6.4

MPEG-1 and MPEG-2 Video Standards

Supavadee Aramvith
Chulalongkorn University, Thailand

Ming-Ting Sun
University of Washington, Seattle

1 MPEG-1 Video Coding Standard

1.1 Introduction

1.1.1 Background and Structure of MPEG-1 Standards Activities

The development of digital video technology in the 1980s has made it possible to use digital video compression in various kinds of applications. The effort to develop standards for coded representation of moving pictures, audio, and their combination is carried out in the Moving Picture Experts Group (MPEG). MPEG is a group formed under the auspices of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). It operates in the framework of the Joint ISO/IEC Technical Committee 1 (JTC 1) on Information Technology, which was formally Working Group 11 (WG11) of Sub-Committee 29 (SC29). The premise is to set the standard for coding moving pictures and the associated audio for digital storage media at about 1.5 Mbit/s so that a movie can be compressed and stored in a VCD (video compact disc). The resultant standard is the international standard for moving picture compression, ISO/IEC 11172 or MPEG-1 (Moving Picture Experts Group-Phase 1). MPEG-1 standards consist of 5 parts, including: systems (11172-1), video (11172-2), audio (11172-3), conformance testing (11172-4), and software simulation (11172-5). In this chapter, we will focus only on the video part.

The activity of the MPEG committee started in 1988 based on the work of ISO JPEG (Joint Photographic Experts Group) [1] and ITU-T (formerly CCITT) Recommendation H.261: "Video Codec for Audiovisual Services at p×64 kbits/s" [2]. Thus, the MPEG-1 standard has much in common with the JPEG and H.261 standards. The MPEG development methodology is similar to that of H.261 and is divided into three phases: requirements, competition, and convergence [3]. The purpose of the requirements phase is to precisely set the focus of the effort and determine the rule for the competition phase. The document of this phase is a "Proposal Package Description" [4] and a test methodology [5]. The next step is the competition phase in which the goal is to obtain state of the art technology from the best of academic and industrial research. The criteria are based on the technical merits and the trade-off between video quality and the cost of implementation [5]. After the competition phase, various ideas and techniques are integrated into one solution in the convergence phase. The solution results in a simulation model, which implements a reference encoder and a decoder. The simulation model is used to carry out simulations to optimize the performance of the coding scheme [6]. A series of fully documented experiments called core experiments are then carried out. The MPEG committee reached the Committee Draft (CD) status in September 1990 and the Committee Draft (CD 11172) was approved in December 1991. International Standard (IS) 11172 for the first three parts was established in November 1992. The IS for the last two parts was finalized in November 1994.

1.1.2 MPEG-1 Target Applications and Requirements

The MPEG standard is a generic standard, which means that it is not limited to a particular application. A variety of digital storage media applications of MPEG-1 have been proposed based on the assumptions that acceptable video and audio quality can be obtained for a total bandwidth of about 1.5 Mbits/s. Typical storage media for these applications include VCD, DAT (digital audio tape), Winchester-type computer disks, and writable optical disks. The target applications are asymmetric applications where the compression process is performed once and the decompression process is required often. Examples of the asymmetric applications include VCD, video on demand, and video games. In these asymmetric applications, the encoding delay is not a concern. The encoders are needed only in small quantities while the decoders are needed in large volumes. Thus, the encoder complexity is not a concern while the decoder complexity needs to be low in order to result in low-cost decoders.

The requirements for compressed video in digital storage media mandate several important features of the MPEG-1 compression algorithm. The important features include normal playback, frame-based random access and editing of video, reverse playback, fast forward/reverse play, encoding high-resolution still frames, robustness to uncorrectable errors, etc. The applications also require MPEG-1 to support flexible picture-sizes and frame-rates. Another requirement is that the encoding process can be performed in reasonable speed using existing hardware technologies and the decoder can be implemented in low cost.

Since MPEG-1 video coding algorithm was developed based on H.261, in the following sections, we will focus only on those parts which are different from H.261.

1.2 MPEG-1 Video Coding vs. H.261

1.2.1 Bi-Directional Motion Compensated Prediction

In H.261, only the previous video frame is used as the reference frame for the motion compensated prediction (forward prediction). MPEG-1 allows the future frame to be used as the reference frame for the motion compensated prediction (backward prediction), which can provide better prediction. For example, as shown in Fig. 1, if there are moving objects, and if only the forward prediction is used, there will be uncovered areas (such as the block behind the car in frame N) for which we may not be able to find a good matching block from the previous reference picture (frame N−1). On the other hand, the backward prediction can properly predict these uncovered areas since they are available in the future reference picture, i.e., frame N+1 in this example. Also shown in the figure, if there are objects moving into the picture (the airplane in the figure), these new objects cannot be predicted from the previous picture, but can be predicted from the future picture. In fact, all the information in the second picture is available from the first and the third picture. Another major advantage of the B-pictures is the denoising capability. In practical situations, the pixel values of an object may not be same, due to various noise effects from lighting changes, shadows, sampling effects, and other noises. Bi-directional prediction could reduce the noise effects due to averaging, or simply due to the fact that there is an extra choice which could provide a better matching.

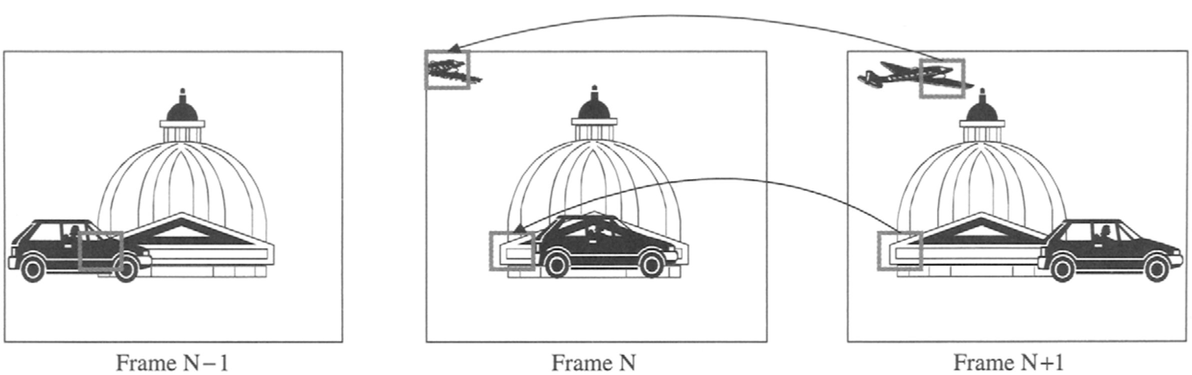

FIGURE 1 A video sequence showing the benefits of bi-directional prediction.

1.2.2 Motion Compensated Prediction with Half-Pixel Accuracy

The motion estimation in H.261 is restricted to only integer-pixel accuracy. However, a moving object often moves to a position which is not on the pixel grid but between the pixels. MPEG-1 allows half-pixel accuracy motion vectors. By estimating the displacement at a finer resolution, we can expect improved prediction and, thus, better performance than motion estimation with integer pixel accuracy. As shown in Fig. 2, since there is no pixel value at the half-pixel locations, interpolation is required to produce the pixel values at the half-pixel positions. Bi-linear interpolation is used in MPEG-1 for its simplicity. As in H.261, the motion estimation is performed only on luminance blocks. The resulting motion vector is scaled by 2 and applied to the chrominance blocks. Motion vectors are differentially encoded with respect to the motion vector in the preceding adjacent macroblock. The reason is that the motion vectors of adjacent regions are highly correlated, as it is quite common to have relatively uniform motion over areas of the picture.

1.3 MPEG-1 Video Structure

1.3.1 Source Input Format (SIF)

The typical MPEG-1 input format is the source input format (SIF). SIF is derived from ITU-R BT 601, formerly CCIR601, a worldwide standard for digital TV studio. ITU-R BT 601 specifies the Y Cb Cr color coordinate where Y is the luminance component (black and white information), and Cb and Cr are two color difference signals (chrominance components). A luminance sampling frequency of 13.5 MHz was adopted. There are several Y Cb Cr sampling formats, such as 4:4:4, 4:2:2, 4:1:1, and 4:2:0. In 4:4:4, the sampling rates for Y, Cb, and Cr are the same. In 4:2:2, the sampling rates of Cb and Cr are half of that of Y. In 4:1:1 and 4:2:0, the sampling rates of Cb and Cr are one quarter of that of Y. The positions of Y Cb Cr samples for 4:4:4, 4:2:2, 4:1:1, and 4:2:0 are shown in Fig. 3.

