

# Smart Classroom – an Intelligent Environment for distant education

Weikai Xie, Yuanchun Shi, Guanyou Xu  
Institute of Human-Computer Interaction and Media Integration  
Department of Computer Science and Technology  
Tsinghua University, P. R. China

## Abstract

In this paper, we present our current work on Intelligent Environment - the Smart Classroom project. It is an augmented classroom used for teachers in the distant education. The teacher can write directly on a wall-size media-board or use speech and gesture to conduct the class discussion involving of the distant students. It is just the experience that the teacher was familiar with in the ordinary classroom educations. This paper explained the motivation of this project and discussed the main challenges of the project, which we considered also have significances to other Intelligent Environment projects. Our current progress and the approaches we used are also described.

**Keywords:** intelligent environment, interactive space, distant education, multimodal, context-awareness

## 1 Introduction

We are steadily moving into a new age of information technology named as ubiquitous computing or pervasive computing, in which computation power and network connection will be embedded and available in the environments, on our bodies, and in the numerous handhold information appliances [1]. The human computer interaction paradigm we currently used on the desktop computers will not be sufficient [2]. Instead of we operating on individual computers and dispatching many trivial commands to separated applications through keyboard and mouse, we should be able to interact with all involved computation devices as a whole, and express our intended tasks in a high abstraction level and by ways as natural as we used to communicate with other peoples in everyday life.

The research titled as Intelligent Environment is just motivated by this vision. General speaking, an Intelligent Environment is an augmented living or working environment which could actively watch and listen to the occupants, recognize their requirements and attentively provide services for them. The occupants could use normal human-being interaction methods such as gesture and voice to interact with the computer system operating the environment.

The researches in this filed are bring into mainstream in the late 1990's by several first-class research groups of the world such as AI Lab and Media Lab in MIT, Xerox PARC, IBM and Microsoft. Currently there are dozens of Intelligent Environment related projects, carried out in research groups from all over the world. Just to mention some most famous ones among them, they are Intelligent Room project from MIT AI Lab [3][4], Aware Home project from GIT [5] and Easy Living project from Microsoft [6][7].

The Smart Classroom project in our group is also a research effort on Intelligent Environment. It demonstrates an intelligent classroom used for teachers involved in distant educations, in which the teachers could use the same ways as they used when teaching in a real classroom to teach the

distant students. The real motivation under the demonstration is we want to identify the key challenges in the research of Intelligent Environment, and at the same time to explore how the Intelligent Environment will influence the way people cooperate with the computers. One reason that we selected the distant education as the test-bed for Intelligent Environment research was another group in our lab had already done years of work on it and had accumulated valuable technologies and experiences. Our project bears some similarities to the Classroom2000 project from GIT [8]. But there are major differences as well. Their project is focused on the automatically recording of the events occurred in the environments and is for traditional classroom education only. However, our project is focused on providing a high abstraction level and multi-modal human-computer interaction possibilities for the teacher in the environment, and the system is used for distant education.

In the following sections, we first introduce the scenario of the Smart Classroom, then comes to the discussion of some key issues in the implementation the Smart Classroom, which we think have significances to other Intelligent Environment projects too, and next we describe our current progress and future work. Finally we end the paper with a conclusion.

## **2 The Scenario**

Almost all the distant education systems developed now are so-called desktop-computing based, where the teacher is required to seat down in front of the desktop computer and use the keyboard or mouse to conduct the distant education course. The teacher's experience is of much difference from teaching in an ordinary classroom, where the teacher could use handwriting to make annotations on the blackboard, use speech and gesture to involve the students in the class discussion and use many natural interaction patterns like that. The different experience always makes the teacher feel uncomfortable and reduced the efficiency of the course as well.

We considered the adopting of Intelligent Environment idea will perfectly resolve this problem by construct a augmented classroom where the teacher could use all those natural interaction patterns they are accustomed to teach the remote students.

