AN OVERVIEW OF TECHNOLOGIES FOR E-MEETING AND E-LECTURE

Berna Erol
Ricoh California Research Center
2882 Sand Hill Rd., Menlo Park, CA 94025, USA
berna_erol@rii.ricoh.com

Ying Li
IBM T. J. Watson Research Center
19 Skyline Dr., Hawthorne, NY 10532, USA
yingli@us.ibm.com

ABSTRACT

Over the past few years, with the rapid adoption of broadband communication and advances in multimedia content capture and delivery, web-based meetings and lectures, also referred to as *e-meeting* and *e-lecture*, have become popular among businesses and academic institutions because of their cost savings and capabilities in providing self-paced education and convenient content access and retrieval. In fact, the technological achievements in capture, analysis, access, and delivery of e-meeting and e-lecture media have already resulted in several working systems that are currently of regular usage. This paper gives an overview of existing work as well as state-of-the-art in these two research areas which are bound to affect the way we teach, learn, and collaborate.

1. INTRODUCTION

Web-based collaboration and education systems are quickly emerging as viable alternatives to traditional face-to-face meetings and lectures. Today, with the advances in multimedia technologies, current e-meeting and e-lecture systems have gone far beyond the stage of plain audio/video capture and display. Instead, more sophisticated technologies involving the capture, analysis, recognition, and structuralization of e-contents are undoubtedly creating new ways to access, search and repurpose both real-time and archived meeting/lecture data.

Much research effort has been directed to the e-meeting and e-lecture areas in recent years. Innovative audio and video capture technologies are being developed which not only capture panoramic video, lecture notes, presentation materials, and whiteboard discussions, and subsequently synchronize them into a well-presented multimedia stream, but also record complex interactions that frequently take place in meeting rooms and lecture halls. Moreover, recent research on intelligent audio/video content analysis, recognition and understanding, as well as novel interfaces that facilitates convenient content visualization and retrieval, has achieved some promising results. Finally, efficient content transmission and delivery systems, especially designed to accommodate e-meeting and e-lecture data, are being actively studied, for which we can expect to see more concrete results in a near future. In a word, all these ongoing research efforts, once combined together, will undoubtedly provide us various much enhanced means for accessing and utilizing e-meeting and e-lecture content.

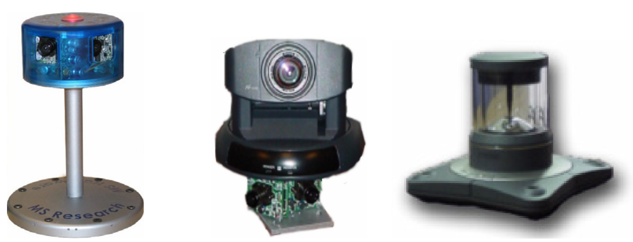
In the rest of this paper, we offer a general overview of existing technologies in both e-meeting and e-lecture areas. More specifically, various audio, video and other media capture technologies are introduced in Section 2. Section 3 investigates related work on the analysis, recognition, synchronization, access, retrieval, and summarization of meeting and lecture videos. Some real-time and on demand e-meeting/e-lecture delivery schemes are presented in Section 4. Finally, we briefly study the user experience with existing e-meeting and e-lecture applications, and discuss the future and challenges that we face in this field in Section 5.

2. CAPTURE SYSTEMS

Meeting and lecture rooms are environments where many complex communications take place via visual and audio interactions, note taking, presentations, document and common workspace sharing, as well as whiteboard discussions. Many researchers have developed specialized technologies that are optimized for capturing these complex communications and artifacts of meetings and lectures. In this section, we review some of these capture technologies.