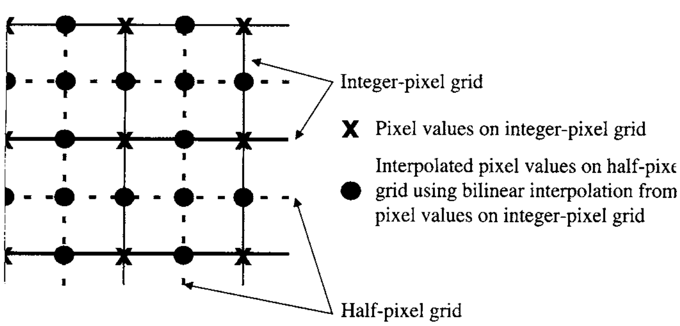

FIGURE 2 Half-pixel motion estimation.

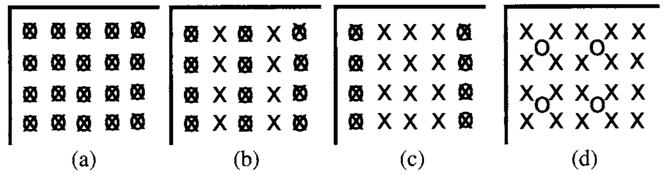

FIGURE 3 Luminance and chrominance samples in (a) 4:4:4 format (b) 4:2:2 format (c) 4:1:1 format (d) 4:2:0 format.

Converting an analog TV signal to digital video with the 13.5 MHz sampling rate of ITU-R BT 601 results in 720 active pixels per line (576 active lines for PAL and 480 active lines for NTSC). This results in a 720×480 resolution for NTSC and a 720×576 resolution for PAL. With 4:2:2, the uncompressed bit-rate for transmitting ITU-R BT 601 at 30 frames/s is then about 166 Mbits/s. Since it is difficult to compress an ITU-R BT 601 video to 1.5 Mb/s with good video quality, in MPEG-1, typically the source video resolution is decimated to a quarter of the ITU-R BT 601 resolution by filtering and sub-sampling. The resultant format is called SIF, which has a 360×240 resolution for NTSC and a 360×288 resolution for PAL. Since in the video coding algorithm, the block-size of 16×16 is used for motion compensated prediction, the number of pixels in both the horizontal and the vertical dimensions should be multiples of 16. Thus, the four left-most and four right-most pixels are discarded to give a 352×240 resolution for NTSC systems (30 frames/s) and a 352×288 resolution for PAL systems (25 frames/s). The chrominance signals have half of the above resolutions in both the horizontal and vertical dimensions (4:2:0, 176×120 for NTSC and 176×144 for PAL). The uncompressed bit rate for SIF (NTSC) at 30 frames/s is about 30.4 Mbits/s.

1.3.2 Group of Pictures (GOPs) and I-B-P Pictures

In MPEG, each video sequence is divided into one or more groups of pictures (GOPs). There are four types of pictures defined in MPEG-1: I-, P-, B-, and D-pictures of which the first three are shown in Fig. 4. Each GOP is composed of one or more pictures; one of these pictures must be an I-picture. Usually, the spacing between two anchor frames (I- or

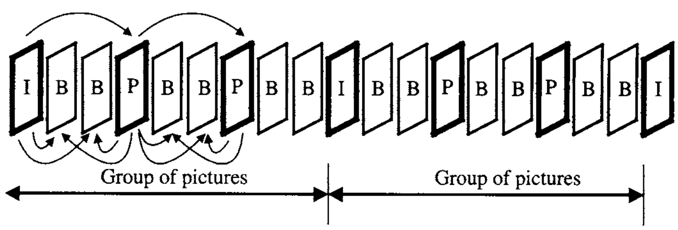

FIGURE 4 MPEG group of pictures.

P-pictures) is referred to as M, and the spacing between two successive I-pictures is referred to as N. In Fig. 4, M = 3 and N = 9. However, it should be noted that a GOP does not need to use a periodical structure.

I-pictures (intra-coded pictures) are coded independently with no reference to other pictures. I-pictures provide random access points in the compressed video data, since the I-pictures can be decoded independently without referencing to other pictures. With I-pictures, an MPEG bit stream is more editable. Also, error propagation due to transmission errors in previous pictures will be terminated by an I-picture since the I-picture does not reference to the previous pictures. Since I-pictures use only transform coding without motion compensated predictive coding, it provides only moderate compression.

P-pictures (predictive-coded pictures) are coded using the forward motion-compensated prediction similar to that in H.261 from the preceding I- or P-picture. P-pictures provide more compression than the I-pictures by virtue of motion-compensated prediction. They also serve as references for B-pictures and future P-pictures. Transmission errors in the I-pictures and P-pictures can propagate to the succeeding pictures since the I-pictures and P-pictures are used to predict the succeeding pictures.

B-pictures (bi-directional predicted pictures) allow macroblocks to be coded using bi-directional motion-compensated prediction from both the past and future reference I- or P-pictures. In the B-pictures, each bi-directional motion-compensated macroblock can have two motion vectors: a forward motion vector which references to a best matching block in the previous I- or P-pictures, and a backward motion vector which references to a best matching block in the next I- or P-pictures as shown in Fig. 5. The motion compensated prediction can be formed by the average of the two referenced motion compensated blocks. By averaging between the past and the future reference blocks, the effect of noise can be decreased. B-pictures provide the best compression compared to I- and P-pictures. I- and P-pictures are used as reference pictures for predicting B-pictures. To keep the structure simple, the B-pictures are not used as reference pictures. Hence, B-pictures do not propagate errors.

D-pictures (dc-pictures) are low-resolution pictures obtained by decoding only the dc coefficient of the discrete cosine transform (DCT) coefficients of each macroblock. They are not used in combination with I-, P-, or B-pictures. D-pictures are rarely used, but are defined to allow fast searches on sequential digital storage media.

The trade-off of having frequent B-pictures is that it decreases the correlation between the previous I- or P-picture and the next reference P- or I-picture. It also causes coding delay and increases the encoder complexity. With the example shown in Figs. 4 and 6, at the encoder, if the order of the incoming pictures is 1, 2, 3, 4, 5, 6, 7, ..., the order of coding the pictures at the encoder will be: 1, 4, 2, 3, 7, 5, 6, ... At the decoder, the order of the decoded pictures will also be 1, 4, 2, 3, 7, 5, 6, ... However, the display order after the decoder should be 1, 2, 3, 4, 5, 6, 7. Thus, frame memories have to be used to put the pictures in the correct order. This picture re-ordering causes delay. The computation and extra memory requirement of bi-directional motion estimation, and the picture re-ordering frame memories increase the encoder complexity.

Encoder Input:

1I 2B 3B 4P 5B 6B 7P 8B 9B 10I 11B 12B 13P 14B 15B 16P

GOP1 GOP2

Decoder Input:

1I 4P 2B 3B 7P 5B 6B 10I 8B 9B 13P 11B 12B 16P 14B 15B

GOP1 GOP2

CLOSED OPEN

FIGURE 6 Frame reordering.

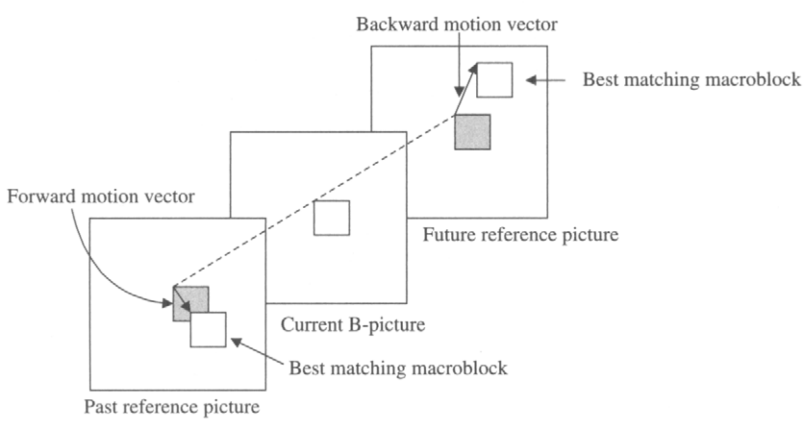

FIGURE 5 Bi-directional motion estimation.

In Fig. 6, two types of GOPs are shown. GOP1 can be decoded without referencing other GOPs. It is called a Closed-GOP. In GOP2, to decode the eighth B- and ninth B-pictures, the seventh P-picture in GOP1 is needed. GOP2 is called an Open GOP which means the decoding of this GOP needs to reference other GOPs.

