

# Digital Video Coding Standards and Their Role in Video Communications

Thomas Sikora

*Heinrich-Hertz-Institute (HHI)*  
*10587 Berlin, Einsteinufer 37, Germany*

**ABSTRACT.** Efficient digital representation of image and video signals has been the subject of considerable research over the past twenty years. With the growing availability of digital transmission links, progress in signal processing, VLSI technology and image compression research, visual communications has become more feasible than ever. Digital video coding technology has developed into a mature field and a diversity of products has been developed—targeted for a wide range of emerging applications, such as video on demand, digital TV/HDTV broadcasting, and multimedia image/video database services. With the increased commercial interest in video communications the need for international image and video coding standards arose. Standardization of video coding algorithms holds the promise of large markets for video communication equipment. Interoperability of implementations from different vendors enables the consumer to access video from a wider range of services and VLSI implementations of coding algorithms conforming to international standards can be manufactured at considerably reduced costs. The purpose of this chapter is to provide an overview of today's image and video coding standards and their role in video communications. The different coding algorithms developed for each standard are reviewed and the commonalities between the standards are discussed.

## 1. Introduction

It took more than 20 years for digital image and video coding techniques to develop from a more or less purely academic research area into a highly commercial business. Modern data compression techniques today offer the possibility to store or transmit the vast amount of data necessary to represent digital images and video in an efficient and robust way. New audio visual applications in the field of communication, multimedia and broadcasting became possible based on digital video coding technology. The importance of these techniques will become even more important in a world where productivity gains through communications will depend on the flexibility, mobility and interoperability of communication equipment—where everybody will be able to communicate with anybody at any place and at any hour.

As manifold as applications for image coding are today, as manifold are the different approaches and algorithms and were the first hardware implementations and even systems in the commercial field, such as private teleconferencing systems [1, 2]. However, with the advances in VLSI-technology it became possible to open more application fields to a larger number of users and therefore the necessity for standards arose, because exchange of compressed video data on national and international bases is mandatory [32]. From the beginning of the 1980's on, standardization activities started within CCITT, followed by CCIR and ISO later on. The outcome of these activities are CCITT Recommendations H.120 and H.261, CCIR Recommendations 721 and 723, ISO 10918 (JPEG) and ISO 11172 (MPEG-1). ISO 13818 (MPEG-2) has just been drafted and ISO MPEG-4 is in its development phase.

This international standardization allows for large scale production of VLSI systems and devices, thus making the products cheaper and therefore more affordable for a wide field

of applications. Most important, it allows for international video data exchange via storage media (*e.g.*, CD-ROM) or via communication networks (*e.g.*, ISDN).

The purpose of this chapter is to provide an overview of today's image and video coding algorithms and standards and their role in video communications. The chapter is organized as follows: In Section 2 we provide an outline of the chronological development of video coding standards and specific target applications. Section 3 reviews the principles of image and video compression and algorithms relevant to the standards coding schemes. In Section 4 commonalities and differences among the diverse coding algorithms are outlined. Furthermore the specific properties of the standards related to their applications are presented. In Section 5 we discuss the performance of the standards and their success in the market place.

## 2. International standardization

International standardization requires collaboration between regions and countries with different infrastructures, different technical background and with different political and commercial interests. Furthermore, since the aim of standardization is to foster implementations of image and video coding equipment, the capability of current state-of-the-art technology needs to be taken into account. Therefore, international standards do not necessarily represent the best technical solutions, but rather attempt to achieve a compromise between the amount of flexibility supported by the standard, the implementation complexity required and compression efficiency achieved.

Although there exist slight differences among the different standardization bodies as far as standardization procedures are concerned, the main steps towards the finalization of a standard can roughly be described as shown in Figure 1. In the first phase the requirements for a specific application or for a field of applications are identified [3]. Next different algorithms are developed by different laboratories and compared [4, 5]. As a result of this comparison a single basic technique will be identified which is then refined in a joint effort during the collaborative phase. At the end of this phase a draft standard will be issued, which has to be validated by compliance testing based on computer simulations or hardware tests and field trials [6, 7]. After successful validation and eventual refinements the final standard is issued.

Study Group (SG) XV of CCITT was the first international committee starting standardization in video coding during its study period 1980-1984 [8]. In 1984 it issued Recommendation H.120 targeted for videoconferencing applications at the digital primary rates of 2.048 and 1.544 Mbit/s for 625/50 and 525/60 TV systems respectively (H1 channel as defined by CCITT Rec. G 702) [9]. This standard consists of 3 parts: Part 1 is intended for regional use in countries supporting systems with 625 lines—50 Hz—2 Mbit/s, Part 2 for international use (625 lines—50 Hz and 525 lines—60 Hz) and Part 3 for regional use in 525 lines—60 Hz—1.5 Mbit/s systems. Unfortunately the algorithms for Part 1 and Part 3 are not identical, although both schemes use essentially the same compression method (temporal DPCM, see Section 3). Therefore the target for a world-wide standard had not really been achieved. Maybe this is one of the reasons why H.120 never became a commercial success, although regular services were established by a few Telecoms [10, 11].

With the advances in image compression research and implementation technology it became possible to consider coding of video at lower data rates. Therefore Study Group XV agreed at the end of 1984 to define a world-wide standard for videophone and videoconferencing applications, targeted for sub-primary rates ( $\leq 2$  Mbit/s) and suitable for 625 lines—50 Hz and 525 lines—60 Hz systems respectively. In 1989 the draft Recommendation H.261 for a codec based on  $p \times p$  kbit/s ( $p = 1 \dots 30$ ) was available. The hybrid DCT/DPCM compression algorithm applied in the Rec. H.261 codec (using Discrete Cosine Transform (DCT) techniques, temporal DPCM and motion compensation, see Section 3 and also [12]), has become a key element for most other image and video coding standards specified later.



FIGURE 1. Steps in international standardization

In parallel to the standardization activities in CCITT, and in co-operation, three other international bodies started standardization of video coding algorithms for compression of digital TV signals with contribution quality, namely CCIR, CMTT [13] and ISO [5]. Within CCIR, SG 11 is responsible for the standardization of video coding, whereas CMTT is responsible for the transmission of TV signals. In order to co-ordinate the work within the different committees and their subgroups, Interim Working Parties (IWP) have been established and it was the task of IWP CMTT/2 to produce Recommendations 721 and 723.

CCIR Rec. 721, issued in 1990, specifies the coding of CCIR Rec. 601 TV-signals at 140 Mbit/s with contribution quality for transmission over H4—channels. Simple spatial DPCM is used for video compression (see Section 3), to assist very simple codec implementations and to achieve a video quality suitable for post-production purposes [14].

CCIR Rec. 723, issued in 1989, standardizes a codec targeted for contribution of CCIR Rec. 601 TV-signals at data rates between 30 and 45 Mbit/s, suitable for transmission over H3 channels. This codec employs a hybrid DCT/DPCM technique similar to the H.261 algorithm, but optimized for the higher data rates envisaged [15]. It is worthwhile to mention that CCIR 723 codecs have been used to code HDTV at rates of 140 Mbit/s and below by using 4 to 6 TV-codecs in parallel, each of them working on vertical [16] or horizontal [17, 18] stripes of the HDTV picture.

Working Group (WG) 8 of Subcommittee (SC) 2 of ISO started standardization activities in 1982, targeted primarily for coding of continuous tone still images. In 1986 members of ISO/SC2/WG8 and CCITT SG VIII joined their activities and formed the Joint Photographic Experts Group—JPEG. This group issued the ISO 10918 draft international standard (DIS) in 1991 and an international standard (IS) in 1992. The core image coding algorithm employs a spatial DCT compression scheme. JPEG provides different modes of operation, *i.e.*, sequential, progressive, hierarchical and lossless modes, the latter mode using spatial DPCM coding instead of DCT [19–21].

In 1988 the Moving Picture Experts Group (MPEG) was founded under ISO/SC2 with the charter to standardize a video coding algorithm targeted for digital storage media and bit rates at up to about 1.5 Mbits/s [22]. Its official denotation is now ISO/IEC/JTC1/SC29/WG11. The first DIS released by the committee, ISO 11172 (MPEG-1), was drafted in 1991 and finally issued as IS in 1992. In contrast to the standards mentioned above, MPEG-1 is intended to be generic (although the initial target applications envisaged and applications parameters defined were constrained to digital storage media). Generic means that the standard is independent of a particular application and therefore comprises mainly a toolbox. It is up to the user to decide which tools to select to suit the particular applications envisaged. This implies

that only the coding syntax is defined and therefore mainly the decoding scheme is standardized [23]. MPEG-1 defines a hybrid DCT/DPCM coding scheme with motion compensation similar to the H.261 and CCIR Rec. 723 coding standards. Further refinements in prediction and subsequent processing were introduced to provide the functionality required for random access in digital storage media.

Studies on MPEG-2 started in 1990 with the initial target to issue a standard for coding of TV-pictures with CCIR Rec. 601 resolution at data rates below 10 Mbit/s. In 1992 the scope of MPEG-2 was enlarged to suit coding of HDTV—thus making an initially planned MPEG-3 phase superfluous. The DIS for MPEG-2 video was issued in early 1994.

The video coding scheme used in MPEG-2 is again generic and similar to that of MPEG-1, however with further refinements and special consideration of interlaced sources. Furthermore, many functionalities such as “scalability” were introduced. In order to keep implementation complexity low for products not requiring the full video input formats supported by the standard (e.g., SIF to HDTV resolutions), so called “Profiles”, describing functionalities, and “Levels”, describing resolutions, were introduced to provide separate MPEG-2 conformance levels [25, 26].

MPEG-4 started activities in 1993 and, in 1998, was still in its definition phase. The target is to standardize video coding schemes for data rates below 64 Kbit/s, suitable for video communication over PSTNs or 2nd generation mobile networks. There was no decision yet whether MPEG-4 will again use a hybrid DCT optimized for very low bit rates [27] or object or model-based approaches [28], which potentially offer higher compression ratios but are not yet mature. Issuing of a draft standard is planned for 1996.

### 3. Video Coding 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 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 sought. In fact, all video coding techniques standardized in the past were developed and optimized with due consideration of the capabilities of “state of the art” (VLSI) technology.

Dependent on the applications requirements we may envisage “lossless” and “lossy” coding of 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 *decoded* image quality is required to be identical to the image quality prior to *encoding*. Important applications for lossless coding techniques are the storage of video or film in the studio environment, primary distribution of video or film and the transmission or storage of highest quality still images, such as satellite imaging and medical imaging.

In contrast, the aim of “lossy” coding techniques 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 *decoded* 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. The ultimate aim of lossy

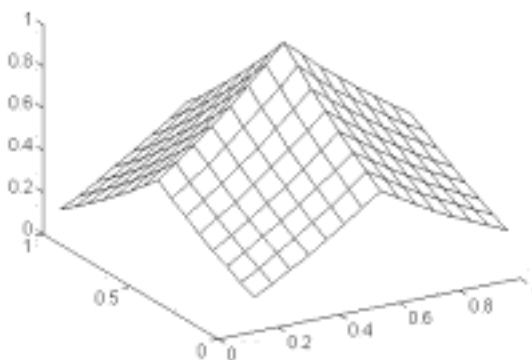


FIGURE 2. Spatial inter-element correlation of “typical” images as calculated using a Gauss Markov image model with high pel-pel correlation. Variables  $x$  and  $y$  describe the distance between pels in horizontal and vertical image dimensions respectively.

coding techniques is to optimize image quality for a given target bit rate subject to “objective” or “subjective” optimization criteria. It should be noted that the degree of image degradation (both the objective degradation as well as the amount of visible artifacts) 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 artifacts may be achieved even with simple compression techniques.

### 3.1. Source Model

All digital image and video coding techniques standardized so far, and which are relevant in the context of this chapter, are statistical in nature. Video sequences usually contain statistical redundancies in both temporal and spatial directions. The basic statistical property upon which image compression techniques rely is inter-element correlation, including the assumption of simple correlated translatory motion between consecutive frames. Thus, 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 and motion estimation). Intuitively, it is clear that in some circumstances, *i.e.*, during scene changes of a video sequence, the temporal correlation between nearby frames is small or even vanishes. In this case Intra-frame coding techniques are appropriate to explore spatial correlation to achieve efficient data compression (see spatial predictive coding and Transform coding). However, if the correlation between nearby frames is high, *i.e.*, in cases where two consecutive frames have similar or identical content, it is desirable to use Inter-frame coding techniques employing temporal prediction (see predictive coding and Transform coding). In practical video coding schemes an adaptive combination of both methods is used to achieve high data compression (see Hybrid Transform/DPCM coding).

Figure 2 depicts an example of Intra-frame pel-to-pel correlation properties of images, here modeled using a rather simple but nevertheless valuable statistical model. The simple model assumption already inherits basic correlation properties of many “typical” images, namely the high correlation between adjacent pixels and the monotonic decay of correlation with increased distance between pels. We will use this model assumption later to demonstrate some of the properties of Transform coding.

### 3.2. Subsampling and Interpolation

Almost all video coding techniques described in the context of this chapter 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. This concept is also used to explore subjective redundancies contained in chrominance data, *i.e.*, the human eye is more sensitive to changes in brightness than to chromaticity changes. Therefore, many of the standardized 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).

### 3.3. Entropy Coding

The pixel color values of digital video frames are usually pre-quantized to fixed-length words with typically 8 bits or 10 bits accuracy per color component. However, for most video material it can be assumed that not all quantized color values occur equally likely within the video scenes. We can reduce the average number of bits per word if color values having lower probability are assigned longer code words, whereas values having higher probability are assigned shorter code words. This method is called variable word-length coding or entropy coding and forms one of the most basic elements of today's video coding standards, especially in combination with transform domain or predictive coding techniques (see sections on Predictive Coding and Transform Domain Coding). If the resulting code words are concatenated to form a stream of binary digits (bits), then correct decoding by a receiver is possible if the code words are uniquely decipherable. Of many possibilities to construct sets of uniquely decipherable codewords, the *Huffman-code* has found widespread application and is used in all standards coding schemes treated in the course of this chapter [31]. The concept of entropy coding can be combined with a *run-length coding* procedure to achieve further data reduction. This method is useful if consecutive pels along a scan line are highly correlated. With run-length coding one codeword is allocated to a pair of input values (run, length) [31, 44], *i.e.*, the number (run) of consecutive pels along a scan line with equal color values (length) can be encoded by transmitting only one codeword.

### 3.4. Predictive Coding

With predictive coding the redundancy in video is determined from the neighboring pels within frames or between frames. This concept is promising if the correlation between adjacent pels that are spatially, as well as temporally, close to each other is strong. In basic predictive coding systems, an approximate prediction of the pel to be coded is made from previously coded information that has been transmitted. The difference between the actual pel and the prediction (prediction error) is usually quantized and entropy coded. This is the well known differential pulse code modulation (DPCM) technique. Predictive methods can be combined with a run-length coding procedure. Only nonzero DPCM pel color values are encoded along with the number (run) of zero pixel values along the scan line. Notice that without quantization of the prediction error lossless compression of images and video can be achieved at relatively low implementation complexity with moderate compression results.

The performance of this method is strongly dependent on the actual predictors used to de-correlate the data. Figure 3 outlines possible locations of pels used for Intra-frame and Inter-field prediction in some of the standards coding schemes discussed in the next section.

For Inter-frame predictive coding the motion compensated prediction (see section below) has found wide application in standards video coding schemes. For pure Intra-frame coding

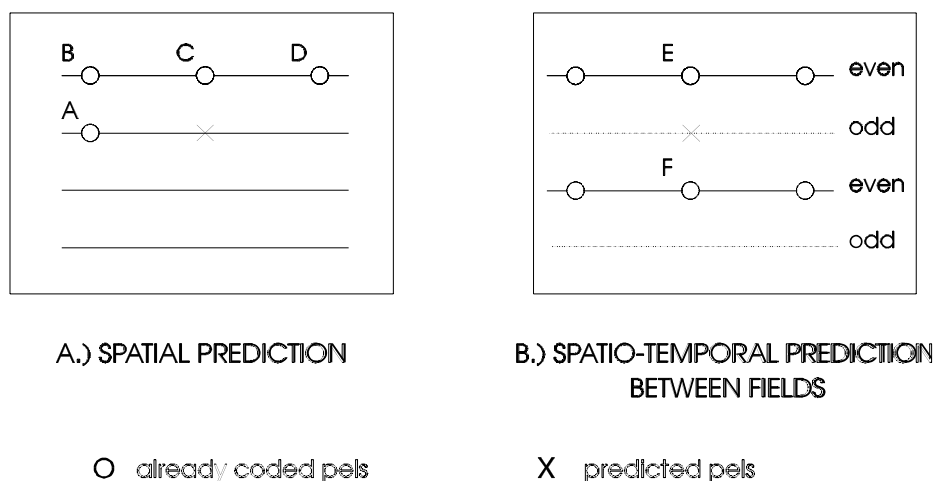


FIGURE 3. Location of predicted pel ( $x$ ) and already coded pels ( $A, B, C, D$  or  $E, F$ ) used for prediction in two separate scenarios. The prediction is calculated as a weighted linear combination of the previously coded pels. A.) predictor geometry used typically with the JPEG lossless coding mode (Intraframe coding). B.) Prediction between fields for coding of digital interlaced TV—pels in odd fields are predicted based on pels already coded in even fields (Inter-field coding).

Transform methods usually outperform DPCM coding techniques. Transform coding results in better video quality for the same bit rate and is used by most video coding standards techniques for lossy coding. However, the JPEG lossless mode compression algorithm for lossless coding of still images uses Intra-frame DPCM due to its low implementation complexity.

### 3.5. Motion Compensation

Motion compensated prediction is a powerful tool to reduce temporal redundancies between frames and is thus used extensively in video coding standards (*i.e.*, H.261, CCIR 723, MPEG-1 and MPEG-2) 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 approximately by a limited number of motion parameters (*i.e.*, estimated motion vectors). 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. Encoding one motion information with each coded 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.*,  $16 \times 16$  pels) and only one motion vector is estimated and coded for each of these blocks (Figure 4).

The motion compensated DPCM technique (used in combination with Transform coding, see section on Hybrid DCT/DPCM coding schemes) has proven to be highly efficient and robust for video data compression and has become a key element for the success of today’s “state-of-the-art” coding standards.

### 3.6. 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 picture 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 (*i.e.*, of size  $N \times N$  pels). The transformation can be represented as a matrix

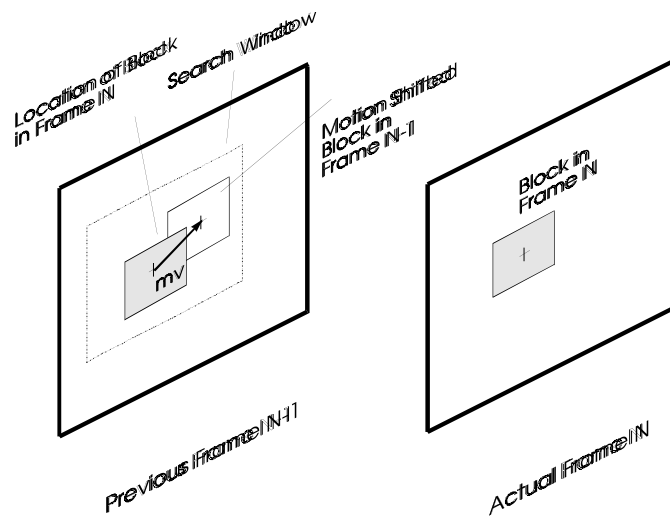


FIGURE 4. Block matching approach for motion compensation: One motion vector ( $mv$ ) is estimated for each block in the actual frame  $N$  to be coded. The motion vector points to a reference block of same size in a previously coded frame  $N - 1$ . The motion compensated prediction error is calculated by subtracting each pel in a block with its motion shifted counterpart in the reference block of the previous frame.

operation using an  $N \times N$  Transform matrix  $A$  to obtain the  $N \times N$  transform coefficients  $c$  based on a linear, separable and unitary *forward* transformation

$$c = AbA^T.$$

Here,  $A^T$  denotes the transpose of the transformation matrix  $A$ . Note that the transformation is reversible, since the original  $N \times N$  block of pels  $b$  can be reconstructed using a linear and separable *inverse* transformation<sup>1</sup>

$$b = A^T cA.$$

Of many possible alternatives the Discrete Cosine Transform (DCT) has become the most successful transform for still image and video coding [35]. 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 5 presents the variance (energy) of Intra-frame DCT coefficients based on the simple statistical model assumption already discussed in Figure 2. Here, the variance for each coefficient represents the variability of each particular 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 depicted in Figure 5, 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.

<sup>1</sup>For a unitary transform the inverse matrix  $A^{-1}$  is identical with the transposed matrix  $A^T$ , that is  $A^{-1} = A^T$ .



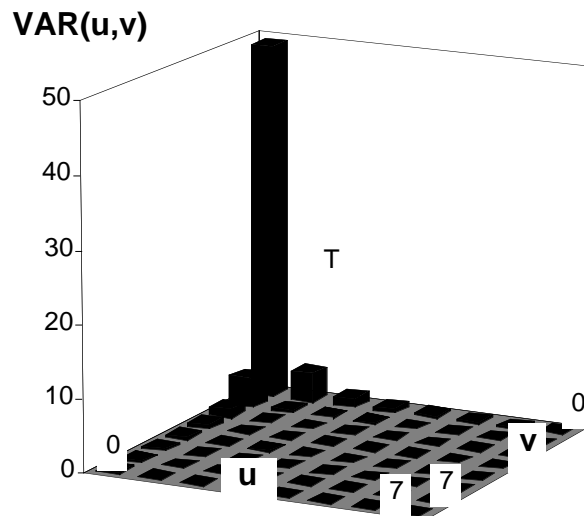


FIGURE 5. The figure depicts the variance distribution of DCT-coefficients “typically” calculated as average over a large number of image blocks. The variance of the DCT coefficients was calculated based on the statistical model used in Figure 2.  $u$  and  $v$  describe the horizontal and vertical image transform domain variables within the  $8 \times 8$  block. Most of the total variance is concentrated around the DC DCT-coefficient ( $u = 0, v = 0$ ).

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.

The DCT is closely related to the 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 many of the 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 desirable to improve the visual quality of the decoded images for a given bit rate.

Usually the DCT-coefficients are quantized prior to coding. Depending on the quantization step-size this will result in a significant number of zero valued coefficients. In all coding standards a run-length coding procedure is employed to efficiently encode the DCT-coefficients which remain nonzero after quantization [24]. To this aim the nonzero coefficients are detected along a scan line and one code word is encoded for each (run, level)-pair. The level indicates the quantization level of the particular nonzero coefficient to be encoded and the run indicates the distance to the previously encoded nonzero coefficient along the scan line [36].

### 3.7. Hybrid DCT/DPCM Coding

A common approach is to combine a temporal DPCM technique with the DCT method into a hybrid coding scheme to efficiently explore spatial as well as temporal redundancies in video scenes. Temporal correlation is reduced first from the video scene, possibly using motion compensated prediction, and in a second step the DCT method is applied to the DPCM prediction error images to explore the remaining spatial redundancies. Finally, the DCT coefficients are quantized and entropy coded. Hybrid DCT/DPCM coding has become the

core method for all recent *video* coding standards. We will discuss this approach in Section 4 in more detail.

## 4. Standards for Image and Video Coding Applications

### 4.1. JPEG—A Standard for Continuous-Tone Still Image Coding

The aim of JPEG (Joint Photographics Experts Group) was to develop a generic image coding standard for coding of continuous-tone still images (greyscale and color)—thus to support a wide range of applications for storage and transmission of digital images. To meet the requirements of the different applications, the JPEG standard specifies both baseline and extended systems for different modes of operation—DCT-based methods for “lossy” coding of images and a predictive method for applications requiring “lossless” compression. The JPEG DCT-based “baseline” method for lossy coding has been the most widely used and implemented method to date. A product conforming to the JPEG standard should, at a minimum, meet the requirements of the baseline system. Notice that, although the JPEG coding standard has been developed for still image compression, it can be used to compress video sequences. This is commonly referred to as MOTION JPEG.

A number of parameters related to the source image and the coding process can be customized to meet the requirements of particular applications. The JPEG standard does not specify or encode any information on pixel aspect ratio or color space. However, any JPEG coding mode supports source images with a size between  $(1 \times 1)$  and  $(65,535 \times 65,535)$  active pixel elements, and each pixel may have between 1 and 255 color components or spectral bands. Pixels can be represented with a precision of 8 or 12 bits for DCT-based coding modes and with a precision between 2 to 16 bits for the predictive lossless coding mode. An important aspect is the possibility to control the quality of the decoded images, or in this respect the number of bits needed to compress the source images, by adjusting quantizer parameters to the specific applications requirements.

**4.1.1. JPEG Baseline Coding Mode.** The block diagram of the JPEG baseline lossy compression mode encoder and decoder is depicted in Figure 6. Here, to simplify the presentation, the figure presents the baseline mode as a single-color component image compression scheme (*i.e.*, greyscale). Color image compression can be approximately regarded as independent processing of multiple greyscale images. At the encoder each image (or each frame of a video sequence) is subdivided into non-overlapping blocks of  $8 \times 8$  pels and the DCT is then applied to each block. After output of the DCT, each of the 64 DCT coefficients is uniformly quantized (Q) with a quantizer value specific to the particular coefficient, to perform perceptual weighting according to human visual criteria. 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<sup>2</sup>. 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.

The concept of “zig-zag” scanning for ordering the coefficients prior to run-length entropy coding using VLC tables is outlined in Figure 7. The non-zero AC coefficient quantizer values (length,  $\alpha$ ) 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

<sup>2</sup>Because there is usually strong correlation between the DC values of adjacent  $8 \times 8$  blocks, the quantized DC coefficient is encoded as the difference between the DC value of the previous block and the actual DC value.

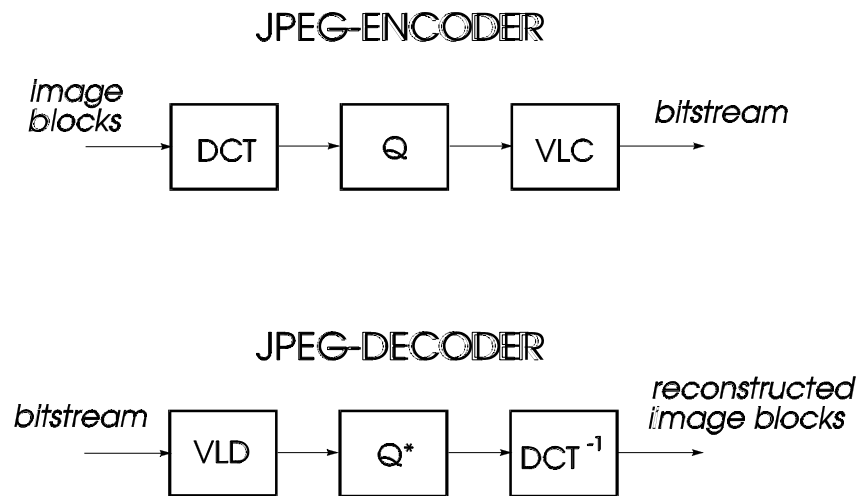


FIGURE 6. JPEG Baseline encoder and decoder structure.

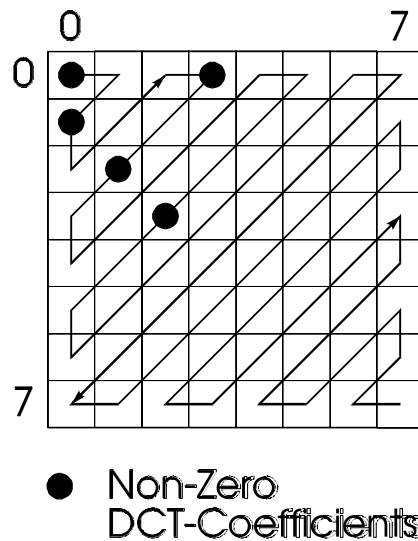


FIGURE 7. “Zig-zag” scanning of the quantized DCT coefficients in an  $8 \times 8$  block. Only the non-zero quantized DCT-coefficients are encoded. The possible locations of non-zero DCT-coefficients are indicated in the figure. The zig-zag scan attempts to trace the DCT-coefficients according to their significance. With reference to Figure 5, the lowest DCT-coefficient  $(0, 0)$  contains most of the energy within the blocks and the energy is concentrated around the lower DCT-coefficients.

low-frequency DCT coefficients (containing most energy) before tracing the high-frequency coefficients<sup>3</sup>.

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 block pixel values are obtained. By processing the entire bit stream all image blocks are decoded and reconstructed.

<sup>3</sup>The location of each non-zero coefficient along the zig-zag scan is encoded relative to the location of the previous coded coefficient. The zig-zag scan philosophy attempts to trace the non-zero coefficients according to their likelihood of appearance to achieve an efficient entropy coding. With reference to Figure 5 the DCT coefficients most likely to appear are concentrated around the DC coefficient with decreasing importance. For many images the coefficients are traced efficiently using the zig-zag scan.

Selection Value	Prediction
0	no prediction
1	$A$
2	$C$
3	$B$
4	$A + B - C$
5	$A + (C - B)/2$
6	$C + (A - B)/2$
7	$(A + C)/2$

TABLE 1. Predictors used for JPEG Intra-Frame Prediction

To allow user controlled quantization, JPEG introduces the concept of a 64-element “Quantization Table”. The values of this Quantization Table can be defined specific to particular applications or image statistics – quantization is performed by dividing each DCT coefficient by its corresponding Quantization Table value and rounding to the nearest integer. This important concept allows the user to tailor the total bits generated by the encoder, and the quality of the decoded images, to the needs of the particular application.

**4.1.2. JPEG Lossless Coding Mode.** Lossless compression of images or video is an important aspect of the JPEG coding standard to ensure exact recovery of every source image pixel. JPEG has chosen a simple predictive method, as already outlined in Section 3, to achieve lossless coding of images. This method is wholly independent of the DCT method described above. The main processing steps for lossless encoding of images consist of prediction and entropy coding, combined with a run-length coding method. Each input source image line is scanned from left to right and a predictor according to Figure 3 (A) combines the values of up to three neighboring pixels ( $A$ ,  $B$  and  $C$ ). One of seven possible predictors listed in Table 1 can be selected to adapt the prediction to the statistics of the actual images. The prediction error is then entropy coded. Although the method employed with the lossless JPEG coders is rather simple to implement, typically a compression ratio of around 2:1 can be achieved for color images with moderately complex scenes.

**4.1.3. JPEG Extended Modes.** The DCT-based “baseline” method for lossy coding and the simple predictive method to achieve lossless compression are the two most widely used JPEG algorithms for coding of still images and video. However, JPEG has standardized a large variety of tools comprising a “tool-box” to suit the different needs of diverse applications. Of the many different techniques standardized two layered coding schemes are worth mentioning:

*Progressive DCT-based method.* This method provides the possibility to progressively build up image quality at the receiver. This technique is suitable for, *e.g.*, transmission of images over low bit rate channels, where the receiver can obtain lower quality images quickly by decoding only parts of the total bit stream. The method is similar to the baseline method but makes use of the interpolation property of the DCT. For each DCT block the quantized DCT coefficient values are partially encoded in multiple scans. Only low frequency DCT coefficients are decoded for each block to reconstruct the images at reduced quality.

*Hierarchical method.* This layered coding technique enables the user to decode images at reduced spatial resolution by extracting and decoding only parts of the bit stream. The method is useful if the receiver is either not interested in or not capable of displaying the images at original spatial resolution. Possible applications are the decoding of very high resolution images for display on smaller resolution terminals. The method is based on ‘pyramidal’ encoding of images in which each encoding stage has a source image of different spatial resolution encoded into a layer [43]. At the lowest resolution layer low spatial resolution

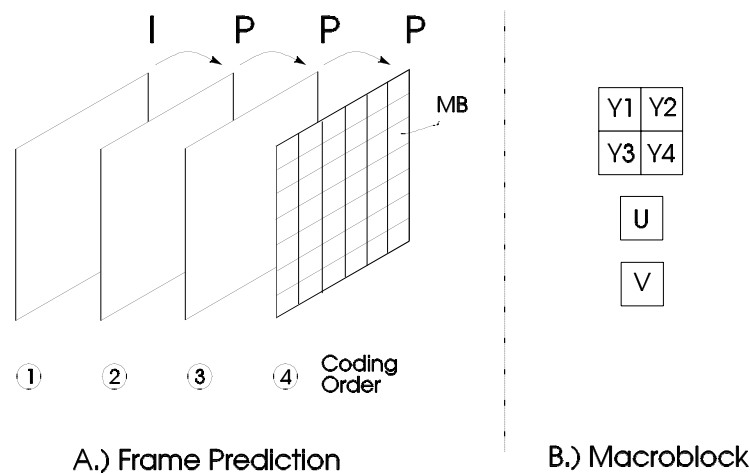


FIGURE 8. A.) Illustration of I-pictures (I) and P-pictures (P) in a video sequence. P-pictures are coded using motion compensated prediction based on the nearest previous frame. Each frame is divided into disjoint “Macroblocks” (MB). B.) With each Macroblock (MB), information related to four luminance blocks (Y1, Y2, Y3, Y4) and two chrominance blocks (U, V) is coded. Each block contains  $8 \times 8$  pels.

versions of the original images are encoded. The higher resolution layers encode refinement information.

#### 4.2. H.261: A Video Coding Standard for ISDN Visual Telephony Applications

The CCITT “Specialist Group on Coding for Visual Telephony” was given the task to standardize a video coding algorithm to support audio-visual teleservices on ISDN, including videophone and videoconferencing applications. The resulting Recommendation H.261 video coding algorithm was designed and optimized for low target bit rate applications suitable for transmission of color video over ISDN at  $p \times 64$  kbits/s with low delay; here  $p$  specifies an integer with values between 1 and 30 to allow transmission over more than one ISDN channel [33]. Although the H.261 video coding standard was developed before the JPEG standard, both algorithms feature common elements. The H.261 standard specifies a Hybrid DCT/DPCM coding algorithm with motion compensation—which can be seen as a straight-forward extension of the JPEG baseline algorithm towards Inter-frame coding.

For an H.261 video coder the input video source consists of non-interlaced color frames of either CIF or quarter CIF (QCIF) format and a frame rate of 29.97 frames per second. Notice that the CIF format specifies frames with  $352 \times 288$  active luminance pixels (Y) and  $176 \times 144$  pixels for each chrominance band (U or V), each pixel represented with 8 bits. H.261 compatible decoders must be able to operate with QCIF frames—the CIF format is optional.

**4.2.1. H.261 Hybrid DCT/DPCM Coding Scheme.** As outlined in Figure 8 (A) the H.261 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 frame is used for prediction. Similar to the JPEG baseline coder, the H.261 algorithm processes the frames of a video sequence block-based. Each color input frame in a video sequence is partitioned into non-overlapping “Macroblocks” as depicted in Figure 8 (B). Each Macroblock contains blocks of data from both luminance and co-sited chrominance bands—four luminance blocks ( $Y_1, Y_2, Y_3, Y_4$ ) and two chrominance blocks (U, V), each with size  $8 \times 8$  pels.

The block diagram of the hybrid DCT/DPCM H.261 encoder and decoder structure is depicted in Figure 9. The previously coded frame  $N - 1$  is stored in a frame store (FS) in both

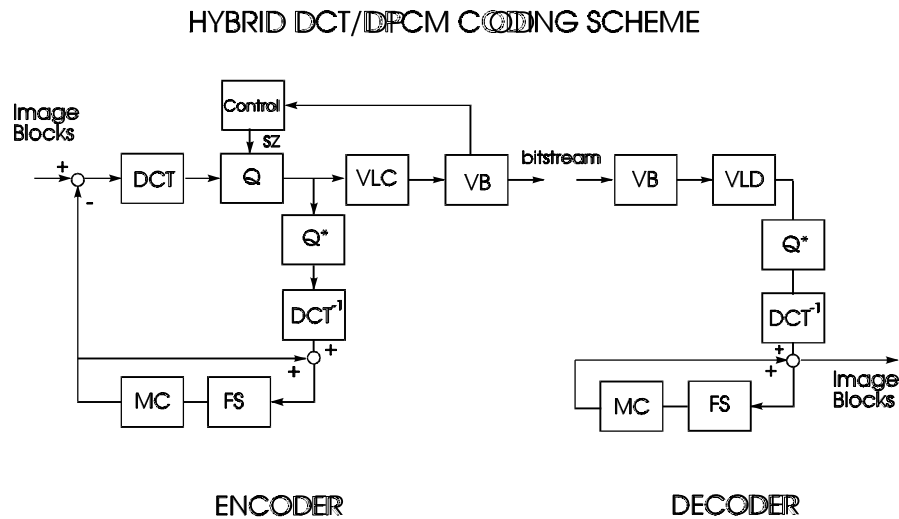


FIGURE 9. Block diagram of the H.261 Hybrid DCT/DPCM encoder and decoder structure.

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. The motion compensated prediction error is calculated by subtracting each pel in a Macroblock with its motion shifted counterpart in the previous frame. A  $8 \times 8$  DCT is then applied to each of the  $8 \times 8$  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 step-size (sz) can be adjusted for each Macroblock in a frame to achieve a given target bit rate and to avoid buffer overflow and underflow [34].

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  $N$ .

**4.2.2. Conditional Replenishment.** The H.261 coding standard is optimized to encode video-phone and videoconferencing scenes at low bit rates ( $> 64$  kbits/s) with sufficient quality. An essential feature supported by the coding scheme 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 low bit rates is the selection of appropriate prediction modes to achieve Conditional Replenishment [34]. Basically, the H.261 standard distinguishes between three different Macroblock coding types (MB types):

*Not coded 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 step-size 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 step-size are transmitted to the receiver.

#### 4.3. 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. Similar to JPEG, 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 of 720 pixels per line, 576 lines per picture, 30 frames per second, and 1.86 Mbits/s. The standard video input consists of a non-interlaced video picture format. It should be noted that the application of MPEG-1 is by no means 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 sought to retain a large degree of commonality 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.

**4.3.1. MPEG-1 Inter-frame Coding Scheme.** The basic MPEG-1 video compression technique is almost identical to the H.261 hybrid DCT/DPCM block-based scheme including Macroblock structure, motion compensation and conditional replenishment, as already outlined in Figure 8 and Figure 9. However, to incorporate the requirements for storage media and to further explore the significant advantages of motion compensation and motion interpolation, the concept of I-pictures and B-pictures (bi-directional predicted/bi-directional interpolated pictures) was introduced by MPEG-1. This concept is depicted in Figure 10 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 the context of the H.261 standard. 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 frame (either 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.

The user can arrange the picture types in a video sequence with a high degree of flexibility to suit diverse application 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 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 10 (I B B P B B P B B I B B P . . .), may achieve high compression and good 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.

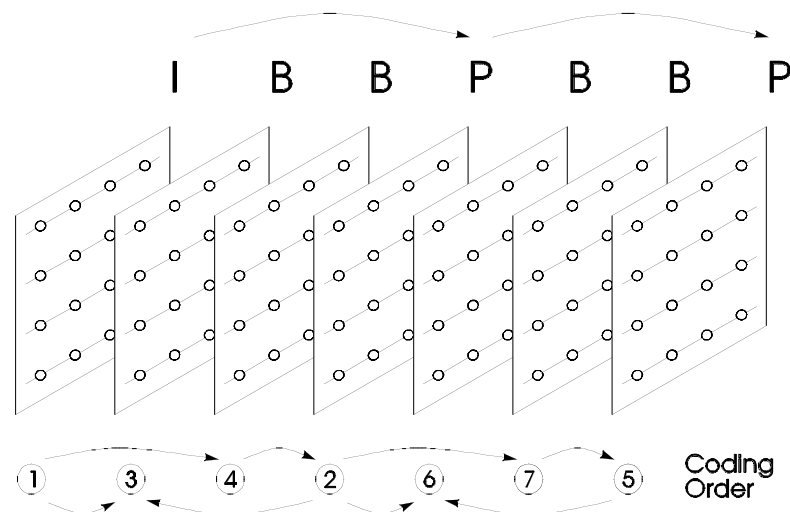


FIGURE 10. I-pictures (I), P-pictures (P) and B-pictures (B) used in an MPEG-1 video sequence. B-pictures can be coded using motion compensated prediction based on the two nearest already coded frames (either I-picture or P-picture). The arrangement of the picture coding types within the video sequence is flexible to suit the needs of diverse applications. The direction for prediction is indicated in the figure.

**4.3.2. Coding of Interlaced Video Sources.** The standard video input format for MPEG-1 is non-interlaced. However, coding of interlaced color 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 color 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 odd field. At the receiver the even field is predicted from the decoded and horizontally interpolated odd field for display, similar to the method already described in Figure 3 (B). 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 [22].

#### 4.4. MPEG-2 and ITU-T-H.262 Standards 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 application 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 successfully 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 was released. 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. It is expected that the ITU-T Experts Group for ATM Video Coding will adapt



Level	Parameters
HIGH	1920 samples/line 1152 lines/frame 60 frames/s 80 Mbit/s
HIGH 1440	1440 samples/line 1152 lines/frame 60 frames/s 60 Mbit/s
MAIN	720 samples/line 576 lines/frame 30 frames/s 15 Mbit/s
LOW	352 samples/line 288 lines/frame 30 frames/s 4 Mbit/s

TABLE 2. Upper bounds of parameters at each level of a profile

the MPEG-2 International Standard, thus the ITU-T H.262 standard for ATM Video Coding will be identical or at least very similar to MPEG-2.

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 a means for defining subsets of the syntax and thus the decoder capabilities required to decode a particular bit stream. This concept is illustrated in Tables 2 and 3.

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.

**4.4.1. 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 9, incorporating a Macroblock structure, motion compensation and coding modes for conditional replenishment of Macroblocks. The

Profile	Algorithms
HIGH	Supports all functionality provided by the Spatial Scalable Profile plus the provision to support <ul style="list-style-type: none"> <li>• 3 layers with the SNR and Spatial scalable coding modes</li> <li>• 4:2:2 YUV-representation for improved quality requirements</li> </ul>
SPATIAL Scalable	Supports all functionality provided by the SNR Scalable Profile plus an algorithm for <ul style="list-style-type: none"> <li>• Spatial scalable coding (2 layers allowed)</li> <li>• 4:0:0 YUV-representation</li> </ul>
SNR Scalable	Supports all functionality provided by the MAIN Profile plus an algorithm for <ul style="list-style-type: none"> <li>• SNR scalable coding (2 layers allowed)</li> <li>• 4:2:0 YUV-representation</li> </ul>
MAIN	Non-scalable coding algorithm supporting functionality for <ul style="list-style-type: none"> <li>• coding interlaced video</li> <li>• random access</li> <li>• B-picture prediction modes</li> <li>• 4:2:0 YUV-representation</li> </ul>
SIMPLE	Includes all functionality provided by the MAIN Profile but <ul style="list-style-type: none"> <li>• does not support B-picture prediction Modes</li> <li>• 4:2:0 YUV-representation</li> </ul>

TABLE 3. Algorithms and functionality supported with each profile

concept of I-pictures, P-pictures and B-pictures as introduced in Figure 10 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 11). 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.

*Field and Frame Prediction.* New motion compensated field prediction modes were introduced by MPEG-2 to efficiently encode field pictures and frame pictures. A simplified example of this new concept is illustrated in Figure 11 for an interlaced video sequence, here

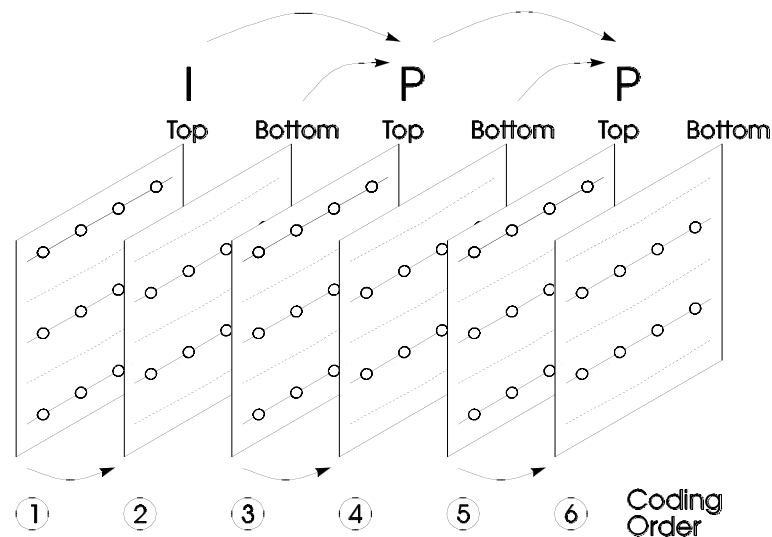


FIGURE 11. The concept of field-pictures and an example of possible field prediction. The top fields and the bottom fields are coded separately. However, each bottom field is coded using motion compensated Inter-field prediction based on the previously coded top field. The top fields are coded using 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.

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 fields, *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 (this technique is similar to the Inter-field prediction already outlined in Figure 3). Generally the Inter-field prediction from the decoded field in the same picture is preferred if no motion occurs between fields. An indication of 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 preferred 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  $16 \times 8$  blocks. A discussion of these methods is beyond the scope of this chapter.

**Chrominance Formats.** MPEG-2 has specified additional Y:U:V luminance and chrominance subsampling ratio formats to assist and enable 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 and 4:4:4 formats suitable for studio video coding applications.

**4.4.2. 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

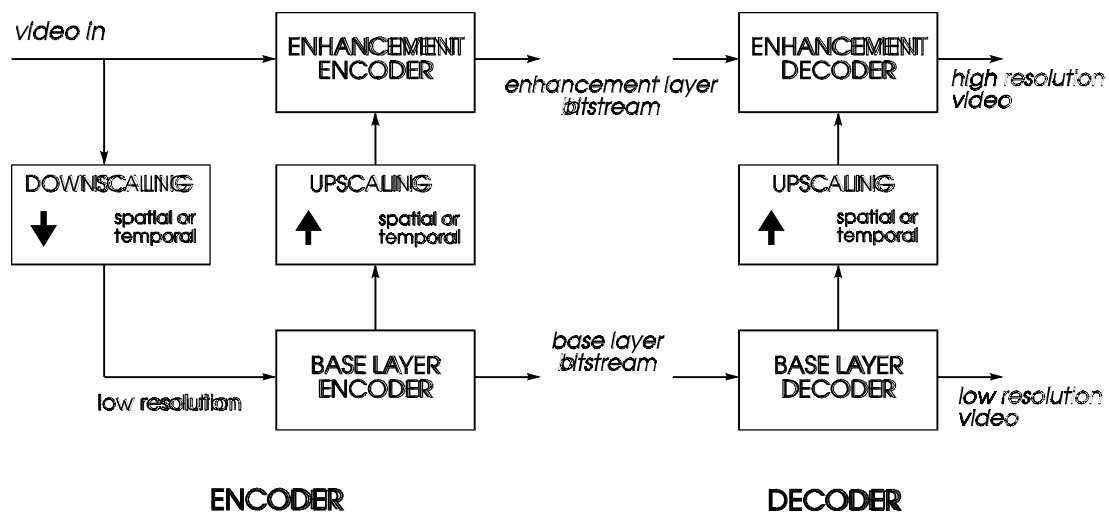


FIGURE 12. Scalable coding of video.

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 12 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 downsampled 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 downsampled 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.

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 constrained 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<sup>4</sup> –

<sup>4</sup>When writing this chapter the Temporal Scalability tool was standardized but was not allocated to any specific MPEG-2 Profile.

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 [26]. The MPEG-2 syntax supports up to three different scalable layers.

*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 and has considerable similarity with the JPEG Progressive DCT-based method, outlined in Section 4.1. Both layers in Figure 12 encode the video signal at the same spatial resolution. At the base layer the DCT coefficients are coarsely quantized to achieve moderate image quality at reduced bit rate. The enhancement layer encodes the difference between the non-quantized DCT-coefficients and the quantized coefficients from the base layer with finer quantization step-size. The method is implemented as a simple and straightforward extension to the MAIN Profile MPEG-2 coder and achieves excellent coding efficiency.

It is also possible to use this method to obtain video with lower spatial resolution at the receiver. If the decoder selects the lowest  $N \times N$  DCT coefficients from the base layer bit stream, non-standard inverse DCT's of size  $N \times N$  can be used to reconstruct the video at reduced spatial resolution [37, 38]. However, depending on the encoder and decoder implementations the lowest layer downsampled video may be subject to drift [39, 40].

*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 migration from a digital TV service to higher spatial resolution HDTV services [25, 26, 42]. The algorithm is based on a classical pyramidal approach for progressive image coding [41, 43]. Spatial Scalability can flexibly support a wide range of spatial resolutions but adds considerable implementation complexity to the MAIN Profile coding scheme.

The *Temporal Scalability* tool was developed with an aim similar to spatial scalability—different frame rates can be supported with a layered bit stream suitable for receivers with different frame rate display capabilities. Layering is achieved by providing a temporal prediction for the enhancement layer based on coded video from 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 [25]. 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.

#### 4.5. CCIR 723- Digital Coding of Component Television Signals for Contribution-Quality Applications at Bit Rates in the Range 34-45 Mbit/s

The CCIR 723 standard specifies a hybrid DCT/DPCM coding scheme with motion compensation, similar to the techniques described above. However, this standard is not generic and the only target application envisaged is the contribution of TV-pictures according to CCIR Rec. 601, preserving sufficient image quality for limited post-production purposes.

The codec is optimized for coding of interlaced TV signals at high bit rates and uses a field-picture structure only. An Intra-field coding mode is employed for prediction as well as two Inter-field modes with and without motion compensated prediction respectively. The modes are selected on a Macroblock basis. Macroblocks consist of two  $8 \times 8$  luminance blocks and two  $8 \times 8$  chrominance blocks for the CR and CB components. The selection criteria as well as Macroblock refreshing procedures (by forced Intra-field coding) are not specified.

In pure Intra-field mode, blocks of  $8 \times 8$  pels are directly DCT transformed, quantized and VLC encoded. In the Inter-field prediction mode without motion compensation a temporal prediction is performed between two adjacent fields (using an  $x = (E + F)/2$  predictor as outlined in Figure 3) and the prediction error is DCT transformed and coded. The motion compensated inter-field mode employs motion compensated prediction based on the last previously coded field. Similar to the video coding standards introduced above, a perceptual weighting of the DCT coefficients according to a visibility matrix can be performed to explore subjective redundancies in the video data.

#### 4.6. CCIR 721—Transmission of Component-Coded Digital Television Signals for Contribution-Quality Applications at Bit Rates Near 140 Mbit/s

As the CCIR 723 standard, the CCIR 721 coding scheme is intended for contribution of CCIR-601 interlaced TV signals, aiming at very high quality. The codec operates on a field-picture structure and only employs one spatial prediction mode. Thus, only a low compression performance can be achieved and consequently the codec operates at a very high bit rate around 140 Mbit/s for coding of TV signals with almost “lossless” quality. Advantages of the CCIR 721 coding scheme are the low complexity for the implementation of both encoder and decoder and the high degree of random access functionality achieved. Due to the field-structure and the pure Intra-field coding mode employed, every field within a video sequence can be accessed and decoded separately, to assist video post-processing in the studio.

Fixed two-dimensional prediction ( $x = (A + B)/2$  as related to Figure 3 (A)) is applied to both the luminance and color-difference components within a field. A non adaptive hybrid DPCM scheme, known as the Van Buul technique [29], in combination with folded quantizers, as proposed by Bostelmann [30], form the main elements of the coding scheme. Nonlinear quantization using 6 bits/sample results in a video bit rate of 124 416 kbit/s for CCIR Rec. 601 input material. The video data is multiplexed with audio and ancillary data and transmitted in a so-called TV container with a data rate of 138 240 kbit/s. This TV-container fits into a channel framing according to CCITT Recommendations G.751 and G.707—G.709. This standard is not generic at all and, in contrast to all standards mentioned above, both encoder and decoder are fully specified.

#### 4.7. H. 120: Codecs for Videoconferencing Using Primary Digital Group Transmission

The H.120 standard specifies coding schemes targeted for the compression of interlaced TV signals for videoconferencing applications. The H.120 Recommendation consists of 3 parts: Part 1 specifying a codec for transmission of video with 625 lines—50 Hz and targeted for 2 Mbit/s, Part 3 aiming for 525 lines—60 Hz systems for transmission with 1.5 Mbit/s, and Part 2 specifies a codec for 625 lines—50 Hz and 525 lines—60 Hz systems at 1.5 Mbit/s.

The Part 1 algorithm uses “Conditional Replenishment”, where only those areas detected as moving are DPCM coded using spatial prediction ( $x = (A + D)/2$  for  $Y$  components and  $x = A$  for  $U$  and  $V$  components, see Figure 3). Additional techniques used are vertical subsampling before coding (to 143 lines/field) and adaptive horizontal and field subsampling in the case of buffer overflow.

The Part 2 coding scheme is specified to support 525 lines—60Hz TV systems but is otherwise identical to the Part 1 algorithm—therefore interoperability between Part 1 and Part 2 encoders and decoders is possible. To achieve compatibility between the two schemes, the 525 lines input pictures are vertically subsampled to 143 lines/field (as for the Part 1 codec) and every 6th field is omitted. In the 525-line, 60 Hz decoder the missing fields are reconstructed using temporal interpolation. Additionally, the Part 2 encoder uses a recursive temporal pre-filter for noise reduction in order to increase the coding efficiency.

The Part 3 algorithm employs a predictive coding scheme with three possible prediction modes: Motion compensated inter-field prediction can be used to encode non-moving and slowly moving areas (the motion vector search area supported is  $\pm 15$  pels horizontally and  $\pm 7$  lines vertically). A background prediction mode suitable for detecting uncovered background using data from a background memory can be employed. An Intra-field prediction mode can be used for coding of rapidly moving areas (using a simple spatial predictor  $x = A$ , see Figure 3).

The prediction modes are selected on a pel-by-pel basis without transmitting side information but by using control signals from the previous pel and the previous line. One motion vector is transmitted for blocks of 16 pels  $\times$  8 lines. The quantized prediction errors and the motion vectors are run-length and VLC coded. Altogether five coding modes including spatial and temporal subsampling are provided, which can be selected on a block, block-line or field basis.

#### 4.8. MPEG-4 and ITU-TS Experts Groups for Coding of Video at Very Low Bit Rates

Recent developments in telecommunications technologies and multimedia systems have prompted the demand for coding of audio-visual information at very low bit rates (5—64 kbit/s) for storage and transmission. It is generally expected that the delivery of video information over existing and future low-bandwidth communication networks will become increasingly important, such as audio-visual services operating over mobile radio networks as well as the PSTN. However, the success of these services in the market place will depend on the ability to encode video at very low bit rates with sufficient image quality. Existing video coding standards (e.g., H.261 or MPEG-1) have been optimized to achieve good video quality at bit rates higher than 64 kbit/s. Accordingly, the video quality provided by these algorithms is not sufficient for the applications envisaged at very low bit rates.

The ITU-TS Experts Group for Very Low Bit Rate Visual Telephony (ITU-TS is the former CCITT SG XV) started activities in 1993 and has targeted its work into two areas: near term work is directed towards a Rec. H.26P coding algorithm and a long term effort towards a H.26P/L coding scheme. It is expected that the first Rec. H.26P standards draft will be frozen in early 1995. To meet the rather short time schedule requirement the H.26P video coding algorithm will be an extension of Rec. H.261. However, to adapt H.261 for videophone applications at very low bit rates between 9.6 kbit/s . . . 28.8 kbit/s, a number of significant changes will be required. Changes discussed at present include extended motion compensation accuracy and smaller motion vector search window sizes compared to H.261. The ITU work to develop Rec. H.26P/L is being accomplished in close collaboration with the ISO MPEG-4 activity.

ISO MPEG-4 started its standardization activities in July 1993 with the charter to develop a generic video coding algorithm mainly targeted for a wide range of low bit rate multimedia applications. In contrast to the near term solution envisaged by the ITU-TS group, it

is planned to incorporate new important multimedia functionalities into the MPEG-4 video coding standard and to finalize the MPEG-4 international standard in late 1998. The additional functionalities attempt to provide a high degree of interactivity, interoperability and flexibility, needed to support universal accessibility and portability of compressed audio and video between emerging future Television and computer and telecommunications multimedia applications. Additional functionalities envisioned attempt to provide

- sufficient temporal random access for very low bitrate audio and video,
- the ability to manipulate or edit the content of audio-visual sequences (segments of arbitrary shape in images) in the compressed domain without the need for transcoding,
- enhanced scalability features to support various multimedia applications that face continuously varying bandwidth, and to efficiently support software decoding of video,
- content-based multimedia data access tools to assist indexing, hyperlinking, querying, browsing, uploading, downloading and deleting of audio-visual content in sequences on the bitstream level,
- coding of multiple concurrent data streams to assist coding of multiple views or soundtracks of a scene efficiently,
- efficient methods for combining synthetic scenes with natural scenes on the bitstream level (*e.g.*, text and graphics overlays),
- robustness in error-prone environments (*e.g.*, targeted for applications over a variety of mobile, wireless and wired networks).

To accomplish these objectives it is expected that a significant change in the video source model is required. A “Call for Proposals” for suitable coding schemes is issued for late 1995 and a competitive test is planned to evaluate the efficiency, robustness and flexibility of the coding schemes proposed based on a large variety of video source material. A wide range of algorithms are currently discussed as possible candidates for the future MPEG-4 and H.26P/L standards. Many of the algorithms are based on so-called “second generation coding techniques” [45]—*e.g.*, object based coding schemes [46], model based coding techniques [47] and segmentation based coding schemes [48]. Many of the schemes diverge significantly from the successful hybrid DCT/DPCM coding concept employed in the current H.261, MPEG-1 and MPEG-2 standards.

## 5. Discussion

International standardization in image coding has evolved remarkably: from a committee-driven process dominated by Telecoms and broadcasters to a market-driven process incorporating industries, Telecoms, network operators, satellite operators, broadcasters and research institutes. With this evolution the actual work of the standardization bodies has also changed considerably, and has evolved from discussion circles of national delegations into international collaborative R&D activities. The standardization process has become significantly more efficient and faster—the reason is that standardization has to follow the accelerated speed of technology development because otherwise standards are in danger of being obsolete before they are agreed upon by the standardization bodies.

The benefits of standardization activities have been recognized world-wide—and with this recognition the basic philosophy behind standardization has changed. Early standards such as H.120 and CCIR 721 used to be designed to meet the requirements for mainly one application, whereas with recent standards such as JPEG, MPEG-1, and MPEG-2 generic algorithms were targeted. Consequently, these standards can find a wider range of applications and therefore have the potential to lead to an economy of scale.

In the telecommunications world, international standards are most important because communication on an international scale is impossible without interoperability of equipment from different manufacturers and vendors. In other areas, such as consumer electronics or



broadcasting, video coding standards seem less important. However, industries involved in manufacturing video communication equipment can benefit significantly. Here, international standardization creates a wider market for compatible video coding technology with large production numbers of VLSI chips and related communication equipment—and therefore leads again to economy of scale. The consumer benefits from the interoperability of the diverse equipment on the market and from the increased accessibility to a broader range of image and video material on storage media. Thus, video communications technology becomes more attractive and innovative for the consumer market which results, in return, in an increased demand for video equipment.

It has to be understood that video coding standards have to rely on compromises between what is theoretically possible and what is technologically feasible. Standards can only be successful in the market place if the cost-performance ratio is well balanced. This is specifically true in the field of image and video coding where a large variety of innovative coding algorithms exist, but may be too complex for implementation with state-of-the-art VLSI technology.

Looking at the acceptance of coding standards we observe that the early standards have been less successful than the more recent ones. Certainly the technology development related to the areas of coding, transmission and storage of video and related to VLSI design has been a major factor in this respect.

The H.120 standard was not successful because it could not be established as a unique world-wide standard and also because the price-performance ratio was not adequate: the costs for the codec equipment and the costs to deliver the video with the relatively high data rate required for transmission were too high to find acceptance.

The H.261 standard was intended to lower at least the transmission costs by giving the freedom to the costumer to choose between video quality and transmission costs. However, at the time the standard was issued only a poor spread of ISDN connections was available and the costs for implementation of the video coding equipment was high due to the limited performance of the VLSI technology. Meanwhile, the spread of ISDN has grown considerably and the advances in technology allow more economic implementations. However, there is a danger that H.261 might become obsolete due to the advent of MPEG-1, which offers a higher functionality and efficiency. Unfortunately, video telephony and multipoint teleconferencing are the only services targeted by H.261. This limited applicability may hinder the success of the H.261 standard because the two services envisaged can also be covered by the MPEG-1 standard.

In this respect the MPEG-1 standard may provide considerable advantages for a variety of multimedia terminals, with the additional flexibility provided for random access of video from storage media and the diverse image source formats assisted. A number of MPEG-1 encoder and decoder chip sets from different vendors are available on the market. Encoder and decoder PC boards have been developed using MPEG-1 chip sets. A number of commercial products use the MPEG-1 coding algorithm for interactive CD applications, such as the CD-I product.

MPEG-2 will become a success because there is a strong commitment from industries, cable and satellite operators and broadcasters to use this standard. Digital TV broadcasting, pay TV, pay-per-view, video-on-demand, interactive TV and many other future video services are the applications envisaged. Although at present the final MPEG-2 standard document has not even been released, many MPEG-2 MAIN Profile at MAIN Level decoder prototype chips are already developed and the real-time testing of the coding algorithm is being performed extensively by many companies involved in the standardization process. The world-wide acceptance of MPEG-2 in consumer electronics will lead to large production scales, making MPEG-2 decoder equipment cheap and therefore also attractive for other related areas, such as video communications and storage and multimedia applications in general.

The JPEG standard has already proven to be a big success. There is a considerable demand to compress still images for archiving and publishing and for multimedia applications in general. The JPEG baseline coding algorithm has found widespread application, and several hardware implementations from different vendors are available on the market today at reasonable prices. The JPEG standard is also attractive because the high computing power of PCs and workstations today allows for software decoding of compressed images without the need to use dedicated hardware. Furthermore, JPEG has found many applications which are beyond the initial target. Examples are motion JPEG for video transmission or video storage and post-processing in professional production and broadcast applications.

The success of the other two standards, CCIR 721 and CCIR 723, which have been established for TV-contribution, is less certain. Although the implementation complexity for CCIR 721 codecs is very low and a high video quality can be achieved, the number of codecs in use will remain small. The main reason is the high data rate of 140 Mbit/s targeted for transmission, which results in very high transmission costs for the delivery of the video.

At least in Europe the situation seems more favorable for CCIR 723 codecs, because the European Broadcast Union decided recently to base their Europe-wide 34 Mbit/s contribution network over satellites on these codecs [49]. However, there is a danger that MPEG-2, which is now also targeting at 4:2:2 and 4:4:4 Y:U:V-representation of video with 10 bit resolution, might offer a higher quality at the same or even lower data rates and therefore make the CCIR 723 standard obsolete before it has even been established.

Finally, the MPEG-4 phase is on its way, but it is hard to predict now which techniques will be used and whether a low bit-rate ( $< 64$  kbit/s) standard will become a success. There is certainly a demand to transmit video data over PSTNs or mobile networks, but the acceptance of such services will very much depend on the achievable image quality. It is not certain whether second generation coding techniques will finally be suitable for MPEG-4. Although non-waveform second generation coding schemes, such as object oriented analysis synthesis coding, have made remarkable progress, DCT-based approaches have been improved as well and acceptable quality has been demonstrated at bit rates of 16 kbit/s. The final standard may again be a compromise with elements from both waveform and non waveform coding schemes. Most certainly, to achieve the many new functionalities targeted will be a great challenge, but this also offers outstanding potential for new compatible products to be developed around a future MPEG-4 standard – the ultimate goal of international video standardization.

## References

- [1] 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.
- [2] B. R. Halhed, "Videoconferencing Codecs: Navigating the MAZE", *Business Communication Review*, Vol. 21, No. 1, pp. 35-40, 1991.
- [3] S. Okubo, "Requirements for high quality video coding standards", *Signal Processing: Image Communication*, Vol. 4, pp. 141-151, 1992.
- [4] T. Hidaka and K. Ozawa, "ISO/IEC JTC1 SC29/WG11, Report on MPEG-2 Subjective Assessment at Kurihama", *Signal Processing: Image Communication*, Vol. 5, Nos. 1-2, pp. 127-157, February 1993.
- [5] H. Yasuda, "Standardization Activities on Multimedia Coding in ISO", *Signal Processing: Image Communication*, Vol. 1, No. 1, pp. 3-16, 1989.
- [6] S. Okubo, M. Wada, M. D. Carr, A. J. Tabatabai, "Hardware Trials for Verifying Recommendation H.261 on p x 64 kbit/s Video Codec", *Signal Processing: Image Communication* Vol. 3, No. 1, pp. 71-78, 1991.
- [7] S. Matsumoto and H. Murakami, "30/45 Mbps Digital Coding System for Transmission of (4:2:2) Digital Component TV Signals Recommended by CCIR", *Journal of Visual Communication and Image Representation*, Vol. 2, No. 4, pp. 314-324, 1991.
- [8] S. Okubo, "Video Codec Standardization in CCITT Study Group XV", *Signal Processing: Image Communication*, Vol. 1, No. 1, pp. 45-54, 1989.
- [9] CCITT Recommendation H.120, "Codecs for Videoconferencing Using Primary Digital Group Transmission", Geneva, 1989.

- [10] R. C. Nicol and N. Mukawa, "Motion Video Coding in CCITT SG XV – The Coded Picture Format", in *Conference Record of IEEE Global Telecommunications Conference and Exhibition—Communications for the Information Age*, Hollywood, Vol. 2, pp. 992-996, 28 Nov. - 1 Dec. 1988.
- [11] N. Mukawa, H. Kuroda, T. Matsuoka, "An Inter-frame Coding System for Video Teleconferencing Signal Transmission at a 1.5 Mbit/s rate", *IEEE Trans. Comm.*, Vol. COM-23, No. 3, pp. 280-287, 1984.
- [12] CCITT Recommendation H.261, "Video Codec for Audiovisual Services at p x 64 kbit/s", Geneva, 1990.
- [13] L. Stenger, "Digital Coding of Television Signals—CCIR Activities for Standardization", *Signal Processing: Image Communication*, Vol.1, No.1, pp. 29-43, 1989.
- [14] CCIR Recommendation 721, "Transmission of Component-Coded Digital Television Signals for Contribution-Quality Applications at Bit Rates Near 140 Mbit/s", Geneva, 1990.
- [15] CCIR Recommendation 723 / ETS 300 174, "Digital Coding of Component Television Signals for Contribution Quality Applications in the Range 34-45 Mbit/s", Sophia Antipolis, 1992.
- [16] M. Barbero, S. Cucchi, M. Stroppiana, "A Codec for HDTV Transmission", in *Proceedings of Fourth International Workshop on HDTV and beyond*, Vol. 1, pp. 5b, Torino, 1991.
- [17] A. D. R. Cotton and N. D. Wells, "Experiments with the HIVITS HDTV Contribution Codec", BBC Research Department, No. 12, 1992.
- [18] C. Guillemot, F. Guirao, C. Perron, R. Schäfer, "Flexible Bit Rate Coding for Satellite Transmission of HDTV", in *Proceedings of International Workshop on HDTV 92*, Kawasaki, Japan, November 18-20, 1992.
- [19] A. Léger, T. Omachi, G. K. Wallace, "JPEG Still Picture Compression Algorithm", *Optical Engineering*, Vol. 30, No. 7, pp. 947-954, 1991.
- [20] G. K. Wallace, "The JPEG Still Picture Compression Standard", *IEEE Transactions on Consumer Electronics*, Vol. 38, No. 1, pp. 18-34, 1992.
- [21] ISO/IEC 10918-1, "Information Technology—Digital Compression and Coding of Continuous-Tone Still Images: Requirements and Guidelines", Geneva, 1994.
- [22] 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.
- [23] D. J. Le Gall, "The MPEG Video Compression Algorithm", *Signal Processing: Image Communication* 4, No. 2, pp. 129-140, 1992.
- [24] C. T. Chen and D. J. Le Gall, "A K-th Order Adaptive Transform Coding Algorithm for High-Fidelity Reconstruction of Still Images", in *Proc. SPIE*, San Diego, August 1989.
- [25] 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.
- [26] J. De Lameillieure and R. Schäfer, "MPEG2 Image Coding for Digital TV", *Fernseh und Kino Technik*, 48. Jahrgang, pp. 99-107, March 1994 (in German).
- [27] D. Lappe et al., "A Proposal for a Reference Model for Mobile Video Communication", in *Proceedings of 1993 Picture Coding Symposium*, Lausanne, 17-19 pp. 19.4, March 1993.
- [28] M. Hötter, "Object-Oriented Analysis-Synthesis Coding Based on Moving Two- Dimensional Objects", *Image Communication*, Vol. 2, No. 4, Dec. 1990.
- [29] M. C. W. van Buul, "Hybrid D-PCM, A Combination of PCM and DPCM", *IEEE Trans. Comm.*, Vol. COM-26, No. 3, pp. 362-368, 1978.
- [30] G. Bostelmann, "A Simple High Quality DPCM Codec", *NTZ*, Vol. 27, No. 3, pp. 115-117, 1974.
- [31] A. N. Netravali and B. G. Haskell "Digital Pictures: Representation and Compression", Plenum Press, New York and London, 1989.
- [32] C. T. Chen, "Video Compression: Standards and Applications", *Visual Communication and Image Representation*, No. 2, pp. 103-111, June 1993.
- [33] CCITT SG XV, "Draft Revision of recommendation H.261", Document 572, March 1990.
- [34] CCITT SG XV, "Description of Reference Model 8 (RM8)", Document 525, June 1989.
- [35] 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.
- [36] R. J. Clark, "Transform Coding of Images", Academic Press, Orlando, Florida, 1985.
- [37] C. Gonzales and E. Viscito, "Flexibly scalable digital video coding", *Signal Processing: Image Communication*, Vol. 5, No. 1-2, February 1993.
- [38] 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.
- [39] H. G. Lim, K. K. Pang, T. K. