Automatic Video Production of Lectures Using an Intelligent and Aware Environment

Michael Bianchi
FOVEAL SYSTEMS, LLC
190 Loantaka Way
Madison, New Jersey 07940-1910
MBianchi@Foveal.com

ABSTRACT

This paper makes the case that much of the promise of ubiquitous multimedia depends on the availability of valued material. In business and academic environments the "presentation in a lecture room with laptop graphics" is a common way of communicating, but making a presentation readily available outside of the room is still a challenge because of the complexities of capturing and distributing the material. The AutoAuditorium™[1] System creates a multi-camera video program of a lecture in real time, without any human control beyond turning the system on and off. The system reduces the opportunity costs of making such a program to the point that it gets used for events previously not seen as candidates for video. Thus an event does not need nearly as many viewers to be considered worth capturing and many more events are seen by many more people.

This paper presents a quick overview of the AutoAuditorium System technology and operational characteristics, a history of it's ancestry, development and use, and some usage experiences that demonstrate its current utility and future potential.

The AutoAuditorium System is an example of an "intelligent and aware environment." In particular, it is:

intelligent	– about creating multi-camera television programs of lectures, in real time, with one or more people on a stage using projected visuals.
aware	– of the motion and gesturing of the people on stage. – of changes in the projected visuals.

Originally created in the early- and mid-1990s as a research project [1] at Bellcore (Bell Communications Research, now Telcordia Technologies), it has been available as a commercial product from Foveal Systems since 1999.

[1] **AutoAuditorium** and **AutoAud** are trademarks of Telcordia Technologies Inc., used under license.

MUM2004 October 27-29, 2004 College Park, Maryland, USA

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Communications Applications — *Videoconferencing*; I.4.9 [**Image Processing and Computer Vision**]: Applications; K.3.1 [**Computer Uses in Education**]: Distance Learning

Keywords

automatic video production, multimedia content creation, multimedia production

1. OVERVIEW

The AutoAuditorium System [2] adds to an ordinary auditorium, lecture hall or classroom the ability to automatically makes video broadcasts and recordings of lectures and talks. Permanently installed in the room, it uses cameras and microphones to be "aware" of what is happening on stage. It televises, in real time, the most common auditorium, lecture hall or conference room talk: people speaking, showing projected visuals to a local audience. The intent is to allow other audiences to see and hear the event as a television program in another place or at another time. Three subsystems create the TV program:

- An automatic Tracking Camera tracks the person or people on stage.
- The automatic Director selects camera shots based on what is happening in the program.
- The automatic Audio Mixing combines stage and audience sound to create a complete sound track, including questions and answers.

Because the system is built into the room the people on stage and in the local audience are not distracted.

Because it is produced using multiple cameras and appropriate video effects (e.g. picture-in-picture), an AutoAuditorium program is often indistinguishable from one produced by a crew.

Because it is completely automatic, it is easy and economical to use, and therefore it is used often. Thus many more people can avoid traveling, or missing talks because of a schedule conflicts. Classes, talks and seminars occurring in one location can be telecast as video programs to other locations and/or recorded for later use. Now the everyday events of business and education can become the feed stock for "ubiquitous multimedia".

2. THE HISTORY OF THE AUTOAUDITORIUM SYSTEM

Reducing Manual Production to One Person. Before the mid-1970s, television technology was only usable by organizations willing to finance a professional operation. The most common of those were television broadcast stations and companies that produced programs for sale to TV stations.

The advent of video tape technology arose from the economies of scale that the transistor and integrated circuit revolutions brought to electronics. Complex functionality was now encapsulated to the point that in-depth engineering knowledge was no longer necessary to use the equipment and create programs. The video cassette meant that virtually anyone could play TV programs when needed and corporations started using video for programs previously presented on 16 millimeter film. The prices were then affordable in many industrial settings.

The home video cassette accelerated the economies of scale and the attendant reduction in prices. By the late 1980s, it became clear that video technology was moving from the "industrial age" into the "consumer age" and technology created for the home market made industrial equipment much more reliable and affordable.

