Module

8

VIDEO
CODING STANDARDS

Lesson

26

H.261 andH.263 Standards

Lesson Objectives

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

1. State the basic objective of H.261 standard.
2. Name the picture format supported by H.261.
3. Show the H.261 bitstream.
4. Show the Group of Block (GOB) arrangements in an H.261 picture.
5. Show the arrangement of macroblock (MB) in a GOB.
6. Show the structure of macroblock layer in H.261.
7. Define the prediction modes of macroblocks in H.261.
8. Show the structure of block layer in H.261.
9. State the basic objectives of H.263 standard.
10. Name the picture formats supported by H.263.
11. State the improved features of H.263 over H.261.
12. Explain the advanced prediction modes in H.263.

26.0 Introduction

In the past three lesson (lesson 23, lesson 24 and lesson-25), we studied three major multimedia standards, namely MPEG-1, MPEG-2 and the more recent as well as advanced MPEG-4. Apart from the MPEG, the International Telecommunication Union- Telecommunications Standards Sector (ITU-T) also evolved the standards for multimedia communications at restricted bit-rate over the wireline and wireless channels. The ITU-T standardization on multimedia first started with H.261, which was developed for ISDN video conferencing. The next standard H.263 supported Plain Old Telephone Systems (POTS) conferencing at very low bit-rates (64 Kbits/sec and lower). The most recent and advanced standard H.264 offers significant coding improvement over its predecessors and supports mobile video applications.

In this lesson, we are going to study the first two of these standards i.e., H.261 and H.263. The next lesson (lesson-27) will focus entirely on the H.264 standard. These standards mostly use the same concepts as those followed in MPEG and instead of repeating information, only the novelties and the special features of these standards will be covered.

26.1 Basic objectives of H.261 standard

The H.261 standard developed in 1988-90 was a fore runner to the MPEG-1 and was designed for video conferencing applications over ISDN telephone lines. The baseline ISDN has a bit-rate of 64 Kbits/ sec and at the higher end, ISDN supports bit-rates having integral multiples (p) of 64 Kbits/sec. For this reason, the standard is also referred to as the p x 64 Kbits/sec standard.

In addition to forming a basis for the MPEG-1 and MPEG-2 standards, the H.261 standards offers two important features:

a) Maximum coding delay of 150 msec. It has been observed that delays exceeding 150 msec do not provide direct visual feed back in bi-directional video conferencing

b) Amenability to VLSI implementation, which is important for widespread commercialization of videophone and teleconferencing equipments.

26.2 Picture formats and frame-types in H.261

The H.261 standard supports two picture formats:

i) Common Intermediate Format (CIF), having 352 x 288 pixels for the luminance channel (Y) and 176 x 144 pixels for each of the two chrominance channels U and V. Four temporal rates, viz, 30 15, 10 or 7.5 frames/ sec are supported. CIF images are used when $p \geq 6$, that is for video conferencing applications.

ii) Quarter of Common Intermediate Format (QCIF) having 176 x 144 pixels for the Y and 88 x 72 pixels each for U and V. QCIF images are normally used for low bit-rates applications like videophones (typically $p = 1$). The same four temporal rates are supported by QCIF images also.

H.261 frames are of two types

- *I-frames*: These are coded without any reference to previously coded frames.
- *P-frames*: These are coded using a previous frame as a reference for prediction.

26.3 H.261 Bit-stream structure

The H.261 bit-stream follows a hierarchical structure having the following layers:

- *Picture-layer*, that includes start of picture code (PSC), time stamp reference (TR), frame-type (I or P), followed by Group of Blocks (GOB) data.

- *GOB layer* that includes a GOB start code, the group number, a group quantization value, followed by macroblocks (MB) data.

- *MB layer*, that includes macroblock address (MBA), macroblock type (MTYPE : intra/inter), quantizer (MQUANT), motion vector data (MVD), the coded block pattern (CBP), followed by encoded block.