1.3.3 Slice, Macroblock, and Block Structures

An MPEG picture consists of slices. A slice consists of a contiguous sequence of macroblocks in a raster scan order (from left to right and from top to bottom). In an MPEG coded bit stream, each slice starts with a slice header which is a clear codeword (a clear codeword is a unique bit pattern which can be identified without decoding the variable length codes in the bit stream). Due to the clear codeword slice header, slices are the lowest level of units which can be accessed in an MPEG coded bit stream without decoding the variable-length codes. Slices are important in the handling of channel errors. If a bit-stream contains a bit error, the error may cause error propagation due to the variable length coding. The decoder can regain synchronization at the start of the next slice. Having more slices in a bit stream allows better error-termination, but the overhead will increase.

A macroblock consists of a 16×16 block of luminance samples and two 8×8 blocks of corresponding chrominance samples as shown in Fig. 7. A macroblock thus consists of four 8×8 Y-blocks, one 8×8 Cb block, and one 8×8 Cr block. Each coded macroblock contains motion-compensated prediction information (coded motion vectors and the prediction errors). There are four types of macroblocks: intra, forward-predicted, backward-predicted, and averaged macroblocks. The motion information consists of one motion vector for forward- and backward-predicted macroblocks and two motion vectors for bi-directionally predicted (or averaged) macroblocks. P-pictures can have intra- and forward-predicted macroblocks. B-pictures can have all four types of macroblocks. The first and last macroblocks in a slice must always be coded. A macroblock is designated as a skipped macroblock when its motion vector is zero and all the quantized DCT coefficients are zero. Skipped macroblocks are not allowed in I-pictures. Non intracoded macroblocks in P- and B-pictures can be skipped. For a skipped macroblock, the decoder just copies the pixel values of the macroblock from the previous picture.

1.4 Summary of the Major Differences between MPEG-1 Video and H.261

As compared to H.261, MPEG-1 video differs in the following aspects:

- MPEG-1 uses bi-directional motion compensated predictive coding with half-pixel accuracy while H.261 has no bi-directional prediction (B-pictures) and the motion vectors are always in integer-pixel accuracy.
- MPEG-1 supports the maximum motion vector range of -512 to $+511.5$ pixels for half-pixel motion vectors and -1024 to $+1023$ for integer-pixel motion vectors while H.261 has a maximum range of only ± 15 pixels.
- MPEG-1 uses visually weighted quantization based on the fact that the human eye is more sensitive to quantization errors related to low spatial frequencies than to high spatial frequencies. MPEG-1 defines a default 64-element quantization matrix, but also allows custom matrices appropriate for different applications. H.261 has only one quantizer for the intra DC coefficient and 31 quantizers for all other coefficients.

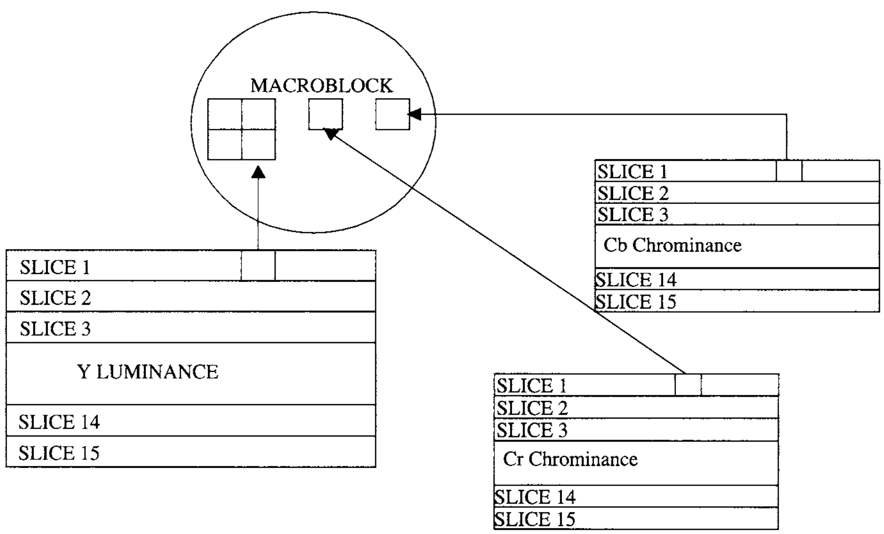

FIGURE 7 Macroblock and slice structures.

TABLE 1 MPEG-1 constrained parameter set

Parameter	Constraint
Horizontal size	$\leq$720 pels
Vertical size	$\leq$576 pels
Total no. of macroblocks/picture	$\leq$396
Total no. of macroblocks/second	$\leq 396 \times 25 = 330 \times 30$
Picture rate	$\leq$30 frames/s
Bit rate	$\leq$1.86 Mbits/s
Decoder buffer	$\leq$376832 bits

- H.261 only specifies two source formats: CIF (common intermediate format, 352 × 288 pixels) and QCIF (quarter CIF, 176 × 144 pixels). In MPEG-1, the typical source format is SIF (352 × 240 for NTSC, and 352 × 288 for PAL). However, the users can specify other formats. The picture size can be as large as 4k × 4k pixels. There are certain parameters in the bit streams that are left flexible, such as the number of lines per picture (less than 4096), the number of pels per line (less than 4096), picture rate (24, 25, and 30 frames/s), and fourteen choices of pel aspect ratios.
- In MPEG-1, I-, P-, and B-pictures are organized as a flexible group of pictures (GOPs).
- MPEG-1 uses a flexible slice structure instead of group of blocks (GOBs) as defined in H.261.
- MPEG-1 has D-pictures to allow the fast-search option.
- In order to allow cost effective implementation of user terminals, MPEG-1 defines a constrained parameter set which lays down specific constraints, as listed in Table 1.

1.5 Simulation Model

Similar to H.261, MPEG-1 specifies only the syntax and the decoder. Many detailed coding options such as the rate control strategy, the quantization decision levels, the motion estimation schemes, and coding modes for each macroblock are not specified. This allows future technology improvement and product differentiation. In order to have a reference MPEG-1 video quality, simulation models were developed in MPEG-1. A simulation model contains a specific reference implementation of the MPEG-1 encoder and decoder including all the details which are not specified in the standard. The final version of the MPEG-1 simulation model is "Simulation Model 3" (SM3) [7]. In SM3, the motion estimation technique uses one forward and/or one backward motion vector per macroblock with half-pixel accuracy. A two-step search scheme which consists of a full search in the range of +/− seven pixels with the integer pixel precision, followed by a search in eight neighboring half-pixel positions, is used. The decision of the coding mode for each macroblock (whether or not it will use motion compensated prediction and intra-inter-coding), the quantizer decision levels, and the rate control algorithm are all specified in the simulation model.

1.6 MPEG-1 Video Bit Stream Structures

As shown in Fig. 8, there are six layers in the MPEG-1 video bit stream: the video sequence, group of pictures, picture, slice, macroblock, and block layers.

- A video sequence layer consists of a sequence header, one or more groups of pictures, and an end-of-sequence code. It contains the setting of the following parameters: the picture size (horizontal and vertical sizes), pel aspect ratio, picture rate, bit rate, the minimum decoder buffer size (video buffer verifier size), constraint parameters flag (this flag is set only when the picture size, picture rate, decoder buffer size, bit rate, and motion parameters satisfy the constraints bound in Table 1), the control for

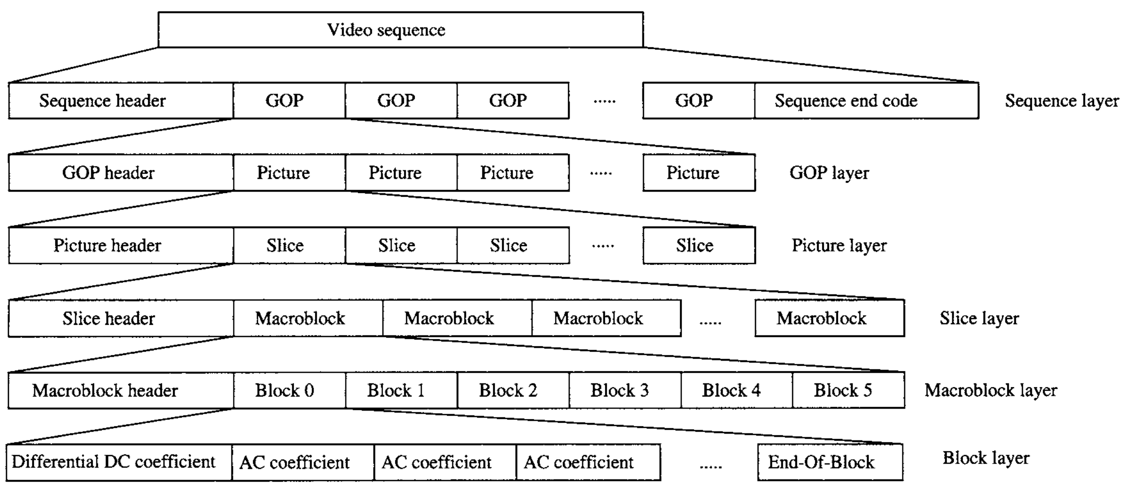

FIGURE 8 MPEG-1 bit-stream syntax layers.

the loading of 64 8-bit values for intra and non-intra quantization tables, and the user data.

