The New EasyLiving Project at Microsoft Research

Steve Shafer, John Krumm, Barry Brumitt, Brian Meyers, Mary Czerwinski, Daniel Robbins
Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Abstract

EasyLiving is a new project in intelligent environments at Microsoft Research. We are working to make computing more accessible and more pervasive than today's desktop computer. More specifically, our goal is to develop a prototype architecture and technologies for building intelligent environments that facilitate the unencumbered interaction of people with other people, with computers, and with devices. This paper describes our goals, design decisions, and applications of EasyLiving.

1. What Is Next in PCs?

Software developers have had a long time to exploit the capabilities of the PC. While new applications for stand-alone PCs are still coming, it is primarily new connected devices that generate new applications and new markets. For instance, inexpensive color printers spawned desktop publishing for the consumer. Digital cameras are creating consumer demand for photo editing software. The World Wide Web (essentially a way of connecting other computers to your own) is giving us unprecedented access to information and a new reason to own a computer.

We are looking at the physical home and work environments as the next things to connect to a PC. Not only will this encourage new applications, it may allow more natural interaction with computers, reducing the barriers of inconvenience that prevent computers from being used for more everyday tasks.

2. Goals of EasyLiving

EasyLiving is a new project at Microsoft Research with its genesis in the Vision Technology Group. Our goal is to develop a prototype architecture and technologies for building intelligent environments that facilitate the unencumbered interaction of people with other people, with computers, and with devices. We are concentrating on applications where we can make computers easier to use for more tasks than the traditional desktop computer. We envision a home or office of the future in which computing is as natural as lighting. It maintains an awareness of its occupants through computer vision, responds to voice and gesture commands, knows its own geometry and capabilities, and can be easily extended. The technology we are developing will, for instance, enable a home's resident to make a phone call by simply speaking his intentions from anywhere that he happens to be. The home will keep track of children and pets automatically. It will allow a user to move from room to room while still maintaining an interactive session with the computer, with the user interface migrating along.

Being new, most of our work to date has been in conceptualizing and planning. This paper describes our goals, plans, and early milestones for EasyLiving. In order to make this pervasive yet unobtrusive style of computing successful, our intelligent environment must have three characteristics that we detail in this section: self-awareness, casual access, and extensibility. Given the luxury (and burden) of building a new intelligent environment from scratch, we are faced with many concrete design decisions such as the system's user interface and software architecture. We present our conclusions on some of these issues in Section 2. Section 3 describes some applications of our system, including the first demonstration that we recently completed

2.1 Self-Aware Spaces

EasyLiving spaces must be aware of their own activity and contents to allow appropriate responses to the movement of people and their requests. Such a "self-aware space" knows its own geometry, the people within it, their actions and preferences, and the resources available to satisfy their requests. Some examples illustrate the importance of self-awareness:

- As people walk around in the space, they will move through the fields of view of the rooms' video cameras. The system must know the 3D regions covered by the cameras to know which room a given person is in and from which other cameras she may soon be visible. This will allow, for instance, ringing the telephone only in the room where the intended callee is located and migrating a user interface along as the user moves from room to room.

- The system should be aware of the identity of the occupants. The system could sense "absolute" identity, *e.g.* "This is George Jetson", or it could, more simply, maintain "relative" identity, *e.g.* "This is the same person I just saw from camera 12." Such knowledge will enable EasyLiving to apply personal preferences for known occupants such as a contact list for telephone calls. It can also be used to block access to certain devices and data.
- EasyLiving must know what hardware and software resources are available to it and how to use them. If a user interface is to move, the system must know how to present it in the user's new location, *i.e.* whether the new location supports audio, speech, pointing, or visual display. It must understand which devices are already in use and whether each device is working or not.

More than just populating a space with intelligent devices, EasyLiving will maintain knowledge of and employ combinations of devices and software to satisfy the users' needs.

