Heinrich-Hertz-Intitut Berlin - Image Processing Department, Thomas Sikora



# MPEG-1 and MPEG-2 Digital Video Coding Standards

The purpose of this page is to provide an overview of the MPEG-1 and MPEG-2 video coding algorithms and standards and their role in video communications. The text is organized as follows: The basic concepts and techniques which are relevant in the context of the MPEG video compression standards are reviewed first. In the following the MPEG-1 and MPEG-2 video coding algorithms are outlined in more detail. Furthermore the specific properties of the standards related to their applications are presented.

### Fundamentals of MPEG Video Compression Algorithms

Generally speaking, video sequences contain a significant amount of *statistical* and *subjective* redundancy within and between frames. The ultimate goal of video source coding is the bit-rate reduction for storage and transmission by exploring both statistical and subjective redundancies and to encode a "minimum set" of information using entropy coding techniques. This usually results in a compression of the coded video data compared to the original source data. The performance of video compression techniques depends on the amount of redundancy contained in the image data as well as on the actual compression techniques used for coding. With practical coding schemes a trade-off between coding performance (high compression with sufficient quality) and implementation complexity is targeted. For the development of the MPEG compression algorithms the consideration of the capabilities of "state of the art" (VLSI) technology foreseen for the lifecycle of the standards was most important.

Dependent on the applications requirements we may envisage "lossless" and "lossy" coding of the video data. The aim of "lossless" coding is to reduce image or video data for storage and transmission while retaining the quality of the original images - the *de*coded image quality is required to be identical to the image quality prior to *en*coding. In contrast the aim of "lossy" coding techniques - and this is relevant to the applications envisioned by MPEG-1 and MPEG-2 video standards - is to meet a given target bit-rate for storage and transmission. Important applications comprise transmission of video over communications channels with constrained or low bandwidth and the efficient storage of video. In these applications high video compression is achieved by degrading the video quality - the *de*coded image "objective" quality is reduced compared to the quality of the original images prior to encoding (i.e. taking the mean-squared-error between both the original and reconstructed images as an objective image quality criteria). The smaller the target bit-rate of the channel the higher the necessary compression of the video data and usually the more coding artefacts become visible. The ultimate aim of lossy coding techniques is to optimise image quality for a given target bit rate subject to "objective" or "subjective" optimisation criteria. It should be noted that the degree of image degradation (both the objective degradation as well as the amount of visible artefacts) depends on the complexity of the image or video scene as much as on the sophistication of the compression technique - for simple textures in images and low video activity a good image reconstruction with no visible artefacts may be achieved even with simple compression techniques.

(A) The MPEG Video Coder Source Model

The MPEG digital video coding techniques are statistical in nature. Video sequences usually contain statistical redundancies in both temporal and spatial directions. The basic statistical property upon which MPEG compression techniques rely is inter-pel correlation, including the assumption of simple correlated translatory motion between consecutive frames. Thus, it is assumed that the magnitude of a particular image pel can be predicted from nearby pels within the same frame (using Intra-frame coding techniques) or from pels of a nearby frame (using Inter-frame techniques). Intuitively it is clear that in some circumstances, i.e. during scene changes of a video sequence, the temporal correlation between pels in nearby frames is small or even vanishes - the video scene then assembles a collection of uncorrelated still images. In this case Intra-frame coding techniques are appropriate to explore spatial correlation to achieve efficient data compression. The MPEG compression algorithms employ Discrete Cosine Transform (DCT) coding techniques on image blocks of 8x8 pels to efficiently explore spatial correlations between nearby pels within the same image. However, if the correlation between pels in nearby frames is high, i.e. in cases where two consecutive frames have similar or identical content, it is desirable to use Inter-frame DPCM coding techniques employing temporal prediction (motion compensated prediction between frames). In MPEG video coding schemes an adaptive combination of both temporal motion compensated prediction followed by transform coding of the remaining spatial information is used to achieve high data compression (hybrid DPCM/DCT coding of video).

Figure 1 depicts an example of Intra-frame pel-to-pel correlation properties of images, here modelled using a rather simple, but nevertheless valuable statistical model. The simple model assumption already inherits basic correlation properties of many "typical" images upon which the MPEG algorithms rely, namely the high correlation between adjacent pixels and the monotonical decay of correlation with increased distance between pels. We will use this model assumption later to demonstrate some of the properties of Transform domain coding.

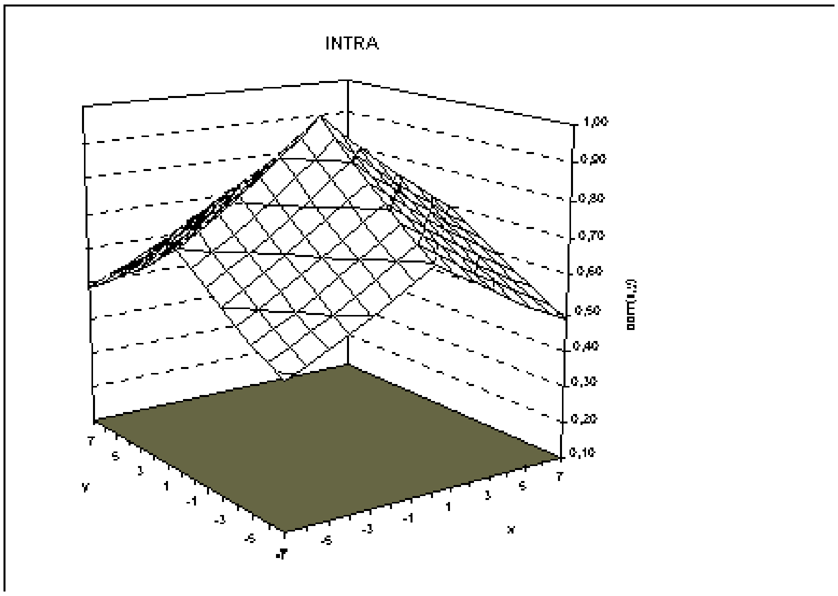


Figure 1: Spatial inter-element correlation of "typical" images as calculated using a AR(1) Gauss Markov image model with high pel-pel correlation. Variables *x* and *y* describe the distance between pels in horizontal and vertical image dimensions respectively.

## (B) Subsampling and Interpolation

Almost all video coding techniques described in the context of this paper make extensive use of subsampling and quantization prior to encoding. The basic concept of subsampling is to reduce the dimension of the input video (horizontal dimension and/or vertical dimension) and thus the number of pels to be coded prior to the encoding process. It is worth noting that for some applications video is also subsampled in temporal direction to reduce frame rate prior to coding. At the receiver the decoded images are interpolated for display. This technique may be considered as one of the most elementary compression techniques which also makes use of specific physiological characteristics of the human eye and thus removes subjective redundancy contained in the video data - i.e. the human eye is more sensitive to changes in brightness than to chromaticity changes. Therefore the MPEG coding schemes first divide the images into YUV components (one luminance and two chrominance components). Next the chrominance components are subsampled relative to the luminance component with a Y:U:V ratio specific to particular applications (i.e. with the MPEG-2 standard a ratio of 4:1:1 or 4:2:2 is used).

