The Challenges and Solutions in Turning HCI from Desktop to Smart Spaces

Enyi CHEN, Yuanchun SHI, Guangyou XU, Weikai XIE, Zhongnan SHEN, Yi CHE, Yanhua MAO

Department of Computer Science and Technology
The State Key Laboratory on Intelligence System and Technology
Tsinghua University, Beijing 100084
chenenyi00@mails.tsinghua.edu.cn

ABSTRACT

Smart Spaces introduce new challenges to interface designers due to a number of differences from traditional desktop-computing model, including multimodal interaction, context sensitive interaction, continuous interaction, dynamically focused interaction, physically embodied interaction and interaction based on dynamic set of device. This paper discusses these challenges and provides the solutions in our Smart Space system – Smart Classroom, an ongoing project in Tsinghua University. The system integrates dozens of hardware devices and several computer vision and speech recognition modules on an infrastructure, Smart Platform.

Keywords: Human-Computer Interaction, Smart Spaces, Ubiquitous Computing, Smart Classroom

1. INTRODUCTION

While interface developers continue to make the desktop PC more perfect, trends in Smart Spaces [3] as a branch of Ubiquitous Computing (UbiComp[1]), are opening up new possibilities for interacting with the real physical world. These trends introduce new challenges in turning HCI from desktop to Smart Spaces.

Smart Spaces (also called Intelligent Environments or IEs [2]) are work spaces embedded with computers, information appliances, and multi-modal sensors that allow people to work efficiently, together or individually, through unprecedented access to information and help from computers. Traditionally, the user only interacts with a desktop PC, which has a relatively standard and fixed set of user-interaction devices (mouse, keyboard, speaker and display). However, Smart Spaces are equipped with cameras as eyes and microphones as ears to make it possible to access the real events occurring in the work space. A multitude of computer-vision and speech-recognition modules then help interpret human-level events, such as what people are saying, where they are standing and what they are doing [5]. So, in Smart Spaces, instead of interacting with a desktop PC, the user maintains an ongoing interaction with a dynamic suite of devices in the work space, including many of which he may not even be aware [9].

There are many research institutes carrying out researches on Smart Spaces with different emphases. Interactive Workspaces in Stanford [10] is the multi-device, multi-user environments for people to work together in technology-rich spaces with computing and interaction devices on many different scales. Smart Rooms in MIT [11] have cameras, microphones, and other sensors, and use these inputs to try to interpret what people are doing in order to help them. Easy Living in Microsoft Research [12] is an intelligent environment for people to live easily. Dream Space in IBM Research [13] allows humans to interact with the real world in natural ways, using the common skills of speaking, gesturing, glancing, moving around and reaching out. Aware Home in Georgia Tech [14] builds a home, which provides services to its residents to enhance their quality of life and help them to maintain independence as they age. Despite of having different purpose, all of them must take consideration of interaction with real world.

We are still interested in exploring the impact of ubiquitous computing on tele-education, which leads to the project of Smart Classroom [6] [7]. It is an augmented physical workspace, a Smart Space type of classroom, which integrates natural human computer interface with CSCW modules to allow lecturers naturally and efficiently use computers to give lessons to distance learning students.

This paper, from the view of Smart Classroom, presents the challenges and solutions in turning HCI from Desktop to Smart Space. It is organized as the following: The next section gives a scenario of Smart Classroom. Section 3 discusses the challenges of interaction presented by shift from desktop to Smart Space. Section 4 provides the overview of Smart Classroom and describes the solutions to these challenges in it. The final section ends with the conclusion and our future work.

2. SCENARIO: SMART CLASSROOM

When Prof. SHI walks into the smart classroom, her biometrics features are automatically detected and the room recognizes her. So the classroom says "Good morning, Prof. Shi, you can give your lecture now." She goes to the blackboard and finds out that her lecture for Distributed Multimedia, displaying on the board just as where she stopped in the last class and the representing image of remote students, who attend her lecture by the Internet, are also displayed on the Student-board. Then Prof. SHI begins her lecture.

