a general setting. Research prototypes for this purpose have experimented with vision-based approaches, and ultrasonic tracking of finger-worn acoustic emitters using three head-worn microphones (Foxlin and Harrington, 2000).

Quite a few mobile input devices tackle continuous 2D pointing. Pointing tasks, the domain of mice in desktop systems, can be performed in the mobile domain by trackballs, track-pads, and gyroscopic mice, many of which transmit data wirelessly to the host computer. It should be mentioned, however, that these devices are popular in large part because, lacking a better mobile UI standard, many researchers currently run common WIMP UIs on their mobile and wearable platforms. Accurate 2D pointing poses a big challenge for a mobile user's motor skills. However, 2D pointing devices can also be used to control cursor-less AR UIs (Feiner *et al.*, 1997). When user interaction mostly relies on discrete 2D pointing events (e.g., selecting from small lists of menu items), then small numeric keypads with arrow keys, or arrow keys only, might provide a solution that is more easily handled on the run, and more easily worn on the body.

Mobile UIs should obviously try to minimize encumbrance caused by UI devices. The ultimate goal is to have a free-to-walk, eyes-free, and hands-free UI with miniature computing devices worn as part of the clothing. As should be clear from our overview so far, this ideal cannot always be reached with current mobile computing and UI technology. Some devices, however, already nicely meet the size and ergonomic constraints of mobility. Auditory UIs, for example, can already be realized in a relatively inconspicuous manner, with small wireless earphones tucked into the ear, and microphones worn as part of a necklace or shoulder pad. There is a growing body of research on wearable audio UIs, dealing with topics such as speech recognition, speech recording for human-to-human interaction, audio information presentation, and audio dialogue. It is clear, however, that a standalone audio UI cannot offer the best possible solution for every situation. Noisy surroundings and environments that demand complete silence pose insurmountable problems to such an approach. On the other hand, audio can be a valuable medium for multimodal and multimedia UIs.

Other devices are more impractical for brief casual use, but have successfully been employed in research prototypes. *Glove-based* input devices, for example, using such diverse technologies as electric contact pads, flex sensors, accelerometers, and even force-feedback mechanisms, can recognize hand gestures, but have the drawback of looking awkward and impeding use of the hands in real-world activities. Nevertheless, the reliability and flexibility of glove gestures has made the computer glove an input device of choice for some MARS prototypes (Thomas and Piekarski, 2002). Starner *et al.* (1997b), on the other hand, explore vision-based hand gesture recognition, which leaves the hands unencumbered, but requires that a camera be worn on a hat or glasses, pointing down to the area in front of the user's body, in which hand gestures are normally made.

We already discussed the use of cameras for vision-based tracking purposes (Section 9.3.3). Apart from that purpose, and the potential of finger and hand gesture tracking, cameras can be used to record and document the user's view. This can be useful as a live video feed for teleconferencing, for informing a remote expert about the findings of AR field-workers, or simply for documenting and storing everything that is taking place in front of the MARS user. Recorded video can be an important element in human-to-human interfaces, which AR technology nicely supports.

A technology with potential for mobile AR is gaze tracking. Eye trackers observe a person's pupils with tiny cameras to determine where that person's gaze is directed. Drawbacks are the additional equipment that needs to be incorporated into the eyewear, the brittleness of the technology (the tracker needs to be calibrated and the cameras must be fixed with respect to the eye), and the overwhelming amount of involuntary eye movement that needs to be correctly classified as such. With the right filters, however, gaze control could provide a very fast and immediate input device. As a pointing device, it could eliminate the need for an entire step of coordinated muscle activity that other pointing devices require in order to move a pointer to a location that was found through eye movement in the first place. Even without gaze control, gaze tracking provides a dynamic history of where a user's attention is directed. As computers gather more and more such knowledge about the user's interests and intentions, they can adapt their UIs better to suit the needs of the current context (see Section 9.4.1).

Other local sensors that can gather information about the user's state include biometric devices that measure heart-rate and bioelectric signals, such as galvanic skin response, electroencephalogram (neural activity), or electromyogram (muscle activity) data. Employing such monitored biological activity for computer UI purposes is an ambitious research endeavor, but the hopes and expectations for future applicability are quite high. *Affective computing* (Picard, 1997) aims to make computers more aware of the emotional state of their users and able to adapt accordingly.

As we can see, UI technology can be integrated with the user more or less tightly. While the previous paragraph hinted at possible future human-machine symbioses (Licklider, 1960), current wearable computing efforts aim to simply make computing available in as unencumbered a form as possible. One item on this agenda is to make clothes more computationally aware; for example, by embroidering electronic circuits (Farringdon *et al.*, 1999). On the other hand, not every UI technology needs to be so tightly integrated with the user. Often, different devices that the user would carry, instead of wear on the body, can support occasionally arising tasks very efficiently. For example, hand-held devices such as palmtop or tablet computers are good choices for reading text, assuming high-contrast and high-resolution displays, and are well suited for pen-based input, using handwriting recognition and marking gestures. *Hybrid user interfaces*, as Feiner and colleagues (1997) explored them for mobile AR purposes, aim to employ different display and input technologies and reap the benefits of each technology for the purposes for which it is best suited. The applicability of a wide variety of input technologies is utilized nicely by *multimodal* UI techniques. Such techniques employ multiple input and output modes in time-synchronized combination (e.g., gestures, speech, vision, sound, and haptics), using different media to present the user with a more natural and robust, yet still predictable, UI.