In this environment, the demand for video recordings of corporate events and functions grew rapidly and the VHS home video format became ubiquitous. The assumption was that everyone had access to a VHS video cassette player, and in technology companies many types of educational and corporate communications were distributed as "video". The "VHS" was understood.

But many of the programs created in the corporate world made poor use of the training and skill of professionals. When the material being presented was little more than a lecture with projected visual aids in the form of 35mm slides, letter-size foils, or (eventually) computer-generated stills, the programs demanded very little of the production crew:

- Follow the person speaking on stage.
- Switch to the projection screen when appropriate.
- Switch back to the person.

Many organizations tried to "roll their own" video programs using just a camcorder, but the results were usually disappointing, tedious to watch, and/or unintelligible. So while many families made home videos using a camcorder, creating the material in corporations remained a "professional" activity, requiring several trained people and professional equipment unavailable to the majority of home videographers.

In the late 1980s, a project called the Distributed Auditorium System at Bellcore reduced the crew requirement for the basic lecture program to one person. The key insights were that:

- TV cameras and related equipment were now inexpensive enough to install permanently in a room.
- TV equipment was acquiring useful automatic features, such as autofocus and automatic white balancing.
- a single, trained person could usually keep up with the requirements of a simple programs.

Figure 1: The Distributed Auditorium Console. Notice the three camera monitors, each above a cluster of controls for that camera.

Although the standard equipment was organized around a one-person-for-each-job model, it was possible to redesign the workspace, using human factor design principles, so most of the controls were usable without looking at them. The operator knew which control they were handling by its shape and direction of action.[2]

A goal of the project was to create a means of recording and telecasting talks that had minimal impact on the people giving the talks and the people watching them. We did not want an environment that felt like or operated like a television studio. Instead we wanted to create a space that looked and felt like an ordinary meeting room and where the technology was unobtrusive (if not invisible) to all involved. So the cameras were hung from the ceiling, added lighting was made as inconspicuous as possible, microphones were also hung from the ceiling, and all evidence of television production was put in a back-room.

Figure 1 shows the operator's console of an original Distributed Auditorium System. The central control panel (two color monitors, three camera monitors and consolidated camera and video mixer controls) enable one person to position, adjust and select the three remotely controlled cameras. Only the most useful controls from the surrounding equipment were duplicated in each camera's control cluster.[3]

Two cameras are hung on the ceiling in the center of the room, pointed at the stage. The first is used primarily for following the person speaking on stage. The second is used primarily to capture the images on the projection screen. The third camera was hung above the corner of the stage opposite the lectern. It was primarily used to look across the stage, taking a shot of the person speaking from the side. It could also look at people in the audience, especially when

[2]For example, the Focus control moved left-and-right while the Zoom control moved up-and-down.

[3]For example, each camera cluster has one button to put that camera's image into the program. Those three buttons are duplicates of ones found among the 75 on the video mixer immediately to the right.

someone was asking a question.

In the Distributed Auditorium environments the operators were seldom people with television production experience. But a well written training manual, some time to practice, and constructive criticism quickly resulted in programs that were worth watching if you were interested in the topic. These were not award winning productions, but to someone faced with either the time and expense of traveling to the event or missing the presentation altogether they were far better than nothing.

Four Distributed Auditorium Systems provided video tapes and live inter-location telecasts for over a decade at Bellcore. The majority of talks important enough to attract a sizable audience were scheduled in the DistAud auditoriums and people's expectations became that they would be televised and taped. Many video cassette recordings were cataloged into the company library.

Genesis of the AutoAuditorium System. The existence and successful use of the Distributed Auditorium Systems as corporate resources motivated the projects which ultimately became the AutoAuditorium System. Although they grew incrementally, the goals of the new projects grew out of the goals successfully implemented in the older one. We wanted the system to be unobtrusive, reliable, valuable, easy to use and frequently used.

Because the operators of the DistAud Systems had other responsibilities, it was not possible to schedule them for every talk. Those missed opportunities motivated experiments in partially automating, and then fully automating, the video production of lectures.