2.2 Casual Access to Computing

Users should not be required to go to a special place (*i.e.* the desktop) to interact with the computer. Nor should they be required to wear special devices or markers to have the computer know where they are. EasyLiving's goal of "casual access to computing" means that the computer will always be available anywhere in an EasyLiving space. Through cameras and microphones, the user will always be able to signal the computer. Since the computer will keep track of users and their contexts, the computer will always be able to signal the users in an appropriate way, and it will know how to avoid being obtrusive. For example, a user watching television could be notified via a superimposed window on the screen, while a sleeping user might not be notified at all, unless the message is important. Information access will be similarly versatile, with the system being able to present, say, an address book entry with whatever output device is available at the user's location in response to whatever input device is available at the user's location.

Combined with self-awareness, the goal of casual access leads to a migrating user interface. When a user moves, the user interface of the application can move with him. This would be useful for carrying on a phone conversation as a user moved throughout the home, or it could be leveraged to move an interactive session to a device with higher fidelity.

2.3 Extensibility

EasyLiving capabilities should grow automatically as more hardware is added. Extending the concept of "plug and play", new devices should be intelligently and automatically integrated. One aspect of extensibility is the view that new devices become new resources that the system can use at will. If, for instance, a CRT is added to the kitchen, it becomes a new way of presenting information in that space, and EasyLiving will automatically take advantage of it. This is an example of extensibility in terms of resources. Another aspect is extensibility in terms of physical space. If a new camera is added, it not only extends the system's resources as a new device, it also extends the system's physical coverage. This means that the system must be able to compute the position and orientation of the camera based only on what it sees through the new camera and any others that share its field of view.

3. Design Issues

Our goals for EasyLiving drive our design. This section discusses some of our particular design decisions for component technologies (sensing & modeling, user interface) and broader issues (software architecture and privacy).

3.1 Sensing and Modeling

EasyLiving spaces must respond to users' actions and words. While there are many types of single use sensors that can monitor people in a room, *e.g.* IR motion sensors and electromagnetic field sensors, video cameras are the most versatile, longest range, and best understood sensing modality for this task. Cameras give rich data that can be used for tracking and identifying people and objects and for measuring them in 3D. In the context of intelligent environments, cameras have been used to track people in Michael Coen's Intelligent Room at MIT's AI Lab[1] and to understand gestures in Mark Lucente's Visualization Space at IBM Research[2].

To accommodate the video sensing demands of EasyLiving, we are building a "vision module" that gives both color and range images. It will consist of between two and four cameras, packaged together, with control and processing done on one PC. We will use the color image to make color histograms, which have been shown to work well for identifying objects[3]. Our own experiments show that color histograms are effective at re-identifying people that have already been seen as long as their clothing doesn't change. The range images will come from passive stereo, and they will be used primarily for image segmentation. We plan to deploy several such vision modules in each room of an EasyLiving space. They will be used to detect motion, identify people, sense gestures, and model the 3D environment. Modeling is important so each vision module knows what parts of the room it can see and its location with respect to the other vision modules. Given this information, a person-tracker can anticipate which camera(s) will give the best view of a moving person.

We will also deploy microphone arrays in EasyLiving spaces. These will be able to "steer" toward people talking using signal processing algorithms.

3.2 User Interface

EasyLiving will use traditional, desktop user interfaces where appropriate, and also more advanced user interfaces where possible. In general, the user will be able to choose his or her own interface mode, constrained only by the devices available in the room. For instance, a user might start an interaction using a wireless keyboard and mouse in the family room and then move to the kitchen, continuing the same interaction but with voice and gestures instead. This goal spawns two research issues: migrating user interfaces and multimodal user interfaces.

The migrating user interface, such as the family-room-to-kitchen example above, requires that an application, or at least its user interface, be able to move smoothly from one room to another. In their work on the Obliq distributed scripting language, Bharat and Cardelli[4] used a software architecture that moves whole applications between computers using agents. The application, including its user interface, is packaged as an agent, and each computer contains software that can receive such an agent and start it running. As implemented, this scheme cannot account for changes in user interface modality beyond a change in screen size.

