Abstract - Using High Definition video distance learning allows large areas of the classroom to be captured at a resolution approaching human visual acuity. This allows instructors and students to interact with significantly less constraints than the classic television based distance learning design. Hi Definition Television (HDTV) is becoming commonplace. This drives cost down, but more significantly it creates an increase in student expectation as HDTV replaces standard television in the home. Implementing a useful distance learning classroom based on HDTV requires significant attention to layout in order to achieve a pedagogically functional classroom, especially when applications beyond a simple lecture are considered. The classrooms considered here use two channel video in each direction, as well as stereo audio. This paper will discuss the issues in capturing an entire instructional area (whiteboard and podium) using multiple HD cameras. Careful attention to audio design is also critical to natural interaction.

Index Terms – Distance Learning, Telepresence, high definition video, natural interaction

INTRODUCTION

Globalization has brought an increase in the distances between different groups working on the same project. Organizations find themselves increasingly spread across multiple cities, countries, and continents. Even small organizations find that their partners, customers, clients and other business collaborations extend beyond the range of easy face-to-face interaction.

Unfortunately, the current paradigm of education does not address the modern workplace’s requirements for producing scientists, engineers and other skilled professionals who are experienced in the issues and practice of collaborating across a distance. Common opportunities for student collaboration in science and engineering education rely primarily on brief teamwork experiences on small-scale lab projects or one-semester major design projects. Likewise, researchers often find current distal collaboration to be more cumbersome than working with local participants. Industry finds that ongoing education of their workforce currently involves lost work time as current approaches require students to be on a college campus or in a separate facility, and thus be separated from their co-workers. This induces a passive learning role rather than encouraging them to participate from the workplace as professionals who can share workplace knowledge with less-experienced fellow students.

Common videoconference and distance learning systems provide students with a traditional television-like view of the instructor and/or the instructor’s materials. This requires the instructor to use specially prepared materials to effectively reach the remote students. The instructor’s view of the remote students, if one is even available, is similar to a television view of a crowd. Individual students’ identities and facial expressions that would show their level of understanding are thus lost due to poor imagery. Furthermore, mediocre audio systems have hampered remote students and instructors ability to communicate verbally. Remote students are thus much less engaged in classroom learning than local students in an existing distance learning facility, even when instruction is specifically tailored to the distance-learning environment. As a result, distance learning and related collaborative facilities have seen only limited use.

The Arbutus Center for Distributed Engineering Education at Georgia Institute of Technology's School of Electrical and Computer Engineering has been working extensively in conjunction classroom technology including distance learning methods[1]-[2] in conjunction with the Georgia Tech Savannah program[3]. The Arbutus Center is currently evaluating a high definition television based classroom prototype for use in the next phase of distance learning classrooms, with potential for impact throughout the state of Georgia as well as through GT’s international campuses. This architecture seeks to approach telepresence, the sense of presence in other than a person’s true location, for classroom application, in a cost effective manner. The resulting compromises may result in a system that does not fully achieve telepresence, but still facilitates natural interaction.

CLASSICAL DELIVERY

Live delivery of an educational setting to a remote classroom has been one form of distance learning. This type of distance learning has generally used classical television camera and display technology and delivered that signal via broadcast, satellite, leased line, and in the last decade via Internet techniques. In the late 1990s the State of Georgia had approximately 300 distance learning classrooms connected via a leased line system through a partnership with BellSouth, known as the Georgia Statewide Academic
and Medical System (GSAMS). With the migration to Internet Protocol (IP) transport there is no longer a formal state-wide infrastructure, but it is believed that roughly the same number of distance learning rooms exist throughout the state.

Unfortunately, the television format is a poor fit for classroom applications. At approximately 640x480 resolution and 30 frame per second, written material, whether printed or handwritten on a whiteboard, is very difficult to transmit via this format. The standards for videoconferencing generally result in image degradation in the lossy compression algorithms used, further hindering the use of written material. Despite this mismatch, the whiteboard continues to be a vital component of any contemporary classroom. Attempts to address this issue include requiring the instructor to use material prepared and distributed prior to class (often via the web), constraining the instructor to very small whiteboards, or paying a camera operator to pan a camera around a larger whiteboard in attempt to follow the instructor at a pace suitable for remote students.

Perhaps more significantly, the students in remote classrooms are often isolated from participating in the class. The instructor usually has one monitor which portrays all of the remote students as very small figures. Typically this monitor is located somewhere at the back of the classroom. The result is that the instructor is frequently unable to even recognize the remote students faces, much less get cues on comprehension, attention, etc. In addition to the limited view of the material, remote students often deal with poor audio quality.

