INFORMATION SCIENCE PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITJ2449

Project Smart Remote Classroom - Providing Novel Real-Time Interactive Distance Learning Technologies*

Yuanchun Shi, Weikai Xie, Guangyou Xu, Peifeng Xiang, and Baopeng Zhang
Tsinghua University, China

ABSTRACT

Real-time interactive virtual classrooms play an important role in distance learning. However currently available systems are insufficient in supporting large-scale user access, and they cannot efficiently support accessing with heterogeneous devices and networks. Furthermore, these systems are usually desktop-based, until the result that the teacher's experience is completely different from teaching in a real physical classroom. This paper discusses the Smart Remote Classroom project that deals with these difficulties using the following novel technologies: 1) A hybrid transport layer multicast protocol called TORM and an adaptive content delivery scheme called AMTM, which work together to enable large-scale users to access a virtual classroom with different devices and networks synchronously. 2) A dedicated software called SameView, which takes use of the proposed TORM and AMTM technology, and provides a rich set of functions for teachers and students to efficiently carry out the real-time interactive tele-education. 3) The Classroom augmented by Smart Space technology called Smart Classroom where the user interfaces of the SameView are incorporated in the classroom space. Thus the teacher can instruct the remote students just like teaching face to face in a conventional classroom. All these technologies have been successfully integrated and demonstrated in the prototype system at Tsinghua University.

Keywords: distance learning, communication technology, multimedia adaptive delivery, smart classroom, human-computer interaction

INTRODUCTION

Motivation

In recent years , distance learning has increasingly become one of the most important applications on the internet and is being discussed and studied by various universities, institutes and companies. The Web/Internet provides relatively easy ways to publish hyper-linked multimedia content for more audiences. Yet, we find that most of the courseware are simply shifted from textbook to HTML files. However, in most cases the teacher's live instruction is very important for catching the attention and interest of the students. That's why Real-Time Interactive Virtual Classroom

This chapter appears in the journal, *International Journal of Distance Education Technology*, edited by Qing Li and Weijia Jia.

(RTIVC) always plays an indispensable role in distance learning, where teachers and students located in different places can take part in the class synchronously through certain multimedia communication systems and obtain real-time and media-rich interactions. However, to provide this type of distance learning in large scale, there still remain some barriers:

1) ***Lack of adequate technologies to cope with large-scale access.*** Most tele-education schools simply adopted commercial videoconference products (usually they are H.32X-based systems) as the operating platform for RTIVC, where all clients should connect to a centered MCU (Multi-Point Controlling Unit) and data initiated from one client is replicated (sometime maybe mixed) and forwarded to all other clients by MCU. However these systems are not scalable, since the maximum user number (usually 10 or more) is rigidly limited by the capacity of the MCU. Moreover, the cost of communication service is very expensive or there is no guarantee for the quality of service. Thereby nowadays most tele-education schools can only operate RTIVC classes including a small number of students. A possible approach to address this scalability issue is leveraged by the IP Multicast technology, where no central data-replicating node like MCU is required. However, the current state of the IP Multicast technology is not full-fledged yet. First, the IP Multicast service provided by the network layer is a best-effort service. This is not tolerable for applications like RTIVC that are sensitive to the loss of messages. For example, the dropping of a single packet at one client will make the state of the whiteboard in a RTIVC system at this client lose consistency with others. Secondly, IP Multicast is not fully supported by many currently deployed routers of the Internet. Today's Internet can be viewed as many Multicast islands that fully support IP Multicast being separated by the Unicast zones which are not capable of IP Multicast. As a result, we can see many applications which directly rely on the network layer. IP Multicast cannot be successfully deployed on current Internet infrastructure.
2) ***Lack of adequate technologies to accommodate students with different network connecting and terminal device conditions in one session.*** Most current RTIVC systems have rigid requirements for the network and device settings of the clients. Clients with inferior device capability or network bandwidth could not join the session and get smooth service quality. On the other hand, clients with superior conditions could not fully take advantage of their extra capabilities. Since handheld devices such as Pocket PC and Smart Phone, and wireless networks such as GPRS and WIFI are becoming more and more popular, it is a natural demand to allow people to take lifelong education with these types of devices and networks, despite the fact that these devices and networks will inevitably possesses diverse capabilities.
3) ***Desktop-based teaching metaphors are not natural enough for many teachers.*** Most current RTIVC systems are desktop-based, i.e., the teacher should remain stationary in front of a desktop computer and use the keyboard or mouse to operate the class or interact with others. This metaphor is particularly unacceptable for the teachers, because their experience here is much different from that in a real classroom. In a real classroom, they can move

freely, talk to the students with hand-gesture and eye contact, and illustrate their ideas conveniently by scribbling on a blackboard. Many teachers involving in tele-education, when talking with us, complained that this divergence of the experience makes them uncomfortable and reduced the efficiency of the teaching and learning activity.

Smart Remote Classroom Project

The Smart Remote Classroom project at our group is a long-term project aimed at providing adequate technologies to overcome the above-mentioned difficulties in the current practice of RTIVC and building an integrated system as an exemplar for the next generation real-time interactive distance learning in China. Currently we have made progresses in the following aspects.

A multimedia communication supporting platform for large-scale real-time interactive distance learning is developed. The work can be divided into two parts: one is a Totally Ordered Reliable Multicast (TORM) protocol (Pei, 1999; Tan, 2000) - a transport-layer multicast protocol locating on top of the conventional network-layer IP-Multicast and Unicast. Thereby it can accommodate large-scale users to access a RTIVC through current Internet infrastructure. The other part is an Adaptive Multimedia Transport Model (AMTM) (Liao, 2000), which can transcode or transform the transferred documents according to the device and network capabilities of each individual client of a RTIVC session, while the consistency of the content remains at the semantic level.

