Module 8

Video Coding Standards

Lesson
23

MPEG-1 standards

Lesson objectives

At the end of this lesson, the students should be able to :

1. Enlist the major video coding standards
2. State the basic objectives of MPEG-1 standard.
3. Enlist the set of constrained parameters in MPEG-1
4. Define the I- P- and B-pictures
5. Present the hierarchical data structure of MPEG-1
6. Define the macroblock modes supported by MPEG-1

23.0 Introduction

In lesson 21 and lesson 22, we studied how to perform motion estimation and thereby temporally predict the video frames to exploit significant temporal redundancies present in the video sequence. The error in temporal prediction is encoded by standard transform domain techniques like the DCT, followed by quantization and entropy coding to exploit the spatial and statistical redundancies and achieve significant video compression. The video codecs therefore follow a hybrid coding structure in which DPCM is adopted in temporal domain and DCT or other transform domain techniques in spatial domain.

Efforts to standardize video data exchange via storage media or via communication networks are actively in progress since early 1980s. A number of international video and audio standardization activities started within the International Telephone Consultative Committee (CCITT), followed by the International Radio Consultative Committee (CCIR), and the International Standards Organization / International Electrotechnical Commission (ISO/IEC). An experts group, known as the Motion Pictures Expects Group (MPEG) was established in 1988 in the framework of the Joint ISO/IEC Technical Committee with an objective to develop standards for coded representation of moving pictures, associated audio, and their combination for storage and retrieval of digital media. The standard was famed in 1992 and was nicknamed MPEG-1 which is the main focus of the current lesson.

This lesson first enlists the different standards proposed under ISO/IEC MPEG and CCITT, which was later renamed as International Telecommunications Union (ITU) and their application domains. This is followed by objectives major features and the modes of MPEG-1 standard.

23.1 Major Video coding initiatives and standards

There have been several initiatives in video coding that have led to a range of video coding standards, as listed below:

a) Video coding for video teleconferencing, which has led to the ITU standards – H..261 for ISDN video conferencing, H.263 for very low bit-rate and Plain Old Telephone Systems (POTS) video conferencing and the latest H.264 for video telephony and video streaming in wireless applications.

b) Video coding for storing movies on CD-ROM, with a bit-rate up to 1.5 M bits/sec, of which 1.2 Mbits/sec is allocated to video and 256 Kbits/sec is allocated to the audio. This is addressed by ISO MPEG-1, the topic for this lesson.

c) Video coding for broadcast and storing video on digital video disks (DVD) with an order of 2-15 Mbits/sec allocated to audio and video coding, which led to the ISO-MPEG-2 standard. This, standard also addressed High Definition Television (HDTV) applications.

d) Coding of separate audio-visual objects, both natural and synthetic, which led to the ISO-MPEG-4 standard.

e) Coding of multimedia metadata, i.e., the data describing the features of multimedia data, which is being addressed in the upcoming MPEG-7 standard.

23.2 Basic objectives of MPEG-1 standard

The MPEG-1 standard was primarily targeted for multimedia CD-ROM applications at a bit rate of 1.5 Mbits/sec.

The standard is generic in the sense that it specifics a syntax for the representation of the encoded bitstream and a method of decoding. Unlike JPEG, MPEG-1 does not stipulate use of specific algorithms for bitstream generation and allows substantial flexibility. The syntax supports operations such as motion estimation; motion compensated prediction; Discrete Cosine transforms (DCT); quantization and variable length coding. The standard supports a number of parameters that can be specified in the bit-stream itself and a variety of picture sizes, aspect ratios etc. are permissible.

In addition, MPEG-1 standard supports the following application specified features:

- *Frame–based random access of video*: This is achieved by allowing independent access-points (I-frames) to the bit-stream.

- *Fast-forward and fast reverse (FF/FR) searches*: This refers to the scanning of the compressed bitstream to search for the desired portions of the video stream.

- *Reverse playback of video*
- *Edit ability of the compressed bit stream*
- *Reasonable coding / decoding delay* of about 1 sec. To give the impression of interactivity.

23.3 Constrained parameters in MPEG-1

MPEG-1 is a generic video coding standard and a diversity of input parameters including flexible picture size and frame rate can be specified by the user. Although many of these parameters are specified in the syntax and are therefore arbitrary, the following set of constrained parameters are specified to aid hardware implementations:

Maximum number of pixels / line   : 720.

Maximum number of lines/picture: 576

Maximum number of picture/sec   : 30

Maximum number of macroblock/ pictures: 396

Maximum number of macroblocks/sec: 9900

Maximum bit-rate; 1.86 Mbit/sec

Maximum decoder buffer size: 376,832 bits.