Automatic Audio Mixing. The first practical automation of the Distributed Auditorium System involved automatically mixing the audio from the stage with the audio from the audience, so the remote viewers could hear both. Experience quickly taught us that no matter what we did to encourage people in the audience to use a specific "question microphone" they would simply shout out their questions or comments during talks. The size of the rooms and the seating plans simply made informal exchanges too easy. We could not enforce any other discipline. The Distributed Auditorium's control console was given an easy-to-use audio mixer so the operator could quickly turn up the microphones hung from the ceiling above the audience, but that often proved unsatisfactory.

After much experimentation and a few false starts, an automatic audio mixer with input priorities was created. The wireless microphones, if used, were given priority over microphones above the stage, which were given priority over the mics above the audience. The result worked well. Polite give-and-take between the primary speaker (wearing the wireless mic), other people who came on stage and the audience resulted in good coverage. If more than one person spoke at once, the priorities decided which person would actually be heard clearly by the remote viewers. The automatic audio mixing tended to work without attention. It also behaved well when the wireless microphone was forgotten, accidentally turned off or suffered a dead battery. It would revert to the microphones mounted over the stage.

The features that make priority audio mixing practical are now common in modern audio mixers, but were not found in the early 1990s.

The automatic audio mixing was often acceptable without any adjustment. In fact, when we could not schedule an operator, we would sometimes set the video mixer on a single picture-in-picture image, combining the projection screen with an inserted image of the lectern, and record the talk with just that image and the automatic audio mix. Since the programs were not intended as entertainment, the audience was often tolerant of less-than-ideal "production values" if they could see most of what they wanted to see and hear most of what they wanted to hear. "*Anything* is better than nothing!".

This success made adding further automation to the system seem practical.

Automatic Camera Tracking of the People on Stage. A common problem in Distributed Auditorium System programs was that the operator could be distracted by motion. A person on stage who paced back and forth would prompt the operator to track them with one of the remote control cameras. This required constant attention, and so the operator sometimes would miss the fact that the *interesting* image was not the person but rather what was on the projection screen.

This is exactly the reason that real television production is done with a person behind every distinct job. When things get busy, each camera operator manages a camera, the person running the video mixer has both hands full, the audio mixing keeps another person busy, while the director does nothing but watch all the monitors and call out instructions to the crew.

Even at a well designed console, having one person trying to serve all those functions inevitably resulted in occasional cognitive overload or running out of hands.

This need motivated a Bellcore research project in motion tracking. The result was an automatic Tracking Camera called the ICU, Intelligent Camera Unit,[4] that was eventually added to one of the Distributed Auditorium Systems. The goal was to give the DistAud System operator an "assistant" which would always follow the person on stage.

In 1994, when this work was being done, there were commercial products that were also intended to track a person automatically, but they required that the person wear a target device that ran on batteries. The ICU avoided that need by using two cameras:

- a Spotting Camera that watched the entire area of interest, and
- a robotic Tracking Camera that followed the moving person.

Software analyzes the Spotting Camera image, looking for moving objects on the presentation stage. It then points the Tracking Camera at those moving objects by robotically moving the camera and adjusting the lens. See **Figure 2**.

While simple to describe, creating a refined ICU took some doing. But eventually the ICU Tracking Camera became useful enough that it proved a great boon to the operator by considerably lightening the work load. The result was fewer cases of not showing the most interesting image.

Reducing the operator's load didn't fix the scheduling problems. But it did offer an alternative.

By using the ICU Tracking Camera for a picture-in-picture shot of the person and the projection screen, we improved

[4] Yes, it is a pun: I-See-You.

