forward. In addition, random access may be possible. The ability of the decoder to implement these modes depends to some extent on the nature of the digital storage medium on which the coded video is stored.

The overall process of encoding and decoding is illustrated below:

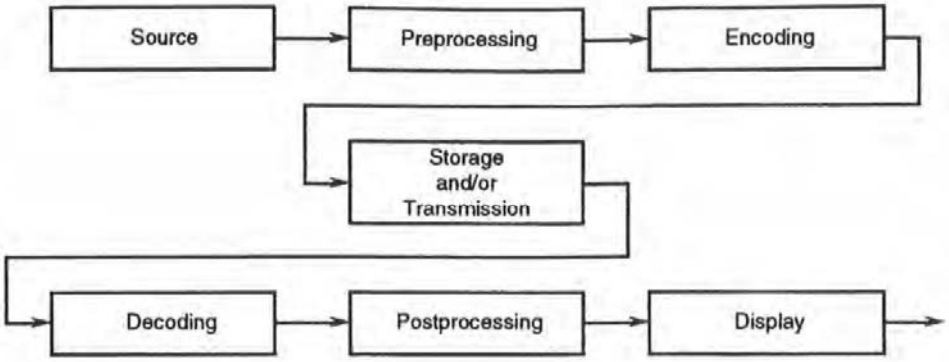


Figure D.1 -- Coding and decoding process

Figure D.1 shows a typical sequence of operations that must be performed before moving pictures can be seen by a viewer. The unencoded source may exist in many forms, such as the CCIR 601 format. Clause D.3 describes how such a source may be converted into the appropriate resolution for subsequent encoding. In the encoding step, the encoder must be aware of the decoder buffer capacity, and the need of the decoder to match the rate of the media to the rate of filling the picture buffer with each successive picture. To this end, a model of the decoder buffer and its overflow and underflow problem is introduced in D.4, and rate control is described in D.6.1 The structure of an ISO/IEC 11172-2 bitstream is covered in D.5, as are the coding operations that compress the video. Following the encoding process, the bitstream may be copied to a storage medium. To view the moving pictures, the decoder accesses the ISO/IEC 11172-2 bitstream, and decodes it as described in D.7. Postprocessing for display is described in D.8.

D.2.2 MPEG video compression techniques

Video is represented as a succession of individual pictures, and each picture is treated as a two-dimensional array of picture elements (pels). The colour representation for each pel consists of three components: Y (luminance), and two chrominance components, Cb and Cr.

Compression of digitized video comes from the use of several techniques: subsampling of the chrominance information to match the sensitivity of the human visual system (HVS), quantization, motion compensation (MC) to exploit temporal redundancy, frequency transformation by discrete cosine transform (DCT) to exploit spatial redundancy, variable length coding (VLC), and picture interpolation.

D.2.2.1 Subsampling of chrominance information

The HVS is most sensitive to the resolution of an image's luminance component, so the Y pel values are encoded at full resolution. The HVS is less sensitive to the chrominance information. Subsampling reduces the number of pel values by systematically combining them with a type of averaging process. This reduces the amount of information to be compressed by other techniques.The International Standard retains one set of chrominance pels for each 2x2 neighbourhood of luminance pels.

D.2.2.2 Quantization

Quantization represents a range of values by a single value in the range. For example, converting a real number to the nearest integer is a form of quantization. The quantized range can be concisely represented as an integer code, which can be used to recover the quantized value during decoding. The difference between the actual value and the quantized value is called the quantization noise. Under some circumstances, the HVS is less sensitive to quantization noise so such noise can be allowed to be large, thus increasing coding efficiency.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.2.2.3 Predictive coding

Predictive coding is a technique to improve the compression through statistical redundancy. Based on values of pels previously decoded, both the encoder and decoder can estimate or predict the value of a pel yet to be encoded or decoded. The difference between the predicted and actual values is encoded. This difference value is the prediction error which the decoder can use to correct the prediction. Most error values will be small and cluster around the value 0 since pel values typically do not have large changes within a small spatial neighbourhood. The probability distribution of the prediction error is skewed and compresses better than the distribution of the pel values themselves. Additional information can be discarded by quantizing the prediction error. In this International Standard predictive coding is also used for the dc-values of successive luminance or chrominance blocks and in the encoding of motion vectors.

D.2.2.4 Motion compensation and interframe coding

Motion compensation (MC) predicts the values of a block pels in a picture by relocating a block of neighbouring pel values from a known picture. The motion is described as a two-dimensional motion vector that specifies where to retrieve a block of pel values from a previously decoded picture that is used to predict pel values of the current block. The simplest example is a scene where the camera is not moving, and no objects in the scene are moving. The pel values at each image location remain the same, and the motion vector for each block is 0. In general however, the encoder may transmit a motion vector for each macroblock. The translated block from the known picture becomes a prediction for the block in the picture to be encoded. The technique relies on the fact that within a short sequence of pictures of the same general scene, many objects remain in the same location while others move only a short distance.

D.2.2.5 Frequency transformation

The discrete cosine transform (DCT) converts an 8 by 8 block of pel values to an 8 by 8 matrix of horizontal and vertical spatial frequency coefficients. An 8 by 8 block of pel values can be reconstructed by performing the inverse discrete cosine transform (IDCT) on the spatial frequency coefficients. In general, most of the energy is concentrated in the low frequency coefficients, which are conventionally written in the upper left corner of the transformed matrix. Compression is achieved by a quantization step, where the quantization intervals are identified by an index. Since the encoder identifies the interval and not the exact value within the interval, the pel values of the block reconstructed by the IDCT have reduced accuracy.

The DCT coefficient in location (0,0) (upper left) of the block represents the zero horizontal and zero vertical frequency and is called the dc coefficient. The dc coefficient is proportional to the average pel value of the 8 by 8 block, and additional compression is provided through predictive coding since the difference in the average value of neighbouring 8 by 8 blocks tends to be relatively small. The other coefficients represent one or more nonzero horizontal or nonzero vertical spatial frequencies, and are called ac coefficients. The quantization level of the coefficients corresponding to the higher spatial frequencies favors the creation of an ac coefficient of 0 by choosing a quantization step size such that the HVS is unlikely to perceive the loss of the particular spatial frequency unless the coefficient value lies above the particular quantization level. The statistical encoding of the expected runs of consecutive zero-valued coefficients of higher-order coefficients accounts for considerable compression gain. To cluster nonzero coefficients early in the series and encode as many zero coefficients as possible following the last nonzero coefficient in the ordering, the coefficient sequence is specified to be a zig-zag ordering; see figure D.30. The ordering concentrates the highest spatial frequencies at the end of the series.

D.2.2.6 Variable-length coding

Variable-length coding (VLC) is a statistical coding technique that assigns codewords to values to be encoded. Values of high frequency of occurrence are assigned short codewords, and those of infrequent occurrence are assigned long codewords. On average, the more frequent shorter codewords dominate, such that the code string is shorter than the original data.

D.2.2.7 Picture interpolation

If the decoder reconstructs a picture from the past and a picture from the future, then the intermediate pictures can be reconstructed by the technique of interpolation, or bidirectional prediction. Blocks in the intermediate pictures can be forward and backward predicted and translated by means of motion vectors. The decoder may reconstruct pel values belonging to a given block as an average of values from the past and future pictures.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.2.3 Bitstream hierarchy

The ISO/IEC 11172-2 coding scheme is arranged in layers corresponding to a hierarchical structure. A **sequence** is the top layer of the coding hierarchy and consists of a header and some number of **groups-of-pictures** (**GOPs**). The sequence header initializes the state of the decoder. This allows decoders to decode any sequence without being affected by past decoding history.

A **GOP** is a random access point, i.e. it is the smallest coding unit that can be independently decoded within a sequence, and consists of a header and some number of **pictures**. The GOP header contains time and editing information.

A **picture** corresponds to a single frame of motion video, or to a movie frame. There are four picture types: **I-pictures**, or *intra coded pictures*, which are coded without reference to any other pictures; **P-pictures**, or *predictive coded pictures*, which are coded using motion compensation from a previous I or P-picture; **B-pictures**, or *bidirectionally* predictive coded pictures, which are coded using motion compensation from a previous and a future I or P-picture, and **D** pictures, or *D* pictures, which are intended only for a fast forward search mode. A typical coding scheme contains a mix of I, P, and B-pictures. Typically, an I-picture may occur every half a second, to give reasonably fast random access, with two B-pictures inserted between each pair of I or P-pictures.

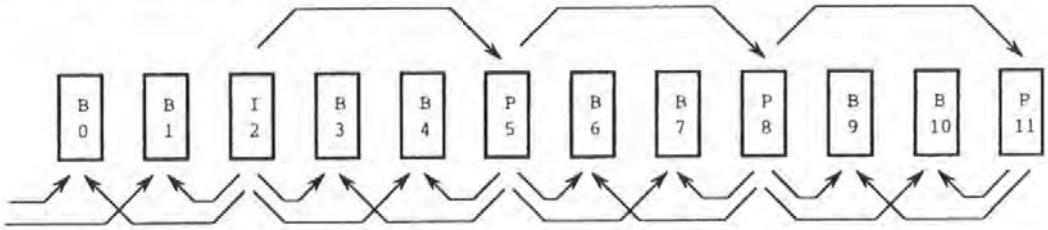


Figure D.2 -- Dependency relationship between I, B, and P-pictures

Figure D.2 illustrates a number of pictures in display order. The arrows show the dependency relationship of the predictive and bidirectionally predictive coded pictures.

Note that because of the picture dependencies, the bitstream order, i.e. the order in which pictures are transmitted, stored, or retrieved, is not the display order, but rather the order which the decoder requires them to decode the bitstream. An example of a sequence of pictures, in display order, might be:

I	B	B	P	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Figure D.3 -- Typical sequence of pictures in display order

whereas the bitstream order would be as shown below:

I	P	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B
0	3	1	2	6	4	5	9	7	8	12	10	11	15	13	14	18	16	17

Figure D.4 -- Typical sequence of pictures in bitstream order

Because the B-pictures depend on the following (in display order) I or P-picture, the I or P-picture must be transmitted and decoded before the dependent B-pictures.

Pictures consist of a header and one or more **slices**. The picture header contains time, picture type, and coding information.

A **slice** provides some immunity to data corruption. Should the bitstream become unreadable within a picture, the decoder should be able to recover by waiting for the next slice, without having to drop an entire picture.

Slices consist of a header and one or more **macroblocks.** At the start of each slice all of the predictors, for dc values and motion vectors, are reset. The slice header contains position and quantizer scale information. This is sufficient for recovery from local corruption.

A **macroblock** is the basic unit for motion compensation and quantizer scale changes.

Each **macroblock** consists of a header and six component 8 by 8 **blocks**: four blocks of luminance, one block of Cb chrominance, and one block of Cr chrominance. See figure D.5. The macroblock header contains quantizer scale and motion compensation information.

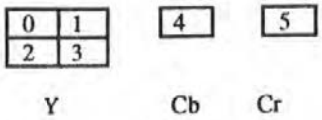


Figure D.5 -- Macroblock structure

A **macroblock** contains a 16-pel by 16-line section of luminance component and the spatially corresponding 8-pel by 8-line section of each chrominance component. A skipped macroblock is one for which no information is stored (see 2.4.4.4).

Note that the picture area covered by the four blocks of luminance is the same as the area covered by each of the chrominance blocks. This is due to subsampling of the chrominance information.

Blocks are the basic coding unit, and the DCT is applied at this block level. Each block contains 64 component **pels** arranged in an 8 by 8 array as shown in figure D.6.

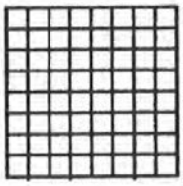


Figure D.6 -- Block structure

© ISO/IEC

D.2.4 Decoder overview

A simplified block diagram of a possible decoder implementation is shown below:

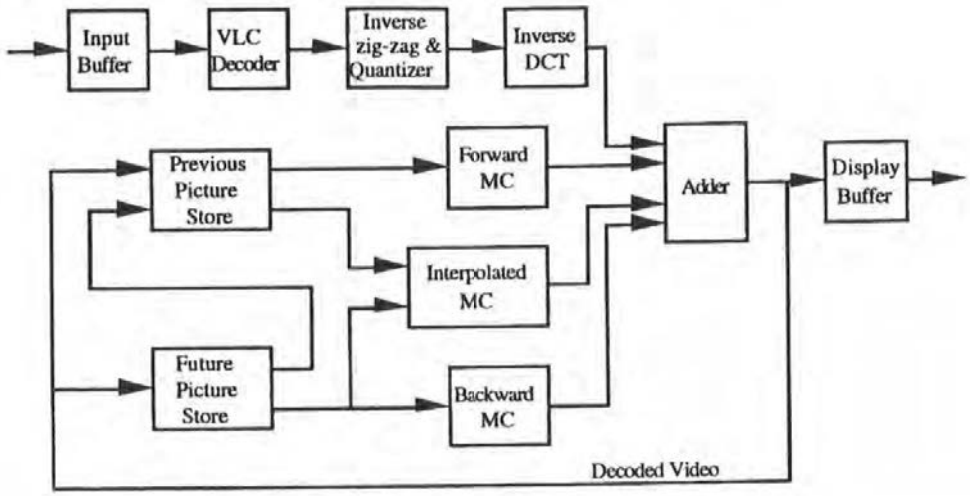


Figure D.7 -- Simplified decoder block diagram

It is instructive to follow the method which the decoder uses to decode a bitstream containing the sequence of pictures given in Fig D.4, and display them in the order given in Fig D.3. The following description is simplified for clarity.

The input bitstream is accumulated in the Input Buffer until needed. The Variable Length Code (VLC) Decoder decodes the header of the first picture, picture 0, and determines that it is an I-picture. The VLC Decoder produces quantized coefficients corresponding to the quantized DCT coefficients. These are assembled for each 8 by 8 block of pels in the image by inverse zig-zag scanning. The Inverse Quantizer produces the actual DCT coefficients using the quantization step size. The coefficients are then transformed into pel values by the Inverse DCT transformer and stored in the Previous Picture Store and the Display Buffer. The picture may be displayed at the appropriate time.

The VLC Decoder decodes the header of the next picture, picture 3, and determines that it is a P-picture. For each block, the VLC Decoder decodes motion vectors giving the displacement from the stored previous picture, and quantized coefficients corresponding to the quantized DCT coefficients of the difference block. These quantized coefficients are inverse quantized to produce the actual DCT coefficients. The coefficients are then transformed into pel difference values and added to the predicted block produced by applying the motion vectors to blocks in the stored previous picture. The resultant block is stored in the Future Picture Store and the Display Buffer. This picture cannot be displayed until B-pictures 1 and 2 have been received, decoded, and displayed.

The VLC Decoder decodes the header of the next picture, picture 1, and determines that it is a B-picture. For each block, the VLC decoder decodes motion vectors giving the displacement from the stored previous or future pictures or both, and quantized coefficients corresponding to the quantized DCT coefficients of the difference block. These quantized coefficients are inverse quantized to produce the actual DCT coefficients. The coefficients are then inverse transformed into difference pel values and added to the predicted block produced by applying the motion vectors to the stored pictures. The resultant block is then stored in the Display Buffer. It may be displayed at the appropriate time.

The VLC Decoder decodes the header of the next picture, picture 2, and determines that it is a B-picture. It is decoded using the same method as for picture 1. After decoding picture 2, picture 0, which is in the Previous Picture Store, is no longer needed and may be discarded.

The VLC Decoder decodes the header of the next picture, picture 6, and determines that it is a P-picture. The picture in the Future Picture Store is copied into the Previous Picture Store, then decoding proceeds as for picture 3. Picture 6 should not be displayed until pictures 4 and 5 have been received and displayed.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The VLC Decoder decodes the header of the next picture, picture 4, and determines that it is a B-picture. It is decoded using the same method as for picture 1.

The VLC Decoder decodes the header of the next picture, picture 5, and determines that it is a B-picture. It is decoded using the same method as for picture 1.

The VLC Decoder decodes the header of the next picture, picture 9, and determines that it is a P-picture. It then proceeds as for picture 6.

The VLC Decoder decodes the header of the next picture, picture 7, and determines that it is a B-picture. It is decoded using the same method as for picture 1.

The VLC Decoder decodes the header of the next picture, picture 8, and determines that it is a B-picture. It is decoded using the same method as for picture 1.

The VLC Decoder decodes the header of the next picture, picture 12, and determines that it is an I-picture. It is decoded using the same method as for picture 0. This process is repeated for the subsequent pictures.

D.3 Preprocessing

The source material may exist in many forms, e.g. computer files or CCIR 601 format, but in general, it must be processed before being encoded. This clause discusses some aspects of preprocessing.

For a given data rate and source material, there is an optimum picture rate and spatial resolution at which to code if the best perceived quality is desired. If the resolution is too high, then too many bits will be expended on the overhead associated with each block leaving too few to code the values of each pel accurately. If the resolution is too low, the pel values will be rendered accurately, but high frequency detail will be lost. The optimum resolution represents a tradeoff between the various coding artifacts (e.g. noise and blockiness) and the perceived resolution and sharpness of the image. This tradeoff is further complicated by the unknowns of the final viewing conditions, e.g. screen brightness and the distance of the viewer from the screen.

At data rates of 1 to 1,5 Mbits/s, reasonable choices are: picture rates of 24, 25 and 30 pictures/s, a horizontal resolution of between 250 and 400 pels, and a vertical resolution of between 200 and 300 lines. Note that these values are not normative and other picture rates and resolutions are valid.

D.3.1 Conversion from CCIR 601 video to MPEG SIF

The two widely used scanning standards for colour television are 525 and 625 lines at 29,97 and 25 frames/s respectively. The number of lines containing picture information in the transmitted signal is 484 for the 525-line system and 576 for the 625-line system. Both use interlaced scanning with two fields per picture.

CCIR Recommendation 601 defines standards for the digital coding of colour television signals in component form. Of these the 4:2:2 standard has become widely adopted; the sampling frequency used for the luminance signal, Y, is 13,5 MHz and the two colour difference signals, Cb or B-Y and Cr or R-Y, are both sampled at 6,75 MHz. The number of luminance samples in the digital active line is 720 but only about 702 will be used in practice by the analogue active line.

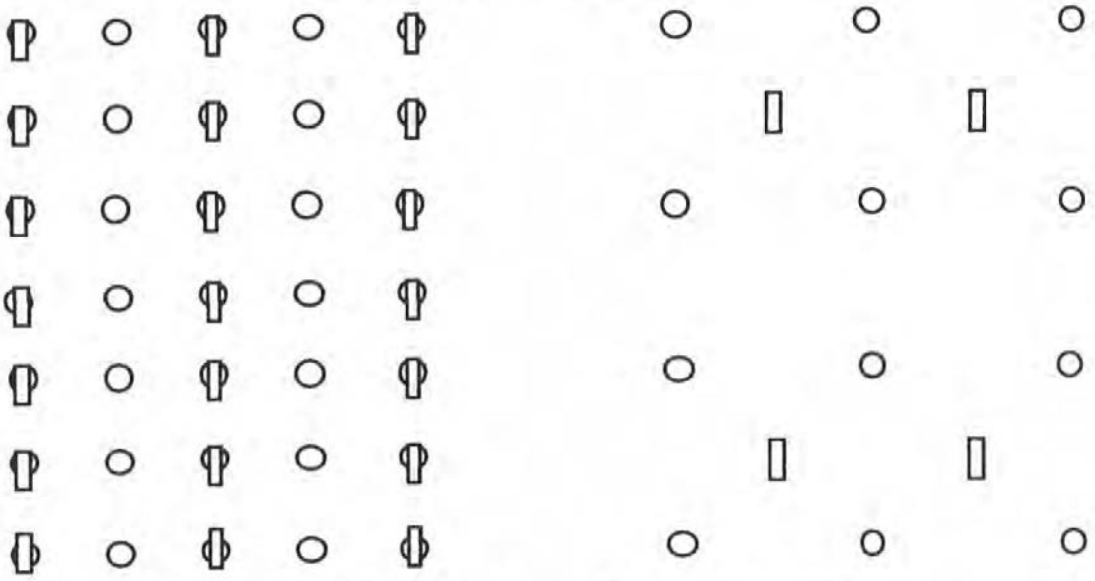
The number of picture elements in the height and width of the picture, in the standards defined above, are too large for effective coding at data rates between 1 and 1,5 Mbit/s. More appropriate values are obtained by decreasing the resolution in both directions to a half. This reduces the pel rate by a factor of four. Interlace should be avoided as it increases the difficulties in achieving low data rates.

One way to reduce the vertical resolution is to use only the odd or the even fields. If the other field is simply discarded, spatial aliasing will be introduced, and this may produce visible and objectionable artifacts. More sophisticated methods of rate conversion require more computational power, but can perceptibly reduce the aliasing artifacts.

The horizontal and vertical resolutions may be halved by filtering and subsampling. Consider a picture in the 4:2:2 format. See the CCIR 601 sampling pattern of figure D.8(a). Such a sampling pattern may be converted to the SIF sampling pattern of figure D.8(b) as follows. The odd field only may be extracted, reducing the number of lines by two, and then a horizontal decimation filter used on the remaining lines to

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

reduce the horizontal resolution by a factor of two. In addition the chrominance values may be vertically decimated. The filters for luminance and chrominance have to be chosen carefully since particular attention has to be given to the location of the samples in the respective International Standards. The temporal relationship between luminance and chrominance must also be correct.



(a) Sampling pattern for 4:2:2 (CCIR 601) (b) Sampling pattern for MPEG (SIF)
Circles represent luminance; Boxes represent Chrominance

Figure D.8 -- Conversion of CCIR 601 to SIF

The following 7-tap FIR filter has been found to give good results in decimating the luminance:

-29	0	88	138	88	0	-29	// 256

Figure D.9 -- Luminance subsampling filter tap weights

Use of a power of two for the divisor allows a simple hardware implementation.

The chrominance samples have to appear in the between the luminance samples both horizontally and vertically. The following linear filter with a phase shift of half a pel may be found useful.

1	3	3	1	// 8

Figure D.10 -- Chrominance subsampling filter tap weights

To recover the samples consistent with the CCIR 601 grid of figure D.8(a), the process of interpolation is used. The interpolation filter applied to a zero-padded signal can be chosen to be equal to the decimation filter employed for the luminance and the two chrominance values in the encoder.

Note that these filters are not part of the International Standard, and other filters may be used.

At the end of the lines some special technique such as renormalizing the filter or replicating the last pel, must be adopted. The following example shows a horizontal line of 16 luminance pels and the same line after filtering and subsampling. In this example the data in the line is reflected at each end.

```
10   12   20   30   35   15   19   11   11   19   26   45   80   90   92   90
   12        32        23         9        12        49        95        92
```

Figure D.11 -- Example of filtering and subsampling of a line of pels

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The result of this filtering and subsampling is a source input format (SIF) which has a luminance resolution of 360 x 240 or 360 x 288, and a chrominance resolution which is half that of the luminance in each dimension.