In our initial demonstration, we achieved a simple migrating user interface using Microsoft Terminal Server, which allows Windows NT 4.0 to host multiple clients with windows appearing on networked PC's. In the end, however, we want the user interfaces of EasyLiving applications to change with the desires of the user and the available devices. This will require that the applications be written with an abstracted user interface, which is very different from the style of GUI programming today. However, we also see EasyLiving as a new way to run current applications, with an intermediate software layer that could, for instance, reinterpret pointing gestures as mouse movements and spoken commands as menu choices.

The other major user interface question asks: If users could use more natural ways of interacting with the computer, say audio and video, how would they do it? Audio and video output to the user is well-understood, while the use of microphones and cameras as input devices is not fully mature. At the extreme, EasyLiving could carefully monitor all the actions and speech of each user, intelligently interpreting what they mean. We don't expect to achieve this, and instead we will require the user to specifically address the computer for most interaction. (One exception is that the system will passively monitor the room with both cameras and microphones to know when it should be alert to possible user commands, avoiding the "push-to-talk" problem. It would then suspend any background activity like periodic geometric modeling to pay more attention to the user(s).)

Used apart, microphone and camera input to programs has been the attention of much research. Speech understanding is available commercially, and many computer vision researchers are working on the tracking and interpretation of human movement such as gestures. The use of speech and gestures simultaneously, however, is a relatively new area of research. Sharon Oviatt has studied the use of speech and gesture in a pen-based, geographic map application[5]. When given the choice to issue commands with a pen and/or voice, users preferred to convey locatives (*i.e.* points, lines, and areas) with the pen, while they preferred speech for describing objects and giving commands. For intelligent environments, this means that perhaps the most important gesture to detect is pointing, while other commands should, initially at least, be left for voice.

We have completed the first in a series of user studies to explore speech and gesture input for EasyLiving. Six subjects were led through a series of exercises using a limited, paper and pencil prototype scenario in which they placed a video conference call from a large display screen on the wall. Given a choice, users preferred speech over gestures, but they could effectively combine both. We collected a broad sampling of the kinds of gestures and speech commands that users generated spontaneously for this task. We have also run paper and pencil walkthroughs of potential 3D user interface designs that could be used for the more advanced user interface directions this project will take. Users provided useful feedback in terms of what metaphors they found to be most meaningful for the "Contact Anyone Anywhere" scenario (Section 4.2) we were exploring. Our goal is to next test the redesigned prototype in a series of "Wizard of Oz" studies using a large wall display.

3.3 Architecture for Extensibility

As described above, one of EasyLiving's goals is automatic extensibility. This appears to be a feature that has not been addressed in other intelligent environment research. Our system will automatically incorporate new devices as they are added. The architecture is shown schematically in Figure 1. At the beginning of an EasyLiving installation, the only "live" device will be a central server. This server will contain all the software necessary for running an entire EasyLiving system. It may, in fact, be remotely located and/or remotely maintained and updated. The central server will also maintain information that is global to the whole system, such as the current time and a directory of people.

Each room of the EasyLiving installation will have its own process called a "room server". When the room