Although these constrained parameters are supported, it may however be noted that to meet the target bit rate of 1.5 Mbits/sec, the input video is first converted into MPEG-1 Standard Intermediate Format (SIF), in which the luminance channel is 352 pixels x 240 lines and 30 frames/sec. Luminance and chrominance components are represented by 8 bits/pixel and the chrominance components are sub-sampled by a factor of 2 in both the horizontal and vertical directions.

23.4 Picture types in MPEG-1

The MPEG-1 standard supports the following three picture types:

- I- picture
- P-picture
- B-picture

23.4.1 Intraframe coded pictures (I-pictures):

These pictures are coded without reference to other pictures in the video sequence. I-pictures therefore do not use any motion estimation and motion compensation and the frames are treated just like still images. The pixel intensity values are DCT encoded in a manner similar to JPEG and compression is achieved by a combination of quantization and run length coding of zero coefficient.

The first frame of every video sequence must necessarily be an I-picture, since it does not have any past reference. In MPEG-1 standard, frames are encoded as I-pictures at regular intervals to enforce updating with the current content. This is done at the beginning of every group of pictures (GOP), which we are going to define in Section-23.5.

Since there is no temporal prediction, I pictures achieve very poor compression performance and require significant bits for encoding. However I-pictures have better reconstruction quality. Further, I-pictures allow random access and fast forward / fast rewind (FF/FR) functionalities in the bitstream.

23.4.2 Interframe **predicted pictures (P-pictures):**

These pictures are coded with reference to the nearest (in temporal order) coded I-picture or P-picture, using motion compensation for prediction. Since these pictures use the temporal redundancy for encoding, they achieve better compression performance as compared to I-pictures. However, these pictures do not allow random access and FF/FR functionalities in the bitstream. Also, temporal prediction does not work where there are scene changes.

23.4.3 Bi-directionally predicted pictures (B-pictures):

These pictures have the best compression performance. They use bi-directional motion estimation with reference to the nearest coded I-picture and / or P-pictures on either side of the B-picture in temporal order. To achieve high compression ratio in the encoded bit stream, most of the frames in a video sequence are encoded as B-pictures. Like the P-pictures, B-pictures also do not allow random access and FF/FR functionalities in the bit-stream.

For P-pictures, as well as B-pictures, the error in prediction through motion compensation is DCT-encoded and the compression is achieved through quantization and run-length encoding for zero coefficients. For both these pictures, entropy-coded motion vectors form part of the bitstream.