- *Block-layer*, that includes zig-zag scanned (*run*, *level*) pair of coefficients, terminated by the end of block (EOB)

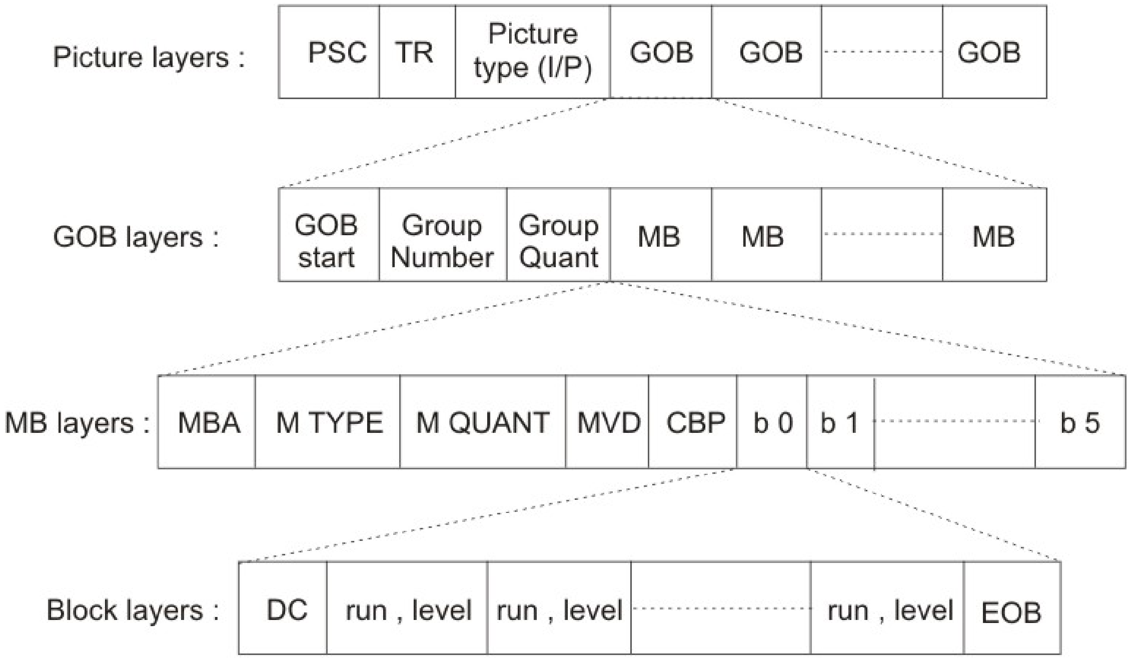
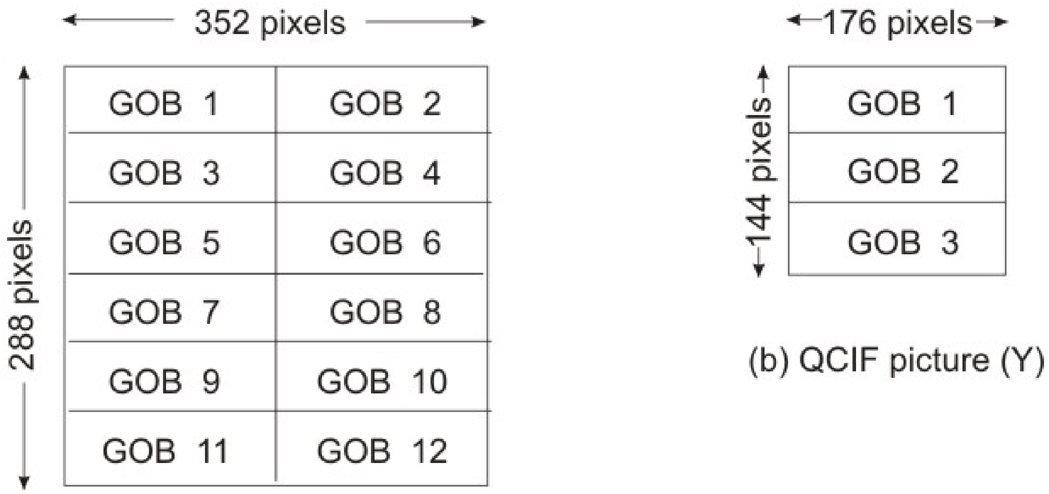


Fig. 26.1 Bit-stream structure of H.261

The H.261 bit-stream structure is illustrated in fig 26.1

It is possible that encoding of some GOBs may have to be skipped and the GOBs considered for encoding must therefore have a group number, as indicated. A common quantization value may be used for the entire GOB by specifying the group quantizer value. However, specifying the MQUANT in the macroblock overrides the group quantization value.