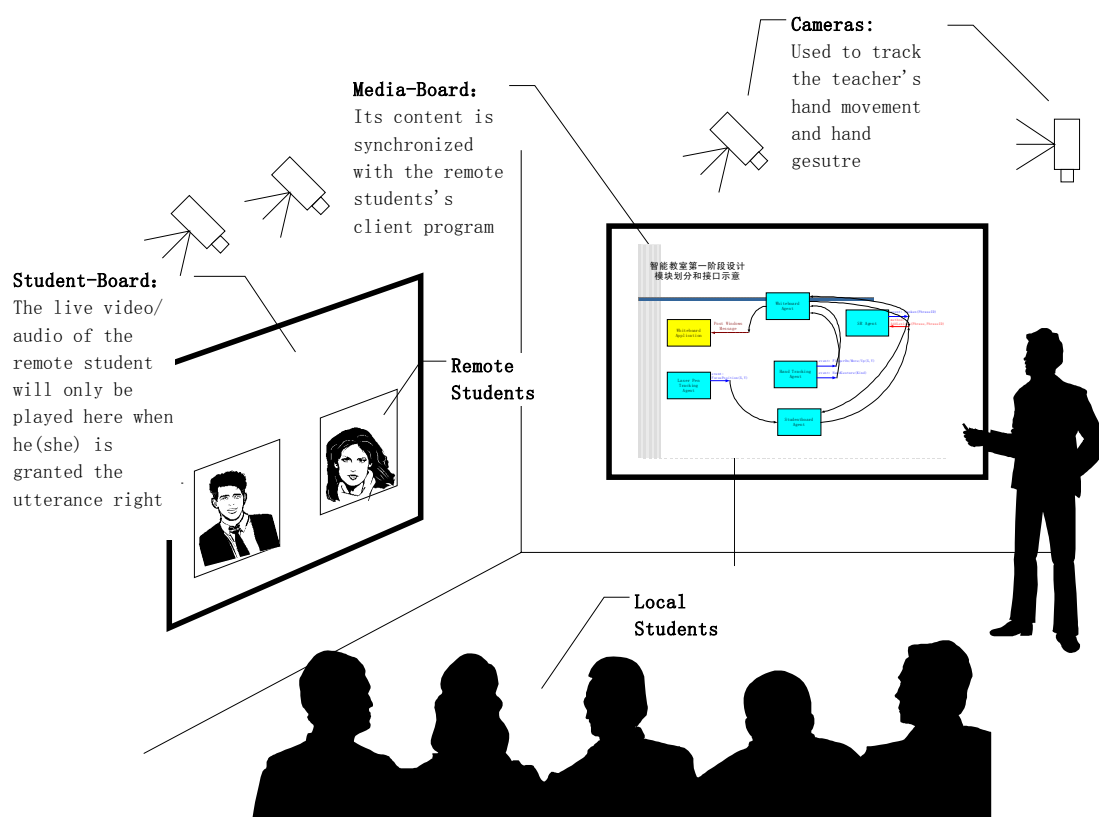
We deployed an experiment system in a room of our lab. We equipped two wall-sized projector systems in it. One is used to display the courseware prepared by teacher – an analog of the blackboard in an ordinary classroom (we call it media-board) and the other is used to display the image of the remote students (we call it student-board). Some cameras and wireless microphone systems are also installed in the proper positions to capture the teacher's video and voice.

The display of the remote students' client software is synchronized with that of the media-board in the classroom, i.e., whatever changes the teacher makes on the media-board will be reflected on the remote student's client software. The captured audio and video in the room are multi-casted to the remote students' sites too. If the teacher wants a remote student to give utterance on the class, he (she) can enable this student's up-link audio/video channels so that the student's audio and video will be sent both to the room and other remote students' sites.

The scenario seems no big difference with many other whiteboard based distant learning systems. However, the magic of the Smart Classroom is the way the teacher using the system – the teacher no longer need to tied up to the desktop computer using the cumbersome keyboard and mouse to interact with the distant education system. Making annotations on the courseware displayed on the media-board is just as easy as writing on an ordinary classroom blackboard, i.e.

the teacher only need to move his finger, then the stroke will be displayed on the media-board and overlapped with the courseware displayed. The teacher could also point to something displayed the media-board and say some words to complete the action he wanted, such as erasing the annotations previously made, scrolling the page up and skipping to another page. Nevertheless, the way to switch the remote students' utterance right is intuitive and easy. Each attended remote student will have an image displayed on the student-board to represent him. Whenever a student requires the utterance right, his corresponding image will start to blink to alert the teacher. In order to grant the utterance right to a selected student, the teacher only need to use a laser-pen to point at the image representing the student.

All these interaction patterns are of much similarity to what happed on an ordinary classroom. In fact, we deliberately hide all the computers running the room's software out of the sight, just to give the teacher a feeling that they are not using some computer programs but teaching in a real classroom.