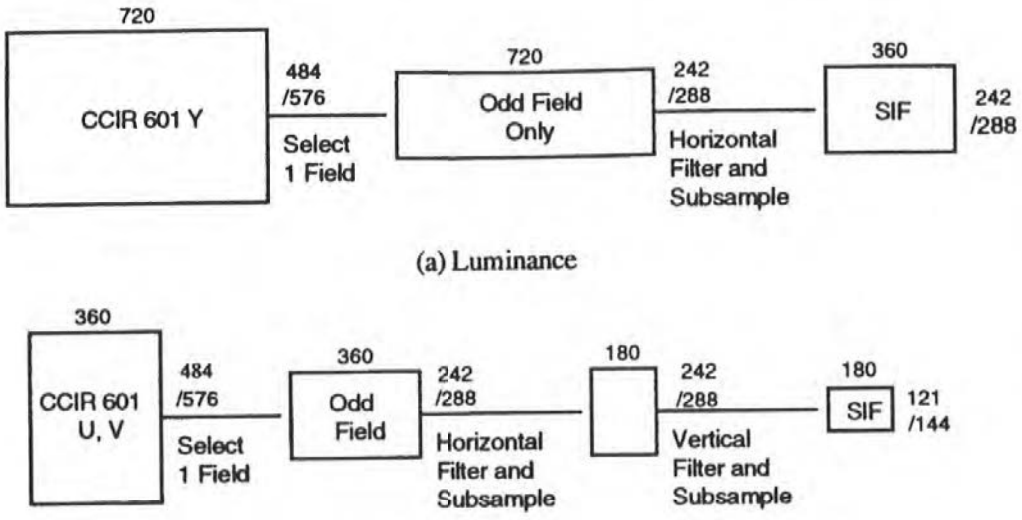


Figure D.12 -- Conversion from CCIR 601 into SIF

The SIF is not quite optimum for processing by MPEG video coders. MPEG video divides the luminance component into macroblocks of 16x16 pels. The horizontal resolution, 360, is not divisible by 16. The same is true of the vertical resolution, 242, in the case of 525-line systems. A better match is obtained in the horizontal direction by discarding the 4 pels at the end of every line of the subsampled picture. Care must be taken that this results in the correct configuration of luminance and chrominance samples in the macroblock. The remaining picture is called the significant pel area, and corresponds to the dark area in figure D.13:

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

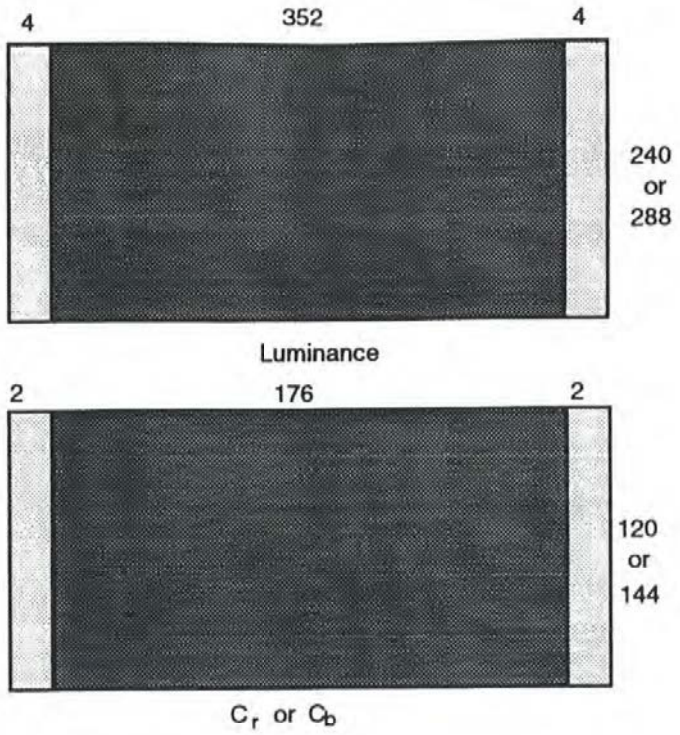


Figure D.13 -- Source input with significant pel area shaded dark

The conversion process is summarized in table D.1.

Table D.1 -- Conversion of source formats

Picture Rate (Hz)	29,97	25
Picture Aspect Ratio (width:height)	4:3	4:3
Luminance (Y)		
CCIR Sample Resolution	720 x 484	720 x 576
SIF	360 x 242	360 x 288
Significant Pel Area	352 x 240	352 x 288
Chrominance (Cb Cr)		
CCIR Sample Resolution	360 x 484	360 x 576
SIF	180 x 121	180 x 144
Significant Pel Area	176 x 120	176 x 144

The preprocessing into the SIF format is not normative, other processing steps and other resolutions may be used. The picture size need not even be a multiple of 16. In this case an MPEG video coder adds padding pels to the right or bottom edges of a picture in order to bring the transmitted resolution up to a multiple of 16, and the decoder discards these after decoding the picture. For example, a horizontal resolution of 360 pels could be coded by adding 8 padding pels to the right edge of each horizontal row bringing the total up to 368 pels. 23 macroblocks would be coded in each row. The decoder would discard the extra padding pels after decoding, giving a final decoded horizontal resolution of 360 pels.

D.3.2 Conversion from film

If film material can be digitized at 24 pictures/s, then it forms an excellent source for an ISO/IEC 11172-2 bitstream. It may be digitized at the desired spatial resolution. The picture_rate field in the video sequence header, see 2.4.2.3, allows the picture rate of 24 pictures/s to be specified exactly.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Sometimes the source material available for compression consists of film material which has been converted to video at some other rate. The encoder may detect this and recode at the original film rate. For example, 24 pictures/s film material may have been digitized and converted to a 30 frame/s system by the technique of 3:2 pulldown. In this mode digitized pictures are shown alternately for 3 and for 2 television field times. This alternation may not be exact since the actual frame rate might be 29,97 frames/s and not the 30 frames/s that the 3:2 pulldown technique gives. In addition the pulldown timing might have been changed by editing and splicing after the conversion. A sophisticated encoder might detect the duplicated fields, average them to reduce digitization noise, and code the result at the original 24 pictures/s rate. This should give a significant improvement in quality over coding at 30 pictures per second, since direct coding at 30 pictures/s destroys the 3:2 pulldown timing and gives a jerky appearance to the final decoded video.

D.4 Model decoder

D.4.1 Need for a decoder model

A coded bitstream contains different types of pictures, and each type ideally requires a different number of bits to encode. In addition, the video may vary in complexity with time, and an encoder may wish to devote more coding bits to one part of a sequence than to another. For constant bitrate coding, varying the number of bits allocated to each picture requires that the decoder have buffering to store the bits not needed to decode the immediate picture. The extent to which an encoder can vary the number of bits allocated to each picture depends on the amount of this buffering. If the amount of the buffering is large an encoder can use greater variations, increasing the picture quality, but at the cost of increasing the decoding delay. Encoders need to know the size of the amount of the decoder's buffering in order to determine to what extent they can vary the distribution of coding bits among the pictures in the sequence.

The model decoder is defined to solve two problems. It constrains the variability in the number of bits that may be allocated to different pictures and it allows a decoder to initialize its buffering when the system is started. It should be noted that Part 1 of this International Standard addresses the initialisation of buffers and the maintenance of synchronisation during playback in the case when two or more elementary streams (for example one audio and one video stream) are multiplexed together. The tools defined in ISO/IEC 11172-1 for the maintenance of synchronisation should be used by decoders when multiplexed streams are being played.

D.4.2 Decoder model

Annex C contains the definition of a parameterized model decoder for this purpose. It is known as a Video Buffer Verifier (VBV). The parameters used by a particular encoder are defined in the bitstream. This really defines a model decoder that is needed if encoders are to be assured that the coded bitstreams they produce will be decodable. The model decoder looks like this:

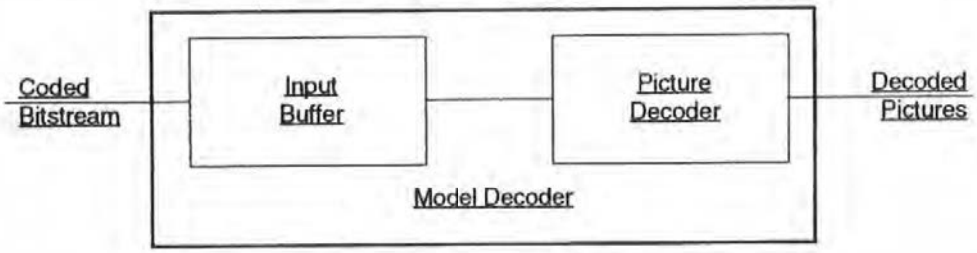


Figure D.14 -- Model decoder

A fixed-rate channel is assumed to put bits at a constant rate into the Input Buffer. At regular intervals, set by the picture rate, the Picture Decoder instantaneously removes all the bits for the next picture from the Input Buffer. If there are too few bits in the Input Buffer, i.e. all the bits for the next picture have not been received, then the Input Buffer underflows and there is an underflow error. If, during the time between picture starts, the capacity of the Input Buffer is exceeded, then there is an overflow error.

Practical decoders differ from this model in several important ways. They may implement their buffering at a different point in the decoder, or distribute it throughout the decoder. They may not remove all the bits required to decode a picture from the Input Buffer instantaneously, they may not be able to control the start

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

of decoding very precisely as required by the buffer fullness parameter in the picture header, and they take a finite time to decode. They may also be able to delay decoding for a short time to reduce the chances of underflow occuring. But these differences depend in degree and kind on the exact method of implementation. To satisfy requirements of different implementations, the MPEG video committee (ISO/IEC JTC1 SC29/WG11) chose a very simple model for the decoder. Practical implementations of decoders must ensure that they can decode the bitstream constrained by this model. In many cases this will be achieved by using an Input Buffer that is larger than the minimum required, and by using a decoding delay that is larger than the value derived from the vbv_delay parameter. The designer must compensate for differences between the actual design and the model in order to guarantee that the decoder can handle any bitstream that satisfies the model.

Encoders monitor the status of the model to control the encoder so that overflow problems do not occur. The calculated buffer fullness is transmitted at the start of each picture so that the decoder can maintain synchronization.

D.4.3 Buffer size and delay

For constant bit rate operation each picture header contains a vbv_delay parameter to enable decoders to start their decoding correctly. This parameter defines the time needed to fill the Input Buffer of figure D.14 from an empty state to the correct level immediately before the Picture Decoder removes all the bits for the picture. This time is thus a delay and is measured in units of 1/90 000 s. This number was chosen because it is almost an exact multiple of the picture durations: 1/24, 1/25, 1/29,97 and 1/30, and because it is comparable in duration to an audio sample.

The delay is given by

$$D = \text{vbv\_delay} \ / \ 90\,000 \text{ s}$$

For example, if vbv_delay were 9 000, then the delay would be 0,1 sec. This means that at the start of a picture the Input Buffer of the model decoder should contain exactly 0,1 s worth of data from the input bitstream.

The bit rate, R, is defined in the sequence header. The number of bits in the Input Buffer at the beginning of the picture is thus given by:

$$B = D * R = \text{vbv\_delay} * R \ / \ 90\,000 \text{ bits}$$

For example, if vbv_delay were 9 000 and R were 1,2 Mbits/s, then the number of bits in the Input Buffer would be 120 000.

The constrained parameter bitstream requires that the Input Buffer have a capacity of 327 680 bits, and B should never exceed this value.

D.5 MPEG video bitstream syntax

This clause describes the video bitstream in a top-down fashion. A sequence is the top level of video coding. It begins with a sequence header which defines important parameters needed by the decoder. The sequence header is followed by one or more groups of pictures. Groups of pictures, as the name suggests, consist of one or more individual pictures. The sequence may contain additional sequence headers. A sequence is terminated by a sequence_end_code. ISO/IEC 11172-1 allows considerable flexibility in specifying application parameters such as bit rate, picture rate, picture resolution, and picture aspect ratio. These parameters are specified in the sequence header.

If these parameters, and some others, fall within certain limits, then the bitstream is called a constrained parameter bitstream.

D.5.1 Sequence

A video sequence commences with a sequence header and is followed by one or more groups of pictures and is ended by a sequence_end_code. Additional sequence headers may appear within the sequence. In each such repeated sequence header, all of the data elements with the permitted exception of those defining quantization matrices (load_intra_quantizer_matrix, load_non_intra_quantizer_matrix and optionally

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

intra_quantizer_matrix and non_intra_quantizer_matrix) shall have the same values as the first sequence header. Repeating the sequence header with its data elements makes random access into the video sequence possible. The quantization matrices may be redefined as required with each repeated sequence header.

The encoder may set such parameters as the picture size and aspect ratio in the sequence header, to define the resources that a decoder requires. In addition, user data may be included.

D.5.1.1 Sequence header code

A coded sequence begins with a sequence header and the header starts with the sequence start code. Its value is:

hex: 00 00 01 B3
binary: 0000 0000 0000 0000 0000 0001 1011 0011

This is a unique string of 32 bits that cannot be emulated anywhere else in the bitstream, and is byte-aligned, as are all start codes. To achieve byte alignment the encoder may precede the sequence start code with any number of zero bits. These can have a secondary function of preventing decoder input buffer underflow. This procedure is called *bit stuffing*, and may be performed before any start code. The stuffing bits must all be zero. The decoder discards all such stuffing bits.

The sequence start code, like all video start codes, begins with a string of 23 zeros. The coding scheme ensures that such a string of consecutive zeros cannot be produced by any other combination of codes, i.e. it cannot be emulated by other codes in the video bitstream. This string of zeros can only be produced by a start code, or by stuffing bits preceding a start code.

D.5.1.2 Horizontal size

This is a 12-bit number representing the width of the picture in pels, i.e. the horizontal resolution. It is an unsigned integer with the most significant bit first. A value of zero is not allowed (to avoid start code emulation) so the legal range is from 1 to 4 095. In practice values are usually a multiple of 16. At 1,5 Mbits/s, a popular horizontal resolution is 352 pels. The value 352 is derived from half the CCIR 601 horizontal resolution of 720, rounded down to the nearest multiple of 16 pels. Otherwise the encoder must fill out the picture on the right to the next higher multiple of 16 so that the last few pels can be coded in a macroblock. The decoder should discard these extra pels before display.

For efficient coding of the extra pels, the encoder should add pel values that reduce the number of bits generated in the transformed block . Replicating the last column of pels is usually superior to filling in the remaining pels with a gray level.

D.5.1.3 Vertical size

This is a 12-bit number representing the height of the picture in pels, i.e. the vertical resolution. It is an unsigned integer with the most significant bit first. A value of zero is not allowed (to avoid start code emulation) so the legal range is from 1 to 4 095. In practice values are usually a multiple of 16. Note that the maximum value of slice_vertical_position is 175 (decimal), which corresponds to a picture height of 2 800 lines. At 1,5 Mbits/s, a popular vertical resolution is 240 to 288 pels. Values of 240 pels are convenient for interfacing to 525-line NTSC systems, and values of 288 pels are more appropriate for 625-line PAL and SECAM systems.

If the vertical resolution is not a multiple of 16 lines, the encoder must fill out the picture at the bottom to the next higher multiple of 16 so that the last few lines can be coded in a macroblock. The decoder should discard these extra lines before display.

For efficient coding, replicating the last line of pels is usually better than filling in the remaining pels with a grey level.

D.5.1.4 Pel aspect ratio

This is a four-bit number which defines the shape of the pel on the viewing screen. This is needed since the horizontal and vertical picture sizes by themselves do not specify the shape of the displayed picture.

The pel aspect ratio does not give the shape directly, but is an index to the following look up table:

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 120 of 341

Table D.2 -- Pel aspect ratio

CODE	HEIGHT/WIDTH	COMMENT
0000	undefined	Forbidden
0001	1,0	square pels
0010	0,6735	
0011	0,7031	16:9 625-line
0100	0,7615	
0101	0,8055	
0110	0,8437	16:9 525-line
0111	0,8935	
1000	0,9157	702x575 at 4:3 = 0,9157
1001	0,9815	
1010	1,0255	
1011	1,0695	
1100	1,0950	711x487 at 4:3 = 1,0950
1101	1,1575	
1110	1,2015	
1111	undefined	reserved

The code 0000 is forbidden to avoid start code emulation. The code 0001 has square pels. This is appropriate for many computer graphics systems. The code 1000 is suitable for displaying pictures on the 625-line 50Hz TV system (see CCIR Recommendation 601).

height / width = 0,75 * 702 / 575 = 0,9157

The code 1100 is suitable for displaying pictures on the 525-line 60Hz TV system (see CCIR Recommendation 601).

height / width = 0,75 * 711 / 487 = 1,0950

The code 1111 is reserved for possible future extensions to this part of ISO/IEC 11172.

The remaining points in the table were filled in by interpolating between these two points 1000 and 1100 using the formula:

$$\text{aspect ratio} = 0{,}5855 + 0{,}044N$$

where N is the value of the code in table D.2. These additional pel aspect ratios might be useful for HDTV where ratios of 16:9 and 5:3 have been proposed.

It is evident that the specification does not allow all possible pel aspect ratios to be specified. We therefore presume that a certain degree of tolerance is allowable. Encoders will convert the actual pel aspect ratio to the nearest value in the table, and decoders will display the decoded values to the nearest pel aspect ratio of which they are capable.

D.5.1.5 Picture rate

This is a four-bit integer which is an index to the following table:

Table D.3 -- Picture rate

CODE	PICTURES PER SECOND
0000	Forbidden
0001	23,976
0010	24
0011	25
0100	29,97
0101	30
0110	50
0111	59,94
1000	60
1001	Reserved
.	.
.	.
1111	Reserved

The allowed picture rates are commonly available sources of analog or digital sequences. One advantage in not allowing greater flexibility in picture rates is that standard techniques may be used to convert to the display rate of the decoder if it does not match the coded rate.

D.5.1.6 Bit rate

The bit rate is an 18-bit integer giving the bit rate of the data channel in units of 400 bits/s. The bit rate is assumed to be constant for the entire sequence. The actual bit rate is rounded up to the nearest multiple of 400 bits/s. For example, a bit rate of 830 100 bits/s would be rounded up to 830 400 bits/s giving a coded bit rate of 2 076 units.

If all 18 bits are 1 then the bitstream is intended for variable bit rate operation. The value zero is forbidden.

For constant bit rate operation, the bit rate is used by the decoder in conjunction with the vbv_delay parameter in the picture header to maintain synchronization of the decoder with a constant rate data channel. If the stream is multiplexed using ISO/IEC 11172-1, the time-stamps and system clock reference information defined in ISO/IEC 11172-1 provide a more appropriate tool for performing this function.

D.5.1.7 Marker bit

The bit rate is followed by a single reserved bit which is always set to 1. This bit prevents emulation of start codes.

D.5.1.8 VBV buffer size

The buffer size is a 10-bit integer giving the minimum required size of the input buffer in the model decoder in units of 16 384 bits (2 048 bytes). For example, a buffer size of 20 would require an input buffer of 20 x 16 384 = 327 680 bits (= 40 960 bytes). Decoders may provide more memory than this, but if they provide less they will probably run into buffer overflow problems while the sequence is being decoded.

D.5.1.9 Constrained Parameter flag

If certain parameters specified in the bitstream fall within predefined limits, then the bitstream is called a constrained parameter bitstream. Thus the constrained parameter bitstream is a standard of performance giving guidelines to encoders and decoders to facilitate the exchange of bitstreams.

The bitrate parameter allows values up to about 100 Mbits/s, but a constrained parameter bitstream must have a bit rate of 1,856 Mbits/s or less. Thus the bit rate parameter must be 3 712 or less.

The picture rate parameter allows picture rates up to 60 pictures/s, but a constrained parameter bitstream must have a picture rate of 30 pictures/s or less.

The resolution of the coded picture is also specified in the sequence header. Horizontal resolutions up to 4 095 pels are allowed by the syntax, but in a constrained parameter bitstream the resolution is limited to 768 pels or less. Vertical resolutions up to 4 095 pels are allowed, but that in a constrained parameter

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 122 of 341

bitstream is limited to 576 pels or less. In a constrained parameter bitstream, the total number of macroblocks per picture is limited to 396. This sets a limit on the maximum area of the picture which is only about one quarter of the area of a 720x576 pel picture. In a constrained parameter bitstream, the pel rate is limited to 2 534 400 pels/s. For a given picture rate, this sets another limit on the maximum area of the picture. If the picture has the maximum area of 396 macroblocks, then the picture rate is restricted to 25 pictures/s or less. If the picture rate has the maximum constrained value of 30 pictures/s the maximum area is limited to 330 macroblocks.

A constrained parameter bitstream can be decoded by a model decoder with a buffer size of 327 680 bits without overflowing or underflowing during the decoding process. The maximum buffer size that can be specified for a constrained parameter bitstream is 20 units.

A constrained parameter bitstream uses a forward_f_code or backward_f_code less than or equal to 4. This constrains the maximum range of motion vectors that can be represented in the bitstream (see table D.7).

If all these conditions are met, then the bitstream is constrained and the constrained_parameters_flag in the sequence header should be set to 1. If any parameter is exceeded, the flag shall be set to 0 to inform decoders that more than a minimum capability is required to decode the sequence.

D.5.1.10 Load intra quantizer matrix

This is a one-bit flag. If it is set to 1, sixty-four 8-bit integers follow. These define an 8 by 8 set of weights which are used to quantize the DCT coefficients. They are transmitted in the zigzag scan order shown in figure D.30. None of these weights can be zero. The first weight must be eight which matches the fixed quantization level of the dc coefficient.

If the flag is set to zero, the intra quantization matrix must be reset to the following default value:

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Figure D.15 -- Default intra quantization matrix

The default quantization matrix is based on work performed by ISO/IEC JTC1 SC29/WG10 (JPEG) [6]. Experience has shown that it gives good results over a wide range of video material. For resolutions close to 350x250 there should normally be no need to redefine the intra quantization matrix. If the picture resolution departs significantly from this nominal resolution, then some other matrix may give perceptibly better results.

The weights increase to the right and down. This reflects the human visual system which is less sensitive to quantization noise at higher frequencies.

D.5.1.11 Load non–intra quantizer matrix

This is a one-bit flag. If it is set to 1, sixty-four 8-bit integers follow in zigzag scan order. None of these integers can be zero.

If the flag is set to zero, the non-intra quantization matrix must be reset to the following default value which consists of all 16s.

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Figure D.16 -- Default non-intra quantization matrix

This flat default quantization matrix was adopted from H.261 which uses a flat matrix for the equivalent of P-pictures [5]. Little work has been performed to determine the optimum non-intra matrix for MPEG video coding, but evidence suggests that it is more dependent on video material than is the intra matrix. The optimum non-intra matrix may be somewhere between the flat default non-intra matrix and the strongly frequency-dependent values of the default intra matrix.

D.5.1.12 Extension data

This start code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B5
binary: 0000 0000 0000 0000 0000 0001 1011 0101

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes are reserved for future extensions to this part of ISO/IEC 11172, and should not be generated by encoders. MPEG video decoders should have the capability to discard any extension data found.

D.5.1.13 User data

A user data start code may follow the optional extension data. This start code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B2
binary: 0000 0000 0000 0000 0000 0001 1011 0010

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes can be used by the encoder for any purpose. The only restriction on the data is that they cannot emulate a start code, even if not byte-aligned. This means that a string of 23 consecutive zeros must not occur. One way to prevent emulation is to force the most significant bit of alternate bytes to be a 1.

In closed encoder-decoder systems the decoder may be able to use the data. In the more general case, decoders should be capable of discarding the user data.

D.5.2 Group of pictures

Two distinct picture orderings exist, the display order and the bitstream order (as they appear in the video bitstream). A group of pictures (gop) is a set of pictures which are contiguous in display order. A group of pictures must contain at least one I-picture. This required picture may be followed by any number of I and P-pictures. Any number of B-pictures may be interspersed between each pair of I or P-pictures, and may also precede the first I-picture.

Property 1. A group of pictures, in bitstream order, must start with an I-picture and may be followed by any number of I, P or B-pictures in any order.

Property 2. Another property of a group of pictures is that it must begin, in display order, with an I or a B-picture, and must end with an I or a P-picture. The smallest group of pictures consists of a single I-picture, whereas the largest size is unlimited.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The original concept of a group of pictures was a set of pictures that could be coded and displayed independently of any other group. In the final version of this part of ISO/IEC 11172 this is not always true, and any B-pictures preceding (in display order) the first I-picture in a group may require the last picture in the previous group in order to be decoded. Nevertheless encoders can still construct groups of pictures which are independent of one another. One way to do this is to omit any B-pictures preceding the first I-picture. Another way is to allow such B-pictures, but to code them using only backward motion compensation.

Property 3. From a coding point of view, a concisely stated property is that a group of pictures begins with a group of pictures header, and either ends at the next group of pictures header or at the next sequence header or at the end of sequence, whichever comes first.

Some examples of groups of pictures are given below:

```
I
I  P  P
I  B  P  B  P
B  B  I  B  P  B  P
B  B  I  B  B  P  B  B  P  B  B  P
B  I  B  B  B  B  P  B  I  B  B  I  I
```

Figure D.17 -- Examples of groups of pictures in display order

These examples illustrate what is possible, and do not constitute a suggestion for structures of groups of pictures.

Group of pictures start code

The group of pictures header starts with the Group of Pictures start code. This code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B8
binary: 0000 0000 0000 0000 0000 0001 1011 1000

It may be preceded by any number of zeros. The encoder may have inserted some zeros to get byte alignment, and may have inserted additional zeros to prevent buffer underflow. An editor may have inserted zeros in order to match the vbv_delay parameter of the first picture in the group.

Time code

A time code of 25 bits immediately follows the group of pictures start code. This encodes the same information as the SMPTE time code [4].

The time code can be broken down into six fields as shown in table D.4.

Table D.4 -- Time code fields

FIELD	BITS	VALUES
Drop frame flag	1	
Hours	5	0 to 23
Minutes	6	0 to 59
Fixed	1	1
Seconds	6	0 to 59
Picture number	6	0 to 60

The time code refers to the first picture in the group in display order, i.e. the first picture with a temporal reference of zero. The SMPTE time code is included to provide a video time identification to applications. It may be discontinuous. The presentation time-stamp in the System layer (Part 1) has a much higher precision and identifies the time of presentation of the picture.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Closed GOP

A one bit flag follows the time code. It denotes whether the group of pictures is open or closed. Closed groups can be decoded without using decoded pictures of the previous group for motion compensation, whereas open groups require such pictures to be available.

A typical example of a closed group is shown in figure D.18a.