We now explain the major elements of the hierarchical data structure.



(a) CIF picture (Y)

(b) QCIF picture (Y)

Fig. 26.2 Arrangement of GOBs in (a) CIF picture (Y) , (b) QCIF picture (Y)

Each picture in H.261 is divided into GOBs as illustrated in fig 26.2

Fig 26.2 Arrangement of GOBs in (a) CIF picture, (b) QCIF picture

Each GOB thus relates to 176 pixels by 48 lines of Y and 88 pixels by 24 lines each of U and V. Each GOB must therefore comprise of 33 macroblocks- 11 horizontally and 3 vertically, as shown in fig 26.3

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

Fig. 26.3 Composition of a GoB. Each box corresponds to a macroblock & the number corresponds to the macroblock number.

Fig 26.3 Composition of a GOB. Each box corresponds to a macroblock and the number corresponds to the macroblock number.

Data for each GOB consists of a GOB header followed by data for macroblocks. Each GOB header is transmitted once between picture start codes in the CIF or QCIF sequence.

26.3.2 Macroblock layer:

As already shown in fig 26.3, each GOB consists of 33 macroblocks. Each macroblock relates to 16 x 16 pixels of Y and corresponding 8 x 8 pixels of each U and V, as shown in fig 26.4.

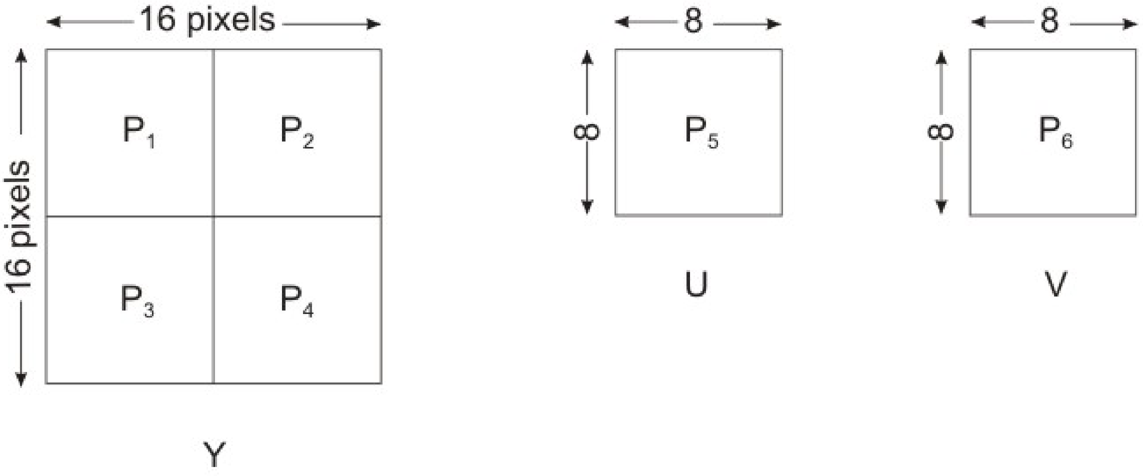


Fig. 26.4 Composition of a macroblock.

Fig 26.4 Composition of a macroblock

Since block is defined as a spatial array of 8 x 8 pixels, each macroblock therefore consists of six blocks – four from Y and one each from U and V. Each macroblock has a header, that includes the following information –

- ***Macroblock address*** **(MBA)-** It is a variable length codeword indicating the position of a macroblock within a group of blocks. For the first transmitted macroblock in a GOB, MBA is the absolute address. For subsequent macroblocks, MBA is the difference between the absolute address of the macroblock and the last transmitted macroblock. MBA is always included in transmitted macroblocks. Macroblocks are not transmitted when they contain no information for that part of the picture they represent.