Finally, fast and reliable text input to a mobile computer is hard to achieve. The standard keyboard, which is the proven solution for desktop computing, requires too much valuable space and a flat typing surface. Small, foldable, or even inflatable keyboards, or virtual ones that are projected by a laser onto a flat surface, are current commercial options or product prototypes. Chording keyboards, which require key combinations to be pressed to encode a single character, such as the one-handed Twiddler2 (Handykey, 2001), are very popular choices for text input in the wearable computing community. Cell phones provide their own alphanumeric input techniques via a numeric keypad. We already mentioned handwriting recognition, pen-based marking, and speech recognition, which experienced major improvements in accuracy and speed over the last decade, but cannot be applied in all situations. Soft keyboards enable text input via various software techniques, but use valuable display screen space for that purpose. Glove-based and vision-based hand gesture tracking do not provide the ease of use and accuracy necessary for serious adoption yet. It seems likely that speech input and some kind of fallback device (e.g., pen-based systems, or special purpose chording or miniature keyboards) will share the duty of providing text input to mobile devices in a wide variety of situations in the near future.

9.3.6 Wireless Communication and Data Storage Technologies

We already discussed the mobility of a computing system in terms of its size, ergonomics, and input/output constraints. Another important question is how connected such a system is in the mobile world. This question concerns the electronic exchange of information with other, mobile or stationary, computer systems. The degree of connectivity can vary from none at all to true global wireless communication. Most likely is a scenario where the mobile client has different options to get connected to the internet, currently ranging in the area covered and connection speed from a fast direct cable connection (when used in a stationary office environment) to slightly slower wireless local area networks (WLANs), which offer full connectivity in building- or campus-sized networks of wireless access points, to wireless phone data connections with nationwide or international coverage, but much slower transmission speeds.

The first packet-based WLAN was ALOHANET at the University of Hawaii in 1971. Today, WLANs provide bandwidths ranging between 2 and 54 Mbps, and are quite common for providing coverage in campuses and homes. At least one US telecommunications consortium has plans for nationwide support of IEEE 802.11b (WiFi) networks (Cometa, 2002). During the first two years of the current century, US phone service providers began to roll out new nationwide networks based on third generation wireless technology (at bandwidths of 144 Kbps, and higher in some selected test areas), which nicely complement smaller sized community WLANs.

For close-range point-to-point connections between different devices, the Bluetooth consortium (Bluetooth, 1998) has established an industry standard for low-power radio frequency communication. Using this technology, wearable computers connect with input/output devices that a user can carry or wear on the body, or walk up to, as in the case of stationary printers. Bluetooth-enabled cellular phones provide access to nationwide wireless connectivity whenever faster networking alternatives are not available.

From the perspective of MARS, the integration of LBS with communication systems is an important issue. Whereas it might be sufficient for special purpose AR systems to store all related material on the client computer, or retrieve it from one single task-related database server, this is not true anymore in the mobile case. For true mobility, the AR client will need to connect to multiple distributed data servers in order to obtain the information relevant to the current environment and situation. Among the data that need to be provided to the client computer are the geometrical models of the environment (see Section 9.3.4), annotation material (object names, descriptions, and links), as well as conceptual information (object categorization and world knowledge) that allows the computer to make decisions about the best ways how to present the data. Some formalism is needed to express and store such meta-knowledge. Markup languages, such as various XML derivatives, are well suited for this purpose. XML offers the advantages of a common base language that different authors and user groups can extend for their specific purposes.

For interactive applications, as required by AR technology, as much as possible of the data that is to be displayed in world overlays should be stored, or cached on the local (worn or carried) client computer. This raises the question of how to upload and "page in" information about new environments that the mobile user is ready to roam and might want to explore. Such information can be loaded preemptively from distributed databases in batches of relative topical or geographical closure (e.g., all restaurants in a certain neighborhood close to the user's current location). We would like to emphasize that currently no coherent global data repository and infrastructure exists that would afford such structured access to data. Instead, different research groups working on mobile AR applications have established their own test infrastructures for this purpose. For example, in our own AR work, data and meta-data is stored and accessed dynamically in relational databases, and distributed to various clients via a data replication infrastructure. The database servers effectively turn into *AR servers*, responsible for providing the material used by the client for overlays to particular locations.

Research from grid computing, distributed databases, middleware, service discovery, indexing, search mechanisms, wireless networking, and other fields will be necessary to build examples of new communication infrastructures that enable such semantically prompted data access. The Internet offers the backbone to experiment with such data distribution on virtually any level of scale.

9.3.7 Summary: A Top-of-the-line MARS Research Platform

Now that we have reviewed the technological and data requirements of mobile AR, here is the summary of a hypothetical MARS research platform made from the most promising components that are available today to researchers. Some of the components currently easily exceed all but the most generous budgets. However, prices keep falling steadily, and new technologies enter the market at a rapid pace.