A dedicated software for RTIVC called SameView (the name comes from the notion that no matter if you are a local student or a remote student and even no matter which kind of network connection you have, you can always get the same view of the content of class) is developed, which runs on the developed TORM/AMTM platform and provides a rich set of interaction channels, including the HTML-capable whiteboard, video/audio communication and chatting for the teacher and remote students participating in a RTIVC class (Pei, 2000). SameView also features in its class-experience recording capability. More specifically, all the activities and events occurred during a class, e.g., the slides teacher presented on the whiteboard, the annotations teachers made as well as the live video and audio, will be captured and recorded in a structured multimedia document. After some simple post-editing when it’s needed, this recorded document can be used as a piece of courseware that can be retrieved and played back after the class.

A real instrumented classroom provided a natural user-interface of the SameView at the teacher side, where teachers can move freely and use the conventional teaching metaphors that they are familiar with to give classes to remote students, instead of the cumbersome desktop-based metaphors used in most current RTIVC systems. This part of technology is called Smart Classroom (Xie, 2001), and since the teacher is in a real classroom, we can accommodate local students presenting in the classroom at the same time. This way we can eliminate the difference of On-campus Learning and Distance Learning activities, because teachers can give classes to local students and remote students simultaneously, which can reduce the required workforce of teachers as compared with separate operation of education on campus and tele-education.

The rest of the paper will be organized as follows: first, we will discuss the multimedia communication supporting platform, the SameView and the Smart Classroom in orderly detail. Then, we will show how these technologies fit into the overall picture of a revolutionary real-time interactive distance learning system. Finally we will give a conclusion.

THE MULTIMEDIA COMMUNICATION SUPPORTING PLATFORM FOR LARGE-SCALE RTIVC

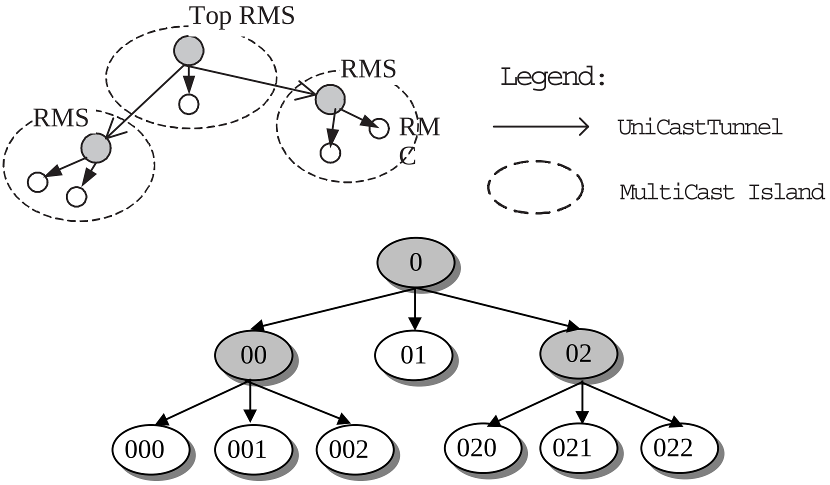
RTIVC is a typical large-scale interactive application, for there may be hundreds or thousands of remote students taking part in a virtual class. So reliable multicast is a useful network service but is also a research issue of challenge for the heterogeneity and the lack of full support of IP multicast in today's Internet infrastructure. Instead of following the traditional end-to-end model for reliable multicast, we developed a Totally Ordered Reliable Multicast (TORM) protocol which uses a hybrid approach that adopts both IP Unicast and IP Multicast to deliver data and the Adaptive Multimedia Transport Model (AMTM) applied to dynamically trans-code the multimedia data to adapt the different device and network capabilities at the users' side.

Totally Ordered Reliable Multicast (TORM)

Large-scale interactive applications have demanding requirements on underlying transport protocols for efficient dissemination of real-time multimedia data over heterogeneous networks (Kuo, 1998). Existing reliable multicast protocols failed to meet these requirements due to following reasons:

1) most protocols presume the existence of multicast fully enabled network infrastructure, which is usually not the case for current Internet;
2) protocols being able to support multiple concurrent data sources only have limited scalability;
3) few of them have implemented an end-

Figure 1: Architecture of TORM and TORM Tree



to-end TCP-friendly congestion control policy.

As Figure 1 shows, we apply a client/server architecture to the design of TORM, in which a Reliable Multicast Servers (RMS) function as a pure controller of a local group. Reliable Multicast Clients (RMC) find the nearest RMS and register their interests for reliable multicast service (Pei, 1999).

Multicast is used as usual inter the multicast island, and Unicast tunnels are created dynamically to connect session members located in separated multicast islands. High scalability is achieved by organizing members of a session into a hierarchical structure. Nodes in a session form a TORM tree where every node has its UID (unique identifier). While disseminating data, the source address is the UID of the source, and the destination address can be unicast, multicast or subcast address. Then a hybrid multicast is implemented as a transport routing which adopts the UID instead of IP address of network layer.

In contrast to most existing tree-based protocols, any concurrent sources in TORM are allowed to exist in a session. In order to support interactive applications that multiple users cooperate based on a shared state, TORM also incorporates two serialization algorithms, i.e., a distributed one and a centralized one, to ensure totally ordered message delivery through all members. These two serialization algorithms are both implemented in the transport layer which can be deployed by applications.

Another eminent feature of TORM is a novel congestion control scheme which is able to respond to congestions in a timely way and fairly share bandwidth with other competing network flows. Here real-time measurement results of reception rate of all receivers, packet losses, and variations of Round-Trip-Time are used as feedback for sender rate adjustments.