- ***Macroblock type (MTYPE)-*** It is also variable length codeword that indicates the prediction mode employed and which data elements are present. The H.261 standard supports the following prediction modes:
    - *intra modes* are adopted for those macroblocks whose content change significantly between two successive macroblocks.
    - *inter modes* employ DCT of the inter-frame prediction error.
    - *inter + MC* modes employ DCT of the motion compensated prediction error.
    - *inter +MC* + fil modes also employ filtering of the predicted macroblock
- ***Quantizer (MQUANT*)**: It is present only if so indicated by the MTYPE. MQUANT overrides the quantizer specified in the GOB and till any subsequent MQUANT is specified, this quantizer is used for all subsequent macroblocks.
- ***Motion vector data (MVD)*** – It is also a variable length codeword (VLC) for the horizontal component of the motion vector, followed by a variable length codeword for the vertical component. MVD is obtained from the macroblock vector by subtracting the vector of the preceding macroblock.
- ***Coded block pattern (CBP)*** – CBP gives a pattern number that signifies which of the block within the macroblock has at least one significant transformation coefficient. The pattern number is given by

$$32P_1 + 16P_2 + 8P_3 + 4P_4 + 2P_5 + P_6$$

where, $P_n = 1$ if any coefficient is present for block n, else 0. The block numberings are as per fig.26.4.

26.3.3 Block Layer:

The block layer does not have any separate header, since macroblock is the basic coding entity. Data for a block consists of codewords of transform coefficients (TCOEFF), followed by an end of block (EOB) marker.

Transform coefficients are always present for intra macroblocks. For inter-coded macro blocks, transform coefficients may or may not be present within the block and their status is given by the *CBP* field in the macroblock layer. TCOEFF encodes the (RUN, LEVEL) combinations using variable length codes, where RUN indicates run of zero coefficient in the zig-zag scanned block DCT array.

26.4 Basic objectives of H.263 standard

Before we talk about the next video coding standard H.263 that was adopted by ITU-T, we would like to tell the readers why we do not talk about any standard called H.262, which should have been logically there in between the H.261 and H.263. Indeed, H.262 project was defined to address ATM/broadband video conferencing applications, but since it scope was included by MPEG-2 standard, no separate ITU-T standard was proposed by the name H.262. The H.261 standard's coding algorithm achieved several major performance breakthroughs, but at the lower extreme of its bit-rate, i.e., at 64 Kbit/sec, serious blocking artifacts produced annoying effects. This was tackled by reduced the frame rate, for example from the usual rate of 30 frames/ sec down to 10 frames/sec by considering only one out of every three frames and dropping the remaining two. However, reduced frame rate reduces temporal resolution, which is also not very desirable for rapidly changing scenes. Reduced frame rate also causes high end-to-end delays, which is also not very desirable. Hence, there was a need to design a coding standard that would provide better performance than the H.261 standard at lower bit-rate. With this requirement evolved the H.263 standard, whose targeted application was POTS video conferencing.

During the development of H.263, the target bit-rate was determined by the maximum bitrate achievable at the general switched telephone network (GSTN), which was 28.8 Kbits/sec at that time. At these bit-rates, it was necessary to keep the overhead information at a minimum. The other requirements of H.263 standardization were:

- Use of available technology
- Interoperability between the other standards, like H.261
- Flexibility for future extensions
- Quality of service parameters, such as resolution, delay, frame-rate etc.
- Subjective quality measurements.

Based on all these requirements an efficient coding scheme was designed. Although it was optimized for 28.8 Kbits/sec, even at higher bit rates up to 600 Kbits/sec, H.263 outperformed the H.261 standard.

