

Internet-Based Videoconferencing Coder/Decoders and Tools for Telemedicine

Wei-Li Liu, M.B.A., Kai Zhang, M.S., Craig Locatis, Ph.D.,
and Michael Ackerman, Ph.D.

Office of High Performance Computing and Communications,
National Library of Medicine, Bethesda, Maryland.

Abstract

Alternative videoconferencing technologies for providing telemedicine via the Internet are described. Background information about how digital video applications have been instantiated using Internet protocols is presented. Specific methods for encoding and decoding video are discussed and video applications that have been tested at the National Library of Medicine are reviewed. This article suggests that no one technology is best and that the appropriateness of a method depends on specific applications. Some technologies, however, have lower, more flexible bandwidth requirements and are more standardized, making them more practical. Still, emerging, yet-to-be-standardized applications offer new capabilities warranting further investigation.

Key words: technology, telecommunications, telemedicine, distance learning

Introduction

Videoconferencing applications range from strictly proprietary to open source, have different video quality, costs, hardware, and network requirements, and need varied levels of expertise to implement and maintain. Two systematic reviews of telemedicine research conducted for the Agency for Healthcare Research and Quality^{1,2} found video most efficacious in specialties wherein verbal interaction is a key part of patient assessment (e.g., psychiatry and neurology), and that outcomes were more variable in other areas. Although telemedicine research is extremely useful, practitioners still have to make choices about specific video technologies, even in specialties where its efficacy is established. These judgments become more salient where the evidence for using the technology is more equivocal.

The National Library of Medicine (NLM) has funded telemedicine projects that have used videoconferencing technologies and has assessed these and other videoconferencing applications employing Internet protocols, including some cutting-edge, experimental ones. Background information is presented in this article about encoding, compressing, and transmitting video as well as other features of videoconferencing applications. Alternative standards for encoding and compressing video are discussed and NLM-tested applications are reviewed, including those built around the H.323

videoconferencing standard, AccessGrid (AG), and ConferenceXP (CXP) and those for transmitting raw video without additional compression.

Background

Most videoconferencing technologies use compression and varied mechanisms for data transmission, controlling audio, securing communication, and sharing software.

COMPRESSION

Video can be viewed on demand or transmitted live, either one way by streaming or two or more ways by point-to-point and multipoint videoconferences. Video data are substantial, and when network bandwidth or data storage are limited, it must be compressed with the goal of maximizing data reduction while maintaining quality. A coder/decoder (CODEC) is either a device or software that compresses. The degree of data reduction is measured by the physical byte size of the file that is stored or the bit rate when data are transmitted, but video quality is measured by resolution or the number of pixels in each image dimension. Television resolution can be standard definition (SD) or high definition (HD). In North America, the SD resolution is 640 × 480 pixels and the HD resolution is either 1280 × 720 or 1920 × 1080 pixels. Video quality can be affected by how the image is scanned and displayed. Interlaced video divides horizontal lines across images into odd and even that refresh alternately, whereas progressive video displays lines in sequence from top to bottom to create a sharper picture. Consequently, the letter i or p is often added to dimensions to define resolution further (e.g., 1280 × 720p).

CODECs may or may not be based on standards and they can be open source or proprietary. Generally, a software CODEC can accomplish sufficient compression with acceptable image quality and latency provided a computer is used with adequate processing power. A hardware CODEC, however, usually will be more efficient. The hardware/software performance gap can be mitigated by using better peripherals such as higher-quality cameras and computers with faster input and output ports. For example, a software CODEC with a higher-quality camera may produce video superior to a hardware system with an inferior one.

Most CODECs employ block-based motion compensation, chroma subsampling, and interframe compression.³ Video is composed of a series of individual images (frames) that when captured and displayed at a rate of 30 per second create the illusion of full motion. Block-based motion compensation divides individual frames into blocks 8 × 8 or 16 × 16 pixels square. Squares that are uniform will be encoded the same, whereas those that are not will be further divided and examined. Chroma subsampling takes advantage of humans

being less sensitive to chroma (color) than to luma (light) to reduce the chroma information within a frame. Some CODECs only perform this intraframe compression, and as every frame must be processed, there can be inefficiencies and latencies. Consequently, other CODECs add interframe compression wherein only the changes between frames get encoded.

