M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 397
To appear: Presence, Special Issue on Augmented Reality, 1997

Augmented Reality Through Wearable Computing

Thad Starner, Steve Mann, Bradley Rhodes, Jeffrey Levine,
Jennifer Healey, Dana Kirsch, Rosalind W. Picard, and Alex Pentland
Room E15-383, The Media Laboratory
Massachusetts Institute of Technology
20 Ames St., Cambridge, MA 02139

Abstract

Wearable computing moves computation from the desktop to the user. We are forming a community of networked wearable computer users to explore, over a long period, the augmented realities that these systems can provide. By adapting its behavior to the user's changing environment, a body-worn computer can assist the user more intelligently, consistently, and continuously than a desktop system. A text-based augmented reality, the Remembrance Agent, is presented to illustrate this approach. Video cameras are used both to warp the visual input (mediated reality) and to sense the user's world for graphical overlay. With a camera, the computer tracks the user's finger, which acts as the system's mouse; performs face recognition; and detects passive objects to overlay 2.5D and 3D graphics onto the real world. Additional apparatus such as audio systems, infrared beacons for sensing location, and biosensors for learning about the wearer's affect are described. Using the input from these interface devices and sensors, a long term goal of this project is to model the user's actions, anticipate his or her needs, and perform a seamless interaction between the virtual and physical environments.

1 Introduction

With the growing acceptance of multimedia and the Internet, desktop computers are becoming all-purpose information appliances, incorporating everyday devices such the telephone, fax, answering machine, and television. However, these computers are still confined to the desk and are not available to the user during most of the day. By designing a networked, multimedia computer that can be worn as clothing or is built into the user's clothes, the power of computing can assist everyday tasks. Following the trend of desktop computing, the networked wearable computer can replace most portable consumer electronics including the wristwatch, cellular phone, fax machine, palmtop, compact disk player, camera, camcorder, and health monitor. Using "heads-up" displays and earphones, wearable computers can provide an improved interface for these features by overlaying graphics, text, and sound on the physical world (see Figure 1). In addition, such augmented realities enable a new range of applications that are location and context specific. For example, hypertext interfaces can be extended to the physical world to aid in on-site repair, construction, museum tours, and interactive games.

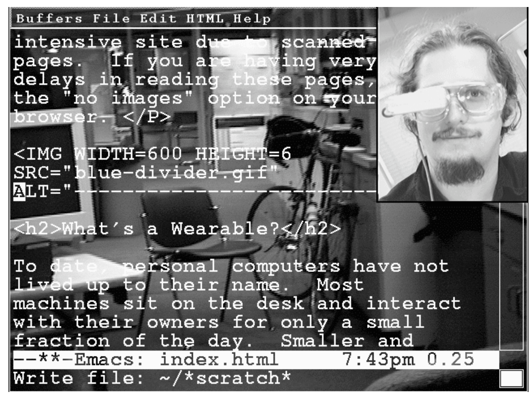

Figure 1: Due to the "sharing" of images between the user's two eyes, monocular displays such as the Private Eye seem to overlay the display on the world as shown. The insert shows the device being worn. For a binocular effect, beam splitters can be used to combine both the real and the virtual for both eyes.

Wearable computers allow a much closer association with the user. In replacing the consumer electronics listed above the sensors added allow the wearable to see what the user sees, hear what the user hears, sense the user's physical state, and analyze what the user is typing. If this information is combined into a user model, an intelligent agent may be able to analyze what the user is doing and try to predict the resources he will need next or in the near future. Using this information, the agent may download files, reserve communications bandwidth, post reminders, or automatically send updates to colleagues to help "smooth" the user's daily interactions. To use a metaphor, such an agent would act as a butler/confidant, who always stands behind the user's shoulder, knows his employer's personal preferences and tastes, and trys to streamline interactions with the rest of the world.

By maintaining a community of wearable computing users (Figure 2) we hope to encourage a diversity of applications and styles. Often the physical interfaces are customized to address the variability in physiology between users. Software interfaces are also adapted to work habits and primary applications. In addition, the networked community of users itself offers research opportunities in computer supported cooperative work (CSCW).

Figure 2: Different styles of wearable computing hardware.