The type of base unit we would select for our MARS depends on the focus of our applications. If detailed 3D graphics is not a strict requirement, we would pick a small wearable computer. According to current product announcements, we will be able in early 2003 to buy hand-held computers with the processing power of a 1GHz Transmeta Crusoe chip, 256 MB main memory, about 10–20 GB disk space, and a full arsenal of I/O interfaces (e.g., OQO Ultra-Personal Computer, Tiqit eightythree). If 3D graphics is a must, we would settle for a (larger) notebook computer with integrated 3D graphics chip. Either solution would come equipped or could be extended with WiFi wide area and Bluetooth personal networking. For nationwide connectivity outside of WiFi networks, we would use a Bluetooth-enabled cell phone with a 3G service plan.

For a display we would pick a Microvision retinal scanning display: monocular, monochromatic, and far larger and heavier than desired, but sufficiently transparent indoors and bright outdoors.

For coarse position tracking, we would use an A-GPS receiver, employing the new chip technology that promises to achieve signal reception even in commonly untracked areas, such as beneath thick foliage, or indoors. Position tracking using WiFi signal quality measures is also a possibility. Higher-precision position tracking indoors is dependent on the availability of special purpose tracking infrastructures. If higher-precision outdoor position tracking is a must, we additionally use an RTK GPS receiver in areas equipped with an RTK GPS base station. For orientation tracking, we would choose fiber optic gyroscope (FOG) sensors (Sawada, 2001) in combination with a magnetometer, if the budget allows it. Small FOG-based sensors, customized for use as a head tracker, can currently cost up to $100,000. A much more affordable and smaller, but far less accurate, backup solution would be a small hybrid inertial and magnetometer-based sensor, such as the Inertiacube2 by InterSense (2002). Ideally, these position and orientation tracking technologies should be used to provide first guess for state-of-the-art vision-based tracking, which requires the addition of one or more tiny cameras.

We select a large set of input devices, so that we can make use of different technologies in different situations: Bluetooth-enabled earphones and microphone, wrist-worn keypad, Twiddler2 chord keyboard and mouse, and, if the selected base computer does not have a pen-operated screen already, or is too big for occasional hand-held use, an additional palm-sized tablet computer/display.

We will also need rechargeable batteries for powering all these devices. By relying on built-in batteries and two additional lithium-ion batteries, we can realistically expect such a system to have an up time of about three hours. For extended operation, we would need to add additional batteries.

9.4 MARS UI CONCEPTS

As described in the previous section, one significant impediment to the immediate widespread use of MARS UIs is technological. However, if we look at the progress in the field over the past ten years, and extrapolate into the future, we can be optimistic that many of the current hardware flaws will be resolved. In the remainder of this chapter, we will assume a faultless MARS device, as far as hardware and tracking is concerned, and take a look at the UI concepts of mobile

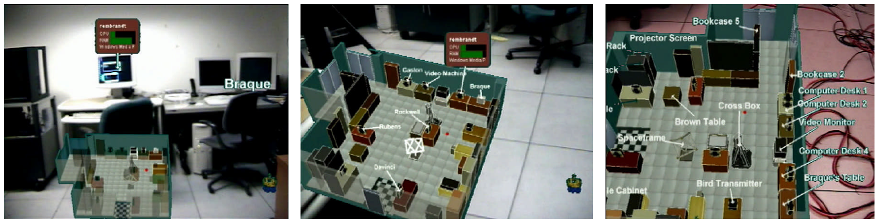
(a) (b)

Figure 9.10 World-stabilized and screen-stabilized UI elements.

AR, as facilitated by the software. The kind of computing that MARS make possible is quite different from the current static work and play environments of our offices and homes. In contrast, the world is the interface, which means that in dealing with a MARS we will rarely focus exclusively on the computer anymore. In fact, while we go about our day-to-day activities, we would not even be able, let alone want, to pay attention to the computer. At the same time, however, we will expect the system to provide assistance and augmentation for many of our tasks. Broll and colleagues (2001) describe a futuristic scenario of using such a mobile helper UI.

9.4.1 Information Display and Interaction Techniques

AR allows the user to focus on computer-supplied information and the real world at the same time. A UI for visual mobile AR can combine screen-stabilized, body-stabilized, and world-stabilized elements (Feiner *et al.*, 1993; Billinghurst *et al.*, 1998). Figure 9.10 shows a UI from the authors' mobile AR work (Höllerer *et al.*, 1999a), photographed through optical see-through head-worn displays. The virtual flags and labels are *world-stabilized* objects, residing in the world coordinate system, denoting points of interest in the environment. They are displayed in the correct perspective for the user's viewpoint, so the user can walk up to and around these objects just like physical objects. The labels face the user and maintain their size irrespective of distance to ensure readability. The blue and green menu bars on the top are *screen-stabilized*, meaning that they occupy the same position on the screen no matter where the user is looking, as is the cone-shaped pointer at the bottom of the screen, which is always pointing towards the currently selected world object. The two images of Figure 9.10 (a) and (b) show the same *in-place* menu options associated with the red flag in front of the columns, realized in two different ways. In part (a), our initial implementation, the menu was arranged in a circular world-stabilized fashion around the flag. This caused problems when the user turned his or her head during menu selection. In the design shown in part (b), the menu is a *screen-stabilized* element, linked back to its associated flag by a



(a) (b) (c)

Figure 9.11 Head-pitch control of WIM. (a) User looking straight ahead. (b) WIM scales up, tilts, and moves up the screen as user tilts head downwards. (c) Near top-down view of WIM as user looks further down.

leader line, so that the user can easily turn back to it. We made the menu semi-transparent, so that the view of other virtual elements is not completely obstructed.