I	B	B	**P**	B	B	**P**	B	B	**P**	B	B	**P**
0	1	2	3	4	5	6	7	8	9	10	11	12

(a) closed group

B	B	**I**	B	B	**P**	B	B	**P**	B	B	**P**	B	B	**P**
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14

(b) open or closed group

Figure D.18 -- Example groups of pictures in display order

A less typical example of a closed group is shown in figure D.18b. In this example, the B-pictures which precede the first I-picture must use backward motion compensation only, i.e. any motion compensation must be based only on picture number 2 in the group.

If the closed_gop flag is set to 0 then the group is open. The first B-pictures that precede the first I-picture in the group may have been encoded using the last picture in the previous group for motion compensation.

Broken link

A one bit flag follows the closed_gop flag. It denotes whether the B-pictures which precede the first I-picture in the GOP can be correctly decoded. If it is set to 1, these pictures cannot be correctly decoded because the I-picture or P-picture from the previous group pictures that is required to form the predictions is not available (presumably because the preceding group of pictures has been removed by editing). The decoder will probably choose not to display these B-pictures.

If the sequence is edited so that the original group of pictures no longer precedes the current group of pictures then this flag normally will be set to 1 by the editor. However, if the closed_gop flag for the current group of pictures is set, then the editor should not set the broken_link flag. Because the group of pictures is closed, the first B-pictures (if any) can still be decoded correctly.

Extension data

This start code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B5
binary: 0000 0000 0000 0000 0000 0001 1011 0101

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes are reserved for future extensions to this part of ISO/IEC 11172, and should not be generated by encoders. MPEG video decoders should have the capability to discard any extension data found.

User data

A user data start code may follow the optional extension data. This start code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B2
binary: 0000 0000 0000 0000 0000 0001 1011 0010

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes can be used by the encoder for any purpose. The only restriction on the data is that they cannot emulate a start code, even if not byte-aligned.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 126 of 341

This means that a string of 23 consecutive zeros must not occur. One way to prevent emulation is to force the most significant bit of alternate bytes to be a 1.

In closed encoder-decoder systems the decoder may be able to use the data. In the more general case, decoders should be capable of discarding the user data.

D.5.3 Picture

The picture layer contains all the coded information for one picture. The header identifies the temporal reference of the picture, the picture coding type, the delay in the video buffer verifier (VBV) and, if appropriate, the range of motion vectors used.

D.5.3.1 Picture header and start code

A *picture* begins with a picture header. The header starts with a picture start code. This code is byte-aligned and is 32 bits long. Its value is:

hex: 00 00 01 00
binary: 0000 0000 0000 0000 0000 0001 0000 0000

It may be preceded by any number of zeros.

D.5.3.2 Temporal reference

The Temporal Reference is a ten-bit number which can be used to define the order in which the pictures must be displayed. It may be useful since pictures are not transmitted in display order, but rather in the order which the decoder needs to decode them. The first picture, in display order, in each group must have Temporal Reference equal to zero. This is incremented by one for each picture in the group.

Some example groups of pictures with their Temporal Reference numbers are given below:

Example (a) in display order	**I** 0	B 1	**P** 2	B 3	**P** 4							
Example (a) in decoding order	**I** 0	**P** 2	B 1	**P** 4	B 3							
Example (b) in display order	B 0	B 1	**I** 2	B 3	B 4	**P** 5	B 6	B 7	**P** 8	B 9	B 10	**P** 11
Example (b) in coded order	**I** 2	B 0	B 1	**P** 5	B 3	B 4	**P** 8	B 6	B 7	**P** 11	B 9	B 10
Example (c) in display order	B 0	**I** 1	B 2	B 3	B 4	B 5	**P** 6	B 7	**I** 8	B 9	B 10	**I** 11	**I** 12
Example (c) in coded order	**I** 1	B 0	**P** 6	B 2	B 3	B 4	B 5	**I** 8	B 7	**I** 11	B 9	B 10	**I** 12

Figure D.19 -- Examples of groups of pictures and temporal references

If there are more than 1024 pictures in a group, then the Temporal Reference is reset to zero and then increments anew. This is illustrated below:

B	B	I	B	B	P	...	P	B	B	P	...	P	B	B	P	display order
0	1	2	3	4	5	...	1 022	1 023	0	1	...	472	473	474	475	

Figure D.20 -- Example group of pictures containing 1 500 pictures

D.5.3.3 Picture coding type

A three bit number follows the temporal reference. This is an index into the following table defining the type of picture.

Table D.5 -- Picture types

CODE	PICTURE TYPE
000	Forbidden
001	I-picture
010	P-picture
011	B-picture
100	D Picture
101	Reserved
110	Reserved
111	Reserved

The various types of pictures are described in D.2.3. Codes 101 through 111 are reserved for future extensions to this part of ISO/IEC 11172. Decoders should be capable of discarding all pictures of this type, and scan for the next picture start code, group start code or sequence start code. Code 000 will never be used to avoid start code emulation.

D.5.3.4 VBV delay

For constant bit rate operation, vbv_delay can be used at the start of decoding and after a random access to ensure that the correct number of bits have been read by the decoder before the first picture is displayed.

The buffer fullness is not specified in bits but rather in units of time. The vbv_delay is a 16-bit number defining the time needed in units of 1/90 000 s to fill the input buffer of the model decoder from an empty state to the correct state at the bit rate specified in the sequence header.

For example, suppose the vbv_delay had a decimal value of 30000, then the time delay would be:

$$D = 30\,000 / 90\,000 = 1 / 3 \text{ s}$$

If the channel bit rate were 1,2 Mbits/s then the contents of the buffer before the picture is decoded would be:

$$B = 1\,200\,000 / 3 = 400\,000 \text{ bits}$$

If the decoder determined that its actual buffer fullness differed significantly from this value, then it would have to adopt some strategy for regaining synchronization.

The meaning of vbv_delay is undefined for variable bit rate operation.

D.5.3.5 Full pel forward vector

This is a one bit flag giving the precision of the forward motion vectors. If it is 1 then the precision of the vectors is in integer pels, if it is zero then the precision is half a pel. Thus if the flag is set to one the vectors have twice the range than they do if the flag set to zero.

This flag is present only in the headers of P-pictures and B-pictures. It is absent in I-pictures and D pictures.

D.5.3.6 Forward f-code

This is a three-bit number and, like the full pel forward vector flag, is present only in the headers of P-pictures and B-pictures. It provides information used for decoding the coded forward vectors and controls the maximum size of the forward vectors that can be coded. It can take only values of 1 through 7; a value of zero is forbidden.

Two parameters used in decoding the forward motion vectors are derived from forward_f_code, forward_r_size and forward_f.

The forward_r_size is one less than the forward_f_code and so can take values 0 through 6.

The forward_f parameter is given by table D.6:

Table D.6 -- f_codes

forward/backward_f_code	forward/backward_f
1	1
2	2
3	4
4	8
5	16
6	32
7	64

D.5.3.7 Full pel backward vector

This is a one bit flag giving the precision of the backward motion vectors. If it is 1 then the precision of the vectors is in integer pels, if it is zero then the precision is half a pel. Thus if the flag is set to one the vectors have twice the range than they do if the flag set to zero.

This flag is only present in the headers of B-pictures. It is absent in I-pictures, P-pictures and D pictures.

D.5.3.8 Backward f-code

This is a three-bit number and, like the full pel backward vector flag, is present only in the headers of B-pictures. It provides information used for decoding the coded backward vectors. It can take only values of 1 through 7; a value of zero is forbidden.

The backward_f parameter is derived from the backward_f_code and is given by table D.6

D.5.3.9 Extra picture information

Extra picture information is the next field in the picture header. Any number of information bytes may be present. An information byte is preceded by a flag bit which is set to 1. Information bytes are therefore generally not byte-aligned. The last information byte is followed by a zero bit. The smallest size of this field is therefore one bit, a 0, that has no information bytes. The largest size is unlimited. The following example has 16 bits of extra information denoted by E:

1 E E E E E E E E 1 E E E E E E E E 0

Where E is an extra information bit.

The extra information bytes are reserved for future extensions to this part of ISO/IEC 11172. The meaning of these bytes is currently undefined, so encoders must not generate such bytes and decoders must be capable of discarding them.

D.5.3.10 Extension data

This start code is byte-aligned and is 32 bits long. Its value is:

hex: 00 00 01 B5
binary: 0000 0000 0000 0000 0000 0001 1011 0101

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes are reserved for future extensions to this part of ISO/IEC 11172, and should not be generated by encoders. MPEG video decoders must be capable of discarding them.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.5.3.11 User data

This start code is byte-aligned and is 32 bits long. Its value is

hex: 00 00 01 B2
binary: 0000 0000 0000 0000 0000 0001 1011 0010

It may be preceded by any number of zeros. If it is present then it will be followed by an undetermined number of data bytes terminated by the next start code. These data bytes can be used by the encoder for any purpose. The only restriction on the data is that they cannot emulate a start code, even if not byte-aligned. One way to prevent emulation is to force the most significant bit of alternate bytes to be a 1.

In closed encoder-decoder systems the decoder may be able to use the data. In the more general case, decoders should be capable of discarding the user data.

D.5.4 Slice

Pictures are divided into slices. Each slice consists of an integral number of macroblocks in raster scan order. Slices can be of different sizes within a picture, and the division in one picture need not be the same as the division in any other picture. Slices can begin and end at any macroblock in a picture subject to the following restrictions. The first slice must begin at the top left of the picture, and the end of the last slice must be the bottom right macroblock of the picture. There can be no gaps between slices, nor can slices overlap. The minimum number of slices in a picture is one, the maximum number is equal to the number of macroblocks.

Each slice starts with a slice start code, the exact value of which defines the vertical position of the slice. This is followed by a code that sets the quantization step-size. At the start of each slice the predictors for the dc coefficient values and the predictors for the vector decoding are all reset. The horizontal position of the start of the slice is given by the macroblock address of the first macroblock in the slice. The result of all this is that, within a picture, a slice can be decoded without information from the previous slices. Therefore, if a data error occurs, decoding can begin again at the subsequent slice.

If the data are to be used in an error free environment, then one slice per picture may be appropriate. If the environment is noisy, then one slice per row of macroblocks may be more desirable, as shown in figure D.21.

1 begin	end 1
2 begin	end 2
3 begin	end 3
4 begin	end 4
5 begin	end 5
6 begin	end 6
7 begin	end 7
8 begin	end 8
9 begin	end 9
10 begin	end 10
11 begin	end 11
12 begin	end 12
13 begin	end 13

Figure D.21 -- Possible arrangement of slices in a 256x192 picture

In this figure and in the next, each strip is one macroblock high, i.e. 16 pels high.

Since each slice header requires 40 bits, there is some penalty for including more than the minimum number of slices. For example, a sequence with a vertical resolution of 240 lines coded at 30 pictures/s requires approximately 40x30 = 1 200 bits/s for the slice headers using one slice per picture, and 40x15x30 = 18 000 bits/s with one slice per row, an additional overhead of 16 800 bits/s. The calculation is approximate and underestimates the impact, since the inclusion of a slice imposes additional requirements that the macroblock immediately before the slice header be coded, as well as the first macroblock in the slice.

The coding structure permits great flexibility in dividing a picture up into slices. One possible arrangement is shown in figure D.22.

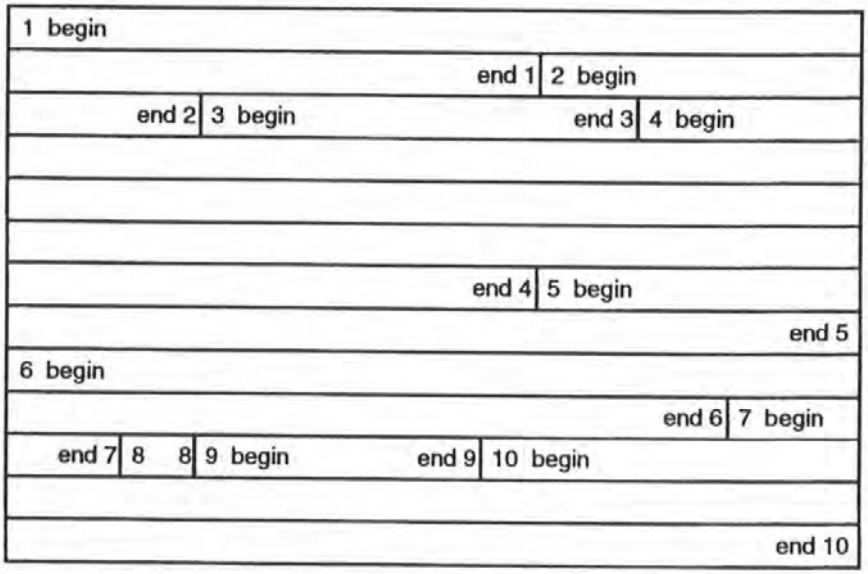


Figure D.22 -- Possible arrangement of slices in a 256x192 picture

This division into slices is given for illustrative purposes only. It is not intended as a suggestion on how to divide a picture into slices.

D.5.4.1 Slice header and start code

Slices start with a slice header. Each slice header starts with a slice start code. This code is byte-aligned and is 32 bits long. The last eight bits can take on a range of values which define the vertical position of the slice in the picture. The permitted slice start codes are:

hex:	from	00 00 01 01
	to	00 00 01 AF
binary:	from	0000 0000 0000 0000 0000 0001 0000 0001
	to	0000 0000 0000 0000 0000 0001 1010 1111

Each slice start code may be preceded by any number of zeros.

The last 8 bits of the slice start code give the slice vertical position, i.e the vertical position of the first macroblock in the slice in units of macroblocks starting with position 1 at the top of the picture. A useful variable is macroblock row. This is similar to slice vertical position except that row 0 is at the top of the picture. Thus

slice vertical position = macroblock row + 1

For example, a slice start code of 00000101 hex means that the first macroblock in the slice is at vertical position 1 or macroblock row 0, i.e. at the top of the picture. A slice start code of 00000120 hex means

that the first macroblock is at vertical position 32 or macroblock row 31, i.e. at the 496th row of pels. It is possible for two or more slices to have the same vertical position.

The maximum vertical position is 175 units. A slice with this position would require a vertical size of 175x16 = 2 800 pels.

The horizontal position of the first macroblock in the slice can be calculated from its macroblock address increment. Thus its position in the picture can be determined without referring to any previous slice or macroblock. Thus a decoder may decode any slice in a picture without having decoded any other slice in the same picture. This feature allows decoders to recover from bit errors by searching for the next slice start code and then resuming decoding.

D.5.4.2 Quantizer scale

The quantizer scale is a five-bit integer which is used by the decoder to calculate the DCT coefficients from the transmitted quantized coefficients. A value of 0 is forbidden, so the quantizer scale can have any value between 1 and 31 inclusive.

Note in addition that the quantizer scale may be set at any macroblock.

D.5.4.3 Extra slice information

Extra slice information forms the last field in the slice header. Any number of information bytes may be present. An information byte is preceded by a flag bit which is set to 1. Information bytes are therefore generally not byte-aligned. The last information byte is followed by a zero bit. The smallest size of this field is therefore one bit, a 0, that has no information bytes. The largest size is unlimited. The following example has 24 bits of extra information denoted by E:

1 E E E E E E E E 1 E E E E E E E E 1 E E E E E E E E 0

The extra information bytes are reserved for future extensions to this part of ISO/IEC 11172. The meaning of these bytes is currently undefined, so encoders must not generate such bytes and decoders must discard them.

The slice header is followed by code defining the macroblocks in the slice.

D.5.5 Macroblock

Slices are divided into macroblocks of 16 x 16 pels. Macroblocks are coded with a header that contains information on the macroblock address, macroblock type, and the optional quantizer scale. The header is followed by data defining each of the six blocks in the macroblock It is convenient to discuss the macroblock header fields in the order in which they are coded.

D.5.5.1 Macroblock stuffing

The first field in the macroblock header is "macroblock stuffing". This is an optional field, and may be inserted or omitted at the discretion of the encoder. If present it consists of any number of 11-bit strings with the pattern "0000 0001 111". This stuffing code is used by the encoder to prevent underflow, and is discarded by the decoder. If the encoder determines that underflow is about to occur, then it can insert as many stuffing codes into the first field of the macroblock header it likes.

Note that an encoder has other strategies to prevent buffer underflow. It can insert stuffing bits immediately before a start code. It can reduce the quantizer scale to increase the number of coded coefficients. It can even start a new slice.

D.5.5.2 Macroblock address increment and macroblock escape

Macroblocks have an address which is the number of the macroblock in raster scan order. The top left macroblock in a picture has address 0, the next one to the right has address 1 and so on. If there are M macroblocks in a picture, then the bottom right macroblock has an address M-1.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The address of a macroblock is indicated by transmitting the difference between the addresses of the current macroblock and the previously coded macroblock. This difference is called the macroblock address increment. In I-pictures, all macroblocks are coded and so the macroblock address increment is nearly always one. There is one exception. At the beginning of each slice the macroblock address is set to that of the right hand macroblock of the previous row. At the beginning of the picture it is set to -1. If a slice does not start at the left edge of the picture, then the macroblock address increment for the first macroblock in the slice will be larger than one. For example, the picture of figure D.22 has 16 macroblocks per row. At the start of slice 2 the macroblock address is set to 15 which is the address of the macroblock at the right hand edge of the top row of macroblocks. If the first slice contained 26 macroblocks, 10 of them would be in the second row, so the address of the first macroblock in slice 2 would be 26 and the macroblock address increment would be 11.

Macroblock address increments are coded using the VLC codes in the table in B.1.

It can be seen that there is no code to indicate a macroblock address increment of zero. This is why the macroblock address is set to -1 rather than zero at the top of a picture. The first macroblock will have an increment of one making its address equal to zero.

The macroblock address increments allow the position of the macroblock within the picture to be determined. For example, assume that a slice header has the start code equal to 00 00 01 0A hex, that the picture width is 256 pels, and that a macroblock address increment code 0000111 is in the macroblock header of the first macroblock in the slice. A picture width of 256 pels implies that there are 16 macroblocks per row in this picture. The slice start code tells us that the slice vertical position is 10, and so the macroblock row is 9. The slice header sets the previous macroblock address to the last macroblock on row 8, which has address 143. The macroblock address increment VLC leads to a macroblock address increment of 8, and so the macroblock address of the first macroblock in the slice is 143 + 8 = 151.

The macroblock row may be calculated from the address:

macroblock row = macroblock address / macroblock width

= 151 / 16

= 9

The division symbol signifies integer truncation, not rounding.

The macroblock column may also be calculated from the address:

macroblock column = macroblock address % macroblock width
= 151 % 16
= 7

Columns are numbered from the left of the picture starting at 0.

There are two special codewords: escape and stuffing.

The escape code means "add 33 to the following macroblock address increment". This allows increments greater than 33 to be coded. For example, an increment of 40 would be coded as escape plus an increment of 7:

0000 0001 0000 0010

An increment of 70 would be coded as two escape codes followed by the code for an increment of 4:

0000 0001 0000 0000 0010 0000 11

The stuffing code is included since the decoder must be able to distinguish it from increment codes. It is used by the encoder to prevent underflow, and is discarded by the decoder.

D.5.5.3 Macroblock types

Each of the picture types I, P, and B, have their own macroblock types. See, respectively, D.6.3, D.6.4, and D.6.5 for the codes and their descriptions.

D.5.5.4 Motion horizontal/vertical forward/backward codes

The interpretation of these codes is explained in D.6.2.3.

D.5.5.5 Motion horizontal/vertical forward/backward R

The interpretation of these codes is explained in D.6.2.3.

D.5.5.6 Coded block pattern

This code describes which blocks within the macroblock are coded and transmitted. The interpretation of this code is explained in D.6.4.2.

D.5.5.7 End of macroblock

This code is used only in D-pictures and is described in D.6.6.

D.5.6 Block

A block is an array of 8 by 8 component pel values, treated as a unit and input to the Discrete Cosine Transform (DCT). Blocks of 8 by 8 pels are transformed into arrays of 8 by 8 DCT coefficients using the two dimensional discrete cosine transform.

D.6 Coding MPEG video

D.6.1 Rate control and adaptive quantization

The encoder must control the bit rate so that the model decoder input buffer neither overflows nor underflows. Since the model decoder removes all the bits associated with a picture from the input buffer instantaneously, it is necessary to control only the total number of bits per picture. The encoder should allocate the total numbers of bits among the various types of pictures so that the perceived quality is suitably balanced. The distribution will vary with the scene content and with the particular distribution of the three picture types (I, P and B-pictures).

Within a picture the encoder should allocate the total number of bits available among the macroblocks to maximize the visual quality of the picture.

One method by which an encoder controls the bit rate is to vary the quantizer scale. This is set in each slice header, and may be set at the beginning of any macroblock, giving the encoder excellent control over the bit rate within a picture.

D.6.1.1 Rate control within a sequence

For a typical coding scheme represented by the following group of pictures in display order:

B B I B B P B B P B B P B B P

it has been found that good results can be obtained by matching the visual quality of the I and P-pictures, and by reducing the code size of the B-pictures to save bits giving a generally lower quality for the B-pictures.

The best allocation of bits among the picture types depends on the scene content. Work of the MPEG video committee suggests that allotting P-pictures about 2-5 times as many bits as B-pictures, and allotting I-pictures up to 3 times as many bits as P-pictures gives good results for typical natural scenes. If there is little motion or change in the video, then a greater proportion of the bits should be allotted to the I-pictures.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

If there is a lot of motion or change, then the proportion allotted to I-pictures should be reduced and most of the savings given to the P-pictures.

A reasonable encoder algorithm is to start with the foregoing estimates, then reallocate bits dynamically depending on the nature of the video.

D.6.1.2 Rate control within a picture

If the buffer is heading toward overflow, the quantizer scale should be increased. If this action is not sufficient to prevent an impending overflow then, as a last resort, the encoder could discard high frequency DCT coefficients and transmit only low frequency ones. Although this would probably produce visible artifacts in the decoded video, it would in no way compromise the validity of the coded bitstream.

If the buffer is heading toward underflow, the quantizer scale should be reduced. If this is not sufficient, the encoder can insert macroblock stuffing into the bitstream, or add leading zeros to start codes.

Under normal circumstances, the encoder calculates and monitors the state of the model decoder buffer and changes the quantizer scale to avert both overflow and underflow problems.

One simple algorithm that helps accomplish this is to monitor the buffer fullness. Assume that the bits have been allocated among the various picture types, and that an average quantizer scale for each picture type has been established. The actual buffer fullness at any macroblock in a picture can be calculated and compared with the nominal fullness, i.e. the value that would be obtained if the bits were uniformly distributed among all the macroblocks in the picture. If the buffer fullness is larger than the nominal value, then the quantizer scale should be set higher than the average, whereas if the buffer fullness is smaller than the nominal, the quantizer scale should be set lower than the average.

If the quantizer scale is kept constant over a picture, then, for a given number of coding bits, the total mean square error of the coded picture will tend to be close to the minimum. However, the visual appearance of most pictures can be improved by varying the quantizer scale over the picture, making it smaller in smooth areas of the picture and larger in busy areas. This technique reduces the visibility of blockiness in smooth areas at the expense of increased quantization noise in the busy areas where, however, it is masked by the image detail.

Thus a good algorithm for controlling the bitrate within a picture adjusts the quantizer scale depending on both the calculated buffer fullness and on the local image content. Examples of techniques for rate control and quantization may be found in [7][8].

D.6.1.3 Buffer fullness

To give the best visual quality, the encoder should almost fill the input buffer before instructing the decoder to start decoding.

D.6.2 Motion estimation and compensation

D.6.2.1 Motion compensation

P-pictures use motion compensation to exploit temporal redundancy in the video. Decoders construct a predicted block of pels from pels in a previously transmitted picture. Motion within the pictures (e.g. a pan) usually implies that the pels in the previous picture will be in a different position from the pels in the current block, and the displacement is given by motion vectors encoded in the bitstream. The predicted block is usually a good estimate of the current block, and it is usually more efficient to transmit the motion vector plus the difference between the predicted block and the current block, than to transmit a description of the current block by itself.

Consider the following typical group of pictures.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

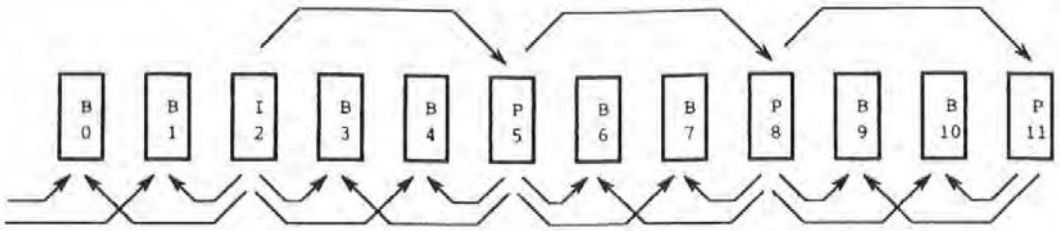


Figure D.23 -- Group of pictures in display order

The I-picture, picture 2, is decoded without requiring any motion vectors. The first P-picture, number 5, is decoded using motion vectors from picture 2. This motion compensation is called forward motion compensation since it is forward in time. Motion vectors define the motion of a macroblock, i.e. the motion of a 16x16 block of luminance pels and the associated chrominance components. Typically, most macroblocks in a P-picture use motion compensation. Non-zero motion vectors are transmitted differentially with reference to the last transmitted motion vector.