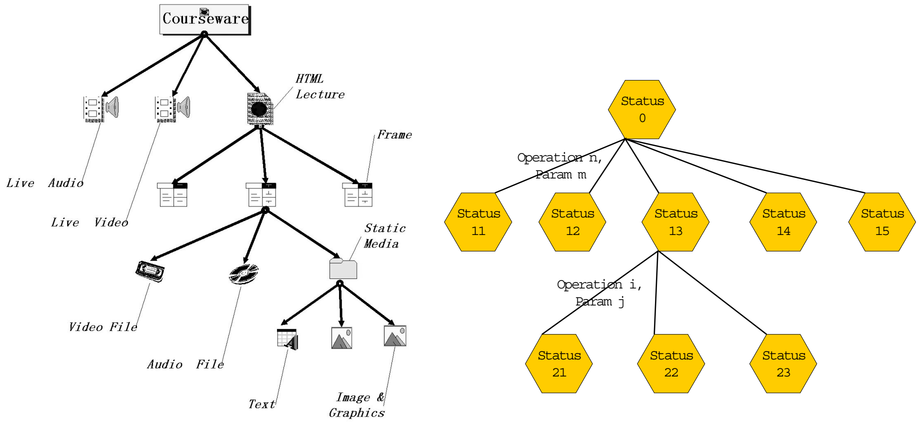
To testify the efficiency of TORM, we compared the loss recovery performance of TORM with SRM/Adaptive (Raman, 1998) through software simulation, more specifically, with NS2 (Network Simulator 2). The recovery latency and duplicate-request number of these two protocols were tested with star, 2-clusters and random topology. The result (Zhuang, 2002) shows that TORM has better performance than SRM as a reliable multicast protocol for large-scale, real-time and interactive applications.

Adaptive Multimedia Transport Model (AMTM)

Adaptive Content Delivery (ACD) (Zhang, 2000) bears the best efficiency among all possible types of Adaptive Multimedia Delivery schemes (Clark, 1999), in which application-layer semantics of the delivered data is coupled with underlying transport mechanisms. In a Smart Remote Classroom project, the live didactical content and recorded courseware are incorporated in a compound multimedia document. We developed AMTM to provide differentiated services for the delivery of this format of data by dynamically transcoding the multimedia data according to the capabilities of the learning environment of each user as mentioned above without data redundancy.

First we proposed the Multimedia Compound Document Semantic Model of AMTM to describe the data organization, PQoS (Perceived QoS) and transformation of compound documents with the notions of Media Object, Content Info Value and Status Space of Transforming respec-

Figure 2: Basic Notions in AMTM



a. The Media Object Tree of a Compound Multimedia Document

b. An Example Transforming Status Tree of a Media Object

tively (Liao,2000). A compound multimedia document is parsed as a structural description of embedded Media Objects as is shown in Figure 2(a).

PQoS is defined as Content Value of the Compound Document:

$$\text{Content - Value}_{\text{whole}} = \sum_i Content - Value_i$$

In the equation above, Content-Value is the Content Value of Media Object i in the document. Equation (1) shows as follows:

$$CV = RCV \times TR \times MTF \times UPF$$

The meanings of equation (1) is as follows.

- CV is Content Value;
- TR is Transport Rate defined as the transport rate of data;
- MTF is Media Type Factor with weight sequence as Text > Image/Graphics > Audio > Binary Data File > Video;
- RCV is Relative Content Value. According to the transcoding or transforming operation on the Media Object, this value can be from 1 to 0;
- UPF is User Preference Factor. But it's still an open problem to measure UPF.

The targets of the adaptive delivery are network, terminal devices and users, i.e., constraints, which fall into two domains, namely Loose Constraints and Hard Constraints. Loose Constraints indicate the least demand for application performance. In contrast to this, Hard Constraints can't be infringed, otherwise the applications will run in error. So loose constraints are considerable. Then the targets of the adaptive delivery can describe that the actual response time should be less than the maximum tolerable response time for users, or the content value of the compound document should be as great as possible. Meanwhile, considering the agility of configuring, easy developing and convenient for extending, the application context is assumed as a HTTP proxy so as to depict conveniently and not lose the generality.

Radically, the former problem is that of "Optimal resource allocation" in that the total time consumed by the compound document can be looked on as some sort of "resource," and it will be optimally distributed to each media object in the document, of course including the document itself, so as to render the most valuable result to the user.

The adaptive process of the AMTM can be described as selecting the optimal resource allocation scheme and associated transformation plan for the embedded Media Objects by searching in the Status Space of Transforming. Our aim is to obtain the largest CV of the document while transforming the Media Object in real-time. To solve the problem, we constructed some formulas to relate the consumed time resource (including the time of media transformation) to the resulting content value, and optimally distributed the resource using Lagrangian method, then searched in the status trees to get the best sequence of trans-coding operations. Figure 2 (b) illustrates a typical transforming status tree, where the original media is labeled with Status0, and converted into the new node Status11 through operation1 with parameter1, while sub-node Status13 is transformed to newer nodes, namely Status21, Status 22, Status 23 and so on. AMTM couples the time/space impact of the media transformation and delivery into the optimal transformation policy to implement the media adaptation. This policy is implemented in the RMS of TORM.

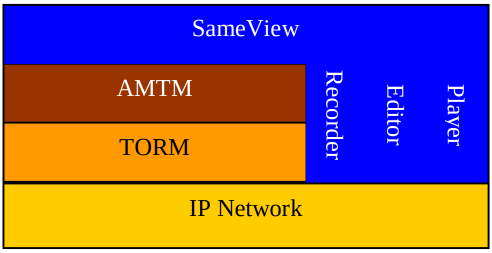
SAMEVIEW – SOFTWARE FOR RTIVC

SameView is a software system developed for real-time interactive distance learning application based on the proposed TORM and AMTM platform. (see Figure 3.)