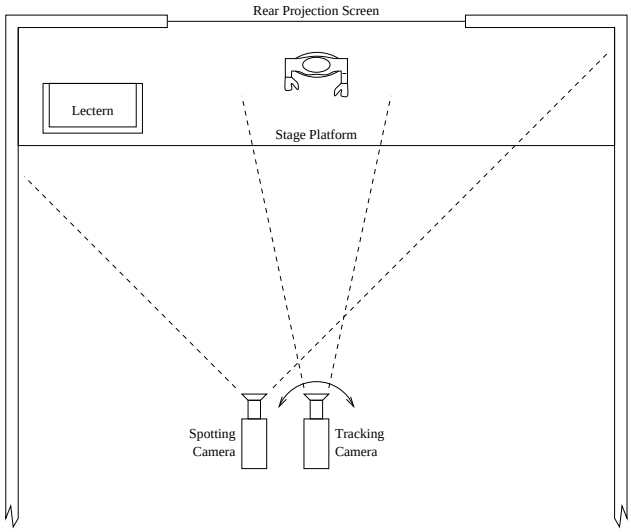

Diagram of the overhead view of the Spotting and Tracking Cameras.

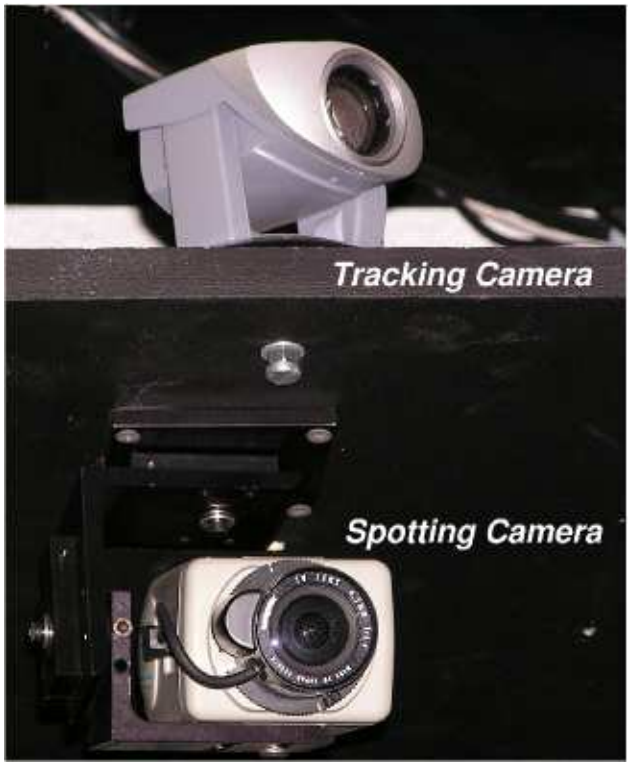

Figure 2: The Tracking and Spotting Cameras are mounted close together to minimize parallax between the Spotting Camera image and the pan, tilt and zoom axises of the Tracking Camera.

Figure 3: Some of the main components of the AutoAuditorium System. The Video and Audio Mixers are not shown.

the operatorless program quite a bit. It was *still* just one picture-in-picture image for the entire program, but there were fewer times when the person on stage was not visible. Again, these programs were not "good television" but the content telecast was so important to some viewers that the effort was praised.

Automatic Camera Selection. Another insight lead to the full automation of the program production. The same technology that made the ICU Tracking Camera work could also determine:

- if the projection screen was blank or not, and
- if the projection screen had changed, and by how much.

From these observations, and a few heuristics and timers, the system acquired the ability to react appropriately to most of the events of a presentation.

By making use of the serial control interface on the Distributed Auditorium System's video mixer, the entire program could be produced in real time without any human control beyond:

- turning the system on,
- connecting the video signal to the corporate television network (if telecasting), and
- starting the VCR (if recording).

The result became known as the AutoAuditorium System (AutoAud for short) and as it became more refined and more accepted, it gradually took work away from the human-controlled Distributed Auditorium System.

The original prototype AutoAuditorium System, considerably updated, still runs at Telcordia Technologies. It still shares the automatic audio mixing function with an original Distributed Auditorium System. See **Figure 3**.

The keyboard, mouse and monitors are only used for maintenance functions. Normal operation is controlled by an ON/OFF switch and an Operating Mode switch.