Actually, this system blurs the differences of the ordinary classroom education and the distant education. The teacher can give a course to the local students on the Smart Classroom and those remotely attended students at the same time.

Another appealing feature of the Smart Classroom is the ability to automatically capture the threads of the class. Any words the teacher said, any annotations the teacher made on the materials, and any discussions took place on the class, plus the original courseware will be recorded in a hypertext documents with time stamps, which could be used to replay the class back for review or to generate a new courseware with a post-edit toolkit.

### 3 Key Research Issues in the Smart Classroom Project

In the process of designing the Smart Classroom, we identified several key research issues of

it, and we think that they will be the main play-fields in the research of Intelligent Environment too.

### **3.1 The right modal of the software structure**

The Smart Classrooms, just like many other similar Intelligent Environment setups, will assemble a good number of hardware and software modules such as projectors, cameras, sensors, face recognition module, speech recognition module and eye-gaze recognition module. It is unimaginable to install all these components in one computer due to the limited computation power and terrible maintenance requirements. Thus, a distributed computing structure is required to implement an Intelligent Environment.

There are several commercial available distributed computing structures such as DCOM, CORBA and EJB. They are all based on the distributed component modal. Under this modal, software should have a central logic and several other peripheral objects offer service to the central logic. The objects run only when invoked by the central logic. We call this software structure as monolithic. We considered this modal is not sufficient as a computing structure for the Intelligent Environment. The reason is

1. The scenarios of an Intelligent Environment are usually very complex. Developing a central logic for it is very difficult even an unpractical matter.
2. The scenarios and configurations in an Intelligent Environment are often very dynamic. New functions will be added, old module will be revised, and all those things are happened continuously. The monolithic approach is very inflexible under this situation, because any trivial modifications will require the whole system to shut down and all modules in the system should be re-linked.
3. The central logic is likely to be the bottleneck of the system.

To summary up, the distributed components modal is a tightly coupled system, which could not accommodate the dynamic and complexity of the Intelligent Environment system. Instead, we need a modal in which the modules are more loosely coupled. Fortunately, we already have one, the multi-agent system (MAS) modal, which have been invented and used for years in the AI domain. According to this modal, a system is constituted by many individual software modules called agents. An agent has its own executing process and is autonomous by itself. Each agent has limited capabilities and limited knowledge of the functions of the whole system, but through communications and cooperations the agents' community will expose a high degree of intelligence and can achieve very complex functions. We considered it a right solution as the software structure for the Intelligent Environment, i.e., the software of an Intelligent Environment should be modeled as a multi-agent system.

### **3.2 Other necessary features of the software architecture**

1. A mechanism to reference modules (agents) by some high-level description

In order to an agent to communicate with each other, it should have a reference to the other peer. In order to make the system more loosely coupled and flexible, the reference binding should be created by some high-level mechanism. A usual implementation is binding by capability. That is to say, on startup, each agent should register its capabilities to some central registry, and when an agent needs some service, it could ask for it by describe the needed capabilities. However, the true challenge here is how to set up a framework for the description

of the capabilities, which could enable a new agents to find out the exact semantic meaning of the capabilities advertised by other agents. For example, if an agent developed by some other group is introduced into an existing agent society and it needs a TTS service, which words should it uses to express its requirement? Is it “say”, “speak” or just “TTS”? Maybe the final solution of this problem only lies in the achievement of the NLP technology.

#### 2. A mechanism to encapsulate legacy programs (, which should be easy to grasp)

When implementing an Intelligent Environment project, we usually want to exploit the capabilities of many legacy programs. After all, to build all the things bottom up is neither necessary nor practical. Thus the software architecture should provide a simple and clear migration path to bring those legacy programs in, making them a legal resident of the agent society.

#### 3. A mechanism to ensure consistent response time

Many perception modules in an Intelligent Environment setup, especially vision algorithm modules, are so-called resource-hungered. They can easily exhaust all the capacity of a latest high-end computer and their response time tend to vary during times. The situation is even worse when multiple modules coexist in one computer. This will be rather bothering to the user. For example, suppose the user points to one object in the environment and asks some information for it. Due to the overload of the hand-tracking module, the system can't give the response until a later time. However, the user might think the system can't recognize his behavior and start to do something else. And when the delayed response eventually comes, its only result is making the user very confused. Therefore, a mechanism to guarantee a consistent response time is very necessary to make an Intelligent System useful in practice.