Such a system totally breaks down when attempting to create group interaction across such a channel. Yet modern education theory suggests that interactive learning, with students engaged in the learning process, is greatly superior to lecture style presentations. Certainly such interaction is to eliminate panning (and the natural interaction is to eliminate panning (and the associated expense of a camera operator), allowing participants to select their point of focus.

Typical human visual acuity is 1/60 of a degree, which corresponds to 20/20 vision. A student sitting 20' from the whiteboard (i.e., near the center of the typical classroom seating), sees detail down to 0.07 inches, or a resolution corresponding to 14 pixels or dots per inch (DPI). While non-square pixels, analog to digital conversion, and image boundary areas effect the exact conversion, Table 1 illustrates the approximate image size for various video standards based on 0.07 inches per pixel. The first entry, Common Intermediate Format (CIF), is the resolution commonly used by traditional videoconferencing H.323 standard videoconference codecs. At a distance of 20 feet and at 14 DPI, a CIF resolution corresponds to a 2 foot x 1.7 foot area. Therefore a CIF based videoconference system’s camera should zoom to cover a whiteboard area of 2’x1.7’ to provide the same detail as a student with average vision sitting 20’ from the whiteboard. If a videoconference system camera zooms out to a larger area, then the video image will loose detail compared to the in-class student positioned 20 feet from the board. Constraining a instructor to a 2’ x 1.7” whiteboard is impractical. Typically the camera is either zoomed out to a larger area and the remote image quality suffers, or the camera is panned to follow the instructor’s writing. A goal of this work in natural interaction is to eliminate panning (and the associated expense of a camera operator), allowing participants to select their point of focus.

### Video Format and Image Field Size at 20 Feet

<table>
<thead>
<tr>
<th>Video Format</th>
<th>Resolution</th>
<th>Image Field Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIF</td>
<td>352 x 288</td>
<td>2’ x 1.7’</td>
</tr>
<tr>
<td>NTSC</td>
<td>720 x 480</td>
<td>4.2’ x 2.8’</td>
</tr>
<tr>
<td>720p</td>
<td>1280 x 720</td>
<td>7.5’ x 4.2</td>
</tr>
<tr>
<td>1080i/p</td>
<td>1920 x 1080</td>
<td>11’ x 6’</td>
</tr>
<tr>
<td>Dual 1080</td>
<td>3840 x 1080</td>
<td>22’ x 6’</td>
</tr>
</tbody>
</table>

Traditional classrooms are configured to support an instructional whiteboard area of at least 16’ wide by 4’ high, with another 4 feet of width dedicated to an instructor podium, for a total of a 20’ wide instructional area. While larger areas are common, this 20’ working area will be assumed to be sufficient for normal instruction. Referring to table 1, Dual 1080 closely matches this area, and therefore provides a reasonable sense of classroom telepresence. Three adjacent monitors at 720p also provide similar width but at a reduced vertical area and increased complexity of three displays and cameras rather than two. While interlacing causes degradation of rapidly moving images, for the relatively slow moving content of a whiteboard the image quality is nearly identical for 1080i (interlaced) versus 1080p (progressive).
An issue is created when dealing with multiple cameras to capture a wider field of view. To ideally capture 20' of whiteboard width two cameras would be placed 10' apart (usually in the back of the class), so that each has a view perpendicular to their perspective whiteboard area. For anything closer to the cameras than the whiteboard, there is a gap between the cameras' field of view. This can result in the instructor completely or partially disappearing as he or she walks across the front of the room, as shown by the small square near the center of figure 1.

While placing the cameras close together but at a different angle as can nearly eliminate this issue as shown in figure 2, the resulting image then has the whiteboards appear as trapezoids due to the increased distance of the outer edges of the whiteboard. One proposed solution to this issue is to use a concave angle between the two remote displays to place the outer edges closer to the viewer. This approach only works correctly for viewers sitting in one location in the remote room, on the centerline between the displays at a proportional distance to the displays. Angled displays may also be prohibited by square room geometry.

For the assumed 20' wide instructional area, the issues with either one of these camera configurations is reduced as the camera is moved farther away from the whiteboard in conjunction with a greater lens magnification. Generally this is limited by the back wall of the classroom. It is also possible to compromise between the two configurations, so that the cameras are placed closer together than 10', but still at some separation. Using cameras 40' from the instructional area, spaced at 5' apart, provides a compromise that only slightly distorts the image, and creates only a few inches gap at the typical distance an instructor would use for a whiteboard. Such a gap is visually tolerated especially in conjunction with the case framing around remote display.