When she gives her presentation, just like in a traditional classroom, she can make annotations on the blackboard with electronic chalk. Besides, she can also control the display of the lecture conveniently by her speech, her gesture or the laser pointer in her hand. When the remote students meet some difficulties on understanding her lecture, they can consult Prof. SHI, just like the local student. For example, the classroom says, "Tom has a question now." Prof. SHI replies "Tom, go ahead." Then Tom's audio and video are presented in the classroom.

The remote students can get the live video and audio of the classroom with different focus according to what the teacher is illustrating. For example, when Prof. SHI makes annotations on the blackboard, the video camera will focus on blackboard; when she explains a realia in her hand, the video camera will focus on the realia; when she interprets a concept on the dais, an overview scene of the classroom will be fed to the remote students.

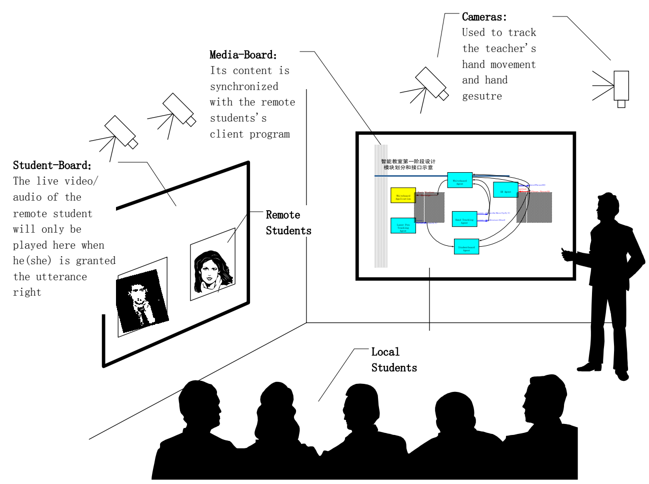

Figure 1. Scenario of Smart Classroom

In the above given scenario, as the figure 1 shows, the user interfaces are not menus, mice and keyboards as in the traditional desktop-computing paradigm, but gesture, speech, context, activity, physical equipment and even biometrics feature instead. By detecting Prof SHI's biometrics features with computer-vision and speaker-verification module, the classroom can identify Prof Shi, and consequently know what course she teaches and the left-point of her last class. Her actions in giving of the lecture, such as what she is saying, where she is standing and what she is doing, are captured and interpreted by a multitude of perception modules. And according to the context, the classroom decides which video should be delivered to remote students.

3. THE CHALLENGES OF INTERACTION IN SMART SPACES

3.1. Multimodal Interaction

In the conventional PC, the user sits in front of the desktop with a display screen, keyboard, mouse, and possibly speakers and microphone ready to hand. Most programs interact with user only via a "GUI" interface using the display, keyboard, and mouse. Thus, the speakers and microphone are primarily to enhance multimedia experiences such as videoconferencing, watching movies, or playing games.

In Smart Spaces, however, the user is not assumed to be sitting in front of a desktop. Therefore, the "hands-on" input devices – keyboard and mouse – cannot be assumed for interaction. There might even be no display available. In such cases, other forms of input and output are required, such as speech conversation. Thus, it is need to create applications that can take advantage of whatever devices are available and use all possible modalities to interact with the user. On the desktop, multimodal interfaces are specialties used for limited purposes; in Smart Spaces, they are norm. In the above scenario, Prof. SHI can make annotation by touching the blackboard when she stands in front of it, and when she stands away from the blackboard, she can still interact with it by speech, gesture and laser pointer.

3.2. Context Sensitive Interaction

Dey defined context as "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application [8]''. The use of context is currently being explored in desktop scenarios, but context can be more richly used in Smart Spaces as proposed by Coen [5]. In smart spaces, context-awareness implies two abilities of a system: the ability to obtain context and the ability to utilize contextual information to detect user intent.