Body-stabilized information, unsurprisingly, is stored relative to the user's body, making it accessible at a turn of the head, independent of the user's location. Note, that in order to store virtual objects relative to the body with respect to yaw (e.g., consistently to the user's left), the body's orientation needs to be tracked in addition to head orientation. One can extend the notion of body stabilized objects to using general head-gestures for virtual object control. Figure 9.11 shows a WIM displayed in front of a mobile user who views the scene through a head-worn display. In part (a) the user is looking straight ahead. As the user looks down towards the ground, the WIM is shown in successively more detail and from an increasingly top-down perspective (parts b and c). Note that this is not a body-stabilized element: the user's head orientation alone is used to control the WIM. In this case, the WIM is always kept visible on the screen, aligned in yaw with the surrounding environment it represents, with head pitch used to control the WIM's pitch, size, position, and level of annotation (Bell *et al.*, 2002).

Mobile AR agrees well with the notion of non-command interfaces (Nielsen, 1993), in which the computer is reacting to sensed user context rather than explicit user commands. For a person trying to focus on a real-world task, and not on how to work a particular computer program, it is desirable that computers understand as much as possible about the task at hand without explicitly being told. Often, much of the needed interaction can be reduced to the user's answering several prompted questions (Pascoe *et al.*, 2000). Some tasks, however, such as placing or moving virtual objects in the environment or modeling them in the first place from physical examples, require extended user interaction with the AR UI. UIs that have been tried for such tasks range from a 2D cursor and head motion (Baillot *et al.*, 2001), to a tracked glove (Thomas and Piekarski, 2002), to a tracked graphics tablet on which UI elements can be overlaid (Reitmayr and Schmalstieg, 2001). Simple interaction with virtual material has also been achieved using vision-based hand tracking (Kurata *et al.*, 2001).

Mobile AR UIs invite collaboration. Several users can discuss and point to virtual objects displayed in a shared physical space (Butz *et al.*, 1999; Reitmayr and Schmalstieg, 2001). At the same time, every participant can see their own private version of the shared data, for example to see annotations optimized for their specific viewing angle (Bell *et al.*, 2001). Multiple users can collaborate in the field, and remote experts with a top-down overview of the user’s environment can communicate and share information with the field worker (Höllerer *et al.*, 1999b).

9.4.2 Properties of MARS UIs

Mobile AR presents a way for people to interact with computers that is radically different from the static desktop or mobile office. One of the key characteristics of MARS is that both virtual and physical objects are part of the UI, and the dynamic context of the user in the environment can influence what kind of information the computer needs to present next. This raises several issues:

Control: Unlike a stand-alone desktop UI, where the only way the user can interact with the presented environment is through a set of well defined techniques, the MARS UI needs to take into account the unpredictability of the real world. For example, a UI technique might rely on a certain object being in the user's field of view and not occluded by other information. Neither of the properties can be guaranteed: the user is free to look away, and other information could easily get in the way, triggered by the user's own movement or an unforeseen event (such as another user entering the field of view). Thus, to be effective, the UI technique either has to relax the non-occlusion requirement, or has to somehow guarantee non-occlusion in spite of possible contingencies.

Consistency: People have internalized many of the laws of the physical world. When using a computer, a person can learn the logic of a new UI. As long as these two worlds are decoupled (as they are in the desktop setting), inconsistencies between them are often understandable. In the case of MARS, however, we need to be very careful to design UIs in which the physical and virtual world are consistent with each other.

Need for embedded semantic information: In MARS, virtual material is overlaid on top of the real world. Thus we need to establish concrete semantic relationships between virtual and physical objects to characterize UI behavior. In fact, since many virtual objects are designed to annotate the real world, these virtual objects need to store information about the physical objects to which they refer (or at least have to know how to access that information).

Display space: In terms of the available display space and its best use, MARS UIs have to deal with a much more complicated task compared to traditional 2D UIs. Instead of one area of focus (e.g., one desktop display), we have to deal with a potentially unlimited display space surrounding the user, only a portion of which is visible at any time. The representation of that portion of augmented space depends on the user's position, head orientation, personal preferences (e.g., filter settings) and ongoing interactions with the augmented world, among other things. Management of virtual information in this space is made even more difficult by constraints that other pieces of information may impose. Certain virtual or physical objects may, for example, need to be visible under all circumstances, and thus place restrictions on the display space that other elements are allowed to obstruct.

The display management problem is further complicated by the possibility of taking into account multiple displays. MARS, as a nonexclusive UI to the augmented world, may seamlessly make use of other kinds of displays, ranging from wall-sized, to desk-top, to hand-held. If such display devices are available and accessible to the MARS, questions arise as to which display to use for what kind of information and how to let the user know about that decision.

Scene dynamics: In a head-tracked UI, the scene will be much more dynamic than in a stationary UI. In MARS, this is especially true, since in addition to all the dynamics due to head motion, the system has to consider moving objects in the real world that might interact visually or audibly with the UI presented on the head-worn display. Also, we have to contend with a potentially large variability in tracking accuracy over time. Because of these unpredictable dynamics, the spatial composition of the UI needs to be flexible and the arrangement of UI elements may need to be changed. On the other hand, traditional UI design wisdom suggests minimizing dynamic changes in the UI composition (Shneiderman, 1998).