Audio and video capture in a meeting, conference, or lecture is particularly challenging because they usually take place in large rooms with many participants. In addition, people occasionally change positions and the focus of attention at a given time may also change easily. To address these issues, many different approaches have been proposed including panoramic video capture, use of microphone arrays, capture with intelligent pan-zoom-tilt (PTZ) cameras, and use of multi-camera systems. Early work on automatic video capture includes the Cornell Lecture Browser [1] that takes the content captured by two cameras and automatically synchronizes and edits it. Also, the eClass project (previously known as Classroom 2000) [2][3] creates lecture videos by combining various data sources collected from whiteboards and cameras. The AutoAuditorium system [4] supports speaker motion tracking with a PTZ camera and proposes a virtual director that automatically selects the view from one of four cameras. In [5], Rui et al. propose an automated director that applies professional videographers' production rules, yielding more visually engaging lecture videos. In [6], Kameda et al. combine visual analysis, sound direction, and outputs of ultrasonic location trackers to determine view selection from 8 PTZ cameras in a lecture room. In the FlySpec system [7], where PTZ and omni-directional cameras are used together, an algorithm is presented that combines inputs from multiple human camera operators with automatic camera controls for better camera management.

Panoramic video capture systems have been shown to be very useful for capturing all the events in a room and providing a user-centric viewpoint [8]-[13]. Examples of such systems are shown in Figure 1. Some of these systems use post-hoc analysis to automatically generate a user-oriented view for playback. In [12], the user-oriented view is automatically determined based on speaker motion. Another method uses microphone arrays for audio capture, computing speaker direction with sound source localization and determining the view point [8][11]. Capturing audio with omni- and uni-directional microphone arrays works very efficiently in small lecture or meeting rooms. For audio capture in larger rooms, a more general approach is to use a wireless microphone for the speaker or instructor, and handheld or spatially distributed microphones for the audience [4][6].



(a) (b) (c)

Figure 1. Examples of meeting/lecture video capture prototypes (a) Ringcam* [10] (b) FlySpec** [7] (c) Portable Meeting Recorder*** [8].

Courtesy of *Microsoft Research, **FX Palo Alto Lab., ***Ricoh California Research Center.

A variety of approaches have been proposed for capturing presentation slides. Some systems capture PowerPoint slides by requiring the presenter to submit presentation slides beforehand or install a plug-in on their computer [6][2]. The plug-in then communicates the slide numbers with a server. Similarly, people have used screen-capture software on the presenter's machine [20] for the same purpose. Comparatively, a less intrusive approach is to dedicate a separate video or digital camera to the presentation slides and saves the output into an individual video stream [4][5][14][15]. Another slide capture approach that achieves the same goal directly taps into the VGA output of the presenter's machine or the projector's output and stores the output in a separate stream [8][16]. On the other hand, slide recording systems are also frequently used for capturing documents and web pages, since they can record everything that is displayed on the presenter's screen.

Personal notes provide excellent access points to meeting and lecture recordings. Consequently, many recording systems support note taking. The Audio Notebook [17], one of the pioneering work in this area, synchronizes handwritten meeting notes with recoded audio. In NotePals [18] and StuPad [19], users compose handwritten notes on PDAs and upload them onto a server where they are cross-indexed with recorded lecture. Other work, such as LiteMinutes [20], let users type notes from a web page and place them in a personal or shared space associated with other captured multimedia data. A Tablet PC and special pads for capturing handwriting are also used in systems such as NoteLook [21], ZenPad in E-class [2], and those proposed in [22][23]. These systems automatically record notes that a user writes on slides during presentations and lectures. In addition, systems such as Minuteaid [24] can link meeting and lecture recordings to notes composed using applications such as Microsoft Word. To support group collaboration, Microsoft's MRAS system [25] provides a discussion forum that augments the recorded presentations.

Whiteboard and blackboard discussions play an important role in both meetings and lectures. Commercial electronic whiteboards that record time-stamped pen coordinates have been successfully employed by many meeting and lecture recording systems [2][26]. Another solution is to take continuous images using digital cameras, then stitch them into one high-resolution image [27][28]. In [29], a whiteboard is captured with a pan-and-tilt video camera and frames are stitched to obtain a high-resolution image. Another video-based whiteboard capture system [30] automatically identifies and zooms into the regions of interest on the board. In [31], Wienecke et al. propose to capture the whiteboard with a video camera, followed by automatic handwriting recognition.

In addition to these basic capture systems, other less traditional meeting capture devices include accelerometers that detect different postures of attendees [32], and wireless meeting and presentation book-marking devices [33].

3. CONTENT ANALYSIS AND RETRIEVAL

To facilitate the access and searching of e-meeting and e-lecture content, a sophisticated analysis and understanding of this content is a prerequisite. So far, many research efforts have been reported along this direction. While some of them directly apply generic video analysis technologies to this domain, others have taken its special characteristics into consideration. Below we give a brief survey of existing work in the second group.

