Situated Documentaries:
Embedding Multimedia Presentations in the Real World

Tobias Höllerer
Steven Feiner
Dept. of Computer Science
Columbia University
New York, NY 10027
{*htobias,feiner*}*@cs.columbia.edu*

John Pavlik
Center for New Media
Graduate School of Journalism
Columbia University
New York, NY 10027
jp35@columbia.edu

Abstract

We describe an experimental wearable augmented reality system that enables users to experience hypermedia presentations that are integrated with the actual outdoor locations to which they are are relevant. Our mobile prototype uses a tracked see-through head-worn display to overlay 3D graphics, imagery, and sound on top of the real world, and presents additional, coordinated material on a hand-held pen computer. We have used these facilities to create several situated documentaries that tell the stories of events that took place on our campus. We describe the software and hardware that underly our prototype system and explain the user interface that we have developed for it.

1. Introduction

Mobile and wearable computing systems provide users access to computational resources even when they are away from the static infrastructure of their offices or homes. One of the most important aspects of these devices is their potential to support *location-aware* or *location-based* computing, offering services and information that are relevant to the user's current locale [1]. Research and commercial location-aware systems have explored the utility of a variety of coarse position-tracking approaches, ranging from monitoring infrared signals emitted by "active badges" [23], to relying on wireless paging cell size to provide local weather and traffic updates [18].

Augmented reality, which demands far more accurate position tracking combined with accurate orientation tracking, can provide an especially powerful user interface for location-aware mobile computing. By supplementing the real world with virtual information, augmented reality can substantially enrich the user's experience of her environment and present her with an integrated user interface for interacting with the surrounding augmented material.

We have been experimenting with using a mobile augmented reality system (MARS) testbed to create location-aware multimedia presentations for outdoor users. Building on our earlier work on a MARS campus tour guide [7], we introduce the concept of a *situated documentary* that embeds a narrated multimedia documentary within the same physical environment as the events and sites that the documentary describes. One of the most important principles of journalism is to locate a story in a physical space. We accomplish this by situating the news consumer literally at the story's location, and layering a multimedia documentary over that space.

As depicted in Figure 1, the user wears an experimental backpack-based system, based on commercial hardware that we have chosen for programmability and power at the expense of comfort and wearability. Graphics and imagery are overlaid on the surrounding world by a see-through head-worn display. Head tracking is accomplished using a centimeter-level real-time kinematic GPS position tracker and an inertial/magnetometer orientation tracker. Audio is presented through the head-worn display's earphones, and coordinated video and other multimedia material are presented on a companion hand-held display. Interaction occurs through a set of selection mechanisms based on positional proximity and gaze orientation, a trackpad that is used with the head-worn display, and a pen-based user interface on the hand-held display.

In this paper, we first discuss how our work relates to previous research in Section 2. Next, in Section 3, we introduce our main application scenario and its user interface techniques: a multimedia documentary of highlights in the history of Columbia's campus. We then briefly describe the hardware and software used for our current testbed in Sec-

Figure 1. Situated documentaries. a) Our backpack-based testbed, with tracked see-through head-worn display and pen-based hand-held computer. b) An image photographed by a video camera that wears our testbed's see-through head-worn display. The labels and virtual flags are part of the user interface, described in Section 3. c) Related information displayed on our hand-held computer.

tion 4. Finally, Section 5 provides our conclusions and a discussion of ongoing and future work.

2. Related Work

As computers continue to shrink in size, researchers have begun to address the development of outdoor location-aware mobile and wearable systems. Some have relied on modifying the environment being explored; for example, Smailagic and Martin [19] label campus information signs with bar codes to provide location-specific information on a hand-held computer equipped with a bar code scanner. In contrast, others have combined GPS and orientation trackers to produce map-based contextual displays [11], to provide audio navigation assistance to blind users [12], and to annotate the world with overlaid textual labels [7, 22, 9].