26.5 Picture formats of H.263

The H.263 standard supports five pictures formats ---

- Sub-QCIF 128 x 96 pixels (Y), 64 x 48 pixels (U,V)
- QCIF 176 x 144 pixels (Y), 88 x 72 pixels (U,V)

- CIF 352 x 288 pixels (Y), 176 x 144 pixel (U,V)
- 4CIF 704 x 576 pixels (Y), 352 x 288 pixel (U,V)
- 16 CIF 1408 x 1152 pixels (Y), 704 x 576 pixel (U,V)

The CIF, 4CIF and 16 CIF picture formats are optional for encoders as well as decoders. It is mandatory for the decoders to support both sub-QCIF and QCIF picture formats. However, for encoders, only one of these two formats (Sub-QCIF or QCIF) is mandatory. In all these formats Y, U and V are sampled in 4: 2:0.

26.6 Improved features of H.263 over H.261

The H.263 offered several major improvements over its predecessor H.261. Some of these are –

- ***Half – pixel motion estimation*:** In H.261, the motion vector were expressed in integer pixel units. This often poses a limitation in motion compensation, since one pixel resolution is often too crude to represent real world motion. Half-pixel motion estimation is explained in the next sub section.

- ***Unrestricted motion vector mode:*** In the default prediction mode of H.263, all motion vectors are restricted so that the pixels referenced by them lie within the coded picture area. In the unrestricted motion vector mode, this restriction is removed and the motion vectors are allowed to point outside the picture.

- ***Advanced prediction mode –*** This is an optional mode that supports four motion vectors per macroblock and overlapped block motion compensation (OBMC). This is explained in detail later.

- ***PB-frames mode –*** This increases the frame-rate without significantly increasing the bitrate. The concept is explained in a later sub-section.

- ***Syntax based arithmetic coding (SAC) mode –*** This mode achieves a better compression, as compared to Huffman coded VLC tables.

26.6.1 Half pixel motion estimation :

Motion estimation with half pixel accuracy requires pixel based interpolation, as illustrated in fig. 26.5.

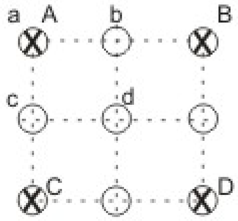


Figure 26.5 Interpolation of pixels on half - pixel grid

In fig 26.5, the integer pixel positions have been indicated by the "+" symbol. A half pixel wide grid is formed in which one out of four pixels in the grid coincides with the integer grid position. The remaining three out of four pixels are generated through interpolation.

The interpolated pixels indicated by "O" symbol lies at the integer as well as the half-pixel positions. Together, the integer and the half pixel positions create a picture that has two times the spatial resolution as compared to the original picture in both the horizontal and the vertical directions. In fig 26.5, the interpolated pixels marked as 'a" 'b" 'c' and 'd' are given by.

A= A ,

b = (A+B) / 2,

C = (A+C)/2,

d = (A+B+C+D)/4.

where A, B,C and D are the pixel intensities at the integer pixel positions.

When the motion estimation is carried out on this improved resolution interpolated image, a motion vector of 1 unit in this resolution corresponds to 0.5 unit with respect to the original resolution. This is the basic principles of half-pixel motion estimation.

26.6.2 Advanced prediction mode in H.263 :

The advanced prediction mode in H.263 has two major aspects –

(a) it uses four motion vectors per macroblock. Each 8 x 8 block is associated with one motion vector instead of one motion vector for the entire macroblock. This results in a better motion representation, but increases the number of bits to encode the motion vector.

(b) it uses *overlapped block motion compensation* (OBMC), which results in overall smoothing of the image and removal of blocking artifacts. OBMC involves using motion vectors of neighboring blocks to reconstruct a block. An 8 x 8 luminance block pixel $\hat{p}(i, j)$ is a weighted sum of three prediction values, as shown below

$$\hat{p}(i, j) = \frac{\left[\sum_{k=0}^{2} p(i+\hat{u}_k, j+\hat{v}_k) H_k(i, j) + 4\right]}{8}$$

where $(\hat{u}_k, \hat{v}_k)$ is the motion vector of the current block (*k* = 0) the block either above or below (*k* =1), or the block either to the left or right of the current block (*k* =2). Here, $\hat{p}(i, j)$ is the reference (previous) frame and $\{H_k(i, j); k = 0,1,2\}$ are the weights defined as

$$H_0 = \begin{bmatrix} 4 & 5 & 5 & 5 & 5 & 5 & 5 & 4 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 6 & 6 & 6 & 6 & 5 & 5 \\ 5 & 5 & 6 & 6 & 6 & 6 & 5 & 5 \\ 5 & 5 & 6 & 6 & 6 & 6 & 5 & 5 \\ 5 & 5 & 6 & 6 & 6 & 6 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\ 4 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \end{bmatrix}, \quad H_1 = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}, \quad H_2 = H_1^T$$