## (C) Motion Compensated Prediction

Motion compensated prediction is a powerful tool to reduce temporal redundancies between frames and is used extensively in MPEG-1 and MPEG-2 video coding standards as a prediction technique for temporal DPCM coding. The concept of motion compensation is based on the estimation of motion between video frames, i.e. if all elements in a video scene are approximately spatially displaced, the motion between frames can be described by a limited number of motion parameters (i.e. by motion vectors for translatory motion of pels). In this simple example the best prediction of an actual pel is given by a motion compensated prediction pel from a previously coded frame. Usually both, prediction error and motion vectors, are transmitted to the receiver. However, encoding one motion information with each coded image pel is generally neither desirable nor necessary. Since the spatial correlation between motion vectors is often high it is sometimes assumed that one motion vector is representative for the motion of a "block" of adjacent pels. To this aim images are usually separated into disjoint blocks of pels (i.e. 16x16 pels in MPEG-1 and MPEG-2 standards) and only one motion vector is estimated, coded and transmitted for each of these blocks (Figure 2).

In the MPEG compression algorithms the motion compensated prediction techniques are used for reducing temporal redundancies between frames and only the prediction error images - the difference between original images and motion compensated prediction images - are encoded. In general the correlation between pels in the motion compensated Inter-frame error images to be coded is reduced compared to the correlation properties of Intra-frames in Figure 1 due to the prediction based on the previous coded frame.

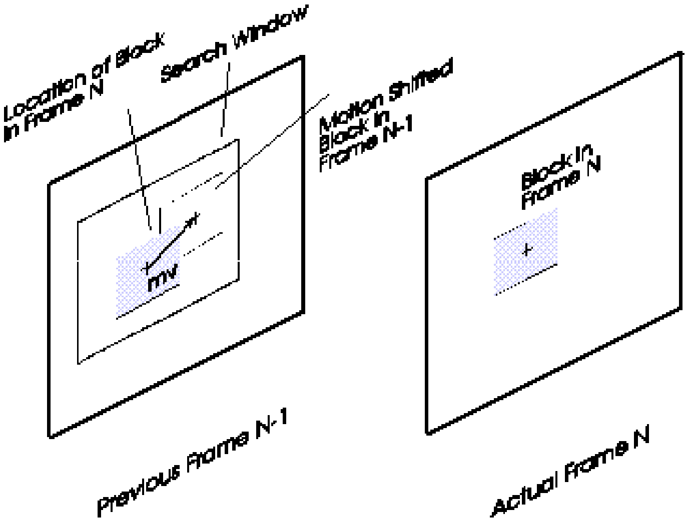



Figure 2: Block matching approach for motion compensation: One motion vector (mv) is estimated for each block in the actual frame N to be coded. The motion vector points to a reference block of same size in a previously coded frame N-1. The motion compensated prediction error is calculated by subtracting each pel in a block with its motion shifted counterpart in the reference block of the previous frame.

(D) Transform Domain Coding

Transform coding has been studied extensively during the last two decades and has become a very popular compression method for still image coding and video coding. The purpose of Transform coding is to de-correlate the Intra- or Inter-frame error image content and to encode Transform coefficients rather than the original pels of the images. To this aim the input images are split into disjoint blocks of pels ***b*** (i.e. of size *NxN* pels). The transformation can be represented as a matrix operation using a *NxN* Transform matrix ***A*** to obtain the *NxN* transform coefficients ***c*** based on a linear, separable and unitary *forward* transformation

$$\boldsymbol{c} = \boldsymbol{A}\,\boldsymbol{b}\,\boldsymbol{A}^{T}.$$

Here, $\boldsymbol{A}^T$ denotes the transpose of the transformation matrix ***A***. Note, that the transformation is reversible, since the original *NxN* block of pels ***b*** can be reconstructed using a linear and separable *inverse* transformation

$$\boldsymbol{b} = \boldsymbol{A}^{T}\boldsymbol{c}\,\boldsymbol{A}.$$

Upon many possible alternatives the Discrete Cosine Transform (DCT) applied to smaller image blocks of usually *8x8* pels has become the most successful transform for still image and video coding [ahmed]. In fact, DCT based implementations are used in most image and video coding standards due to their high decorrelation performance and the availability of fast DCT algorithms suitable for real time implementations. VLSI implementations that operate at rates suitable for a broad range of video applications are commercially available today.

A major objective of transform coding is to make as many Transform coefficients as possible small enough so that they are insignificant (in terms of statistical and subjective measures) and need not be coded for transmission. At the same time it is desirable to minimize statistical dependencies between coefficients with the aim to reduce the amount of bits needed to encode the remaining coefficients. Figure 3 depicts the variance (energy) of a *8x8* block of Intra-frame DCT coefficients based on the simple statistical model assumption already discussed in Figure 1. Here, the variance for each coefficient represents the variability of the coefficient as averaged over a large number of frames. Coefficients with small variances are less significant for the reconstruction of the image blocks than coefficients with large variances. As may be depicted from Figure 3, on average only a small number of DCT coefficients need to be transmitted to the receiver to obtain a valuable approximate reconstruction of the image blocks. Moreover, the most significant DCT coefficients are concentrated around the upper left corner (low DCT coefficients) and the significance of the coefficients decays with increased distance. This implies that higher DCT coefficients are less important for reconstruction than lower coefficients. Also employing motion compensated prediction the transformation using the DCT usually results in a compact representation of the temporal DPCM signal in the DCT-domain - which essentially inherits the similar statistical coherency as the signal in the DCT-domain for the Intra-frame signals in Figure 3 (although with reduced energy) - the reason why MPEG algorithms employ DCT coding also for Inter-frame compression successfully [schaf].

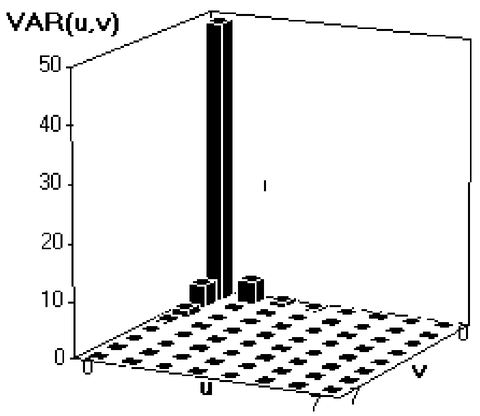



Figure 3: The figure depicts the variance distribution of DCT-coefficients "typically" calculated as average over a large number of image blocks. The variance of the DCT coefficients was calculated based on the statistical model used in Figure 1. *u* and *v* describe the horizontal and vertical image transform domain variables within the *8x8* block. Most of the total variance is concentrated around the DC DCT-coefficient (*u=0*, *v=0*).

The DCT is closely related to Discrete Fourier Transform (DFT) and it is of some importance to realize that the DCT coefficients can be given a frequency interpretation close to the DFT. Thus low DCT coefficients relate to low spatial frequencies within image blocks and high DCT coefficients to higher frequencies. This property is used in MPEG coding schemes to remove subjective redundancies contained in the image data based on human visual systems criteria. Since the human viewer is more sensitive to reconstruction errors related to low spatial frequencies than to high frequencies, a frequency adaptive weighting (quantization) of the coefficients according to the human visual perception (perceptual quantization) is often employed to improve the visual quality of the decoded images for a given bit rate.

The combination of the two techniques described above - temporal motion compensated prediction and transform domain coding - can be seen as the key elements of the MPEG coding standards. A third characteristic element of the MPEG algorithms is that these two techniques are processed on small image blocks (of typically 16x16 pels for motion compensation and 8x8 pels for DCT coding). To this reason the MPEG coding algorithms are usually refered to as hybrid block-based DPCM/DCT algorithms.