The transmitted vectors usually have a precision of half a pel. The maximum range of the vector is set by the forward_f parameter in the picture header. Sometimes, if the motion is unusually large, the range may be doubled and the accuracy reduced to integer pels, by the full_pel_forward_vector bit in the picture header.

A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pels in the referenced picture which are spatially to the right or below the pels being predicted.

Not all macroblocks in a P-picture necessarily use motion compensation. Some macroblocks, as defined by the transmitted macroblock type (see table B.2b), may be intra-coded, and these are reconstructed without motion compensation. Full details defining the method of decoding the vectors and constructing the motion-compensated macroblock are given in 2.4.4.2.

P-picture 8 in figure D.23 uses forward motion compensation from picture 5. P-pictures always use forward motion compensation from the last transmitted I or P-picture.

B-pictures may use motion compensation from the previous I or P-picture, from the next (in display order) I or P-picture, or both; i.e., from the last two transmitted I or P-pictures.

Prediction is called forward if reference is made to a picture in the past and called backward if reference is made to a picture in the future. For example, B-picture 3 in figure D.23 uses forward motion compensation from I-picture 2, and backward motion compensation from P-picture 5. B-pictures may use both forward and backward motion compensation and average the result. This operation is called interpolative motion compensation.

All three types of motion compensation are useful, and typically are used in coding B-pictures. Interpolative motion compensation has the advantage of averaging any noise present. Forward or backward motion compensation may be more useful near the edges of pictures, or where a foreground object is passing in front of a fixed or slow moving background.

Note that this technique of coding with P and B-pictures increases the coding efficiency. B-pictures can have greater errors of reconstruction than I or P-pictures to conserve coding bits, but since they are not used as the basis of motion compensation for future pictures, these errors may be tolerated.

D.6.2.2 Motion estimation

Motion compensation in a decoder is straightforward, but motion estimation which includes determining the best motion vectors and which must be performed by the encoder, presents a formidable computational challenge.

Various methods are available to the encoder. The more computationally intensive methods tend to give better results, so there is tradeoff to be made in the encoder: computational power, and hence cost, versus coded video quality.

Using a search strategy the encoder attempts to match the pels in a macroblock with those in a previous or future picture. The vector corresponding to the best match is reported after the search is completed.

D.6.2.2.1 Block matching criteria

In seeking a match, the encoder must decide whether to use the decoded past and future pictures as the reference, or use the original past and future pictures. For motion estimation, use of the decoded pictures by the encoder gives the smallest error in the error picture, whereas use of the original pictures gives the most accurate motion vectors. The choice depends on whether the artifacts of increased noise, or greater spurious motion are judged to be the more objectionable. There is usually little or no difference in quality between the two methods. Note that the decoder does not perform motion estimation. It performs motion compensated prediction and interpolation using vectors calculated in the encoder and stored in the bitstream. In motion compensated prediction and interpolation, both the encoder and decoder must use the decoded pictures as the references.

Several matching criteria are available. The mean square error of the difference between the motion-compensated block and the current block is an obvious choice. Another possible criterion is the mean absolute difference between the motion-compensated block and the current block.

For half pel shifts, the pel values could be interpolated by several methods. Since the decoder uses a simple linear interpolation, there is little reason to use a more complex method in the encoder. The linear interpolation method given in this part of ISO/IEC 11172 is equivalent to the following. Consider four pels having values A, B, D and E as shown in figure D.24:

```
A   h   B
v   c
D       E
```

Figure D.24 -- Interpolation of half pel shifts

The value of the horizontally interpolated pel is

$$h = (A + B) // 2$$

where the double division symbol means division with rounding to the nearest integer. Half integer values are to be rounded to the next higher value. Thus if A = 4 and B = 9 then h = 6.5 which is rounded up to 7.

The value of the vertically interpolated pel is

$$v = (A + D) // 2$$

The value of the central interpolated pel is

$$c = (A + B + D + E) // 4$$

D.6.2.2.2 Search range

Once a block matching criterion has been selected, some kind of search strategy must be adopted. This must recognize the limitations of the VLC tables used to code the vectors. The maximum range of the vector depends upon forward_f_code or backward_f_code. The motion vector ranges are given in table D.7.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Table D.7 -- Range of motion vectors

forward_f_code or backward_f_code	Motion vector range	
	full_pel=0	full_pel=1
1	-8 to 7,5	-16 to 15
2	-16 to 15,5	-32 to 31
3	-32 to 31,5	-64 to 63
4	-64 to 63,5	-128 to 127
5	-128 to 127,5	-256 to 255
6	-256 to 255,5	-512 to 511
7	-512 to 511,5	-1 024 to 1 023

The range depends on the value of full_pel_forward_vector or full_pel_backward_vector in the picture header. Thus if all the motion vectors were found to be 15 pels or less, the encoder would usually select half pel accuracy and a forward_f_code or backward_f_code value of 2.

The search must be constrained to take place within the boundaries of the decoded reference picture. Motion vectors which refer to pels outside the picture are not allowed. Any bitstream which refers to such pels does not conform to this part of ISO/IEC 11172.

D.6.2.2.3 2-D search strategy

There are many possible methods of searching another picture for the best match to a current block, and a few simple ones will be described.

The simplest search is a full search. Within the chosen search range all possible displacements are evaluated using the block matching criterion.

The full search is computationally expensive, and practical encoders may not be able to afford the time required for a full search.

A simple modification of the full search is to search using only integer pel displacements. Once the best integer match has been found, the eight neighbouring half-integer pel displacements are evaluated, and the best one selected as illustrated below:

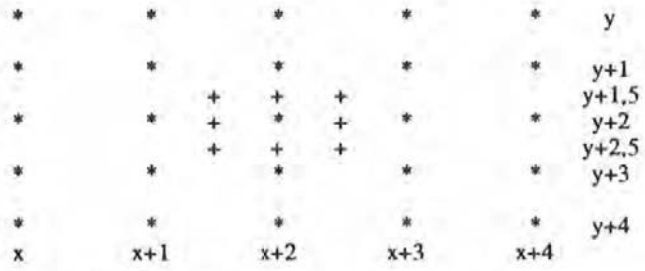


Figure D.25 -- Integer pel and half pel displacements

Assume that the position x+2,y+2 gives the best integer displacement matching using the selected block matching criterion, then the encoder would evaluate the eight positions with half pel displacements marked by + signs in figure D.25. If one of them were a better match then it would become the motion vector, otherwise the motion vector would remain that of the integer displacement x+2,y+2.

If during the integer pel search, two or more positions have the same block matching value, the encoder can adopt a consistent tie-breaking rule.

The modified full search algorithm is approximately an order of magnitude simpler than the full search. Using only integer displacements for the first stage of the search reduces the number of evaluations by a factor of four. In addition, the evaluations are simpler since the pel differences can be calculated directly and do not have to be interpolated.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

For some applications even the modified full search may be too time consuming, and a faster search method may be required. One such method is the logarithmic search.

D.6.2.2.4 Logarithmic search

In this search method, grids of 9 displacements are examined, and the search continued based on a smaller grid centered on the position of the best match. If the grids are reduced in size by a factor of 3 at each step then the search is maximally efficient in the sense that any integer shift has a unique selection path to it. This method will find the best match only for a rather limited set of image types. A more robust method is to reduce the size of the grids by a smaller factor at each step, e.g. by a factor of 2. The scaling factors can also be adjusted to match the search ranges of table D.7.

The method will be illustrated with an example. Consider the set of integer shifts in figure D.26:

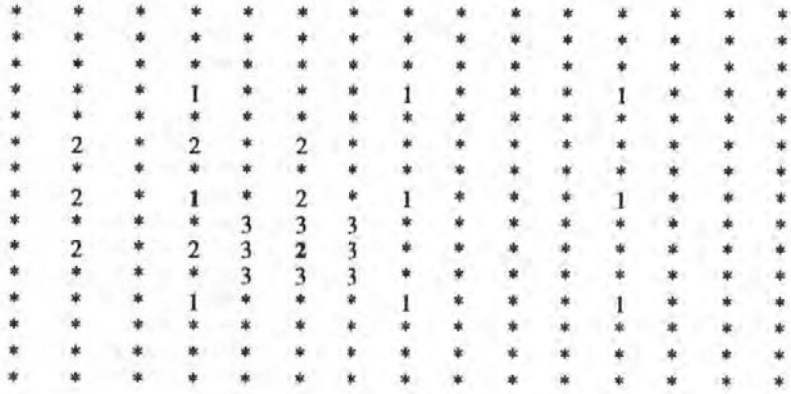


Figure D.26 -- Logarithmic search method for integer pel shifts

The first grid has a spacing of 4 pels. The first step examines pels at shifts of 0, 4, or -4 pels in each direction, marked 1 in figure D.26. The best position is used as the center point of the second grid. Assume it is the pel marked 1 directly to the left of the center pel. The second grid has a spacing of 2 pels. The second step examines pels at shifts of 0, 2, or -2 pels in each direction from the center of the new grid, marked 2 in the figure. The best position is used as the center point of the third grid, assume it is the lower right pel of the second grid. The third grid has a spacing of 1 pel. The third step examines pels at shifts of 0, 1, or -1 pels in each direction from the center of the grid. The best position is used as the center point of the fourth grid. The fourth grid has a spacing of 1/2 pel. The fourth step examines pels at shifts of 0, 1/2, or -1/2 pels in each direction from the center of the grid using the same method as in the modified full search. The best position determines the motion vector.

Some possible grid spacings for various search ranges are given in table D.8.

Table D.8 -- Grid spacings for logarithmic searches

forward_f_code	RANGE	STEPS	GRID SPACINGS
1	±7,5	4	4 2 1 1/2
2	±15,5	5	8 4 2 1 1/2
3	±31,5	6	16 8 4 2 1 1/2

For P-pictures only forward searches are performed, but B-pictures require both forward and backward searches. Not all the vectors calculated during the search are necessarily used. In B-pictures either forward or backward motion compensation might be used instead of interpolated motion compensation, and in both P and B-pictures the encoder might decide that a block is better coded as intra, in which case no vectors are transmitted.

D.6.2.2.5 Telescopic search

Even with the faster methods of the modified full search, or the logarithmic search, the search might be quite expensive. For example, if the encoder decides to use a maximum search range of 7 pels per picture

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

interval, and if there are 4 B-pictures preceding a P-picture, then the full search range for the P-picture would be 35 pels. This large search range may exceed the capabilities of the encoder.

One way of reducing the search range is to use a telescopic search technique. This is best explained by illustrating with an example. Consider the group of pictures in figure D.27.

I	B	B	B	P	B	B	B	P	B	B	B	P
0	1	2	3	4	5	6	7	8	9	10	11	12

Figure D.27 -- Example group of pictures in display order

The encoder might proceed using its selected block matching criterion and D search strategy. For each P-picture and the preceding B-pictures, it first calculates all the forward vectors, then calculates all the backward vectors. The first set of pictures consists of pictures 0 through 4.

To calculate the complete set of forward vectors, the encoder first calculates all the forward vectors from picture 0 to picture 1 using a 2-D search strategy centered on zero displacement. It next calculates all the forward vectors from picture 0 to picture 2 using a 2-D search strategy centered on the displacements calculated for the corresponding block of picture 1. It next calculates all the forward vectors from picture 0 to picture 3 using a 2-D search strategy centered on the displacements calculated for the corresponding block of picture 2. Finally, it calculates all the forward vectors from picture 0 to picture 4 using a 2-D search strategy centered on the displacements calculated for the corresponding block of picture 3.

To calculate the complete set of backward vectors, the encoder first calculates all the backward vectors from picture 4 to picture 3 using a 2-D search strategy centered on zero displacement. It next calculates all the backward vectors from picture 4 to picture 2 using a 2-D search strategy centered on the displacements calculated for the corresponding block of picture 3. Finally, it calculates all the backward vectors from picture 4 to picture 1 using a 2-D search strategy centered on the displacements calculated for the corresponding block of picture 2.

Further methods of motion estimation are given by Netravali and Haskell [1].

D.6.2.3 Coding of motion vectors

The motion vector of a macroblock tends to be well correlated with the vector of the previous macroblock. For example, in a pan all vectors would be roughly the same. Motion vectors are coded using a DPCM technique to make use of this correlation.

In P-pictures the motion vector used for DPCM, the prediction vector, is set to zero at the start of each slice and at each intra-coded macroblock. Note that macroblocks which are coded as predictive but which have no motion vector, also set the prediction vector to zero.

In B-pictures there are two motion vectors, forward and backward. Each vector is coded relative to the predicted vector of the same type. Both motion vectors are set to zero at the start of each slice and at each intra-coded macroblock. Note that predictive macroblocks which have only a forward vector do not affect the value of the predicted backward vector. Similarly, predictive macroblocks which have only a backward vector do not affect the value of the predicted forward vector.

The range of the vectors is set by two parameters. The full_pel_forward_vector and full_pel_backward_vector flags in the picture header determine whether the vectors are defined in half-pel or integer-pel units.

A second parameter, forward_f_code or backward_f_code, is related to the number of bits appended to the VLC codes in table D.9.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Table D.9 -- Differential motion code.

VLC code	Value
0000 0011 001	-16
0000 0011 011	-15
0000 0011 101	-14
0000 0011 111	-13
0000 0100 001	-12
0000 0100 011	-11
0000 0100 11	-10
0000 0101 01	-9
0000 0101 11	-8
0000 0111	-7
0000 1001	-6
0000 1011	-5
0000 11	-4
0001 1	-3
0011	-2
011	-1
1	0
010	1
0010	2
0001 0	3
0000 110	4
0000 1010	5
0000 1000	6
0000 0110	7
0000 0101 10	8
0000 0101 00	9
0000 0100 10	10
0000 0100 010	11
0000 0100 000	12
0000 0011 110	13
0000 0011 100	14
0000 0011 010	15
0000 0011 000	16

Advantage is taken of the fact that the range of displacement vector values is constrained. Each VLC represents a pair of difference values. Only one of the pair will yield a motion vector falling within the permitted range.

The range of the vector is limited to the values shown in table D.7. The values obtained by decoding the differential values must be kept within this range by adding or subtracting a modulus which depends on the f value as shown in table D.10.

Table D.10 -- Modulus for motion vectors

forward_f_code or backward_f_code	MODULUS
1	32
2	64
3	128
4	256
5	512
6	1 024
7	2 048

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The use of the modulus, which refers only to the numbers in tables D.8 through D.10, will be illustrated by an example. Assume that a slice has the following vectors, expressed in the units set by the full pel flag.

3 10 30 30 -14 -16 27 24

The range is such that an f value of 2 can be used. The initial prediction is zero, so the differential values are

3 7 20 0 -44 -2 43 -3

The differential values are reduced to the range -32 to +31 by adding or subtracting the modulus 64 corresponding to the forward_f_code of 2.

3 7 20 0 20 -2 -21 -3

To create the codeword, (mvd + (sign(mvd)*(forward_f-1))) is divided by forward_f. The signed quotient of this division is used to find a variable length codeword from table D.9. Then the absolute value of the remainder is used to generate a fixed length code that is concatenated with the variable length code. The codes generated by this example are shown below:

Value	VLC Code
3	0010 0
7	0000 1100
20	0000 0100 101
0	1
20	0000 0100 101
-2	0111
-21	0000 0100 0110
-3	0011 0

D.6.3 Coding I-pictures

In coding I-pictures, the encoder has two main decisions to make that are not mandated by this part of ISO/IEC 11172. These are: how to divide the picture up into slices, and how to set the quantizer scale.

D.6.3.1 Slices in I-pictures

Division of the picture into slices is described in D.5.4.

D.6.3.2 Macroblocks in I-pictures

D.6.3.2.1 Macroblock types in I-pictures

There are two types of macroblock in I-pictures. Both use intra coding. One uses the current quantizer scale, whereas the other defines a new value for the quantizer scale. They are identified in the coded bitstream by the VLC codes given in table D.11.

Table D.11 -- Macroblock type VLC for I-pictures (table B.2a.)

TYPE	QUANT	VLC
intra-d		1
intra-q	1	01

The types are referred to names in this annex. Intra-d is the default type where the quantizer scale is not changed. Intra-q sets the quantizer scale.

In order to allow for possible future extension to MPEG video, the VLC for intra-q is 01 rather than 0. Additional types could be added to this table without interfering with the existing entries. The VLC table is thus open for future additions, and not closed. A policy of making the coding tables open in this way was adopted by in developing this part of ISO/IEC 11172. The advantage of future extension was judged to be worth the slight coding inefficiency.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 142 of 341

D.6.3.2.2 Quantizer scale

If the macroblock type is intra-q, then the macroblock header contains a five-bit integer which defines the quantizer scale. This is used by the decoder to calculate the DCT coefficients from the transmitted quantized coefficients. A value of 0 is forbidden, so the quantizer scale can have any value between 1 and 31 inclusive.

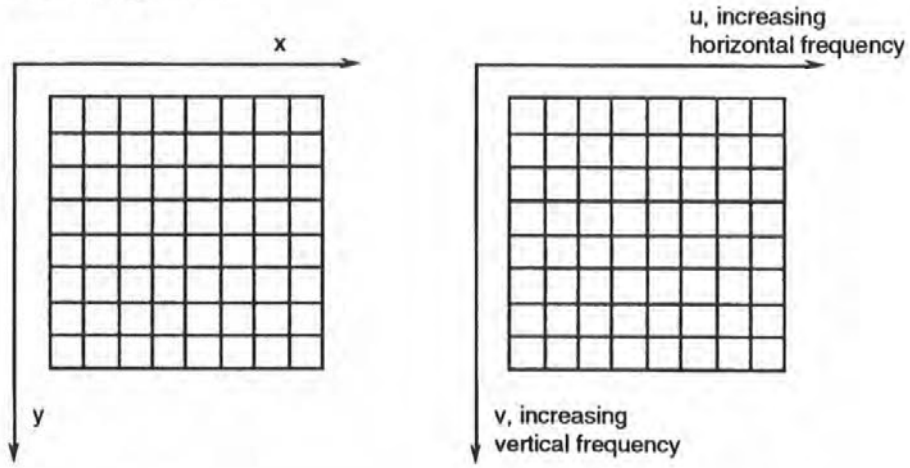
Note that also the quantizer scale is set in a slice header.

If the block type is intra-d, then no quantizer scale is transmitted and the decoder uses the previously set value. For a discussion on strategies encoders might use to set the quantizer scale, see D.6.1.

Note that the cost of transmitting a new quantizer scale is six bits: one for the extra length of the macroblock type code, and five to define the value. Although this is normally a small fraction of the bits allocated to coding each macroblock, the encoder should exercise some restraint and avoid making a large number of very small changes.

D.6.3.3 DCT transform

The DCT is illustrated in figure D.28.



(a) Pels (b) DCT Coefficients

Figure D.28 -- Transformation of pels to coefficients

The pels are shown in raster scan order, whereas the coefficients are arranged in frequency order. The top left coefficient is the dc term and is proportional to the average value of the component pel values. The other coefficients are called ac coefficients. The ac coefficients to the right of the dc coefficient represent increasing horizontal frequencies, whereas ac coefficients below the dc coefficient represent increasing vertical frequencies. The remaining ac coefficients contain both horizontal and vertical frequency components. Note that an image containing only vertical lines contains only horizontal frequencies.

The coefficient array contains all the information of the pel array and the pel array can be exactly reconstructed from the coefficient array, except for information lost by the use of finite arithmetic precision.

The two-dimensional DCT is defined as

$$F(u,v) = \frac{1}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos(\pi(2x+1)u/16) \cos(\pi(2y+1)v/16)$$

with: u, v, x, y = 0, 1, 2, ... 7
where x, y = spatial coordinates in the pel domain
u, v = coordinates in the transform domain
C(u) = $1/\sqrt{2}$ for u = 0
C(v) = $1/\sqrt{2}$ for v = 0
= 1 otherwise

This transform is separable, i.e. a one-dimensional DCT transform may be applied first in the horizontal direction and then in the vertical direction. The formula for the one dimensional transform is:

$$F(u) = \frac{1}{2} C(u) \sum_{x=0}^{7} f(x)\cos(\pi(2x+1)u/16)$$

$$C(u) = 1/\sqrt{2} \text{ for } u = 0$$
$$= 1 \text{ otherwise}$$

Fast DCT transforms exist, analogous to fast Fourier transforms. See reference [3].

The input pel values have a range from 0 to 255, giving a dynamic range for the dc coefficient from 0 to 2 040. The maximum dynamic range for any ac coefficient is about -1 000 to 1 000. Note that for P and B-pictures the component pels represent difference values and range from -255 to 255. This gives a maximum dynamic range for any coefficient of about -2 000 to 2 000. The encoder may thus represent the coefficients using 12 bits whose values range from -2 048 to 2 047.

D.6.3.4 Quantization

Each array of 8 by 8 coefficients produced by the DCT transform operation is quantized to produce an 8 by 8 array of quantized coefficients. Normally the number of non-zero quantized coefficients is quite small, and this is one of the main reasons why the compression scheme works as well as it does.

The coefficients are quantized with a uniform quantizer. The characteristic of this quantizer, only for I-blocks, is shown below:

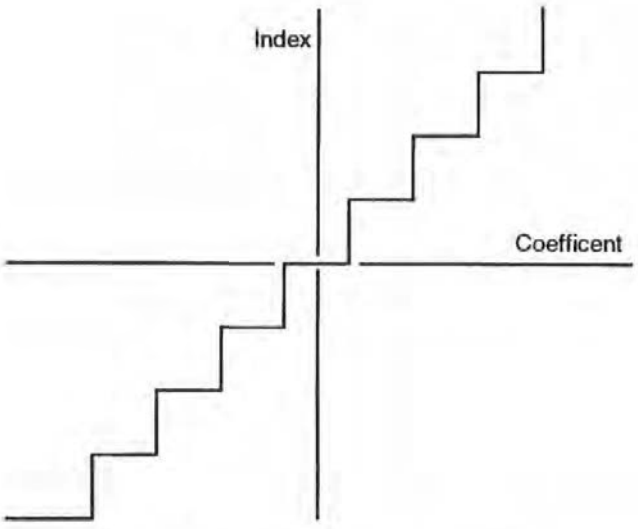


Figure D.29. -- Uniform quantizer characteristics

The value of the coefficient is divided by the quantizer step size and rounded to the nearest whole number to produce the quantized coefficient. Half integer values may be rounded up or down without directly affecting image quality. However, rounding towards zero tends to give the smallest code size and so is preferred. For example, with a step size of 16 all coefficients with values between 25 and 40 inclusive would give a quantized coefficient of 2.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

The quantizer step size is derived from the quantization matrix and the quantizer scale. It can thus be different for different coefficients, and may change between macroblocks. The only exception is the dc coefficient which is treated differently.

The eye is quite sensitive to large area luminance errors, and so the accuracy of coding the dc value is fixed. The quantizer step size for the dc coefficients of the luminance and chrominance components is fixed at eight. The dc quantized coefficient is obtained by dividing the dc coefficient by eight and rounding to the nearest whole number. This effectively quantizes the average dc value to one part in 256 for the reconstructed pels.

For example, a dc coefficient of 21 is quantized to a value of 3, independent of the value of the quantizer scale.

The ac coefficients are quantized using the intra quantization matrix. The quantized coefficient i[u,v] is produced by quantizing the coefficient c[u,v] for I-blocks. One equation is given by the formula:

$$i[u,v] = 8 * c[u,v] \;//\; (q * m[u,v])$$

where m[u,v] is the corresponding element of the intra quantization matrix, and q is the quantizer scale. The quantized coefficient is limited to the range -255 to +255.

The intra quantization matrix might be the default matrix, or it might have been downloaded in the sequence header.

D.6.3.5 Coding of quantized coefficients

The top left coefficient in figure D.28b is called the dc coefficient, the remainder are called ac coefficients. The dc coefficient is correlated with the dc coefficient of the preceding block, and advantage is taken of this in coding. The ac coefficients are not well correlated, and are coded independently.