In the above scenario, the biometrics feature of the occupant and the time are important contexts because different lecturers may give presentations on different subjects, and even the same lecturer may give different presentations at different time. For example, when Prof. SHI enters Smart Classroom, the slides for the Distributed Multimedia course will be displayed on the Smart Board. But when Prof. XU enters the Smart Classroom, what it shows might be the slides for Computer Vision course. For Prof SHI, she gives the Distributed Multimedia class on Monday and Ubiquitous Computing class on Friday. Location information is also an important context. For example, when Prof. SHI makes annotations on the blackboard, the delivered video should be focused on the blackboard's content; when she shows a realia, the delivered video should be switched to the realia.

3.3. Continuous Interaction

Just as extending the availability of computing beyond the traditional desktop fundamentally changes the relationship between humans and computers, providing continuous interaction changes computing from a localized tool to a constant companion [4]. Designing for continuous interaction requires addressing interruption and resumption of interaction, representing passages of time and providing associative storage models.

In the above scenario, Prof SHI may give the same lecture at different Smart Classrooms or at the different time. It will be convenient for her if the presentation of slides can begin from where she left last time. She does not need to care about where her lecture slides are stored and not need to memorize the previous state. As long as she enters the Smart Classroom, all the parameters are configured automatically, so she can focus her attention on the lecture itself, rather than on recovering the configuration.

3.4. Dynamically Focused Interaction

In a desktop scenario, the user is typically focused on a single screen where all current tasks are displayed. A single computational entity manages all the interface devices and routes events to the appropriate place. So the user has a conceptual model of where information relevant to the current task will be displayed. However, in Smart Spaces, the user is likely to be interacting with a set of devices that are connected loosely across a network. Thus, the knowledge and interaction are distributed.

The use of contextual information is crucial to routing the message to the user in the most appropriate way. In the previous scenario, the interaction's focal point is different at different time or different location. If Prof SHI makes annotations on blackboard by touching, the focal point is on the board. But when she shows the realia, the focal point is on the device. Another focal point is the input event. For example, if a remote audience has a question to ask Prof. SHI, the student should give a sound tip to her. But if the Prof SHI is busy in discussing with the local students, the tip should be given later to avoid confusion in the smart classroom.

3.5. Physically Embodied Interaction

In additions to explicit communication with the computer through traditional means, Smart Spaces allow physical objects become additional UI devices. A physical object is used in its normal manner of operation, but this use is enhanced by the Smart Spaces. In other word, the physical object itself is used for a specific physical purpose by the user, but Smart Spaces add a UI role to it.

In the previous scenario, it is a normal manner for Prof. SHI to take a realia, but the computer vision module detects the event and switches the realia video to remote audience. Thus the realia becomes input UI device through it is only a physical object without any sensors.

3.6. Interaction Based on Dynamic Set of Devices

A PC's operating system typically routes the signal from the attached devices to the appropriate application. This routing takes place inside the PC via the built-in software and/or hardware architecture. However, in Smart Spaces, this routing is much more complex. The devices involved in a given interaction can have separate network identities, and the signals may need to be routed access the network to the machine where the actual application resides.

Dynamic set of devices implies the need for interfaces that reconfigure themselves depending on the set of devices available. For example, currently, an instant messaging client on a cell phone looks very different from the one PC. Ideally, it would be possible for user to seamlessly move an instant message conversation from a PC to a cell phone， to a voice system in their car， or to a handheld device, with state preserved across the transitions. Context is critical to manage these transitions so that the appropriate interface can be provided based on the available devices and without requiring manual configuration by the user. (This issue is not prominent at the current stage, but it is the future work of our Smart Classroom. So this challenge is posed here.)