Commercial Applications of the AutoAuditorium System. During the late 1990s the audio/video industry started to commercialize the technologies that made the AutoAuditorium System possible. Originally it required several hardware "hacks" to put the right combinations of remotely controlled automatic focus, remote control zoom with position feedback, camera and video mixer synchronization (aka "gen-lock"), and video mixer remote control into the same application. Television equipment of that era often assumed that a person was ultimately in control of the

equipment. Remote control interfaces often did not provide feedback of the equipment state. The assumption was that the human operator could always see the image or which buttons were lit. Getting around those limitations took either considerable imagination or a willingness to void the warranties by opening up the equipment and soldering wires where needed.

Slowly, oh so slowly, the industry began to understand that more and more of their equipment would be controlled by computers instead of by people, and the interfaces became more complete.[5]

IBM Watson Research became the first commercial customer of the AutoAuditorium System in December 1999. By the end of 2001 they had three AutoAud Systems. Originally purchased in support of their e-Seminar [9] research project, it has since become a corporate service managed by the Audio/Video group.

The first IBM installation was in a newly renovated 110 seat lecture hall, designed to support studio-quality videos with permanently present cameras and microphones. Several fortuitous coincidences made it relatively easy to install the AutoAuditorium System as an alternative way of making videos in that room. They had selected camera equipment and pan/tilt heads that were compatible with AutoAuditorium control software, and their equipment placement was also suitable. When the renovations were completed, they quickly started using the AutoAuditorium System to make videos. The AutoAud System did such a good job (even with some of its first-installation learning curve issues) that the need for manual productions was greatly reduced.

The second IBM system was installed in their 300 seat formal auditorium as a retrofit. One installation challenge was an existing movable wall that could split the room in half. That required that there be two AutoAuditorium configurations, each quite different. A single rotary switch controlled the entire system, selecting between:

- System Off
- Full Room
- Half Room
- Maintenance Mode

The fact that the system could be controlled with single switch made user training minimal. AutoAuditorium videos made in that room also became popular.

IBM's third system uses one controller and one video mixer but two sets of cameras and microphones, one set in each of two rooms. A video+audio routing switcher under AutoAuditorium System control determines which set is used at a particular time. Again, a single rotary switch selected between the two operating modes.

Boeing Phantom Works also has an AutoAuditorium System in a room which can seat up to 80 people. It is very similar to the first IBM installation.

The original AutoAuditorium prototype system is still at Telcordia Technologies. It is now over 8 years old and was recently moved to a larger auditorium and updated.

[5]Today it is still common to find "computer-controlled" equipment where action commands do not have corresponding status queries, and vice versa, or where status cannot be requested or is not reported correctly while an action is in progress. "Can't you see what is happening?" No, I cannot. The good news is that some of the manufacturers do a good job of providing computer interfaces as capable as the human controls.

3. RECENT EXPERIENCES

The best measurements of AutoAuditorium System use come from IBM Watson Research. They have three systems in two locations in down-state New York; the first is at their laboratory in Hawthorne and the other two are in their Yorktown Heights research headquarters, about 10 miles away. These are connected by high quality video links so talks given in one location's AutoAuditorium room can be seen easily in the other locations.

Prior to the e-Seminar project and the AutoAuditorium System installation, the Audio/Video group made about 50 crew-based video productions a year.

When the AutoAud Systems were installed at IBM Watson Research, they were used to create source material for their e-Seminar project, which was already in progress. (In 2002 the project was renamed "Research Media Portal" [6].) Over time they moved away from manually recording programs, which resulted in more programs being captured. The room reservation system was given an additional check box to request an AutoAuditorium recording. At first the recordings were made on video tape, which were later encoded for delivery over IP networks. IBM has eight laboratories around the world, and these encoded files were sent via FTP to IBM VideoCharger™ servers at each lab.[6]

As their experience grew and the e-Seminar system improved, the roles reversed: the encoding was performed in real time and video tape was used as a safety backup. If the encoding was deemed good, the tape was used again for the next lecture.

By 2003 the interconnection between the Research Media Portal and the AutoAuditorium Systems had matured to the point where everything worked smoothly. The expectations of the users of the systems and the sponsors of talks were being met and most talks of consequence were recorded. That year the Audio/Video group made 65 talks produced using crews, only a few of which were recordings of talks in the AutoAuditorium rooms. During the same year, 233 AutoAuditorium videos were created, which is just shy of one every business day. The Research Media Portal library of recordings then contained over 600 presentations, the majority of which created in the AutoAuditorium rooms.

4. LONG TERM EXPERIENCE

Between the Distributed Auditorium and AutoAuditorium Systems, we have over 15 years of experience with business videos in research lab and business settings. While the following observations are mostly anecdotal, they point to some useful insights.

First and foremost, that not all recorded lectures are created equal, because not all lectures are created equal. The ones that were regarded as uninteresting by the live audiences generally were not viewed as recordings. Even a well-made video cannot rescue a badly given presentation.

On the other hand, less-than-perfect videos are still worth watching if the material is interesting. Some of our "locked-down" Distributed Auditorium videos were very tedious as television programs, but nonetheless valued by interested viewers because the material was presented well, the projection screen was always visible and the automatic audio mixing covered both the stage and the audience reasonably

[6]Streaming video is rarely sent to distant labs, because of the many timezones that separate them.

well. "I just could not get there that day. Thank goodness that recording was made."

The AutoAuditorium System does not require any attention from the people making presentations. It works best when people forget it is there and simply give their presentations as they normally would to their local audience.

The AutoAuditorium System design recognizes the capabilities and limitations of the technologies and each installation is customized to emphasize their strengths and accommodate their weaknesses. Similarly each installation is adjusted to the particular environment it operates in and to the most common uses we expect to see there. In short, a great deal of attention to detail is required before the system performs as a hands-off capability.

The AutoAuditorium Systems' customers use them mostly as a means to capture and share more information with more people, rather than as ways to reduce the costs of video productions. Sometimes their use is unscheduled, such as when a "little event" attracts more people than expected. More than once, the AutoAuditorium System has been turned on and the program routed to rooms down the hall where the overflow crowd is then able to watch in relative comfort.

5. AUTOAUDITORIUM PRODUCTION HEURISTICS

Modern automatically produced videos still fall short when compared to the best manually produced videos, but they often have advantages. Because they are automated, they follow some simple heuristics when deciding what to do next. But a tired, bored, uninterested or distracted operator will make mistakes, such as failing to show the screen because they did not notice when the projection changed. The automatic production heuristics are very unlikely to make that mistake.

And the people in the production crew may have other jobs. More than once we've heard of a person recording a lecture and being called away for another task. "What could I do? I just locked down a shot and walked away."

Automatically produced video can be superior to those that are hand-made. We've seen "professionally" produced videos with some glaring flaws:

- Some camera operators don't know when the screen is important and when the person is. We've seen programs with long shots of the back of a person who is clearly talking about what is on the screen, but we never see it. Because the speaker is moving a lot, pointing at the screen, the operator perceives the person as being "where the action is".

  The automation doesn't know when the screen is more important than the person either, but the AutoAuditorium heuristics know enough to show the screen every time it changes and to hold the screen shot much longer if the change affects most of the projected image.

- Shots of the screen are sometimes held too little time to read. The lack of motion on the screen is seen as a reason not to show it.

  Because of comments from our audiences, the AutoAuditorium heuristics tend to hold shots of the projection screen longer than we've seen in live-crew productions. Also, if the screen is unchanged for a long period of time, it is reshown periodically so the remote audiences can refresh their memories.

- Television production "wisdom" dictates that showing people watching a presentation is interesting, and changing camera angles and moving the camera constantly is "good television." But constant switching of camera angles and shots of the audience do not necessarily help the remote audiences understand the material being presented. The AutoAuditorium heuristics favor the projection screen and the people on stage far above other shots and camera angles.

