19 ITU-T Video Coding Standards H.261 and H.263

This chapter introduces ITU-T video coding standards H.261 and H.263, which are established mainly for videophony and videoconferencing. The basic technical detail of H.261 is presented. The technical improvements with which H.263 achieves high coding efficiency are discussed. Features of H.263+, H.263++, and H.26L are presented.

19.1 INTRODUCTION

Very low bit rate video coding has found many industry applications such as wireless and network communications. The rapid convergence of standardization of digital video-coding standards is the reflection of several factors: the maturity of technologies in terms of algorithmic performance, hardware implementation with VLSI technology, and the market need for rapid advances in wireless and network communications. As stated in the previous chapters, these standards include JPEG for still image coding and MPEG-1/2 for CD-ROM storage and digital television applications. In parallel with the ISO/IEC development of the MPEG-1/2 standards, the ITU-T has developed H.261 (ITU-T, 1993) for videotelephony and videoconferencing applications in an ISDN environment.

19.2 H.261 VIDEO-CODING STANDARD

The H.261 video-coding standard was developed by ITU-T study group XV during 1988 to 1993. It was adopted in 1990 and the final revision approved in 1993. This is also referred to as the P × 64 standard because it encodes the digital video signals at the bit rates of P × 64 Kbps, where *P* is an integer from 1 to 30, i.e., at the bit rates 64 Kbps to 1.92 Mbps.

19.2.1 Overview of H.261 Video-Coding Standard

The H.261 video-coding standard has many features in common with the MPEG-1 video-coding standard. However, since they target different applications, there exist many differences between the two standards, such as data rates, picture quality, end-to-end delay, and others. Before indicating the differences between the two coding standards, we describe the major similarity between H.261 and MPEG-1/2. First, both standards are used to code similar video format. H.261 is mainly used to code the video with the common intermediate format (CIF) or quarter-CIF (QCIF) spatial resolution for teleconferencing application. MPEG-1 uses CIF, SIF, or higher spatial resolution for CD-ROM applications. The original motivation for developing the H.261 video-coding standard was to provide a standard that can be used for both PAL and NTSC television signals. But later, the H.261 was mainly used for videoconferencing and the MPEG-1/2 was used for digital television (DTV), VCD (video CD), and DVD (digital video disk). The two TV systems, PAL and NTSC, use different line and picture rates. The NTSC, which is used in North America and Japan, uses 525 lines per interlaced picture at 30 frames/second. The PAL system is used for most other countries, and it uses 625 lines per interlaced picture at 25 frames/second. For this purpose, the CIF was adopted as the source video format for the H.261 video coder. The CIF format consists of 352 pixels/line, 288 lines/frame, and 30 frames/second. This format represents half the active lines of the PAL signal and the same picture rate of the NTSC signal. The PAL systems need only perform a picture rate conversion and NTSC systems need only perform a line number conversion. Color pictures consist of one luminance and two color-difference components (referred to as Y C_b C_r format) as specified by the CCIR601 standard. The C_b and C_r components are the half-size on both horizontal and vertical directions and have 176 pixels/line and 144 lines/frame. The other format, QCIF, is used for very low bit rate applications. The QCIF has half the number of pixels and half the number of lines of CIF format. Second, the key coding algorithms of H.261 and MPEG-1 are very similar. Both H.261 and MPEG-1 use DCT-based coding to remove intraframe redundancy and motion compensation to remove interframe redundancy.

Now let us describe the main differences between the two coding standards with respect to coding algorithms. The main differences include:

- H.261 uses only I- and P-macroblocks but no B-macroblocks, while MPEG-1 uses three macroblock types, I-, P-, and B-macroblocks (I-macroblock is in intraframe-coded macroblock, P-macroblock is a predictive-coded macroblock, and B-macroblock is a bidirectionally coded macroblock), as well as three picture types, I-, P-, and B-pictures as defined in Chapter 16 for the MPEG-1 standard.
- There is a constraint of H.261 that for every 132 interframe-coded macroblocks, which corresponds to 4 GOBs (group of blocks) or to one-third of the CIF pictures, it requires at least one intraframe-coded macroblock. To obtain better coding performance at low-bit-rate applications, most encoding schemes of H.261 prefer not to use intraframe coding on all the macroblocks of a picture, but only on a few macroblocks in every picture with a rotational scheme. MPEG-1 uses the GOP (group of pictures) structure, where the size of GOP (the distance between two I-pictures) is not specified.
- The end-to-end delay is not a critical issue for MPEG-1, but is critical for H.261. The video encoder and video decoder delays of H.261 need to be known to allow audio compensation delays to be fixed when H.261 is used in interactive applications. This will allow lip synchronization to be maintained.
- The accuracy of motion compensation in MPEG-1 is up to a half-pixel, but is only a full-pixel in H.261. However, H.261 uses a loop filter to smooth the previous frame. This filter attempts to minimize the prediction error.
- In H.261, a fixed picture aspect ratio of 4:3 is used. In MPEG-1, several picture aspect ratios can be used and the picture aspect ratio is defined in the picture header.
- Finally, in H.261, the encoded picture rate is restricted to allow up to three skipped frames. This would allow the control mechanism in the encoder some flexibility to control the encoded picture quality and satisfy the buffer regulation. Although MPEG-1 has no restriction on skipped frames, the encoder usually does not perform frame skipping. Rather, the syntax for B-frames is exploited, as B-frames require much fewer bits than P-pictures.