Most research on visual analysis of meeting/lecture videos falls into the following five categories: speaker tracking for PTZ camera control [4][5][6], human motion analysis for gesture recognition and activity classification [34][35][36], meeting participant segmentation [37], face detection and recognition for attendee identification [38][39][8], and gaze detection for focus-of-attention determination [40].

There has also been significant work on employing audio cues for content understanding. The ICSI project at UC Berkeley distributes a large corpus of annotated meeting audio recordings for researching into this area [41][42]. A scene detection and classification scheme is proposed in [43] and [44] which identifies discussion scenes in education videos, and subsequently categorizes them into either 2-speaker discussions (e.g. Q&A) or multi-speaker discussions (e.g. classroom discussion). On the other hand, the sound source localization technique has been used by many projects to locate speakers using microphone arrays [8][10][45].

The joint audio and visual analysis has been successfully employed for high-level event detection and structuring of meeting and lecture videos. In the IDIAP project [46], McCowan et al. model the audio-visual interaction between participants to recognize high-level events such as presentations, general discussions, and consensus. To identify video segments with meaningful discussion topics, Mahmood and Srinivasan propose a method to jointly analyze the contents of slide and audio recordings [47][48]. In a similar manner, Ngo et al. [49] extract discussion topics from electronic slides via video text analysis and speech recognition. Instead, Haubold and Kender classify video keyframes into six distinct categories including blackboard, class, computer, illustration, podium and sheet [50]. In [51], Dorai et al. propose to understand the narrative structure of educational videos by distinguishing frames containing a narrator, slide, web page and whiteboard using color and textual information. Targeted at identifying sections with topic changes in instructional videos, Phung et al. propose a density function based on video shot changes in [52]. They also propose to recognize narrative structures, such as discussions and on-screen narration scenes, using a set of audiovisual features extracted from lecture videos [53]. In [54], Jain et al. present a way to assign a higher-level structure to meetings via semiautomatic tagging, so that the meeting viewing experience could be improved. Some other work on semantic topic segmentation of lecture recordings can be found in [55][56].

With the semantics extracted from analysis modules, the video content can then be effectively accessed, browsed and searched. In [57], Jaimes et al. conduct a study in which 519 participants are surveyed to determine what artifacts people access after meetings. The conclusion is that documents, personal notes and whiteboard images are the most frequently accessed, while speaker locations and identifications are the most frequently remembered. A prototype system presented in [58] uses explicit artifacts such as notes and derived artifacts like speaker IDs, as pointers into meeting records. It has also surveyed various ways of creating indices into meeting records and introduced the notion of creating indices based on user interaction with domain-specific artifacts. Gross et al. present a multimodal approach to create meeting records based on speech recognition, face detection and people tracking in CMU's Meeting Room System [59]. A similar system which attempts to index, and subsequently allows users to browse and search video conference archives with automatically identified participants and titles is described in [60]. Moreover, an interface called MuVIE Client is described in [8], which allows browsing of meeting recordings based on speaker transition, audio and visual activity, slide, whiteboard, note and transcript content.

Synchronization of different media streams according to their content has also been a topic of interest since it opens new ways for cross-media indexing. Liu et al. propose a system where original electronic slides are synchronized with a video recording [61] and presentation videos are enhanced by using the original slides in the content area. In the e-meeting system reported in [62], images taken by a digital camera during presentations are synchronized with those of JPEG images captured by a presentation recorder based on image analysis. In [63], handwritten annotations are synchronized and registered with projected slide images for a better playback experience. Various other work on synchronization of different media channels for content access and retrieval can be found in [1][64][65][66].

Summarization has been an efficient and effective way to facilitate quick overview of presentation and meeting recordings [67]. Consequently, quite a few investigators have pursued along this direction. For instance, Waibel et al. [68] generate keyword-based meeting summaries based on relevance ranking and topic segmentation. In [69], He et al. describe a method to summarize audio-video presentations using slide transitions and pitch activity cues. In [70], Erol et al. propose a multimodal summarization technique that combines visual and audio activity with TF-IDF (Term Frequency-Inverse Document Frequency) analysis of speech transcripts. Focusing more on lecture videos with handwritten slides and blackboard presentations, Liu et al. present a system that selects optimal keyframes and generates a mosaic summary by detecting and stitching related content regions together [71][72].