- The GOP layer consists of a set of pictures that are in a continuous display order. It contains the setting of the following parameters: the time code which gives the hours-minutes-seconds time interval from the start of the sequence, the closed GOP flag which indicates whether the decoding operation needs pictures from the previous GOP for motion compensation, the broken link flag which indicated whether the previous GOP can be used to decode the current GOP, and the user data.
- The picture layer acts as a primary coding unit. It contains the setting of the following parameters: the temporal reference which is the picture number in the sequence and is used to determine the display order, the picture types (I/P/B/D), the decoder buffer initial occupancy which gives the number of bits that must be in the compressed video buffer before the idealized decoder model defined by MPEG decodes the picture (it is used to prevent the decoder buffer overflow and underflow), the forward motion vector resolution and range for P- and B-pictures, the backward motion vector resolution and range for B-pictures, and the user data.
- The slice layer acts as a resynchronization unit. It contains the slice vertical position where the slice starts, and the quantizer scale that is used in the coding of the current slice.
- The macroblock layer acts as a motion compensation unit. It contains the setting of the following parameters: the optional stuffing bits, the macroblock address increment, the macroblock type, quantizer scale, motion vector, and the coded block pattern which defines the coding patterns of the six blocks in the macroblock.
- The block layer is the lowest layer of the video sequence and consists of coded 8×8 DCT coefficients. When a macroblock is encoded in the intra-mode, the DC coefficient is encoded similar to that in JPEG (the DC coefficient of the current macroblock is predicted from the DC coefficient of the previous macroblock). At the beginning of each slice, predictions for DC coefficients for luminance and chrominance blocks are reset to 1024. The differential DC values are categorized according to their absolute values and the category information is encoded using VLC (variable length code). The category information indicates the number of additional bits following the VLC to represent the prediction residual. The AC coefficients are encoded similar to that in H.261 using a VLC to represent the zero-run-length and the value of the non zero coefficient. When a macroblock is encoded in non intra modes, both the DC and AC coefficients are encoded similar to that in H.261.

Above the video sequence layer, there is a system layer in which the video sequence is packetized. The video and audio bit streams are then multiplexed into an integrated data stream. These are defined in the Systems part.

1.7 Summary

MPEG-1 is mainly for storage media applications. Due to the use of B-picture, it may result in long end-to-end delay. The MPEG-1 encoder is much more expensive than the decoder due to the large search range, the half-pixel accuracy in motion estimation, and the use of the bi-directional motion estimation. The MPEG-1 syntax can support a variety of frame-rates and formats for various storage media applications. Similar to other video coding standards, MPEG-1 does not specify every coding option (motion estimation, rate control, coding modes, quantization, pre-processing, post-processing, etc.). This allows continuing technology improvement and product differentiation.

2 MPEG-2 Video Coding Standard

2.1 Introduction

2.1.1 Background and Structure of MPEG-2 Standards Activities

The MPEG-2 standard represents the continuing efforts of the MPEG committee to develop generic video and audio coding standards after their development of MPEG-1. The idea of this second phase of MPEG work came from the fact that MPEG-1 is optimized for applications at about 1.5 Mb/s with input source in SIF, which is a relatively low-resolution progressive format. Many higher quality higher bit rate applications require a higher resolution digital video source such as ITU-R BT 601, which is an interlaced format. New techniques can be developed to code the interlaced video better.

The MPEG-2 committee started working in late 1990 after the completion of the technical work of MPEG-1. The competitive tests of video algorithms were held in November 1991, followed by the collaborative phase. The Committee Draft (CD) for the video part was achieved in November 1993. The MPEG-2 standard (ISO/IEC 13818) [8] consists of nine parts. The first five parts are organized in the same fashion as MPEG-1: systems, video, audio, conformance testing, and simulation software technical report. The first three parts of MPEG-2 reached International Standard (IS) status in November 1994. Parts 4 and 5 were approved in March 1996. Part 6 of the MPEG-2 standard specifies a full set of Digital Storage Media Control Commands (DSM-CC). Part 7 is the specification of Advanced Audio Coding (AAC). Part 8 was originally planned to be the coding of 10-bit video but was discontinued. Part 9 is the specification of real-time interface (RTI) to transport stream decoders which may be utilized for adaptation to all appropriate networks carrying MPEG-2 transport streams. Part 10 is the specification of conformance testing part of DSM-CC. Part 6 and Part 9 have already been approved as International Standards in July 1996. Like the MPEG-1 video standard, MPEG-2 video coding standard specifies only the bit stream syntax and the semantics of the decoding process. Many encoding options were left unspecified to encourage continuing technology improvement and product differentiation.

MPEG-3, which was originally intended for HDTV (high definition digital television) at higher bit-rates, was merged with MPEG-2. Hence there is no MPEG-3. MPEG-2 video coding standard (ISO/IEC 13818-2) was also adopted by ITU-T as ITU-T Recommendation H.262 [9].

2.1.2 Target Applications and Requirements

MPEG-2 is primarily targeted at coding high-quality video at 4–15 Mb/s for video on demand (VOD), digital broadcast television, and Digital Storage Media such as DVD (digital versatile disc). It is also used for coding HDTV, cable/satellite digital TV, video services over various networks, 2-way communications, and other high-quality digital video applications.

The requirements from MPEG-2 applications mandate several important features of the compression algorithm. Regarding picture quality, MPEG-2 needs to be able to provide good NTSC quality video at a bit rate of about 4–6 Mbits/s and transparent NTSC quality video at a bit rate of about 8–10 Mbits/s. It also needs to provide the capability of random access and quick channel-switching by means of I-pictures in GOPs. Low-delay mode is specified for delay-sensitive visual communications applications. MPEG-2 has scalable coding modes in order to support multiple grades of video quality, spatial resolutions, and frame-rates for various applications. Error resilience options include intramotion vector, data partitioning, and scalable coding. Compatibility with the existing MPEG-1 video standard is another prominent feature provided by MPEG-2. For example, MPEG-2 decoders should be able to decode MPEG-1 bit streams. If scalable coding is used, the base layer of MPEG-2 signals can be decoded by a MPEG-1 decoder. Finally, it should allow reasonable complexity encoders and low-cost decoders be built with mature technology. Since MPEG-2 video is based heavily on MPEG-1, in the following sections, we will focus only on those features which are different from MPEG-1 video.

2.2 MPEG-2 Profiles and Levels

MPEG-2 standard is designed to cover a wide range of applications. However, features needed for some applications may not be needed for other applications. If we put all the features into one single standard, it may result in an overly expensive system for many applications. It is desirable for an application to implement only the necessary features to lower the cost of the system. To meet this need, MPEG-2 classified the groups of features for important applications into profiles. A profile is defined as a specific subset of the MPEG-2 bit-stream syntax and functionality to support a class of applications (e.g., low-delay video conferencing applications, or storage media applications). Within each profile, levels are defined to support applications which have different quality requirements (e.g., different resolutions). Levels are specified as a set of restrictions on some of the parameters (or their combination) such as sampling rates, frame resolutions, and bit rates in a profile. Applications are implemented in the allowed range of values of a particular profile at a particular level.

Table 2 shows the combination of profiles and levels that are defined in MPEG-2. MPEG-2 defines seven distinct profiles: simple, main, snr scalable, spatially scalable, high, 4:2:2, and multiview. The last two profiles were developed after the final approval of MPEG-2 video in November 1994. Simple profile is defined for low-delay video conferencing applications using only I- and P-pictures. Main profile is the most important and widely used profile for general high-quality digital video applications such as VOD, DVD, digital TV, and HDTV. SNR (signal-to-noise ratio) scalable profile supports multiple grades of video quality. Spatially scalable profile supports multiple grades of resolutions. High profile supports multiple grades of quality, resolution, and chroma formats. Four levels are defined within the profiles: low (for SIF resolution pictures), main (for ITU-R BT 601 resolution pictures), high-1440 (for European HDTV resolution pictures), and high (for North America HDTV resolution pictures). The 11 combinations of profiles and levels in Table 2 define the MPEG-2 conformance points which cover most practical MPEG-2 target applications. The numbers in each conformance point indicate the maximum bound of the parameters. The number in the first line indicates the luminance rate in samples/s. The number in the second line indicates bit rate in bits/s. Each conformance point is a subset of the conformance point at the right or above. For example, a main profile main level decoder should also decode simple profile main level and main profile low level bit streams. Among the defined profiles and levels, main profile at main level (MP@ML) is used for digital television broadcast in ITU-R BT 601 resolution and DVD video. The main profile at high level (MP@HL) is used for HDTV. The 4:2:2 profile at main level (422P@ML) is defined for professional video production environments, which supports a higher bit-rate of up to 50 Mbit/s with 4:2:2 color subsampling, and higher precision in DCT coding. Although the high profile supports 4:2:2 also, a high profile codec needs to support SNR scalable profile and spatially scalable profile. This makes the high profile codec expensive. The 4:2:2 profile does not need to support the scalabilities and thus will be much cheaper to implement. Multiview profile is defined to support the efficient encoding for the applications involving two video sequences from

