Overview of the px64 kbit/s Video Coding Standard

Ming Liou

Recognizing the need for providing ubiquitous video services using the Integrated Services Digital Network (ISDN) [3], CCITT (International Telegraph and Telephone Consultative Committee) Study Group XV established a Specialist Group on Coding for Visual Telephony in 1984 with the objective of recommending a video coding standard for transmission at $m \times 384$ kbit/s ($m = 1,2, \ldots, 5$). Later in the study period, standardization at $n \times 64$ kbit/s ($n = 1,2, \ldots, 5$) was also considered. With new discoveries in video coding techniques, it became clear that a single standard, $p \times 64$ kbit/s ($p = 1,2, \ldots, 30$), can cover the entire ISDN channel capacity. After more than five years of intensive deliberation, CCITT Recommendation H.261, Video Codec for Audiovisual Services at $p \times 64$ kbit/s, was completed and approved in December 1990 [1, 7, 9–11]. A slightly modified version of this Recommendation was also adopted for use in North America [2].

The intended applications of this international standard are for videophone and videoconferencing. Therefore, the recommended video coding algorithm has to be able to operate in real time with minimum delay. For $p = 1$ *or* 2, due to severely limited available bit rate, only desktop face-to-face visual communication (often referred to as videophone) is appropriate. For $p \geq 6$, due to the additional available bit rate, more complex pictures can be transmitted with better quality. This is, therefore, more suitable for videoconferencing. This article presents a brief overview of the H.261 video coding algorithm. Additional information related to this article can be found in [4] and [12].

Video Format

The CCITT has adopted the Common Intermediate Format (CIF) and Quarter-CIF (QCIF) as the video formats for visual telephony. The parameters for these formats are listed in Table 1. All codecs must be able to operate with QCIF. Use of CIF is optional. The maximum picture rate for both formats is 30000/1001 (approximately 29.97) frames per second with the provision that either 1, 2, or 3 frames may be dropped at an encoder between transmitted frames.

The uncompressed bit rates for transmitting CIF and QCIF at 29.97 frames/sec are 36.45 and 9.115 *Mbit/s* respectively. A great deal of bit-rate reduction is required to transport these video signals using ISDN channels ($p \times 64$ *kbit/s*, $p = 1, 2, \ldots, 30$). The choice of CIF or QCIF depends on the availability of channel capacity. For $p = 1$ *or* 2, QCIF is normally used for desktop videophone applications. Even with QCIF operated at 10 frames per second, a bit-rate reduction of approximately 47.5 to 1 is needed to transport that signal using a 64 *kbit/s* channel. This is an extremely difficult task to achieve.

TABLE 1.

Parameters for CCITT Video Formats

	CIF		QCIF	
	Lines/Frame	Pixels/Line	Lines/Frame	Pixels/Line
Luminance (Y)	288	360 (352)	144	180 (176)
Chrominance (C_B)	144	180 (176)	72	90 (88)
Chrominance (C_R)	144	180 (176)	72	90 (88)

Note: The numbers in the parentheses represent coded pixels.

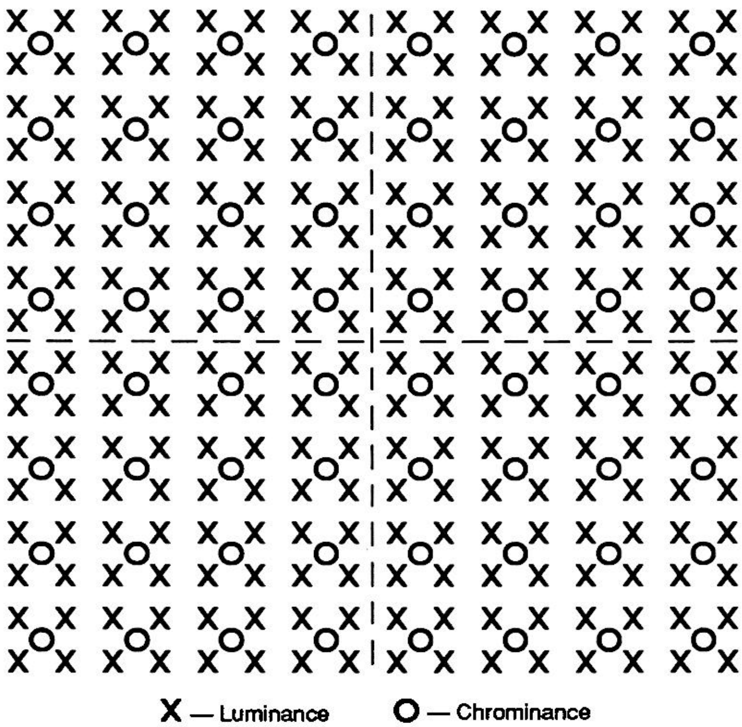
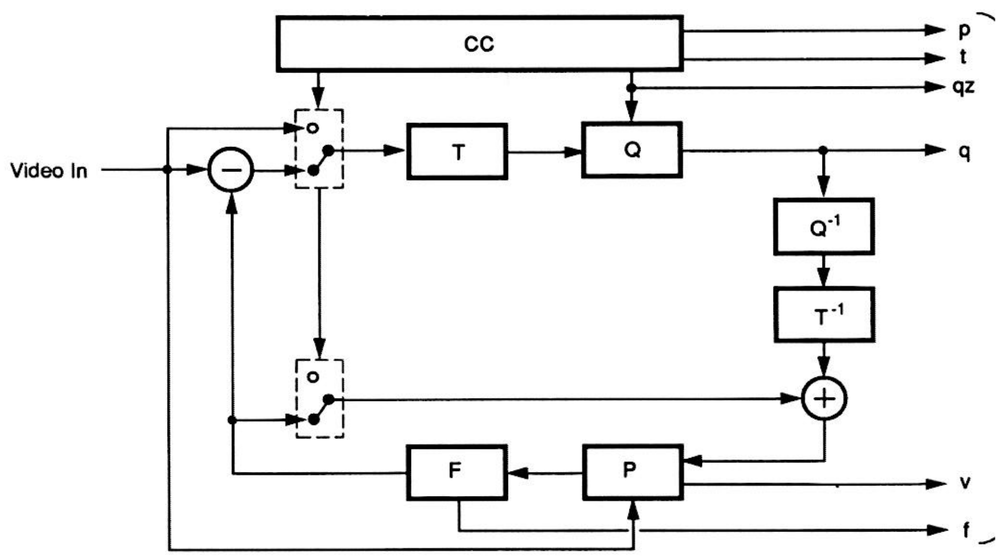


FIGURE 1.
Composition of a Macro Block

For $p \geq 6$, CIF may be used since there are more bits available to code a picture. Due to its increased resolution, CIF is more appropriate for videoconferencing applications.

For the video coding algorithm recommended by the CCITT, the CIF and QCIF are divided into a hierarchical block structure consisting of Pictures, Groups of Blocks (GOBs), Macro Blocks (MB), and Blocks. The composition of a Macro Block is shown in Figure 1. Each Macro Block is composed of four 8×8 luminance (Y) Blocks and two 8×8 chrominance (C_B and C_R) Blocks. A Group of Blocks is composed of 3×11 Macro Blocks. A QCIF Picture has 3 GOBs while CIF has four times that number. This elaborate hierarchical block structure is essential for the high-compression video coding algorithm which will be briefly described in the following section.



T — Transform
Q — Quantizer
P — Picture memory with motion-compensated variable delay
F — Loop filter
CC — Coding control

p — Flag for INTRA/INTER
t — Flag for transmitted or not
qz — Quantizer indication
q — Quantizing index for transform coefficients
v — Motion vector
f — Switching on/off of the loop filter

FIGURE 2.
Block Diagram of the Source Coder

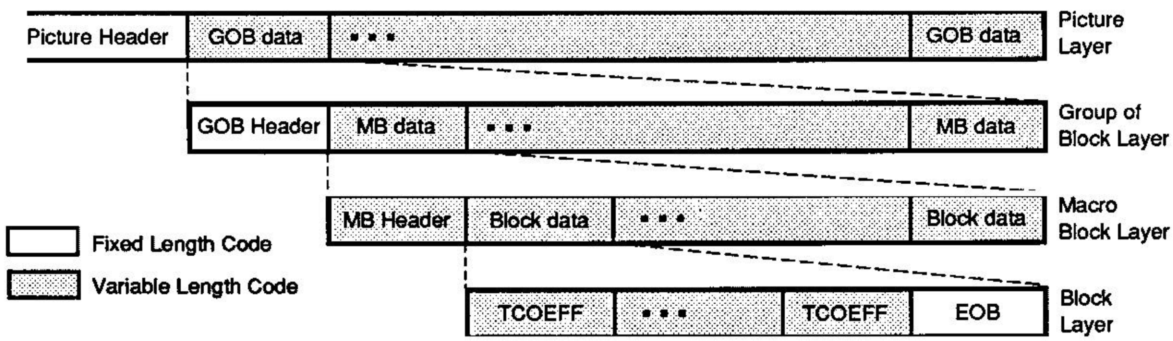


FIGURE 3.
A Simplified Data Structure of the Video Multiplex Coder

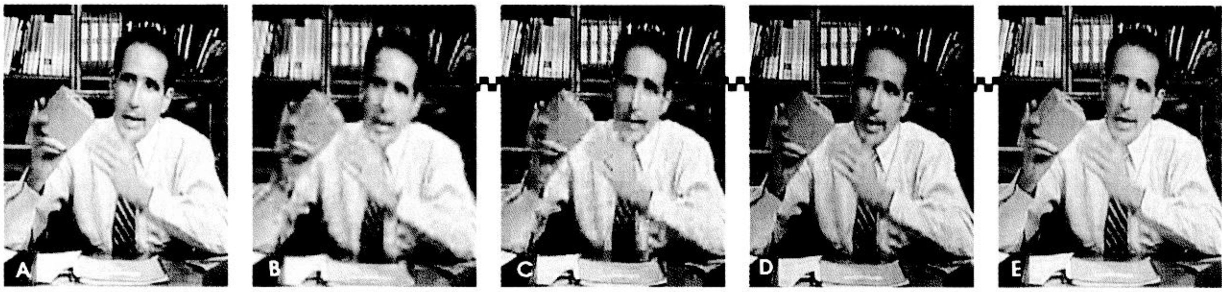


FIGURE 4.
Comparison of Picture Quality
(a) Original
(b) Coded Picture—Format: QCIF, frame rate: 15, bit rate 64 kbit/s
(c) Coded Picture—Format: CIF, frame rate: 10, bit rate: 64 kbit/s
(d) Coded Picture—Format: CIF, frame rate: 15, bit rate 320 kbit/s
(e) Coded Picture—Format: CIF, frame rate: 30, bit rate: 1,472 kbit/s

Video Coding Algorithm

The basic objective of video coding is to reduce the bit rate by removing redundant information. There are two major categories of coding schemes, i.e., source coding and entropy coding. Source coding deals with source material and yields results which are lossy, i.e., picture quality is degraded. Source coding can be further divided into intraframe and interframe coding. Intraframe coding is used for the first picture and for later pictures after a change of scene. Interframe coding is for sequences of similar pictures, including those containing moving objects. Intraframe coding removes only the spatial redundancy within a picture, whereas interframe coding removes also the temporal redundancy between pictures. Entropy coding achieves bit-rate reduction by using the statistical properties of the signals and, in theory, is lossless. The CCITT H.261 video coding algorithm uses both the intraframe and interframe coding schemes.

A block diagram of the CCITT H.261 source encoder is shown in Figure 2. It is a hybrid of DCT (Discrete Cosine Transform) and DPCM (Differential Pulse Code Modulation) schemes with motion estimation [6, 8]. In intraframe mode, the DPCM is not operative. Every 8×8 block in a picture frame is transformed into DCT coefficients, linearly quantized, and then sent to the video multiplex coder. The same picture frame is also recovered (through the inverse quantizer and inverse transform) and stored in the picture memory for interframe coding.

During the interframe coding mode, the DPCM is in operation. The prediction is based on motion estimation by comparing every Macro Block (luminance only) of the current frame with the Macro Blocks in the neighborhood of the corresponding Macro Block in the previous frame. If the difference between the current and the predicted Macro Blocks is less than a certain threshold, no data is transformed for that Macro Block. Otherwise, the difference is DCT transformed, linearly quantized, and then sent to the video multiplex coder together with motion vector information. A loop filter can be switched on and off to improve picture quality by removing high-frequency noise when needed. The step size of the linear quantizer can be adjusted depending on the fullness of the transmission buffer of an encoder. When the transmission buffer is close to full, the step size will be increased so that less information needs to be coded. This, of course, will result in a degraded picture. On the other hand, the step size will be decreased to improve picture quality when the transmission buffer is not full.

To further increase coding efficiency, variable word-length entropy coding is used in the video multiplex coder which immediately follows the source coder. There are five variable word-length coding tables for the quantized DCT coefficients and various side information. The output of the video multiplex coder is sent to a transmission buffer, which regulates the flow of video information to a constant bit rate by controlling the step size of the linear quantizer.

Video Data Structure

One of the most important aspects of a video coding standard is to define a data structure so that a decoder can decode the received bit stream without any ambiguity. A simplified hierarchical structure with four layers of video data is shown in Figure 3.

Data for each picture consists of a picture header followed by data for GOBs. The picture header includes a 20-bit picture start code and other information such as video format (CIF or QCIF), temporal reference (frame number), etc.

The GOB layer consists of a GOB header followed by data for MBs. The GOB header includes a 16-bit GOB start code and other information such as the position of the GOB, and quantizer information for the GOB until overridden by any subsequent MB quantizer information, etc.

The MB layer consists of an MB header followed by data for Blocks. The MB header includes a variable length code (VLC) for the MB address. It is followed by a VLC for MB type indicating whether it is intraframe or interframe, with or without motion estimation and/or loop filter. Depending on a particular MB type, various combinations of video side information may follow. When motion estimation is accurate to within a given specification, no block data for DCT coefficients needs to be transmitted. It should also be noted that not every MB in a GOB needs to be transmitted if it contains no information for that part of the picture.

The Block layer contains the DCT coefficients (TCOEFF) of a block followed by a fixed-length code, EOB, to indicate the end of a block. The coefficients are coded using a two-dimensional VLC. Not every block in an MB needs to be transmitted.

Comparison of Picture Quality

The CCITT H.261 video coding algorithm covers a wide range of bit rates for various real-time visual applications. The picture quality as well as motion effect varies depending on the bit rate used. While it is difficult to demonstrate motion effect in printed form, image quality can be roughly compared at a particular frame of a video sequence for CIF and QCIF at various bit rates.

Figure 4 shows the original and coded pictures under various conditions. The 115th frame of the "Salesman" test sequence was used for illustration. Clearly, the picture quality improves as the bit rate increases. Additional post-processing can be used to further improve picture quality. However, this is not a part of the standard and can be implemented in many different ways by codec manufacturers.

Conclusion

A brief overview of CCITT Recommendation H.261, Video Codec for Audiovisual Services at $p \times 64$ kbit/s, has been presented. The recommended video coding algorithm has been fully tested with experimental prototypes constructed by various organizations involved in the standardization process. VLSI implementation of several key signal processing modules, such as DCT, motion estimation, and variable length codec, have already appeared in the market [5]. It is expected that the cost of such a codec will drop significantly in the near future.

A similar video coding standard for multimedia applications at about 1.5 Mbit/s has been recently approved by the Moving Picture Expert Group of the International Standards Organization [4]. Although the two standards are not compatible, many of the key signal processing modules mentioned in this article are used in both standards. Consequently, the cost of these signal processing modules will drop even further due to increased volume.

Providing a ubiquitous $p \times 64$ kbit/s audiovisual service requires the following four essential ingredients: the availability of worldwide ISDN; advanced video compression techniques; low-cost VLSI technology; and an international video coding standard. It appears that all of these ingredients are present today and visual telephony using ISDN will soon become a reality.

Acknowledgments

The author wishes to thank numerous experts for their invaluable contributions toward the successful completion of this important video coding standard. He is also grateful to Ali Tabatabai for providing the coded pictures shown in this article.

References

1. Carr, M. et al. Motion video coding in CCITT SGXV—The video multiplex and transmission coding. *IEEE Global Telecomm. Conf. Rec.*, (Nov. 1988.), 1005–1010.
2. Digital processing of video signals—Video coder/decoder for audiovisual services at 56 to 1,536 kbit/s. American National Standard T1.p64–199x (Draft), American National Standards Institute, Inc., Oct. 1990.
3. *IEEE Commun. Magazine 28*, 4 (Apr. 1990), Special Issue on ISDN.
4. LeGall, D.J. MPEG: A video compression standard for multimedia applications. *Commun. ACM 34*, 4 (Apr. 1991).
5. Liou, M.L. Visual telephony as an ISDN application. *IEEE Commun. Magazine 28*, 2 (Feb. 1990), 30–38.
6. Netravali, A.N. and Haskell, B.G. *Digital Pictures—Representation and Compression.* Plenum Press, New York and London, 1988.
7. Nicol, R.C. and Mukawa, N. Motion video coding in CCITT SGXV—The coded picture format. *IEEE Global Telecomm. Conf. Rec.* (Nov. 1988), 992–996.
8. Plompen, R. Motion Video Coding for Visual Telephony. PTT Research Neher Laboratories, The Netherlands, 1989.
9. Plompen, R. et al. Motion video coding in CCITT SGXV—The video source coding. *IEEE Global Telecomm. Conf. Rec.* (Nov. 1988), 997–1004.
10. Tabatabai, A., Mills, M., and Liou, M.L. A review of CCITT $p \times 64$ *kpbs* video coding and related standards. *International Electronic Imaging Exposition and Conference*, (Oct. 1990), 58–61.
11. Video Codec for Audiovisual Services at $p \times 64$ kbit/s. CCITT Recommendation H.261, CDM XV-R 37-E, International Telegraph and Telephone Consultative Committee (CCITT), Aug. 1990.
12. Wallace, G. The JPEG still picture compression standard. *Commun ACM 34*, 4 (Apr. 1991).

CR Categories and Subject Descriptors: C.2.0 [**Computer-Communication Networks**]: General—*Data communications;* I.4.2 [**Image Processing**]: Compression (coding)—*Approximate methods*

General Terms: Design, Standardization

Additional Key Words and Phrases: px64 kbit/s video format

About the Author:

MING L. LIOU is a district manager at Bellcore. His current research interests include video processing, image compression techniques, VLSI architecture, and implementation for video applications. **Author's Present Address:** Bellcore, RM 3X-303, 331 Newman Springs Road, Red Bank, NJ 07701-7040. email: liou@nyquist.bellcore.com