4. CONTENT DELIVERY

Recently, on-demand streaming of meetings and lectures, as well as systems that support interactive real-time communication, have become popular due to rapid advancement of video compression and transmission technologies. Consequently, many research efforts have emerged in this area. Some of this work is reviewed below.

Most e-meeting and e-lecture systems support on-demand streaming of lectures with synchronized voice, video, and slides. In addition to on-demand and real-time streaming, some existing research and commercial e-lecture and e-meeting systems also support *real-time and interactive* sharing of slides and whiteboards, as well as video and web conferencing [75][76] for both remote meeting and distributed education settings. Use of real-time video conferencing is particularly challenging when there are multiple clients concurrently using the system. A multimedia distance learning system is reported in [77][78], which describes a real-time interactive virtual classroom at the University of Washington. This system allows a remote participant to not only receive a live class feed, but also interact with the class by asking questions with real-time audio and video. In [79], a videoconference system called Virtual Auditorium is developed which supports dialog-based distance learning. Specifically, the instructor can monitor students on a tiled display and maintain eye contact with any student as shown in Figure 2. An extension of this system to multiparty videoconferencing is provided in [80][81], which automatically determines whether students are speaking, making gestures, or twisting in their seats.

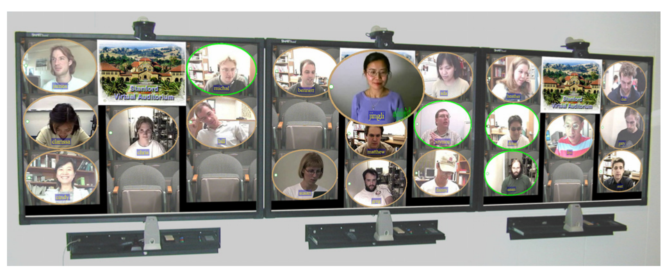


Figure 2. Virtual auditorium system [79].

Courtesy of Milton Chen, VSee Lab.

Recently, mobile learning has become gradually popular with the ubiquity of wireless devices that have multimedia playback capabilities. Nevertheless, mobile transmission of e-meeting and e-lecture data is challenging as it involves transmission of rich media content with limited bandwidth. Addressing this problem, an adaptive transmission scheme for lecture videos is presented in [82][83] for mobile applications. Specifically, a content-based analysis method and a buffer-based model detect salient content regions, select optimal keyframes, and transmit compressed keyframes using an adaptive feedback control scheme. This accommodates dynamically changing bandwidths in wireless environments. Readers can find various other mobile learning systems in [84][85].

5. LESSONS LEARNED AND OUTLOOK

So far, many commercial and experimental e-meeting and e-lecture recording systems are in regular use, including IBM Research's e-Seminar [15] and Georgia Tech's eClass [3]. Other examples can be found in [3][4][8][86]-[91]. Consequently, there has been considerable work on evaluating the effectiveness and usefulness of these systems. For instance, a study carried out at Microsoft Research shows that most of evaluation participants agree that watching recordings of missed meetings is very helpful [10]. Georgia Tech has used its eClass system [2][3] to record classroom lectures since 1997, and it is reported that lecture recordings are frequently accessed by students, especially right before exams [3]. A study of Berkley's BIBS lecture broadcasting system [87] also reaches the similar conclusion.

There is no question that e-meeting and e-learning systems are bound to change the way we teach, learn, conduct business, and share information. On one hand, students can now take fewer notes during the class and focus more on the lecture [3]; likewise, corporate employees can attend fewer presentations without sacrificing the learning opportunity [15]. On the other hand, lecturers need take extra effort to make sure that their presentation styles are suitable for on-demand viewing [92].

Before we conclude this paper, we would like to suggest readers some conferences or workshops that address various aspects of e-meeting, e-lecture or e-learning research. It is our belief that, with e-meeting, e-lecture and e-learning systems rapidly gaining popularity, researchers in this field could benefit from these conferences and workshops.