TABLE 2 Profiles and levels

Level	Profile: Simple 4:2:0	Main 4:2:0	SNR Scalable 4:2:0	Spatially Scalable 4:2:0	High 4:2:0 or 4:2:2
High 1920 × 1152 (60 frames/s)		62.7 Ms/s 80 Mbits/s			100 Mbit/s for 3 layers
High-1440 1440 × 1152 (60 frames/s)		47 Ms/s 60 Mbit/s		47 Ms/s 60 Mbit/s for 3 layers	80 Mbits/s for 3 layers
Main 720 × 576 (30 frames/s)	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s for 2 layers		20 Mbit/s for 3 layers
Low 352 × 288 (30 frames/s)		3.04 Ms/s 4 Mbit/s	3.04 Ms/s 4 Mbit/s for 2 layers		

two cameras shooting the same scene with a small angle between them.

2.3 MPEG-2 Video Input Resolutions and Formats

Although the main concern of the MPEG-2 committee is to support the ITU-R BT 601 resolution which is the digital TV resolution, MPEG-2 allows a maximum picture size of 16k × 16k pixels. It also supports the frame rates of 23.976, 24, 25, 29.97, 30, 50, 59.94 and 60 Hz as in MPEG-1. MPEG-2 is suitable for coding the progressive video format as well as the interlaced video format. As for the color subsampling formats, MPEG-2 supports 4:2:0, 4:2:2, and 4:4:4. MPEG-2 uses the 4:2:0 format as in MPEG-1 except that there is a difference in the positions of the chrominance samples as shown in Figs. 9(a) and 9(b).

In MPEG-1, a slice can cross macroblock row boundaries. Therefore, a single slice in MPEG-1 can be defined to cover the entire picture. On the other hand, slices in MPEG-2 begin and end in the same horizontal row of macroblocks. There are two types of slice structure in MPEG-2: the general and the restricted slice structures. In the general slice structure, MPEG-2 slices need not cover the entire picture. Thus, only the regions enclosed in the slices are encoded. In the restricted slice structure, every macroblock in the picture shall be enclosed in a slice.

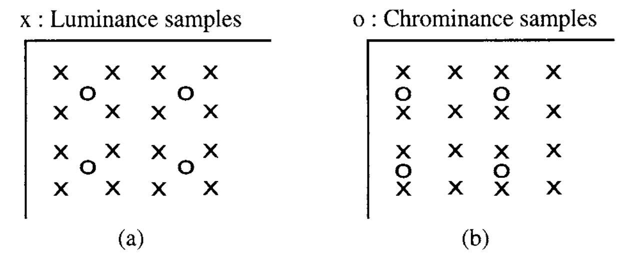

FIGURE 9 The position of luminance and chrominance samples for 4:2:0 format in (a) MPEG-1 (b) MPEG-2.

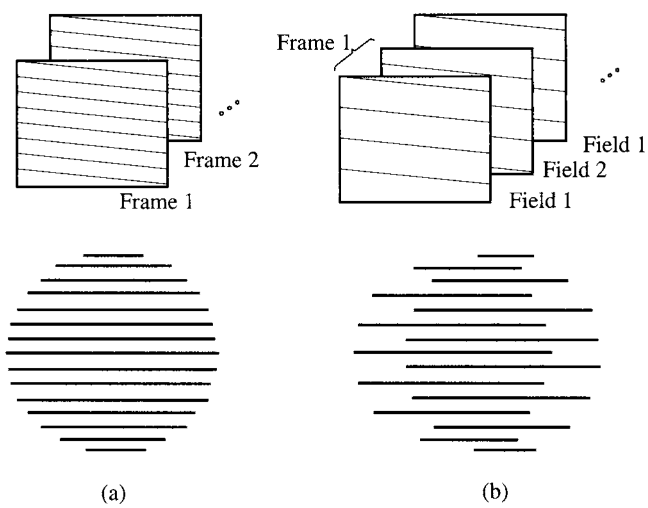

FIGURE 10 (a) Progressive scan. (b) Interlaced scan.

2.4 MPEG-2 Video Coding Standard Compared to MPEG-1

2.4.1 Interlaced vs. Progressive Video

Figure 10 shows the progressive and interlaced video scan. In the interlaced video, each displayed frame consists of two interlaced fields. For example, frame 1 consists of field 1 and field 2, with the scanning lines in field 1 located between the lines of field 2. On the contrary, the progressive video has all the lines of a picture displayed in one frame. There are no fields or half pictures as with the interlaced scan. Thus, progressive video requires a higher picture rate than the frame rate of an interlaced video, to avoid a flickery display.

The main disadvantage of the Interlaced format is that when there are object movements, the moving object may appear distorted when we merge two fields into a frame. For example, Fig. 10 shows a moving ball. In the interlaced format, since the moving ball will be at different locations in the two fields, when we put the two fields into a frame, the ball will look distorted. Using MPEG-1 to encode the distorted objects in the frames of the interlaced video will not produce the optimal results.

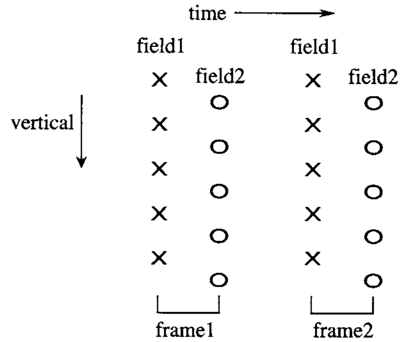

FIGURE 11 Interlaced video format.

2.4.2 Interlaced Video Coding

Figure 11 shows the interlaced video format. As explained earlier, an interlaced frame is composed of two fields. From the figure, the top field (Field 1) occurs earlier in time than the bottom field (Field 2). Both fields together form a frame. In MPEG-2, Pictures are coded as I-, P-, and B-pictures like in MPEG-1. To optimally encode the interlaced video, MPEG-2 can encode a picture either as a field picture or a frame picture. In the field picture mode, the two fields in the frame are encoded separately. If the first field in a picture is an I-picture, the second field in the picture can be either I- or P-pictures as the second field can use the first field as a reference picture. However, if the first field in a picture is a P- or B-field picture, the second field has to be the same type of picture. In a frame-picture, two fields are interleaved into a picture and coded together as one picture. In MPEG-2, a video sequence is a collection of frame pictures and field pictures.

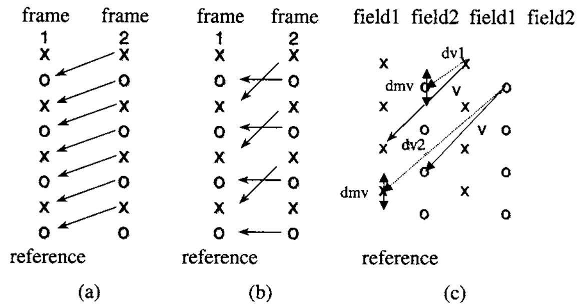

FIGURE 12 Three types of motion compensated prediction.

2.4.2.1 Frame-based and Field-based Motion Compensated Prediction. In MPEG-2, an interlaced picture can be encoded as a frame picture or as field pictures. MPEG-2 defines two different motion compensated prediction types: frame-based and field-based motion compensated prediction. Frame-based prediction forms a prediction based on the reference frames. Field based prediction is made based on reference fields. For the simple profile where the bi-directional prediction cannot be used, MPEG-2 introduced a dual-prime motion compensated prediction to efficiently explore the temporal redundancies between fields. Figure 12 shows the three types of motion compensated predictions. Note that all motion vectors in MPEG-2 are specified with a half-pixel resolution.

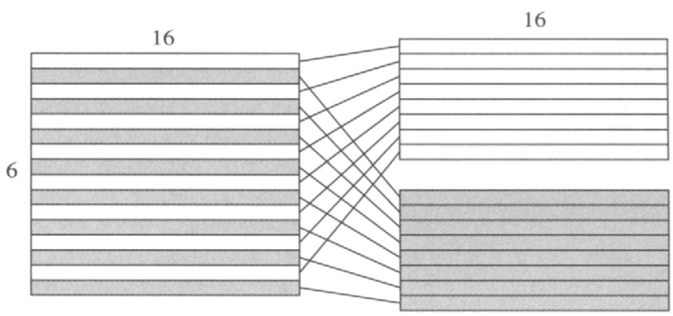

FIGURE 13 Blocks for frame-/field-based prediction.