Interaction Approaches

SameView provides a set of interaction approaches for the teacher and students to efficiently gain their ends of teaching and learning.

Figure 3: Architecture of SameView



1) **Mediaboard,** which is a shared whiteboard, is capable of displaying multimedia contents in HTML format. The teacher can show any materials for the class on the board as long as it is in HTML format or can be embedded in an HTML page. Or it can even open an Internet URL here. Moreover, she can add annotations or scribble on the slides on the fly. All actions the teacher makes on the board, such as jumping between slides, scrolling the slides and writing on the slides, will be synchronously displayed on each student's client. When permitted by the teacher, a student can also dominate the Mediaboard, for example, write down his solution to a problem issued by the teacher.
2) **Audio/Video.** The students can hear the audio and see the video come from the teacher side. In addition, a student can also broadcast his audio and video to the others, when permitted by the teacher, for example, when the teacher asks him to give comments on a topic. The audio/video compressing and decompressing are developed on MS DirectX SDK, so any DirectX compliant codec can be utilized by the system, for instance, DivX or Intel Indeo.
3) **Chat.** In addition, teachers and students can communicate by text messages.

Session Management

The users participating in a class through SameView will play different roles in the class. The possible roles include Chair, Addresser, Audience and Anonymous. Chart 1 shows their respective privileges.

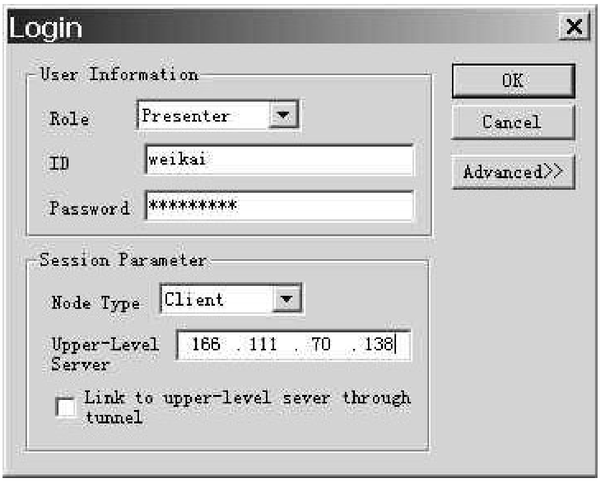
The teacher usually plays the role of Chair, while students are always initially assigned to the role of Audience. As a class going, the teacher can dynamically change the role of a student if necessary, for example, to invite the student to give comments on a topic. An audience can also send a request to the Chair if he/she wants to be an addresser. If more than one participant broadcasts Audio/Video at the same time, only the one who speaks loudest will get the right of speaking.

Remote students use a client program to take part in a RTIVC held with SameView. On running this program, the student will first see a login dialog as in Figure 4. The upper part of this dialog is for getting the information related to student's personal information, such as his/her name and desired role in the session. The lower part of this dialog is for configuring the parameters related to the under-

Chart 1: Privileges of Different Roles

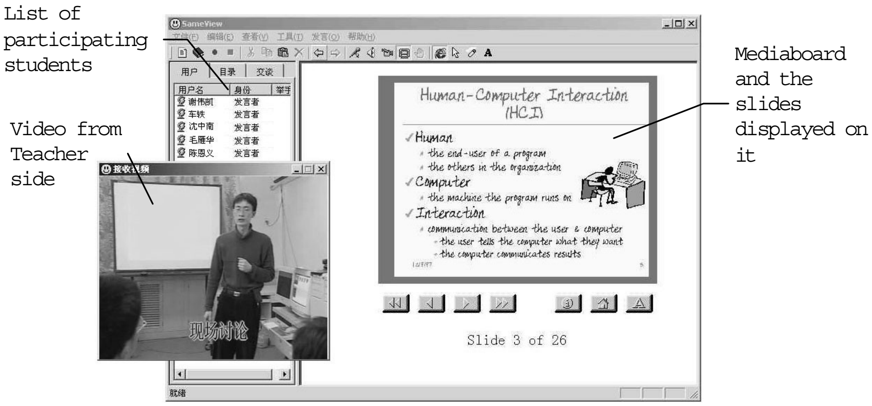
Role	Change other users' role	Action on Media-board	Broad-cast Audio/ Video to others	User number with this role	Listed on Partici-pant List
Chair	Allowed	Allowed	Allowed	1	Yes
Addresser	NA	Allowed	Allowed	>=0	Yes
Audience	NA	NA	NA	>=0	Yes
Anonymous	NA	NA	NA	>=0	No

Figure 4: Login Window of SameView Client Application



Role: Presenter
ID: weikai
Password: *********
Node Type: Client
Upper-Level Server: 166 . 111 . 70 . 138
☐ Link to upper-level sever through tunnel

Figure 5: The Main User-Interface of the SameView Client



用户名	身份	举手
谢伟凯	发言者	
车铁	发言者	
沈中南	发言者	
毛雁华	发言者	
陈恩义	发言者	

Human-Computer Interaction (HCI)

- Human
  - the end-user of a program
  - the others in the organization
- Computer
  - the machine the program runs on
- Interaction
  - communication between the user & computer
    - the user tells the computer what they want
    - the computer communicates results

Slide 3 of 26

lying TORM session, such as whether this program should function as an RMS or an RMC and the address of the direct upper-stream server in the TORM hierarchy. If this node is not in the same IP-Multicast-capable subnet as the upper-stream server, the “link to upper-level server through tunnel” should be checked to establish tunnel with Unicast.