## MPEG-1 - A Generic Standard for Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s

The video compression technique developed by MPEG-1 covers many applications from interactive systems on CD-ROM to the delivery of video over telecommunications networks. The MPEG-1 video coding standard is thought to be generic. To support the wide range of applications profiles a diversity of input parameters including flexible picture size and frame rate can be specified by the user. MPEG has recommended a constraint parameter set: every MPEG-1 compatible decoder must be able to support at least video source parameters up to TV size: including a minimum number of 720 pixels per line, a minimum number of 576 lines per picture, a minimum frame rate of 30 frames per second and a minimum bit rate of 1.86 Mbits/s. The standard video input consists of a non-interlaced video picture format. It should be noted that by no means the application of MPEG-1 is limited to this constrained parameter set.

The MPEG-1 video algorithm has been developed with respect to the JPEG and H.261 activities. It was seeked to retain a large degree of commonalty with the CCITT H.261 standard so that implementations supporting both standards were plausible. However, MPEG-1 was primarily targeted for multimedia CD-ROM applications, requiring additional functionality supported by both encoder and decoder. Important features provided by MPEG-1 include frame based *random access* of video, *fast forward/fast reverse (FF/FR)* searches through compressed bit streams, *reverse playback* of video and *editability* of the compressed bit stream.

### (A) The Basic MPEG-1 Inter-Frame Coding Scheme

The basic MPEG-1 (as well as the MPEG-2) video compression technique is based on a Macroblock structure, motion compensation and the conditional replenishment of Macroblocks. As outlined in Figure 4a the MPEG-1 coding algorithm encodes the first frame in a video sequence in Intra-frame coding mode (I-picture). Each subsequent frame is coded using Inter-frame prediction (P-pictures) - only data from the nearest previously coded I- or P-frame is used for prediction. The MPEG-1 algorithm processes the frames of a video sequence block-based. Each colour input frame in a video sequence is partitioned into non-overlapping "Macroblocks" as depicted in Figure 4b. Each Macroblock contains blocks of data from both luminance and co-sited chrominance bands - four luminanceblocks ($Y_1$, $Y_2$, $Y_3$, $Y_4$) and two chrominance blocks (U, V), each with size *8 x 8* pels. Thus the sampling ratio between Y:U:V luminance and chrominance pels is 4:1:1.

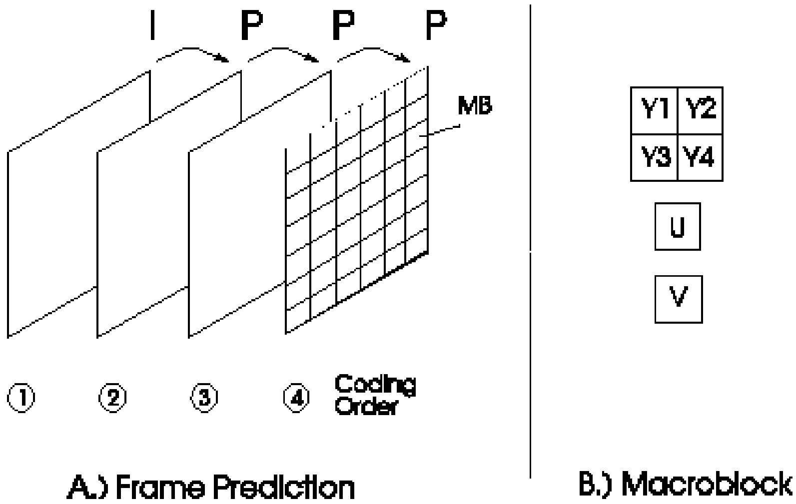



Figure 4: A.) Illustration of I-pictures (I) and P-pictures (P) in a video sequence.

P-pictures are coded using motion compensated prediction based on the nearest previous frame. Each frame is divided into disjoint "Macroblocks" (MB).

B.) With each Macroblock (MB), information related to four luminance blocks (Y1, Y2, Y3, Y4) and two chrominance blocks (U, V) is coded. Each block contains *8x8* pels.

The block diagram of the basic hybrid DPCM/DCT MPEG-1 encoder and decoder structure is depicted in Figure 5. The first frame in a video sequence (I-picture) is encoded in INTRA mode without reference to any past or future frames. At the encoder the DCT is applied to each *8 x 8* luminance and chrominance block and, after output of the DCT, each of the 64 DCT coefficients is uniformly quantized (Q) . The quantizer stepsize (sz) used to quantize the DCT-coefficients within a Macroblock is transmitted to the receiver. After quantization, the lowest DCT coefficient (DC coefficient) is treated differently from the remaining coefficients (AC coefficients). The DC coefficient corresponds to the average intensity of the component block and is encoded using a differential DC prediction method. The non-zero quantizer values of the remaining DCT coefficients and their locations are then "zig-zag" scanned and run-length entropy coded using variable length code (VLC) tables.

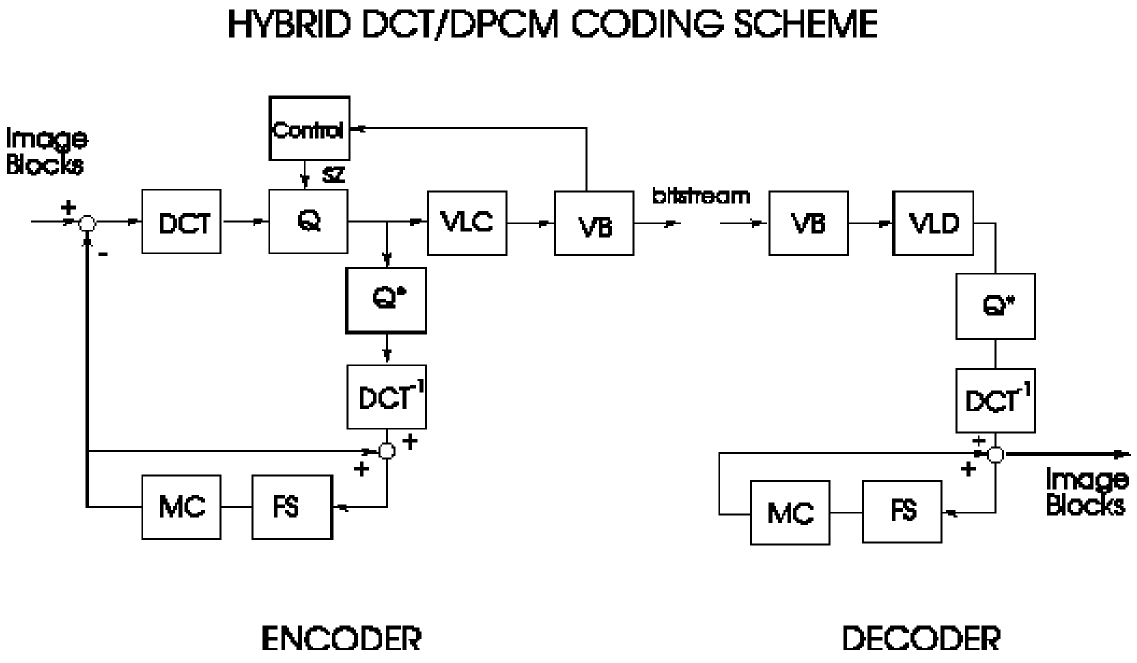



Figure 5: Block diagram of a basic hybrid DCT/DPCM encoder and decoder structure.