One possible solution to this dilemma lies in the careful application of automated UI management techniques.

9.4.3 UI Management

In our own work on MARS, we adapt and simplify the UI through a set of management techniques, including the following steps: information filtering, UI component design, and view management.

The large amount of virtual information that can be displayed, coupled with the presence of a richly complex physical world, creates the potential for clutter. Cluttered displays can overwhelm the user with unneeded information, impacting her ability to perform her tasks effectively. Just as in desktop information visualization (Shneiderman, 1998), we address clutter through information filtering. For our MARS work, *information filtering* (Julier *et al.*, 2000) means the act of culling the information that can potentially be displayed by identifying and prioritizing what is relevant to a user at a given point in time. The priorities can be based on the user's tasks, goals, interests, location, or other user context or environmental factors.

While information filtering determines the subset of the available information that will be displayed, it is still necessary to determine the format in which this information is to be communicated, and how to realize that format in detail. Registration accuracy, or how accurately the projected image of a virtual object can be positioned, scaled, and oriented relative the real world, is an important factor in choosing the right UI format. Registration accuracy is determined by tracking system accuracy, which, as the mobile user moves about, may vary for a variety of reasons that depend on the tracking technologies used. Therefore, if information is always formatted in a way that assumes highly accurate registration, that information will not be presented effectively when registration accuracy decreases. To address this issue, *UI component design* (Höllerer *et al.*, 2001b) determines the format in which information should be conveyed, based on contextual information, such as the available display resources and tracking accuracy. This technique determines the concrete elements that comprise the UI and information display.

Filtering and formatting information is not enough—the information must be integrated with the user's view of the physical world. For example, suppose that annotations are simply projected onto the user's view of the world such that each is colocated with a physical object with which it is associated. Depending on the user's location in the world (and, thus, the projection that they see), annotations might occlude or be occluded by other annotations or physical objects, or appear ambiguous because of their proximity to multiple potential referents. *View management* (Bell *et al.*, 2001) attempts to ensure that the displayed information is arranged appropriately with regard to the projections on the view plane of it and other objects; for example, virtual or physical objects should not occlude others that are more important, and relationships among objects should be as unambiguous as possible.

More detail about this suggested UI management pipeline can be found in Höllerer *et al.* (2001a).

9.5 CONCLUSIONS

In this chapter, we presented an overview of the field of mobile AR, including historical developments, future potential, application areas, challenges, components and requirements, state-of-the-art systems, and UI concepts. AR and wearable computing are rapidly growing fields, as exemplified by the soaring number of research contributions and commercial developments since the mid 1990s. We have reached an important point in the progress toward mobile AR, in that the available technology is powerful enough for an increasing number of impressive research prototypes, but not yet sufficiently reliable, general, and comfortable for mass adoption. Compared to other applications described in this book, which are immediately realizable using today's technology, it will take more time for mobile AR to reach the computing mainstream. However, mobile AR will have an enormous impact when it becomes commonplace.

We are looking forward to further progress in the areas of computing hardware miniaturization, battery design, display technology, sensor technology, tracking accuracy and reliability, general vision-based tracking and scene understanding, and overall comfort. We anticipate the emergence of distributed data infrastructures for context-based computing in general, and AR in particular, leading to much improved data access capabilities for mobile users.

Finally, we hope that the benefits of mobile AR will be achieved without compromising privacy and comfort (Feiner, 1999). Research and development in a field that could have a significant impact on social structures and conventions should be accompanied by careful consideration of how the commendable aspects of our social equilibrium can be protected and strengthened.

9.6 ACKNOWLEDGMENTS

The research described here is funded in part by ONR Contracts N00014-99-1-0249, N00014-99-1-0394, and N00014-99-1-0683, NSF Grant IIS-00-82961, and gifts from Intel, Microsoft Research, Mitsubishi Electric Research Laboratories, and Pacific Crest. We would also like to thank Blaine Bell, Drexel Hallaway, Ryuji Yamamoto, Hrvoje Benko, Sinem Güven, Tiantian Zhou, Alex Olwal, Gabor Bláskó, Simon Lok, Elias Gagas, Tachio Terauchi, and Gus Rashid for their invaluable contributions to Columbia University's MARS infrastructure over the years.

9.7 REFERENCES

Abdelguerfi, M. 2001. *3D Synthetic Environment Reconstruction*. Kluwer Academic Publishers, New York, NY.

Azuma, R. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385.

Azuma, R. 1999. The challenge of making augmented reality work outdoors. In Ohta, Y. and Tamura, H., editors, *Mixed Reality, Merging Real and Virtual Worlds*, pages 379–390. Ohmsha/Springer, Tokyo/New York.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. 2001. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47.

Azuma, R., Hoff, B., Neely III, H., and Sarfaty, R. 1999. A motion-stabilized outdoor augmented reality system. In *Proc. IEEE Virtual Reality '99*, pages 242–259.

Bahl, P. and Padmanabhan, V. 2000. RADAR: an in-building RF-based user location and tracking system. In *Proc. IEEE Infocom 2000*, pages 775–784, Tel Aviv, Israel.