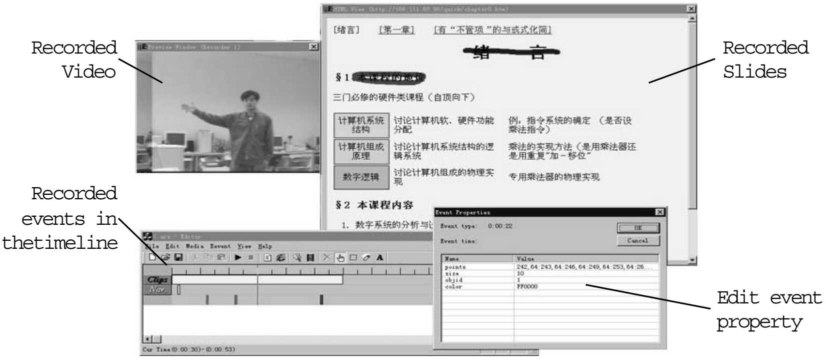
Figure 5 shows the main user-interface of the SameView Client application after login into a session. On the right is the window for the Mediaboard. As shown, a HTML page is displayed on it. The panel on the left lists all the participants of this RTIVC session and their state, such as their role and whether a role-upgrading request is pending for that participant. Another separated floating window is used to show the live video of other participants who are broadcasting their video, according to selection policy described in the previous section.

Class Recording

SameView can capture the exact process of a class by recording everything that happened on all interaction approaches, such as the slides the teacher showed, the annotations made on the slides, as well as the live video and audio, and integrate into a structured compound document in a synchronized manner. Actually, this recorded document is a good starting point for developing a courseware for e-learning. Toward this direction, we provide a post-edit toolkit for the teacher to edit the recorded document as necessary, such as to correct some mistakes made in the class, or to add some tags and indices to the document so that retrieval of the content of the document becomes more efficient.

Figure 6 shows the interface of the post-editing tool. It is composed of four main windows. The upper-right window displays the contents and events that appeared on the Mediaboard, including the loaded HTML page, the added annotations and the browsing sequence. The upper-left window displays the recorded live video. The lower-right window is for modifying properties of a recorded event. The lower-left window displays the recorded events in a time-line style. The timing relation between each event is obvious through this view. The cursor on the timeline can be easily dragged right and left to locate the wanted time point, and the content in the other three windows will jump to the correct position with the moving of the cursor.

Fig. 6: Post-Editing Tool for the Same View Recorded Class



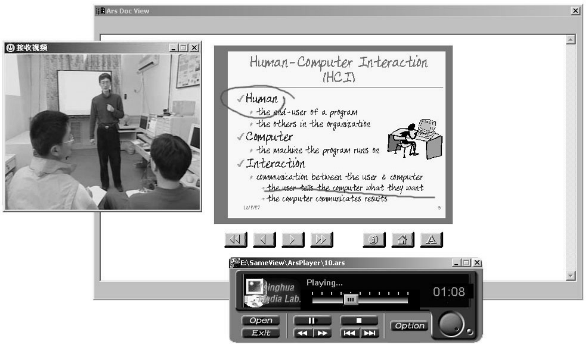
After post-edition, the recorded document can be put on a website, and students can download and play this document through a provided SameView Player program at any time they want after the class. Figure 7 shows the SameView Player program, which is similar to the SameView Client program, except for an additional VCR-like panel for playback control.

For the students to rapidly find the information they need while viewing the recorded class, two kinds of retrieval technique are provided. The first one is seeking by time, similar to a VTR control. The second one is seeking by keyword. For this kind of retrieval to function, the teacher should have manually tagged each event-of-interest in the recorded document in advance with the post-edit tool. We are currently considering using certain natural-language processing technologies such as digesting to automatically generate index tags for each recorded event in the document.

SMART CLASSROOM – A NATURAL FRONT-END OF SAMEVIEW

As analyzed in the introduction section, the desktop-based experience in most of today's RTIVC systems is not accept-

Figure 7: SameView Player



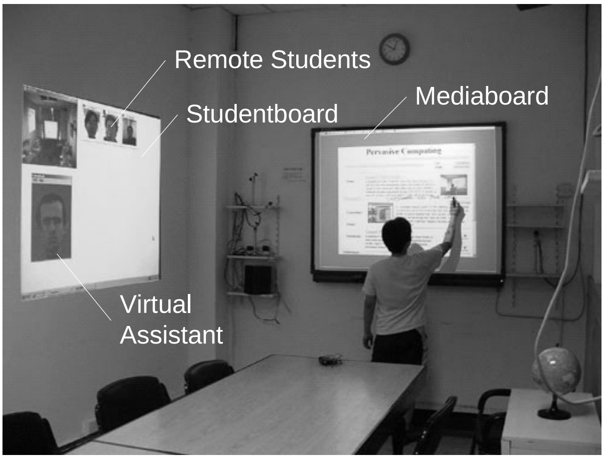
able for many teachers. For them, face-to-face classroom education is most familiar and efficient. Therefore, a question for researchers of RTIVC technologies is whether it is possible for the teachers to have the same experience as in a traditional classroom. Our answer to this is the Smart Classroom.

Smart Classroom is inspired by the research filed of Smart Space. A Smart Space is a richly instrumented physical environment where people can get transparent access to information and assistance from computers while performing their ordinary tasks in this space. (Refer to NIST, an overview of the research filed of Smart Space.) Smart Classroom is just such an effort to turn an ordinarily classroom into a Smart Space for Tele-education (Xie, 2001; Xie, 2002 Sep). We equipped an ordinary classroom with wall-sized displays, sensors, cameras and the associated software modules such as gesture-tracking module and speech-recognition module so as to allow the teacher presenting in the classroom to access the SameView system transparently, rather than appeal to a desktop computer. Through Smart Classroom, we actually extend the user interface of the SameView for the teachers from a desktop computer into the 3-D space of the classroom.