The concept of "zig-zag" scanning of the coefficients is outlined in Figure 6. The scanning of the quantized DCT-domain 2-dimensional signal followed by variable-length code-word assignment for the coefficients serves as a mapping of the 2-dimensional image signal into a 1-dimensional bitstream. The non-zero AC coefficient quantizer values (length, ) are detected along the scan line as well as the distance (run) between two consecutive non-zero coefficients. Each consecutive (run, length) pair is encoded by transmitting only one VLC codeword. The purpose of "zig-zag" scanning is to trace the low-frequency DCT coefficients (containing most energy) before tracing the high-frequency coefficients.

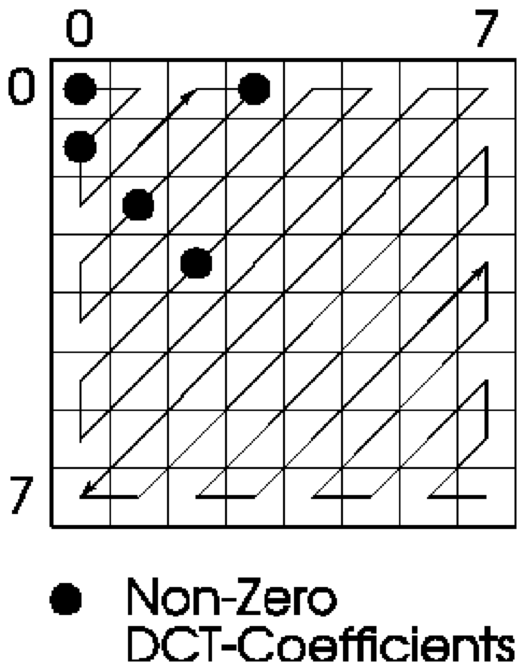



Figure 6: "Zig-zag" scanning of the quantized DCT coefficients in an *8x8* block. Only the non-zero quantized DCT-coefficients are encoded. The possible locations of non-zero DCT-coefficients are indicated in the figure. The zig-zag scan attempts to trace the DCT-coefficients according to their significance. With reference to Figure 3, the lowest DCT-coefficient (*0*,*0*) contains most of the energy within the blocks and the energy is concentrated around the lower DCT-coefficients.

The decoder performs the reverse operations, first extracting and decoding (VLD) the variable length coded words from the bit stream to obtain locations and quantizer values of the non-zero DCT coefficients for each block. With the reconstruction (Q*) of all non-zero DCT coefficients belonging to one block and subsequent inverse DCT (DCT-1) the quantized block pixel values are obtained. By processing the entire bit stream all image blocks are decoded and reconstructed.

For coding P-pictures, the previously I- or P-picture frame *N-1* is stored in a frame store (FS) in both encoder and decoder. Motion compensation (MC) is performed on a Macroblock basis - only one motion vector is estimated between frame *N* and frame *N-1* for a particular Macroblock to be encoded. These motion vectors are coded and transmitted to the receiver. The motion compensated prediction error is calculated by subtracting each pel in a Macroblock with its motion shifted counterpart in the previous frame. A *8x8* DCT is then applied to each of the *8x8* blocks contained in the Macroblock followed by quantization (Q) of the DCT coefficients with subsequent run-length coding and entropy coding (VLC). A video buffer (VB) is needed to ensure that a constant target bit rate output is produced by the encoder. The quantization stepsize (sz) can be adjusted for each Macroblock in a frame to achieve a given target bit rate and to avoid buffer overflow and underflow.

The decoder uses the reverse process to reproduce a Macroblock of frame *N* at the receiver. After decoding the variable length words (VLD) contained in the video decoder buffer (VB) the pixel values of the prediction error are reconstructed (Q*-, and DCT-1-operations). The motion compensated pixels from the previous frame *N-1* contained in the frame store (FS) are added to the prediction error to recover the particular Macroblock of frame *N*.

The advantage of coding video using the motion compensated prediction from the previously reconstructed frame *N-1* in an MPEG coder is illustrated in Figures 7a - 7d for a typical test sequence. Figure 7a depicts a frame at time instance *N* to be coded and Figure 7b the reconstructed frame at instance *N-1* which is stored in the frame store (FS) at both encoder and decoder. The block motion vectors (mv, see also Figure 2) depicted in Figure 7b were estimated by the encoder motion estimation procedure and provide a prediction of the translatory motion displacement of each Macroblock in frame *N* with reference to frame *N-1*. Figure 7b depicts the pure frame difference signal (frame *N* - frame *N-1*) which is obtained if no motion compensated prediction is used in the coding process - thus all motion vectors are assumed to be zero. Figure 7d depicts the motion compensated frame difference signal when the motion vectors in Figure 7b are used for prediction. It is apparent that the residual signal to be coded is greatly reduced using motion compensation if compared to pure frame difference coding in Figure 7c.


**FIGURE 7a**

**FIGURE 7b**

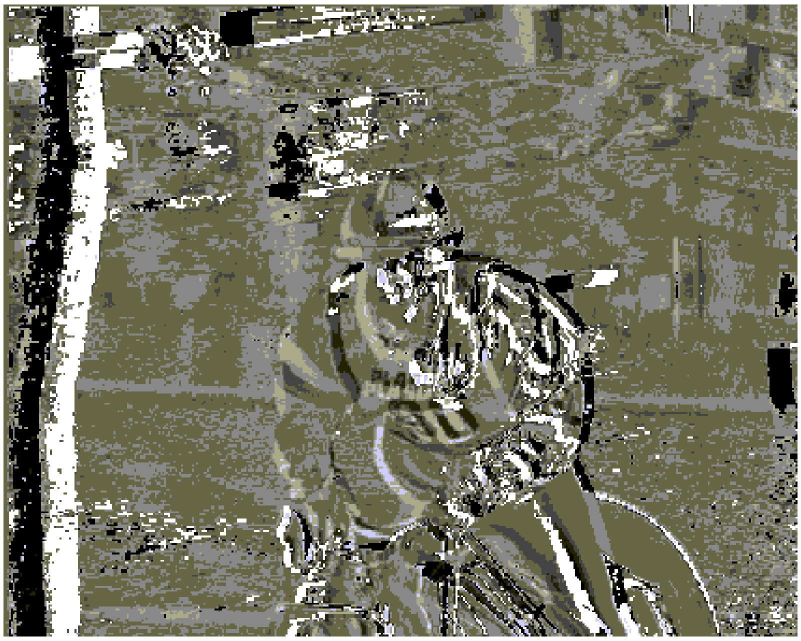


**FIGURE 7c**

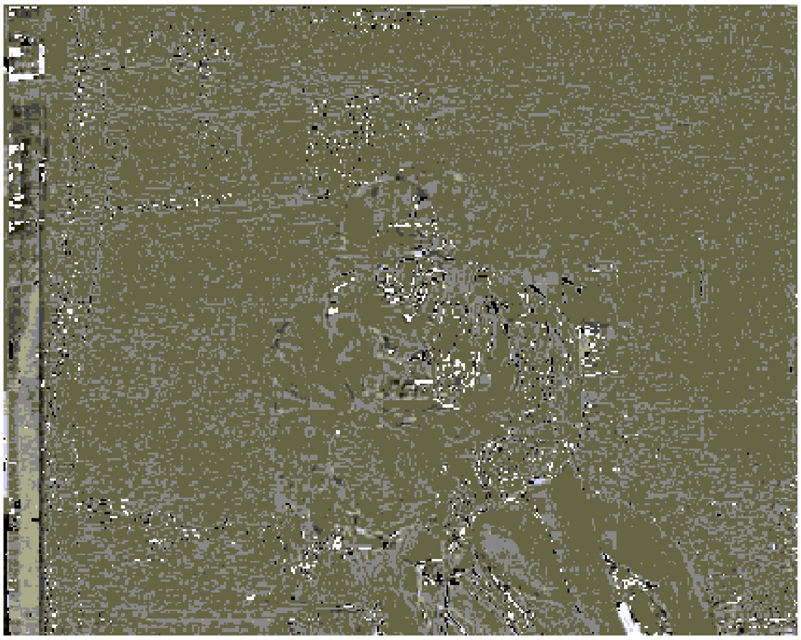


**FIGURE 7d**

Figure 7: (A) Frame at time instance *N* to be coded. (B) Frame at instance *N-1* used for prediction of the content in frame *N* (note that the motion vectors depicted in the image are not part of the reconstructed image stored at the encoder and decoder). (C) Prediction error image obtained without using motion compensation - all motion vectors are assumed to be zero. (D) Prediction error image to be coded if motion compensated prediction is employed.

