Evaluation of Potential Effectiveness of Desktop Remote Video Conferencing for Interactive Seminars

Engr. Faryal Zia

Abstract

This research paper discusses various aspects of desktop videoconferencing and conducts an experiment utilizing this technology to deliver a weekly seminar class to participants distributed throughout the state of North Carolina as the remote audience. The main purpose of this research was to investigate the current state of desktop video conferencing technology and evaluate the potential effectiveness of this technology for delivering interactive seminars to a remote audience.

Introduction

Enabling technologies for desktop video conferencing that are discussed in this research work include the compression of audio and video data and the transmission of this data over various communication channels. Features of currently available video conferencing systems and emerging interoperability standards are addressed. Non-technical aspects of desktop video conferencing systems such as the value added by audio, video and integrated computer application and the importance of usable user interfaces are evaluated. To demonstrate the use of desktop videoconferencing for distance learning, a weekly seminar class was broadcast, using Internet MBone, to remote participants located throughout the state of North Carolina. Delivery of this demonstration project and observations about this project are evaluated in detail. This demonstration project is judged whether it is a successful demonstration of the potential application of desktop videoconferencing for distance learning or not. Based on the results, the suggestions / recommendations for future projects of this nature are made.

Room videoconferencing has been used for some time as a mean to deliver interactive classes to geographically distributed audiences. One example of the successful use of room videoconferencing for class delivery is the North Carolina Research and Education Network (NC-REN). Tele-classrooms and conference rooms across the state of North Carolina are linked together through NC-REN's private microwave facilities. Video communications Services at North Carolina State University (NCSU) and similar facilities throughout the state produce programs that can be sent to any of the 19 universities, medical schools, and research organizations that are connected via NC-REN.

Advances in computer technology such as faster processors and better data compression schemes have made it possible to integrate audio and video data into the computing environment. A new type of videoconferencing, desktop videoconferencing, has become possible. Unlike room videoconferencing, which requires specially equipped rooms with expensive hardware, desktop videoconferencing can be achieved by adding software and hardware to standard desktop computers.

In accordance with the goals of Deliverable Team 5 (DT-5) of the Southeastern University and College Coalition for Engineering Education (SUCCEED), the main objective of this research project was to implement a demonstration of electronic connectivity, using desktop videoconferencing, for educational delivery and interaction among NCSU and other SUCCEED institutions. To achieve this objective, existing technology was surveyed, infrastructure was put into place, and a demonstration project was performed and evaluated.

Technological Components for Experiment

Audio and video must be captured from their analog form and stored digitally to be manipulated by the computer. If the data is uncompressed, it would require massive amounts of bandwidth to transmit; therefore the data is compressed before it is sent over communication channels. All this must happen in real-time to facilitate communication and interaction.

• Audio Component for Experiment

The frequency of sound waves is measured in Hertz (Hz), meaning cycles per second. The human ear can typically perceive frequencies between 20 Hz to 20 kHz. Human voice can typically produce sound frequencies between 40 Hz to 4 kHz. These limits are important factors to remember when discussing digital audio encoding. Desktop videoconferencing systems are typically designed to handle speech quality audio which encompasses a much smaller range of frequencies than the range perceptible to humans. Audio is delivered to computer equipment through various types of connectors. Digital audio data is usually described using three parameters: sampling rate, bits per sample and number of channels.

• Audio Sampling used in Experiment

An audio signal has amplitude values that continuously vary with time. To encode this signal digitally, the amplitude value of the signal is measured at regular intervals. This is called sampling. According to the Nyquist theory of signal processing, to faithfully represent a signal of a certain frequency, the sampling rate must be at least twice that of the highest frequency present in the signal. Using this theory, sampling is lossless since the original signal can be reconstructed based on the samples. To avoid aliasing distortion, the signal is low-pass filtered to remove any high frequencies, that cannot be represented by the sampling rate.

• Audio Quantizing for the Experiment

Sampled values representing the amplitude of the signal at the sample time are quantized into a discrete number of levels. The number of levels depends on how many bits are used to store the sample value. For digital audio, this precision usually ranges from 8 bits per sample (256 levels) to 16 bits per sample (65536 levels). Quantization includes error into the data because no matter how many bits of precision are used, it is impossible to represent an infinite number of amplitude values with a finite number of increments. Uniform Pulse Code Modulation (PCM) encoding is an encoding method where the quantizer values are uniformly spaced.