In the advanced prediction mode, motion vectors are allowed to cross the picture boundaries, just like the unrestricted motion vector mode.

26.6.3 PB-frame mode :

The H.263 standard has introduced a new concept, called PB-frames, which combines a P-frame and a B-frame to encode as one unit. The prediction mechanism employed in PB-frame is illustrated in fig.26.6.

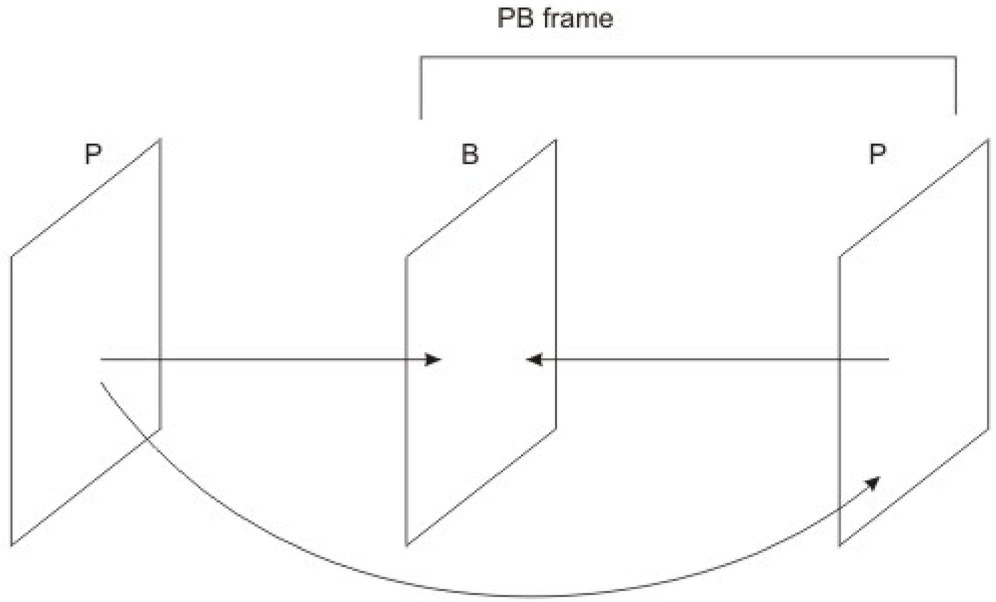


Fig. 26.6 Prediction in PB frame.

The P-picture within the PB-frame is predicted from the previously decoded P-picture and the B-picture is bi-directionally predicted from the previous P-picture, as well as the P-picture currently being decoded.

Information from the P-picture and the B-picture within the PB-frame is interleaved at the macroblock level. The P-macroblock information is directly followed by the B-macroblock information.

The transmission of bit overhead is much higher if P and B pictures are encoded separately. For a given P-picture rate, the use of PB-frames causes no extra delay.

26.7 H.263 + Extension

Subsequent to the H.263 recommendations, some extension features were added to it and these are referred to as the H.263 + features. These include some new types of pictures, such as scalability pictures, improved PB-frames, custom source formats etc; new coding methods such as advanced intra coding through spatial filtering, deblocking filters etc.

26.8 Conclusions

This lesson has extensively covered the ITU-T video coding standards H.261 and H.263. The latter has better performance figures than the former and can address very low bit-rate coding, having a bit rate lower than 64 kbits/sec.

In lesson–27, we are going to study H.264, the latest in the ITU-T standard, which outperforms the two earlier standards and caters for a wide range of applications and networks.