23.5 Hierarchical data structure in MPEG-1

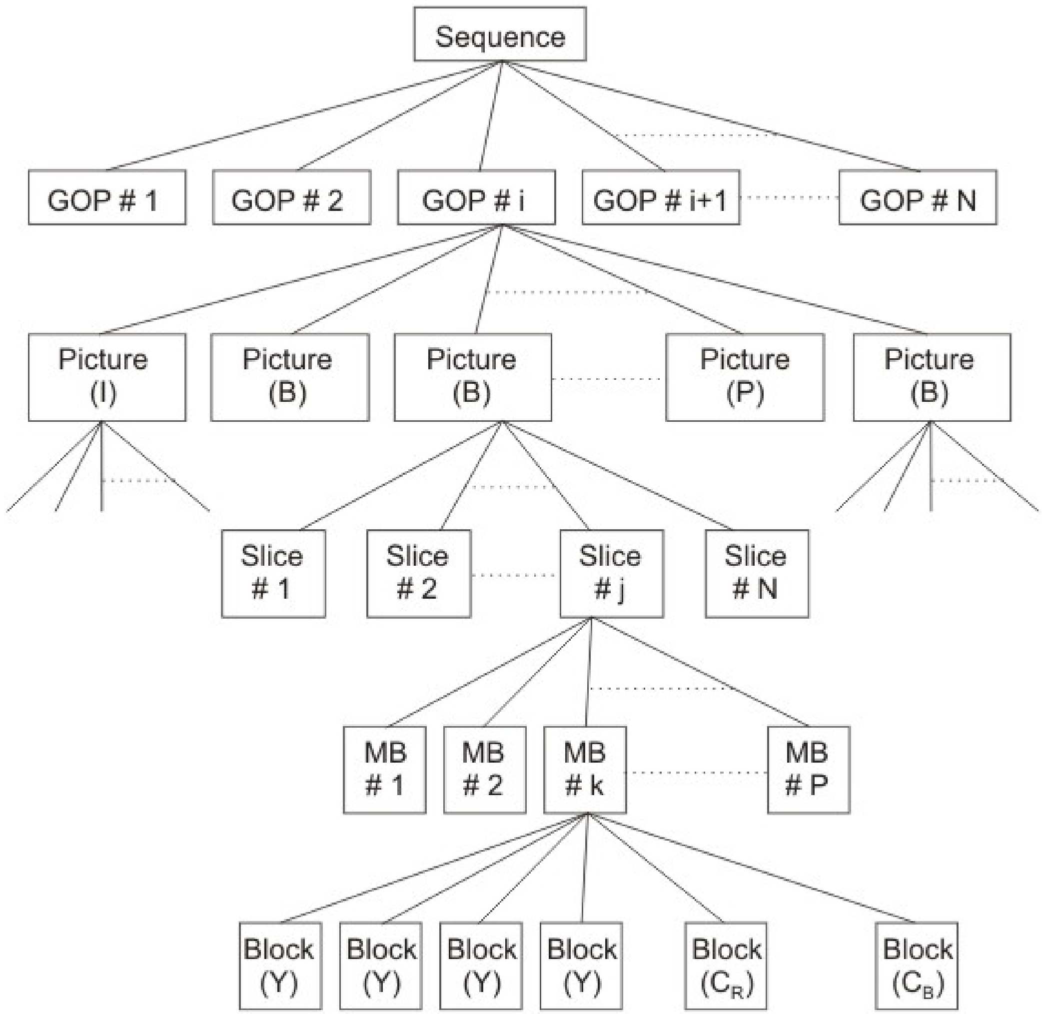

Fig. 23.1: Hierarchical data structure in MPEG-1.

The MPEG-1 data stream follows a six-layer hierarchical structure, as illustrated in fig 23.1.

At the top most level of the data structure, we have the video sequence itself, which consists of several group of pictures (GOP) in the next level. Each GOP begins with an I-picture and taking this as the reference, P-picture is encoded N-frames later than the I-picture in temporal order. The (N-1) frames in between are encoded as B-pictures. The P-pictures in turn predict the next P-picture that occurs N-frames later. Typically N is a small number of three or four frames. Larger value of N introduces more B-pictures and we may therefore think that it would be more efficient in terms of compression. However, larger value of N makes prediction worse in the generation of P-pictures which in turn require more number of bits to encode the prediction error. Typically, N=3 or 4 for GOP s. Fig 23.2 illustrates the composition of a GOP.

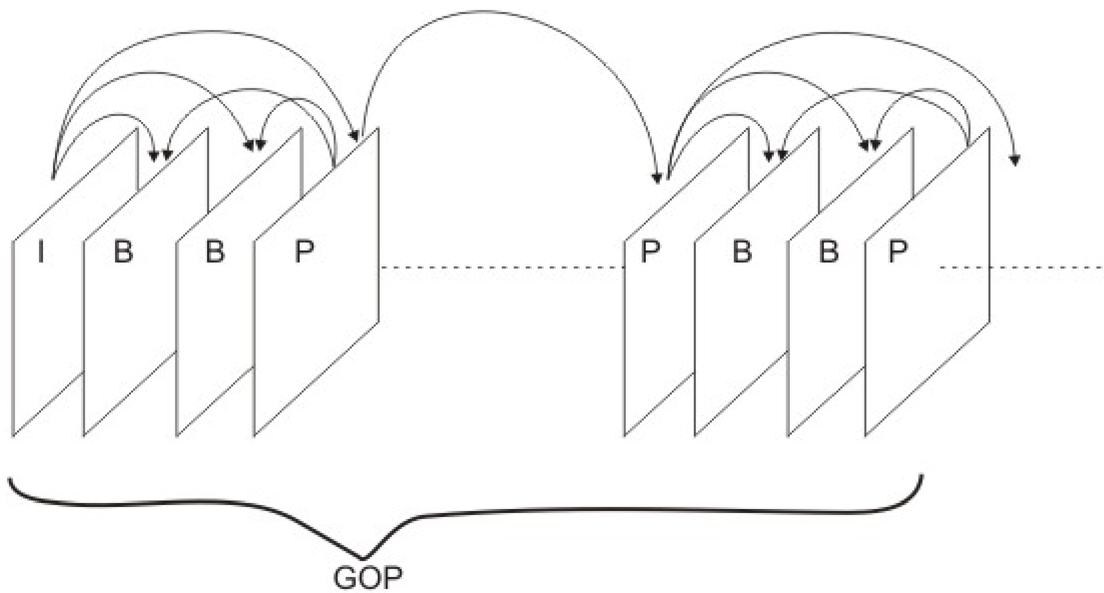

Fig 23.2 Compression of a GOP

At the next level of hierarchy, pictures are composed of slices, which are essentially sequence of macroblocks in raster scan order and are designed for error recovery. Below the slice layer are the macroblocks, whose composition is illustrated in fig 23.3.