(B) Conditional Replenishment

An essential feature supported by the MPEG-1 coding algorithm is the possibility to update Macroblock information at the decoder only if needed - if the content of the Macroblock has changed in comparison to the content of the same Macroblock in the previous frame (Conditional Macroblock Replenishment). The key for efficient coding of video sequences at lower bit rates is the selection of appropriate prediction modes to achieve Conditional Replenishment. The MPEG standard distincts mainly between three different Macroblock coding types (MB types):

*skipped MB* - prediction from previous frame with zero motion vector. No information about the Macroblock is coded nor transmitted to the receiver.

*Inter MB* - motion compensated prediction from the previous frame is used. The MB type, the MB address and, if required, the motion vector, the DCT coefficients and quantization stepsize are transmitted.

*Intra MB* - no prediction is used from the previous frame (Intra-frame prediction only). Only the MB type, the MB address and the DCT coefficients and quantization stepsize are transmitted to the receiver.

(C) Specific Storage Media Functionalities

For accessing video from storage media the MPEG-1 video compression algorithm was designed to support important functionalities such as random access and fast forward (FF) and fast reverse (FR) playback functionalities. To incorporate the requirements for storage media and to further explore the significant advantages of motion compensation and motion interpolation, the concept of B-pictures (bi-directional predicted/bi-directional interpolated pictures) was introduced by MPEG-1. This concept is depicted in Figure 8 for a group of consecutive pictures in a video sequence. Three types of pictures are considered: Intra-pictures (I-pictures) are coded without reference to other pictures contained in the video sequence, as already introduced in Figure 4. I-pictures allow access points for random access and FF/FR functionality in the bit stream but achieve only low compression. Inter-frame predicted pictures (P-pictures) are coded with reference to the nearest previously coded I-picture or P-picture, usually incorporating motion compensation to increase coding efficiency. Since P-pictures are usually used as reference for prediction for future or past frames they provide no suitable access points for random access functionality or editability. Bi-directional predicted/interpolated pictures (B-pictures) require both past and future frames as references. To achieve high compression, motion compensation can be employed based on the nearest past and future P-pictures or I-pictures. B-pictures themselves are never used as references.

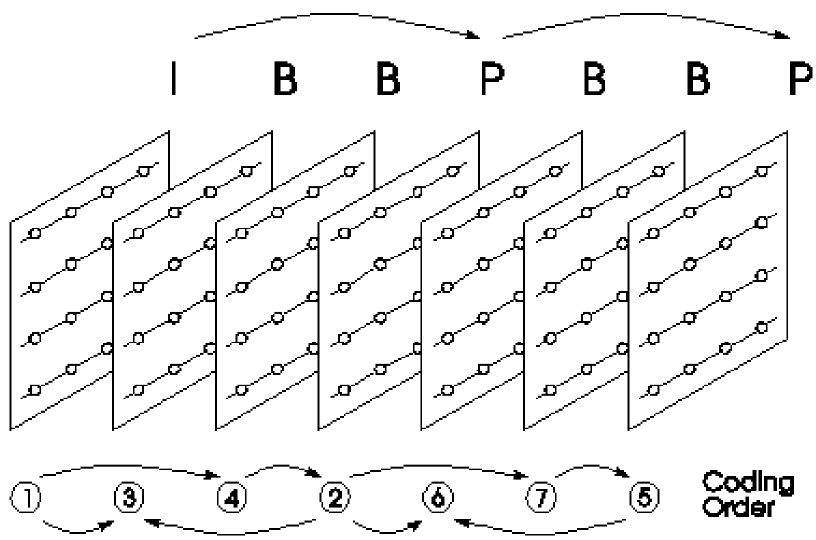



Figure 8: I-pictures (I), P-pictures (P) and B-pictures (B) used in a MPEG-1 video sequence. B-pictures can be coded using motion compensated prediction based on the two nearest already coded frames (either I-picture or P-picture). The arrangement of the picture coding types within the video sequence is flexible to suit the needs of diverse applications. The direction for prediction is indicated in the figure.

The user can arrange the picture types in a video sequence with a high degree of flexibility to suit diverse applications requirements. As a general rule, a video sequence coded using I-pictures only (I I I I I I .....) allows the highest degree of random access, FF/FR and editability, but achieves only low compression. A sequence coded with a regular I-picture update and no B-pictures (i.e I P P P P P P I P P P P ...) achieves moderate compression and a certain degree of random access and FF/FR functionality. Incorporation of all three pictures types, as i.e. depicted in Figure 8 (I B B P B B P B B I B B P ...), may achieve high compression and reasonable random access and FF/FR functionality but also increases the coding delay significantly. This delay may not be tolerable for e.g. videotelephony or videoconferencing applications.

## (D) Rate Control

An important feature supported by the MEPG-1 encoding algorithms is the possibility to tailor the bitrate (and thus the quality of the reconstructed video) to specific applications requirements by adjusting the quantizer stepsize (sz) in Figure 5 for quantizing the DCT-coefficients. Coarse quantization of the DCT-coefficients enables the storage or transmission of video with high compression ratios, but, depending on the level of quantization, may result in significant coding artefacts. The MPEG-1 standard allows the encoder to select different quantizer values for each coded Macroblock - this enables a high degree of flexibility to allocate bits in images where needed to improve image quality. Furthermore it allows the generation of both constant and variable bitrates for storage or real-time transmission of the compressed video.

Compressed video information is inherently variable in nature. This is caused by the, in general, variable content of successive video frames. To store or transmit video at constant bit rate it is therefore necessary to buffer the variable bitstream generated in the encoder in a video buffer (VB) as depicted in Figure 5. The input into the encoder VB is variable over time and the output is a constant bitstream. At the decoder the VB input bitstream is constant and the output used for decoding is variable. MPEG encoders and decoders implement buffers of the same size to avoid reconstruction errors.

A rate control algorithm at the encoder adjusts the quantizer stepsize sz depending on the video content and activity to ensure that the video buffers will never overflow - while at the same time targeting to keep the buffers as full as possible to maximize image quality. In theory overflow of buffers can always be avoided by using a large enough video buffer. However, besides the possibly undesirable costs for the implementation of large buffers, there may be additional disadvantages for applications requiring low-delay between encoder and decoder, such as for the real-time transmission of conversational video. If the encoder bitstream is smoothed using a video buffer to generate a constant bit rate output, a delay is introduced between the encoding process and the time the video can be reconstructed at the decoder. Usually the larger the buffer the larger the delay introduced.