After the dc coefficient of a block has been quantized it is coded losslessly by a DPCM technique. Coding of the luminance blocks within a macroblock follows the raster scan order of figure D.5, 0 to 3. Thus the dc value of block 3 becomes the dc predictor for block 0 of the following macroblock. The dc value of each chrominance block is coded using the dc value of the corresponding block of the previous macroblock as a predictor. At the beginning of each slice, all three dc predictors for Y, Cb and Cr, are set to 1 024 (128*8).

The differential dc values thus generated are categorized according to their absolute value as shown in table D.12.

Table D.12. -- Differential dc size and VLC

DIFFERENTIAL DC (absolute value)	SIZE	VLC CODE (luminance)	VLC CODE (chrominance)
0	0	100	00
1	1	00	01
2 to 3	2	01	10
4 to 7	3	101	110
8 to 15	4	110	1110
16 to 31	5	1110	1111 0
32 to 63	6	1111 0	1111 10
64 to 127	7	1111 10	1111 110
128 to 255	8	1111 110	1111 1110

The size is transmitted using a VLC. This VLC is different for luminance and chrominance since the statistics are different.

The size defines the number of additional bits required to define the level uniquely. Thus a size of 6 is followed by 6 additional bits. These bits define the level in order, from low to high. Thus the first of these extra bits gives the sign: 0 for negative and 1 for positive. A size of zero requires no additional bits.

The additional codes are given in table D.13.

Table D.13. -- Differential dc additional code

DIFFERENTIAL DC	SIZE	ADDITIONAL CODE
-255 to -128	8	00000000 to 01111111
-127 to -64	7	0000000 to 0111111
-63 to -32	6	000000 to 011111
-31 to -16	5	00000 to 01111
-15 to -8	4	0000 to 0111
-7 to -4	3	000 to 011
3 to -2	2	00 to 01
-1	1	0
0	0	
1	1	1
2 to 3	2	10 to 11
4 to 7	3	100 to 111
8 to 15	4	1000 to 1111
16 to 31	5	10000 to 11111
32 to 63	6	100000 to 111111
64 to 127	7	1000000 to 1111111
128 to 255	8	10000000 to 11111111

For example, a luminance dc change of 10 would be coded as 1101010. table D.12 shows that the first three bits 110 indicate that the size is 4. This means that four additional bits are required to define the exact value. The next bit is a 1, and table D.13 shows that the differential dc value must be somewhere between 8 and 15 inclusive. The last three bits, 010, show that the exact value is 10.

The decoder reconstructs dc quantized coefficients by following the inverse procedure.

The ac quantized coefficients are coded using a run length and level technique. The quantized coefficients are first scanned in the zigzag order shown in figure D.30.

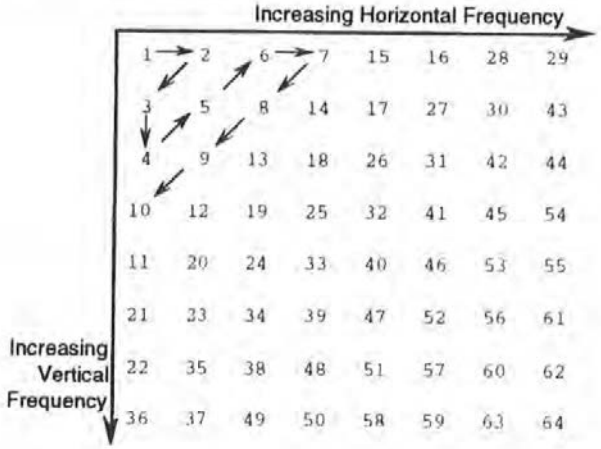


Figure D.30. -- Quantized coefficient block in zigzag scan order

The scanning order starts at 1, passes through 2, 3 etc in order, eventually reaching 64 in the bottom right corner. The length of a run is the number of zero quantized coefficients skipped over. For example, the quantized coefficients in figure D.31 produce the list of run lengths and levels in table D.14.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

1	0	0	0	0	0	0	0	0
2	-3	0	0	0	0	0	0	0
4	-5	0	0	0	0	0	0	0
1	0	0	130	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Figure D.31. -- Example quantized coefficients

Table D.14. -- Example run lengths and levels

RUN-LENGTH	LEVEL
1	2
0	4
0	-3
3	-5
0	1
14	130
end	

The scan starts at position 2 since the top left quantized coefficient is coded separately as the dc quantized coefficient.

Using a zig zag scan rather than a raster scan is more efficient as it gives fewer runs and can be coded with shorter VLC codes.

The list of run lengths and levels is coded using table D.15 Not all possible combinations of run length and level are in these tables, only the more common ones. For combinations not in the tables, an escape sequence is used. In table D.15, the last bit 's' denotes the sign of the level; 0 means a positive level and 1 means a negative level. The escape code is used followed by the run length derived from table D.16 and then the level from table D.17.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Table D.15. -- Combination codes for DCT quantized coefficients. s = 0 for positive level, s = 1 for negative level

RUN	LEVEL	VLC CODE
EOB		10
0	1	1s IF 1st COEFF
0	1	11s NOT 1st COEFF
0	2	0100 s
0	3	0010 1s
0	4	0000 110s
0	5	0010 0110 s
0	6	0010 0001 s
0	7	0000 0010 10s
0	8	0000 0001 1101 s
0	9	0000 0001 1000 s
0	10	0000 0001 0011 s
0	11	0000 0001 0000 s
0	12	0000 0000 1101 0s
0	13	0000 0000 1100 1s
0	14	0000 0000 1100 0s
0	15	0000 0000 1011 1s
0	16	0000 0000 0111 11s
0	17	0000 0000 0111 10s
0	18	0000 0000 0111 01s
0	19	0000 0000 0111 00s
0	20	0000 0000 0110 11s
0	21	0000 0000 0110 10s
0	22	0000 0000 0110 01s
0	23	0000 0000 0110 00s
0	24	0000 0000 0101 11s
0	25	0000 0000 0101 10s
0	26	0000 0000 0101 01s
0	27	0000 0000 0101 00s
0	28	0000 0000 0100 11s
0	29	0000 0000 0100 10s
0	30	0000 0000 0100 01s
0	31	0000 0000 0100 00s
0	32	0000 0000 0011 000s
0	33	0000 0000 0010 111s
0	34	0000 0000 0010 110s
0	35	0000 0000 0010 101s
0	36	0000 0000 0010 100s
0	37	0000 0000 0010 011s
0	38	0000 0000 0010 010s
0	39	0000 0000 0010 001s
0	40	0000 0000 0010 000s
1	1	011s
1	2	0001 10s
1	3	0010 0101 s
1	4	0000 0011 00s
1	5	0000 0001 1011 s
1	6	0000 0000 1011 0s
1	7	0000 0000 1010 1s
1	8	0000 0000 0011 111s
1	9	0000 0000 0011 110s
1	10	0000 0000 0011 101s
1	11	0000 0000 0011 100s
1	12	0000 0000 0011 011s
1	13	0000 0000 0011 010s
1	14	0000 0000 0011 001s
1	15	0000 0000 0001 0011s
1	16	0000 0000 0001 0010s
1	17	0000 0000 0001 0001s
1	18	0000 0000 0001 0000s

RUN	LEVEL	VLC CODE
2	1	0101 s
2	2	0000 100s
2	3	0000 0010 11s
2	4	0000 0001 0100 s
2	5	0000 0000 1010 0s
3	1	0011 1s
3	2	0010 0100 s
3	3	0000 0001 1100 s
3	4	0000 0000 1001 1s
4	1	0011 0s
4	2	0000 0011 11s
4	3	0000 0001 0010 s
5	1	0001 11s
5	2	0000 0010 01s
5	3	0000 0000 1001 0s
6	1	0001 01s
6	2	0000 0001 1110 s
6	3	0000 0000 0001 0100s
7	1	0001 00s
7	2	0000 0001 0101 s
8	1	0000 111s
8	2	0000 0001 0001 s
9	1	0000 101s
9	2	0000 0000 1000 1s
10	1	0010 0111 s
10	2	0000 0000 1000 0s
11	1	0010 0011 s
11	2	0000 0000 0001 1010s
12	1	0010 0010 s
12	2	0000 0000 0001 1001s
13	1	0010 0000 s
13	2	0000 0000 0001 1000s
14	1	0000 0011 10s
14	2	0000 0000 0001 0111s
15	1	0000 0011 01s
15	2	0000 0000 0001 0110s
16	1	0000 0010 00s
16	2	000 0000 0001 0101s
17	1	0000 0001 1111 s
18	1	0000 0001 1010 s
19	1	0000 0001 1001 s
20	1	0000 0001 0111 s
21	1	0000 0001 0110 s
22	1	0000 0000 1111 1s
23	1	0000 0000 1111 0s
24	1	0000 0000 1110 1s
25	1	0000 0000 1110 0s
26	1	0000 0000 1101 1s
27	1	0000 0000 0001 1111s
28	1	0000 0000 0001 1110s
29	1	0000 0000 0001 1101s
30	1	0000 0000 0001 1100s
31	1	0000 0000 0001 1011s
ESCAPE		0000 01

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Table D.16. -- Zero run length codes

RUN-LENGTH	CODE
0	0000 00
1	0000 01
2	0000 10
.	.
62	1111 10
63	1111 11

Table D.17. -- Level codes for DCT quantized coefficients

LEVEL	CODE
-256	FORBIDDEN
-255	1000 0000 0000 0001
-254	1000 0000 0000 0010
.	.
-129	1000 0000 0111 1111
-128	1000 0000 1000 0000
-127	1000 0001
-126	1000 0010
.	.
-2	1111 1110
-1	1111 1111
0	FORBIDDEN
1	0000 0001
2	0000 0010
.	.
126	0111 1110
127	0111 1111
128	0000 0000 1000 0000
129	0000 0000 1000 0001
.	.
254	0000 0000 1111 1110
255	0000 0000 1111 1111

Using tables D.15 through D.17 we can derive the VLC codes for the example of table D.14:

Table D.18. -- Example run lengths, values, and VLC codes

RUN	VALUE	VLC CODE	COMMENT
1	2	0001 100	
0	4	0000 1100	
0	-3	0010 11	
3	-5	0000 0100 0011 1111 1011	esc seq
0	1	110	
14	130	0000 0100 1110 0000 0000 1000 0010	esc seq
EOB		10	

The first three codes in table D.18 are derived directly from table D.15. The next code is derived indirectly since table D.15 does not contain an entry corresponding to a run length of 3 and a level of 5. Instead the escape code 000001 is used. This is followed by the six-bit code 000011 from table D.16 indicating a run length of 3. Lastly the eight-bit code from table D.17 (11111011 - indicating a level of -5) is appended.

After the last coefficient has been coded, an EOB code is used to inform the decoder that there are no more quantized coefficients in the current 8 by 8 block. This EOB code is used even if the last quantized coefficient is at the bottom right of the block.

There are two codes for the 0,1 run length, level combination, as indicated in table D.15. Intra block coding always has the first quantized coefficient, the dc quantized coefficient, coded using the dc size method. Consequently intra blocks always use the code 11s to denote a run length, level combination of 0,1. It will

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

be seen later that predictively coded blocks code the dc quantized coefficient differently, and may use the shorter code.

D.6.4 Coding P-pictures

As in I-pictures, each P-picture is divided up into one or more slices, which are, in turn, divided into macroblocks. Coding is more complex than for I-pictures, since motion-compensated macroblocks may be constructed. The difference between the motion compensated macroblock and the current macroblock is transformed with a two-dimensional DCT giving an array of 8 by 8 transform coefficients. The coefficients are quantized to produce a set of quantized coefficients. The quantized coefficients are then encoded using a run-length value technique.

As in I-pictures, the encoder needs to store the decoded P-picture since this may be used as the starting point for motion compensation. Therefore, the encoder will reconstruct the image from the quantized coefficients.

In coding P-pictures, the encoder has more decisions to make than in the case of I-pictures. These decisions are: how to divide the picture up into slices, determine the best motion vectors to use, decide whether to code each macroblock as intra or predicted, and how to set the quantizer scale.

D.6.4.1 Slices in P-pictures

P-pictures are divided into slices in the same way as I-pictures. The same considerations as to the best method of dividing a picture into slices apply, see D.5.4.

D.6.4.2 Macroblocks in P-pictures

Slices are divided into macroblocks in the same way as for I-pictures. The major difference is the complexity introduced by motion compensation.

The macroblock header may contain stuffing. The position of the macroblock is determined by the macroblock address. Whereas the macroblock address increment within a slice for I-pictures is restricted to one, it may be larger for P-pictures. Any macroblocks thus skipped over are called "skipped macroblocks". The decoder copies them from the previous picture into the current picture. Skipped macroblocks are as predicted macroblocks with a zero motion vector for which no additional correction is available. They require very few bits to transmit.

The next field in the macroblock header defines the macroblock type.

D.6.4.2.1 Macroblock types in P-pictures

There are eight types of macroblock in P-pictures:

Table D.19 -- Macroblock type VLC for P-pictures (table B.2b)

TYPE	VLC	INTRA	MOTION FORWARD	CODED PATTERN	QUANT
pred-mc	1	0	1	1	0
pred-c	01	0	0	1	0
pred-m	001	0	1	0	0
intra-d	0001 1	1	0	0	0
pred-mcq	0001 0	0	1	1	1
pred-cq	0000 1	0	0	1	1
intra-q	0000 01	1	0	0	1
skipped	N/A				

Not all possible combinations of motion compensation, coding, quantization, and intra coding occur. For example, with intracoded macroblocks, intra-d and intra-q, motion vectors are not transmitted.

Skipped macroblocks have no VLC code. Instead they are coded by having the macroblock address increment code skip over them.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.6.4.2.2 Quantizer scale

If the macroblock type is pred-mcq, pred-cq or intra-q, i.e. if the QUANT column in table D.19 has a 1, then a quantizer scale is transmitted . If the macroblock types are pred-mc, pred-c or intra-d, then the DCT correction is coded using the previously established value for the quantizer scale.

D.6.4.2.3 Motion vectors

If the macroblock type is pred-m, pred-mc or pred-mq, i.e. if the MOTION FORWARD column in table D.19 has a 1, then horizontal and vertical forward motion vectors are transmitted in succession..

D.6.4.2.4 Coded block pattern

If the macroblock type is pred-c, pred-mc, pred-cq or pred-mcq, i.e. if the CODED PATTERN column in table D.19 has a 1, then a coded block pattern is transmitted. This informs the decoder which of the six blocks in the macroblock are coded, i.e. have transmitted DCT quantized coefficients, and which are not coded, i.e. have no additional correction after motion compensation.

The coded block pattern is a number from 0 to 63 that indicates which of the blocks are coded, i.e. have at least one transmitted coefficient, and which are not coded. To understand the structure of the coded block pattern, we refer to figure D.5 and introduce the variables PN to indicate the status of each of the six blocks. If block N is coded then PN has the value one, if it is not coded then PN is zero. The coded block pattern is defined by the equation:

$$CBP = 32*P0 + 16*P1 + 8*P2 + 4*P3 + 2*P4 + P5$$

This is equivalent to the definition given in 2.4.3.6.

For example, if the top two luminance blocks and the Cb block are coded, and the other three are not, then P0 = 1, P1 = 1, P2 = 0, P3 = 0, P4 = 1, and P5 = 0. The coded block pattern is:

$$CBP = 32*1 + 16*1 + 8*0 + 4*0 + 2*1 + 0 = 50$$

Certain patterns are more common than others. Advantage is taken of this fact to increase the coding efficiency and transmit a VLC representing the coded block pattern, rather than the coded block pattern itself. The VLC codes are given in table D.20.

Table D.20 -- VLC table for coded block pattern

CBP	VLC CODE
60	111
4	1101
8	1100
16	1011
32	1010
12	1001 1
48	1001 0
20	1000 1
40	1000 0
28	0111 1
44	0111 0
52	0110 1
56	0110 0
1	0101 1
61	0101 0
2	0100 1
62	0100 0
24	0011 11
36	0011 10
3	0011 01
63	0011 00

CBP	VLC CODE
5	0010 111
9	0010 110
17	0010 101
33	0010 100
6	0010 011
10	0010 010
18	0010 001
34	0010 000
7	0001 1111
11	0001 1110
19	0001 1101
35	0001 1100
13	0001 1011
49	0001 1010
21	0001 1001
41	0001 1000
14	0001 0111
50	0001 0110
22	0001 0101
42	0001 0100
15	0001 0011

CBP	VLC CODE
51	0001 0010
23	0001 0001
43	0001 0000
25	0000 1111
37	0000 1110
26	0000 1101
38	0000 1100
29	0000 1011
45	0000 1010
53	0000 1001
57	0000 1000
30	0000 0111
46	0000 0110
54	0000 0101
58	0000 0100
31	0000 0011 1
47	0000 0011 0
55	0000 0010 1
59	0000 0010 0
27	0000 0001 1
39	0000 0001 0

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Thus the coded block pattern of the previous example, 50, would be represented by the code "00010110".

Note that there is no code representing the state in which none of the blocks are coded, a coded block pattern equal to zero. Instead, this state is indicated by the macroblock type.

For macroblocks in I-pictures, and for intra coded macroblocks in P and B-pictures, the coded block pattern is not transmitted, but is assumed to have a value of 63, i.e. all the blocks in the macroblock are coded.

The use of coded block patterns instead of transmitting end of block codes for all blocks follows the practice in CCITT Recommendation H.261.

D.6.4.3 Selection of macroblock type

An encoder has the difficult task of choosing between the different types of macroblocks.

An exhaustive method is to try coding a macroblock to the same degree of accuracy using each type, then choose the type that requires the least number of coding bits.

A simpler method, and one that is computationally less expensive, is to make a series of decisions. One way to order these decisions is:

1: motion compensation or no motion compensation, i.e. is a motion vector transmitted or is it assumed to be zero.
2: intra or non intra coding, i.e. is the macroblock type intra or is it predicted using the motion vector found in step 1.
3: if the macroblock type is non-intra, is it coded or not coded, i.e. is the residual error large enough to be coded using the DCT transform.
4: decide if the quantizer scale is satisfactory or should be changed.

These decisions are summarized in figure D.32.

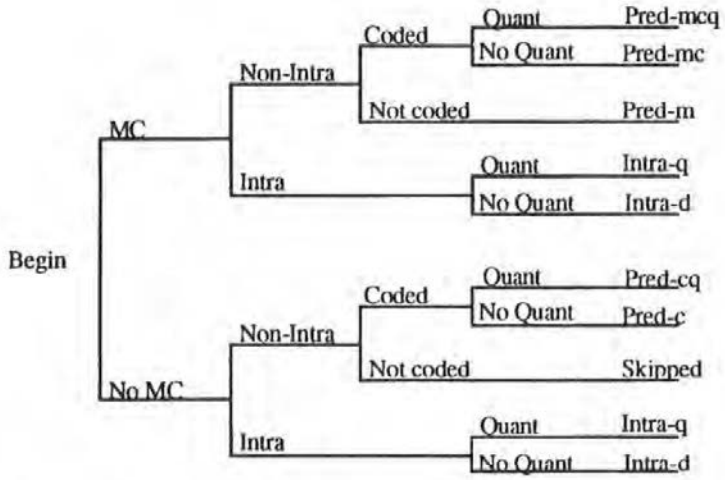


Figure D.32 -- Selection of macroblock types in P-pictures

The four decision steps are discussed in the next four clauses.

D.6.4.3.1 Motion compensation decision

The encoder has an option whether to transmit motion vectors or not for predictive-coded macroblocks. If the motion vector is zero then some code may be saved by not transmitting the motion vectors. Thus one algorithm is to search for the best match and compare the error of the predicted block with that formed with a zero vector. If the motion-compensated block is only slightly better than the uncompensated block, using the selected block matching criterion, then the zero vector might be used to save coding bits.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

An algorithm used in the development of both CCITT Recommendation H.261 and this part of ISO/IEC 11172 was as follows.

The block-matching criterion is the sum of absolute differences of all the luminance pels in a macroblock, when compared with the motion-compensated macroblock. If the sum is M for the motion-compensated block, and Z for the zero vector, then the decision of whether to make use of the motion vector is defined by figure D.33.

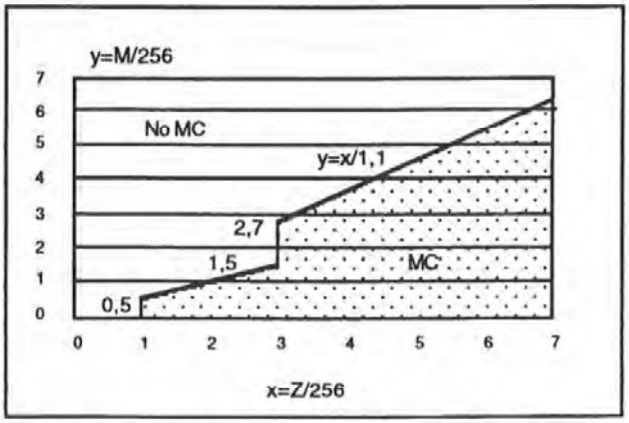


Figure D.33 -- Characteristic MC/No MC

Points on the line dividing the No MC (no motion compensation, i.e. zero vector), from the MC (motion compensation) regions, are regarded as belonging to the no motion compensation region.

It can be seen that if the error is sufficiently low, then no motion compensation should be used. Thus a way to speed up the decision is to examine the zero vector first and decide if it is good enough.

The foregoing algorithm was designed for telecommunications sequences in which the camera was fixed, and in which any movement of the background caused by the "drag along effect" of nearby moving objects was very objectionable. Great care was taken to reduce this spurious motion, and this accounts for the curious shape of the boundary between the two regions in figure D.33.

D.6.4.3.2 Intra/non-intra coding decision

After the encoder has determined the best motion vector, it is in a position to decide whether to use it, or disregard it entirely and code the macroblock as intra. The obvious way to do this is to code the block as intra, and compare the total number of bits required when coded as motion compensated plus correction with the same quantizer scale. The method using the fewest bits may be used.

This may be too computationally expensive for the encoder to do, and a faster algorithm may be required. One such algorithm, used in the simulation model during the development of this part of ISO/IEC 11172, was based on the variance of the luminance component of the macroblock. The variance of the current macroblock and of the difference macroblock (current - motion-compensated previous) is compared. It is calculated using the method represented by the following C program fragment. Note that in calculating the variance of the difference macroblock, the average value is assumed to be zero.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

```
int pelp[16][16];        /* Pel values in the Previous macroblock after motion compensation */
int pelc[16][16];        /* Pel values in the Current macroblock */
long dif;                /* Difference between two pel values */
long sum;                /* Sum of the current pel values */
long vard;               /*  Variance of the Difference macroblock */
long varc;               /*Variance of the Current macroblock */
int x,y;                 /* coordinates */

sum = 0;
vard = 0;
varc = 0;
for (y=0;y<16;y++) {
        for (x=0;x<16;x++) {
                sum = sum + pelc[y][x];
                varc = varc + (pelc[y][x]*pelc[y][x]);

                dif = pelc[y][x] - pelp[y][x];
                vard = vard + (dif*dif);
        }
}
vard = vard/256;    /* assumes mean is close to zero */
varc = ( varc/256 ) - ( (sum/256)*(sum/256) );
```

The decision as to whether to code as intra or non intra is then based on figure D.34.

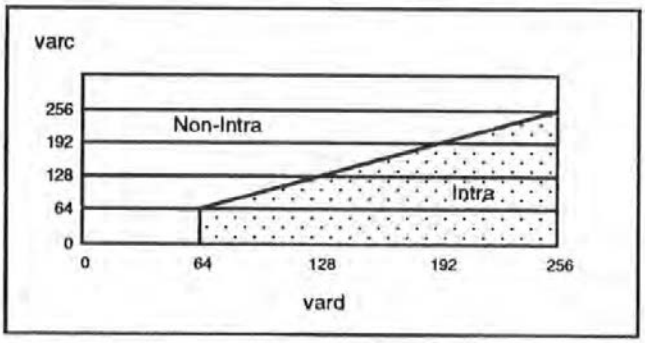


Figure D.34 -- Characteristic intra/non-intra

Points on the line dividing the non-intra from the intra regions, are regarded as belonging to the non-intra region.

D.6.4.3.3 Coded/not coded decision

The choice of coded or not coded is a result of quantization; when all coefficients are zero then a block is not coded. A macroblock is not coded if no block in it is coded, else it is coded.

D.6.4.3.4 Quantizer/no quantizer decision

Generally the quantizer scale is changed based on local scene content to improve the picture quality, and on the buffer fullness of the model decoder to prevent overflow and underflow.

D.6.4.4 DCT transform

Coefficients of intra blocks are transformed into quantized coefficients in the same way that they were for intra blocks in I-pictures. Prediction of the dc coefficient differs, however. The dc predicted values are all set to 1 024 (128*8) for intra blocks in P and B-pictures, unless the previous block was intra coded.