19.2.2 Technical Detail of H.261

The key technologies used in the H.261 video-coding standard are the DCT and motion compensation. The main components in the encoder include DCT, prediction, quantization (Q), inverse DCT (IDCT), inverse quantization (IQ), loop filter, frame memory, variable-length coding, and coding control unit. A typical encoder structure is shown in Figure 19.1.

The input video source is first converted to the CIF frame and then is stored in the frame memory. The CIF frame is then partitioned into GOBs. The GOB contains 33 macroblocks, which are 1/12 of a CIF picture or 1/3 of a QCIF picture. Each macroblock consists of six 8×8 blocks among which four are luminance (Y) blocks and two are chrominance blocks (one of C_b and one of C_r).

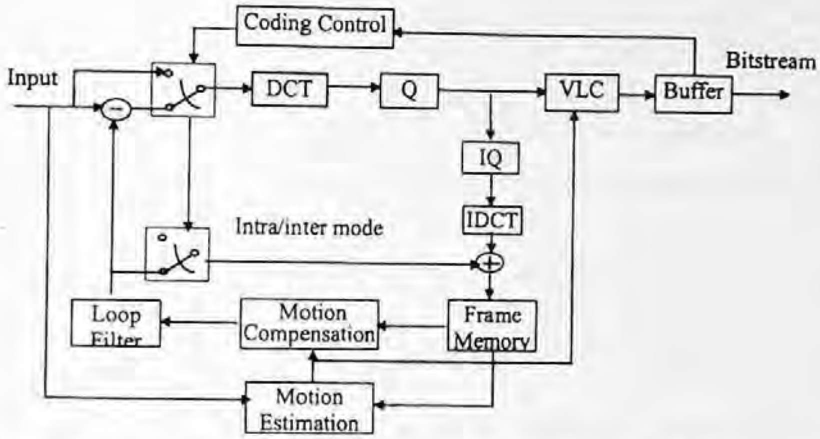

FIGURE 19.1 Block diagram of a typical H.261 video encoder. (From ITU-T Recommendation H.261, March 1993. With permission.)

For the intraframe mode, each 8 × 8 block is first transformed with DCT and then quantized. The variable-length coding (VLC) is applied to the quantized DCT coefficients with a zigzag scanning order such as in MPEG-1. The resulting bits are sent to the encoder buffer to form a bitstream.

For the interframe-coding mode, frame prediction is performed with motion estimation in a similar manner to that in MPEG-1, but only P-macroblocks and P-pictures, no B-macroblocks and B-pictures, are used. Each 8 × 8 block of differences or prediction residues is coded by the same DCT coding path as for intraframe coding. In the motion-compensated predictive coding, the encoder should perform the motion estimation with the reconstructed pictures instead of the original video data, as it will be done in the decoder. Therefore, the IQ and IDCT blocks are included in the motion compensation loop to reduce the error propagation drift. Since the VLC operation is lossless, there is no need to include the VLC block in the motion compensation loop. The role of the spatial filter is to minimize the prediction error by smoothing the previous frame that is used for motion compensation.

The loop filter is a separable 2-D spatial filter that operates on an 8 × 8 block. The corresponding 1-D filters are nonrecursive with coefficients ¼, ½, ¼. At block boundaries, the coefficients are 0, 1, 0 to avoid the taps falling outside the block. It should be noted that MPEG-1 uses subpixel accurate motion vectors instead of a loop filter to smooth the anchor frame. The performance comparison of two methods should be interesting.

The role of coding control includes the rate control, the buffer control, the quantization control, and the frame rate control. These parameters are intimately related. The coding control is not the part of the standard; however, it is an important part of the encoding process. For a given target bit rate, the encoder has to control several parameters to reach the rate target and at the same time provide reasonable coded picture quality.

Since H.261 is a predictive coder and the VLCs are used everywhere, such as coding quantized DCT coefficients and motion vectors, a single transmission error may cause a loss of synchronization and consequently cause problems for the reconstruction. To enhance the performance of the H.261 video coder in noisy environments, the transmitted bitstream of H.261 can optionally contain a BCH (Bose, Chaudhuri, and Hocquengham) (511,493) forward error-correction code.