- ICME 2005 Special session on "Technologies for E-meeting and E-lecture", 2005
- NIST ICASSP 2004 Meeting Recognition Workshop, 2004
- The International Conference on Web-based Learning (ICWL)
- ACM Conference on Computer Supported Cooperative Work (CSCW)
- IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE)
- Frontiers in Education Conference
- IEEE International Conference on Advanced Learning Technologies (ICALT)
- IEEE International Workshop on Multimedia Technologies in E-Learning and Collaboration (WOMTEC)
- APRU Distance learning and the Internet Conference (DLI)
- MLEARN: Learning with mobile devices
- European Workshop on Mobile and Contextual Learning
- International Workshop on Multimedia Distance Education Systems and Technologies (MDEST)
- International Conference on Intelligent Multimedia and Distance Education (ICCIMADE)

The fact that there are many e-lecture and e-meeting systems currently in regular usage demonstrates the technological achievements in various areas related with e-media such as capture, access, and delivery. Nevertheless, this is not to say that challenges do not remain. In fact, creative solutions are still in great need regarding the content analysis, recognition and structuring, as well as easy authoring, sharing and re-use of multimodal media that support learning and decision-making. Yet it is our belief that, with people's requirement for self-directed education and flexible meeting and learning schedules keeps increasing, e-meetings and e-lectures will inevitably become viable complements to traditional face-to-face meetings and lectures.

6. ACKNOWLEDGEMENTS

Authors would like to thank Dar-Shyang Lee and Jonathan J. Hull for their valuable feedback in preparation of this tutorial paper.

7. REFERENCES

[1] Mukhopadhyay, S., and Smith, B., "Passive capture and structuring of lectures", ACM Multimedia pp. 477-487, 1999.

[2] J. A. Brotherton, J. R. Bhalodia, and G. D. Abowd, "Automated capture, integration, and visualization of multiple media streams," *IEEE Multimedia*, 1998.

[3] J. A. Brotherton and G. D. Abowd, "Lessons learned from eClass: Assessing automated capture and access in the classroom," *ACM Trans. on CHI*, pp. 121 - 155, 2004.

[4] M. Bianchi, "AutoAuditorium: A fully automatic, multi-camera system to televise auditorium presentations," *Joint DARPA/NIST Smart Spaces Tech. Workshop*, 1998.

[5] Y. Rui, A. Gupta, J. Grudin, and L. He, "Automating lecture capture and broadcast: Technology and videography," *ACM Multimedia Systems Journal*, 2004.

[6] Y. Kameda, S. Nishiguchi, and M. Minoh, "CARMUL: Concurrent automatic recording for multimedia lecture," *ICME*, 2003.

[7] Q. Liu, D. Kimber, J. Foote, L. Wylcox, and J. Boreczky, "FLYSPEC: A multi-user video camera system with hybrid human and automatic control," *ACM Multimedia*, 2002.

[8] D. S. Lee, B. Erol, J. Graham, J. J. Hull, and N. Murata, "Portable meeting recorder," *ACM Multimedia*, pp. 493-502, 2002.

[9] J. Foote and D. Kimber, "FlyCam: Practical panoramic video and automatic camera control," *ICME*, vol.3, pp.1419-1422, 2000.

[10] R. Cutler, Y. Rui, A. Gupta, J. J. Cadiz, and I. Tashev, "Distributed meetings: A meeting capture and broadcasting system," *ACM Multimedia*, 2002.

[11] Y. Rui, A. Gupta, and J. Cadiz, "Viewing meetings captured by an omni-directional camera," *ACM CHI*, pp. 450-457, Seattle, March 31- April 4, 2001.

[12] X. Sun, J. Foote, D. Kimber, and B. Manjunath, "Panoramic video capturing and compressed domain virtual camera control," *ACM Multimedia*, pp. 229-238, 2001.

[13] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, "The Diver project: Interactive digital video

repurposing," *IEEE Multimedia Magazine*, vol. 11, no. 1, pp. 54-61, 2004.

[14] W. H. Leung, T. Chen, F. Hendriks, X. Wang, and Z. Shae, "eMeeting: A multimedia application for interactive meeting and seminar," *IEEE Globecom*, Vol. 3, pp.2994-2998, 2002.