MPEG has defined a minimum video buffer size which needs to be supported by all decoder implementations. This value is identical to the maximum value of the VB size that an encoder can use to generate a bitstream. However, to reduce delay or encoder complexity, it is possible to choose a virtual buffer size value at the encoder smaller than the minimum VB size which needs to be supported by the decoder. This virtual buffer size value is transmitted to the decoder before sending the video bitstream.

The rate control algorithm used to compress video is not part of the MPEG-1 standard and it is thus left to the implementers to develop efficient strategies. It is worth emphasizing that the efficiency of the rate control algorithms selected by manufacturers to compress video at a given bit rate heavily impacts on the visible quality of the video reconstructed at the decoder.

## (E) Coding of Interlaced Video Sources

The standard video input format for MPEG-1 is non-interlaced. However, coding of interlaced colour television with both 525 and 625 lines at 29.97 and 25 frames per second respectively is an important application for the MPEG-1 standard. A suggestion for coding Rec.601 digital colour television signals has been made by MPEG-1 based on the conversion of the interlaced source to a progressive intermediate format. In essence, only one horizontally subsampled field of each interlaced video input frame is encoded, i.e. the subsampled top field. At the receiver the even field is predicted from the decoded and horizontally interpolated odd field for display. The necessary pre-processing steps required prior to encoding and the post-processing required after decoding are described in detail in the Informative Annex of the MPEG-1 International Standard document [MPEG1].

## MPEG-2 Standard for Generic Coding of Moving Pictures and Associated Audio

World-wide MPEG-1 is developing into an important and successful video coding standard with an increasing number of products becoming available on the market. A key factor for this success is the generic structure of the standard supporting a broad range of applications and applications specific parameters. However, MPEG continued its standardization efforts in 1991 with a second phase (MPEG-2) to provide a video coding solution for applications not originally covered or envisaged by the MPEG-1 standard. Specifically, MPEG-2 was given the charter to provide video quality not lower than NTSC/PAL and up to CCIR 601 quality. Emerging applications, such as digital cable TV distribution, networked database services via ATM, digital VTR applications and satellite and terrestrial digital broadcasting distribution, were seen to benefit from the increased quality expected to result from the new MPEG-2 standardization phase. Work was carried out in collaboration with the ITU-T SG 15 Experts Group for ATM Video Coding and in 1994 the MPEG-2 Draft International Standard (which is identical to the ITU-T H.262 recommendation) was released [hal]. The specification of the standard is intended to be generic - hence the standard aims to facilitate the bit stream interchange among different applications, transmission and storage media.

Basically MPEG-2 can be seen as a superset of the MPEG-1 coding standard and was designed to be backward compatible to MPEG-1 - every MPEG-2 compatible decoder can decode a valid MPEG-1 bit stream. Many video coding algorithms were integrated into a single syntax to meet the diverse applications requirements. New coding features were added by MPEG-2 to achieve sufficient functionality and quality, thus prediction modes were developed to support efficient coding of *interlaced video.* In addition *scalable video* coding extensions were introduced to provide additional functionality, such as embedded coding of digital TV and HDTV, and graceful quality degradation in the presence of transmission errors.

However, implementation of the full syntax may not be practical for most applications. MPEG-2 has introduced the concept of "Profiles" and "Levels" to stipulate conformance between equipment not supporting the full implementation. Profiles and Levels provide means for defining subsets of the syntax and thus the decoder capabilities required to decode a particular bit stream. This concept is illustrated in Table II and III.

As a general rule, each Profile defines a new set of algorithms added as a superset to the algorithms in the Profile below. A Level specifies the range of the parameters that are supported by the implementation (i.e. image size, frame rate and bit rates). The MPEG-2 core algorithm at MAIN Profile features non-scalable coding of both progressive and interlaced video sources. It is expected that most MPEG-2 implementations will at least conform to the MAIN Profile at MAIN Level which supports non-scalable coding of digital video with approximately digital TV parameters - a maximum sample density of 720 samples per line and 576 lines per frame, a maximum frame rate of 30 frames per second and a maximum bit rate of 15 Mbit/s.

### (A) MPEG-2 Non-Scalable Coding Modes

The MPEG-2 algorithm defined in the MAIN Profile is a straight forward extension of the MPEG-1 coding scheme to accommodate coding of interlaced video, while retaining the full range of functionality provided by MPEG-1. Identical to the MPEG-1 standard, the MPEG-2 coding algorithm is based on the general Hybrid DCT/DPCM coding scheme as outlined in Figure 5, incorporating a Macroblock structure, motion compensation and coding modes for conditional replenishment of Macroblocks. The concept of I-pictures, P-pictures and B-pictures as introduced in Figure 8 is fully retained in MPEG-2 to achieve efficient motion prediction and to assist random access functionality. Notice, that the algorithm defined with the MPEG-2 SIMPLE Profile is basically identical with the one in the MAIN Profile, except that no B-picture prediction modes are allowed at the encoder. Thus the additional implementation complexity and the additional frame stores necessary for the decoding of B-pictures are not required for MPEG-2 decoders only conforming to the SIMPLE Profile.

*Field and Frame Pictures:* MPEG-2 has introduced the concept of *frame pictures* and *field pictures* along with particular *frame prediction* and *field prediction* modes to accommodate coding of progressive and interlaced video. For interlaced sequences it is assumed that the coder input consists of a series of odd (top) and even (bottom) fields that are separated in time by a field period. Two fields of a frame may be coded separately (field pictures, see Figure 9). In this case each field is separated into adjacent non-overlapping Macroblocks and the DCT is applied on a field basis. Alternatively two fields may be coded together as a frame (frame pictures) similar to conventional coding of progressive video sequences. Here, consecutive lines of top and bottom fields are simply merged to form a frame. Notice, that both frame pictures and field pictures can be used in a single video sequence.

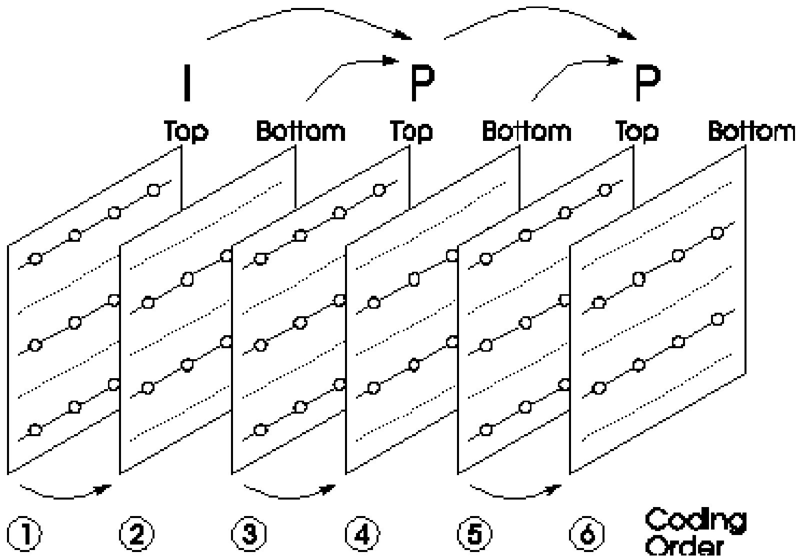



Figure 9: The concept of field-pictures and an example of possible field prediction. The top fields and the bottom fields are coded separately. However, each bottom field is coded using motion compensated Inter-field prediction based on the previously coded top field. The top fields are coded using motion compensated Inter-field prediction based on either the previously coded top field or based on the previously coded bottom field. This concept can be extended to incorporate B-pictures.