TRANSMISSION

The transmission control protocol and user datagram protocol are the most common methods for transmitting data on the Internet.⁴ Data are broken down into packets, which are sent from the machine or device having the data to one or more requesting it. The transmission control protocol guarantees delivery by checking packet transmission and resending those lost. It is used to transmit Web pages, but not video because more data are involved, loss checking takes time, and unacceptable latencies are introduced. The user datagram protocol is used to stream video and audio as there is no packet checking, speeding transfer. Transmission rates are expressed in thousands, millions, or billions of bits per second: kilobits per second (Kbps), megabits per second (Mbps), or gigabits per second (Gbps).

Packets are distributed by unicast and multicast.⁵ Unicast establishes separate streams for every end point participating in a videoconference, each of which will have a unique Internet address or alias (extension) of such an address. If three sites are participating at a bit rate of 384 Kbps, then 1,152 Kbps of bandwidth is consumed on parts of the network. When multiple end points participate by unicast, a device called a multipoint control unit is often needed to manage the streams. Multicast allows transmitting video and audio streams to a single unique multicast address. Routers on the network detecting stream requests will route them to appropriate end points, whereas those not detecting requests will drop transmission. This feature and the use of single streams improve bandwidth efficiency, but end points may be on networks having routers without multicast capability or that do not have it enabled. Consequently, multicast is less common, except on advanced research and education networks.

Videoconferencing technologies having the same resolution (SD or HD) may not have the same image quality if different CODECs are employed or if data are transmitted at different rates. CODECs used by applications will apply different degrees of intraframe compression and some will add interframe. Those allowing varied rates of transmission will apply more compression at lower transmission rates, sometimes to the point the video becomes less clear or frames are dropped, introducing jitter. Video transmitted to a device capable of 1080p resolution at a lower bit rate will look worse than that transmitted at higher ones.

AUDIO, SECURITY, AND APPLICATION SHARING

Videoconferencing hardware devices normally include one or more camera and auxiliary video inputs and microphone and auxiliary audio inputs. They usually have built-in echo cancellation to block any audio sent to a distant end point that is picked up by its microphones and sent back. Videoconferencing software requires video input to the computer through an installed digital video (DV) card or

the computer's IEEE 1394 or universal serial bus (USB) ports. Some software applications will have built-in echo cancellation, but others may require use of headsets or additional echo cancellation hardware.

Videoconferencing applications may offer security by imposing passwords and providing encryption.⁶ The very act of videoconferencing raises network security issues because end points need access to each other's networks that firewalls may block. Some videoconferencing tools accommodate firewalls better because they use fewer ports or more common ones (e.g., port 80 for Web access) that are usually kept open. When built-in security is lacking, it can be provided by other means, for example, by running virtual private network⁷ software and offering a conference over it.

Videoconferencing software may have built-in ways to share slide, word processing, and other applications on a computer, whereas videoconferencing hardware devices allow users to input data from a computer (e.g., a slide presentation) or other information source (e.g., an ultrasound device or microscope) and transmit it as higher-resolution video. If mechanisms for content sharing are lacking, additional software can be installed on computers to provide it and two connections can be made, one for the videoconference and one between the computers sharing data.

CODER/DECODERS

The International Telecommunication Union-T Video Coding Experts Group and the International Standards Organization Motion Picture Experts Group (MPEG) have focused on the H.26x family of compression standards and the MPEG series of compression standards, respectively (e.g., H.261, H.263, MPEG-1, and MPEG-2). The two organizations formed a joint video team and developed the currently popular H.264/MPEG-4 Advanced Video Coding⁸ standard for transmitting and storing video. Other widely used nonstandard, proprietary CODECs are Windows Media Video, Quicktime, and Real Video from Microsoft, Apple, and Real Networks.⁹ Standard CODECs are often preferred because they are intensely reviewed, widely adopted, and more often cross-platform (see summary in *Table 1*). Tools for displaying video using proprietary CODECs usually accommodate standard ones. Video CODECs are integrated into other videoconferencing standards that included additional specifications besides compression (e.g., how end points call each other).