[15] A. Steinmetz, “Media and distance: A learning experience,” *IEEE Multimedia*, pp. 8-10, 2001.

[16] P. Chiu, A. Kapuskar, S. Reitmeier, and L. Wilcox, “Room with a rear view: Meeting capture in a multimedia conference room,” *IEEE Multimedia Magazine*, pp. 48-54, vol. 7, no. 4, 2000.

[17] E. Stifelman, B. Arons, and C. Schmandt, “The audio notebook: Paper and pen interaction with structured speech,” *ACM CHI*, pp. 182-189, 2001.

[18] R.C. Davis, J.A. Landay, and et.al, “NotePals: Lightweight note sharing by the group, for the group,” *CHI*, pp.338-345, 1999.

[19] K.N. Truong and G.D. Abowd, “StuPad: Integrating student notes with class lectures,” *CHI*, pp.208-209, 1999.

[20] P. Chiu, J. Boreczky, A. Girgensohn, and D. Kimber, “LiteMinutes: An Internet-based system for multimedia meeting minutes,” *World Wide Web*, pp.140-149, 2001.

[21] P. Chiu, A. Kapuskar, S. Reitmeier and L. Wilcox, “NoteLook: Taking notes in meetings with digital video and ink,” *ACM Multimedia*, 1999.

[22] R. Anderson, R. Anderson, C. Hoyer, B. Simon, and et al., “Lecture presentation from the Tablet PC,” *Workshop on Advanced Collaborative Environments*, 2003.

[23] N. Joukov, and T. Chiue, "LECTERN II: A multimedia lecture capturing and editing system," *ICME*, 2003.

[24] D. Lee, J. J. Hull, B. Erol, and J. Graham, “MinuteAid: Multimedia note-taking in an intelligent meeting room,” *ICME*, 2004.

[25] D. Bargeron, J. Grudin, A. Gupta, E. Sanocki, F. Li, and S. LeeTiernan, “Asynchronous collaboration around multimedia applied to on-demand education,” *HICSS*, 2001.

[26] G. Friedland, L. Knipping, R. Rojas, and E. Tapia, "Teaching with an intelligent electronic chalkboard," *ACM Workshop on Effective Telepresence*, 2004.

[27] Z. Zhang and L. He, “Notetaking with a camera: Whiteboard scanning and image enhancement,” *ICASSP*, 2004.

[28] L. He, Z. Liu, and Z. Zhang, "Why take notes? Use the whiteboard capture system," *ICASSP*, 2003.

[29] E. Saund, “Image mosaicing and a diagrammatic user interface for an office whiteboard scanner,” Technical Report, Xerox Palo Alto Research Center, 1999.

[30] M. Onishi, M. Izumi, and K. Fukunaga, “Production of video image by computer controlled camera operation based on distribution of spatiotemporal mutual information,” *ICPR*, pp. 102-105, 2000.

[31] M. Wienecke, G. Fink, and G. Sagerer, "Towards automatic video-based whiteboard reading,” *ICDAR*, 2003.

[32] N. Kern, B. Schiele, and et al., "Wearable sensing to annotate meeting recordings," *The 6th Int. Symposium on Wearable Computing (ISWC)*, Seattle, 2002.

[33] R. Nair, “Calculation of an aggregated level of interest function for recorded events,” *ACM Multimedia*, 2004.

[34] T. Nishimura, H. Yabe, and R. Oka, ``Indexing of human motion at meeting room by analyzing time-varying images of omni-directional camera,’’ *Asia Conference on Computer Vision*, vol. 1, pp.1-4, 2000.

[35] A. Hakeem and M. Shah, “Ontology and taxonomy collaborated framework for meeting classification,” *ICPR*, 2004.

[36] I. Mikic, K. Huang, and M. Trivedi, “Activity monitoring and summarization for an intelligent meeting room,” *IEEE Workshop on Human Motion*, 2000.

[37] D. Lee, B. Erol, and J. Hull, "Segmenting people in panoramic meeting videos using mixture background and object models," *IEEE Pacific-Rim Conference on Multimedia*, 2002.

[38] R. Gross, J. Yang, and A. Waibel, “Face recognition in a meeting room,” *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2000.