2 Extending the Traditional Computer Interface with Wearable Augmented Reality

Augmented and virtual reality often focus on extending the human-computer interface through sophisticated 3D graphics. However, while graphics have had a significant impact on the computing world, the written word still dominates computer use. In general, 95% of human-computer time is spent word processing. Through the augmented environment made available by wearable computers, word processing can occur almost anywhere. With the heads-up display, the computer has the ability to display messages unobtrusively or urgently grab the user's attention. In addition, the wearable computer can display text or graphics in physically meaningful locations in the visual field, provided that it is aware of where the user is looking. Using these concepts, word processing can be significantly enhanced with a sort of augmented memory. In the system described below, a one-handed chording keyboard is used for entering text (see Section 5), though the system can be adapted easily for those instances in which speech recognition is appropriate.

2.1 Augmented Memory

Computers perform well at storing data and executing repetitive functions quickly. Humans, on the other hand, make intuitive leaps and recognize patterns and structure, even when passive. Thus, an interface in which the wearable computer helps the user remember and access information seems profitable. While word processing may comprise the majority of computer use, it requires only a small fraction of the computer's processing power. Instead of wasting the remaining power, an information agent can use the time to search the user's personal text database for information relevant to the current task. The names and short excerpts of the closest matching files can then be displayed. If the search engine is fast enough, a continuously changing list of matches can be maintained, increasing 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. In order to explore such a work environment, the Remembrance Agent (Rhodes and Starner, 1996) was created.

2.1.1 The Remembrance Agent

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

The Remembrance Agent can also 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 content (e.g. archives-linux and linux-archives). The first trial of the prototype RA revealed many such inconsistencies within the sample database and 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 through constant re-training. A goal of the RA is to allow co-workers to access the "public" portions of this database conveniently 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 (e.g. thadra@media.mit.edu). The RA can then return its best guess at an appropriate file. Thus, the user is not interrupted by the query, and he never has to format his knowledge explicitly, as with HTML. Knowledge transfer may occur in a similar fashion. When an engineer trains his successor, he can also transfer his RA's database of knowledge on the subject so that his replacement may continually receive 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.

2.1.2 Implementation

The Remembrance Agent (RA) is comprised of two parts, the user interface and the search engine. The user interface continuously watches what the user types and reads, and it sends this information to the search engine. The search engine finds old email, notes files, and on-line documents that are relevant to the user's context. This information is then unobtrusively displayed in the user's field of view.

Currently, the user interface runs in elisp under Emacs-19, a UNIX based text editor that can also be used for applications such as email, netnews, and web access. The Remembrance Agent displays one-line suggestions at the bottom of the emacs display buffer, along with a numeric rating indicating the relevancy of the document. These items contain just enough information to represent the contents of the full document being suggested. For example, the suggestion line for a piece of email includes the subject line, the sender, and the date received. The suggestion line for a notes file contains the file name, owner, date last modified, and the first few words of the file. With a simple key combination, the user can display the full text of a suggested document. The RA may suggest files based on the most recent 500 words of the user's current document, the most recent 50 words, the most recent 10 words, or it may display suggestions in a hierarchy of all 3 contexts. In this way, the RA can be relevant both to the user's overarching context and to any immediate associations. Each user may customize the number of suggestions displayed, the type of documents referenced, and the update frequency of the display.

The current implementation uses Savant, an in-house information retrieval system, which is similar to the SMART information retrieval program (Salton, 1971). This search engine determines document similarity based on the frequency of words common to both the query and reference documents. In order to improve its speed, Savant creates an index of all words in each document source nightly. While Savant is not as sophisticated as many more modern information-retrieval systems, it does not require human pre-processing of the documents to be indexed.

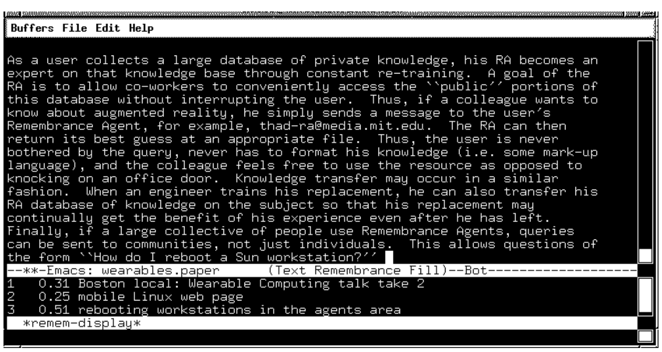

Figure 3: An example of Remembrance Agent output while editing an early version of this document. The first number on each line provides a file label for convenience, while the second number is the "relevance measure" of the message.