H.261 AND H.263

H.261 was originally designed for transmitting video over ISDN lines at bit rates from 64 Kbps to 1.2 Mbps to support two video resolutions¹⁰ (352×288 and 176×144 pixels). H.263 added resolutions 128×96, 704×576, and 1408×1152 pixels and had bit rates from 128 Kbps to 2 Mbps.¹¹

MPEG-2 AND H.264/MPEG-4

MPEG-2, also known as H.262, is a video- and audio-coding standard used for recording digital video discs (DVDs) and, with enhancements, for transmitting HD television.¹² It is not optimized for transmission at rates below 1 Mbps. The rate for DVD quality is up to 10 and 25 Mbps for HD commercial television quality at resolutions

Table 1. Coder/Decoder Comparison Resolution and Bit Rates

	BIT RATE	RESOLUTION
H.261	64 Kbps–1.2 Mbps	176×144, 352×288
H.263	64 Kbps–1.2 Mbps	128×96, 704×576, 1408×1152
MPEG-2/H.262	5–25 Mbps	352×420 to 1920×1080
H.264/MPEG-4 AVC	40 kbps–10 Mbps	176×144 to 1920×1080
DV	30 Mbps	720×480
HDV	25 Mbps	1080i

from 352×420 to 1920×1080 pixels. H.264/MPEG-4 Advanced Video Coding is the latest H.26x CODEC standard providing the same video quality as MPEG-2 at up to half of the bandwidth,¹³ delivering video at rates from 40 Kbps to 10 Mbps for resolutions ranging between 176×144 and 1920×1080 pixels. Resolutions beyond 1080p are possible at even higher data transfer rates. The MPEG-4 CODEC incorporates the H.264 CODEC and other levels of compression.

DIGITAL VIDEO

DV is a format created by camera producers for storing video digitally at a native resolution of 720×480, which degrades slightly on SD displays.¹⁴ DV only applies intraframe compression, improving video quality and allowing easier editing. The DV specification defines the CODEC, the MiniDV tape storage mechanism, and the recording bit rate. A high-definition DV format called HDV uses MPEG-2 compression for a resolution of ~1080i. Some applications for streaming video without additional compression use the DV format.

Applications

NLM has experimented with SD and HD videoconferencing. SD applications generally are more stable, easier to deploy, and less costly, while providing sufficient video quality for many telemedicine situations. HD videoconferencing can provide greater resolution when higher image quality is needed. NLM has experimented with three applications using one or more CODECs previously discussed: (1) H.323 applications, (2) AG, and (3) CXP. In addition, NLM has tested uncompressed video. Criteria for choosing videoconferencing applications are summarized in *Table 2*.

H.323

H.323 is a videoconferencing standard widely adopted by commercial equipment manufacturers and incorporates H.261, H.263, and H.264 video CODECs, the latter providing HD resolution in recent products. The metaphor for H.323 videoconferencing is the point-to-point call or conference call, wherein an Internet protocol address of a videoconferencing end point or multipoint conferencing unit or its alias is used to initiate communication.¹⁵ In addition to CODECs and call initiation, the specification includes standards for sharing content, remotely controlling cameras, and so on. These elements work together to provide interoperable video, audio, and data communication by unicast or multicast,¹⁶ although unicast is more prevalent.