Baillot, Y., Brown, D., and Julier, S. 2001. Authoring of physical models using mobile computers. In *Proc. ISWC '01 (Fifth IEEE Int. Symp. on Wearable Computers)*, pages 39–46, Zürich, Switzerland.

Behringer, R. 1999. Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors. In *Proc. IEEE Virtual Reality '99*, pages 244–251.

Bell, B., Feiner, S., and Höllerer, T. 2001. View management for virtual and augmented reality. In *Proc. ACM UIST 2001 (Symp. on User Interface Software and Technology)*, pages 101–110, Orlando, FL. (CHI Letters, vol. 3, no. 2).

Bell, B., Höllerer, T., and Feiner, S. 2002. An annotated situation-awareness aid for augmented reality. In *Proc. ACM UIST 2002 (Symp. on User Interface Software and Technology)*, pages 213–216, Paris, France.

Billinghurst, M., Bowskill, J., Dyer, N., and Morphett, J. 1998. Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications*, 18(6):24–31.

Broll, W., Schäfer, L., Höllerer, T., and Bowman, D. 2001. Interface with angels: The future of VR and AR interfaces. *IEEE Computer Graphics and Applications*, 21(6):14–17.

Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S. 2000. Easyliving: Technologies for intelligent environments. In *Proc. Second Int. Symp. on Handheld and Ubiquitous Computing (HUC 2000)*, pages 12–29, Bristol, UK. Springer Verlag.

Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C. 1999. Enveloping users and computers in a collaborative 3D augmented reality. In *Proc. IWAR '99 (IEEE and ACM Int. Workshop on Augmented Reality)*, pages 35–44, San Francisco, CA.

Castro, P., Chiu, P., Kremenek, T., and Muntz, R. 2001. A probabilistic room location service for wireless networked environments. In *Proc. ACM UbiComp 2001: Ubiquitous Computing*, volume 2201 of *Lecture Notes in Computer Science*, pages 18–35.

Caudell, T. P. and Mizell, D. W. 1992. Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes. In *Proceedings of 1992 IEEE Hawaii International Conference on Systems Sciences*. IEEE Press.

Cheverst, K., Davies, N., Mitchell, K., and Blair, G. S. 2000. Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In *Proceedings of ACM CHI' 00*, Netherlands.

Cometa Networks 2002. Wholesale nationwide broadband wireless internet access. http://www.cometanetworks.com.

Drascic, D. and Milgram, P. 1996. Perceptual issues in augmented reality. *Proc. SPIE: Stereoscopic Displays and Virtual Reality Systems III*, 2653:123–134.

Ekahau 2002. Location in wireless networks. http://www.ekahau.com.

Ericsson, IBM, Intel, Nokia, and Toshiba 1998. Bluetooth mobile wireless initiative. http://www.bluetooth.com.

Farringdon, J., Moore, A. J., Tilbury, N., Church, J., and Biemond, P. D. 1999. Wearable sensor badge and sensor jacket for context awareness. In *Proc. ISWC '99 (Third Int. Symp. on Wearable Computers)*, pages 107–113, San Francisco, CA.

Feiner, S. 1999. The importance of being mobile: Some social consequences of wearable augmented reality systems. In *Proc. IWAR '99 (IEEE and ACM Int. Workshop on Augmented Reality)*, pages 145–148, San Francisco, CA.

Feiner, S. 2002. Augmented Reality: A new way of seeing. *Scientific American*, 286(4):48–55.

Feiner, S., MacIntyre, B., Haupt, M., and Solomon, E. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proc. UIST '93 (ACM Symp. on User Interface Software and Technology)*, pages 145–155, Atlanta, GA.

Feiner, S., MacIntyre, B., and Höllerer, T. 1999. Wearing it out: First steps toward mobile augmented reality systems. In Ohta, Y. and Tamura, H., editors, *Mixed Reality: Merging Real and Virtual Worlds*, pages 363–377. Ohmsha (Tokyo)–Springer Verlag, Berlin.

Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. 1997. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Proc. ISWC '97 (First IEEE Int. Symp. on Wearable Computers)*, pages 74–81, Cambridge, MA.

Feiner, S. and Shamash, A. 1991. Hybrid user interfaces: virtually bigger interfaces for physically smaller computers. In *Proc. ACM UIST '91*, pages 9–17. ACM press.

Foxlin, E. and Harrington, M. 2000. Weartrack: A self-referenced head and hand tracker for wearable computers and portable vr. In *Proc. ISWC '00 (Fourth IEEE Int. Symp. on Wearable Computers)*, pages 155–162, Atlanta, GA.

Fuchs, H., Livingston, M., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J., Rademacher, P., Drake, S., and Meyer, A. 1998. Augmented Reality Visualization for Laparoscopic Surgery. In *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*.

Furmanski, C., Azuma, R., and Daily, M. 2002. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Proc. ISMAR '02 (Int. Symposium on Augmented Reality)*, pages 215–224, Darmstadt, Germany.

Furness, T. 1986. The super cockpit and its human factors challenges. In *Proc. Human Factors Society 30th Annual Meeting*, pages 48–52, Santa Monica, CA.

Getting, I. 1993. The global positioning system. *IEEE Spectrum*, 30(12):36–47.

Global Locate 2002. Global Locate/Fujitsu GL-16000 Indoor GPS chip. http://www.globallocate.com/.