[39] J. Yang, X. Zhu, R. Gross, J. Kominek, Y. Pan, and A. Waibel, “Multimodal people ID for a multimedia meeting browser,” *ACM Multimedia*, 1999.

[40] R. Stiefelhagen, J. Yang, and A. Waibel, "Estimating focus of attention based on gaze and sound," *ACM Workshop on Perceptive User Interfaces*, pp. 1-9, 2001.

[41] A. Janin, J. Ang, S. Bhagat, R. Dhillon, and et al , “The ICSI meeting project: Resources and research,” *NIST ICASSP 2004 Meeting Recognition Workshop*, May 2004.

[42] A. Stolcke, C. Wooters, N. Mirghafori, T. Pirinen, and et al, “Progress in meeting recognition: The ICSI-SRI-UW Spring 2004 evaluation system,” *NIST ICASSP 2004 Meeting Recognition Workshop*, May 2004.

[43] Y. Li and C. Dorai, “Detecting discussion scenes in instructional videos,” *ICME*, 2004.

[44] Y. Li and C. Dorai, “Analyzing discussion scene contents in instructional videos,” *ACM Multimedia*, 2004.

[45] S. Nishiguchi, K. Higashi, Y. Kameda, and M. Minoh, “A sensor-fusion method of detecting a speaking student,” *ICME*, 2003.

[46] I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang, “Automatic analysis of multimodal group actions in meetings,” *IEEE Transactions on PAMI*, Vol. 27, No. 3, Mar. 2005.

[47] T. Mahmood and S. Srinivasan, “Detecting topical events in digital video,” *ACM Multimedia*, pp. 85-94, November 2000.

[48] T. Mahmood, “Indexing for topics in videos using foils,” *CVPR*, 2000.

[49] C. Ngo, F. Wang, and T. Pong, “Structuring lecture videos for distance learning applications,” *ISMSE*, 2003.

[50] A. Haubold and J. Kender, “Analysis and interface for instructional video,” *ICME*, 2003.

[51] C. Dorai, V. Oria, and V. Neelavalli, "Structuralizing educational videos based on presentation content," *ICIP*, 2003.

[52] D. Phung, S. Venkatesh, and C. Dorai, "High level segmentation of instructional videos based on content density," *ACM Multimedia*, 2002.

[53] D. Phung, C. Dorai, and S. Venkatesh, "Narrative structure analysis with education and training videos for E-learning," *ICPR*, 2002.

[54] R. Jain, P. Kim, and Z. Li, "Experiential meeting system," *ACM SIGMM Workshop on Experiential Telepresence*, pp. 1-12, 2003

[55] A. Stewart, P. Wolf, and M. Hemmje, "Media and metadata management for capture and access systems in electronic lecturing environments," *ICME*, 2003.

[56] E. Altman, Y. Chen, and W. Low, "Semantic exploration of lecture videos," *ACM Multimedia*, pp. 416-417, 2002.

[57] A. Jaimes, K. Omura, T. Nagamine, and K. Hirata, "Memory cues for meeting video retrieval," *ACM CARPE Workshop*, 2004.

[58] W. Geyer, H. Richter, and G. Abowd, "Making multimedia meeting records more meaningful," *ICME*, 2003

[59] R. Gross, M. Bett, H. Yu, X. Zhu, Y. Pan, J. Yang, and A. Waibel, "Towards a multimodal meeting record," *ICME*, pp. 1593-1596, New York, 2000.

[60] J. Song, M. Lyu, J. Hwang, and M. Cai, "PVCAIS: A personal videoconference archive indexing system," *ICME*, 2003.

[61] T. Liu, R. Hjelsvold, and J. R. Kender, "Analysis and enhancement of videos of electronic slide presentations," *ICME*, 2002.

[62] B. Erol, J. J. Hull, and D. Lee, "Linking multimedia presentations with their symbolic source documents: Algorithm and applications," *ACM Multimedia*, 2003.

[63] W. Li, H. Tang, and Z. Zhu, "Automated registration of high resolution images from slide presentation and whiteboard handwriting via a video camera," *CVPRW*, Vol. 11, pp. 168-172, 2004

[64] D. Franklin, S. Bradshaw, and K. Hammond, "Jabberwocky: You don't have to be a rocket scientist to change slides for a hydrogen combustion lecture," *Intelligent User Interfaces*, pp. 98-105, 2000.

[65] P. Chiu, J. Foote, A. Girgensohn, and J. Boreczky, "Automatically linking multimedia meeting documents by image matching," *Proceedings of Hypertext'00*, ACM Press, pp. 244-245, 2000.

[66] Y. Chen and W. Heng, "Automatic synchronization of speech transcript and slides in presentation," *ISCAS*, 2003.

[67] L. He, E. Sanocki, A. Gupta, and J. Grudin, "Comparing presentation summaries: Slides vs. reading vs. listening," *CHI,* pp. 177-184, 2000.

[68] A. Waibel, M. Bett, and et al., "Advances in automatic meeting record creation and access," *ICASSP*, 2001.

[69] L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-summarization of audio-video presentations," *ACM Multimedia*, 1999.

[70] B. Erol, D. Lee, J. Hull, "Multimodal summarization of meeting recordings," *ICME*, 2003.

[71] T. Liu and J. R. Kender, "Rule-based semantic summarization of instructional videos," *ICIP*, 2002.

[72] T. Liu and J. R. Kender, "Semantic mosaic for indexing and compressing instructional videos," *ICIP*, 2003.

[73] L. Rowe and J. Gonzlez, "BMRC Lecture Browser Demo," at http://bmrc.berkeley.edu/frame/projects/lb/index.html.

[74] C. Dorai, P. Kermani, and A. Stewart, "Elm-n: e-learning media navigator," *ACM Multimedia*, pp 634–635, 2001.

[75] Webex Web conferencing , http://www.webex.com/

[76] Polycom video conferencing systems, http://www.polycom.com/

[77] J. Hwang, S. Deshpande, and M. Sun, "A virtual classroom for real-time interactive distance learning," *ISCAS*, vol. 3, pp. 611-614, 1998.

[78] S. Deshpande and J. Hwang, "A real-time interactive virtual classroom – Multimedia distance learning system," *IEEE Trans. on Multimedia*, vol. 3, no. 4, 2001.

[79] M. Chen, "Design of a virtual auditorium," *ACM Multimedia*, 2001.

[80] M. Chen, "Achieving effective floor control with a low-bandwidth gesture-sensitive videoconferencing system," *ACM Multimedia*, 2002.

[81] M. Chen, "Visualizing the pulse of a classroom," *ACM Multimedia*, 2003.

[82] T. Liu and C. Choudary, "Real-time content analysis and adaptive transmission of lecture videos for mobile applications," *ACM Multimedia*, 2004.

[83] T. Liu and C. Choudary, "Content-aware streaming of lecture videos over wireless networks," *IEEE Int. Workshop on Wireless Multimedia*, 2004.

[84] C. Leung and Y. Chan, "Mobile learning: A new paradigm in electronic learning," *Int. Conference on Advanced Learning Technologies*, pp. 76-80, 2003.

[85] A. Fong, and S. Hui, "Low-bandwidth Internet streaming of multimedia lectures," *Engineering Science and Education Journal*, pp. 212-218, 2001.

[86] S. Rosenschein, "Quindi meeting companion: A personal meeting-capture tool," *ACM CARPE Workshop*, 2004.

[87] L.A. Rowe, and et.al. "BIBS: A lecture webcasting system," *Technical Report*, Berkeley Multimedia Research Center, U.C. Berkeley, May 2001.

[88] R. Müller and T. Ottmann, "The 'authoring on the fly' system for automated recording and replay of (Tele)presentations," *ACM/Springer Multimedia Systems Journal*, Vol. 8, No. 3, pp. 158-176, 2000.

[89] A. Waibel, and et. al., "SmaRT: The smart meeting room task at ISL," *ICASSP*, 2003.

[90] P. Chiu, A. Kapuskar, and L. Wilcox, "Meeting capture in a media enriched conference room," *The 2nd International Workshop on Cooperative Buildings*, pp.79-88, 1999.

[91] Multi-university Research Laboratory, murl.microsoft.com

[92] L. He, J. Grudin, and A. Gupta, "Designing presentations for on-demand viewing," A*CM 2000 Conf. on Computer Supported Cooperative Work*, Dec. 2000.