4. SOLUTIONS IN SMART CLASSROOM

4.1. Overview of the Smart Classroom

We have built up two Smart Classrooms in our laboratory. The majority of each room is laid out like an ordinary classroom. It has two overhead projectors in addition to a speaker and a wireless microphone. Seven video cameras, which are used by computer vision modules, are mounted at various places in the room.

Separated from the classroom area by a small partition, there are eight PCs at the corner of the room, which perform the room's computation. This section of the room is not interactive, but having it adjacent to the interactive area can simplify wiring, implementation and debugging.

The classroom contains a series of computer controlled devices, such as smart board [17], video cameras, projectors, curtains, a set of wireless microphone equipments, audio multiplexer and etc, and software modules, such as face recognition module, speech recognition module.

4.2. Infrastructure of the Smart Classroom

Like many other similar Smart Spaces, our Smart Classroom is a number of computer controlled devices and software modules, so it is unimaginable to install all these components only in one computer due to the limited computation power and terrible maintenance requirement. Thus, a Smart Platform [15], which is a Multi-Agent System, was developed. Corresponding to other infrastructure in the smart space, such as OAA [18], our infrastructure has some new mechanisms, such as spontaneously directory server, combination of peer-to-peer and delegating communication, automatic dependency management and auto-loading codes, load balance and move management.

By collaborating and coordinating different modules, the infrastructure maintains multimodal interaction. After a sentence is analyzed into a command by speech recognition module, the Smart Classroom will be triggered to display the result of the command by propagating messages among agents. And it is similar that user's gesture or laser pointer's movement, detected by computer vision recognition module, also causes Smart Classroom's responses.

Context sensitive interaction is an important attribute of our Smart Classroom. Smart Platform integrates the room's myriad subsystems and from them produces a coherent whole. So any modules can get the state information of every controlled device or software modules, and Smart Classroom provides the necessary context information as well.

Also, other interactions' features are related to the infrastructure by transferring messages between corresponding agents.

4.3. Vision and Speech Subsystems

In our Smart Classroom, there are several Computer Vision subsystems and two Speech modules. The Computer Vision subsystems include Face Recognition subsystem, Laser Pointer Tracking subsystem, and Gesture Recognition subsystem.

Speech subsystems include Speaker Recognition subsystem and Speech Recognition subsystem. Through these subsystems on the Smart Platform, as the Figure 2 shows, not only nearly every device but also the natural speech and action can become interfaces to our Smart Classroom.

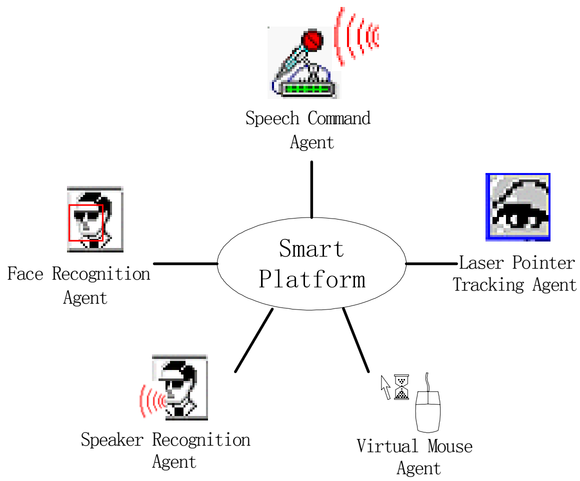

Figure 2. Vision and Speech Subsystems on Smart Platform in the Smart Classroom

As Mark Weiser had explained in ubiquitous computing, to rescue people's energy from irrelevant interaction with computer to the intentioned goal, allowing user's interaction with computer as naturally as possible is vital. In our Smart Classroom, many interactions are natural user interfaces through these Computer Vision subsystems and Speech subsystems. The first one, Gesture Recognition subsystem, is called "Virtual Mouse". In this system, several cameras installed on the room together with a skin color consistency based algorithm are used to recognize the 3D movement parameters of lecturer's hand, as well as some simple actions of the lecturer's palm such as open, close and push [16]. By this system, the movement of Prof. Shi's hand is explained as dragging of the pointer on the screen and the actions of her palm are explained as clicks on the mouse button. In this way, Prof. Shi can easily make annotations on the blackboard or select a remote student on the student-board with her hand, just like in a real classroom.

