M.I.T. Media Lab Vision and Modeling Group Technical Report No. 355, Nov. 1995
Submitted to Presence special issue on Augmented Reality, Nov. 1995

Wearable Computing and Augmented Reality

Thad Starner, Steve Mann, Bradley Rhodes, Jennifer Healey, Kenneth B. Russell, Jeffrey Levine, and Alex Pentland

The Media Laboratory
Massachusetts Institute of Technology
Room E15-394, 20 Ames St., Cambridge MA 02139
Email: thad@media.mit.edu, steve@media.mit.edu

ABSTRACT: Wearable computing will change the current paradigms of human-computer interaction. With heads-up displays, unobtrusive input devices, personal wireless local area networks, and a host of other context sensing and communication tools, wearable computing can provide the user with a portable augmented reality where many aspects of everyday life can be electronically assisted. This paper focuses on several such situations, such as an academic or business conference, classroom note-taking, office communication, maintenance, or a visit to a museum.

1 Introduction

The recent push for smaller and faster notebook computers is indicative of a major trend in computing. Users want computers that are as portable and convenient as possible to help them with daily activities. In order to accommodate this trend, keyboard-less Personal Digital Assistants (PDA's) were introduced. Current attempts at a PDA revolve around pen computing. While handwriting recognition will improve, these systems will always require shifting one's gaze and using both hands (or, at least a hand and a wrist) for input. In addition, a useable writing surface will always be larger than the dimensions of the typical pocket.

However, another PDA effort has been underway since before the much publicized introduction of the pen computers. These systems use head mounted displays (HMD's) to provide privacy and convenience. Their CPU's are designed to be small and unobtrusive [Platt, 1993, Martin & Siewiorek 1994], and alternative input devices have been developed to utilize these machines in just about any context. Gradually, a common goal is emerging among the independent inventors responsible for these devices: a personal computer should be worn, much as eyeglasses or clothing is worn, to provide access to computing power at all times. These new machines are now mature enough to provide personal, portable, augmented realities. This capability promises to deliver where the pen-based PDA's are faltering, in providing a truly ubiquitous personal assistant.

1.1 Paper Overview

While advances in hardware make it difficult to talk about a particular wearable computing platform in a timely fashion, it is usually the first question asked of our "cyborgs." In addition, there are many design issues and misconceptions about wearable computing equipment. In order to address some of these, Sections 2 and 3 discuss the current hardware and some experiential notes about daily use of these systems. Those with experience in the field may want to skip to Section 4 which discusses directions for future hardware development. Finally, those interested in new software systems and applications of this type of augmented reality will find Sections 5-7 on typical applications of the base equipment, augmented memory, and camera-based augmented and 'mediated' realities more interesting.

2 Current Hardware

Two different styles of wearable computing hardware are supported: local processing and remote processing. The local processing systems are intended for constant, everyday use while the remote systems allow more flexibility for experimentation.

2.1 Local processing system

Our current high-end local processing system is marketed by the Phoenix Group Inc. (Figure 1). The computer is approximately 8.5cm x 16cm x 12cm and contains a 66Mhz 486 CPU; 32M of RAM; 775M of hard disk; a type 2 and a type 4 PCMCIA slot; support for Private Eye (TM), LCD flat panel, and SVGA displays; and various serial, parallel, and SCSI ports. Currently, the system is used (by the first author) with a a 720x280 red monochrome Private Eye display mounted into a pair of safety glasses (Figure 2), or (by the second author) with a greyscale VGA display. Two greyscale VGA displays are used, the commercial Kopin product (Figure 3), and an earlier CRT-based system built into a pair of sunglasses. Newer 1024x768 systems should be available soon. Handykey's Twiddler (TM), a one-handed chording keyboard and mouse, is used for input (Figure 4).

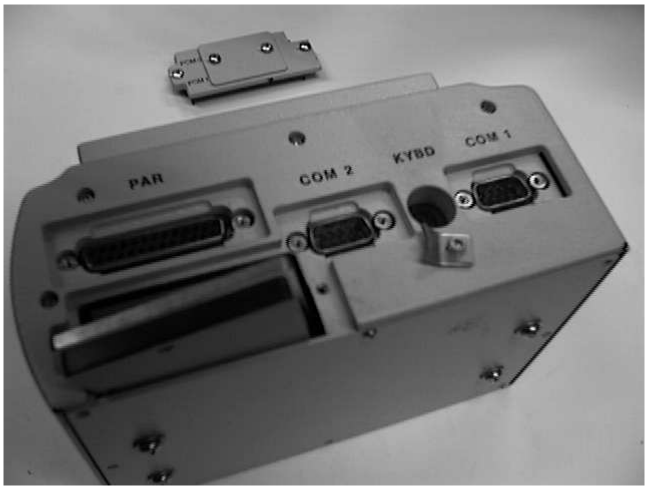

Figure 1: Phoenix 2 wearable computer base unit.

The first author has been using a similar system, constructed from PC104 stackable boards, for approximately three years (Figure 5). The PC104 standard [PC104 Corp.] has made upgrades and adaptations for different needs easy. However, this standard is size-limited to approximately 3"x3" boards due to its connector specifications. Even so, the support for this standard makes such systems ideal for fast, useable prototypes.

Linux was chosen as the operating system for the wearable computers due to its community support, source code availability, small size (can run in 2M), ease of porting, installation flexibility,

Figure 2: Private Eye (TM) display mounted on safety glasses.