Situated documentaries rely in part on the idea of creating hypertextual links between physical and virtual objects or locations. In earlier indoor work, using short-range, magnetic and ultrasonic tracking systems, we developed a hypermedia system that supports linking arbitrary X11 windows, displayed on a tracked see-through head-worn dis-

play, to a variety of targets, including 3D world locations and tracked objects [6]. Wearable systems by Rekimoto et al. [17] and Starner et al. [21] allow people to register digital data with visually-coded or infrared-tagged objects. Billinghurst et al. [2] use similar visual fiducials to position texture-mapped representations of participants in an augmented reality teleconference. Mann [15] and Jebara et al. [10] associate information with untagged objects using visual recognition algorithms. Pascoe [16] uses a hand-held display and GPS to allow an ecologist to link observation notes to the locations at which they are written. All these projects can be seen as leading towards the goal articulated in Spohrer's proposal for a "WorldBoard" [20]: the creation of a world-wide spatial hypertext of information anchored to physical locations and objects.

Our work is built on top of a new version of the backpack-based wearable MARS testbed that we developed for our earlier "Touring Machine" [7]. This system uses a campus database to overlay labels on buildings seen through a tracked head-worn display. Users can request additional overlaid information, such as the names of a building's departments, and can view related information, such as a de-

Figure 2. Virtual flags denoting points of interest, photographed from the top of a campus building.
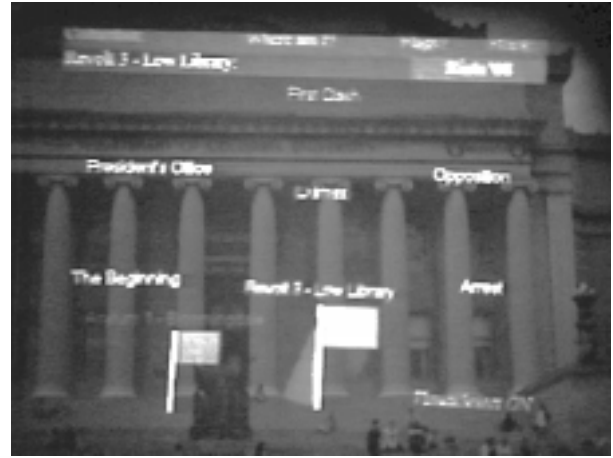
partment's web page, on a hand-held display. The situated documentaries that we describe here extend this previous work in several ways:

- Rather than linking individual labels or web pages to locations, we support context-dependent, narrated multimedia presentations that combine audio, still images, video, 3D graphics, and omnidirectional camera imagery.

- We make extensive use of overlaid 3D graphics for both the user interface (e.g., 3D widgets for user guidance) and the presentation content (e.g., *in situ* reconstructions of buildings that no longer exist and views of visually obstructed infrastructure).

- We embed the informational elements in an early version of a new *physical hypermedia* user interface that guides users through a presentation, while giving them the freedom to follow their own trails through the material.

3. User Interface

Our user stands in the middle of Columbia's campus, wearing our experimental backpack computer system and a see-through head-worn display, and holding a tablet computer (Figure 1a). As the user moves about, their position and head orientation are tracked, and through the head-worn display they see the campus environment overlaid with virtual material, such as that shown in Figures 1(b) and 2.

The user can interact with the surrounding environment in different ways. On the hand-held computer, which is net-

a)

b)

Figure 3. Two different menu designs for listing multimedia snippets about the student revolt. a) World-stabilized circular menu around Low Library (photographed through an earlier, low-resolution, see-through, head-worn display). b) Head-stabilized list with anchor to its flag (screen dump of the system running in indoor test mode, with an omnidirectional image as a backdrop).

worked to the backpack computer that drives the head-worn display, the user can view and interact with information, and input data with a stylus. All information on the hand-held display is presented using a standard web browser. Items seen on the head-worn display can be selected with an approximation to gaze-oriented selection described below. A menu on the head-worn display can be manipulated using a two-button trackpad mounted on the back of the hand-held computer for easy "reach-around" selection.