Room Setting

The space of a classroom usually can be divided into two parts - the teaching area that is mostly used and occupied by the teacher, and the audience area where local students reside. The most significant physical instrumentations in a Smart Classroom, as illustrated in Figure 8, are two large projector screens in the teaching area. The one on the front wall is a touch-sensitive screen, replacing the usual blackboard or whiteboard found in a classroom. It functions as a physical embodiment of the shared Mediaboard component of the SameView software at the teacher side. (We also call this physical screen a Mediaboard. Whether this name refers to the software component in the SameView system or this physical screen will be clear according to the context.) Another screen called Student Board is on the sidewall. It is the window for remote students, on which the image of remote students participating in a SameView session with Ad-

Figure 8: A Snap Shot of the Smart Classroom



dresser roles will be displayed and the video and audio of the remote student who takes the floor will be played here too.

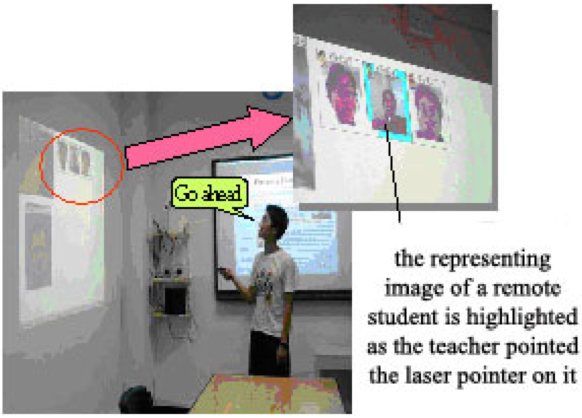
Besides these two obvious facilities, there are about a half-dozen cameras, each with different usage, installed at different places in the classroom. Some are used to recognize the action of the teacher and the result is interpreted as an interaction command to the underlying SameView system. Others are used to capture the live video of the classroom for broadcasting to the remote students. In addition, the teacher wears a wireless microphone to capture his speech.

Unrestricted Natural Teaching Experience

In Smart Classroom, in order to give classes to remote students, the teacher no longer needs to remain stationary in front of a desktop computer and, for most common tasks involved in a class, the teacher no longer needs to use the keyboard and mouse either. For this aim, the following technologies are developed and integrated in the Smart Classroom.

1. **Pen-Based UI.** As mentioned above, the Mediaboard is displayed on a touch-sensitive screen, which is actually a commercial product called SmartBoard. Most functions of a mouse found in a desktop setting can be accomplished by operating directly on this board. For example teachers can control the display of the slides. Moreover, using the provided pens and erasers, teachers can write comments and scribbles on the slides or wipe the strokes, as illustrated in Figure 7, which provides just the same experience as using a real blackboard or whiteboard.
2. **Speech-Capable Virtual Assistant.** The Smart Classroom incorporates a speech-recognition module as well as a text-to-speech module. Therefore the teacher can complete several common tasks in a class by voice command such as “Give the floor to Tom” or “Jump to the previous page.” The system can also use the synthesized voice to notify the teacher of certain events. For example when a remote student named Peter asks for the floor, the system will alert the teacher, “Peter is asking for the floor.” But interacting with a dummy room by voice will seem funny in some sense, so

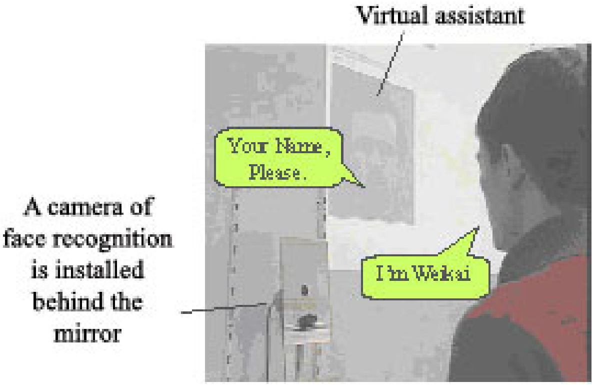
Figure 9: Give the Floor to a Remote Student by a Laser Pointer Plus Voice Command



we introduced a Virtual Assistant figure into the Smart Classroom to impersonate the classroom. The Virtual Assistant shows a face of a virtual person, which is displayed on the StudentBoard as illustrated in Figure 8 as well as in Figure 10, and its face expression and lips movement is synchronized with the synthesized voice of the system. Thus the teacher can see a vivid virtual assistant that can understand her/his voice command and give notifications or feedback through speech, just as if there is a real assistant.

3. **Laser Pointer as An Interactive Tool.** Laser pointers are widely used as a tool for indicating the focus nowadays. Besides this common utility, in Smart Classroom the teacher can even use a laser pointer as an interactive tool. The function is provided by a computer vision module that can track the movement of the spot of the laser pointer and recognize its certain movement patterns. This recognition result is interpreted differently according to whether the spot is on the Mediaboard or the Studentboard. If on the Mediaboard, the movement of the spot is interpreted as the mouse movement event and circling the spot around a point is interpreted as the mouse-click event, so that the teacher can select and open a link on the HTML page by laser pointer. Although the same task can be completed by manipulating on the Mediaboard directly, this is useful, for the teacher does not need to approach the Mediaboard every time he/she needs to open a link. If the spot is on the Studentboard, the recognition result is interpreted as the current selection of a remote student, and as an indication, this remote student's representing image will be highlighted. Together with voice command, it provides a convenient method for the teacher to control the floor among remote students. For example, the teacher can point the laser pointer at a remote student's representing image and say "Go ahead" (as shown in Figure 9), which will make the system give the floor to this remote student.
4. **Login in to the Classroom Based on Biometric Characteristic.** Since a Smart Classroom is a public space, each teacher who gives a class in it needs to identify him/herself as the member of the system in order to be authorized to use the facilities in the classroom. The common way of authentication on a desktop computer is inputting one's ID and Password with keyboard. In Smart Classroom, we use a combination of