#### 4. Feasible facilities for debugging

As mentioned above, the software in the Intelligent Environment is composed of many software modules (agents) running in their own processes. Misbehaviors of one agent are usually related to the states of many other agents. This makes debugging of the system very difficult because you should bring all related agents into their exact states as the error occurring, in order to replay the error. According to our knowledge, currently there is no debug tool available that can address this problem. Therefore, the software architecture should build some debug facilities in it, such as centralized logging and centralized agents management interface.

### **3.3 Multi-modal processing capability**

People use multiple modalities to communicate with each other in everyday life, such as speaking, pointing and gestures. The multi-modal interaction capability is a fundamental requirement of the Intelligent Environment, because a single modality is often semantic incomplete. For example, when one say “move it to right”, without the recognition of the hand pointing modality we could not tell which object is referred by the speaker. Another benefit of the multi-modal processing is the information from other modalities often helps to improve the recognition accuracy of a single modality. For example, suppose in a noise environment, a not so smart SR algorithm maybe has difficulty to decide what the user said is “move to right” or “move to left”. But after referring the information from the hand gesture recognition module, the system

could eventually make the right choice.

This also happens to Smart Classroom. In the scenario we designed, the teacher can use speech and hand gesture to make annotations on the media-board, to manage the content in the media-board and to control the utterance right of the remote students.

Currently, some advances have been made in the multi-modality processing research. The most famous approach is the one used in the Quickset project [9]. It essentially takes the multi-modality integration process as a kind of language parsing process, i.e., each separate action in a single modality is considered as a phrase structure in the multi-modal language grammar, they are grouped and induced to generate a new higher level phrase structure. The process is repeated until a semantic completed sentence is found. The process could also help to correct the wrong recognition result of one modality. If one phrase structure could not be grouped with any other phrase structure according to the grammar, it will be regarded as a wrong recognition result and will be abandoned. Although the method is focused on the pen-based computer, it could be used into the Intelligent Environment research as well after some modifications.

### **3.4 Take context into account**

In order to determine the exact intention of the user in the Intelligent Environment, the context information should also be taken into account. For example, in the Smart Classroom, when the teacher says “turn to the next page”, does he mean the courseware displayed on the media-board should be switched to the next page or just tells the students to turn their textbooks to the next page? Here the multi-modal integration mechanism could not resolve the ambiguity because the teacher just says the command without any additional gestures. However, we could easily tell the teacher’s exact intention by recognizing where he stands and what he faces.

We consider that the context in an Intelligent Environment generally include the following factors:

1. Who is there?

A natural and sophisticated way to acquire this information is through the person’s biology character, such as face, acoustic, foot print and so on.

2. Where is the person located?

This information could be acquired by vision tracking. In order to the result from the vision tracking module could be interpreted by other modules in the system, a geometric modeling of the environment is also needed.

3. What are ongoing in the environment?

This information including the action the user explicitly takes and others implied in the user’s behaviors. The former one could be acquired by multi-modal recognition and processing. However, there is no formal approach available dealing with the latter problem indeed using our current AI technologies except some ad-hoc methods.

## **4 Our approach and current progress**

We have currently completed a first-stage demo of the Smart Classroom. The system is composed of the following key components.

1. The multi-agent software platform.

We adopted a public available multi-agent system, OAA (Open Agent Architecture), as the software platform for the Smart Classroom. It was developed by SRI and has been used by

many research groups [10]. We fixed some errors of the implementation provided by SRI to make it more robust. All the software modules in the Smart Classroom are implemented as the agents in the OAA, and using the capability provided by it to communicate and cooperate with each other.

2. The distant education supporting system

It is based on the distant education software called Same View from another group in our lab. The system is constituted of three layers. The most upper layer is a multimedia whiteboard application we called media-board and associated utterance right control mechanism. As mentioned above, one interesting feature of the media-board is it can record all the actions on it, which could be used to aid the creation of the courseware. The middle layer is a self-adapting content transform layer, which could automatically re-authorizing or trans-coding the content sent to the user according to the user's bandwidth and device capability [11]. The media-board use this feature to ensure that students at different sites could get the contents on the media-board at a maximum quality according to their different network and hardware conditions. The lowest layer is a reliable multicast transport layer called Totally Ordered Reliable Multicast (TORM), which could be used in a WAN environment where sub-networks capable of or incapable of multicast coexist [12][13]. The media-board uses this layer to improve its scalability across large networks like Internet.