Frame predictions in frame pictures: in the frame-based prediction for frame-pictures, as shown in Fig. 12(a), the whole interlaced frame is considered as a single picture. It uses the same motion compensated predictive coding method used in MPEG-1. Each 16×16 macroblock can have only one motion vector for each forward or backward prediction. Two motion vectors are allowed in the case of the bi-directional prediction.

Field prediction in frame-pictures: the field-based prediction in frame pictures considers each frame picture as two separate field pictures. Separate predictions are formed for each 16×8 block of the macroblock as shown in Fig. 13. Thus, the field-based prediction in a frame picture needs two sets of motion vectors. A total of four motion vectors are allowed in the case of bi-directional prediction. Each field-prediction may select either the field 1 or the field 2 of the reference frame.

Field prediction in field pictures: in field-based prediction for field pictures, the prediction is formed from the two most recently decoded fields. The predictions are made from reference fields, independently for each field, with each field considered as an independent picture. The block size of prediction is 16×16; however, it should be noted that the 16×16 block in the field picture corresponds to a 16×32 pixel-area in the frame picture. Field-based prediction in field picture needs only one motion vector for each forward- or backward-prediction. Two motion vectors are allowed in the case of the bi-directional prediction.

16×8 prediction in field pictures: two motion vectors are used for each macroblock. The first motion vector is applied to the 16×8 block in field 1 and the second motion vector is applied to the 16×8 block in field 2. A total of four motion vectors are allowed in the case of bi-directional prediction.

Dual-prime motion compensated prediction can be used only in P-pictures. Once the motion vector "v" for a macroblock in a field of given parity (field 1 or field 2) is known relative to a reference field of the same parity, it is extrapolated or interpolated to obtain a prediction of the motion vector for the opposite parity reference field. In addition, a small correction is also made to the vertical component of the motion vectors to reflect the vertical shift between lines of the field 1 and field 2. These derived motion vectors are denoted dv1 and dv2 (represented by dash lines) in Fig. 12(c). Next, a small refinement differential motion vector, called "dmv," is added. The choice of dmv values $(-1, 0, +1)$ is determined by the encoder. The motion vector "v" and its corresponding "dmv" value are included in the bit-stream so that the decoder can also derive dv1 and dv2. In calculating the pixel values of the prediction, the motion compensated predictions from the two reference fields are averaged which tends to reduce the noise in the data.

Dual-prime prediction is mainly for low-delay coding applications such as videophone and video conferencing. For low-delay coding using simple profile, B-pictures should not be used. Without using bi-directional prediction, dual-prime prediction is developed for P-pictures to provide a better prediction than the forward prediction.

2.4.2.2 Frame/Field DCT. MPEG-2 has two DCT modes: frame-based and field-based DCT as shown in Fig. 14. In the frame-based DCT mode, a 16×16-pixel macroblock is divided into four 8×8 DCT blocks. This mode is suitable for the blocks in the background or in a still image that have little motion because these blocks have high correlation between pixel values from adjacent scan lines. In the field-based DCT mode, a macroblock is divided into four DCT blocks where the pixels from the same field are grouped together into one block. This mode is suitable for the blocks that have motion because as explained, motion causes distortion and may introduce high-frequency noises into the interlaced frame.

2.4.2.3 Alternate Scan. MPEG-2 defines two different zigzag scanning orders: zigzag and alternate scans as shown in Fig. 15. The zigzag scan used in MPEG-1 is suitable for progressive images where the frequency components have equal importance in each horizontal and vertical direction. In MPEG-2, an alternate scan is introduced based on the fact that interlaced images tend to have higher frequency components in the vertical direction. Thus, the scanning order weighs more on the higher vertical frequencies than the same horizontal frequencies. In MPEG-2, the selection between these two zigzag scan orders can be made on a picture basis.

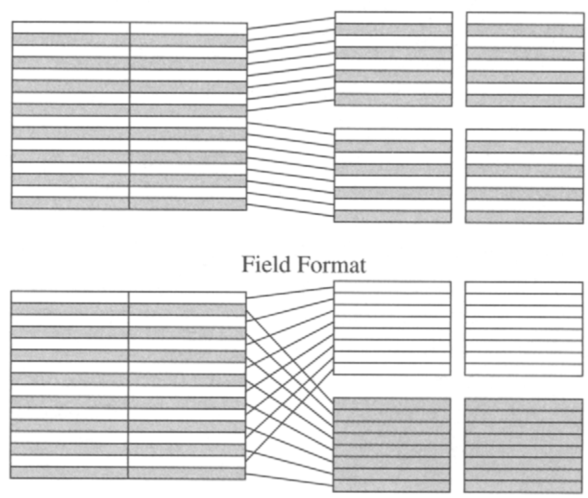

FIGURE 14 Frame/field format block for DCT.

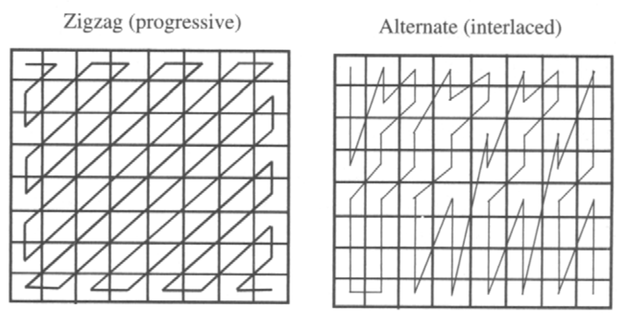

FIGURE 15 Progressive/Interlaced scan.

2.4.3 Quantization

In MPEG-2, for intra DC coefficients, 8–11 bits of precision is allowed after quantization, while in MPEG-1, only 8-bit precision is allowed. In MPEG-2, intra AC coefficients and all non intra DC and AC coefficients can be quantized to [−2048, 2047], while in MPEG-1, the range is [−256, 255]. The finer quantization can reduce the quantization error to improve the reconstructed video quality.

2.5 Scalable Coding

Scalable coding is also called layered coding. In scalable coding, the video is coded in a base layer and several enhancement layers. If only the base layer is decoded, basic video quality can be obtained. If the enhancement layers are also decoded, enhanced video quality (e.g., higher signal-to-noise ratio, higher resolution, higher frame rate) can be achieved. Scalable coding is useful for transmission over noisy channel since the more important layers (e.g., the base layer) can be better protected and sent over a channel with better error performance. Scalable coding is also used in video transport over variable bit rate channels. When the channel bandwidth is reduced, the less important enhancement layers may not be transmitted. It is also useful for progressive transmission which means the users can get rough representations of the video fast with the base layer and then the video quality will be refined as more enhancement data arrive. Progress transmission is useful for database browsing and image transmission over the Internet.

MPEG-2 supports three types of scalability modes: SNR (signal-to-noise ratio), spatial, and temporal scalability. Different scalable modes can be combined into hybrid coding schemes such as hybrid spatial-temporal and hybrid spatial-SNR scalability. In a basic MPEG-2 scalability mode, there can be two layers of video: lower and enhancement layers. The hybrid scalability allows up to three layers.

2.5.1 SNR Scalability

MPEG-2 SNR scalability provides two different video quality from a single video source while maintaining the same spatial and temporal resolutions. A block diagram of the two-layer SNR scalable encoder and decoder is shown in Figs. 16(a) and (b), respectively. In the base layer, the DCT coefficients are coarsely quantized and the coded bit stream is transmitted with moderate quality at a lower bit rate. In the enhancement layer, the difference between the nonquantized DCT coefficients and the coarsely quantized DCT coefficients from the lower layer is encoded with finer quantization step-sizes. By doing this, the moderate video quality can be achieved by decoding only the lower layer bit streams while the higher video quality can be achieved by decoding both layers.

2.5.2 Spatial Scalability

With spatial scalability, the applications can support users with different resolution terminals. For example, the compatibility between SDTV (standard definition TV) and HDTV can be achieved with the SDTV being coded as the base layer. With the enhancement layer, the overall bit stream can provide the HDTV resolution. The input to the base layer usually is created by downsampling the original video to create a low-resolution video for providing the basic spatial resolution. The choice of video formats such as frame sizes, frame rates, or chrominance formats is flexible in each layer.

A block diagram of the two-layer spatial scalable encoder and decoder is shown in Figs. 17(a) and (b), respectively. In the base layer, the input video signal is downsampled by spatial decimation. To generate a prediction for the enhancement layer video signal input, the decoded lower layer video signal is upsampled by spatial interpolation and is weighted and combined with the motion-compensated prediction from the enhancement layer. The selection of weights is done on a macroblock basis and the selection information is sent as a part of the enhancement layer bit stream.