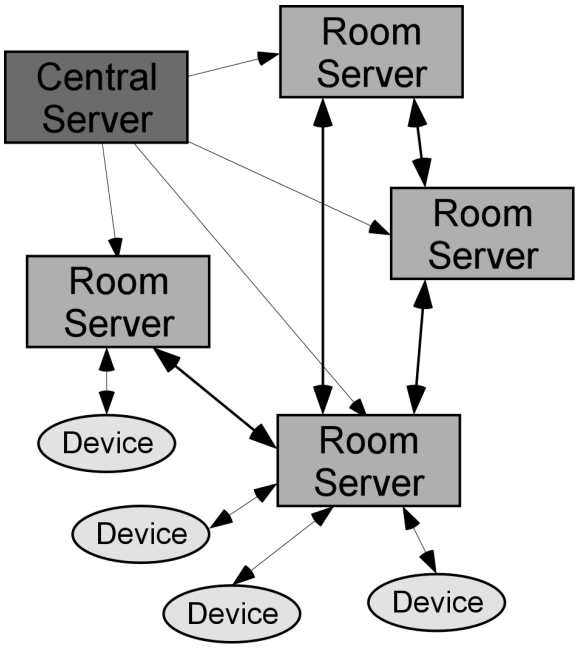

Figure 1: Software architecture, based on rooms and devices, allows for extensibility.

server is activated, it will announce itself as such to the central server and download the required software to make it a room server. This includes all the software necessary for the other processes that may run in the room. As other rooms are added to EasyLiving, they will also have room servers that start up in the same way. Each device added to a room, for instance a vision module, will connect to its room's room server and download the necessary software.

The room server will contain a model of the room, including its geometry, its contents, and locations of people. It will be connected to the room servers of adjacent rooms. These connections will be used to exchange information about overlapping fields of view of the rooms' cameras and to alert adjacent rooms that someone is about to enter.

This architecture will simplify the addition of new devices and new spaces to an EasyLiving system.

3.4 Privacy

In any intelligent environment, there is a tradeoff between privacy and convenience. The more the system knows about you, the more it can do for you, but the more it may reveal to someone else. Just having cameras and microphones in the room begs the question of "Who might be watching and listening right now?" There will also be symbolic data in the system that can be used to infer the users' habits and preferences. These concerns are made worse by the fact that the system could be passively gathering the data even while the it is not being actively used.

One way to make the system more secure with respect to outside snooping is encryption. We expect that as e-commerce becomes more common, it will provide publicly trusted encryption methods for transmitting data over networks. We may also want to enforce a policy of not transmitting any video over the EasyLiving network, choosing instead to do all computer vision at the camera and only transmitting results. If the user desires, the system can be set up such that it will not try to identify anyone unless they actively request it by using a password, cardkey, or biometrics. This means the system will not know who is in the space.

In general, privacy must be deeply rooted in the system with the tradeoffs made clear to users. There is not a single good answer to the question of making the system actually private and convincing users of the same.

4. Applications

We have completed our first demonstration of the EasyLiving system, and we have several more applications planned. We realize that we cannot predict what will be the most useful and popular applications for EasyLiving. We are confident, though, that the combination of capabilities that we provide will spark new applications.

4.1 Migrating Windows

Our first demonstration (July 1998) showed an implementation of a migrating user interface. We set up an office with three video cameras monitoring three "hotspots" – 3D regions of interest in the room where a user could go to interact with one of three video displays in the room. The locations of the hotspots were drawn as rectangles in the three camera views prior to running, as shown in Figure 2. To be considered in a hotspot, a user had to appear in the hotspot in at least two camera views. Each user began by logging into the system and starting an application at one of the displays. Based on this login, we knew which hotspot the user was in, so the system stored the color histograms of the corresponding image regions. Identifying users with their color histograms meant that the system could accommodate more than one user in the scene simultaneously. The system continuously monitored each hotspot, and when the histograms matched the stored histograms, the application window was moved to that display. The application's window was moved using Microsoft Terminal Server. Each camera had its own, dedicated PC, and a fourth PC ran the terminal server. The PC's communicated via sockets.

Our next demonstration (October 1998) will use the vision module (Section 3.1) to accomplish the same end.