3. The hand-tracking agent.

The agent could track the 3D movement parameters of the teacher's hand using a skin color consistency based algorithm. It could also recognize some simple actions of the teacher's palm such as open, close and push [14]. The same recognition engine had been successfully used in a project in Intel China Research Center (ICRC), which we have taken part in.

4. The multi-modal unification agent.

The agent is based on the work in the project of ICRC mentioned above, under a collaboration agreement. The approach is essentially based on the one used in Quickset as mentioned above.

5. The speech recognition module.

The speech recognition agent is developed with a simplified Chinese version of ViaVoice SDK from IBM. We carefully designed the interface to make any agents who need the SR capability could dynamically add or delete the recognizable phrases together with the associated action (an OAA resolve request indeed) when recognized. Thus the vocabulary in the SR agent is always kept to a minimum size according to the context of the time. It is very important to improve the recognition rate and accuracy.

In the next stage, we planed to add other computer vision capabilities needed in the scenario, such as face tracking, pointing tracking. Another important thing under consideration is to develop a framework for the agents in the system to exchange their knowledge in order to make a better use of the context information. We also want to introduce the acoustic-recognition and face-recognition program developed by other groups in our lab into the scenario. These technologies could be used to identify the teacher automatically and then provide personalized services to him (her). An example is the classroom could automatically resume the context (such as the content on the media-board) where the teacher stopped at the last class.

## 5 Conclusion

We believe the Intelligent Environment will be the right metaphor for people to interact with the computer systems in the ubiquitous computing age. The Smart Classroom project is started as a test-bed for the research issues of the Intelligent Environment and also as an illustration of its application. We are just achieved the first stage of the work and will continue our effort in future work.

## References

1. Mark Weiser. The Computer for the 21 st Century. Scientific American. pp. 94-10, September 1991.
2. Mark Weiser. The world is not a desktop. Interactions, pages 7--8, January 1994.
3. Coen, M. Design Principles for Intelligent Environments. In Proceedings of The Fifteenth National Conference on Artificial Intelligence. (AAAI98). Madison , Wisconsin. 1998.
4. Coen, M. The Future Of Human-Computer Interaction or How I learned to stop worrying and love My Intelligent Room. IEEE Intelligent Systems. March/April. 1999.
5. Kidd, Cory D., Robert J. Orr, Gregory D. Abowd, et al. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In the Proceedings of the Second International Workshop on Cooperative Buildings - [CoBuild'99](#), Position paper, October 1999.
6. Steven Shafer, et al. The New EasyLiving Project at Microsoft Research. Proceedings of the 1998 DARPA/NIST Smart Spaces Workshop, July 1998, pp.127-130.
7. Brumitt, B. L., Meyers, B., Krumm, J., et al. EasyLiving: Technologies for Intelligent Environments. Handheld and Ubiquitous Computing, 2nd Intl. Symposium, September 2000, pp. 12-27.
8. G. D. Abowd. Classroom 2000: An experiment with the instrumentation of a living educational environment. IBM Sysms Journal, Vol. 38. No.4
9. Michael Johnston .Unification-based multimodal parsing. In the Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 98), August 98, ACL Press, 624-630.
10. <http://www.ai.sri.com/~oaa/>
11. Liao Chunyuan, Shi Yuanchun, Xu Guangyou. AMTM - An Adaptive Multimedia Transport Model. In proceeding of SPIE International Symposia on Voice, Video and Data Communication, Boston, Novermeber 5-8, 2000.
12. Pei Yunzhang, Liu Yan, Shi Yuanchun, Xu Guangyou. Totally Ordered Reliable Multicast for Whiteboard Application. In proceedings of the 4<sup>th</sup> International Workshop on CSCW in Design, Paris, France, 1999.
13. Tan Kun, Shi Yuanchun, Xu Guangyou. A practical semantic reliable multicast architecture. In proceedings of the third international conference on multimodal interfaces, BeiJing, China, 2000.
14. Haibing Ren, Yuanxin Zhu, Guangyou Xu, Xueyin Lin, Xiaoping Zhang. Spatio-temporal appearance modeling and recognition of continuous dynamic hand gestures. Chinese Journal of Computers (in Chinese), 1999. Vol 23, No. 8 , Agu. 2000,

pp. 824-828.