• Digital Audio Compression Techniques

Typically these techniques can achieve real-time compression and decompression in software or inexpensive hardware. Some techniques apply to general audio signals and some are designed specifically for speech signals.

i. Mu-law and A-law PCM

The Mu-law and A-law PCM encoding methods are formally specified in the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Recommendation G.711, "Pulse Code Modulation (PCM) of voice frequencies". The Mu-law PCM encoding format is common in North America and Japan for digital telephony with the Integrated Services Digital Network (ISDN). The A-law PCM encoding format is common with ISDN in other countries.

ii. ADPCM

The Adaptive Differential Pulse Code Modulation (ADPCM) encoding method computes the difference between each sample and its predicted value and encodes the difference. Fewer bits (typically 4) are needed to encode the difference than the complete sample value.

iii. LPC and CELP

These encoding methods are designed typically for speech. A Linear Predictive Coding (LPC) encoder fits speech signals to a simple analytic model of the vocal tract. The best-fit parameters are transmitted and used by the decoder to generate synthetic speech that is similar to the original. Groupe Speciale Mobile (GSM) encoding uses a variation of LPC called Regular Pulse Excited - LPC with a long term predictor loop.; A Code Excited Linear Prediction (CELP) encoder does the same vocal tract modeling as an LPC encoder. In addition, it computes the error between the input speech data and the model and transmits the model parameters and a representation of the errors.

• Video

Video is a sequence of still images. When presented at a high enough rates, the sequence of images (frames) gives the illusion of fluid motion. Desktop videoconferencing uses video as an input. This video may come from a camera, VCR or other video device. An analog video signal must be encoded in the digital form so that it can be manipulated by a computer.

• Video Formats

There are two widely used formats for analog video: NTSC and PAL. National Television Standards Committee (NTSC) format is used in the America and Japan. This standard was approved in 1953 by the Federal Communications Committee (FCC) for commercial broadcasting. NTSC format has a resolution of 525 lines per frame and 60 interlaced frames per second. With interlacing, two fields make up a complete frame. Phase Alteration Line (PAL) format is used in Western Europe and Australia. PAL format has a resolution of 625 lines per frame and 50 interlaced frames per second, resulting in 25 frames per second.

There exists a third video format used in France, Russia and Eastern Europe known as SECAM (Sequential Couleur A Memoire; meaning sequential color with memory). It has the same resolution as PAL, but the video information is encoded differently.

• Video Delivery

Video is commonly delivered in composite NTSC or PAL format through phono or BNC (Bayonet, Non-Continuous) connectors. Either format can also be delivered by an S-Video (Y/C) connector. S-Video delivers the luma and chroma components separately. The Y and C signals, if summed together, will form a legal NTSC or PAL signal. S-Video provides a sharper image with better color separation.

• Digital Video Compression Techniques

Analog video is digitized so that it may be manipulated by a computer. Each frame of video becomes a two dimensional array of pixels. A complete color image is composed of three image frames, one for each color components.

Video compression is typically lossy, means some of the information is lost during the compression step. This is acceptable though, because encoding algorithms are designed to discard information that is not perceptible to humans or information that is redundant. There are some basic techniques common to most video compression algorithms, including color space sampling and redundancy reduction

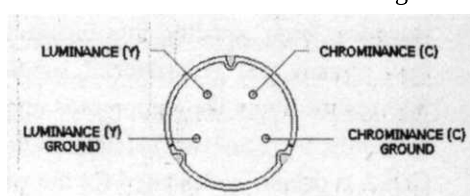

Fig 1: Digital Video Compression

Some important video encoding and compression techniques related to desktop videoconferencing are dotted below:

i. MJPEG

JPEG is an encoding standard for still images developed by the Joint Photographic Experts Group (JPEG). Although designed for still images, with special hardware it is possible to encode and decode a series of JPEG images in real-time to achieve motion video. This use of JPEG encoding is typically referred to as Motion JPEG or MJPEG.

However, no official MJPEG standard exists.