The second computer vision subsystem, Laser Pointer Tracking subsystem, is used to track the laser cursor on the blackboard based on frame-difference algorithm. And the cursor movement of laser pointer is delivered to remote students so that they can feel they are in the local classroom. By this system, when Prof. Shi emphasizes her lecturing point far away from blackboard, the Smart Classroom can capture the events and record them.

The third computer vision subsystem, Face Recognition subsystem, and the Speaker Recognition System are used to automatically identify the lecturer. The interaction of authorization with Smart Classroom is performed by these two systems instead of inputting ID and password from the keyboard. Facilitated adoption of the authorization process is taken consideration of when we designed and implemented it. The authorization process is completed as long as the lecturer looks into the mirror and speaks out who he is

The Speech Recognition subsystem is a flexible system, which takes consideration of not only the syntax of legal phrases but also the semantic effect of the phrases. The semantic effect refers to an Inter-Agent-Language message that can trigger wanted actions of other agents in the system. Using this grammar, other agents could add vocabularies into the speech recognition agent on the runtime. Nevertheless, the vocabularies could be dynamically enabled and disabled by other agents according to their knowledge of current context in order to keep the effective vocabulary at a minimum size, which is very important for improving the recognition speed and accuracy rate.

These modules exist in the Smart Classroom as agents of Smart Platform. By these software modules together with the hardware devices, such as smart board [17], the interaction in the Smart Classroom becomes richer and more effective. First, the interaction is multimodal, such as touching on the Smart Board, speech command, movement of lecturer's hand, gesture of lecturer's palm, and laser pointer's movement. Second, physically embodied interaction comes into true. When Prof. Shi takes and shows a relia, the physical object becomes an additional UI device although it is used in its normal manner of operation.

4.4. Maintenance of Continuous Task

As described in the previous section, a lecturer can begin his lecture from where he left previously every time, but it is not necessary to care for where his lecture is stored and to memorize the previous state. This function is implemented by States Maintainer, an agent in the Smart Platform.

After this agent is notified of the lecturer's biometrics features, it will automatically find the storing location of his lecture (may be accessed by Internet or may be in the portable note carried by the lecturer.) At the same time, the agent also opens the lecture at where the user left previously by retrieving the states stored last time.

4.5. Coordination of Interaction's Focal Point

In the Smart Classroom, the interaction has lack of single focal point. As the previous scenario shows, the remote students can get the same live instructing video and audio as the local audience get. This function is maintained by Smart Camera module, an agent of Smart Platform. This agent deduces the current focal point of interactions from the context messages of hardware devices and other software modules. When Prof. Shi makes her annotation on the media board by electric chalk, Smart Camera gets the touching event. In consequence, the Smart Camera deduces the current focal point of interaction is on the media board, so media board's video is delivered to the remote audience. However, when Prof. Shi takes the realia, the realia become an additional UI input device by the Computer Vision System and the Smart Camera get the context message from Smart Platform. As a result, the Smart Camera deduces the active interaction point is the realia, so the realia's video is given to remote audience. Similarly, when the lecturer is discussing with the local audience, the Smart Camera also can get context messages, such as lecturer's position, orientation, pose and etc from other modules on the Smart Platform. Consequently, the Smart Camera deduces the whole scene is the local point of interaction and releases the video of the whole scene to remote audience.