The head-worn user interface consists of a screen-

stabilized part and a world-stabilized part. The menu bars on top of the screen and the cone-shaped pointer at the bottom (shown most clearly in Figure 3b) are screen-stabilized and therefore always visible. World-stabilized material is visually registered with specific locations on campus. World-stabilized 3D elements are displayed in the correct perspective for the user's viewpoint, so the user can walk up to these elements just as they can to physical objects

3.1   Application Scenario

Our situated documentary begins with a narrated introduction, explaining that the user will be able to learn about events related to the campus, and referring the user to the hand-held display for an overview. Before turning to the hand-held computer, the user looks around and sees virtual flags with textual labels denoting points of interest, positioned around the campus (see Figures 1b and 2). The virtual flags are world-stabilized user-interface elements that are iconic representations of the topmost *group nodes* in a hierarchical presentation.

The hand-held display provides an overview of the material embedded in the surrounding environment. Three main topics are currently available: a description of the Bloomingdale Asylum for the Insane, which once occupied the current campus before Columbia's move in the late 19th century, a documentary on the Columbia student revolt of 1968, and a tour of Columbia's extensive underground tunnel system. Looking at the surrounding flags, the user can see how the different stories are distributed over the campus area. The labeled flags come in three different colors: red for the student revolt, blue for the tunnel system, and green for the Bloomingdale Asylum.

The user can select a flag in several different ways. One method, which works when the user is in the system's *VisualSelect* mode, is to look in the flag's direction, orienting one's head so the desired flag's projection is closer than any other to the center of the head-worn display and within a fixed target area. When these criteria are met, the flag's label changes color to yellow. If the criteria hold for a half second, then the flag is selected and its label changes color to green. (This approximation of gaze selection was originally developed for selection of building tags in [7].) Flags are selectable from any distance. Although the flags scale with distance, their textual labels do not, so there is always a visible anchor that is selectable.

A second selection method is based on positional proximity. A menu item allows the user to ask the system to select the flag to which they are currently closest (or to select another flag by name), and the cone-shaped pointer on the head-worn display will point towards that flag, guiding the user to it. Finally, a flag can be selected automatically by following a link in the presentation.

When a flag is selected, it starts to wave gently, and all flags of a different color are dimmed (reduced in intensity). Therefore, when a user looks around while a flag is selected, the other flags in its category stand out. The cone-shaped pointer always points toward the selected flag, so that the user can be guided back to it should they look away.

Selecting a flag causes the second menu bar (the green *context* menu below the blue top-level menu) to display that flag's label plus additional entries that are available for its group node (e.g., links to other group nodes). All these entries can be selected using the trackpad. The group nodes (and their corresponding flags) have a default numbering corresponding to an order set forth in the presentation description. A button click on the trackpad directs the user to the next node in this order; however, at all times the user can choose to select a different flag using any of the methods mentioned above.

In our case, the user selects the entry for the student revolt from the overview menu on the hand-held computer. The cone-shaped arrow on the head-worn display points to a red flag, which starts waving, in front of Low Library, which is about 150 yards away. This flag is the starting point for information on the student revolt.

Once a flag is selected, the user can display an overlaid *in-place* menu (see Figure 3), which lists the parts of the presentation associated with the flag's group node. (Section 3.3 discusses the in-place menus further.) The in-place menu for Low Library's revolt flag provides access to background information on how the student revolt started, grouped into five segments.

Selecting an entry in this menu using the trackpad starts that entry's part of the multimedia presentation, each of which ranges in length from seconds to minutes in our current material. Here, the user selects the entry labeled *First Clash*. This results in a narrated description of how the students and the police clashed for the first time on the steps of Low Library, where the user is now looking. The presentation includes coordinated still images that are overlaid on the scene (Figure 4a) and videos that are played on the hand-held computer (Figure 4b).