Coefficients of non-intra blocks are coded in a similar way. The main difference is that the coefficients to be transformed represent differences between pel values rather than the pel values themselves. The differences are obtained by subtracting the motion-compensated pel values from the previous picture from

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

the pel values in the current macroblock. Since the coding is of differences, there is no spatial prediction of the dc term.

D.6.4.5 Quantization of P-pictures

Intra macroblocks in P and B-pictures are quantized using the same method as described for I-pictures.

Non-intra macroblocks in P and B-pictures are quantized using the quantizer scale and the non-intra quantization matrix. Both dc and the ac coefficients are quantized the same way.

The following quantization formula was derived by inverting the reconstruction formula given in 2.4.4.2. Note that the divisor indicates truncation towards zero.

```
int coefforig;      /* original coefficient */
int coeffqant;      /* quantized coefficient */
int coeffrec;       * reconstructed coefficient */
int niqmatrix;      /* non-intra quantization matrix */
int quantscale;     /* quantizer scale */

coeffqant = (8 * coefforig) / (quantscale * niqmatrix);
```

The process is illustrated below:

niqmatrix	16	16	16	16	16
quantscale	10	10	10	10	10
coefforig	-39—20	-19~19	20~39	40~59	60~79
coeffqant	-1	0	1	2	3
coeffrec	-29	0	29	49	69

The last line shows the reconstructed coefficient values. The following diagram shows the characteristics of this quantizer. The flat spot around zero gives this type of quantizer its name: a dead-zone quantizer.

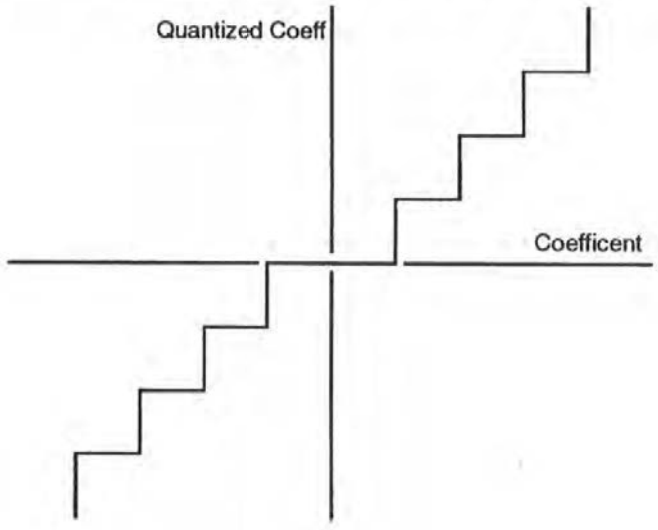


Figure D.35 -- Dead zone quantizer characteristic

D.6.4.6 Coding of quantized coefficients

D.6.4.6.1 Coding of intra blocks

Intra blocks in P-pictures are coded the same way as intra blocks in I-pictures. The only difference lies in the prediction of the dc coefficient. The dc predicted value is 128, unless the previous block was intra coded.

D.6.4.6.2 Coding of non-intra blocks

The coded block pattern is transmitted indicating which blocks have coefficient data. These are coded in a similar way to the coding of intra blocks except that the dc coefficient is coded in the same way as the ac coefficients.

D.6.5 Coding B-pictures

As in I and P-pictures, each B-picture is divided up into one or more slices, which are, in turn, divided into macroblocks. Coding is more complex than for P-pictures, since several types of motion compensated macroblock may be constructed: forward, backward, and interpolated. The difference between the motion-compensated macroblock and the current macroblock is transformed with a two-dimensional DCT giving an array of 8 by 8 transform coefficients. The coefficients are quantized to produce a set of quantized coefficients. The quantized coefficients are then encoded using a run-length value technique.

The encoder does not need to store the decoded B-pictures since they will not be used for motion compensation.

In coding B-pictures, the encoder has more decisions to make than in the case of P-pictures. These decisions are: how to divide the picture up into slices, determine the best motion vectors to use, decide whether to use forward or backward or interpolated motion compensation or to code as intra, and how to set the quantizer scale.

D.6.5.1 Slices in B-pictures

B-pictures are divided into slices in the same way as I and P-pictures. Since B-pictures are not used as a reference for motion compensation, errors in B-pictures are slightly less important than in I or P-pictures. Consequently, it might be appropriate to use fewer slices for B-pictures.

D.6.5.2 Macroblocks in B-pictures

Slices are divided into macroblocks in the same way as for I-pictures.

The macroblock header may contain stuffing. The position of the macroblock is determined by the macroblock address. Whereas the macroblock address increment within a slice for I-pictures is restricted to one, it may be larger for B-pictures. Any macroblocks thus skipped over are called "skipped macroblocks". Skipped macroblocks in B-pictures differ from skipped macroblocks in P-pictures. Whereas in P-pictures skipped macroblocks have a motion vector equal to zero, in B-pictures skipped macroblocks have the same motion vector and the same macroblock type as the previous macroblock, which cannot be intra coded. As there is no additional DCT correction, they require very few bits to transmit.

The next field in the macroblock header defines the macroblock type.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.6.5.2.1 Macroblock types in B-pictures

There are 12 types of macroblock in B-pictures:

Table D.21 -- Macroblock type VLC for B-pictures (table B.2d)

TYPE	VLC	INTRA	MOTION FORWARD	MOTION BACKWARD	CODED PATTERN	QUANT
pred-i	10	0	1	1	0	0
pred-ic	11	0	1	1	1	0
pred-b	010	0	0	1	0	0
pred-bc	011	0	0	1	1	0
pred-f	0010	0	1	0	0	0
pred-fc	0011	0	1	0	1	0
intra-d	0001 1	1	0	0	0	0
pred-icq	0001 0	0	1	1	1	1
pred-fcq	0000 11	0	1	0	1	1
pred-bcq	0000 10	0	0	1	1	1
intra-q	0000 01	1	0	0	0	1
skipped	N/A					

Compared with P-pictures, there are extra types due to the introduction of the backward motion vector. If only a forward motion vector is present, then the motion compensated-macroblock is constructed from a previous picture, as in P-pictures. If only a backward motion vector is present, then the motion-compensated macroblock is constructed from a future picture. If both forward and backward motion vectors are present, then motion-compensated macroblocks are constructed from both previous and future pictures, and the result is averaged to form the "interpolated" motion-compensated macroblock.

D.6.5.2.2 Quantizer scale

If the macroblock type is pred-icq, pred-fcq, pred-bcq, or intra-q, i.e. if the QUANT column in table D.21 has a 1, then a quantizer scale is transmitted. For the remaining macroblock types, the DCT correction is coded using the previously established value for the quantizer scale.

D.6.5.2.3 Motion vectors

If the MOTION FORWARD column in table D.21 has a 1, then horizontal and vertical forward motion vectors are transmitted in succession. If the MOTION BACKWARD column in table D.21 has a 1, then horizontal and vertical backward motion vectors are transmitted in succession. If both types are present then four component vectors are transmitted in the following order:

horizontal forward
vertical forward
horizontal backward
vertical backward

D.6.5.2.4 Coded block pattern

If the CODED PATTERN column in table D.21 has a 1, then a coded block pattern is transmitted. This informs the decoder which of the six blocks in the macroblock are coded, i.e. have transmitted DCT quantized coefficients, and which are not coded, i.e. have no additional correction after motion compensation.

D.6.5.3 Selection of macroblock type

The encoder has more types of macroblock to choose from in B-pictures, than in P-pictures, and consequently its job is a little harder.

For the simulation model used during development of this part of ISO/IEC 11172, the following sequential decision algorithm was used:

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

1: motion compensation mode, i.e. is forward or backward or interpolative motion compensation best? What of the vector values?
2: intra or non intra coding, i.e. is the macroblock type intra or is it motion compensated using mode and the vectors found in step 1?
3: if the macroblock type is non-intra, is it coded or not coded, i.e. is the residual error large enough to be coded using the DCT transform.
4: decide if the quantizer scale is satisfactory or should be changed.

These decisions are summarized in the following diagram:

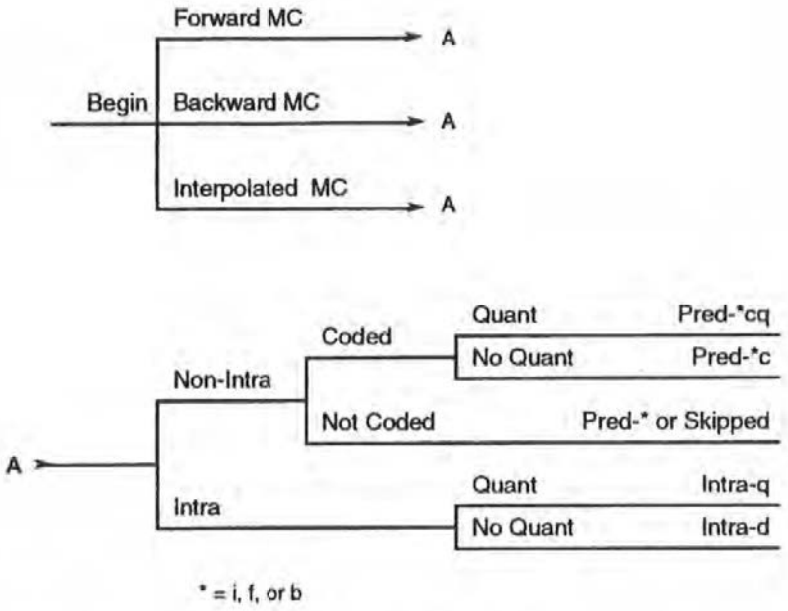


Figure D.36 -- Selection of macroblock type in B-pictures

The four decision steps are discussed in the next four clauses.

D.6.5.3.1 Selecting motion-compensation mode

An encoder should attempt to code B-pictures using skipped macroblocks if possible. This suggests that the encoder should first examine the case where the motion compensation is the same as for the previous macroblock. If the previous macroblock was non-intra, and if the motion-compensated block is good enough, there will be no additional DCT correction required and the block can be coded as skipped.

If the macroblock cannot be coded as skipped, then the following procedure may be followed.

For the simulation model, the selection of a motion compensation mode for a macroblock was based on the minimization of a cost function. The cost function was the MSE of the luminance difference between the motion-compensated macroblock and the current macroblock. The encoder calculated the best motion-compensated macroblock for forward motion compensation. It then calculated the best motion-compensated macroblock for backward motion compensation by a similar method. Finally it averaged the two motion-compensated macroblocks to produce the interpolated macroblock. It then selected the one that had the smallest mean square difference between it and the current macroblock. In the event of a tie, interpolative mode was chosen.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

D.6.5.3.2 Intra/non–intra coding decision

Based on the smallest MSE, a decision is made between the best of the three possible prediction modes and the Intra mode. The calculation is similar to that of P-pictures. The variances of the difference macroblock, vard, and of the current macroblock, varc, are calculated.

In the simulation model the final decision was based on simply the macroblock type with the smallest variance. If the two variances were equal, non-intra coding was chosen.

D.6.5.3.3 Coded/not–coded decision

The choice of coded or not coded is a result of quantization, when all coefficients are zero then a block is not coded. A macroblock is not coded if no block in it is coded, else it is coded.

D.6.5.4 DCT transform

Coefficients of blocks are transformed into quantized coefficients in the same way that they are for blocks in P-pictures.

D.6.5.5 Quantization of B-pictures

Blocks in B-pictures are quantized in the same way as for P-pictures.

D.6.5.6 Coding quantized coefficients

Blocks in B-pictures are coded the same way as blocks in P-pictures.

D.6.6 Coding D-pictures

D pictures contain only low frequency information. They are intended to be used for fast visible search modes. It is intended that the low frequency information they contain is sufficient for the user to locate the desired video.

D pictures are coded as the dc coefficients of blocks. There is a bit transmitted for the macroblock type, although only one macroblock type exists. In addition there is a bit denoting end of macroblock.

D.6.7 Coding at lower picture rates

This part of ISO/IEC 11172 does not allow pictures to be dropped at the encoder. This differs from the case of CCITT Recommendation H.261 [5] where temporal sub-sampling may be done by omitting coded pictures from the sequence. This part of ISO/IEC 11172 requires that all source pictures must be encoded and that coded pictures must be inserted into the bitstream nominally at the rate defined by the **picture_rate** field in the sequence header.

Despite this requirement it is possible for encoders to operate at a lower effective picture rate than the one defined in the sequence header by using P-pictures or B-pictures that consist entirely of macroblocks that are copied from a neighbouring reference picture with no DCT information. This creates a flexible method of temporal sub-sampling and picture repetition that may be implemented in the encoder by inserting a defined block of data. For example, to encode at an effective rate of 12,5 Hz in a 25 Hz bitstream, alternate pictures can be copied from the preceding picture by inserting the block of data in table D.22.

Table D.22 -- Example of the coded data elements needed to generate repeated pictures

Value (bits)	Mnemonic	Length (bits)
0000 0000 0000 0000 0000 0001 0000 0000	picture_start_code	32 bits
xxxx xxxx xx	temporal_reference	10 bits
010	picture_coding_type	3 bits
xxxx xxxx xxxx xxxx	vbv_delay	16 bits
0	full_pel_forward_code	1 bit
001	forward_f_code	3 bits
0000 000	stuffing	7 bits
0000 0000 0000 0000 0000 0001 0000 0001	slice_start_code	32 bits
0000 1	quantizer_scale	5 bits
1	macroblock_address_increment	1 bit
001	macroblock_type	3 bits
0	motion_horizontal_forward_code	1 bit
0	motion_vertical_forward_code	1 bit
0000 0001 000 (x 11)	macroblock_escape (x11)	121 bits
0000 0011 001	macroblock_address_increment	11 bits
001	macroblock_type	3 bits
0	motion_horizontal_forward_code	1 bit
0	motion_vertical_forward_code	1 bit
0000	stuffing	4 bits
	Total	256 bits

D.7 Decoding MPEG video

D.7.1 Decoding a sequence

D.7.1.1 Decoding for forward playback

At the beginning of a sequence, the decoder will decode the sequence header including the sequence parameters. If a parameter exceeds the capability of the decoder, then the decoder should report this. If the decoder determines that it can decode the bitstream, then it will set up its parameters to match those defined in the sequence header. This will include the horizontal and vertical resolutions and aspect ratio, the bit rate, and the quantization matrices.

Next the decoder will decode the group of pictures header, including the closed_gop and broken_link information, and take any appropriate action. It will decode the first picture header in the group of pictures and read the vbv_delay field. If the decoder uses the vbv_delay information to start-up decoding rather than the information in the system stream (ISO/IEC 11172-1) then it must delay displaying pictures until after a time determined by the vbv_delay information and a knowledge of the decoder's architecture.

If the closed-gop flag is 0, indicating that the group is open, and the broken_link flag is 1, then any B-pictures preceding (in display order) the first I-picture in the group cannot be decoded. The decoder may adopt one of several strategies. It may display the first I-picture during the time that the undecodable B-pictures would be displayed. This strategy maintains audio synchronization and buffer fullness. However it is likely that the broken link has occured because of post coding editing, in which case audio may be discontinuous. An alternative strategy might be to discard the B-pictures entirely, and delay decoding the I-picture until the buffer fullness is within limits.

If playback begins from a random point in the bitstream, the decoder should discard all the bits until it finds a sequence start code, a group of pictures start code, or a picture start code which introduces an I-picture. The slices and macroblocks in the picture are decoded and written into a display buffer, and perhaps into another buffer. The decoded pictures may be post processed and displayed in the order defined by the temporal reference at the picture rate defined in the sequence header.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 160 of 341

Subsequent pictures are processed at the appropriate times to avoid buffer overflow and underflow.

D.7.1.2 Decoding for fast playback

Fast forward can be supported by D pictures. It can also be supported by an appropriate spacing of I-pictures in a sequence. For example, if I-pictures were spaced regularly every 10 pictures, then a decoder might be able to playback the sequence at 10 times the normal speed by decoding and displaying only the I-pictures. This simple concept places considerable burdens on the media and the decoder. The media must be capable of speeding up and delivering 10 times the data rate, the decoder must be capable of accepting this higher data rate and decoding the I-pictures. Since I-pictures typically require significantly more bits to code than P or B-pictures, the decoder will have to decode significantly more than 10% of the data, even if it can search for picture start codes and discard the data for P and B-pictures.

For example, a sequence might be coded as follows:

I B P B P B P B P B I B P B P B P B P B I ...

Assume that the average code size per picture is C, that each B-picture requires 0,3C, that each P-picture requires 1,5C, and that each I-picture requires 2,5C, then the I-pictures require 25% of the code for their 10% of the display time.

Another way to achieve fast forward in a constant bit rate application, is for the media itself to sort out the I-pictures and transmit them. This would allow the data rate to remain constant. Since this selection process can be made to produce a valid ISO/IEC 11172-2 bitstream, the decoder should be able to decode it. If every I-picture of the preceding example were selected, then one I-picture would be transmitted every 2.5 picture periods, and the speed up rate would be 10/2,5 = 4 times. The decoder might be able to display the I-pictures at exactly 2,5 periods, or it might alternate displays at 2 and 3 periods.

If alternate I-pictures of the preceding example were selected, then one I-picture would again be transmitted every 2,5 picture periods, but the speed up rate would be 20/2,5 = 8 times.

If one in N I-pictures of the preceding example were selected, then the speed up rate would be 10N/2,5 = 4N times.

D.7.1.3 Decoding for pause and step modes

Decoding for pause requires the decoder to be able to control the incoming bitstream, and display a decoded picture without decoding any additional pictures. If the decoder has full control over the bitstream, then it can be stopped for pause and resumed when playback resumes. If the decoder has less control, as in the case of a CD ROM, then there may be a delay before playback can be resumed.

D.7.1.4 Decoding for reverse playback

To decode a bitstream and playback in reverse, the decoder must decode each group of pictures in the forward direction, store the decoded pictures, then display them in reverse order. This places severe storage requirements on the decoder in addition to any problems in gaining access to the coded bitstream in the correct order.

To reduce decoder memory requirements, groups of pictures should be small. There is no mechanism in the syntax for the encoder to state what the decoder requirements are in order to playback in reverse.

The amount of display buffer storage may be reduced by reordering the pictures, either by having the storage unit read and transmit them in another order, or by reordering the coded pictures in a decoder buffer. To illustrate the savings, consider the following typical group of pictures:

B	**B**	**I**	**B**	**B**	**P**	**B**	**B**	**P**	**B**	**B**	**P**	pictures in display order
0	1	2	3	4	5	6	7	8	9	10	11	temporal reference
I	**B**	**B**	**P**	**B**	**B**	**P**	**B**	**B**	**P**	**B**	**B**	pictures in coded order
2	0	1	5	3	4	8	6	7	11	9	10	temporal reference
I	**P**	**P**	**P**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	pictures in new order
2	5	8	11	10	9	7	6	4	3	1	0	temporal reference

Figure D.37 -- Example group of pictures

The decoder would decode the pictures in the new order, and display them in the reverse of the normal display order. Since the B-pictures are not decoded until they are ready to be displayed, the display buffer storage is minimized. The first two B-pictures, 0 and 1, would remain stored in the input buffer until the last P-picture in the previous group of pictures is decoded.

D.8 Post processing

D.8.1 Editing

Editing of a video sequence is best performed before compression, but situations arise where only the coded bitstream is available. One possible method would be to decode the bitstream, perform the required editing, and recode the bitstream. This usually leads to a loss in video quality, and it is better, if possible, to edit the coded bitstream itself.

Although editing may take several forms, the following discussion pertains only to editing at the picture level: deletion of coded video material from a bitstream, insertion of coded video material into a bitstream, or rearrangement of coded video material within a bitstream.

If editing is anticipated, e.g. clip video is provided analogous to clip art for still pictures, then the video can be encoded with well defined cutting points. These cutting points are places at which the bitstream may be broken apart or joined. Each cutting point should be followed by a closed group of pictures. This allows smooth playback after editing.

An editor must take care to ensure that the bitstream it produces is a legal bitstream. In particular it must ensure that the new bitstream complies with the requirements of the video buffering verifier. This is a difficult task and in general it will not be possible to edit together arbitrary sections of bitstreams that comply with this part of ISO/IEC 11172 to produce another bitstream that also complies with this part of ISO/IEC 11172 (see for example figure D.38).

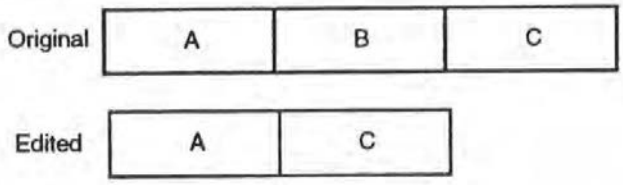


Figure D.38 -- Sequences

It may however be possible to deliberately encode bitstreams in a manner that allows some editing to occur. For instance, if all Groups of Pictures had the same number of pictures and were encoded with the same number of bits, then many of the problems of complying with the video buffering verifier would be solved.

The easiest editing task is to cut at the beginning of groups of pictures. If the group of pictures following the cut is open, which can be detected by examining the closed_gop flag in the group of pictures header, then the editor must set the broken_link bit to 1 to indicate to the decoder that the previous group of pictures cannot be used for decoding any B-pictures.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 162 of 341

D.8.2 Resampling

The decoded bitstream may not match the picture rate or the spatial resolution of the display device. In this quite frequent situation, the decoded video must be resampled or scaled.

One example, considered under preprocessing, is the case where the decoded video has SIF resolution and must be converted to CCIR 601 resolution.

D.8.2.1 Conversion of MPEG SIF to CCIR 601 format

A SIF is converted to its corresponding CCIR 601 format by spatial upsampling. A linear phase FIR filter is applied after the insertion of zeroes between samples. A filter that can be used for upsampling the luminance is shown in figure D.39:

-12	0	140	256	140	0	-12	// 256

Figure D.39 -- Upsampling filter for luminance

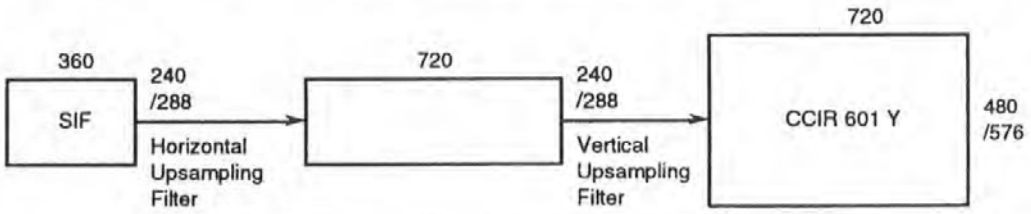
At the end of the lines some special technique, such as replicating the last pel, must be adopted.

According to CCIR Rec. 601 the chrominance samples need to be co-sited with the luminance samples 1, 3, 5,... In order to achieve the proper location, the upsampling filter should have an even number of taps,as shown in figure D.40.

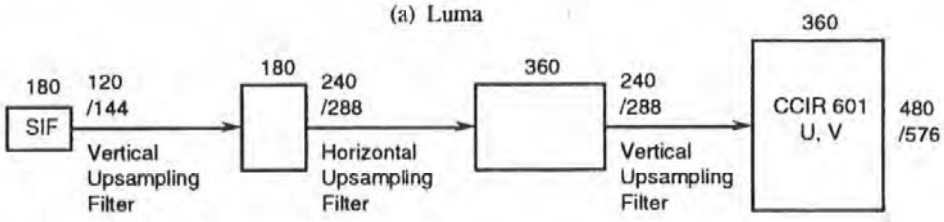
1	3	3	1	// 4

Figure D.40 -- Upsampling filter for chrominance

The SIF may be reconstructed by adding four black pels to each end of the horizontal luminance lines in the decoded bitmap, and two gray pels to each end of the horizontal chrominance lines. The luminance SIF may then be upsampled horizontally and vertically. The chrominance SIF should be upsampled once horizontally and twice vertically. This process is illustrated by the following diagram:



(a) Luma



(b) Chroma

Figure D.41 -- Simplified decoder block diagram

D.8.2.2 Temporal resampling

Since the picture rates are limited to those commonly used in the television industry, the same techniques may be applied. For example, conversion from 24 pictures/s to 60 fields/s may be achieved by the technique of 3:2 pulldown.

Video coded at 25 pictures/s can be converted to 50 fields/s by displaying the original decoded lines in the odd CCIR 601 fields, and the interpolated lines in the even fields. Video coded at 29,97 or 30 pictures/s may be converted to a field rate twice as large using the same method.

Video coded at 23,976 or 24 pictures/s may be converted to 50 fields/s by speeding it up by about 4% and decoding it as if it had been encoded at 25 pictures/s. The decoded pictures could be displayed in the odd fields, and interpolated pictures in the even fields. The audio must be maintained in synchronization, either by increasing the pitch, or by speeding it up without a pitch change.