5. CONCLUSION AND FUTURE WORKS

We believe the Smart Spaces will be the right metaphor for people to interact with the computer systems in the ubiquitous computing age. After giving the challenges of interaction in the Smart Spaces, this paper describes how to solve these challenges by our infrastructure (Smart Platform), several computer visions and speech modules, States Maintainer, Smart Camera and etc. The solution is implemented in our Smart Classroom project, a test-bed for researchers on the Smart Spaces as well as illustration of its application, which takes us last three year. Currently, we can launch and run our Smart Classroom, just as the demo videos on http://media.cs.tsinghua.edu.cn/~pervasive shown.

In current stage, there is only one user (lecturer) naturally interacting with the Smart Classroom, other attendants are just observers or listeners and not able to exploit the fascinating features of natural interaction. Because a class is bound to have multiple participants, to make a qualified Smart Classroom, we need to enhance the classroom's support for multi-user interaction. In our next step, we will empower the classroom with capability to track and identify more than one user dynamically, and enable Smart Classroom's in-place service to every user in the room.

With the development of computing and network technology, the portable computing device will be proliferated. The lecturer may take a portable computer or PDA into the Smart Classroom, it is interesting how to configure interface depending on the set of devices available. How to seamlessly add computing device into the Smart Classroom is another future work.

REFERENCES

1. Weiser Mark. "The Computer for the Twenty-First Century", in Scientific American, September 1991, pp. 94-100.
2. Michael H. Coen, "The future of human-computer interaction, or how I learned to stop worrying and love my intelligent room", in IEEE Intelligent System, March/April 1999, pp 8-10.
3. http://www.itl.nist.gov/iaui/Projects/smart_space.html.
4. GREGORY D. ABOWD and ELIZABETH D. MYNATT, "Charting Past, Present, and Future Research in Ubiquitous Computing," in ACM Transactions on CHI, Vol. 7, No. 1, March 2000, pp. 29–58.
5. Cohen, M. "Design principles for intelligent environments", in Proceedings of the AAAI Symposium on Intelligent Environments, 1998, pp. 36-43.
6. Jiang Changhao, Shi Yuanchun, Xu Guangyou, Xie Weikai "Classroom in the Era of Ubiquitous Computing - Smart Classroom," in ICWLHN2001, Dec 5-7, 2001, Singapore, pp. 14-26.

7. Weikai Xie, Yuanchun Shi, Guanyou Xu "Smart Classroom – an Intelligent Environment for distant education" in PCM2001, Springer LNCS2195, pp. 662-668.
8. Anind K. Dey, Gregory D. Abowd and Daniel Salber, "A Conceptual Framework and Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," in HCI Journal, 2001; (or be available in http://www.cc.gatech.edu/fce/ctk/pubs/HCIJ16.pdf)
9. Shafer, S.A.N., Brumitt, B., and Cadiz, J.J. "Interaction Issues in Context-Aware Interactive Environments", to appear in HCI Journal, or be available in http://www.research.microsoft.com/easyliving/Documents/2001 03 Shafer.doc
10. Interactive Workspaces in Stanford, http://graphics.stanford.edu/projects/iwork/
11. Smart Rooms in MIT, http://vismod.www.media.mit.edu/vismod/demos/smartroom/
12. Easyliving in Microsoft Research, http://www.research.microsoft.com/easyliving/
13. Dream Spaces in IBM Research, http://www.research.ibm.com/natural/dreamspace/
14. Aware Home in Georgia Tech, http://www.cc.gatech.edu/fce/ahri/
15. Weikai Xie, Yuanchun Shi, Guangyou Xu, Yanhua Mao, "Smart Platform – A Software Infrastructure for Smart Space (SISS) ", in submission to UbiComp 2002
16. Ren, H.B., Zhu, Y.X., Xu, G.Y., and et al. "Spatial-temporal appearance modeling and recognition of continuous dynamic hand gestures", in Chinese Journal of Computer (In Chinese), Vol 23, No.8, August 2000, Pages 824-828
17. Smart Board, http://www.ask.no
18. OAA, http://www.ai.sri.com/~oaa