# Augmented Realities Integrating User and Physical Models

Thad Starner, Bernt Schiele, Bradley J. Rhodes, Tony Jebara
Nuria Oliver, Joshua Weaver and Alex Pentland
{testarne,bernt,rhodes,jebara,nuria,joshw,sandy}@media.mit.edu
Media Laboratory, Massachusetts Institute of Technology

**Abstract**

Besides the obvious advantage of mobility, wearable computing offers intimacy with the user for augmented realities. A model of the user is as important as a model of the physical world for creating a seamless, unobtrusive interface while avoiding "information overload." This paper summarizes some of the current projects at the MIT Media Laboratory that explore the space of user and physical environment modeling.

## 1 Introduction

Wearable computers have the potential to "see" as the user sees, "hear" as the user hears, and experience the life of the user in a "first-person" sense. They are also physically and mentally more intimate with the user and are used for extended periods of time, allowing a unique opportunity to model user's usage patterns and habits. This increase in user and environmental information can lead to more intelligent and fluid interfaces that use the physical world as part of the interface.

This paper summarizes several of the current wearable computing and augmented reality research projects at the MIT Media Laboratory that explore the dimensions of user and physical modeling. For more complete information on a particular project and related work, the reader is encouraged to refer to the original papers on these projects.

## 2 Remembrance Agent

The Remembrance Agent is a program that continuously "watches over the shoulder" of the wearer of a wearable computer and displays one-line

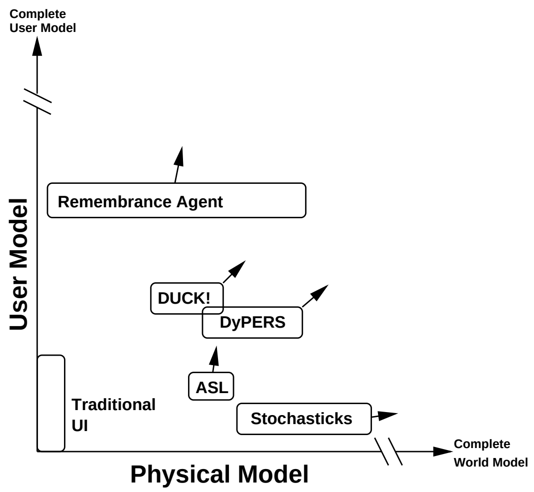


Figure 1: Wearable computing projects mentioned in this paper positioned along conceptual axes of how much each project involved a user model versus how much the project models the physical environment of the user. The arrows indicate the direction of current research.

summaries of notes-files, old email, papers, and other text information that might be relevant to the user's current context[Rho97]. These summaries are listed on a head-up display so the wearer can view the information with a quick glance. The full text can be retrieved using a one-handed keyboard.

The wearable version of the RA uses physical sensors to model the user's environment. The RA continuously watches sensors on the wearable as well as the notes being entered by the user to suggest the documents from a set of pre-indexed text that are "most relevant" to the user's current situation. For example, a user's context might be described by a combination of the current time of day and day of the week (provided by the wearable's system clock), location (provided by an infrared beacon in the room), who is being spoken to (provided by an active badge), and the subject of the conversation (as indicated by the notes being taken). The suggestions provided by the RA are based by a combination of all these elements, using text-retrieval techniques similar to those used in web search engines.

## 2.1 Augmented Reality Remembrance Agent

The wearable RA uses an overlay display, but does not register its annotations with specific objects or locations in the real world as one might expect from a full augmented reality system. In many cases such a "real-world fixed" display wouldn't even make sense, since suggestions often are conceptually relevant to the current situation without being relevant to a specific object or location. To examine augmented reality interfaces, a different version of the RA was implemented using a desktop computer, HUD, and head-mounted camera. The wearer of the system viewed the world through the camera, and the camera output also went to an SGI reality engine. Around the room were colored tags, which a vision system could detect. The size and the shape of the tags are used to determine distance and orientation of the tag (see Figure 2). Color-coding was used to identify each tag.


Figure 2: Multiple graphical overlays aligned through visual tag tracking. The color code in each tag provides a unique ID for the object. In addition, the vision process tracks each tag in 2.5D as the head-mounted camera moves.

Whenever a tag was detected the code-number was looked up in a table and a pre-computed message was overlayed on top of the object being viewed. Because the system could detect the distance of the tag, more information would be displayed as the user came closer to a tagged object. Thus, the act of approaching a tag was equivalent to clicking on a hypertext link.

On top of this system, the Remembrance Agent showed the top suggestions for an individual tag. This created a two-level information system, with some information being provided by the infrastructure (tied to the tag via a lookup table), and some information provided from the user's own files via the Remembrance Agent.

## 2.2 Dynamic Personal Enhanced Reality Agent

A recent extension of the system described above uses a generic object recognizer to identify objects instead of tags. The system, called "Dynamic Personal Enhanced Reality System" (DyPERS, [JSOP99]), retrieves audio and video clips based on associations with real objects. The generic object recognizer is based on a probabilistic recognition system [SC96] which is

Figure 3: A DyPERS user listening to a guide during a test art gallery tour

Figure 4: Left: The Patrol cap with two cameras. The larger, visible camera is mounted facing downward. The second camera faces forward and is hidden by the brim. Right: Images from the Patrol cap. The left and right images are from the downward-looking and forward-looking cameras respectively.

capable of discriminating more than 100 objects in the presence of major occlusions, scalings and rotations. While 100 objects is not enough to be practical in an unconstrained environment, the number of possible objects can be significantly reduced using the location of the user, time of day and other available information.

## 3 User-observing wearable cameras

In the previous section, head-mounted camera systems face forward, observe the same region as the user's own eyes. By changing the angle of the camera to point down, the user's own body can be tracked. This allows the user's hands, feet, torso, and lips to be observed without the gloves or body suits associated with virtual reality gear.

Two projects currently use this camera orientation. The first attempts to translate American Sign Language to English by tracking the user's hands via the downward looking camera. The wearable ASL recognition system outperforms equivalent desk-based camera systems and most dataglove-based systems in recognition accuracy. For five word sentences comprised from a forty word lexicon, the system achieves 96.8% word accuracy with an unrestricted grammar (any word is possible, any number of times, in any order) [SWP98].

The second system "DUCK!" begins to demonstrate how such methods may be useful to augmented realities. Using only two video views, DUCK! tracks the wearer's current location and task. DUCK!'s domain is limited to the real-space game Patrol which is played by MIT students every weekend in a campus building. Participants are divided into teams and aggressively hunt each other with rubber suction-cup dart guns through 14 rooms or areas. DUCK! monitors the average color and luminance values of the floor, the scene in front of the player, and the player's nose as a lighting calibration image. These images are pulled from the two video streams at 6 frames per second to determine location. With 24.5 minutes of training video and 19.3 minutes of test video, the system performs with 82% accuracy [SSP98]. DUCK! also attempts to discriminate between the player's tasks by identifying hand gestures representing aiming/shooting, reloading, and "everything else" through a combination of a generic object recognition system [SC96] and the HMM's used in the ASL task above. Preliminary results show 86% accuracy in distinguishing these classes. Other user actions such as standing, walking, running, and scanning of the environment can be considered as tasks than can run concurrently with other actions. Future work will address these tasks.

While preliminary, the systems described above suggest how context perception may be used in augmented reality interfaces. Through head-up displays, the players can keep track of the team's positions, even to the extent of "seeing through walls" to what may be occurring several rooms away. If aim and reload gestures are recognized with a particular player's system, his position can be highlighted in the rest of the team's displays indicating he needs aid. Furthermore, when the computer recognizes a player to be in a battle, the computer should inhibit his interface to avoid interruptions.

# 4 Stochasticks

Stochasticks is a practical application of wearable computing and augmented reality which enhances the game of billiards [JEW+97]. Probabilistic color models and symmetry operations are used to localize the table, pockets and balls through a video camera near the user's eyes. Classification of the objects of interest is performed and each possible shot is ranked

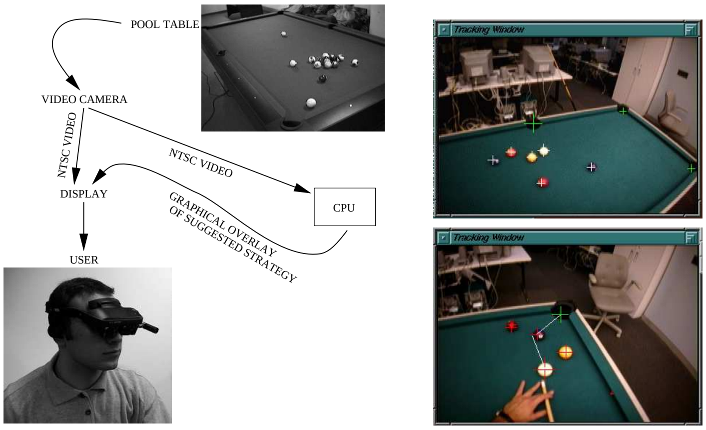


Figure 5: Left: The system components. Right Top: Finding the balls. system. Right Bottom: Suggested shot.

in order to determine its relative usefulness. The system allows the user to proceed through a regular pool game while it automatically determines strategic shots. The resulting trajectories are rendered as graphical overlays on a head mounted live video display. The wearable video output and the computer vision system provide an integration of real and virtual environments which enhances the experience of playing and learning the game of billiards without encumbering the player.

## 4.1 The System

A wearable computer is the hardware platform for the system and it includes a head-mounted display, head-mounted video camera and central processing unit. The head-mounted camera is a miniature ELMO 2.2mm video camera, mounted on the head-up display and aligned with the orientation of the eyes. Thus, the user's head direction will automatically direct the camera to areas of interest in the scene. The head-mounted display consists of a Virtual I/O 3D display (or a Sony Glasstron) which allows the CPU to project semi-transparent imagery into each eye via two separate CRTs at about 10Hz.

Once ball position is known, the easiest possible shot for a given player is computed, and the shot trajectory is projected onto the user's eye. At this point, we are undertaking a performance analysis of the overall system. The reliability of the algorithm is being investigated as well as its accuracy for both 2D and 3D overlays.

## 5 Conclusion

We have demonstrated how computer vision can be incorporated into augmented realities. Self-observing wearable camera systems were discussed which identify the user's gestures and location in a variety of conditions. Finally, through the projects presented, we have shown how both modeling of the user and of the physical world play an important role in augmented realities.

## References


[JEW+97] T. Jebara, C. Eyster, J. Weaver, T. Starner, and A. Pentland. Stochasticks: Augmenting the billards experience with probabilistic vision and wearable computers. In *International Symposium on Wearable Computers*, 1997.

[JSOP99] T. Jebara, B. Schiele, N. Oliver, and A. Pentland. Dypers: dynamic and personal enhanced reality system. In *Submitted to Intl. Conference on Computer Vision Systems*, 1999. also TR 463, M.I.T. Media Lab, Perceptual Computing Section.

[Rho97] B. Rhodes. The wearable Remembrance Agent: A system for augmenting memory. *Personal Technologies*, 1(1), 1997.

[SC96] B. Schiele and J Crowley. Probabilistic object recognition using multidimensional receptive field histograms. In *International Conf. on Pat. Rec.*, volume B, pages 50–54, August 1996.

[SSP98] T. Starner, B. Schiele, and A. Pentland. Visual conextual awareness in wearable computing. In *Second International Symposium on Wearable Computers*, 1998.

[SWP98] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer-based video. *IEEE Trans. Patt. Analy. and Mach. Intell.*, To appear 1998.