Figure 3 shows the output of the Remembrance Agent (in the bottom buffer) while editing an earlier version of this document (top buffer). The reference database for the RA was the third author's e-mail archives. Personal email and notes files are good sources for an RA to reference, because these files are already personalized and automatically change with the user's interests. This personal reference database is much richer in potentially relevant information than, for example, the Web. Such personalization also helps solve many "homonym" problems. For example, when an AI researcher mentions "agents" in a memo or paper, that person's RA will suggest work on autonomous agents. A chemist's RA, on the other hand, will bring up lists of various solvents and other chemical agents. Additionally, because the reference database is the user's own work, the one-line summary may be all that is necessary to remind the user of a document's relevancy. Finally, given the greater opportunity to record personal and experiential knowledge using a wearable computer, the user generates a fertile reference database for the RA.

While the Remembrance Agent gains most of its contextual information from typed text, wearable computers have the potential to provide a wealth of contextual features similar to those discussed in (Lamming and Flynn, 1994). With the systems described in later sections, additional sources of user context information may include time of day, location, biosignals, face recognition, and visual object tags. Thus, the RA begins to show the advantages of wearable, context-driven augmented reality. However, with a more comprehensive agent, the wearable computer 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.

2.2 Finger-tracking as a Pointing Device

Many different user interfaces become possible when a video camera is added to the wearable computer. While appropriate wearable video digitizers are just now becoming available, a wireless remote processing system is currently being used to experiment with this modality (see Section 5). A camera system allows the user to have an intuitive and convenient replacement for the mouse. Figure 4 shows tracking of the user's finger while he selects an item from a pull-down menu. Here, the computer tracks the color of the user's hand, though template matching techniques are also possible at higher computational expense. Thus, the finger can replace the mouse whenever a pointing device is desired. Such a pointing device itself takes no extra room, can not be lost, and can be used almost anywhere there is light.

Figure 4: Using the finger as a mouse to control the system's pointer.

The systems described above augment the traditional computer interface. However, much more can be done when the context-sensing nature of the Remembrance Agent is extended to the physical world. Much of a user's life is spent away from a desk. Thus, the computer must be adapted to take advantage of user mobility.

3 Camera-based Realities

Wearable camera systems give remote users a first-hand view of a local problem (Mann, 1994; Kraut et al., 1996; Baum et al., 1996). Such "over the shoulder" telepresence has applications in repair and maintenance, medicine, courtroom proceedings, security, conferencing, and many other fields. Camera systems can also enable users with visual disabilities. When a wearable system has access to a digitizer and the CPU power to process the images, the camera becomes a sensor which can be integrated into the user interface itself. In addition, wearable computers enable a unique, first-person viewpoint for computer vision researchers concentrating on understanding human gesture and context.

3.1 Physically-based Hypertext

Museum exhibit designers often have the dilemma 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 accommodated. Each room could have an inexpensive computer embedded in its walls, say in a light switch or power outlet. When a visitor enters the room, the wall computer can wirelessly download museum information to the visitor's computer. Then, as the visitor explores the room, graphics and text overlay the exhibits according to his interests. Taking this example farther, such a system can be use to create a physically-based extension of the "Web." With augmented reality, hypertext links can be associated with physical objects detailing instructions on use, repair information, history, or information left by a previous user. Such an interface can make more efficient use of workplace resources, guide tourists through historical landmarks, or overlay a role-playing game environment on the physical world.

Figure 5: Multiple graphical overlays aligned through visual tag tracking. Such techniques as shown in the following 3 figures can provide a dynamic, physically-realized extension to the World Wide Web.

In order to experiment with such an interface, the head-mounted camera and display system as shown in Figure 5 is used. Visual "tags" uniquely identify each active object. These tags consist of two red squares bounding a pattern of green squares representing a binary number unique to that room. A similar identification system has been demonstrated by (Nagao and Rekimoto, 1995) for a tethered, hand-held system. These visual patterns are robust in the presence of similar background colors and can be distinguished from each other in the same visual field. Once an object is identified, text, graphics, or a texture mapped movie can be rendered on top of the user's visual field as shown in Figure 5. Since the visual tags have a known height and width, the visual tracking code can recover orientation and distance, providing 2.5D information to the graphics process. Thus, graphics objects can be rotated and zoomed to match their counterparts in the physical world. This system is used to give mini-tours of the laboratory space as shown in Figures 6 - 8. Active LED tags are shown in this sequence, though the passive tags work as well. Whenever the camera detects a tag, it renders a small red arrow on top of that object indicating a hyperlink (Figure 6). If the user is interested in that link and turns to see it, the object is labeled with text (Figure 7). Finally, if the user approaches the object, 3D graphics or a texture mapped movie are rendered on the object to demonstrate its function (Figure 8). Using this strategy, the user is not overwhelmed upon walking into a room but can explore interesting objects at leisure.

Figure 6: When a tag is first located, a red arrow is used to indicate a hyperlink.

By recognizing and tracking physical objects, the wearable computer can assign computation to passive objects. The virtual version of the object maintained in the wearable computer (or on a wireless network) can then perform tasks on behalf of the user, communicate with other objects or users, or keep track of its own position and status. For example, the plant in Figure 5 may "ask" passers-by for water based on a time schedule maintained by its virtual representation. This method is an effective way to gain the benefits of ubiquitous computing (Weiser, 1991) with

Figure 7: If the user shows interest, the appropriate text labels are displayed.

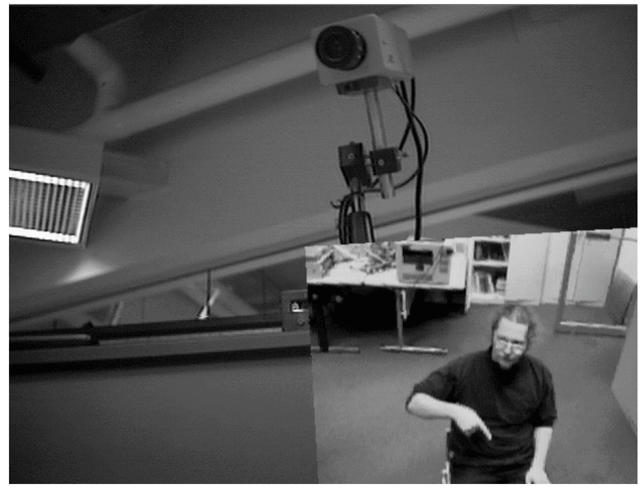

Figure 8: If the user approaches the object, 3D graphics or movie sequences are displayed.

a sparse infrastructure.

Unfortunately, the visual tag system described above has a limited number of unique codes. To avoid running out of identifiers for objects, an additional sense of location is needed. Outdoors, the Global Positioning System can be used to subdivide the space. However, for indoor use, a system of low-cost, infrared, light-powered beacons was developed to serve this purpose (Figure 9) (Poor, 1996). Each of these beacons consists of a low-power microprocessor, an infrared transmitter, and an infrared receiver. A solar cell is used to avoid constant battery replacement. These systems are typically mounted under fluorescent light fixtures where they can draw power and effectively cover a region.

By listening to these IR transmitters, the user's wearable computer can determine its location and load the appropriate set of tag identifiers for the region. In addition, the IR receiver and microprocessor enable location based information uploading. Users can leave location-based, encrypted "Post-it" notes or graphics for each other, thus extending the physical hypertext system.

Such a beacon architecture protects user privacy. While the user's computer listens to the beacons to determine its position, it does not reveal its position without explicit instruction by the user. If the user desires to reveal his position, a location daemon can be run over the wearable's wireless data network to process inquiries. Significantly, since the beacons themselves are not networked even in upload mode, there is no remotely monitorable network traffic to reveal the presence of a user at a particular node.

3.2 3D Graphical Overlays and Active Tags

When three or more tags are used on a rigid object, and the relative positions of the tags are known, 3D information about the object can be recovered using techniques developed in (Azarbayejani and Pentland, 1995). Registered 3D graphics can be then be overlaid on the real object. Such registered graphics can be very useful in the maintenance of machinery. Extending a concept by (Feiner et al., 1993), Figure 10 shows 3D animated images demonstrating repair instructions for a laser printer. The registration method becomes increasingly stable with additional known feature points. Since the tags have known dimensions, two feature points can be recovered: the right and left-hand sides of the tag. More recently, 2D visual tags have been developed which encode 32 to 128 bits of information in a 7 by 5 grid of color squares. Due to the inherent planar structure of these tags, only one tag is needed to a align a 3D graphics overlay.

Note that the visual tags shown in Figure 10 consist of small LED alphanumeric displays. For expensive machinery such as an aircraft, a manufacturer may want to embed such tags to aid in repair diagnostics. Such displays may indicate error codes (similar to some of today's printers and copiers) that the technician's wearable computer can sense. Thus, appropriate graphical instructions can automatically overlay the user's visual field. In addition, active tags may blink in spatial-temporal patterns to communicate with the wearable computer or to aid tracking in visually complex environments. Adding infrared or radio communications between the repair object and the wearable computer may allow more complicated cooperative diagnostics or repair instructions tailored to the user's level of expertise. Of course, the features of these more advanced systems must be weighed against the low cost of the passive tags discussed earlier.

3.3 Augmenting Reality Using Inherent Visual Features

One of the traditional goals of computer vision is the identification and tracking of objects by their video images. While this is a hard problem in general, constrained situations may be addressed with sufficient processing power. For example, (Uenohara and Kanade, 1994; Baum et al., 1996) have shown tethered systems which can overlay graphics on computer internals for repair or marked human legs for surgery. This section describes attempts to bring such vision sensing to wearable computing-based

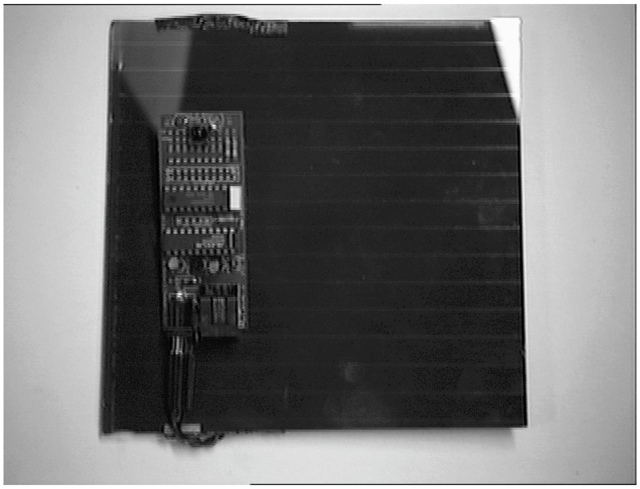

Figure 9: Environmentally-powered, microcontroller-based IR transponders.

Figure 10: A maintenance task using 3D animated graphics. 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.

augmented realities.

Using face recognition methods developed by (Turk and Pentland, 1991; Pentland et al., 1994), a face can be compared against an 8000 face database in approximately one second on a 50Mhz 486 class wearable computer. Aligning the face in order to perform this search is still costly (on the order of a minute on R4400-based machines). 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 search can be limited to faces that are within conversational distance. In the current implementation, the user further assists the system by centering the head of his conversant on a mark 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 as in Figure 11.

Without visual tags, other methods of aligning the real and virtual are necessary. Using the "pencigraphic" imaging approach (Mann, 1995), the virtual image of a rigid planar patch (Huang and Netravali, 1984) may be superimposed on to the wearer's real world visual field, creating the illusion of the virtual image floating in 3D space. Figure 12 shows six frames of video from a processed image sequence. The computer recognizes the cashier and superimposes a previously entered shopping list on her. When the wearer turns his head to the right, not only does the shopping list move to the left on the wearer's screen, but its "chirp-rate" and keystoning are manipulated automatically to follow the flowfield of the video imagery coming from the camera. Note that the tracking (initially triggered by automatic face-recognition) continues even when the cashier's face is completely outside the camera's visual field, because the tracking is sustained by other objects in the room, such as the counters, walls, row of flourescent lights, and the video surveillance cameras installed on the ceiling. In this way, the shopping list appears attached to a plane on the cashier. Such techniques demonstrate how computer vision can directly aid the registration problem in augmented reality.

Figure 11: Upon correct identification by face recognition, information about the conversant can be overlaid on the visual field.

Figure 12: The "pencigraphic" imaging approach is used to properly register, keystone, and chirp a shopping list in space using the video flowfield.

3.4 Aids for the Visually Disabled

Wearable computers and augmented reality techniques can assist users by adapting their physical senses as well as augmenting them. Approximately 2 million Americans are affected by low vision, a set of conditions which can not be corrected with normal eyeglasses and severely affects the individual's sight. However, some of their needs may be addressed by remapping the visual input. To do this remapping, a camera is mounted on an opaque head-mounted display. The image is wirelessly transmitted to a remote computer, processed, and sent back to the head-mounted display. The process of completely controlling the user's visual field has been termed "mediated reality" (Mann, 1994) to distinguish it from the "see-through" effect generally associated with augmented reality. With off-the-shelf SGI hardware, the incoming video may be remapped arbitrarily in real time. Figure 13 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 recognizable while still providing the context cues of the surrounding imagery. Figure 14 shows how the same technique can be used to map around scotomas ("blind spots"). Until self-contained systems such as (Baker, 1994) can include the processing power necessary to perform this amount of computation, such systems can provide a general experimental platform for testing theories of low vision aids. If considerable wireless bandwidth is made available to the public, as per (Nagel and Lovette, 1995), then this system may become practical. Since only cameras, a HMD, and a transmitter/receiver pair are needed, the apparatus can be made lightweight from off-the-shelf components. Furthermore, this approach may improve battery life by using just enough power to transmit the video to the nearest repeater instead of trying to process the video locally.

Figure 13: In order to help those with low vision, the visual field is remapped through a "hyper-fisheye." Note that while the letters in the center of the screen are enlarged, enough of the rest of the field is included to provide context.

Note that the techniques from previous sections can also be used in such contexts. The wearable computer acts as a guide to physical spaces and also adapts for its user's abilities. For example, for a user with low vision, the location and visual tracking systems of the wearable computer may actively direct the user to objects that would be difficult to find or manipulate even with the remapped visual field.

4 Current Efforts: User Sensing and Modeling

Augmented reality attempts to provide informative or entertaining overlays on the physical world. However, it is easy to cross the boundary between useful information and overwhelming clutter. A copier repairman does not need diagrams to replace the most commonly broken belt. In fact, such interference would be considered annoying. In order to assist the user unobtrusively, the wearable computer must model its user's knowledge, actions, goals, and

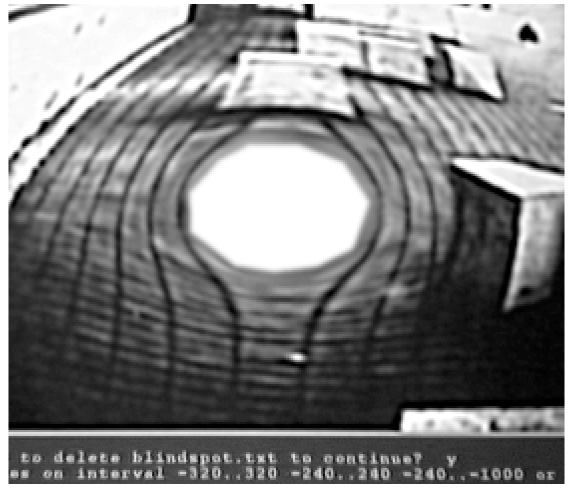

Figure 14: Visual remapping around a scotoma or "blind-spot." Note the desired distortion in the cobblestones as the image in the blind spot is pushed to either side.

even emotions. This paper has presented systems to track the user's position, visual field, and current interests as revealed by what is being typed. However, a more personal and striking interface may be possible if the user's emotional affect can be sensed as well.

Emotional affect plays a large part in everyday life. In fact, there is evidence that without affect, even rational intellect is impaired (Damasio, 1994). To date, computer interfaces have mostly ignored human affect. However, wearable computers, which are in contact with their users in many different contexts, allow an unprecedented opportunity for affect sensing. Picard discusses ways in which computers might recognize affect (Picard, 1995), as well as a number of potential applications of affective wearables. To this end, we have begun to interface temperature, blood volume pressure, galvanic skin response, foot pressure, and electromyogram biosensors with our wearable computers. While simply providing a body "status line" overlay for the user's benefit is an interesting application, we hope to create a sophisticated model of the user by combining affect and environment sensing as well as pattern recognition techniques similar to (Orwant, 1993).

Through sensor data and the user model, the wearable computer can track the state of its user and adjust its behavior accordingly. For example, suppose the user is attending an important business lunch. The user's computer, realizing that the user does not want to be disturbed, should take messages if phone or electronic messages arrive. However, in the case of an emergency message, the computer should understand enough of the context to grab the user's attention immediately. The computer should also be able to identify urgent or time-critical messages (Schmandt, 1994) and wait for a break in the conversation to post a summary message discreetly onto the user's heads-up display.

A user model should also predict the user's next action or state. Such information can be used to allocate resources pre-emptively. For example, suppose the user enjoys music while he's working. When working on a late night project, the user likes hard rock to keep him alert. However, during the day, the user prefers classical music to lower his stress. Through having learned these preferences, knowing the time of day, and sensing the user's stress level, the wearable computer can predict what the user may want to listen to next and can download potential selections over the wireless network.

This predictive ability becomes vital when network-stored video clips are to be used for overlays during repair tasks. For example, the wearable computer may begin downloading diagnostic tools for near-future use while displaying an informative overlay for a current repair task.

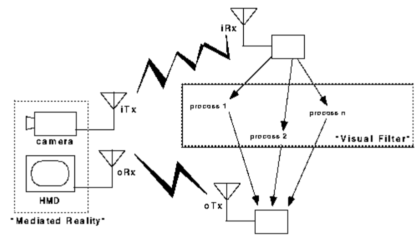

Figure 15: Remote video processing system.

5 Apparatus

While a standard hardware set allows apparatus to be shared amoung the members of the community, each user is encouraged to customize his wearable computer. The result is a constantly changing and improving environment. This section gives a brief overview of the systems used for the work in this paper, but for detailed or recent information, the reader is encouraged to visit the MIT wearable computing web site (Starner et al., 1995).

A standard system consists of a Private Eye display, Twiddler one-handed keyboard, PC-104 based 50MHz 486 computer, 16M of RAM, and approximately 1G of hard disk. The PC-104 3.6" by 3.8" board standard allows each user to add off-the-shelf 16-bit sound boards, video digitizers with on-board DSP's, PCMCIA adaptors, higher end 586 processors, or extra communications ports as desired. Cameras, biosensors, CRT or LCD displays, extra disk capacity, and custom clothing are also incorporated. Network connectivity in the region is achieved either through commercial digital cellular data service or a custom wireless system supported by the second author. When on-board processing power is not sufficient for analyzing video for the above techniques, a self contained, full duplex amateur television transmitter/receiver (Figure 15) is used in conjunction with HP or SGI workstations. Video is transmitted from the head mounted camera, analyzed remotely, and returned to the user's display with appropriate graphics overlaid on the image (Mann, 1994). This method emulates the experience of having significant processing power locally and allows rapid prototyping. As techniques are proven, the code is optimized and moved to the processor(s) of the local wearable computer, if possible. Thus, concept demonstrations are gradually integrated into everyday use. As of this writing, the face recognition code has been ported, and ports of the finger tracking and visual tag code are underway.

6 Discussion

Our purpose in creating a wearable computer community is to explore the various, and often unanticipated, uses of the augmented reality interface. The diversity of the community is reflected by the range of projects, but what are the common threads that make wearable computing-based augmented reality advantageous over virtual reality or other forms of portable multimedia?

Wearable computing allows graphics and text to be associated with physical objects and locations. Through overlay displays, the user can concentrate on the task instead of constantly looking down as is the case with notebook computers or pen-based systems. Using the techniques shown above to sense at what object the user is looking, the computer may deduce the user's intention and update the display automatically.

Wearable computers are private and are more discreet in many situations than virtual reality gear or even such consumer devices as cellular phones, pagers, or hand-held PDA's. The user can concentrate on physical reality but have immediate access to the virtual if needed. Conversely, the user can engage the virtual world but still be aware of his physical surroundings. Unlike virtual realities which use opaque displays, the augmented reality of wearable computers degrades gracefully when sensors or displays fail, defaulting to the user's unaided senses. Such a feature is particularly important in time critical applications such as aircraft control or navigation aids.

Finally, wearable computing augmented realities are truly personal. The user grows to expect his interface to be accessible continually and unchanging, unless specified otherwise. Such a persistent and consistent interface encourages trust. With experience, the user personalizes his system to ensure fluid use in everyday contexts. As a result, the user's personal system becomes a mediator for other computers and interfaces, providing a familiar, dependable interface and set of tools complementing the abilities the infrastructure provides (more processor power, additional sensing, etc.). In addition, this interface filters and manages the otherwise overwhelming flow of information. With sophisticated user models and corresponding software agents, such an interface can be extended to recognize and predict the resources needed by the user.

7 Conclusion

Wearable computing provides an exciting way to explore augmented realities and begins to fulfill the promise of a truly personal digital assistant. While wearable computing augmented realities imply the potential for significant additional informational sources for the user, the implementations described also provide better ways of managing such information. The result is a rich combination of physical and virtual realities that can intelligently assist the user instead of overwhelming him.

8 Acknowledgements

The authors would like to thank Ali Azarbayejani, Baback Moghaddam, and Kenneth Russell for providing access to

their 3D shape recovery, face recognition, and graphics code respectively. Thanks also to Michal Hlavac for his deft skill in 3D modeling and Fred Martin and Rob Poor for their IR transponder designs. Last, but certainly not least, thanks to Tavenner Hall for her constructive criticism and proofing early copies of this document.

References

Azarbayejani, A. and Pentland, A. (1995). Recursive estimation of motion, structure, and focal length. *IEEE PAMI*, 17(6).

Baker, F. (1994). Assistive technology for the visually impaired. *http:// www.wilmer.jhu.edu/ low_vis/ lves.htm.*

Baum, W., Ettinger, G., White, S., Lozano-Pérez, T., Wells, W., and Kikinis, R. (1996). An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Trans. Medical Imaging*, 15(2):129–140.

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

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

Huang, T. and Netravali, A. (1984). Motion and structure from feature correspondences: a review. *Proc. of IEEE.*

Kraut, R., Miller, M., and Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. In *forthcoming ACM Conference on Computer Supported Cooperative Work (CSCW)*, Boston, MA.

Lamming, M. and Flynn, M. (1994). Forget-me-not: Intimate computing in support of human memory. In *FRIEND21: Inter. Symp. on Next Generation Human Interface*, pages 125–128, Meguro Gajoen, Japan.

Mann, S. (1994). Mediated reality. Technical Report 260, MIT Media Lab, Perceptual Computing Group.

Mann, S. (1995). Video orbits of the projective group; a new perspective on image mosaicing. Technical Report 338, MIT Media Lab, Perceptual Computing Group.

Nagao, K. and Rekimoto, J. (1995). Ubiquitous talker: Spoken language interaction with real world objects. In *Proc. of Inter. Joint Conf. on Artifical Intelligence (IJCAI)*, pages 1284–1290, Montreal.

Nagel, D. and Lovette, J. (1995). Petition for rulemaking: NII band. *Open letter to the FCC: http:// www.research.apple.com/ research/ proj/ niiband/ petition/ NII_Petition.html.*

Orwant, J. (1993). Doppelganger goes to school: Machine learning for user modeling. Master's thesis, MIT, Media Laboratory.

Pentland, A., Moghaddam, B., and Starner, T. (1994). View-based and modular eigenspaces for face recognition. In *IEEE Conference on Computer Vision & Pattern Recognition (CVPR'94)*, Seattle, WA.

Picard, R. (1995). Affective computing. Technical Report 321, MIT Media Lab, Perceptual Computing Group.

Poor, R. (1996). iRX 2.0. *http:// ttt.www.media.mit.edu/ pia/ Research/ iRX2/ index.html.*

Rhodes, B. and Starner, T. (1996). Remembrance agent: A continuously running automated information retreival system. In *Proc. of Pract. App. of Intelligent Agents and Multi-Agent Tech. (PAAM)*, London.

Salton, G. (1971). *The SMART Retrieval System - experiments in automatic document processing.* Perntice-Hall, Inc., Englewood Cliffs, NJ.

Schmandt, C. (1994). *Voice Communication with Computers.* Van Nostrand Reinhold, New York.

Starner, T., Mann, S., and Rhodes, B. (1995). The MIT wearable computing web page. *http:// wearables.www.media.mit.edu/ projects/ wearables.*

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Uenohara, M. and Kanade, T. (1994). Vision-based object registration for real-time image overlay. Technical report, Carnegie Mellon University.

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