The H.261 video decoder performs the inverse operations of the encoder. After optional error correction decoding, the compressed bitstream enters the decoder buffer and then is parsed by the variable-length decoder (VLD). The output of the VLD is applied to the IQ and IDCT where the data are converted to the values in the spatial domain. For the interframe-coding mode, the motion

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

FIGURE 19.2 Arrangement of macroblocks in a GOB. (From ITU-T Recommendation H.261, March 1993. With permission.)

compensation is performed and the data from the macroblocks in the anchor frame are added to the current data to form the reconstructed data.

19.2.3 Syntax Description

The syntax of H.261 video coding has a hierarchical layered structure. From the top to the bottom the layers are picture layer, GOB layer, macroblock layer, and block layer.

19.2.3.1 Picture Layer

The picture layer begins with a 20-bit picture start code (PSC). Following the PSC, there are temporal reference (5-bit), picture type information (PTYPE, 6-bit), extra insertion information (PEI, 1-bit), and spare information (PSPARE). Then the data for GOBs are followed.

19.2.3.2 GOB Layer

A GOB corresponds to 176 pixels by 48 lines of Y and 88 pixels by 24 lines of C_b and C_r. The GOB layer contains the following data in order: 16-bit GOB start code (GBSC), 4-bit group number (GN), 5-bit quantization information (GQUANT), 1-bit extra insertion information (GEI), and spare information (GSPARE). The number of bits for GSPARE is variable depending on the set of GEI bits. If GEI is set to "1," then 9 bits follow, consisting of 8 bits of data and another GEI bit to indicate whether a further 9 bits follow, and so on. Data of the GOB header are then followed by data for macroblocks.

19.2.3.3 Macroblock Layer

Each GOB contains 33 macroblocks, which are arranged as in Figure 19.2. A macroblock consists of 16 pixels by 16 lines of Y that spatially correspond to 8 pixels by 8 lines each of C_b and C_r. Data in the bitstream for a macroblock consist of a macroblock header followed by data for blocks. The macroblock header may include macroblock address (MBA) (variable length), type information (MTYPE) (variable length), quantizer (MQUANT) (5 bits), motion vector data (MVD) (variable length), and coded block pattern (CBP) (variable length). The MBA information is always present and is coded by VLC. The VLC table for macroblock addressing is shown in Table 19.1. The presence of other items depends on macroblock type information, which is shown in the VLC Table 19.2.

19.2.3.4 Block Layer

Data in the block layer consists of the transformed coefficients followed by an end of block (EOB) marker (10 bits). The data of transform coefficients (TCOEFF) is first converted to the pairs of RUN and LEVEL according to the zigzag scanning order. The RUN represents the number of successive zeros and the LEVEL represents the value of nonzero coefficients. The pairs of RUN and LEVEL are then encoded with VLCs. The DC coefficient of an intrablock is coded by a fixed-length code with 8 bits. All VLC tables can be found in the standard document (ITU-T, 1993).

TABLE 19.1
VLC Table for Macroblock Addressing

MBA	Code	MBA	Code	MBA	Code
1	1	13	0000 1000	25	0000 0100 000
2	011	14	0000 0111	26	0000 0011 111
3	010	15	0000 0110	27	0000 0011 110
4	0011	16	0000 0101 11	28	0000 0011 101
5	0010	17	0000 0101 10	29	0000 0011 100
6	0001 1	18	0000 0101 01	30	0000 0011 011
7	0001 0	19	0000 0101 00	31	0000 0011 010
8	0000 111	20	0000 0100 11	32	0000 0011 001
9	0000 110	21	0000 0100 10	33	0000 0011 000
10	0000 1011	22	0000 0100 011	MBA stuffing	0000 0001 111
11	0000 1010	23	0000 0100 010	Start code	0000 0000 0000 0001
12	0000 1001	24	0000 0100 001		

TABLE 19.2
VLC Table for Macroblock Type

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter+MC		x			0000 0000 1
Inter+MC		x	x	x	0000 0001
Inter+MC	x	x	x	x	0000 0000 01
Inter+MC+FIL		x			001
Inter+MC+FIL		x	x	x	01
Inter+MC+FIL	x	x	x	x	0000 01

Notes:
1. "x" means that the item is present in the macroblock.
2. It is possible to apply the filter in a non-motion-compensated macroblock by declaring it as MC+FIL but with a zero vector.

19.3 H.263 VIDEO-CODING STANDARD

The H.263 video-coding standard (ITU-T, 1996) is specifically designed for very low bit rate applications such as practical video telecommunication. Its technical content was completed in late 1995 and the standard was approved in early 1996.

19.3.1 Overview of H.263 Video Coding

The basic configuration of the video source coding algorithm of H.263 is based on the H.261. Several important features that are different from H.261 include the following new options: unrestricted motion vectors, syntax-based arithmetic coding, advanced prediction, and PB-frames. All these features can be used together or separately for improving the coding efficiency. The H.263

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.