  The AutoAuditorium heuristics only show the audience and other covering shots when the projection screen has not shown anything new in a while. Then other shots are added to the program to avoid just cycling between the person and the unchanging screen.[7]

Reasonably well made videos of talks, lectures and seminars fill a very real need. The promise of ubiquitous multimedia is that "place shouldn't matter. You can get what you want, when you want it, wherever you are." But that is only true if what you want exists inside the network. If it doesn't, the rest is irrelevant. The ability to capture lectures, talks and seminars automatically with reasonable fidelity is crucial to meeting the needs of people who must attend from another place or at another time.

6. RELATED DEVELOPMENTS

The idea of automatically capturing talks, lectures and seminars as multimedia has been researched for years. For instance:

- the Declarative Camera Control Language[4]
- the eClass project[3],
- the E-Seminar project[9],
- design issues of capturing collaboration[10]
- intelligent camera management[8]
- the Smart Classroom[11],
- the Virtual Director project[5]
- the Virtual Videography project[7]

7. CONCLUSIONS

The promise of ubiquitous multimedia is the possibility of having access to all *forms* of information almost anywhere. But it isn't the *form* of the information that really interests us; it is the *content* that demands our attention.

The AutoAuditorium System addresses the problem of making a common form of information sharing, namely a presentation to a group of people with projected visual aids, readily available for distribution over the multimedia networks of today and tomorrow. It does so by using existing technology (cameras, video mixers, microphones, audio mixers, computers and image processing software) in existing environments (classrooms, lecture halls and auditoriums) to capture the events that are frequently held there (presentations, talks and seminars).

Users of the AutoAuditorium Systems are already moving towards the day when most presentations are available at a

[7]It is possible to adjust the heuristics to put more variety and "production values" into the programs if the customer feels it makes them look more professional.

distant place or a distant time. As we have seen, when a customer creates and keeps an AutoAuditorium program every business day, it suggests that they have achieved the ease-of-use necessary to make these every-day-events into multimedia communication assets.

8. REFERENCES

[1] Bianchi, M. H. AutoAuditorium: a fully automatic, multi-camera system to televise auditorium presentations. In *Joint DARPA/NIST Smart Spaces Technology Workshop* (1998). www.AutoAuditorium.com/nist/autoaud.html.

[2] Bianchi, M. H. Autoauditorium system home page, 2004. www.AutoAuditorium.com.

[3] Brotherton, J. A. *Enriching Everyday Activities through the Automated Capture and Access of Live Experiences; eClass: Building, Observing and Understanding the Impact of Capture and Access in an Educational Domain*. PhD thesis, Georgia Institute of Technology, July 2001.

[4] Christianson, e. a. Declarative camera control for automatic cinematography. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (1996), pp. 148–155.

[5] E. Machnicki, L. R. Virtual director: Automating a webcast. In *Proceedings of SPIE Vol. 4673: Multimedia Computing and Networking* (2002), pp. 208–225.

[6] Kernmani, P. The IBM research media portal, December 2002. www.poly.edu/Podium/eef2002.cfm.

[7] Michael Gleicher, J. M. Towards virtual videography. In *ACM Multimedia 2000, Los Angeles, CA* (November 2000). www.cs.wisc.edu/graphics/Papers/Gleicher/Video/vv.pdf.

[8] Rui, Y., He, L., Gupta, A., and Liu, Q. Building an intelligent camera management system. In *Proceedings of ACM Multimedia 2001* (2001), pp. 2–11.

[9] Steinmetz, A. A. E-seminar lecture recording and distribution system: a bottom up approach from video to knowledge streaming. In *Proceedings of SPIE Volume 4312* (2001). www.spie.org/web/meetings/programs/pw01/confs/4312.html.

[10] Truong, K., Abowd, G., and Brotherton, J. Who, what, when, where, how: Design issues of capture and access applications. In *Ubicomp 2001 Conference Proceedings* (2001), ACM Press, pp. 209–224.

[11] Yuanchun Shi, e. a. The smart classroom: Merging technologies for seamless tele-education. In *IEEE Pervasive Computing Magazine* (April-June 2003).