There are four defined modes of operation for JPEG: Sequential, Progressive, Lossless and Hierarchical. Typically only sequential mode is implemented.

ii. ITU-T Recommendation H.261

H.261 is a ITU-T video coding standard, ratified in November 1988. It is the first member of the H.26x family of video coding standards in the domain of the ITU-T Video Coding Experts Group (VCEG), and was the first video codec that was useful in practical terms. H.261 was originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbit/s. The coding algorithm was designed to be able to operate at video bit rates between 40 kbit/s and 2 Mbit/s. The standard supports two video frame sizes: CIF (352x288 luma with 176x144 chroma) and QCIF (176x144 with 88x72 chroma) using a 4:2:0 sampling scheme. It also has a backward-compatible trick for sending still picture graphics with 704x576 luma resolution and 352x288 chroma resolution (which was added in a later revision in 1993).

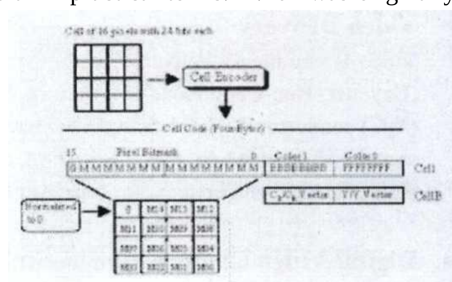

Fig 2: Video Coding

iii. CellB

The Cell image compression algorithm is a variable bit-rate video coding scheme. It provides "high" quality, low bit-rate image compression at low computational cost. The byte stream that is produced by the Cell encoder consists of instructional codes and information about the compressed image.

Currently, there are two versions of the Cell compression technology: Cell A and Cell B. Cell A is primarily designed for the encoding of stored video intended for local display. Cell B, derived from Cell A, has been optimized for network-based video applications. It is computationally symmetric in both encode and decode. Cell B utilizes a fixed color map and vector quantization techniques in the YUV color space to achieve compression.

iv. Network Video

Network Video is an internet videoconferencing tool developed at Xerox / PARC. It is the most commonly used video tool on the Internet MBone. The native nv-coding technique utilizes spatial (intraframe) and temporal (interframe) compression. The first step of the nv-algorithm compares the current frame to the previous frame and marks the areas that have changed significantly.

v. CU-SeeMe

Developed at Cornell University, this internet videoconferencing tool utilizes spatial and temporal compression with a few twists to optimize performance on a Macintosh, its original platform. CU-SeeMe represents video input in 16 shades of grey using 4 bits per pixel. The default transmitting bandwidth setting is 80 kbps.

vi. Indeo

Indeo is a video compression technique designed by Intel. It is evolved from Digital Video Interactive (DVI) technology. Indeo starts off with YUV input; U and V sub sampled 4:1 both horizontally and vertically. Indeo supports motion estimation, using the previous frame to predict values for the current frame and only transmitting data if the difference is significant.

• Communication Channels

Channels can be classified as either circuit-switched or packet-switched.

i. Circuit-Switched Communication Channels

Circuit switching is a methodology of implementing a telecommunications network in which two network nodes establish a dedicated communications channel (circuit) through the network before the nodes may communicate. The circuit guarantees the full bandwidth of the channel and remains connected for the duration of the communication session. :

ii. Packet-Switched Communication Channel

Packet switching is a digital networking communications method that groups all transmitted data - regardless of content, type, or structure - into suitably sized blocks, called packets. Packet switching features delivery of variable-bit-rate data streams (sequences of packets) over a shared network.

iii. Broadband ISDN

BISDN is an extension of ISDN in terms of capabilities, i.e. it not only has the narrowband capability of ISDN but also the broadband capability. The Goal of BISDN is to achieve complete integration of services, ranging from low bit-rate bursty signals to high bit-rate continuous real-time signals.

Modes of Conferencing

Desktop videoconferencing systems communicate in a variety of different ways.

• POTS Conferencing

Plain old telephone service (POTS) is the voice-grade telephone service that remains the basic form of residential and small business service connection to the telephone network in many parts of the world. While POTS provides limited features, low bandwidth and no mobile capabilities, it provides greater reliability than other telephony systems (mobile phone, VoIP, etc.).