Handykey 2001. Handykey Twiddler2 chord keyboard. http://www.handykey.com.

Hasvold, P. 2002. In-the-field health informatics. In *The Open Group conference*, Paris, France. http://www.opengroup.org/public/member/q202/documentation/plenary/hasvold.pdf.

Höllerer, T., Feiner, S., Hallaway, D., Bell, B., Lanzagorta, M., Brown, D., Julier, S., Baillot, Y., and Rosenblum, L. 2001a. User interface management techniques for collaborative mobile augmented reality. *Computers and Graphics*, 25(5):799–810.

Höllerer, T., Feiner, S., and Pavlik, J. 1999a. Situated documentaries: Embedding multimedia presentations in the real world. In *Proc. ISWC '99 (Third IEEE Int. Symp. on Wearable Computers)*, pages 79–86, San Francisco, CA.

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

Höllerer, T., Hallaway, D., Tinna, N., and Feiner, S. 2001b. Steps toward accommodating variable position tracking accuracy in a mobile augmented reality system. In *2nd Int. Workshop on Artificial Intelligence in Mobile Systems (AIMS '01)*, pages 31–37.

Hua, H., Gao, C., Brown, L.D., Ahuja, N., and Rolland, J.P. 2001. Using a head-mounted projective display in interactive augmented environments. In *Proc. ISAR 2001 (IEEE and ACM Int. Symp. on Augmented Reality 2001)*, pages 217–223, New York, NY.

InterSense 2002. InterSense Inc., IS-900 Wide Area Precision Motion Tracker, IS300 and InertiaCube2 orientation sensors. http://www.isense.com.

Ishii, H. and Ullmer, B. 1997. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proc. Conference on Human Factors in Computing Systems (CHI '97), (Atlanta, March 1997)*, pages 234–241. ACM Press.

Julier, S. and Bishop, G., editors 2002. *IEEE Computer Graphics and Applications, Special Issue on Tracking*, volume 6(22), pages 22–80. IEEE Computer Society.

Julier, S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., Höllerer, T., and Sestito, S. 2000. Information filtering for mobile augmented reality. In *Proc. ISAR '00 (IEEE and ACM Int. Symposium on Augmented Reality)*, pages 3–11, Munich, Germany.

Kasai, I., Tanijiri, Y., Endo, T., and Ueda, H. 2000. A forgettable near eye display. In *Proc. ISWC '00 (Fourth Int. Symp. on Wearable Computers)*, pages 115–118, Atlanta, GA.

Klinker, G., Creighton, O., Dutoit, A. H., Kobylinski, R., Vilsmeier, C., and Brügge, B. 2001. Augmented maintenance of powerplants: a prototyping case study of a mobile AR system. In *Proc. ISAR '01 (IEEE and ACM Int. Symposium on Augmented Reality)*, pages 124–133, New York, NY.

Kurata, T., Okuma, T., Kourogi, M., Kato, T., and Sakaue, K. 2001. VizWear: Toward human-centered interaction through wearable vision and visualization. *Lecture Notes in Computer Science*, 2195:40–47.

Licklider, J. 1960. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics,* HFE-1, 4–11.

Loomis, J., Golledge, R., and Klatzky, R. 1993. Personal guidance system for the visually impaired using GPS, GIS, and VR technologies. In *Proc. Conf. on Virtual Reality and Persons with Disabilities*, Millbrae, CA.

Mann, S. 1997. Wearable computing: A first step toward personal imaging. *IEEE Computer*, 30(2):25–32.

Mizell, D. 2001. Boeing's wire bundle assembly project. In Barfield, W. and Caudell, T., editors, *Fundamentals of Wearable Computers and Augumented Reality*, pages 447–467. Lawrence Erlbaum Assoc, Mahwah, NJ.

Nielsen, J. 1993. Noncommand user interfaces. *Communications of the ACM*, 36(4):82–99.

Nusser, S. M., Miller, L. L., Clarke, K., and Goodchild, M. F. 2003. Geospatial IT for mobile field data collection. *Communications of the ACM,* 46(1):63–64.

Pascoe, J., Ryan, N., and Morse, D. 2000. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*, 7(3):417–437.

Pausch, R. 1991. Virtual reality on five dollars a day. *Proc. ACM CHI '91 Conference on Human Factors in Computing Systems*, pages 265–270.

Pavlik, J. V. 2001. *Journalism and New Media*. Columbia University Press, New York, NY.

Petrie, H., Johnson, V., Strothotte, T., Raab, A., Fritz, S., and Michel, R. 1996. MoBIC: Designing a travel aid for blind and elderly people. *Jnl. of Navigation*, 49(1):45–52.

Picard, R. W. 1997. *Affective Computing*. MIT Press, Cambridge, MA.

Piekarski, W. and Thomas, B. 2001. Tinmith-Metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In *Proc. ISWC '01 (Fifth IEEE Int. Symp. on Wearable Computers)*, pages 31–38, Zürich, Switzerland.

Point Research Corporation 2002. Dead reckoning module DRM-III. Accurate navigation for personnel on foot. http://www.pointresearch.com.

Raab, F. H., Blood, E. B., Steiner, T. O., and Jones, H. R. 1979. Magnetic position and orientation tracking system. *IEEE Transaction on Aerospace and Electronic Systems*, AES-15(5):709–717.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. 1998. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proc. SIGGRAPH '98*, pages 179–188.