Figure 10: A Teacher Login to the Smart Classroom



face-recognition and speaker-verification technology to automatically identify the teacher, and provide the teachers an unrestricted experience while logging into the classroom. As the teacher enters the Smart Classroom, he first should show up in front of a mirror (behind which a camera for capturing the teacher's face is installed) and speak out his name. If both the face-recognition and the speaker-verification are passed, the Virtual Assistant will greet him, indicating the Smart Classroom is now ready to serve him. This procedure is shown in Figure 10. Moreover, the system will use the teacher's identification information to load the right voice model for the teacher into the speech-recognition module if the teacher has trained such a model in advance, which is beneficial for the accuracy rate of the recognition result.

Smart Cameraman

When taking a class in a real classroom, the students will change the focus of their sights as the context of the class changes. For example, when the teacher is writing a formula on the blackboard, the students will focus their sight on the formula, while when the teacher is showing a model in his hand, the students will focus their sight on the model. However, in most current RTIVC systems, the students can only get the teacher-side video with a fixed scene no matter how the context of the class changes, which significantly decreases the efficiency of understanding of the teacher's instruction.

To overcome this problem, a facility called Smart Cameraman is introduced into Smart Classroom, which can distinguish among several kinds of contexts in a class by observing some cues in the classroom, and then select a camera with proper view according the context from an array of available ones as the source of the live video to remote students. Currently, this module can successfully distinguish the following three kinds of contexts:

1) **Teacher Writing on Mediaboard**, where the teacher is writing or scribbling on the Mediaboard. In this case, the camera which focuses on the Mediaboard will be selected, as Figure 11(a) shows.
2) **Teacher Showing a Model**, where the teacher is holding a model in his hand. In this case, the camera which follows the teacher's hand will be selected as Figure 11(b) shows.
3) **Others,** for all other situations. In this case, the camera with a overview of the whole classroom will be selected, as Figure 11(c) shows.

Figure 11: Different Scenes Remote Students Get According to Context of the Class

(a) The teacher was writing on the Mediaboard

(b) The teacher was showing a model

(c) The teacher was discussing with local students

The cues used by the Smart Cameraman to estimate the current context includes the output of a Person-Tracking module which tracks the position of the teacher, a Gesture-Recognition module that can decide whether the teacher is holding something in his hand and the touch event as reported by the SmartBoard.

THE SOFTWARE OF SMART CLASSROOM

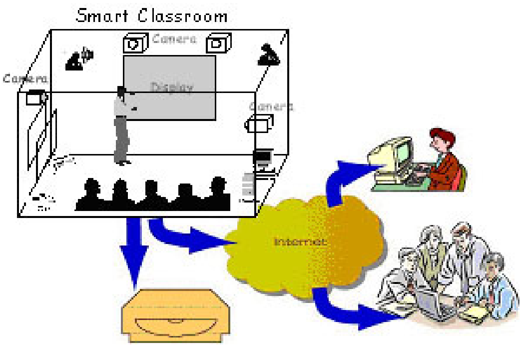
As we have seen, besides the SameView software, there are many other software modules running in the Smart Classroom that provide periphery services for the SameView system such as speech-recognition module, text-to-speech module, hand gesture tracking module and so on. In total we have about a dozen modules running on eight distributed computers and we developed a multi-agent system called Smart Platform (Xie, 2002 Oct) to interconnect and coordinate these modules. Here each module is running in a separate process; they communicate with each other through predefined XML format messages. For example, the module in charge of the laser pointer tracking will periodically send a message to the module in charge of the Mediaboard to update the location of the cursor while detecting a laser pointer spot.

Since the Smart Classroom is a fairly complicated system, its reliability is an important issue that needs careful consideration. Besides each composing modules are carefully designed and developed, the interconnection of modules takes a loose-coupling policy. First, we use message passing instead of RPC so that one module will not block the completion of a communication request. This also reduces the blocking possibility of another module which is communicating with a halting module. Secondly, we use a mediated communication scheme, in that all the modules only need to maintain a connection with a centralized message dispatcher module and all the inter-module communications are dominated by this dispatcher. This structure reduces the recovery work needed when a module is restarted after failure.

THE OVERALL SCENARIO

Figure 12 depicts how the above-described parts of the Smart Remote Classroom project are integrated into an overall scenario to enable a revolutionary real-time interactive distance learning practice. In this scenario, a teacher gives a class with natural ways in a Smart Classroom, where there might be local students, while the remote students connected by Internet ac-

Figure 12: Smart Remote Classroom System



cess this classroom with SameView clients. The remote students can see the present class materials, the annotations made during the lecture, the live audio/video of the classroom and also can take the initiative to interact with the teacher and local students in the classroom, just like attending the classroom locally. Furthermore, the process of the lecture will be recorded as a multimedia courseware for review after class.

CONCLUSIONS

We developed a set of key technologies for a real-time interactive distance learning system and made a new paradigm of real-time interactive distance learning with the following characteristics possible: 1) unifying the face-to-face education and tele-education in the Smart Classroom, which on one hand provides a consistent teaching experience for teachers and on the other hand reduces the teacher workforce needed, for the teacher does not need to give the same class for the on-campus students and remote students separately; 2) accepting large-scale users to access the virtual classroom simultaneously with different network and device capabilities; 3) allowing the class to be recorded and turned into a piece of courseware for review of the class of E-learning.