• Switched 56 Conferencing

Switched 56 is another neat little name used to denote the use of 56Kbps circuits as a Wide Area Net (WAN) technology or strategy. It is really just a souped-up version of plain old telephone system (POTS) service. The main differences between POTS service and switched 56 service are the speeds that can be obtained.

Problems

Some observers argue that three outstanding issues have prevented videoconferencing from becoming a standard form of communication, despite the ubiquity of videoconferencing-capable systems. These issues are:

1. Eye Contact

Eye contact plays a large role in conversational turn-taking, perceived attention and intent, and other aspects of group communication. Some telepresence systems have cameras located in the screens that reduce the amount of parallax observed by the users. This issue is also being addressed through research that generates a synthetic image with eye contact using stereo reconstruction.

2. Appearance Consciousness

A second psychological problem with videoconferencing is being | on camera, with the video stream possibly even being recorded. The burden of presenting an acceptable onscreen appearance is not present in audio-only communication. Early studies by Alphonse Chapanis found that the addition of video actually impaired communication, possibly because of the consciousness of being on camera.

3. Signal latency

The information transport of digital signals in many steps needs time. In a telecommunicated conversation, an increased latency larger than about 150-300 ms becomes noticeable and is soon observed as unnatural and distracting. Therefore, next to a stable large bandwidth, a small total round-trip time is another major technical requirement for the communication channel for interactive videoconferencing

Recommendation for Future Work

Desktop video conferencing has great potential in the area of distance learning. Videoconferencing provides students with the opportunity to learn by participating in two-way communication forums. Furthermore, teachers and lecturers worldwide can be brought to remote or otherwise isolated educational facilities. Students from diverse communities and backgrounds can come together to learn about one another, although language barriers will continue to persist. Such students are able to explore, communicate, analyze and share information and ideas with one another. Through videoconferencing students can visit other parts of the world to speak with their peers, and visit museums and educational facilities. Such virtual field trips can provide enriched learning opportunities to students, especially those in geographically isolated locations, and to the economically disadvantaged. Small schools can use these technologies to pool resources and provide courses, such as in foreign languages, which could not otherwise be offered.

Conclusion

In this research, Remote Video Conferencing Technology is tested in the North Carolina Research and Education Network (NC-REN). Tele-classrooms and conference rooms across the state of North Carolina are linked together through NC-REN's private microwave facilities. Video communications Services at North Carolina State University (NCSU) and similar facilities throughout the state produce programs that can be sent to any of the 19 universities, medical schools, and research organizations that are connected via NC-REN.

The main purpose of this research paper was to enlist the existing technology, infrastructure and a demonstration project was performed and evaluated to document the features and benefits attain from Remote Video Conferencing.

For this reason, we first enlisted the audio compression and encoding techniques best suited to our requirement. And then we categorized the video compression techniques and procedures by discussing various communication channels and modes of conferencing. Constraints in remote video conferencing technology were also classified and finally, the recommendations for future work were suggested.

Reference(s)

[1] The North Carolina Information Highway; *Part 1: Understanding the Network.* Ver. 1.1 Latest Ver. January 2003

[2] Mulbach et al, 1995. Pg. 291.

[3] Wolfe, Mark. *Broadband videoconferencing as knowledge management tool*, Journal of Knowledge Management 11, no. 2 (2007)

[4] Jim Van Meggelen 2005, *The problem with video conferencing.*

[5] Vertegaal, *Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization?* ACM Conference on Computer Supported Cooperative Work, 2002.

[6] Computer vision approaches to achieving eye contact appeared in the 1990s, such as Teleconferencing Eye Contact Using a Virtual Camera, ACM CHI 1993. More recently gaze correction systems using only a single camera have been shown, such as. Microsoft's GazeMaster system.

[7] Lawson, Stephen. *Vidyo Packages Conferencing For Campuses,* IDG News Service, February 16, 2010. Retrieved via Computerworld.com's website, February 18, 2010

[8] Davis, Andrew W. & Weinstein, Ira M. *The Business Case for Videoconferencing,* Wainhouse Research, March 2005.

[9] Hoffman, Jan. *When Your Therapist Is Only a Click Away,* New York Times, September 25, 2011, pg. STL Also published September 23, 2011 online at www.nytimes.com.