NLM has used H.323 products from Tandberg, LifeSize, and Polycom, employing customized digital signal processing hardware for video compression, although there are software implementations also deployed by NLM. The bit rates for SD video are those for the H.261 and H.263 CODECs (64 Kbps to 1.2 Mbps), whereas those for HD range from 512 Kbps to 4 Mbps at 720p or 1080p resolution, the actual appearance depending on bit rate. As additional compression is applied in videoconferencing HD implementations, image quality

Table 2. Videoconferencing Application Selection Criteria

- Does the videoconferencing application have the image quality required for the telemedicine context?
Some applications may be completely inadequate for certain types of telemedicine. For example, in dermatology, higher video resolution may be needed to show the patient's skin more clearly. However, others may provide more resolution than needed, often at additional costs (in terms of either pricing or technical support).
- Does the bandwidth available accommodate the videoconferencing application?
Some applications accommodate a range of bandwidths but reduce quality when set for lower bit rates. Consequently, the bit rate the network can accommodate has to be compared to the image quality required (criterion 1). Some applications, especially those uncompressed, may exceed available bandwidth.
- Does the videoconferencing application have the functionality and performance required?
For example, if providers need to control cameras at the distant end point and view input from other devices, does the application allow this? Similarly, some applications may have greater latency, making communication difficult, or they may be more vulnerable to network congestion.
- Is the videoconferencing application easy to use and is there technical support?
Generally, commercially available applications are easier to use and have company support. Open source applications evolve based on the work done by the community using them and in-house technical expertise may be needed to support them. Even when in-house technical expertise is available, some applications (e.g., uncompressed video) are very new and not well documented.
- Is the cost of the videoconferencing application appropriate?
An application's price must fit one's budget and hidden costs must be taken into account. A free, open-source software is attractively "priced," but the technical support that is often needed may cost more than a commercial product. Moreover, commercial products often offer technical support and service contracts at a marginal additional expense.

is lower than commercial HD television. The standard includes the T.120 protocol for sharing applications when H.323 software is installed on computers, whereas standalone systems usually have computer inputs and employ the H.239 protocol to dual-stream computer output as higher-resolution video. Most commercial H.323 products have echo cancellation and provide encryption.

ACCESSGRID

AG is a free, open-source videoconferencing software developed by Argonne National Laboratory (ANL), enabling large-scale collaboration and providing interactive application sharing. Its principle components are the videoconferencing tool (VIC)¹⁷ for video and the Robust Audio Tool¹⁸ for audio. Unlike call-initiated H.323 applications, the AG is based on the concept of a venue or virtual place. Users can establish their own venue servers or use public ones at ANL and the National Center for Supercomputing Applications. Client software accesses a venue server “lobby” where users can navigate to other venues established by server managers. When users access the same lobby or subvenue, they see and hear each other because the venue sets multicast or unicast addresses (usually both) for transmitting video and audio. Users have to access venues in the same mode (multicast or unicast) to communicate because of unicast and multicast address differences.^{19,20}

AG clients work with late model computers, but they need to have mechanisms for inputting video and audio and for echo cancellation. Like H.323, AG’s VIC tool supports multiple CODECs, including H.261 and H.263, from SD analog cameras input through video capture cards or from USB cameras. An enhanced version of VIC supports H.264 and MPEG-4 video, and recently at HD resolution. The MPEG-4 compression provides full-motion video, but the H.264 compression (as tested at NLM) was jittery because greater compression is applied and the computer used was not robust enough. Optional components can be added to clients to transmit uncompressed DV. Network bandwidth requirements depend on which CODEC is used, whether video is transmitted uncompressed, and the number of end points participating. When H.263 is used, the rule of thumb is 2.5 Mbps per video stream.

Clients can run under Windows, Mac, Linux, and additional application sharing programs that work within it, including those for sharing PowerPoint slides and Web browsers and recording and playing back videoconference sessions. IOCOM is a commercial videoconferencing application built on the AG toolkit with enhanced proprietary features. The IOCOM client can interoperate with AG clients when proprietary features are not used and the company provides a bridge to public AG venues at ANL and National Center for Supercomputing Applications as well as bridges allowing H.323 systems, cell phones, and other devices to interoperate with its system. AG and IOCOM provide encryption.

CONFERENCEXP

CXP, originally developed by Microsoft Research, is a free, open-source collaboration software maintained by the University of Washington.²¹ Like AG, it uses public or private venue servers and has additional programs for sharing PowerPoint, browser and other

applications, and software for recording and playing back videoconferences. Similarly, it needs additional mechanisms for inputting video and providing echo cancellation. Unlike the AG, CXP runs only under Windows and, although open source, uses Windows Media technology. The technology compresses video in Microsoft’s proprietary format and other standardized ones. Default resolution for SD video is 320×240, but it can be set at 640×480. CXP supports both unicast and multicast and provides encryption.

