SMART ROOMS, DESKS, AND CLOTHES

Alex Pentland

The Media Laboratory
Massachusetts Institute of Technology

ABSTRACT

We are working to develop smart networked environments that can help people in their homes, offices, cars, and when walking about. Our research is aimed at giving rooms, desks, and clothes the perceptual and cognitive intelligence needed to become active helpers.

1. THE OPPORTUNITY

Inanimate things are coming to life. However these stirrings are not Shelley's Frankenstein or the humanoid robots dreamed of in artificial intelligence laboratories. This new awakening is more like Alice in Wonderland or Walt Disney: the simple objects that surround us are gaining sensors, network connections, computational powers, and actuators. Desks and doors, TVs and telephones, cars and trains, eyeglasses and shoes, and even the shirts on our backs....all are changing from static, inanimate objects into adaptive, reactive systems that are more useful and efficient.

Some examples: Imagine a house (or a city!) that always knows where your kids are, and tells you when they might be getting in trouble. Or an office that knows when you are in the middle of an important conversation, and shields you from interruptions. Or a car that knows when you are sleepy, checks the network to find a nearby open coffee shop, and suggests that you stop there for a while. Or glasses that can capture the face of person you just met, query the network for their name, and whisper it in your ear.

These examples are not a long-range fantasy: my research group and I, with the help of Profs. Rosalind Picard and Pattie Maes, are building such research prototypes today. The key to changing inanimate objects like offices, houses, cars, or glasses into smart, active helpmates is to teach them how to pay attention to us the way another person (or even a dog!) would. That way they can adapt their behavior to us, rather than the other way around.

1.1. The Challenge: Making things smart and attentive

How can common objects become smart and attentive? After all, Artificial Intelligence researchers have been trying to build intelligent machines for thirty years, with little success outside of abstract mathematical domains or carefully controlled factory environments.

We believe that the main problem has been that current computers are dumb, deaf and blind. The are dumb because they are *static*. They do only what they were programmed to do, and do not learn to adapt to changing circumstances. The are deaf and blind because they mostly experience the world around them through a slow serial line to a keyboard and mouse. Even "multimedia computers," which can handle signals like sound and image, do so mainly as a transport device that knows nothing of the signals' content.

If you imagine raising a learning-impaired child in a closed, dark, soundproof box with only a telegraph connection to the outside world, you can quickly realize how difficult it is for computers to become intelligent and helpful. They exist in a world that is almost completely disconnected from ours, so how can they understand and help us?

We believe that there are three keys to making the objects around us smart and helpful:

First, they need to share our perceptual environment before they can be really helpful. They need to be situated in the same world that we are; they need to know much more than just the text of our words. They also need to know who we are, see our expressions and gestures, and hear the tone and emphasis of our voice.

Second, they need to learn. They need to watch what we do in each situation, and learn our preferences and habits. Filtering and matchmaking agents need to know your likes and dislikes; remembrance agents need to know what you've seen and heard before; and negotiating agents need to know your goals and values.

Third, they have to share knowledge with each other. Without communications between these learning, perceptive agents, they can get only a narrow view of the world, and consequently will be limited in what they can learn and what they can do. Moreover, they have to communicate to share knowledge, computational resources, and actuators. Consequently, smart objects have to communicate with each other in order to be smarter, more efficient, and more effective.

Each of these elements are critical. Without a good representation of the situation (e.g., perception) it is impossible to know what action to take. Without learning about the user's habits and preferences, it is impossible to be helpful in any but the most general way. And without communication, it is impossible to know enough about the situation to learn or act efficiently.

2. THREE EXPERIMENTAL TESTBEDS

To conduct research on making things pay attention to people, we have created a series of experimental testbeds. These testbeds are "smart environments" that help users navigate the World Wide Web, to teleconference in virtual worlds, to find multimedia information, and to interact with artifical life agents. They accomplish this by using new technologies like software agents for negotiation and memory augmentation, and computer vision for face recognition and head/hand/eye tracking.

Figure 1. Smart Clothes: The Cyborgs of the M.I.T. Media Laboratory.

The testbeds can be divided into three main types: smart rooms, smart desks, and smart clothes. The idea of a smart room is a little like having a butler; that is, a passive observer who usually stands quietly in the corner but who is constantly looking for opportunities to help, and who knows your preferences so well that they can act without detailed instructions. A smart desk is similar, but aimed at the office work environment rather than at living spaces in general; it is perhaps more like a good secretary. Smart clothes, on the other hand, are more like a personal assistant. That is, they are like a person who travels with you, seeing and hearing everything that you do, and who tries to anticipate your needs and generally smooth your way. Some of our prototype smart clothes are shown in Figure 1.

Smart rooms, desks, and clothes all have embedded computers. Unlike todays computers, however, they are also instrumented with sensors (mainly cameras and microphones, but also biosensors for heartrate and muscle action), which allow the computer to see, hear, and interpret users' actions. They are networked together by IR ethernet, wire ethernet, and ISDN.

People in a smart room or at a smart desk can control programs, browse multimedia information, and experience shared virtual environments without keyboards, special sensors, or special goggles. People in smart clothes can obtain personalized information about their environment, such as the names of people they meet or directions to the next meeting, and can replace most computer and consumer electronics. The key idea is that because the room, desk or clothing knows something about your situation, it can react intelligently.

3. CURRENT DEMONSTRATIONS

3.1. No Wires, Keyboards, or Mice

In many applications it is desirable to have an interface that is controlled by gesture rather than by a keyboard or mouse. In a smart room or desk the position of the user and the configuration of the user's appendages can be mapped into a control space, while sounds and gestures made by the user are used to change the operating mode.

This allows the user to control an application with their body directly. This interface has been used to navigate in networked 3-D virtual game environments, and as an interface to various visualization applications. Figure 2(a) shows playing a shoot-em-up virtual game M.I.T., the other participant is at BT in Martlesham; they are connected by an ISDN line.

3.2. Playing With Artificial Creatures

The cameras, microphones and other sensors of a smart room or smart desk can be the eyes and ears for artifical life creatures. For instance, in the Artificial Life Interactive Video Environment (ALIVE) the smart room's description of the user's shape is used to composite a video model of the user into a virtual reality scene populated with computer-generated artificial life forms, as illustrated in Figure 2(b). Information about the users' gestures, sounds, and position are used by the artificial life forms to make decisions about how to interact with the user.

3.3. Talking With Deaf People

American Sign Language (ASL) is a sophisticated set of hand gestures that allow deaf people to communicate more naturally. When hearing people first discover ASL, they find it surprising because it allows communication as complex and rapid as speaking. ASL, therefore, is a good test for our smart room's ability to interpret rapid, complex human gesture.