*Field and Frame Prediction:* New motion compensated field prediction modes were introduced by MPEG-2 to efficiently encode field pictures and frame pictures. An example of this new concept is illustrated simplified in Figure 9 for an interlaced video sequence, here assumed to contain only three field pictures and no B-pictures. In field prediction, predictions are made independently for each field by using data from one or more previously decoded field, i.e. for a top field a prediction may be obtained from either a previously decoded top field (using motion compensated prediction) or from the previously decoded bottom field belonging to the same picture. Generally the Inter-field prediction from the decoded field in the same picture is prefered if no motion occurs between fields. An indication which reference field is used for prediction is transmitted with the bit stream. Within a field picture all predictions are field predictions.

Frame prediction forms a prediction for a frame picture based on one or more previously decoded frames. In a frame picture either field or frame predictions may be used and the particular prediction mode prefered can be selected on a Macroblock-by-Macroblock basis.

It must be understood, however, that the fields and frames from which predictions are made may have themselves been decoded as either field or frame pictures.

MPEG-2 has introduced new motion compensation modes to efficiently explore temporal redundancies between fields, namely the "Dual Prime" prediction and the motion compensation based on 16x8 blocks. A discussion of these methods is beyond the scope of this paper.

*Chrominance Formats:* MPEG-2 has specified additional Y:U:V luminance and chrominance subsampling ratio formats to assist and enfoster applications with highest video quality requirements. Next to the 4:2:0 format already supported by MPEG-1 the specification of MPEG-2 is extended to 4:2:2 formats suitable for studio video coding applications.

(B) MPEG-2 Scalable Coding Extensions

The scalability tools standardized by MPEG-2 support applications beyond those addressed by the basic MAIN Profile coding algorithm. The intention of scalable coding is to provide interoperability between different services and to flexibly support receivers with different display capabilities. Receivers either not capable or willing to reconstruct the full resolution video can decode subsets of the layered bit stream to display video at lower spatial or temporal resolution or with lower quality. Another important purpose of scalable coding is to provide a layered video bit stream which is amenable for prioritized transmission. The main challenge here is to reliably deliver video signals in the presence of channel errors, such as cell loss in ATM based transmission networks or co-channel interference in terrestrial digital broadcasting.

Flexibly supporting multiple resolutions is of particular interest for interworking between HDTV and Standard Definition Television (SDTV), in which case it is important for the HDTV receiver to be compatible with the SDTV product. Compatibility can be achieved by means of scalable coding of the HDTV source and the wasteful transmission of two independent bit streams to the HDTV and SDTV receivers can be avoided. Other important applications for scalable coding include video database browsing and multiresolution playback of video in multimedia environments.

Figure 10 depicts the general philosophy of a multiscale video coding scheme. Here two layers are provided, each layer supporting video at a different scale, i.e. a multiresolution representation can be achieved by downscaling the input video signal into a lower resolution video (downsampling spatially or temporally). The downscaled version is encoded into a base layer bit stream with reduced bit rate. The upscaled reconstructed base layer video (upsampled spatially or temporally) is used as a prediction for the coding of the original input video signal. The prediction error is encoded into an enhancement layer bit stream. If a receiver is either not capable or willing to display the full quality video, a downscaled video signal can be reconstructed by only decoding the base layer bit stream. It is important to notice, however, that the display of the video at highest resolution with reduced quality is also possible by only decoding the lower bit rate base layer. Thus scalable coding can be used to encode video with a suitable bit rate allocated to each layer in order to meet specific bandwidth requirements of transmission channels or storage media. Browsing through video data bases and transmission of video over heterogeneous networks are applications expected to benefit from this functionality.

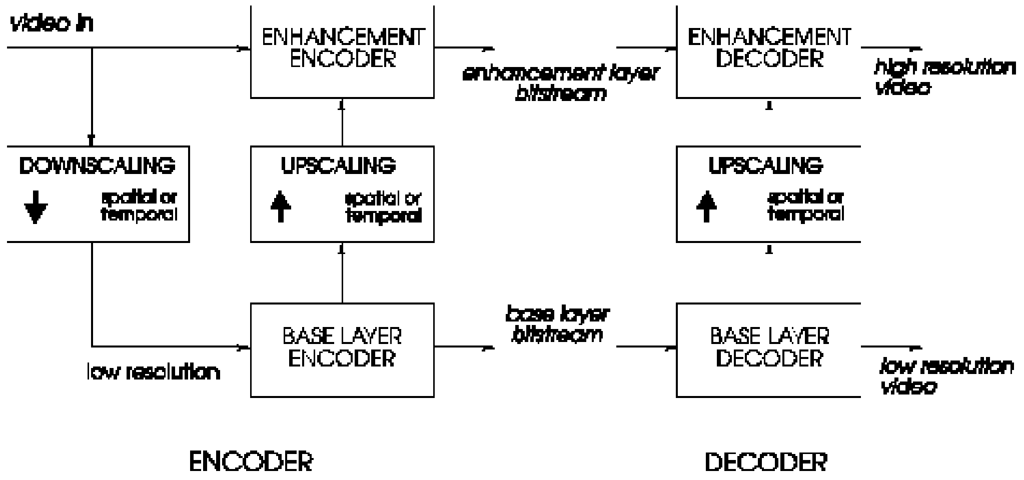



Figure 10: Scalable coding of video.

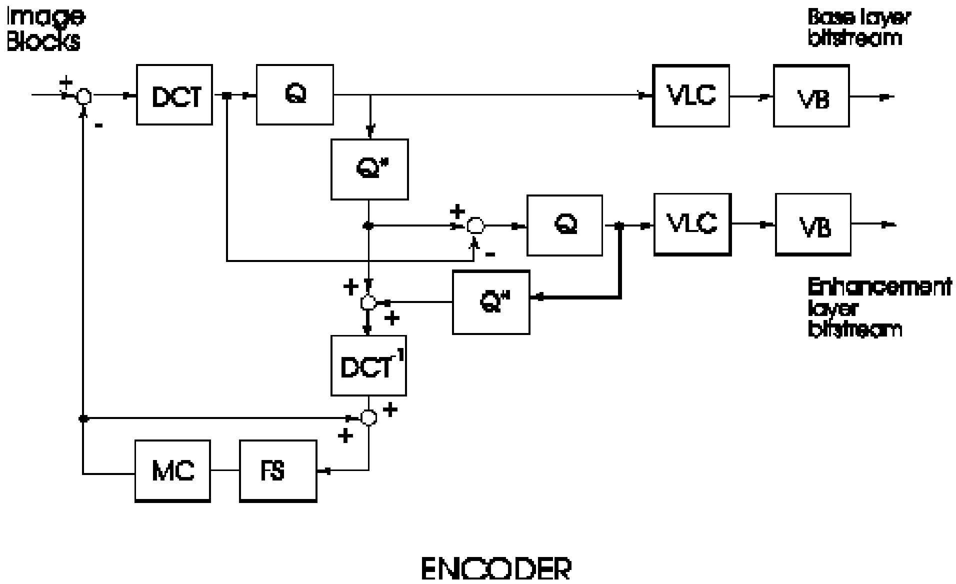
During the MPEG-2 standardization phase it was found impossible to develop one generic scalable coding scheme capable to suit all of the diverse applications requirements envisaged. While some applications are constricted to low implementation complexity, others call for very high coding efficiency. As a consequence MPEG-2 has standardized three scalable coding schemes: SNR (quality) Scalability, Spatial Scalability and Temporal Scalability - each of them targeted to assist applications with particular requirements. The scalability tools provide algorithmic extensions to the non-scalable scheme defined in the MAIN profile. It is possible to combine different scalability tools into a hybrid coding scheme, i.e. interoperability between services with different spatial resolutions *and* frame rates can be supported by means of combining the Spatial Scalability and the Temporal Scalability tool into a hybrid layered coding scheme. Interoperability between HDTV and SDTV services can be provided *along* with a certain resilience to channel errors by combining the Spatial Scalability extensions with the SNR Scalability tool [lam]. The MPEG-2 syntax supports up to three different scalable layers.