UNCOMPRESSED VIDEO

There is interest in transmitting DV without applying additional compression to attain higher image quality and there are software programs for doing this with SD and HD video. They include the DV transport system (DVTS), iHDTV, UltraGrid, and HD CXP. With the exception of HD CXP, the programs are designed to work independently of any applications discussed earlier, although others have written or are in process of writing programs to include some in the AG toolkit. Generally, the video quality of these programs exceeds applications employing compression, but the programs do not currently provide encryption and have higher bit rates.

DVTS is a software that “packetizes” input from a miniDV camcorder through the computer’s IEEE 1394, firewire, port to provide video essentially equivalent to the camera’s.²² An alternative version, HD video transport system, works similarly, except that an HD miniDV camcorder is required and MPEG-2 compression is applied to achieve a resolution of ~1080i HD video. Both have bit rates of about 30 Mbps.²³

iHDTV, UltraGrid, and HD CXP are all open-source packetizing software for transmitting HD uncompressed video directly from HD video cameras using HD video capture cards.^{24,25} Uncompressed iHDTV and UltraGrid have 1920×1080i resolution and 1.5 gigabits per second bit rate. iHDTV requires embedding hardware to synchronize audio and video, whereas UltraGrid requires a separate audio mechanism. UltraGrid can compress HD video to a rate of 250 Mbps. HD CXP requires video inputs similar to iHDTV and UltraGrid, but it can capture audio from the computer’s sound card or external source. It purports to support 1920×1080p video resolution uncompressed, a resolution NLM has yet to test, while also allowing video compression for bit rates of 1–5 Mbps.

Conclusions

NLM has investigated a range of videoconferencing applications for telemedicine. Commercial products using the H.323 standard are the most widely deployed, but the AG and, to a lesser extent, CXP have been deployed at educational and research institutions on advanced networks, such as Internet2 and National Lambda Rail. The latter have been mostly used for distance learning because they are concentrated in academia and can accommodate many end points. Uncompressed video is very experimental and has been deployed only in a limited number of research centers having sufficient bandwidth, but may become more widely used as the technology matures and available bandwidth increases. Although bandwidth requirements are a limiting factor, the increased resolution offered may make these applications especially attractive for telemedicine. As bandwidth is constantly improving, future investigation is warranted

of higher definition video without compression and with varied degrees of compression applied. Although it is logical to assume that compressed and uncompressed HD will perform better in telemedicine applications where the efficacy of SD video is well established (e.g., telepsychiatry), the image quality that can be realized when HD video is uncompressed or with varied degrees of compression applied needs further study in laboratory and clinical settings for other applications.

Disclosure Statement

No competing financial interests exist.

REFERENCES

- Hersh WR, Hickam DH, Severance SM, Dana TL, Krages KP, Helfand M. Telemedicine for the Medicare Population: Update. Evidence Report/Technology Assessment No. 131 (Prepared by the Oregon Evidence-based Practice Center under Contract No. 290-02-0024.) AHRQ Publication No. 06-E007. Rockville, MD: Agency for Healthcare Research and Quality, 2006.
- Hersh WR, Wallace JA, Patterson PK, et al. Telemedicine for the Medicare Program. Evidence Report/Technology Assessment No. 24 (Prepared by Oregon Health Sciences University, Portland, OR under Contract No. 290-97-0018). AHRQ Publication No. 01-E012. Rockville, MD: Agency for Healthcare Research and Quality, 2001.
- Thomas W, Gary JS, Gisle B, Ajay L. Overview of the H.264/AVC video coding standard. *IEEE Trans Circuits Syst Video Technol* 2003;13:560-576.
- Lakshmanan S, Venkadachalam S. Integrated proactive admission control technique for both UDP and TCP traffic flows. *J Netw* 2007;2:1-8.
- Carlos ASO, Panos MP. A survey of combinatorial optimization problems in multicast routing. *Comput Oper Res* 2005;32:1953-1981.
- Haifeng Q, Yuan Z, Zhibin L, Zecheng W, Bing Z. Efficient public key encryption with smallest ciphertext expansion from factoring. *Des Codes Cryptogr* 2008;49:233-249.
- Venkateswaran R. Virtual private network. *IEEE Potentials* 2001;20:11-15.
- Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec). H.264/ISO/IEC 14 496-10 AVC, in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT- G050, 2002. 8th Meeting: Geneva, Switzerland, May 23-27, 2003.
- Hung-Jen Y, Hsiao-Chih L, Yeong-Diann W, Lung-Hsing K. Designing and constructing live streaming system for broadcast. Proceedings of the 2010 American Conference on Applied Mathematics 2010;1:266-271.
- Video Codec for Audiovisual Services at px64 kbit/s ITU-T Recommendation H.261, Version 1, ITU-T, ITU-T Recommendation H.261 Version 1, 1988. International Telecommunications Union: Geneva, Switzerland.
- Video Coding for Low Bit Rate Communication, ITU-T, ITU-T Recommendation H.263 version 1, 2005. International Telecommunications Union: Geneva, Switzerland.
- Zhi Z, Shijun S, Shawamin L, Ming-Ting S. Motion information and coding mode reuse for MPEG-2 to H.264 transcoding. *IEEE Int Symp Circ Syst* 2005;2:1230-1233.
- Apple, Inc. Technology Brief, QuickTime and MPEG-4: Now Featuring H.264 [Internet]. Available at http://images.apple.com/br/quicktime/pdf/H264_Technology_brief.pdf (last accessed September 6, 2010)
- Akimichi O, Katsushi K, Kazunori S, Osamu N, Jun M. Design and implementation of DV based video over RTP, 2000. Keio University, Japan.
- Glasmann J, Kellerer W, Muller H. Service architectures in H.323 and SIP: A comparison. *Commun Surv Tutorials IEEE* 2003;5:32-47.
- Obrazzka K. Multicast transport mechanisms: A survey and taxonomy. *IEEE Commun Mag* 1998;36:94-102.
- McCanne S, Jacobson V. Vic: A flexible framework for packet video. Proceedings of the Third ACM International Conference on Multimedia, 1995;1:511-522.
- Hardman V, Sasse MA, Kouvelas I. Successful multiparty audio communication over the Internet. *Commun ACM* 1998;41:74-80.
- Childers L, Disz T, Olson R, Papka ME, Stevens R, Udeshi T. Access Grid: Immersive group-to-group collaborative visualization. 4th International Immersive Projection Technology Workshop, Ames, IA, 2000.
- Silverstein JC, Dech F, Binns J, Jones D, Papka ME, Stevens R. Distributed collaborative radiological visualization using Access Grid. *Stud Health Technol Inform* 2005;111:477-481.
- Needham T. ConferenceXP and advanced collaborative scenarios. *Int Symp Collab Technol Syst* 2006;1:392.
- Ogawa A, Kobayashi K, Sugiura K, Nakamura O, Murai J. Design and implementation of DV based video over RTP. Packet Video Workshop, Sardinia, Italy, 2000.
- Lee J, Im H, Chon K. High-definition digital video over IP: System design, experiments and applications. Proceedings of the APAN Network Research Workshop, Cairns, Australia, 2004.
- iHDTV project, ResearchChannel [Internet]. Available at <http://ihdtv.sourceforge.net/> (last accessed October 15, 2010).
- Holub P, Matyska L, Liška M, Hutanuc A, Paruchuric R, Radila J, Hladkása E. High-definition multimedia for multiparty low-latency interactive communication. *Future Generation Comput Syst* 2006;22:856-861.

Address correspondence to:
Wei-Li Liu, M.B.A.

Office of High Performance Computing and Communications
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

E-mail: wliu@mail.nih.gov

Received: November 4, 2010
Revised: December 21, 2010
Accepted: December 23, 2010