Thad Starner and I therefore set out to build a real-time system to read ASL. We started building HMM models of each sign by observing many examples of the pattern of hand measurements that are output by Pfinder (e.g., hand position, orientation, and width/height ratio). We then found that by comparing these HMM models to Pfinder measurements of a ASL signing we could obtain 99.2% accurate classification of a forty-word subset of ASL in real time. Thad Starner is shown using this system in Figure 3. The ability to recognize ASL in real time opens the possibility of interfaces for deaf people that will match the speech recognition systems now being introduced for the hearing.
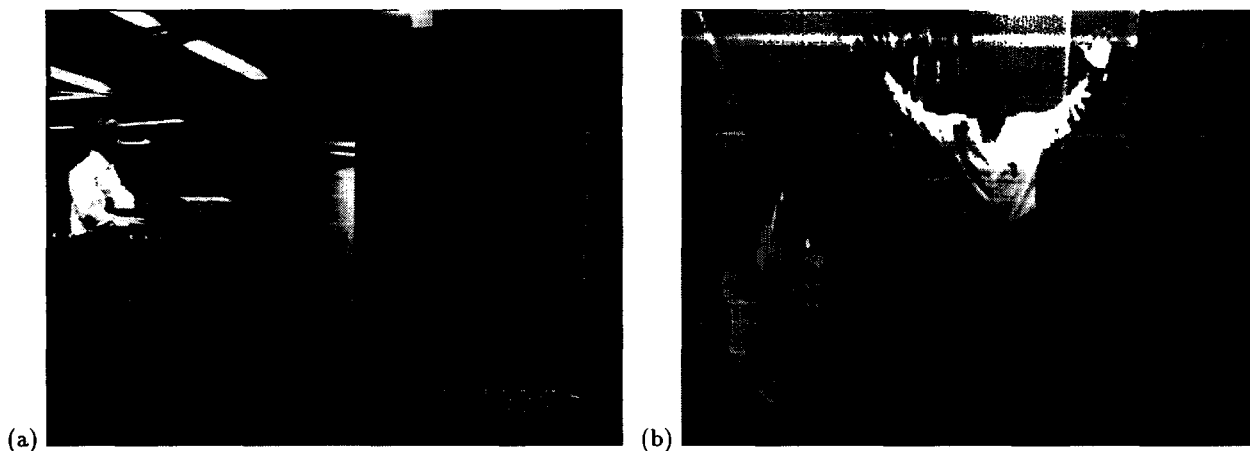
(a)

(b)

Figure 2. (a) navigating through a virtual world while playing a shoot-'em-up game, (b) Chris Wren playing with Bruce Blumberg's virtual dog.

Figure 3. Real-time reading of American Sign Language (with Thad Starner doing the signing)

3.4. The Cyborgs Are Coming

In a smart room the cameras and microphones passively watch people move around. However when we build them into a person's clothes, the computer's view moves from a passive third person to an active first-person vantage point. This means that smart clothes can be more intimately and actively involved in the user's activities, making them potentially a real Personal (Digital) Assistant.

Augmenting human memory is a major application for smart clothes. For instance, when we built a camera into eyeglasses, we found we could use our face recognition software to help remember the names of people. When we met a person again, the computer would recognize them and whisper their name in our ear.

Similarly we found that a computer built into your clothes can automatically remind you of important facts that are related to your current conversation. For instance, if you have a discussion about "the Megadeal contract," your smart clothes computer can automatically project Megadeal's finances onto the display built into your glasses.

Our current smart clothes prototypes use off-the-shelf head mounted displays (HMD's) to provide privacy and convenience. Their CPU's are designed are small and unobtrusive, and alternative input devices have been developed to utilize these machines in just about any context. They use IR and wireless ethernet to stay in contact with the entire World Wide Web at all times.

The goal is that eventually such devices will be so small

and light that they will be worn constantly, much as eyeglasses and clothing are now today, thus providing access to computing power at all times. Today's smart clothes are not yet inconspicuous, particularly the HMDs, so they project a rather CyberPunk look (see Figure 1). However the coming of continuous computing is not far off; two of the Media Lab Cyborgs, Thad Starner at the extreme left and Steve Mann at the extreme right, already wear their devices all day every day.

4. CONCLUSION

It is now possible to track people's motion, identify them by facial appearance, and recognize their actions in real time using only modest computational resources. By using this perceptual information we have been able to build smart rooms and smart clothes that have the potential to recognize people, understand their speech, allow them to control computer displays without wires or keyboards, communicate by sign language, and warn them they are about to make a mistake.

We are now beginning to apply such perceptual intelligence to a much wider variety of situations. For instance, we are planning prototypes of displays that know if you are watching, credit cards that recognize their owners, chairs that adjust to keep you awake and comfortable, and shoes that know where they are. We imagine building a world where the distinction between inanimate and animate objects begins to blur, and the objects that surround us be-

come more like helpful assistants or playful pets than insensible tools.

5. REFERENCES

For additional information, including technical papers and on-line demonstrations, please visit our web site at http://vismod.www.media.mit.edu/vismod