Figure 3: Kopin (TM) display.

and modern features.

While "docked" the above wearable computers can be connected to other systems through serial lines, parallel ports, or PCMCIA ethernet cards. However, mobile wide area networks tend to be much slower. Currently, a cellular phone and modem are used for field data connections. Data rates can range from 1200 to 28,800 baud using bit-rate fallback if needed. However, the reliability of the connection is very poor. Off-the-shelf amateur packet (HAM) radio is more reliable than cellular but is also slow. Standard amateur packet radio operates at 1200 baud, but effective through-rates of 300 baud are more typical when considering turn-around and settling time of the audio channel. While 56kbps amateur radio links are possible, it is still not sufficient for some of the full-motion bi-directional video experiments that were planned. Thus, a different system had to be developed.

Figure 4: The Twiddler one-handed chording keyboard.

Figure 5: Wearable computer made from PC104 boards.

2.2 Remote processing system

One of the goals of the project is to experiment with computer vision algorithms in the context of wearable computing. However, the current generation of wearable computers does not have the CPU power to run many of the desired algorithms. Instead, these algorithms are developed and tested on powerful workstations, such as those made by Silicon Graphics. In order to simulate this amount of processing power on a wearable computer, a full duplex amateur television system was created [Mann, 1994]. In particular, this "reality mediator" (RM) consists of a high-quality communications link which is used to send the video from the user's cameras to the remote computer(s), while a lower quality communications link is used to carry the signal back from the computer to the HMD. This apparatus is depicted in Figure 6. Ideally both channels would be of high-quality, but the machine-vision algorithms were found to be much more susceptible to noise than the wearer's vision.

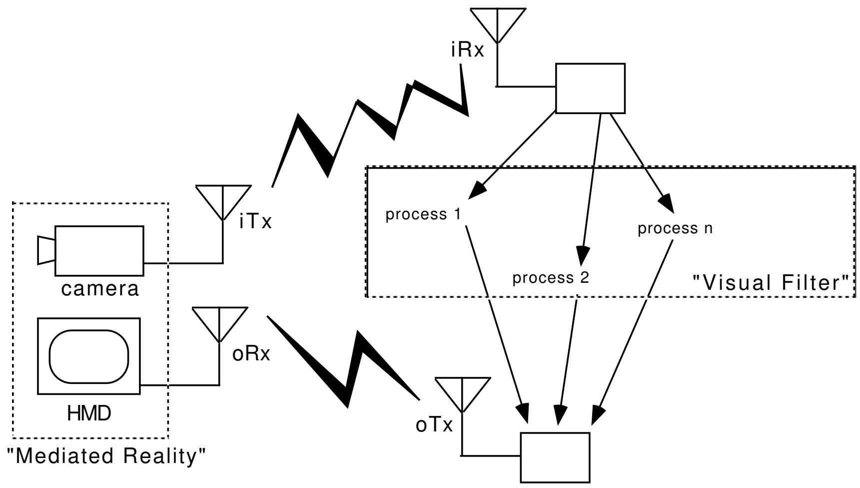

Figure 6: Remote processing system: implementation of a 'reality mediator (RM)'. The camera sends video to one or more computer systems over a high-quality microwave communications link. The computer system(s) send back the processed image over a UHF communications link. Note the designations "i" for inbound (e.g. iTx denotes inbound transmitter), and "o" for outbound.

3 Experiential Notes

Three years of wearing a computer as part of daily life resulted in surprises on many levels, both in successes and failures. In addition, questions from the curious public illuminated many misconceptions of the hardware. This section will address some of these misconceptions and give an overview of practical design concerns for wearable computing.

3.1 Monocular Displays

Many misconceptions center around monocular displays such as the Private Eye (with which we have the most experience), the Kopin display, and the monocular Virtual Image Displays unit (formerly VirtualVision). The Private Eye distinguishes itself from the others by not using the LCD technology common to head mounted displays. Instead, it consists of 280 LED's arranged in a column and a vibrating mirror that scans quickly across the eye. By switching the LED's on and off in conjunction with the position of the mirror, a fully addressable virtual image can be created. A focusing element allows this image to be moved from 10" to infinity. A valid concern is how robust such a system can be. Experience has shown that of the displays listed, the Private Eye is the most robust in harsh environments. Not only has the display been repeatedly dropped on concrete (as a failure mode of several clothing designs), but it has also been subjected to extreme temperatures and precipitation.

Eye strain is often voiced as a concern about using monocular displays. However, by adjusting the focus depth of the virtual image to match that of the real world, eye strain is avoided. In very little time a new user learns how to adjust the focus to match whatever context he is in. All three of these displays have adjustable focus. In fact, it has been our experience that head-mounted displays actually causes less eye strain than normal CRT monitors. A reason for this may be the adjustable focus. When working in long sessions, say to write a paper, the focus can be changed to relieve some of the eye muscle strain associated with holding a constant focus as with a normal computer monitor. In addition, the exceedingly crisp monochrome high contrast image provided by the Private Eye avoids the slightly-out-of-convergence effect of most color monitors, making it the preferred display for editing text.

Another misconception is that such displays act as an "eye patch" since they are not transparent. The thought is that the virtual image on one eye overrides the image of the real world on the other. In actuality, the images from the eyes are "shared" so that it looks as if both eyes see the real world and the virtual at the same time (assuming normal, healthy eyes). However, if the virtual and the real are widely disparate, say text free-floating over a hiking trail, the user must choose which image will be primarily attended. Even so, one can comfortably navigate a busy conference or city sidewalk while jotting notes on the day's events.

A final misconception, often found with vision scientists, is that there is an adaption period when putting on or taking off the display. However, if properly focused, these displays can be put on or taken off at will without any noticeable adaptation effects.

There are many design improvements that can be made to the Private Eye and its cousins. A simple improvement is to remove the "dead" zones (the area taken up by the casing of the display) associated with wearing the display in front of the eye. A beam-splitter arranged as described by [Feiner, MacIntyre, and Seligmann, 1993] creates a see-through system and thus reduces the effect of these dead zones. However, a penalty is paid in brightness and contrast. Some form of focus referent, whether it be a background pattern or text, should remain on the screen at all times to avoid the system acting as an inactive eye patch. Additional improvements would include auto-focus, auto-intensity, auto-chromatic correction, and full color. Auto-focus would change the focus of the image in the virtual field of view to match the active focus of the real world. Auto-intensity would change the average image brightness of the display to match lighting conditions. This relatively simple improvement would avoid the virtual image overwhelming the real at night and improve light adaptation of the eyes when moving between variably lit areas. Finally, auto-chromatic correction would change the color of the display slightly depending on conditions to provide contrast with the real world.

3.2 Keyboards

A common misconception with chording keyboards is that they are difficult to learn or somehow inefficient. In general, both of these preconceptions are wrong. In fact, chording keyboards are much easier to learn and can produce typing speeds much faster than traditional "QWERTY" keyboards, which were optimized for the constraints of mechanical typewriters. For example, a typical learning curve for the Twiddler is 5 minutes to learn the alphabet, 1 hour to begin to touch type, and typing rates of 10 words per minute in two days of practice. Several other chording keyboards can boast similar learning rates, and there have been reports of typists reaching speeds significantly faster than speech (around 160 wpm). Courtroom stenography is a common example of high speed typing. With the Twiddler, typing rates of 50 wpm have been attained. However, experiments are underway to create an optimized macro package based on useage to explore the upper limit of this design.

Critics of the Twiddler suggest that the positioning of the fingers on the keyboard may cause repetitive stress injury. However, these opinions are often expressed by people who have only used one for an hour. Just like when learning to play an instrument, using a new type of keyboard may feel very awkward. However, in the case of the Twiddler, the natural use position keeps the wrist straight and unstressed and involves a different range of motion for the fingers, thus providing a possible alternative for those who are experiencing difficulty with "normal" keyboards.

3.3 Clothing

The components of our early computer proved surprisingly robust. As a point in case, the original 85M Integral hard drive, still in use, has never failed to boot, even after uncountable drops and power brown outs while spinning. However, connectors and clothing have proved a continual problem. In fact, the weakest point in the system was also the simplest: the power cable. Finding a compromise between a connector system that disconnects instead of breaking during catastrophes (falling by the cord) but still remains intact during normal use (running for the subway, for example) has been difficult.

Creating suitable computing clothing is very user dependent. Some users find weight bearing belts appropriate for the CPU box and battery, while others find shoulder strap systems or vests to be much more comfortable. As a specific example, the safety glasses mount for the Private Eye was found to be the most comfortable and convenient of any system tried for long term use. The glasses can be put on or taken off quickly, are very stable even when walking, provide surprisingly good weight distribution, and can be folded and hooked on to the shirt collar, much like sunglasses, when not in use. However, such a solution may not be tenable for a user who has a nose that is sensitive to the weight (for example, due to previous breakage). Even more to the point, with this mounting technique, each new user must have the display custom mounted to account for eyeglasses, facial features, and taste. Generally, the display is mounted so that the top line of text is centered on the center of focus of the unobstructed eye at conversational distance. This enables both the reading of a few words of text while attending a conversational partner without noticeable eye saccades. Thus, time-critical messages can be delivered without interrupting the flow of conversation. Mounting the display in this way also allows both eyes to directly see the ground. This feature can be helpful in rough terrain (i.e. mountain climbing). In summary, we have discovered that wearable computers must be tailored to the individual for successful long term use.

4 Future Hardware Efforts

While hardware development was not an initial goal of this project, the expertise gained in prototyping the initial systems has suggested some important design changes which have meshed well with other projects at the laboratory. For more information, see the references provided in the bibliography.

4.1 BodyNet

There should be no connecting wires between the components of a wearable computer. Each module of the system should have one task, whether it be external communication, processing, user input, sensing, or output. These devices should be independent and interchangeable with the rest of the body system [Hawley, 1993]. To date, Thomas Zimmerman, formerly of the Physics and Media group, has developed a system that provides body-based network capability [Zimmerman, 1995]. The advantages of this system are that it is inexpensive, noise resistant (spread spectrum), and hard to attack without being in physical contact with the user. In addition, transmission in a particular band can enable inter-person networks (providing the users are in close proximity) without interrupting the person's body network. Thus, a simple mechanism exists for sharing information. A current project is to use this research to remove the wire between the Twiddler keyboard and the CPU.

4.2 Human Powered Computing

Average battery life for the PC104 system (disk spinning, no CPU slowdown) is between 5-8 hours with a Panasonic 12V 3.4Ahr rchargeable lead gel cell, depending on the configuration of the system (286 vs. 486, 5W vs. 7.5W). While this battery life was considered extremely good versus the laptops of the time, and battery technology has improved, the user is still required to carry around significantly more weight and bulk than the electronics alone. In addition, the user must switch out his battery repeatedly during the day, which ties him to some external electrical system. Instead, why not generate power directly from the excess of the user? While seeming fanciful at first, the calculations from [Starner, 1995] show that this is a viable avenue of research. Depending on the technique used, between 5 and 17 Watts might be recovered from the user's walking without significant loading. In fact, the spring system proposed may increase the efficiency of the user's walking, even after power has been tapped for the computer! Simple mock-ups of this system, worn over several days, reinforce this idea. Since the CPU and long term memory storage require the most power, they should be placed in the shoe to take advantage of the local power. Input and output devices may generate their own power. For example, a keyboard might generate enough power from keystrokes to announce them to the shoe-based CPU.

4.3 Biosensors

Emotional affect plays a large part in our lives. In fact, there is eveidence that without affect, intellect is impaired [Damasio, 1994]. To date, computer interfaces have mostly ignored affect. However, if a computer can begin to recognize human moods and stress levels, interfaces can adapt accordingly (for example, help interfaces). Since wearable computers are in contact with their users in many contexts, affect sensing becomes an important feature to help the computer adapt to those contexts. To this end we have begun to interface temperature, blood volume pressure, galvanic skin response, foot pressure, and EMG biosensors to our wearable computers.

4.4 Displays

Finally, a head mounted display that is less obtrusive than the Private Eye is desireable. While there are several well-planned projects for such a device [Tidwell *et al*, 1995; Alvelda, 1995], a scanning system in the style of the Private Eye, whose optics are mounted on the ear pieces of sunglasses, may provide a temporary solution until the new systems are commercially viable.

5 Typical Applications

While the simple text overlay mode of a wearable computer may be considered an inferior augmented reality by some, it has proven to be one of the most important ways in augmenting the everyday world. We have identified over 20 distinct application domains for wearable computing using the current local processing system. While space does not allow a complete treatment of the subject, these applications can be separated into three loose categories: data storage, real-time data access, and heads-up display clients. Examples from each category will be discussed.

5.1 Data storage

Students are a typical example of those who can use wearable computing for data storage. In fact, note-taking is the most commonly used mode of the prototype systems. Wearable computing allows this process to become much more fluid than with any other system. The students maintain electronic copies of their textbooks, problems sets, and solutions, which they can reference and annotate at anytime. The heads-mounted display here means that students no longer have to look down at a notebook computer to verify what they are typing. In addition, the chording keyboard is much quieter than normal notebook computers and can be used under the table where it is less distracting to the class. The notes are also private, in that no one else can see what is being typed. Due to these properties, wearable computers are often allowed when other computers are not (for example, in classrooms where laptops are prohibited due to keyclick noise).

Wearable computers work just as well while standing in a laboratory or walking through a conference poster session. Information that is gained by chance meetings of a colleague in a hallway or discussions over the dinner table are no longer lost but permanently recorded without interrupting the flow of conversation. Finally, such a system allows work to be performed anywhere at anytime. By taking advantage of this functionality, an amazing amount of previous "dead" time used. For example, the few minutes spent walking between classes, standing in a cafeteria line, waiting for class to start, or traveling on public transportation might be spent writing a paper.

Medical physicians are also attracted by the properties of note-taking on wearable computers. As part of the doctor-patient relationship, physicians are sometimes taught not to write in front of their patients but instead to commit the examination to short term memory so as to write the report where the patient can not see the chart or preliminary prognosis. Unfortunately, information is lost during this process. Instead, with the privacy ensured by a wearable computer, the doctor can record all his thoughts instantly and maintain all records electronically. Electronic records result in less errors in care, provide the potential for automatic interactive diagnosis tools during examination, allow safety cross checking systems for medication, and help to identify chronic health problems. In addition, diagnosis equipment can be embedded in wearable computers, much like the "trichorder" in Star Trek. In this manner, human error in recording readings can be reduced.

Note that in neither of the above applications is speech recognition appropriate. In fact, while wearable computers, like desktop computers, are used for word processing 95% of the time, speech recognition is only applicable for a small fraction of that time. In meetings, conferences, classrooms, and private conversations, only the user's voice could be recognized by today's systems (though the technology has become quite advanced). Storing both parties' speech for later review is possible with the amount of local disk storage available, but the awkwardness of reviewing speech, selecting the important parts, and translating it to text is prohibitive (depending on knowing *a priori* when an interesting utterance will occur is unreliable). However, speech does have its uses in many other application domains. See [Schmandt, 1994] for a good treatment of when speech and speech recognition are appropriate.

5.2 Real-time Access

Another major category of wearable computer users are those who need access to real-time data, at any given time of day. For example, financial investors, in order to remain competitive, have become more and more dependent on news sources from around the world. Thus, news that happens after normal trading hours or during lunch may require immediate attention and preparation. With wearable computing, when such acrisis occurs, the proper actions can be taken with the minimum amount of interruption (no running to the phone or to the office on business). Similar scenarios can be constructed for computer systems administators, lawyers, medical doctors, or news reporters. An even stronger need for such a system occurs in the military, where access to real-time data of troop and supply movement, both of friend and foe, is crucial. The U.S. military has recognized this need and has begun field testing wearable computing hardware in both front line and support positions [CPSI, 1995].

5.3 Heads-up information displays

The final major category of wearable computer users are those who desire information overlays on the real world. For example, sports binoculars could be designed that automatically overlay the name and current statistics of the current baseball player at bat. News reporters can keep notes and check references while maintaining eye contact with an interviewee. Surgeons could watch a patient's heart and breath rate while operating. Public speakers can keep virtual notes in front of their eyes while walking among their listeners. A speaker can maintain every talk he has ever given on his system, making extemporaneous speeches much more manageable. If the talk is technical, the speaker can also keep all of his supporting material in case conflicting results are reported by his audience.

In each of these cases, a see-through graphics overlay is superior to a simple laptop computer display, due to the inconvenience of managing two physically disparate visual inputs at once. The see-though overlay is also superior to video compositing in most of these cases due to the limited resolution current video compositing techniques imply.

Note that while several markets are identified in the above examples, only the most primitive type of augmented reality is used. With more sophisticated augmented reality techniques, these markets expand and create new applications never explored before. Also, none of the previous examples mentioned what happens when the wearable computer has a concept of the context of its user. The following sections will begin to explore these possibilities.

6 Augmented Memory

We do not use our computers to their full potential. Computers are very good at storing data and performing repetitious functions, like search, very quickly. Humans, on the other hand, can be very good at intuitive leaps and recognizing patterns and structure, even when passive. Thus, an interface where the wearable computer helps the user remember and access information seems profitable. As mentioned earlier, 95% of general computer time is dedicated to word processing. With such convenient access to a keyboard, this percentage may be even higher for the wearable computer. However, word processing requires about 1% of the processing power of the system. Instead of wasting the remaining 99%, an information agent can use the time to search the user's personal text database for information relevant to the current context. The names and short excerpts of the closest matching files could then be displayed. If the search engine is fast enough, a continuously changing list of matches could be maintained, which will increase the probability that a useful piece of information will be recovered. Thus, the agent can act as a memory aid. Even if the user mostly ignores the agent, he will still tend to glance at it whenever there is a short break in his work. Thus, serendipity has a much better chance of happening. In order to explore such a work environment, the Remembrance Agent [Starner, 1993] was created.

6.1 The Remembrance Agent

The benefits of the Remembrance Agent (RA) are many. First, the RA provides timely information. If writing a paper, the RA might suggest other references that are relevant. If reading email and scheduling an appointment, the RA may happen to suggest relevant constraints. If holding a conversation with a colleague at a conference, the RA might bring up relevant work based on the notes taken. Since the RA "thinks" differently that its user, it often suggests combinations that the user would never put together. Thus, the RA can act as a constant "brain-storming" system.

The Remembrance Agent can help with personal organization. As new information arrives, the RA, by its nature, suggests files with similar information. Thus, the user gets suggestions on where to store the new information, avoiding the common phenomenon of multiple files with similar notes (i.e. archives-linux and linux-archives). The first trial of the prototype RA revealed many such inconsistencies on the sample "notes" database as well as suggested a new research project by its groupings.

As a user collects a large database of private knowledge, his RA becomes an expert on that knowledge base through constant re-training. A goal of the RA is to allow co-workers to conveniently access the "public" portions of this database without interrupting the user. Thus, if a colleague wants to know about augmented reality, he simply sends a message to the user's Remembrance Agent (for example, thad-ra@media.mit.edu). The RA can then return its best guess at an appropriate file. Thus, the user is never bothered by the query, never has to format his knowledge (i.e. some mark-up language), and the colleague feels free to use the resource as opposed to knocking on an office door. Knowledge transfer may occur in a similar fashion. When an engineer trains his replacement, he can also transfer his RA database of knowledge on the subject so that his replacement may continually get the benefit of his experience even after he has left. Finally, if a large collective of people use Remembrance Agents, queries can be sent to communities, not just individuals. This allows questions of the form "How do I reboot a Sun workstation?" to be sent to 1000 co-workers whose systems, in their spare cycles, may send a response. The questioner's RA, who knows how the user "thinks," can then organize the responses into a top 10 list for convenience.

6.2 Implementation

The current Remembrance Agent uses the SMART information retrieval system developed at Cornell University [Buckley, 1985; Buckley and Stalton, 1988], though different search engines can be substituted easily. Other systems under consideration include an in-house variant of [Deerwester *et al*, 1990]. The Remembrance Agent runs through emacs, a popular text editor. The user interface is programmed in elisp, and the results are presented as a three line buffer at the bottom of the window. Several considerations have gone into the design of the RA. First, the RA should not be distracting from normal work unless unusual circumstances arise. To that end, the RA does not use boldface or highlighting and is run at a low priority. Secondly, if the RA recovers something of interest to the user, the full text is accessible with a quick key combination. Most importantly, the RA searches on local, medial, and global contexts. In particular, the RA searches on the last 5 words, last 50 words, and last 1000 words and returns the results of these searches on the last, middle, and first line of its text buffer respectively. These values are configurable due to different needs with different text databases. To conserve computer power, local context searches occur when the user completes each word, while the other contexts are searched at every carriage return. A command line interface to the RA is also provided so that e-mail systems such as described above can be implemented.

Figure 7 shows the output of the Remembrance Agent (bottom buffer) while editing an earlier version of this document (in the top buffer). The reference database for the RA was the third author's e-mail archives. The first number on each line of the RA output is simply a file label for convenience. For example, to view message 2, the user would simply press "Control-2" The second number on each line refers to the relevance measure of the message.

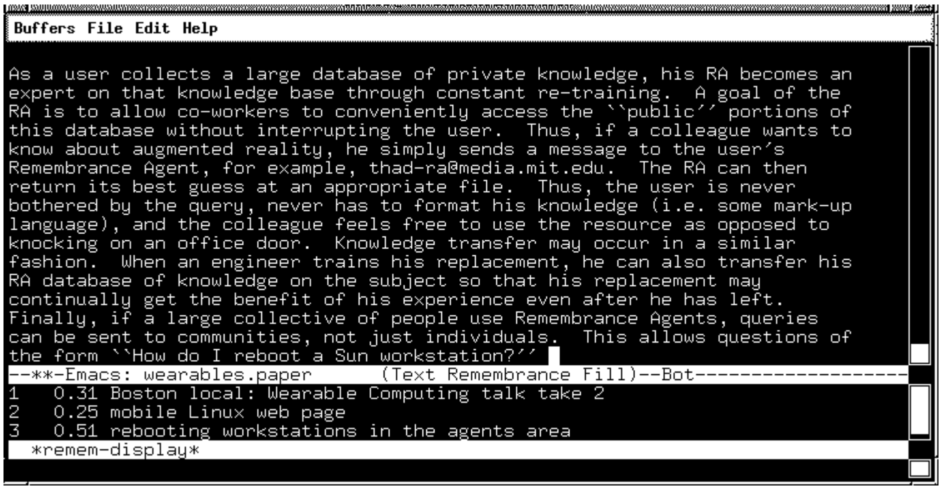

Figure 7: An example of Remembrance Agent output while editing this document.

While the Remembrance Agent can certainly be run on desktop computers, it is much more poignant to use it on wearable computing platforms due to the many more contexts in which this additional knowledge recovery can be useful. In addition, the notes that are generated on wearable computers tend to be the personal and experiential knowledge that is often hard to convey. By providing some form of access to this knowledge, a colleague may discover unexpected synergies with past experiences. Forgotten messages of "I need to tell Bob about this new research I heard about on my last trip" now have a chance for serendipity.

The Remembrance Agent currently limits itself to text. However, wearable computers have the potential to provide a wealth of contextual features [Lamming *et al*, 1994]. Additional sources of information may include time of day, location (GPS), emotional affect, face recognition, and informational tags as described in later sections. With such context, the RA may be able to uncover trends in the user's everyday life, predict the user's needs, and pre-emptively gather resources for upcoming tasks.

7 Camera-Assisted Augmented and Mediated Realities

Adding a camera to a wearable computer adds much more functionality than image capture. With a real time digitizer and the CPU power to process the images, the camera becomes an interface device. While fast, wearable digitizers are just now becoming available, the remote processing systems described earlier allow prototyping of interfaces that assume such functionality. This process, which mediates visual reality and possibly inserts "virtual" objects, is what is referred to as the "Visual Filter" in [Mann, 1994].

7.1 Finger Tracking

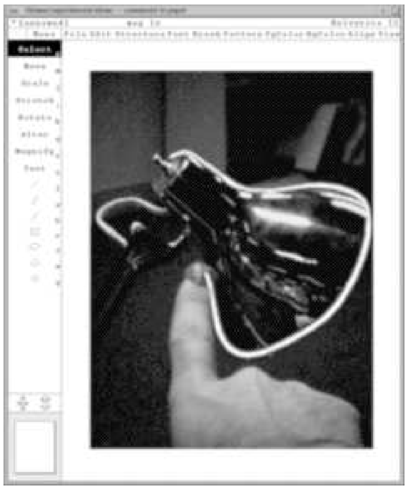

Figure 8: Using the finger as a mouse to outline an object.

Contrary to the pen computing industry's slogan, the pen is not the most intuitive pointing interface. The user's finger is. Figure 8 shows a system that tracks the user's finger while he outlines a shape. In this case, the red color of the user's fingertip is used for the tracking, though template matching methods are viable with specialized vision hardware. Thus, the finger can replace the mouse whenever a pointing device is prefered. Note that alignment is not as much an issue with a completely video based system ("mediated reality") than with a see-through system. See [Mann, 1994] for a discussion of partially-transparent versus video-based systems.

7.2 Aids for the Visually Disabled

Figure 9: Using the visual filter to enlarge individual letters while still providing a sense of context.

A video system may be preferred when creating aids for the visually disabled. Figures 9 and 10 show visual effects that run in real-time on an SGI Onyx with Reality Engine and Sirius Video boards. Through this real-time video texture mapping capability and the ability to send and receive

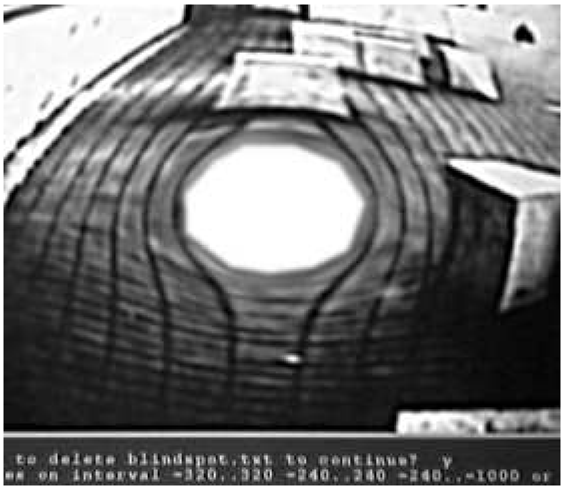

Figure 10: Mapping around a visual scotoma. Note the desired distortion in the cobblestones.

video wirelessly, those handicapped by low vision may be helped, at least while within the limited range of the communications apparatus. Figure 9 shows how text can be magnified by applying a simple 2D 'hyper-fisheye' coordinate transformation. This allows individual letters to be magnified so as to be recognizeable while still providing the context cues of the surrounding imagery. Figure 10 shows how the same technique can be used to map around scotomas ("blind spots"). Until self contained systems such as [Visionics, 1995] can include the processing power to perform this amount of computation, this relatively simple apparatus can provide a general experimental platform for testing theories of low vision aids. If large amounts of wireless bandwidth are made available to the public, as per [Nagel, 1995], then this system may become practical. Since only cameras, a HMD, and a transmitter/receiver pair are needed, the apparatus could be made lightweight from off-the-shelf components. With current technology, another benefit would be improved battery life by using just enough power to transmit the video to the nearest repeater instead of trying to locally process 28Mbytes of data each second.

7.3 Face Recognition

Recent results in face recognition allow a face to be compared against an 8000 face database in approximately one second on a 50Mhz 486 class machine [Moghaddam and Pentland, 1994]. Aligning the face in order to perform this search is still costly (on the order of a minute). However, if the search can be limited to a particular size and rotation, the alignment step is much more efficient. In the case of wearable computing, the user can assist the alignment process by limiting the search to faces that are within conversational distance. To further increase the speed of the system, the user can center the eyes of his conversant on marks provided by the system. The system can then rapidly compare the face versus images stored in the database. Given the speed of the algorithm, the system can constantly assume a face in the proper position, return the closest match, and withhold labeling until its confidence measure reaches a given threshold. Upon proper recognition, the system can overlay the returned name and useful information about the person being addressed, as simulated in Figure 11. Face recognition, combined with the Remembrance Agent as discussed earlier, should make a powerful combination for conference and business meeting scenarios. At present, the face recognition system from [Moghaddam and Pentland, 1994] has been ported to the Linux platform where proof of concept tests have been successfully performed. However, an appropriate digitization system has yet to be implemented.

Figure 11: Simulation of the face recognition system under development.

7.4 2.5D Graphics Overlay Tag System

Musuem exhibit designers often have the dilema of balancing too much text for the easily bored public with too little text for an interested visitor. With wearable computers, large variations in interests can be accomodated. Assuming that each exhibit has small bar codes, the visitor's wearable computer can upload the relevant information for a particular room from that room's network computer, possibly embedded in the wall socket or light switch. A more sophisticated system would have the visitor's computer downloading the user's stated (or learned) preferences before the room computer selects the information to send. Then, as the visitor's wearable computer camera observes the various tags in the room, the relevant information can be attached to that virtual point in space. Since the bar codes have known sizes and have a primary moment, 2D rotation and zoom can be recovered. Thus, text can be rotated and overlaid on to the real world to match the orientation of the tag as shown in Figures 12. This demonstration system can be used to give a self guided tour of a room of the laboratory. Using the visual filter hardware, one SGI is used to locate and identify the bar codes, while a second SGI composites the 3D text with the video stream. As the user passes a tagged object, the system overlays the text explaining the purpose of the object. Movies are being added to further aid the explanation. While the current apparatus bulky, it is now being reduced significantly.

Note that varying amounts of information can be attached to the tags as the user shows more interest. For example, when a visitor is distant from a tag, only the word "tag" may be overlaid on the object to alert the visitor of a hyperlink. As the visitor shows interest by getting closer, more descriptive text is overlaid on the object in question. Finally, if the user shows enough interest to stand in front of the object for a few seconds, a movie could be overlaid on the object explaining its function. In this example case, the object in question is the camera used for real-time recognition of American Sign Language (ASL). Thus, the movie could depict stored images taken from that camera and used to train the ASL recognizer. While the current demonstration system does not exchange tag data with a wall computer and currently requires the remote-processing system, it

Figure 12: A text message can associated with the tagged object. In this case, the object is a camera used to train a sign language recognizer demonstration.

shows the viability of the concept.

Previous work by [Nagao & Rekimoto, 1995] used red and blue alternating colors to identify objects using a wired, hand held system. However for this simple implementation, the remote computer looks for patterns of a particular color of red to identify a bar code. In order to avoid noise caused by other red objects in the room, simple pattern recognition algorithms are used to verify the geometry of the bar code. The tags are LED driven [PhenomenArts, Inc.] to enable the creation of a variety of codes. However, experiments have shown that red marker tape can also be used to create the codes. The advantage of passive markers are that they do not require batteries, while active tags might, in the future, listen for queries from the user's computer and flash to allow identification at greater distances or in extremely noisy environments. Note that passive tags are also limited to the spatial resolution of NTSC cameras. Variations on this system using infared reflective markers or long range bar code scanners may remove some of these constraints.

This tag system could be used to exchange object centered messages between co-workers in an office environment instead of tacky note paper. In addition, a tag architecture as described above can begin to bring the benefits of hypertext and network linking to the physical world. Privacy signatures can be attached to these links to guarantee certain messages will be seen only by a particular person or group of people. Such a system also allows the user to control how much of his personal information is used by the environment to order the relevance of the tags to him.

7.5 3D Graphics Overlay Tag System

When three or more tags are used to indicate a rigid object, and the relative positions of the tags are known, 3D information about the object can be recovered using a simplified form of [Azarbayejani and Pentland, 1995]. If the geometry of the object is known, 3D graphics can be overlaid on the real object. This concept can be very useful in maintenance. Extending a concept by [Feiner, MacIntyre, and Seligmann, 1993], Figure 13 shows 3D animated images instructing how to fix a laser printer. Note that this system involves no wires which might encumber the user and no specialized tracking hardware. Again, passive tags can be used instead of active ones. However, by using active tags, the laser printer could encode error information for the repair technician's wearable computer. The

Figure 13: A maintenance task using 3D animated instructions. The left side shows the laser printer to be repaired. The right side shows the same printer with the overlying transparent instructions showing how to reach the toner cartridge.

computer could then overlay appropriate instructions automatically. An appropriate criticism of this system is that the active tags may not be working. Instead, products could be made with single LED's embedded at advantageous locations. These would be connected to a power and communication jack which could interface directly to the repair technician's computer or through a "key" made of a small battery and wireless RF pack that could be carried by the technician. The advantage of the latter option, of course, is that the technician would not be limited in his movement by a tether. The computer and LED system could then be in interactive contact. As different areas are serviced, different LED's could be flashed, ensuring a much more robust overlay system.

8 Conclusion

Wearable computing provides an exciting way to explore augmented realities and begins to fulfill the promise of a truly personal digital assistant. While many applications and problem domains were discussed here, the potential for this field is just beginning to be realized. In the next few years, as the hardware becomes more common and accepted, the applications will challenge how the world currently thinks about computing.

9 Acknowledgements

The authors would like to thank Prof. Alex Pentland, Prof. Michael Hawley, and Prof. Rosalind Picard for their support and suggestions. In addition, a sincere thank you to Ali Azarbayejani and Baback Moghaddam for providing access to their 3D shape recovery and face recognition code. As always, thanks to the members of the Media Laboratory and the wearables mailing list for early suggestions and feedback.

References

[1] P. Alvelda (1995). VLSI Microdisplay Status Update. http://www.ai.mit.edu/people/alvelda/microdisplay.html.

[2] A. Azarbayejani and A. Pentland (1995). Recursive Estimation of Motion, Structure, and Focal Length. *IEEE Trans. Pattern Analysis and Machine Intelligence.* 17(6):562–575, June, 1995.

[3] C. Buckley. Implementation of the SMART information retrieval system. TR 85-686, CS Dept., Cornell University, Ithaca, NY, 1985

[4] C. Buckley and G. Stalton (1988). Improving Retrieval Performance by Relevance Feedback. TR 88-898. Dept. of CS, Cornell University, Ithica, NY 14853-7501. Feb 1988

[5] Apache Helicopter Maintenance Application Information Sheet (1995). Computer Products and Services, Inc. Fairfax, Virginia.

[6] A. Damasio (1994). *Descartes' Error.* G. P. Putnam's Sons, New York.

[7] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, R. Harshman (1990). Indexing by Latent Semantic Analysis. *J. American Society for Information Science.* 41(6):391-407, 1990.

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

[9] M. Hawley (1993) BodyTalk and the BodyNet, Executive Summary. http://clark.lcs.mit.edu/bodynet.html.

[10] M. Lamming, P. Brown, K. Carter, M. Eldridge, M. Flynn, G. Louie, P. Robinson, A. Sellen (1994). The Design of a Human Memory Prosthesis. To appear, *Computer Journal.*

[11] Litesign by PhenomenArts, Inc. Lexington, MA USA.

[12] S. Mann Mediated Reality (1994). *Vision & Modeling Technical Report #260.* MIT Media Laboratory.

[13] T. Martin and D. P Siewiorek (1994). Wearable computers. *IEEE Potentials.*, August, 36–38.

[14] B. Moghaddam and A. Pentland (1994) Face recognition using view-based and modular eigenspaces. *SPIE Conf. on Automatic Systems for Identification & Inspection of Humans.* San Diego, July 1994.

[15] K. Nagao and J. Rekimoto (1995). Ubiquitous Talker: Spoken Language Interaction with Real World Objects. In *Proceedings IJCAI '95.*

[16] D. Nagel (1995). NII Band: FCC Petition for Rulemaking. Apple Computer, Inc. Cupertino, CA

[17] What is PC/104? As reprinted on the AIM16-1/104 specification sheet by Analogic, Wakefield, MA. Originally printed by the PC/104 Corporation.

[18] D. Platt (1993). Presentation of the HiP PC at the MIT Media Laboratory.

[19] Christopher Schmandt (1994). *Voice Communication with Computers.* Van Nostrand Reinhold, New York.

[20] T. Starner (1993). The Remembrance Agent. Intelligent Agents Class Project. Taught by P. Maes and H. Lieberman, Fall 1993.

[21] T. Starner (1995). Human Powered Wearable Computing. To appear IBM Systems Journal. Vision and Modeling Technical Report #328 MIT Media Laboratory

[22] M. Tidwell, R.S. Johnston, D. Melville, T. A. Furness III (1995). The Virtual Retinal Display - A Retinal Scanning Imaging System. In *Proceedings of Virtual Reality World '95.* (pp. 325-334). Munich, Germany

[23] LVES Information Sheet (1995). Visionics Corp. Golden Valley, MN. http://www.wilmer.jhu.edu/low_vis/low_vis.htm

[24] T. Zimmerman (1995). Personal Area Networks (PAN): Near-Field Intra-Body Communication. Master's Thesis, MIT Media Laboratory, September 1995.