Image processing and rendering techniques could be used to horizontally warp the image to eliminate that effect, much like the vertical keystone correction in modern computer projectors. Such rendering techniques are not available in current codec or display implementations, and may require additional resolution to avoid image artifacts. Image processing techniques could also be used to correct for minor misalignments in the cameras in order to create one large continuous image.

Camera resolution drives the usable field of view for the image capture, and remote displays must be able to accept an image of that resolution. Once the resolution is fixed, the ideal display size is determined by the distance to the viewer. The display should provide 1 pixel per 1/60° degree of field of view to again match human visual acuity. Consistent with table 1, for 1080i/p, this works out to a monitor horizontal dimension approximately ½ the distance from the viewer to the monitor. Since display pricing is non-linear with display size, it is often more cost effective to use multiple smaller display areas around the classroom rather than one large centrally located image. Since the dual camera image is typically rendered on two physical displays, the term ‘display area’ will be used to identify the combined physical displays that render one complete view of the remote site.

**NATURAL INTERACTION**

While this discussion has focused heavily on the display of a classroom whiteboard, the ability to render the equivalent instructional area provides the necessary facility for classroom instructors to work un-impeled by the distance learning technology, therefore achieve a true sense of classroom telepresence. In addition to visual acuity, other issues that must be addressed to promote natural interaction include:

1. **Eye Contact**

Eye contact is an important part of human interaction, and therefore it is necessary to preserve the geometries that facilitate eye contact in an interactive classroom. For appropriate eye contact, the camera must be as close as possible to the display, so that looking at the display of the remote room equates to looking into the camera. This often conflicts with the desire to use multiple smaller display
areas, as a single camera view of the room cannot coincide with the location of the different display areas.

II. Audio

The other component which is often poorly managed is the classroom audio. Many current distance learning classrooms mute the microphones in remote classrooms to reduce ambient noise at the instructor’s site. While this practice may be necessitated by poor acoustics, this dramatically reduces the sense of presence, and therefore pursuing appropriate open microphone audio technology is highly beneficial. Table or desk top microphones tend to accentuate noises from papers, books, pencil tapping, etc. Ceiling mounted microphones provide significantly less noise pickup and generally cover a greater number of students per microphone, reducing the total number of microphones required.

Initial testing also shows significant benefits from using stereo audio connectivity. Human perception uses stereo audio to determine the location of a sound source. This ability has also been shown to improve voice recognition in the presence of noise[4]. This also allows for voice localization, and is therefore critical in highly interactive scenarios.

III. Latency

The encoding, transmission, and decoding of the transmitted signal involves processing, queuing, and propagation time. It is critical that audio and video latencies are equivalent as human perception of speakers naturally relies on comparing lip motion to perceived sounds (commonly known as lip-sync issues). The actual amount of latency that is tolerated depends on the level of interaction, with less latency required for increased interaction. Non-interactive presentations can be completely pre-recorded (a very high latency). Based on telephone latency requirements, highly interactive conversation requires latency limited to about 200ms to preserve human speech cues for interrupting other speakers. It would appear that a question and answer session based on students raising hands, for example, could tolerate longer delays but this has not been well studied.

Session T3C

The remote students’ view of the instructional area is addressed by the previous section, with tradeoffs between parallel camera views and single point camera views. The instructor should have a view of the remote students at an image size that is consistent with the image size of students in the class, and within the instructors normal field of view. Displays on the far wall of the classroom need to be extremely large to accomplish this, as remote students would have to be shown at twice life size to have the same visual impact on the instructor as a student sitting in the middle of the room. Placing a small pair of monitors in the podium achieves the same visual angle with a much smaller pair of displays, two 19” widescreen monitors in a podium provide a sufficient display for someone standing at that podium, but do not remain in the instructor’s field of view when working across a moderate sized whiteboard. Mounting the monitors below the desktop of the first row of students, or near the ceiling at that distance, allows a moderate sized pair of displays to remain in the instructor’s field of view from within the normal instructional area.

II. Lecture Receiver:

Students watch a lecture or presentation from a remote location, with limited interaction (question and answer).

In this mode the students need good display of the remote classroom. While LCD screens generally provide the best brightness, dimensions of the current large size screens place the optimal viewing distance at 10’, which requires multiple display areas to provide good visibility to even a moderate sized classroom. An alternative is to use projectors which can create very large images. Unfortunately most of the current 1080i/p capable projectors are designed for home theater use and have insufficient intensity to provide a large useable image in conjunction with standard classroom lighting.