Video coded at 23,976 or 24 pictures/s may be converted to 59,94 or 60 fields/s using the technique of 3:2 pull down.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Annex E

(informative)

Bibliography

[1] Arun N. Netravali & Barry G. Haskell *Digital Pictures, representation and compression* Plenum Press, 1988.

[2] Didier Le Gall *MPEG: A Video Compression Standard for Multimedia Applications* Trans ACM, April 1991.

[3] C Loeffler, A Ligtenberg, G S Moschytz *Practical fast 1-D DCT algorithms with 11 multiplications* Proceedings IEEE ICASSP-89, Vol. 2, pp 988-991, Feb. 1989.

[4] IEC Standard Publication 461, Second edition 1986 *Time and control code for video tape recorders.*

[5] CCITT Recommendation H.261 *Codec for audiovisual services at px64 kbit/s* Geneva, 1990.

[6] ISO/IEC DIS 10918-1 *Digital compression and coding of continuous-tone still images - Part 1: Requirements and guidelines.*

[7] E Viscito and C Gonzales *A Video Compression Algorithm with Adaptive Bit Allocation and Quantization*, Proc SPIE Visual Communications and Image Proc '91 Boston MA November 10-15 Vol 1605 205, 1991.

[8] A Puri and R Aravind *Motion Compensated Video Coding with Adaptive Perceptual Quantization*, IEEE Trans on Circuits and Systems for Video Technology, Vol 1 pp 351 Dec 1991.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440

Annex F

(informative)

List of patent holders

The user's attention is called to the possibility that - for some of the processes specified in this part of ISO/IEC 11172 - compliance with this International Standard may require use of an invention covered by patent rights.

By publication of this part of ISO/IEC 11172, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. However, each company listed in this annex has filed with the Information Technology Task Force (ITTF) a statement of willingness to grant a license under such rights that they hold on reasonable and nondiscriminatory terms and conditions to applicants desiring to obtain such a license.

Information regarding such patents can be obtained from :

AT&T
32 Avenue of the Americas
New York
NY 10013-2412
USA

Aware
1 Memorial Drive
Cambridge
02142 Massachusetts
USA

Bellcore
290 W Mount Pleasant Avenue
Livingston
NJ 07039
USA

The British Broadcasting Corporation
Broadcasting House
London
W1A 1AA
United Kingdom

British Telecommunications plc
Intellectual Property Unit
13th Floor
151 Gower Street
London
WC1E 6BA
United Kingdom

CCETT
4 Rue du Clos-Courtel
BP 59
F-35512
Cesson-Sevigne Cedex
France

CNET
38-40 Rue du General Leclerc
F-92131 Issy-les-Moulineaux
France

Compression Labs, Incorporated
2860 Junction Avenue
San Jose
CA 95134
USA

CSELT
Via G Reiss Romoli 274
I-10148 Torino
Italy

CompuSonics Corporation
PO Box 61017
Palo Alto
CA 94306
USA

Daimler Benz AG
PO Box 800 230
Epplestrasse 225
D-7000 Stuttgart 80
Germany

Dornier Gmbh
An der Bundesstrasse 31
D-7990 Friedrichshafen1
Germany

Fraunhofer Gesselschaft zur Foerderung der Angerwandten Forschung e.V.
Leonrodstrasse 54
8000 Muenchen 19
Germany

Hitachi Ltd
6 Kanda-Surugadai 4 chome
Chiyoda-ku
Tokyo 101
Japan

Institut für Rundfunktechnik Gmbh
Florianmühlstraße 60
8000 München 45
Germany

International Business Machines Corporation
Armonk
New York 10504
USA

KDD Corporation
2-3-2 Nishishinjuku
Shinjuku-ku
Tokyo
Japan

Licentia Patent-Verwaltungs-Gmbh
Theodor-Stern-Kai &
D-6000 Frankfurt 70
Germany

Massachusetts Institute of Technology
20 Ames Street
Cambridge
Massachusetts 02139
USA

Matsushita Electric Industrial Co. Ltd
1006 Oaza-Kadoma
Kadoma
Osaka 571
Japan

Mitsubishi Electric Corporation
2-3 Marunouchi
2-Chome
Chiyoda-Ku
Tokyo
100 Japan

NEC Corporation
7-1 Shiba 5-Chome
Minato-ku
Tokyo
Japan

Nippon Hoso Kyokai
2-2-1 Jin-nan
Shibuya-ku
Tokyo 150-01
Japan

Philips Electronics NV
Groenewoudseweg 1
5621 BA Eindhoven
The Netherlands

Pioneer Electronic Corporation
4-1 Meguro 1-Chome
Meguro-ku
Tokyo 153
Japan

Ricoh Co, Ltd
1-3-6 Nakamagome
Ohta-ku
Tokyo 143
Japan

Schawartz Engineering & Design
15 Buckland Court
San Carlos, CA 94070
USA

© ISO/IEC

Sony Corporation
6-7-35 Kitashinagawa
Shinagawa-ku
Tokyo 141
Japan

Symbionics
St John's Innovation Centre
Cowley Road
Cambridge
CB4 4WS
United Kingdom

Telefunken Fernseh und Rundfunk GmbH
Gottinger Chaussee
D-3000 Hannover 91
Germany

Thomson Consumer Electronics
9, Place des Vosges
La Défense 5
92400 Courbevoie
France

Toppan Printing Co, Ltd
1-5-1 Taito
Taito-ku
Tokyo 110
Japan

Toshiba Corporation
1-1 Shibaru 1-Chome
Minato-ku
Tokyo 105
Japan

Victor Company of Japan Ltd
12 Moriya-cho 3 chome
Kanagawa-ku
Yokohama
Kanagawa 221
Japan

APPENDIX ITU01



H.263 : Video coding for low bit rate communication

Recommendation H.263 (02/98)

Approved in 1998-02-06

Status : Superseded

Table of Contents

Summary

Access : Freely available items

Available languages and formats :

Click on the selected format and language to get the document

	Format	Size	Posted	Article Number
English	PDF (acrobat)	1261298 bytes	1998-10-22	E 13923
Español	PDF (acrobat)	1327345 bytes	1998-10-22	S 13925
Français	PDF (acrobat)	1342950 bytes	1998-10-22	F 13924

Access : TIES users

Available languages and formats :

Click on the selected format and language to get the document

	Format	Size	Posted	Article Number	TIES users
English	Word	2393088 bytes	1998-10-22	E 13923	
Español	Word	2350592 bytes	1998-10-22	S 13925	
Français	Word	2367488 bytes	1998-10-22	F 13924	

 Copyright © ITU 2008 All Rights Reserved

INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.263

(02/98)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Coding of moving video

---

Video coding for low bit rate communication

ITU-T Recommendation H.263

(Previously CCITT Recommendation)

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 172 of 341

ITU-T H-SERIES RECOMMENDATIONS

AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Characteristics of transmission channels used for other than telephone purposes	H.10–H.19
Use of telephone-type circuits for voice-frequency telegraphy	H.20–H.29
Telephone circuits or cables used for various types of telegraph transmission or simultaneous transmission	H.30–H.39
Telephone-type circuits used for facsimile telegraphy	H.40–H.49
Characteristics of data signals	H.50–H.99
CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100–H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	
General	H.200–H.219
Transmission multiplexing and synchronization	H.220–H.229
Systems aspects	H.230–H.239
Communication procedures	H.240–H.259
Coding of moving video	**H.260–H.279**
Related systems aspects	H.280–H.299
Systems and terminal equipment for audiovisual services	H.300–H.399

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION H.263

VIDEO CODING FOR LOW BIT RATE COMMUNICATION

Summary

This Recommendation specifies a coded representation that can be used for compressing the moving picture component of audio-visual services at low bit rates. The basic configuration of the video source coding algorithm is based on Recommendation H.261 and is a hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy. The source coder can operate on five standardized video source formats: sub-QCIF, QCIF, CIF, 4CIF and 16CIF, and can also operate using a broad range of custom video formats.

The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder. Half pixel precision is used for the motion compensation, as opposed to Recommendation H.261 where full pixel precision and a loopfilter are used. Variable length coding is used for the symbols to be transmitted.

In addition to the basic video source coding algorithm, sixteen negotiable coding options are included for improved compression performance and the support of additional capabilities. Additional supplemental information may also be included in the bitstream for enhanced display capability and for external usage.

Source

ITU-T Recommendation H.263 was revised by ITU-T Study Group 16 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on the 6th of February 1998.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 1998

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 175 of 341

CONTENTS

		Page
1	Scope	1
2	References	1
3	Brief specification	2
3.1	Video input and output	2
3.2	Digital output and input	2
3.3	Sampling frequency	2
3.4	Source coding algorithm	2
3.4.1	Continuous Presence Multipoint and Video Multiplex mode	3
3.4.2	Unrestricted Motion Vector mode	3
3.4.3	Syntax-based Arithmetic Coding mode	3
3.4.4	Advanced Prediction mode	3
3.4.5	PB-frames mode	3
3.4.6	Forward Error Correction	3
3.4.7	Advanced INTRA Coding mode	4
3.4.8	Deblocking Filter mode	4
3.4.9	Slice Structured mode	4
3.4.10	Supplemental enhancement information	4
3.4.11	Improved PB-frames mode	4
3.4.12	Reference Picture Selection mode	4
3.4.13	Temporal, SNR and Spatial Scalability mode	5
3.4.14	Reference Picture Resampling mode	5
3.4.15	Reduced-Resolution Update mode	5
3.4.16	Independent Segment Decoding mode	5
3.4.17	Alternative INTER VLC mode	5
3.4.18	Modified Quantization mode	5
3.5	Bit rate	6
3.6	Buffering	6
3.7	Symmetry of transmission	6
3.8	Error handling	6
3.9	Multipoint operation	7
4	Source coder	7
4.1	Source format	7
4.2	Video source coding algorithm	9
4.2.1	GOBs, slices, macroblocks and blocks	9
4.2.2	Prediction	11

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 176 of 341

		Page
4.2.3	Motion compensation	12
4.2.4	Quantization	12
4.3	Coding control	13
4.4	Forced updating	13
4.5	Byte alignment of start codes	13
5	Syntax and semantics	13
5.1	Picture layer	22
5.1.1	Picture Start Code (PSC) (22 bits)	22
5.1.2	Temporal Reference (TR) (8 bits)	23
5.1.3	Type Information (PTYPE) (Variable Length)	23
5.1.4	Plus PTYPE (PLUSPTYPE) (Variable Length)	24
5.1.5	Custom Picture Format (CPFMT) (23 bits)	27
5.1.6	Extended Pixel Aspect Ratio (EPAR) (16 bits)	28
5.1.7	Custom Picture Clock Frequency Code (CPCFC) (8 bits)	28
5.1.8	Extended Temporal Reference (ETR) (2 bits)	28
5.1.9	Unlimited Unrestricted Motion Vectors Indicator (UUI) (Variable length)	29
5.1.10	Slice Structured Submode bits (SSS) (2 bits)	29
5.1.11	Enhancement Layer Number (ELNUM) (4 bits)	29
5.1.12	Reference Layer Number (RLNUM) (4 bits)	29
5.1.13	Reference Picture Selection Mode Flags (RPSMF) (3 bits)	29
5.1.14	Temporal Reference for Prediction Indication (TRPI) (1 bit)	30
5.1.15	Temporal Reference for Prediction (TRP) (10 bits)	30
5.1.16	Back-Channel message Indication (BCI) (Variable length)	30
5.1.17	Back-Channel Message (BCM) (Variable length)	30
5.1.18	Reference Picture Resampling Parameters (RPRP) (Variable length)	30
5.1.19	Quantizer Information (PQUANT) (5 bits)	30
5.1.20	Continuous Presence Multipoint and Video Multiplex (CPM) (1 bit)	31
5.1.21	Picture Sub-Bitstream Indicator (PSBI) (2 bits)	31
5.1.22	Temporal Reference for B-pictures in PB-frames (TR_B) (3/5 bits)	31
5.1.23	Quantization information for B-pictures in PB-frames (DBQUANT) (2 bits)	31
5.1.24	Extra Insertion Information (PEI) (1 bit)	31
5.1.25	Supplemental Enhancement Information (PSUPP) (0/8/16 ... bits)	31
5.1.26	Stuffing (ESTUF) (Variable length)	32
5.1.27	End Of Sequence (EOS) (22 bits)	32
5.1.28	Stuffing (PSTUF) (Variable length)	32
5.2	Group of Blocks Layer	32
5.2.1	Stuffing (GSTUF) (Variable length)	32

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 177 of 341

		Page
5.2.2	Group of Block Start Code (GBSC) (17 bits)	33
5.2.3	Group Number (GN) (5 bits)	33
5.2.4	GOB Sub-Bitstream Indicator (GSBI) (2 bits)	33
5.2.5	GOB Frame ID (GFID) (2 bits)	33
5.2.6	Quantizer Information (GQUANT) (5 bits)	33
5.3	Macroblock layer	33
5.3.1	Coded macroblock indication (COD) (1 bit)	34
5.3.2	Macroblock type & Coded Block Pattern for Chrominance (MCBPC) (Variable length)	34
5.3.3	Macroblock mode for B-blocks (MODB) (Variable length)	37
5.3.4	Coded Block Pattern for B-blocks (CBPB) (6 bits)	37
5.3.5	Coded Block Pattern for luminance (CBPY) (Variable length)	37
5.3.6	Quantizer Information (DQUANT) (2 bits/Variable Length)	37
5.3.7	Motion Vector Data (MVD) (Variable length)	38
5.3.8	Motion Vector Data (MVD_{2-4}) (Variable length)	40
5.3.9	Motion Vector Data for B-macroblock (MVDB) (Variable length)	40
5.4	Block layer	40
5.4.1	DC coefficient for INTRA blocks (INTRADC) (8 bits)	41
5.4.2	Transform Coefficient (TCOEF) (Variable length)	41
6	Decoding process	44
6.1	Motion compensation	44
6.1.1	Differential motion vectors	44
6.1.2	Interpolation for subpixel prediction	46
6.2	Coefficients decoding	46
6.2.1	Inverse quantization	46
6.2.2	Clipping of reconstruction levels	46
6.2.3	Zigzag positioning	47
6.2.4	Inverse transform	47
6.3	Reconstruction of blocks	47
6.3.1	Summation	47
6.3.2	Clipping	48
Annex A – Inverse transform accuracy specification		48
Annex B – Hypothetical Reference Decoder		49

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 178 of 341

		Page
Annex C – Considerations for multipoint		51
C.1	Freeze picture request	51
C.2	Fast update request	51
C.3	Freeze picture release	51
C.4	Continuous Presence Multipoint and Video Multiplexing (CPM)	51
	C.4.1 End Of Sub-Bitstream code (EOSBS) (23 bits)	52
	C.4.2 Ending Sub-Bitstream Indicator (ESBI) (2 bits)	53
Annex D – Unrestricted Motion Vector mode		53
D.1	Motion vectors over picture boundaries	53
	D.1.1 Restrictions for motion vector values	54
D.2	Extension of the motion vector range	54
Annex E – Syntax-based Arithmetic Coding mode		56
E.1	Introduction	56
E.2	Specification of SAC encoder	57
E.3	Specification of SAC decoder	58
E.4	Syntax	58
E.5	PSC_FIFO	59
E.6	Header layer symbols	59
E.7	Macroblock and Block layer symbols	60
E.8	SAC models	61
Annex F – Advanced Prediction mode		64
F.1	Introduction	64
F.2	Four motion vectors per macroblock	64
F.3	Overlapped motion compensation for luminance	65
Annex G – PB-frames mode		68
G.1	Introduction	68
G.2	PB-frames and INTRA blocks	69
G.3	Block layer	69
G.4	Calculation of vectors for the B-picture in a PB-frame	69
G.5	Prediction of a B-block in a PB-frame	70
Annex H – Forward error correction for coded video signal		72
H.1	Introduction	72
H.2	Error correction framing	72
H.3	Error correcting code	72

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 179 of 341

Page

H.4 Relock time for error corrector framing 72

Annex I – Advanced INTRA Coding mode 73

I.1 Introduction 73

I.2 Syntax 74

I.3 Decoding process 74

Annex J – Deblocking Filter mode 80

J.1 Introduction 80

J.2 Relation to UMV and AP modes (Annexes D and F) 81

J.3 Definition of the deblocking edge filter 81

Annex K – Slice Structured mode 85

K.1 Introduction 85

K.2 Structure of slice layer 86

K.2.1 Stuffing (SSTUF) (Variable length) 86

K.2.2 Slice Start Code (SSC) (17 bits) 86

K.2.3 Slice Emulation Prevention Bit 1 (SEPB1) (1 bit) 86

K.2.4 Slice Sub-Bitstream Indicator (SSBI) (4 bits) 86

K.2.5 Macroblock Address (MBA) (5/6/7/9/11/12/13/14 bits) 87

K.2.6 Slice Emulation Prevention Bit 2 (SEPB2) (1 bit) 87

K.2.7 Quantizer Information (SQUANT) (5 bits) 87

K.2.8 Slice Width Indication in Macroblocks (SWI) (3/4/5/6/7 bits) 87

K.2.9 Slice Emulation Prevention Bit 3 (SEPB3) (1 bit) 88

Annex L – Supplemental Enhancement Information Specification 88

L.1 Introduction 88

L.2 PSUPP format 88

L.3 Do Nothing 89

L.4 Full-Picture Freeze Request 89

L.5 Partial-Picture Freeze Request 89

L.6 Resizing Partial-Picture Freeze Request 90

L.7 Partial-Picture Freeze-Release Request 90

L.8 Full-Picture Snapshot Tag 90

L.9 Partial-Picture Snapshot Tag 91

L.10 Video Time Segment Start Tag 91

L.11 Video Time Segment End Tag 91

L.12 Progressive Refinement Segment Start Tag 91

L.13 Progressive Refinement Segment End Tag 91

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 180 of 341

		Page
L.14	Chroma Keying Information	92
L.15	Extended Function Type	94
Annex M – Improved PB-frames mode		94
M.1	Introduction	94
M.2	B_{PB}-macroblock prediction modes	95
M.2.1	Bidirectional prediction	95
M.2.2	Forward prediction	95
M.2.3	Backward prediction	95
M.3	Calculation of vectors for bidirectional prediction of a the B-macroblock	95
M.4	MODB table	95
Annex N – Reference Picture Selection mode		96
N.1	Introduction	96
N.2	Video source coding algorithm	97
N.3	Channel for back-channel messages	97
N.3.1	Separate logical channel mode	98
N.3.2	Videomux mode	98
N.4	Syntax	98
N.4.1	Forward channel	98
N.4.2	Back-Channel Message (BCM) syntax	100
N.5	Decoder process	101
Annex O – Temporal, SNR, and Spatial Scalability mode		102
O.1	Overview	102
O.1.1	Temporal scalability	102
O.1.2	SNR scalability	103
O.1.3	Spatial scalability	104
O.1.4	Multilayer scalability	105
O.2	Transmission order of pictures	106
O.3	Picture layer syntax	108
O.4	Macroblock layer syntax	108
O.4.1	Coded macroblock indication (COD) (1 bit)	110
O.4.2	MBTYPE/MCBPC (VLC)	110
O.4.3	Coded Block Pattern for Chrominance (CBPC) (Variable length)	112
O.4.4	Coded Block Pattern for Luminance (CBPY) (Variable length)	112
O.4.5	Quantizer Information (DQUANT) (2 bits/Variable length)	113
O.4.6	Motion vector data (MVDFW, MVDBW) (Variable length)	113

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 181 of 341

		Page
O.5	Motion vector decoding	113
	O.5.1 Differential motion vectors	113
	O.5.2 Motion vectors in direct mode	113
O.6	Interpolation filters	113
Annex P – Reference picture resampling		116
P.1	Introduction	116
P.2	Syntax	119
	P.2.1 Warping Displacement Accuracy (WDA) (2 bits)	119
	P.2.2 Warping parameters (Variable length)	119
	P.2.3 Fill Mode (FILL_MODE) (2 bits)	120
	P.2.4 Fill Color Specification (Y_FILL, C_B_EPB, C_B_FILL, C_R_EPB, C_R_FILL) (26 bits)	120
P.3	Resampling algorithm	121
P.4	Example of implementation	124
	P.4.1 Displacements of virtual points	124
	P.4.2 Resampling algorithm	124
P.5	Factor-of-4 resampling	126
	P.5.1 Factor-of-4 upsampling	126
	P.5.2 Factor-of-4 downsampling	128
Annex Q – Reduced-Resolution Update mode		129
Q.1	Introduction	129
Q.2	Decoding procedure	130
	Q.2.1 Reference preparation	131
	Q.2.2 Macroblock layer decoding	131
	Q.2.3 Picture store	132
	Q.2.4 Display	132
Q.3	Extension of referenced picture	133
Q.4	Reconstruction of motion vectors	134
Q.5	Enlarged overlapped motion compensation for luminance	136
Q.6	Upsampling of the reduced-resolution reconstructed prediction error	138
	Q.6.1 Upsampling procedure for the pixels inside a 16 × 16 reconstructed prediction error block	140
	Q.6.2 Upsampling procdedure for the pixels at the boundary of 16 × 16 reconstructed prediction error block	141
Q.7	Block boundary filter	141
	Q.7.1 Definition of the default block boundary filter	142

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 182 of 341

Page

Q.7.2 Definition of the block boundary filter when Deblocking Filter mode is used 142

Annex R – Independent Segment Decoding mode 143

R.1 Introduction 143

R.2 Mode operation 143

R.3 Constraints on usage 144

R.3.1 Constraint on segment shapes 144

R.3.2 Constraint on changes of segment shapes 144

Annex S – Alternative INTER VLC mode 145

S.1 Introduction 145

S.2 Alternative INTER VLC for coefficients 145

S.2.1 Encoder action 145

S.2.2 Decoder action 145

S.3 Alternative INTER VLC for CBPY 146

Annex T – Modified Quantization mode 146

T.1 Introduction 146

T.2 Modified DQUANT Update 146

T.2.1 Small-step QUANT alteration 146

T.2.2 Arbitrary QUANT selection 147

T.3 Altered quantization step size for chrominance coefficients 147

T.4 Modified coefficient range 148

T.5 Usage restrictions 148

Appendix I – Error tracking 149

I.1 Introduction 149

I.2 Error tracking 149

Appendix II – Recommended Optional Enhancement 150

II.1 Introduction 150

II.2 Levels of preferred mode support 151

II.2.1 Level 1 preferred modes 151

II.2.2 Level 2 preferred modes 152

II.2.3 Level 3 preferred modes 152

II.3 Picture formats and picture clock frequencies 153

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 183 of 341

Recommendation H.263

VIDEO CODING FOR LOW BIT RATE COMMUNICATION

(revised in 1998)

1 Scope

This Recommendation specifies a coded representation that can be used for compressing the moving picture component of audio-visual services at low bit rates. The basic configuration of the video source coding algorithm is based on Recommendation H.261. Sixteen negotiable coding options are included for improved performance and increased functionality. This Recommendation contains Version 2 of Recommendation H.263, which is fully compatible with the original Recommendation, adding only optional features to the original Version 1 content of the Recommendation.

2 References

The following Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below.

[1] ITU-R Recommendation BT.601-5 (1995), *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios.*

Reference [1] is referenced herein to define the (Y, C_B, C_R) colour space and its 8-bit integer representation for pictures used by video codecs designed according to this Recommendation. (Reference [1] is not used to define any other aspects of this Recommendation.)

The following additional ITU-T Recommendations are mentioned for illustration purposes in this text; however, they do not constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

[2] ITU-T Recommendation H.223 (1996), *Multiplexing protocol for low bit rate multimedia communication.*

[3] ITU-T Recommendation H.242 (1997), *System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s.*

[4] ITU-T Recommendation H.245 (1998), *Control protocol for multimedia communication.*

[5] ITU-T Recommendation H.261 (1993), *Video codec for audiovisual services at $p \times 64$ kbit/s.*

[6] ITU-T Recommendation H.262 (1995) | ISO/IEC 13818-2:1996, *Information technology – Generic coding of moving pictures and associated audio information: Video.*

[7] ITU-T Recommendation H.324 (1998), *Terminal for low bit rate multimedia communication.*

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 184 of 341

3 Brief specification

An outline block diagram of the codec is given in Figure 1.

3.1 Video input and output

To permit a single Recommendation to cover use in and between regions using 625- and 525-line television standards, the standardized source formats on which the source coder operates are based on a Common Intermediate Format (CIF). It is also possible using external negotiation (for example, Recommendation H.245), to enable the use of a wide range of optional custom source formats. The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the source coding format are not subject to recommendation.

3.2 Digital output and input

The video coder provides a self-contained digital bitstream which may be combined with other multi-facility signals (for example as defined in Recommendation H.223). The video decoder performs the reverse process.

3.3 Sampling frequency

Pictures are sampled at an integer multiple of the video line rate. This sampling clock and the digital network clock are asynchronous.