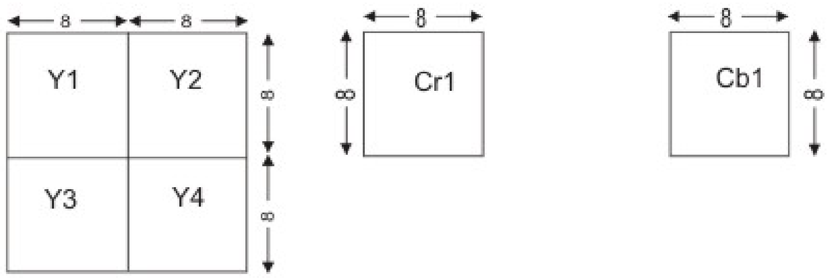

Fig 23.3 Composition of a macro block

MPEG-1 standard works on colour images in Y-Cr-Cb format and every R-G-B sequence must be converted to Y-Cr-Cb format before encoding through MPEG-1. Each macroblock is composed of 16 x 16 pixels of luminance (Y) channel and one block of 8 x 8 pixels from each of the Cr and Cb channels. The groups of 16x16 luminance pixels are further subdivided into four blocks of 8x8 pixels. In MPEG-1 standard, motion compensation is applied on 16x16 pixels and the DCT is applied on 8 x 8 pixels.

Each of the six layers of hierarchical data structure have headers to uniquely specific the data that follows.

23.6 Macroblock types supported by MPEG-1 standard.

The MPEG-1 standard supports the following macroblock types, depending upon the picture-types, ie, I, P- and B.

23.6.1 Macroblock types for I-pictures :

There are two types of macroblocks (MB) in the I-picture : “Intra” MBs are coded with the current quantization matrix x and “Intra-A’ (Intra-Adaptive) MBs are coded with the current quantization matrix elements divided by a quantization scale parameter MQUANT, which is transmitted as a part of MB header. In Intra-A, each macroblock can have different quantization step-size, unlike “Intra” MBs. This mode is preferred for images, whose level of details varies significantly from one region of the image to the other.

The quantization step size of the AC coefficients in the DCT array varies according to psycho-visual characteristics, based on which the elements of the quantization matrix are specified in the standard.

The “Intra-A” macroblock types are used for the bit-rate control and MQUANT may be varied on a macroblock-to-macroblock basis. Due to the provision of

"Intra-A" mode in MPEG, the encoding of I-pictures in MPEG is 30% more efficient than JPEG.

23.6.2 Macroblock types for P-pictures :

The allowable MB types for P-pictures are as follows :

- Intra
- Intra-A
- Intra-D
- Inter-DA
- Inter-F
- Inter-FD
- Inter-FDA
- Skipped

The "Intra" and "Intra-A" modes are the same as those discussed under I-pictures, There are some macroblocks in a P-picture where no reliable temporal reference exists due to change in scene content, occlusion/ uncovering of objects, objects moving out of frame, new objects moving into the frame etc. In such cases, those macroblocks may be coded as "Intra" or "Intra-A", although it may still be a P-picture.

MBs classified as "Inter" are interframe coded and the temporal prediction may use motion compensation and / or adaptive quantization. The letters that follows "Inter" are defined as

D : DCT of the prediction error will be coded.

F : Forward motion compensation is on.

A : Adaptive quantization is on.

Only "Inter-F" mode indicates that only motion compensation is satisfactory and motion vectors are to be transmitted, but no encoding of prediction error will be done. On the other hand, "Inter-D" mode encodes the frame-to frame differences at the current macroblock positions without performing any motion compensation. Inter-D mode is therefore preferred in cases where frame-to-frame displacements of the macroblock are insignificant, but there may be some changes in the pixel intensity values. The "skipped" modes is used for stationary macroblocks which neither undergo any noticeable displacements nor have any noticeable changes in pixel intensity values.

23.6.3 Macroblock types for B-pictures :

In MPEG-1 encoded sequences, almost two thirds of the frames are B-pictures. Bi-directional motion prediction offers several advantages over only forward prediction, as employed for P-picture, namely

(a) if an object is going to be covered or uncovered, it may still be predicted from one of the two references: I-picture or P-picture.

(b) averaging of motion compensation from two references provide better SNR as compared to P-picture

(c) fewer bits are required for encoding of B-pictures as compared to P-pictures, since the prediction is better.

B-pictures allow al the MB types supported by P-pictures. In addition to forward motion compensation (“F”-type), B-pictures allow backward motion compensation (B-type) and bi-directional (interpolative) motion compensation (I-type). Hence, following MB-types are supported--.

- Intra * Inter-B * Inter-I
- Intra-A * Inter-BD * Inter-ID
- Inter-F * Inter-BDA * Inter-IDA
- Inter-FD * Skipped
- Inter-FDA

23.7 Summary and Conclusions

MPEG-1 was the first ISO/IEC standard that was primarily targeted for a bit-rate up to 1.5 Mbits/sec for digital storage of video on CDs. The standard was adopted in 1992. As newer application areas like video broadcasting and HDTV emerged, there was a need to develop new standards. In the next few lessons, we shall cover more recent video coding standards like MPEG-2, MPEG-4, ITU-T, H.263, H.264 etc.