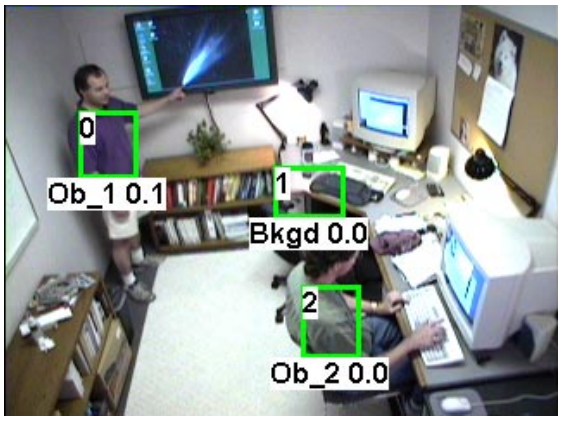

Figure 2: A view of the office from one camera, with three "hotspots". The user interface migrates between the three screens (including the large one on the back wall) as the user moves from hotspot to hotspot. Based on color histograms, the program has identified the contents of the hotspots as object 1, background, and object 2.

Using stereo and color together, we will track users instead of relying on hotspots.

4.2 Contact Anyone Anywhere

We will demonstrate several EasyLiving capabilities with our "Contact Anyone Anywhere" demonstration in April 2000. In this scenario, a user in an EasyLiving space will signal that she wants to place a call to someone else in an EasyLiving space. This signal may be given using a keyboard, mouse, voice, or gesture command. Since EasyLiving will know the user's identity, it will present her personal phone list, and the user will indicate which person she wants to call, using one of the same set of command modalities. EasyLiving will place the call, signaling the callee in the appropriate place and in an appropriate way. During the call, the user will switch to another UI device, move to another room, and finally pick up the call on a mobile device, all while maintaining the conversation.

4.3 Child Care Assistant

One attractive application of EasyLiving is to aid in caring for a child or a pet. EasyLiving could act as an enhanced child monitor, checking for dangerous conditions (*e.g.* near top of stairs), monitoring protected spaces, information, and vital signs. The system could also monitor a babysitter's time with children. If any condition required attention from a parent, the system could notify them in an appropriate way, including making a cellular telephone call. We plan a demonstration of these ideas in April 2000.

4.4 Vision-Based Home Automation

Having cameras in the room invites many interesting applications. For instance, cameras could make a video history of the space, recording during those times when motion occurs. The video history could be used to answer questions of the type, "What happened?" For instance, "Where did I leave my keys?" "What did the burglars look like and what did they take?" "How long has that vase been missing?"

Cameras could also be used to adjust light levels appropriately. If a person starts reading, the cameras can measure the ambient light and adjust lamps until the light is bright enough. Used as motion detectors, the cameras could cause the system to turn off lights in unoccupied rooms.

5. Summary

EasyLiving is looking beyond the desktop as the next step in computing for the everyday user. By connecting the home and work environment to the PC, we can provide casual access to computing. We are building an architecture and technologies to investigate and demonstrate this concept.

Acknowledgements

Thank you Pierre De Vries, Director of the Advanced Products Group at Microsoft for helping to get EasyLiving started from the beginning.

Thank you also to Charles P. Thacker, Director of Advanced Systems at Microsoft Research in Cambridge, England for contributing concepts to EasyLiving, in particular the ideas of self-aware spaces and casual access to computing.

References

[1] M. H. Coen, "Design Principals for Intelligent Environments," presented at AAAI Spring Symposium on Intelligent Environments, Stanford, CA, 1998.

[2] M. Lucente, G.-J. Zwart, and A. George, "Visualization Space: A Testbed for Deviceless Multimodal User Interface," presented at AAAI Spring Symposium on Intelligent Environments, Stanford, CA, 1998.

[3] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, pp. 11-32, 1991.

[4] K. A. Bharat and L. Cardelli, "Migratory Applications," presented at UIST '95, Pittsburgh, PA, 1995.

[5] S. Oviatt, "Multimodal Interactive Maps: Designing for Human Performance," *Human-Computer Interaction*, vol. 12, pp. 93-129, 1997.