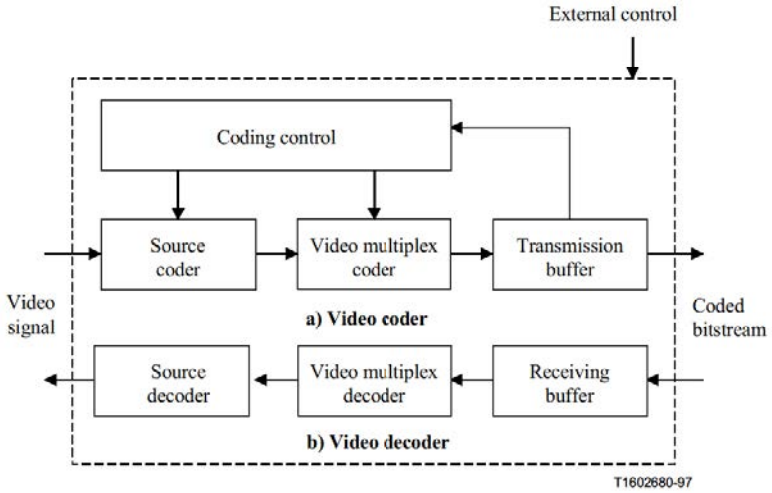


Figure 1/H.263 – Outline block diagram of the video codec

3.4 Source coding algorithm

A hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy is adopted. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder. Half pixel precision is used for the motion compensation, as opposed to Recommendation H.261 where full pixel precision and a loopfilter are used. Variable length coding is used for the symbols to be transmitted.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 185 of 341

In addition to the core H.263 coding algorithm, sixteen negotiable coding options can be used, either together or separately (subject to certain restrictions). Additional supplemental information may also be included in the bitstream for enhanced display capability and for external usage. A forward error correction method for application to the resulting video bitstream is provided for use when necessary. The negotiable coding options, forward error correction, and supplemental information usage are described in the subsequent subclauses.

3.4.1 Continuous Presence Multipoint and Video Multiplex mode

In this optional mode, up to four separate video "Sub-Bitstreams" can be sent within the same video channel. This feature is designed for use in continuous presence multipoint application or other situations in which separate logical channels are not available, but the use of multiple video bitstreams is desired (see also Annex C).

3.4.2 Unrestricted Motion Vector mode

In this optional mode, motion vectors are allowed to point outside the picture. The edge pixels are used as prediction for the "non-existing" pixels. With this mode a significant gain is achieved if there is movement across the edges of the picture, especially for the smaller picture formats (see also Annex D). Additionally, this mode includes an extension of the motion vector range so that larger motion vectors can be used. This is especially useful in case of camera movement and large picture formats.

3.4.3 Syntax-based Arithmetic Coding mode

In this optional mode, arithmetic coding is used instead of variable length coding. The SNR and reconstructed pictures will be the same, but significantly fewer bits will be produced (see also Annex E).

3.4.4 Advanced Prediction mode

In this optional mode, Overlapped Block Motion Compensation (OBMC) is used for the luminance part of P-pictures (see also Annex F). Four 8×8 vectors instead of one 16×16 vector are used for some of the macroblocks in the picture. The encoder has to decide which type of vectors to use. Four vectors use more bits, but give better prediction. The use of this mode generally gives a considerable improvement. A subjective gain is achieved because OBMC results in less blocking artifacts.

3.4.5 PB-frames mode

A PB-frame consists of two pictures being coded as one unit. The name PB comes from the name of picture types in Recommendation H.262 where there are P-pictures and B-pictures. Thus a PB-frame consists of one P-picture which is predicted from the previous decoded P-picture and one B-picture which is predicted from both the previous decoded P-picture and the P-picture currently being decoded. The name B-picture was chosen because parts of B-pictures may be bidirectionally predicted from the past and future pictures. With this coding option, the picture rate can be increased considerably without substantially increasing the bit rate (see also Annex G). However, an Improved PB-frames mode is also provided (see also Annex M). The original PB-frames mode is retained herein only for purposes of compatibility with systems made prior to the adoption of the Improved PB-frames mode.

3.4.6 Forward Error Correction

A forward error correction method is specified for use when necessary for the protection of the video bitstream for use when appropriate. The method provided for forward error correction is the same BCH code method specified also in Recommendation H.261 (see also Annex H).

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 186 of 341

3.4.7 Advanced INTRA Coding mode

In this optional mode, INTRA blocks are first predicted from neighbouring INTRA blocks prior to coding (see also Annex I). Separate Variable Length Code (VLC) tables are defined for the INTRA blocks. The technique is applied to INTRA-macroblocks within INTRA pictures and to INTRA-macroblocks within INTER pictures. This mode significantly improves the compression performance over the INTRA coding of the core H.263 syntax.

3.4.8 Deblocking Filter mode

In this optional mode, a filter is applied across the 8 × 8 block edge boundaries of decoded I- and P-pictures to reduce blocking artifacts (see also Annex J). The purpose of the filter is to mitigate the occurrence of block edge artifacts in the decoded picture. The filter affects the picture that is used for the prediction of subsequent pictures and thus lies within the motion prediction loop.

3.4.9 Slice Structured mode

In this optional mode, a "slice" layer is substituted for the GOB layer of the bitstream syntax (see also Annex K). The purposes of this mode are to provide enhanced error resilience capability, to make the bitstream more amenable to use with an underlying packet transport delivery, and to minimize video delay. A slice is similar to a GOB, in that it is a layer of the syntax that lies between the picture layer and the macroblock layer. However, the use of a slice layer allows a flexible partitioning of the picture, in contrast with the fixed partitioning and fixed transmission order required by the GOB structure.

3.4.10 Supplemental enhancement information

Additional supplemental information may be included in the bitstream to signal enhanced display capability or to provide information for external usage (see also Annex L). This supplemental information can be used to signal a full-picture or partial-picture freeze or freeze-release request with or without resizing, can also be used to tag particular pictures or sequences of pictures within the video stream for external usage, and can be used to convey chroma key information for video compositing. The supplemental information may be present in the bitstream even though the decoder may not be capable of providing the enhanced capability to use it, or even to properly interpret it – simply discarding the supplemental information is allowable by decoders unless a requirement to provide the requested capability has been negotiated by external means.

3.4.11 Improved PB-frames mode

This optional mode represents an improvement compared to the PB-frames mode option (see also Annexes G and M). The main difference between the two modes is that in the Improved PB-frames mode, each B-block may be forward predicted using a separate motion vector or backward predicted with a zero vector. This significantly improves coding efficiency in situations in which downscaled P-vectors are not good candidates for B-prediction. The backward prediction is particularly useful when there is a scene cut between the previous P-frame and the PB-frame.

3.4.12 Reference Picture Selection mode

An optional mode is provided which improves the performance of real-time video communication over error-prone channel by allowing temporal prediction from pictures other than the most recently-sent reference picture (see also Annex N). This mode can be used with back-channel status messages which are sent back to the encoder to inform it about whether its bitstream is being properly received. In error-prone channel environments, this mode allows the encoder to optimize its video encoding for the conditions of the channel.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 187 of 341

3.4.13 Temporal, SNR and Spatial Scalability mode

In this optional mode, there is support for temporal, SNR, and spatial scalability (see also Annex O). Scalability implies that a bitstream is composed of a base layer and one or more associated enhancement layers. The base layer is a separately decodable bitstream. The enhancement layers can be decoded in conjunction with the base layer to increase perceived quality by either increasing the picture rate, increasing the picture quality, or increasing the picture size. SNR scalability refers to enhancement information to increase the picture quality without increasing picture resolution. Spatial scalability refers to enhancement information to increase the picture quality by increasing picture resolution either horizontally, vertically, or both. There is also support for temporal scalability by the use of B-pictures. A B-picture is a scalability enhancement containing pictures that can be bidirectionally predicted from two pictures in a reference layer, one temporally previous to the current picture and one temporally subsequent. B-pictures allows enhancement layer information to be used to increase perceived quality by increasing the picture rate of the displayed enhanced video sequence. This mode can be useful for heterogenous networks with varying bandwidth capacity and also in conjunction with error correction schemes.

3.4.14 Reference Picture Resampling mode

A syntax is provided for supporting an optional mode for which the reference picture used for video image prediction is processed by a resampling operation prior to its use in forming a predictor for the current input picture (see also Annex P). This allows efficient dynamic selection of an appropriate image resolution for video encoding, and can also support picture warping for use as a global motion compensator or special-effect generator.

3.4.15 Reduced-Resolution Update mode

An optional mode is provided which allows reduced-resolution updates to a reference picture having a higher resolution (see also Annex Q). This mode is expected to be used when encoding a highly active scene, and allows an encoder to increase the picture rate at which moving parts of a scene can be represented, while maintaining a higher resolution representation in more static areas of the scene.

3.4.16 Independent Segment Decoding mode

An optional mode is provided which allows a picture to be constructed without any data dependencies which cross GOB or multi-GOB video picture segments or slice boundaries (see also Annex R). This mode provides error robustness by preventing the propagation of erroneous data across the boundaries of the video picture segment areas.

3.4.17 Alternative INTER VLC mode

An optional mode is provided which improves the efficiency of INTER picture coding when significant changes are evident in the picture (see also Annex S). This efficiency improvement is obtained by allowing a VLC code originally designed for INTRA pictures to be used for some INTER picture coefficients as well.

3.4.18 Modified Quantization mode

An optional mode is provided which improves the bit-rate control ability for encoding, reduces chrominance quantization error, expands the range of representable DCT coefficients, and places certain restrictions on coefficient values (see also Annex T). This mode modifies the semantics of the differential quantization step size parameter of the bitstream by broadening the range of step size changes that can be specified. It also reduces the quantization step size used for chrominance data. The range of DCT coefficient levels is broadened to ensure that any possible coefficient value can be

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 188 of 341

encoded to within the accuracy allowed by the step size. Certain restrictions are also placed on coefficients in this mode to increase error detection performance and minimize decoder complexity.

3.5 Bit rate

The transmission clock is provided externally. The video bit rate may be variable. In this Recommendation no constraints on the video bit rate are given; constraints will be given by the terminal or the network.

3.6 Buffering

The encoder shall control its output bitstream to comply with the requirements of the hypothetical reference decoder defined in Annex B. Video data shall be provided on every valid clock cycle. This can be ensured by MCBPC stuffing (see Tables 7 and 8) or, when forward error correction is used, also by forward error correction stuffing frames (see Annex H).

The number of bits created by coding any single picture shall not exceed a maximum value specified by the parameter BPPmaxKb which is measured in units of 1024 bits. The minimum allowable value of the BPPmaxKb parameter depends on the largest picture size that has been negotiated for use in the bitstream (see Table 1). The picture size is measured as the picture width times height for the luminance (Y) component, measured in pixels. An encoder may use a larger value for BPPmaxKb than as specified in Table 1, provided the larger value is first negotiated by external means, for example Recommendation H.245.

When the Temporal, SNR, and Spatial Scalability mode (Annex O) is employed, the number of bits sent for each picture in each enhancement layer shall not exceed the maximum value specified by BPPmaxKb.

Table 1/H.263 – Minimum BPPmaxKb for different source picture formats

Y picture size in pixels	Minimum BPPmaxKb
up to 25 344 (or QCIF)	64
25 360 to 101 376 (or CIF)	256
101 392 to 405 504 (or 4CIF)	512
405 520 and above	1024

3.7 Symmetry of transmission

The codec may be used for bidirectional or unidirectional visual communication.

3.8 Error handling

Error handling should be provided by external means (for example, Recommendation H.223). If it is not provided by external means (for example, in Recommendation H.221) the optional error correction code and framing as described in Annex H can be used.

A decoder can send a command to encode one or more GOBs (or slices if Annex K is employed) of its next picture in INTRA mode with coding parameters such as to avoid buffer overflow. A decoder can also send a command to transmit only non-empty GOB headers if the Slice Structured mode (see Annex K) is not in use. The transmission method for these signals is by external means (for example, Recommendation H.245).

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 189 of 341

3.9 Multipoint operation

Features necessary to support switched multipoint operation are included in Annex C.

4 Source coder

4.1 Source format

The source coder operates on non-interlaced pictures having a source format which is defined in terms of:

1) the picture format, as determined by the number of pixels per line, the number of lines per picture, and the pixel aspect ratio; and

2) the timing between pictures, as determined by the Picture Clock Frequency (PCF). For example, the Common Intermediate Format (CIF) has 352 pixels per line, 288 lines, a pixel aspect ratio of 12:11, and a picture clock frequency of 30 000/1001 pictures per second.

The source coder operates on non-interlaced pictures occurring at a Picture Clock Frequency (PCF) of 30 000/1001 (approximately 29.97) times per second, termed the CIF PCF. It is also possible to negotiate the use of an optional custom PCF by external means. A custom PCF is given by 1 800 000/(clock divisor * clock conversion factor) where clock divisor can have values of 1 through 127 and clock conversion factor can either be 1000 or 1001. The tolerance on the picture clock frequency is ± 50 ppm.

Pictures are coded as luminance and two colour difference components (Y, C_B and C_R). These components and the codes representing their sampled values are as defined in ITU-R Recommendation BT.601-5.

- Black = 16;
- White = 235;
- Zero colour difference = 128;
- Peak colour difference = 16 and 240.

These values are nominal ones and the coding algorithm functions with input values of 1 through to 254.

There are five standardized picture formats: sub-QCIF, QCIF, CIF, 4CIF and 16CIF. It is also possible to negotiate a custom picture format. For all of these picture formats, the luminance sampling structure is dx pixels per line, dy lines per picture in an orthogonal arrangement. Sampling of each of the two colour difference components is at dx/2 pixels per line, dy/2 lines per picture, orthogonal. The values of dx, dy, dx/2 and dy/2 are given in Table 2 for each of the standardized picture formats.

Table 2/H.263 – Number of pixels per line and number of lines for each of the standardized H.263 picture formats

Picture format	Number of pixels for luminance (dx)	Number of lines for luminance (dy)	Number of pixels for chrominance (dx/2)	Number of lines for chrominance (dy/2)
sub-QCIF	128	96	64	48
QCIF	176	144	88	72
CIF	352	288	176	144
4CIF	704	576	352	288
16CIF	1408	1152	704	576

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 190 of 341

For all picture formats, colour difference samples are sited such that their block boundaries coincide with luminance block boundaries as shown in Figure 2. The pixel aspect ratio is the same for each of the standardised picture formats and is the same as defined for QCIF and CIF in Recommendation H.261: (288/3):(352/4), which simplifies to 12:11 in relatively prime numbers. The picture area covered by all of the standardized picture formats except the sub-QCIF picture format has an aspect ratio of 4:3.

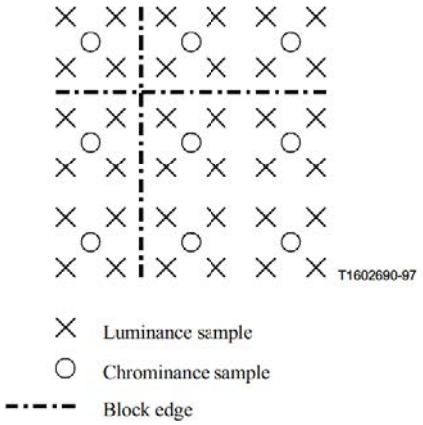


Figure 2/H.263 – Positioning of luminance and chrominance samples

Custom picture formats can have a custom pixel aspect ratio as described in Table 3, if the custom pixel aspect ratio use is first negotiated by external means. Custom picture formats can have any number of lines and any number of pixels per line, provided that the number of lines is divisible by four and is in the range [4, ..., 1152], and provided that the number of pixels per line is also divisible by four and is in the range [4, ..., 2048]. For picture formats having a width or height that is not divisible by 16, the picture is decoded in the same manner as if the width or height had the next larger size that would be divisible by 16 and then the picture is cropped at the right and the bottom to the proper width and height for display purposes only.

Table 3/H.263 – Custom pixel aspect ratios

Pixel aspect ratio	Pixel width:Pixel height
Square	1:1
CIF	12:11
525-type for 4:3 picture	10:11
CIF for 16:9 picture	16:11
525-type for 16:9 picture	40:33
Extended PAR	m:n, m and n are relatively prime

All decoders and encoders shall be able to operate using the CIF picture clock frequency. Some decoders and encoders may also support custom picture clock frequencies. All decoders shall be able to operate using the sub-QCIF picture format. All decoders shall also be able to operate using the QCIF picture format. Some decoders may also operate with CIF, 4CIF or 16CIF, or custom picture formats. Encoders shall be able to operate with one of the picture formats sub-QCIF and QCIF. The

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 191 of 341

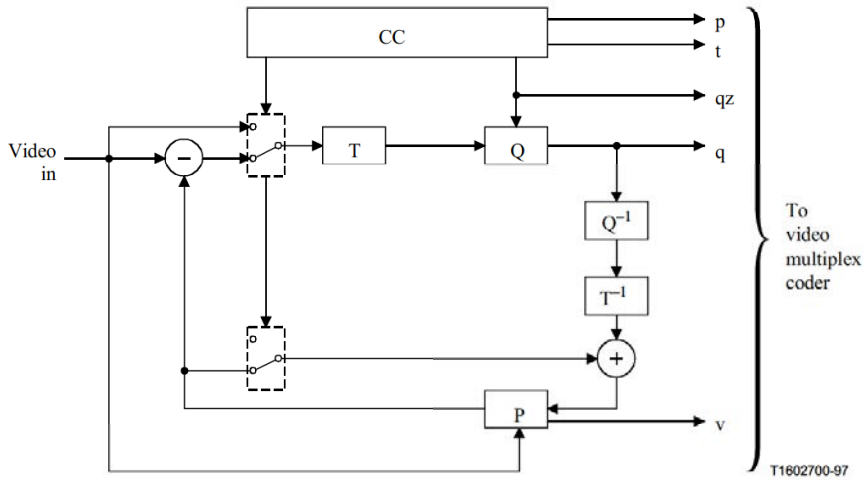
encoders determine which of these two formats are used, and are not obliged to be able to operate with both. Some encoders can also operate with CIF, 4CIF, 16CIF, or custom picture formats. Which optional formats and which picture clock frequencies can be handled by the decoder is signalled by external means, for example Recommendation H.245. For a complete overview of possible picture formats and video coding algorithms, refer to the terminal description, for example Recommendation H.324.

NOTE – For CIF, the number of pixels per line is compatible for practical purposes with sampling the active portions of the luminance and colour difference signals from 525- or 625-line sources at 6.75 and 3.375 MHz respectively. These frequencies have a simple relationship to those in ITU-R Recommendation BT.601-5.

Means shall be provided to restrict the maximum picture rate of encoders by having a minimum number of non-transmitted pictures between transmitted ones. Selection of this minimum number shall be by external means (for example, Recommendation H.245). For the calculation of the minimum number of non-transmitted pictures in PB-frames mode, the P-picture and the B-picture of a PB-frames unit are taken as two separate pictures.

4.2 Video source coding algorithm

The source coder is shown in generalized form in Figure 3. The main elements are prediction, block transformation and quantization.



T	Transform
Q	Quantizer
P	Picture Memory with motion compensated variable delay
CC	Coding control
p	Flag for INTRA/INTER
t	Flag for transmitted or not
qz	Quantizer indication
q	Quantizing index for transform coefficients
v	Motion vector

Figure 3/H.263 – Source coder

4.2.1 GOBs, slices, macroblocks and blocks

Each picture is divided either into Groups Of Blocks (GOBs) or into slices.

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 192 of 341

A Group Of Blocks (GOB) comprises of up to k * 16 lines, where k depends on the number of lines in the picture format and depends on whether the optional Reduced-Resolution Update mode is in use (see Annex Q). The dependencies are shown in Table 4. If the number of lines is less than or equal to 400 and the optional Reduced-Resolution Update mode is not in use, then k = 1. If the number of lines is less than or equal to 800 and the optional Reduced-Resolution Update mode is in use or the number of lines is more than 400, then k = 2. If the number of lines is more than 800, then k = 4. When using custom picture sizes, the number of lines in the last (bottom-most) GOB may be less than k * 16 if the number of lines in the picture is not divisible by k * 16. However, every GOB in each of the standardized picture formats has k * 16 lines, as the number of lines in each standardized picture format is an integer multiple of k * 16. Thus, for example, if the optional Reduced-Resolution mode is not in use, the number of GOBs per picture is 6 for sub-QCIF, 9 for QCIF, and 18 for CIF, 4CIF and 16CIF. The GOB numbering is done by use of vertical scan of the GOBs, starting with the upper GOB (number 0) and ending with the bottom-most GOB. An example of the arrangement of GOBs in a picture is given for the CIF picture format in Figure 4. Data for each GOB consists of a GOB header (may be empty) followed by data for macroblocks. Data for GOBs is transmitted per GOB in increasing GOB number.

Table 4/H.263 – Parameter k for GOB size definition

Number of Lines dy	Value of k when not in RRU mode	Value of k when in RRU mode
4,…,400	1	2
404,…,800	2	2
804,…,1152	4	4

The Slice Structured mode is described in Annex K. Slices are similar to GOBs in that they are a multi-macroblock layer of the syntax, but slices have a more flexible shape and usage than GOBs, and slices may appear in the bitstream in any order, under certain conditions.

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Figure 4/H.263 – Arrangement of Groups of Blocks in a CIF picture

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 193 of 341

Each GOB is divided into macroblocks. The macroblock structure depends on whether the optional Reduced-Resolution Update (RRU) mode is in use (see Annex Q). Unless in RRU mode, each macroblock relates to 16 pixels by 16 lines of Y and the spatially corresponding 8 pixels by 8 lines of C_B and C_R. Further, a macroblock consists of four luminance blocks and the two spatially corresponding colour difference blocks as shown in Figure 5. Each luminance or chrominance block thus relates to 8 pixels by 8 lines of Y, C_B or C_R. Unless in RRU mode, a GOB comprises one macroblock row for sub-QCIF, QCIF and CIF, two macroblock rows for 4CIF and four macroblock rows for 16CIF.

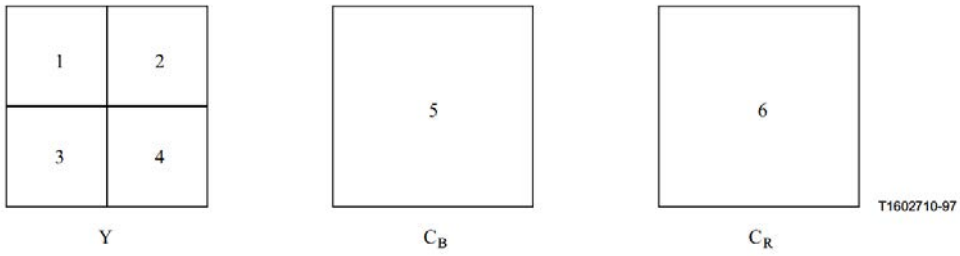


Figure 5/H.263 – Arrangement of blocks in a macroblock

When in RRU mode, a macroblock relates to 32 pixels by 32 lines of Y and the spatially corresponding 16 pixels by 16 lines of C_B and C_R, and each luminance or chrominance block relates to 16 pixels by 16 lines of Y, C_B or C_R. Furthermore, a GOB comprises one macroblock row for CIF and 4CIF, and two macroblock rows for 16CIF.

The macroblock numbering is done by use of horizontal scan of the macroblock rows from left to right, starting with the upper macroblock row and ending with the lower macroblock row. Data for the macroblocks is transmitted per macroblock in increasing macroblock number. Data for the blocks is transmitted per block in increasing block number (see Figure 5).

The criteria for choice of mode and transmitting a block are not subject to recommendation and may be varied dynamically as part of the coding control strategy. Transmitted blocks are transformed and resulting coefficients are quantized and entropy coded.

4.2.2 Prediction

The primary form of prediction is inter-picture and may be augmented by motion compensation (see 4.2.3). The coding mode in which temporal prediction is applied is called INTER; the coding mode is called INTRA if no temporal prediction is applied. The INTRA coding mode can be signalled at the picture level (INTRA for I-pictures or INTER for P-pictures) or at the macroblock level in P-pictures. In the optional PB-frames mode, B-pictures are always coded in INTER mode. The B-pictures are partly predicted bidirectionally (refer to Annex G).

In total, H.263 has seven basic picture types (of which only the first two are mandatory) which are defined primarily in terms of their prediction structure:

1) INTRA: A picture having no reference picture(s) for prediction (also called an I-picture);
2) INTER: A picture using a temporally previous reference picture (also called a P-picture);
3) PB: A frame representing two pictures and having a temporally previous reference picture (see Annex G);
4) Improved PB: A frame functionally similar but normally better than a PB-frame (see Annex M);

Snap - Exhibit 1016
Snap v. Nokia - IPR2025-01440
Page 194 of 341