Reitmayr, G. and Schmalstieg, D. 2001. Mobile collaborative augmented reality. In *Proc. ISAR '01 (Int. Symposium on Augmented Reality)*, pages 114–123, New York, NY.

Rekimoto, J., Ayatsuka, Y., and Hayashi, K. 1998. Augment-able reality: Situated communication through physical and digital spaces. In *Proc. ISWC '98 (Second Int. Symp. on Wearable Computers)*, pages 68–75, Cambridge, MA.

Rekimoto, J. and Nagao, K. 1995. The world through the computer: Computer augmented interaction with real world environments. In *Proc. ACM Symposium on User Interface Software and Technology*, Virtual and Augmented Realities, pages 29–36.

Rolland, J. P., Davis, L. D., and Baillot, Y. 2001. A survey of tracking technologies for virtual environments. In Barfield, W. and Caudell, T., editors, *Fundamentals of Wearable Computers and Augumented Reality*. Lawrence Erlbaum Assoc, Mahwah, NJ.

Sato, K., Ban, Y., and Chihara, K. 1999. MR aided engineering: Inspection support systems integrating virtual instruments and process control. In Ohta, Y. and Tamura, H., editors, *Mixed Reality, Merging Real and Virtual Worlds*, pages 347–361. Ohmsha/Springer, Tokyo/New York.

Satoh, K., Anabuki, M., Yamamoto, H., and Tamura, H. 2001. A hybrid registration method for outdoor augmented reality. In *Proc. ISAR '01 (Int. Symposium on Augmented Reality)*, pages 57–76, New York, NY.

Sawada, K., Okihara, M., and Nakamura, S. 2001. A wearable attitude measurement system using a fiber optic gyroscope. In *Proc. ISMR '01 (Second Int. Symp. on Mixed Reality)*, pages 35–39, Yokohama, Japan.

Sawhney, N. and Schmandt, C. 1998. Speaking and listening on the run: Design for wearable audio computing. In *Proc. ISWC '98 (Second Int. Symp. on Wearable Computers)*, pages 108–115, Cambridge, MA.

Shneiderman, B. 1998. *Designing the User Interface*. Addison-Wesley, third edition.

Spohrer, J. 1999. Information in places. *IBM Systems Journal*, 38(4):602–628.

Starner, T., Leibe, B., Singletary, B., and Pair, J. 2000. MIND-WARPING: towards creating a compelling collaborative augmented reality game. In *Proc. Int. Conf. on Intelligent User Interfaces (IUI '00)*, pages 256–259.

Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R., and Pentland, A. 1997a. Augmented reality through wearable computing. *Presence*, 6(4):386–398.

Starner, T., Weaver, J., and Pentland, A. 1997b. A wearable computing based american sign language recognizer. In *Proc. ISWC '97 (First Int. Symp. on Wearable Computers)*, pages 130–137, Cambridge, MA.

Stoakley, R., Conway, M., and Pausch, R. 1995. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings of Human Factors in Computing Systems (CHI '95)*, pages 265–272.

Sutherland, I. 1968. A head-mounted three dimensional display. In *Proc. FJCC 1968*, pages 757–764, Washington, DC. Thompson Books.

Tappert, C. C., Ruocco, A. S., Langdorf, K. A., Mabry, F. J., Heineman, K. J., Brick, T. A., Cross, D. M., and Pellissier, S. V. 2001. Military applications of wearable computers and augmented reality. In Barfield, W. and Caudell, T., editors, *Fundamentals of Wearable Computers and Augumented Reality*, pages 625–647. Lawrence Erlbaum Assoc, Mahwah, NJ.

Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., and Piekarski, W. 2000. ARQuake: An outdoor/indoor augmented reality first person application. In *Proc. ISWC '00 (Fourth Int. Symp. on Wearable Computers)*, pages 139–146, Atlanta, GA.

Thomas, B. H. and Piekarski, W. 2002. Glove based user interaction techniques for augmented reality in an outdoor environment. *Virtual Reality: Research, Development, and Applications*, 6(3):167–180. Springer-Verlag London Ltd.

Thorp, E. O. 1998. The invention of the first wearable computer. In *Proc. ISWC '98 (Second Int. Symposium on Wearable Computers)*, pages 4–8.

Trimble Navigation Ltd 2002. Trimble MS750 RTK receiver for precise dynamic positioning. http://www.trimble.com/ms750.html.

Vlahakis, V., Ioannidis, N., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., Gleue, T., Daehne, P., and Almaida, L. 2002. Archeoguide: An augmented reality guide for archaeological sites. *IEEE Computer Graphics and Applications*, 22(5):52–59.

Weiser, M. 1991. The computer for the 21st century. *Scientific American*, 3(265):94–104.

Zhang, X., Genc, Y., and Navab, N. 2001. Taking AR into large scale industrial environments: Navigation and information access with mobile computers. In *Proc. ISAR '01 (Int. Symposium on Augmented Reality)*, pages 179–180, New York, NY.

Zhang, X., Navab, N., and Liou, S.-P. 2000. E-Commerce direct marketing using augmented reality. In *Proc. ICME2000 (IEEE Int. Conf. on Multimedia and Exposition)*, New York, NY.