Cameras position at the front of the room, in a single point configuration, provides a good view of the students.

III. Group Interaction:

In a group interaction scenario, students at either end interact together, with or without the participation of an instructor or discussion coordinator. Configuring two rooms in the lecture receiver mode allows both groups to see each other, and interact accordingly. Unfortunately in this configuration, it is not easy to use the whiteboard.

IV. Full Interaction:

A flexible configuration can be created by arranging a room similar to figure 3. This allows two rooms of students to see each other, and an instructor, and instructor whiteboard. This configuration basically places the instructional area, display area, and student seating in a triangular configuration. For eye-contact purposes, it is desirable to place the cameras at the center of the display area, but moving the cameras slightly counterclockwise in the figure improves the view of the board. This configuration

LAYOUT AND CLASSROOM MODALITY

The optimal configuration of a distance learning classroom depends strongly on the mode of instruction (interaction) used. Ideally the room should be flexible enough to accommodate changes in the modality during the progression of a single course. Some common modes are:

I. Lecture source:

An instructor presents to the entire course, both local and remote. Student to student interaction is limited to the need to overhear questions. The students need to see the instructor and presented material. The instructor needs to see both local and remote students.
generally also provides less seating in the room than a traditional configuration.

This configuration needs to switch between to camera viewpoints. Cameras mounted on the opposing wall provide optimal view of the whiteboard for intensive written material, as shown by the fine dotted lines in figure 3. For general group interaction, the wider angle camera view set in the center of the display provide a full view of instructor and students, but distort the whiteboard by being off center, and provide significantly less resolution for material on the whiteboard as a significant portion of the cameras field of view is focused on seating.

![Interactive Triangular Configuration](image)

**CURRENT TECHNOLOGY**

Several vendors are releasing high definition videoconference codecs. 720p resolution is more commonplace than 1080i/p. Many of these systems implement a single video channel and the simultaneous use of two video streams requires two codecs. Other vendors provide the capability to encode a second video signal, but often at a low frame rate, with the intention of using this second video channel to support a remote computer display.

Encoding of high definition signals commonly yields between 2-5 Mb/s of network traffic per video channel, depending on resolution (720p, 1080i, or 1080p), the complexity and amount of activity in the scene, and the codec implementation. Such traffic can be both disruptive to other traffic, and sensitive to disruption from other network traffic. Therefore, careful planning by knowledgeable network architects is necessary for successful and reliable use. Certainly these systems are not deployable across the current public Internet, although they can be used in conjunction with a Internet traffic on a well designed network.

Display technology is a limiting factor in the design of a high definition classroom. Using undersized displays eliminates the benefits of improved resolution. Current plasma and LCD technology provides sizes up to approximately 5 foot diagonal measurements. Such size places the ideal viewing distance at 10' from the display, significantly less than the dimensions of even a small classroom. Larger displays are available with projection technology, but sufficient intensity for classroom use is lacking. Life sized images could be created with projection technology, but would equate with the center of a modest sized classroom. (This is not coincidental since that 20 foot distance was used for the camera field of view and resolution).

**SUMMARY**

Commercial telepresence systems are designed for an ideal viewing distance in the neighborhood of 10 feet from user to display. These are generally designed to be consistent with a small conference room where people seated around a table are rarely more than 20 feet apart (twice the user to screen distance). Classroom applications deal with larger audiences, and larger distances, with twice the above distances being appropriate for a small classroom. Legibility of written material such as on a whiteboard is frequently a driving factor in classrooms, although the ability to discern facial expression is important. Two adjacent high definition displays at 1080i/p resolution, in conjunction with matching cameras provide a baseline visual experience that closely matches typical in-room seating.

There are multiple modes for such classrooms, and supporting multiple modes in one room results in both complexities and trade-offs in the implementation. Complexities generally increase the costs, and trade-offs generally reduce overall quality of some element of the distance learning experience. Well planned audio is often overlook, but should receive at least as much attention as the video configuration.

**REFERENCES**


AUTHOR INFORMATION

Randal Abler  Associate Director of the Arbutus Center and Assistant Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, randal.abler@gatech.edu

Sean Brennan  Technical Manager, Academic and Research Technologies, Georgia Institute of Technology, sean.brennan@oit.gatech.edu

Joel Jackson  Director, Digital Media Laboratory, Georgia Institute of Technology, jacksonj@ece.gatech.edu