The head-worn display's menu bar allows the user to display an overview of the student revolt on the hand-held computer or to follow links to other places directly by selecting them with the trackpad to learn more about about the revolt and what happened at other campus buildings.

At this point, the user has found a description of how the students used Columbia's tunnel system to occupy buildings guarded aboveground by the police. The user decides to follow a link to learn more about the tunnels by exploring the blue flags. Since the real tunnels are difficult (and illegal) to enter, the user can vicariously explore portions of them through a set of $360°$ omnidirectional camera photo-
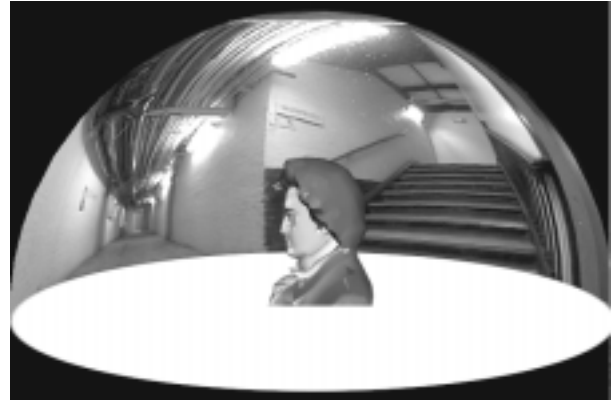
a)
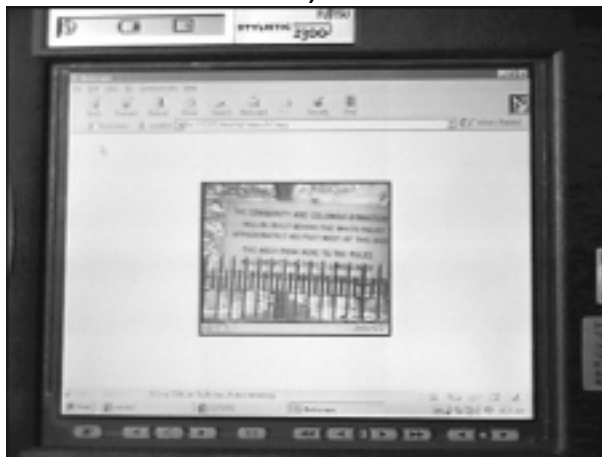
b)

Figure 4. Imagery documenting the student revolt in 1968: a) Still image, overlaid on top of Low Library, b) video material displayed on the hand-held computer

a)

b)

Figure 5. Exploring Columbia's tunnel system: a) Schematic view of how a user experiences an omnidirectional camera image. b) The omnidirectional camera image seen from a user's perspective.

graphic images (Figure 5) that temporarily teleport the user underground, supplemented by maps and blueprints.

The presentation mentions that the oldest parts of the tunnel system preceded Columbia's move to the area and were originally built for the Bloomingdale Asylum. Intrigued, our user turns to the green flags to find out where the main asylum buildings were situated, and is shown a 3D model of the buildings overlaid in place on the campus, in conjunction with historical images (see Figure 6). The documentary mentions that one building built for the asylum is still standing and is now known as Buell Hall, and points the user toward it.

3.2. Multimedia Presentations

The multimedia material in each presentation node is a coordinated media stream (see Section 4.2) that typically, but not necessarily, makes use of both the hand-held display and the head-worn display, and which includes an audio track. The different media that can be freely combined to create a multimedia presentation are:

- *Audio material on the head-worn display.* Audio is played over the head-worn display's earphones, and includes both narration and non-speech audio (e.g., recordings of the 1968 revolt).

- *Images on the head-worn display.* Images (e.g., Figure 4a) are displayed as world- or head-stabilized 3D textured polygons that can make use of simple ani-

5

a)
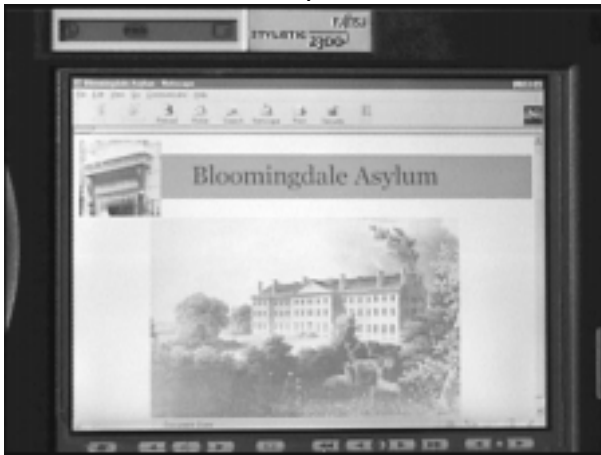
b)

Figure 6. a) A simplified 3D model of the main Bloomingdale asylum building overlaid on Columbia's campus by the see-through head-worn display. b) Documentary material displayed on the hand-held computer.

mated effects. For example, we often "flip up" head-stabilized images from a horizontal position until they fill the screen.

- *Web pages that include static images, video material, and applets on the hand-held display.* Figures 4(b) and 6(b) show examples of images and video, created by calling up related material on the hand-held browser using our communication infrastructure (see Section 4.2).

- *3D models.* Figure 6(a) shows a simple example. Models are shown full-size and world-stabilized in their actual location.

- 360° *omnidirectional camera surround views.* These allow us to immerse the user in an environment that is not physically available. We use a commercial omnidirectional camera [5]: a digital camera pointing at a parabolic mirror that captures a 360° hemispherical surround view in a single image. Each of these anamorphic images is texture-mapped onto a hemisphere displayed around the user, as depicted schematically in Figure 5(a), so that the user can look around (Figure 5b). The see-through head-worn display's opacity is controlled by a dial, allowing us to make the display opaque when viewing these images. (Unfortunately, the display's opacity cannot be set in software.)

3.3. Exploratory UI Design

We also use omnidirectional images as backdrops for indoor demonstrations of our system and for exploratory development of new user interface elements and variants. Figure 3 demonstrates this approach. Part (a) shows our original version of an in-place menu, shot outdoors through a low-resolution see-through head-worn display; part (b) shows our current version of the same menu, captured as a screen dump of the system running indoors, using an omnidirectional image of the campus as a backdrop. In the latter design, the menu is a head-stabilized element, rather than the world-stabilized circular menu of part (a). A leader line links the menu to its associated flag, allowing it to be followed back if the user turns away from the flag, an approach that we used to direct users to objects that were not within their field of view in an earlier indoor augmented reality system for maintenance and repair [8].

4. System Design

4.1. Hardware

Our current backpack is an updated version of our first outdoor MARS testbed [7], with the following changes:

Head-worn Display: We use a Sony LDI-100B color display with 800×600 triad resolution. It has a dial to adjust its opacity from nearly totally opaque to about 20% transparent. In our experience, under a bright cloudy sky the preferred setting is close to the most opaque. We have just begun to experiment with a stereo version of this display, the Sony LDI-D100B.

The images in this paper were shot directly through the LDI-100B display worn by a dummy head containing an embedded NTSC camera. Images 3a) and 4a) stem from earlier footage, shot through a Virtual I/O i-glasses display with 263×230 triad resolution.

Hand-held Computer: The hand-held computer shown in Figures 1, 4(b), and 6(b) is a Fujitsu Stylistic 2300 with

a 233 MHz Pentium MMX CPU and a transflective 800 × 600 color display, designed to be readable in bright sunlight. The Fujitsu's performance is adequate for playing MPEG movies of up to VGA resolution at reasonable frame rates, but it is heavier than we would like (3.9 pounds). We have just switched to a 2.2 pound Mitsubishi AmITY CP pen-based computer with a 166 MHz Pentium MMX CPU and 640 × 480 color display.

Orientation Tracker: We use an Intersense IS-300Pro inertial/magnetometer orientation tracker with a single sensor mounted rigidly on a head band that we attached to the head-worn display's temple pieces, as shown in Figure 1(a).

Position Tracker: Position tracking is done with an Ashtech GG24 Surveyor real-time kinematic differential GPS system, which uses both US GPS and Russian Glonass satellite constellations to increase the number of visible satellites. We have installed a base station on campus, from which we broadcast correction signals via radio modem. This system provides centimeter-level accuracy in open areas, such as those depicted in the figures, where we have line-of-sight to more than six satellites. However, tracking degradation and loss remain a problem when we pass too close to tall buildings or beneath trees.

4.2. Software

We extended the software architecture of our previous prototype [7], which is based on our COTERIE distributed virtual environment infrastructure [13, 14]. We run a custom-built HTTP server on the hand-held computer, allowing it to communicate with the backpack computer and accept user input from any web-based interface, including Java applets.

The multimedia information to be conveyed through the augmented reality interface has to be arranged and locally distributed over the target region. For this purpose we designed several authoring tools.

To create the multimedia presentations, we developed a simple extension to the interpreted language *Repo*, our extended variant of the lexically scoped interpreted language *Obliq* [4]. Each multimedia presentation is stored as a *Repo* script, referencing by filename the multimedia "chunks" (images, video segments, audio snippets, 3D animations, omnidirectional views) it uses. Each chunk is stored on the computer (backpack or hand-held) on which it is to be played; additional material to be presented on the hand-held computer can be obtained from the web using a wireless network interface.

Students in a graduate Journalism class taught by the third author used our multimedia prototyping environment to break the footage they had collected into chunks and wrote scripts to create our multimedia presentations. Syn-

chronization takes place purely at the level of these relatively coarse-grain media chunks by exchanging *Repo* messages between the main server on the backpack computer and the HTTP server on the hand-held computer.

All location-based information is stored in a campus database on the backpack computer. This database contains the complete structure of the situated documentaries, including the contents of all context-menus and links to the multimedia presentation scripts.

We used an early version of a map-based tool we are developing to place 3D objects at any specified latitude–longitude. For this project, we scanned in a high-resolution map of Columbia's campus that provides a placement resolution of about 6 inches in latitude or longitude.

5. Conclusions and Future Work

Although most of our user experience has been limited to the authors of this paper and to the students who helped construct the presentations, our system has been demonstrated informally in several Journalism classes, to visitors to our lab, and to attendees of a Department of Defense seminar who tried the indoor version. While feedback has been encouraging, users understandably cite the current prototype's form factor, weight (about forty pounds), and appearance as drawbacks. We are confident, however, that these issues will be addressed by the commercial development of sufficiently small wearable devices.

For the near term, we note that much of our backpack's weight is due to its computer, which together with its external battery weighs about twenty-two pounds. We selected this machine (Fieldworks 7600) for the programming comfort of the system's developers, rather than the physical comfort of its wearers. Its flexibility and extensibility (expansion ports for six PCI and ISA cards, and the ability to run a desktop operating system and programming environment) have been invaluable during development and testing. We are investigating options for replacing it with a lighter, more powerful laptop, but require high-performance support for the OpenGL 3D graphics API that we use, which is not yet offered by current laptops. To provide a lighter hand-held display, we are beginning to experiment with the Casio Cassiopeia E-100 running Windows CE, a palm-top computer with a 240 × 320 16-bit color display.

There are many directions that we are currently exploring to further develop our software. For example, our system currently provides no reasonable facilities for end-user authoring. We are especially interested in developing this kind of support, with emphasis on how such a system might be used by journalists in the field to develop stories. We are also working on an interface between our backpack system and an indoor multi-user augmented reality system [3]

to make possible collaboration among indoor and outdoor users. Using a 3D model of the environment, indoor users create virtual objects and highlight real objects for outdoor users to see, and maintain histories of outdoor users' activities. In turn, outdoor users point out interesting objects and events for indoor users to view.

6. Acknowledgements

References

[1] H. Beadle, B. Harper, G. Maguire Jr., and J. Judge. Location aware mobile computing. In *Proc. ICT '97 (IEEE/IEE Int. Conf. on Telecomm.)*, Melbourne, Australia, 1997.

[2] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. A wearable spatial conferencing space. In *Proc. ISWC '98 (Second Int. Symposium on Wearable Computers)*, pages 76–83, 1998.

[3] A. Butz, T. Höllerer, S. Feiner, B. MacIntyre, and C. Beshers. Enveloping users and computers in a collaborative 3D augmented reality. In *Proc. IWAR '99 (Int. Workshop on Augmented Reality)*, San Francisco, CA, October 20–21 1999.

[4] L. Cardelli. A language with distributed scope. *Computing Systems*, 8(1):27–59, Jan 1995.

[5] Cyclovision Technologies, Inc. ParaShot—One Shot 360 Degree Surround View Images. http://www.cyclovision.com, 1998.

[6] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon. Windows on the world: 2D windows for 3D augmented reality. In *Proc. UIST '93*, pages 145–155, 1993.

[7] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Proc. ISWC '97 (First Int. Symp. on Wearable Computers)*, pages 74–81, Cambridge, MA, October 13–14 1997.

[8] S. Feiner, B. MacIntyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[9] B. Jang, J. Kim, H. Kim, and D. Kim. An outdoor augmented reality system for GIS applications. In Y. Ohta and H. Tamura, editors, *Mixed Reality, Merging Real and Virtual Worlds*, pages 391–399. Ohmsha/Springer, Tokyo/New York, 1999.

[10] T. Jebara, B. Schiele, N. Oliver, and A. Pentland. DyPERS: Dynamic personal enhanced reality system. In *Proc. 1998 Image Understanding Workshop*, Monterey, CA, November 1998.

[11] S. Long, D. Aust, G. Abowd, and C. Atkenson. Cyberguide: Prototyping context-aware mobile applications. In *CHI '96 Conference Companion*, pages 293–294, April 1996.

[12] J. Loomis, R. Golledge, and R. Klatzky. 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, June 17–18 1993.

[13] B. MacIntyre and S. Feiner. Language-level support for exploratory programming of distributed virtual environments. In *Proc. UIST '96*, pages 83–94, Seattle, WA, November 6–8 1996.

[14] B. MacIntyre and S. Feiner. A distributed 3D graphics library. In *Computer Graphics (Proc. ACM SIGGRAPH '98)*, Annual Conference Series, pages 361–370, Orlando, FL, July 19–24 1998.

[15] S. Mann. Wearable computing: A first step toward personal imaging. *IEEE Computer*, 30(2), February 1997.

[16] J. Pascoe. Adding generic contextual capabilities to wearable computers. In *Proc. ISWC '98 (Second Int. Symp. on Wearable Computers)*, pages 92–99, Cambridge, MA, October 19–20 1998.

[17] J. Rekimoto, Y. Ayatsuka, and K. Hayashi. 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, October 19–20 1998.

[18] Seiko Communications. Seiko message watch documentation. http://www.messagewatch.com, 1998.

[19] A. Smailagic and R. Martin. Metronaut: A wearable computer with sensing and global communication capabilities. In *Proc. ISWC '97 (First Int. Symp. on Wearable Computers)*, pages 116–122, Cambridge, MA, October 13–14 1997.

[20] J. Spohrer. WorldBoard—What Comes After the WWW? http://www.worldboard.org/pub/spohrer/wbconcept/default.html, 1997.

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

[22] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther. A wearable computer system with augmented reality to support terrestrial navigation. In *Proc. ISWC '98 (Second Int. Symp. on Wearable Computers)*, pages 168–171, Pittsburgh, PA, October 19–20 1998.

[23] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Trans. on Information Systems*, 10(1):91–102, January 1992.