*Spatial Scalability* has been developed to support displays with different spatial resolutions at the receiver - lower spatial resolution video can be reconstructed from the base layer. This functionality is useful for many applications including embedded coding for HDTV/TV systems, allowing a migration from a digital TV service to higher spatial resolution HDTV services [MPEG2, lascha]. The algorithm is based on a classical pyramidal approach for progressive image coding [puri, burt]. Spatial Scalability can flexibly support a wide range of spatial resolutions but adds considerable implementation complexity to the MAIN Profile coding scheme.

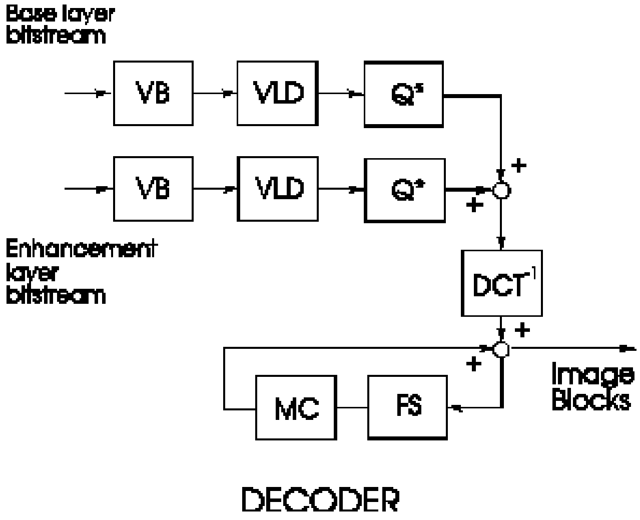
*SNR Scalability:* This tool has been primarily developed to provide graceful degradation (quality scalability) of the video quality in prioritized transmission media. If the base layer can be protected from transmission errors, a version of the video with gracefully reduced quality can be obtained by decoding the base layer signal only. The algorithm used to achieve graceful degradation is based on a frequency (DCT-domain) scalability technique. Both layers in Figure 11 encode the video signal at the same spatial resolution. A detailed outline of a possible implementation of a SNR scalability encoder and decoder is depicted in Figures 11a and 11b. The method is implemented as a simple and straightforward extension to the MAIN Profile MPEG-2 coder and achieves excellent coding efficiency.

At the base layer the DCT coefficients are coarsely quantized and transmitted to achieve moderate image quality at reduced bit rate. The enhancement layer encodes and transmits the difference between the non-quantized DCT-coefficients and the quantized coefficients from the base layer with finer quantization stepsize. At the decoder the highest quality video signal is reconstructed by decoding both the lower and the higher layer bitstreams.

It is also possible to use this method to obtain video with lower spatial resolution at the receiver. If the decoder selects the lowest *NxN* DCT coefficients from the base layer bit stream, non-standard inverse DCT's of size *NxN* can be used to reconstruct the video at reduced spatial resolution [gon, siko2]. However, depending on the encoder and decoder implementations the lowest layer downscaled video may be subject to drift [john].




# FIGURE 11 (A)




# FIGURE 11 (B)

Figure 11: (A) A possible implementation of a two layer encoder for SNR-scalable coding of video. (B) Decoder

The *Temporal Scalability* tool was developed with an aim similar to spatial scalability - steroscopic video can be supported with a layered bit stream suitable for receivers with stereoscopic display capabilities. Layering is achieved by providing a prediction of one of the images of the stereoscopic video (i.e. left view) in the enhancement layer based on coded images from the opposite view transmitted in the base layer.

*Data Partitioning* is intended to assist with error concealment in the presence of transmission or channel errors in ATM, terrestrial broadcast or magnetic recording environments. Because the tool can be entirely used as a post-processing and pre-processing tool to any single layer coding scheme it has not been formally standardized with MPEG-2, but is referenced in the informative Annex of the MPEG-2 DIS document [MPEG2]. The algorithm is, similar to the SNR Scalability tool, based on the separation of DCT-coefficients and is implemented with very low complexity compared to the other scalable coding schemes. To provide error protection, the coded DCT-coefficients in the bit stream are simply separated and transmitted in two layers with different error likelihood.

## References

[ahmed] N.Ahmed, T.Natrajan and K.R.Rao, "Discrete Cosine Transform", IEEE Trans. on Computers, Vol. C-23, No.1, pp. 90-93, December 1984.

[burt] P.J. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code", IEEE Trans. COM, Vol. COM-31, pp. 532-540, 1983

[chen] W. Chen and D. Hein, "Motion Compensated DXC System", in Proceedings of 1986 Picture

Coding Symposium, Vol. 2-4, pp. 76-77, Tokyo, April 1986

[gon] C.Gonzales and E.Viscito, "Flexibly scalable digital video coding", Signal Processing: Image Communication, Vol. 5, No. 1-2, February 1993.

[hal] B.R. Halhed, "Videoconferencing Codecs: Navigating the MAZE", Business Communication Review, Vol. 21, No. 1, pp. 35-40, 1991

[john] A.W.Johnson, T.Sikora, T.K.Tan and K.N.Ngan, "Filters for Drift Reduction in Frequency Scalable Video Coding Schemes", Electronic Letters, Vol. 30, No.6, pp. 471-472**,** 1994

[lam] J. De Lameillieure and R. Schäfer, "MPEG-2 Image Coding for Digital TV", Fernseh und Kino Technik, 48. Jahrgang, pp. 99-107, March 1994 (in German)

[lascha] J.De Lameilieure and G.Schamel, "Hierarchical Coding of TV/HDTV within the German HDTVT Project", Proc. Int. Workshop on HDTV’93, pp. 8A.1.1 - 8A.1.8, Ottowa, Canada, October 1993.

[schaf] R.Schäfer and T.Sikora, "Digital Video Coding Standards and Their Role in Video Communications", Proceedings of the IEEE Vol. 83, pp. 907-923, 1995.

[puri] A.Puri and A.Wong, "Spatial Domain Resolution Scalable Video Coding", Proc. SPIE Visual Communications and Image Processing, Boston, MA, November 1993.

[sik] T.Sikora, "MPEG Digital Video Coding Standards", In Digital Electronics Consumer Handbook, McGraw Hill Company,

Ed. R.Jurgens, to be published 1997.

[siko1] T.Sikora, "The MPEG-1 and MPEG-2 Digital Video Coding Standards", IEEE Signal Processing Magazine, to be published.

[siko2] T.Sikora, T.K.Tan and K.N.Ngan, "A performance comparison of frequency domain pyramid scalable coding schemes", Proc. Picture Coding Symposium, Lausanne, pp. 16.1 - 16.2, March 1993.

[MPEG1] ISO/IEC 11172-2, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s - Video", Geneva, 1993

[MPEG2] ISO/IEC JTC1/SC29/WG11 N0702 Rev, "Information Technology - Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262", Draft International Standard, Paris, 25 March 1994.

---

Tables

---