Currently we have made concrete achievements in each part of the project and the prototype system runs quite well in initial informal evaluations. (Welcome to http://media.cs.tsinghua.edu.cn/~pervasive.) We are planning a formal user study of this system in the near future. We are also cooperating with the Distance Learning School of Tsinghua University for the large-scope user experiments of this system.

ENDNOTE

* This paper is the extended version of the preliminary paper the appeared in ICWL2002 conference proceedings. The related research is supported by NSFC, “863” High-tech Plan, The Ministry of Education, China, and IBM China Research Lab.

REFERENCES

Clark, D.D. & Tennenhouse, D, L. (1999). Architectural Considerations for a New Generation of Protocols. In *Proceedings of the ACM Symposium on Communications Architectures and Protocols*, ACM Press, 200-208.

Liao, C.Y., Shi, Y.C. & Xu, G.Y. (2000). AMTM: an Adaptive Multimedia Transport Model. In *Proceedings of SPIE Vol.4210: Internet Multimedia Management Systems, IT2000*, 141-149.

NIST. Availble online at: http://www.nist.gov/smartspace

NS2. Availble online at: http://www.isi.edu/nsnam/ns

Kuo, F., Effelsberg, W. & Garcia-Luna-Aceves, J. (1998). *Multimedia Communications: Protocols and Applications.* Prentice Hall PTR.

Pei, Y.Z., Liu, Y., Shi, Y.C. & et al. (1999). Totally Ordered Reliable Multicast for Whiteboard Application. In *Proceedings of the 4th International Workshop on CSCW in Design*, Compigne, France, Sept. 1999, 391-395.

Pei, Y.Z., Shi, Y.C. & Xu G.Y. (2000). SameView: Real Time Cooperative Design on the Internet. In *Proceedings of the 5th International Workshop on CSCW in Design*, Hongkong, China: Nov. 2000. 210-214.

Raman, S. & McCanne, S. (1998). Scalable Data Naming for Application Level Framing in Reliable Multicast. In *Proceedings of The Sixth ACM International Conference on Multimedia,* 1998 , Bristol, United Kingdom, 391-400.

SmartBoard. Availble online at: http://www.smarttech.com/

Smart Remote Classroom Project. Availble online at: http://media.cs.tsinghua.edu.cn/~pervasive

Tan K., Shi, Y.C. & Xu, G.Y. (2000). A Pragmatic Semantic Reliable Multicast Architecture. In *Proceedings of The Third International Conference On Multimodal Interfaces,* Beijing, China, Springer-Verlag, LNCS 1948: 658-665.

Xie, W.K., Shi, Y.C. & Xu G.Y. (2001). Smart Classroom - an Intelligent Environment for Tele-education. In *Proceedings of The Second Pacific-Rim Conference on Multimedia (PCM 2001),* Beijing, China. Springer LNCS2195, 662-668.

Xie, W.K., Shi, Y.C. & Xu G.Y. (2002, Sep). Toward A Better User Experience in Tele-education–Recent Advance in Smart Classroom Project. *The Fourth International Conference on Ubiquitous Computing (UbiComp 2002),* Sep 2002, Goteborg, Sweden, Adjunct Proceedings, 51-52. Available at http://media.cs.tsinghua.edu.cn/~pervasive/projects/classroom/ubicomp2002.pdf.

Xie, W.K., Shi, Y.C., Xu, G.Y. & Mao, Y,H. (2002, Oct). Smart Platform - A Software Infrastructure for Smart Space (SISS). The *Fourth International Conference on Multimodal Interfaces (ICMI2002),* Oct 2002, Pittsburgh, USA, Proceedings, 429-434.

Zhuang, Z.Y., Pei, Y.Z. & Shi, Y.C. (2002). Loss Recovery Performance Comparison of Two Reliable Multicast Protocols (SRM and TORM). *Computer Engineering* (in Chinese). Vol.28, No.11, 2002, pp170-173.

Zhang, H.J. (2000). Adaptive Content Delivery: A New Research Area in Media Computing. *Proceedings of The International Workshop on Multimedia Data Storage,* Retrieval, Integration and Applications 2000, Hong Kong, Jan. 13-15.

Yuanchun Shi received all her Ph.D., M.Sc. and B.Sc. degrees from the Department of Computer Science, Tsinghua University. She is now a Professor of the department of computer science, Tsinghua University. Her research interests include pervasive computing, distributed multimedia and e-learning technology. She was a Senior Visiting Scholar at MIT AI lab during 2001-2002. She can be reached at: shiyc@tsinghua.edu.cn.

Weikai Xie is a Ph.D. candidate of the Department of Computer Science, Tsinghua University. He researches on Smart Space and Context-Aware Computing. He earned his B.Sc. from the Department of Computer Science, Tsinghua University in 1998.He can be reached at: xwk@media.cs.tsinghua.edu.cn.

Guangyou Xu is a Chair Professor of the Department of Computer Science, Tsinghua University. His research interests include computer vision and pervasive *computing. He was a Visiting Professor at Purdue University in 1982 and Illinois University in 1993. He is an IEEE senior member. He can be reached at: xgy-dcs@tsinghua.edu.cn.*

Peifeng Xiang is a Ph.D. candidate of the Department of Computer Science, Tsinghua University. He researches on context-aware computing and e-learning systems. He earned his B.Sc. from the Department of Computer Science, Tsinghua University in 2001. He can be reached at: xpf97@mails.tsinghua.edu.cn.

Baopeng Zhang is a Ph.D. candidate of the Department of Computer Science, Tsinghua University. He researches on cyber foraging and adaptive multimedia transforming. He earned his M.Sc. from North Jiaotong University and BSc from Tianjing Commerce University. He can be reached at: zbp02@mails.tsinghua.edu.cn.