The base and enhancement layer coded bit streams are then transmitted over the channel. At the decoder, the lower layer bit streams are decoded to obtain the lower resolution video. The lower-resolution video is interpolated and then weighted and added to the motion compensated prediction from the enhancement layer. In the MPEG-2 video standard, the spatial interpolator is defined as a linear interpolation or a simple averaging for missing samples.

2.5.3 Temporal Scalability

The temporal scalability is designed for video services which require different temporal resolutions or frame rates. The target applications include video over wireless channel where the video frame rate may need to be dropped when the channel condition is poor. It is also intended for stereoscopic video and coding of future HDTV formats in which the baseline is to make the migration from the lower temporal

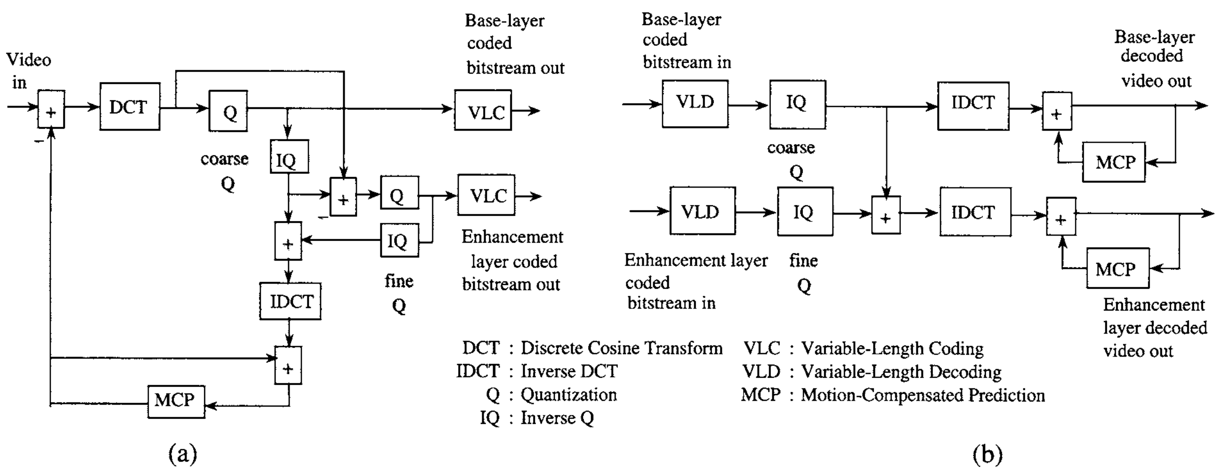

FIGURE 16 (a) SNR scalable encoder. (b) SNR scalable decoder.

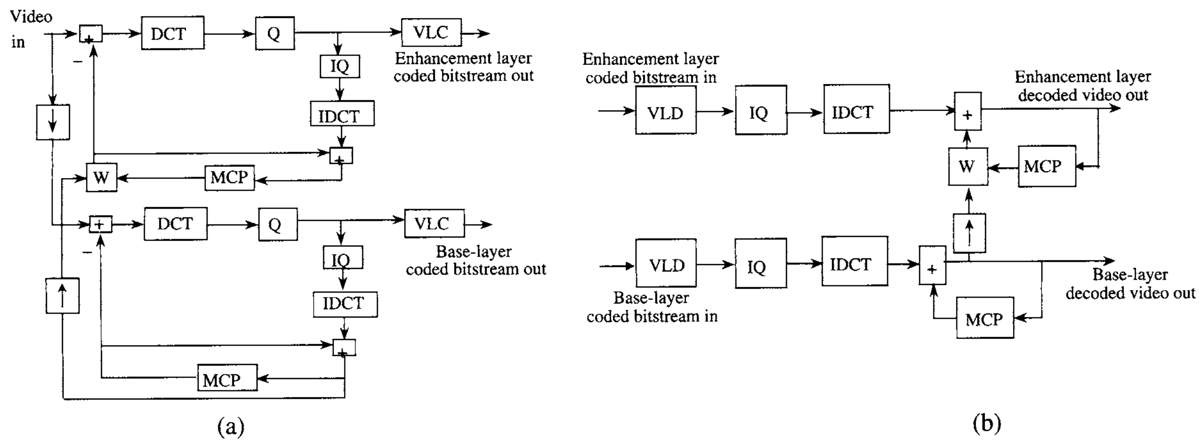

FIGURE 17 (a) Spatial scalable encoder. (b) Spatial scalable decoder.

resolution systems to the higher temporal resolution systems possible. In temporal scalable coding, the base layer is coded at a lower frame rate. The decoded base layer pictures provide motion compensated predictions for encoding the enhancement layer.

2.5.4 Hybrid Scalability

Two different scalable modes from the three scalability types, SNR, spatial, and temporal, can be combined into hybrid scalable coding schemes. Thus, it results in three combinations: hybrid of SNR and spatial, hybrid of spatial and temporal, and hybrid of SNR and temporal. Hybrid scalability supports up to three layers: the base layer, enhancement layer 1, and enhancement layer 2. The first combination, hybrid of SNR and Spatial scalabilities, is targeted at applications such as HDTV/SDTV or SDTV/videophone at two different quality levels. The second combination, hybrid spatial and temporal scalability, can be used for applications such as high temporal resolution progressive HDTV with basic interlaced HDTV and SDTV. The last combination, hybrid SNR and temporal scalable mode, can be used for applications such as enhanced progressive HDTV with basic progressive HDTV at two different quality levels.

2.6 Data Partitioning

Data partitioning is designed to provide more robust transmission in an error-prone environment. Data partitioning splits the block of 64 quantized transform coefficients into partitions. The lower partitions contain more critical information such as low frequency DCT coefficients. To provide more robust transmission, the lower partitions should be better protected or transmitted with a high priority channel with low probability of error while the upper partitions can be transmitted with a lower priority. This scheme has not been formally standardized in MPEG-2 but was specified in the information annex of the MPEG-2 DIS document [7]. One thing to note is that the partitioned data is not backward compatible with other MPEG-2 bit streams. Therefore, it requires a decoder which supports the decoding of data-partitioning. Using the scalable coding and data partitioning may result in mismatch of reconstructed pictures in the encoder and the decoder and thus cause drift in video quality. In MPEG-2, since there are I-pictures which can terminate error propagation, depending on the application requirements, it may not be a severe problem.

2.7 Other Tools for Error-Resilience

The effect of bit errors in MPEG-2 coded sequences varies depending on the location of the errors in the bit stream. Errors occurring in the sequence header, picture header, and slice header can make it impossible for the decoder to decode the sequence, the picture, or the slice. Errors in the slice data that contains important information such as macroblock header, DCT coefficients, and motion vectors can cause the decoder to lose synchronization or cause spatial and temporal error propagation. There are several techniques to reduce the effects of errors besides the scalable coding. These include concealment motion vectors, the slice structure, and temporal localization by the use of intra pictures/slices/macroblocks.

The basic idea of concealment motion vector is to transmit motion vectors with the intra-macroblocks. Since the intra-macroblocks are used for future predictions, they may cause severe video quality degradations if they are lost or corrupted by transmission errors. With a concealment motion vector, a decoder can use the best-matching block indicated by the concealment motion vector to replace the corrupted intra-macroblock. This improves the concealment performance of the decoder.

In MPEG, each slice starts with a slice header which is a unique pattern that can be found without decoding the variable length codes. These slice headers represent possible re-synchronization markers after a transmission error. A small slice size, i.e., a small number of macroblocks in a slice, can be chosen to increase the frequency of synchronization points, thus reducing the effects of the spatial propagation of each error in a picture. However, this can lead to a reduction in coding efficiency as the slice header overhead information is increased.

The temporal localization is used to minimize the extent of error propagation from picture to picture in a video sequence, e.g., by using intra-coding modes. For the temporal error propagation in an MPEG video sequence, the error from an I- or a P-picture will stop propagating when the next error-free I picture occurs. Therefore, increasing the number of I-pictures/slices/macroblocks in the coded sequence can reduce the distortion caused by the temporal error propagation. However, more I-pictures/slices/macroblocks will result in reduction of coding efficiency and it is more likely that errors will occur in the I-pictures which will cause error propagation.

2.8 Test Model

Similar to other video coding standards such as H.261 and MPEG-1, MPEG-2 only specifies the syntax and the decoder. Many detailed coding options are not specified. In order to have a reference MPEG-2 video quality, test models were developed in MPEG-2. The final test model of MPEG-2 is called "Test Model 5" (TM5) [10]. TM5 was defined only for main profile experiments. The motion compensated prediction techniques involve frame, field, dual-prime prediction and have forward and backward motion vectors as in MPEG-1. The dual-prime was kept in main profile but restricted to P-pictures with no intervening B-pictures. Two-step search, which consists of an integer pixel full search followed by a half-pixel search, is used for motion estimation. The mode decision (intra/inter-coding) is also specified. Main profile was restricted to only two quantization matrices: the default table specified in MPEG-1 and the nonlinear quantizer tables. The traditional zigzag scan is used for inter-coding while the alternate scan is used for intra-coding. The rate control algorithm in TMN5 consists of three layers operating at the GOP, the picture, and the macroblock levels. A bit-allocation per picture is determined at the GOP layer and updated based on the buffer fullness and the complexity of the pictures.

2.9 MPEG-2 Video and System Bit Stream Structures

A high-level structure of the MPEG-2 video bit stream is shown in Fig. 18. Every MPEG-2 sequence starts with a sequence header and ends with an end-of-sequence. MPEG-2 syntax is a superset of the MPEG-1 syntax. The MPEG-2 bit stream is based on the basic structure of MPEG-1 (refer to Fig. 8). There are two bit stream syntax allowed: ISO/IEC 11172-2 video sequence syntax or ISO/IEC 13818-2 (MPEG-2) video sequence syntax.

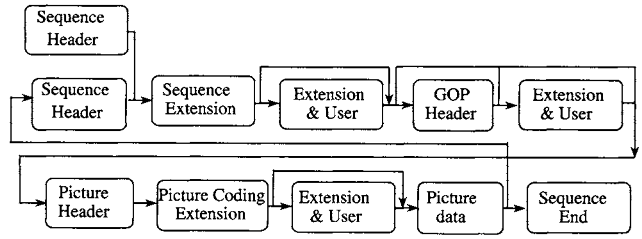

FIGURE 18 MPEG-2 data structure and syntax.

If the sequence header is not followed by the sequence extension, the MPEG-1 bit stream syntax is used. Otherwise, the MPEG-2 syntax is used which accommodates more features but at the expense of higher complexity. The sequence extension includes a profile/level indication, a progressive/interlaced indicator, a display extension including choices of chroma formats and horizontal/vertical display sizes, and choices of scalable modes. The GOP header is located next in the bit stream syntax with at least one picture following each GOP header. The picture header is always followed by the picture coding extension, the optional extension and user data fields, and picture data. The picture coding extension includes several important parameters such as the indication of intra-DC precision, picture structures (choices of the first/second fields or frame pictures), intra VLC format, alternate scan, choices of updated quantization matrix, picture display size, display size of the base layer in the case of the spatial scalability extension, and indicator of the forward/backward reference picture in the base layer in the case of the temporal scalability extension. The picture data consists of slices, macroblocks, and data for the coded DCT blocks. MPEG-2 defines six layers as MPEG-1. However, the specification of some data elements is different. The details of MPEG-2 syntax specification are documented in [8].

2.10 Summary

MPEG-2 is mainly targeted at general high quality video applications at bit rates greater than 2 Mbit/s. It is suitable for coding both progressive and interlaced video. MPEG-2 uses frame/field adaptive motion compensated predictive coding and DCT. Dual prime motion compensation for P-pictures is used for low-delay applications with no intervening B-picture. In addition to the default quantization table, MPEG-2 defines a nonlinear quantization table with increased accuracy for small values. Alternate scan and new VLC tables are defined for DCT coefficient coding. MPEG-2 also supports compatibility and scalability with the MPEG-1 standard. MPEG-2 syntax is a superset of MPEG-1 syntax and can support a variety of rates and formats for various applications. Similar to other video coding standards, MPEG-2 defines only syntax and semantics. It does not specify every encoding options (preprocessing, motion estimation, quantizer, rate quality control, and other coding options) and decoding options (post processing and error concealment) to allow continuing technology improvement and product differentiation. It is important to keep in mind that different implementations may lead to different quality, bit rate, delay, and complexity tradeoffs with different cost factors. An MPEG-2 encoder is much more expensive than an MPEG-2 decoder, since it needs to perform many more operations (e.g., motion estimation, coding mode decisions, and rate control). An MPEG-2 encoder is also much more expensive than an H.261 or an MPEG-1 encoder due to the higher resolution and more complicated motion estimations (e.g., larger search range, frame/field bi-directional motion estimation). MPEG-1 has been successfully used in VCD. MPEG-2 has been used in DVD, digital television including SDTV and HDTV broadcasting over satellites or cable TV (CATV) networks. They have been used in high quality video networking and streaming applications for transmitting live and pre-recorded streams over broadband Internet, corporate Intranets and virtual private networks. References [11]–[26] provide further information on the related MPEG-1 and MPEG-2 topics.

References

[1] ISO/IEC JTC1 CD 10918. Digital compression and coding of continuous-tone still images. International Organization for Standardization (ISO), 1993.

[2] ITU-T Recommendation H.261. Line transmission of non-telephone signals. Video codec for audio visual services at $p \times 64$ kbits/s, March 1993.

[3] S. Okubo, "Reference Model Methodology – A tool for the collaborative creation of video coding standards," *Proceedings of the IEEE*, 83, 2, 139–150, February 1995.

[4] MPEG proposal package description. Document ISO/WG8/MPEG/89–128 (July 1989).

[5] T. Hidaka, K. Ozawa, "Subjective assessment of redundancy-reduced moving images for interactive applications: Test methodology and report," *Signal Processing: Image Commun.*, 2, 201–219, Aug. 1990.

[6] ISO/IEC JTC1 CD 11172. Coding of moving pictures and associated audio for digital storage media up to 1.5 Mbits/s. International Organization for Standardization (ISO), 1992.

[7] ISO/IEC JTC1/SC2/WG11, "MPEG Video Simulation Model Three (SM3)," MPEG 90/041, July 1990.

[8] ISO/IEC JTC1/SC29/WG11. MPEG-2: Generic coding of moving pictures and associated audio information. ISO/IEC International Standard, August 2000.

[9] ISO/IEC 13818-2-ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio information: Video," 1995.

[10] ISO/IEC JTC1/SC29/WG11, "Test Model 5", MPEG 93/457, Document AVC-491, April 1993.

[11] M. L. Liou, "Visual Telephony as an ISDN application," *IEEE Commun Magazine*, 28, 30–38, Feb 1990.

[12] A. Tabatabai, M. Mills, and M. L. Liou, "A review of CCITT $p \times 64$ kbps video coding and related standards," *Intl. Electronic Imaging Exposition and Conf.*, 58–61, Oct. 1990.

[13] D. J. Le Gall, "MPEG: A video compression standard for multimedia applications," *Commun. of the ACM*, 34, 47–58, April 1991.

[14] D. J. Le Gall, "The MPEG video compression algorithm," *Signal Process. Image Commun.*, 4, 129–140, Apr. 1992.

[15] L. Chiariglione, "Standardization of moving picture coding for interactive applications," *GLOBECOM'89*, 559–563, Nov. 1989.

[16] A. Puri, "Video Coding Using the MPEG-1 Compression Standard," Society for Information Display International Symposium, Boston, Mass., 123–126, May 1992.

[17] A. Puri, "Video Coding using the MPEG-2 compression standard," *SPIE/VCIP*, Cambridge, MA, 2094, 1701–1713, Nov. 1993.

[18] S. Okubo, K. McCann, and A. Lippman, "MPEG-2 requirements, profiles, and performance verification," *Proceedings International Workshop on HDTV'93*, Ottawa, Canada, Oct 1993.

[19] A. Puri, R. Aravind, and B. Haskell, "Adaptive frame/field motion compensated video coding," *Signal Process: Image Commun.*, 5, 39–58, Feb. 1993.

[20] T. Naveen et al., "MPEG 4:2:2 profile: high-quality video for studio applications," *Photonics East, SPIE*, CR60, Philadelphia, PA, Oct. 1995.

[21] A. Puri, Compression of stereoscopic video using MPEG-2, *Photonics East, SPIE*, CR60, Philadelphia, PA, Oct. 1995.

[22] R. J. Clarke, *Digital Compression of Still Images and Video*, Academic Press, 1995.

[23] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, Kluwer Academic Publishers, Boston, 1995.

[24] J. L. Mitchell, W. B. Pennebaker, and D. J. Le Gall, *The MPEG Digital Video Compression Standard*, New York, NY: Van Nostrand Reinhold, 1996.

[25] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*, Prentice Hall, 1996.

[26] K. Konstantinides, C.-T. Chen, T.-C. Chen, H. Cheng, and F.-T. Jeng, "Design of an MPEG-2 Codec," *IEEE Signal Processing Magazine*, 32–41, July 2002.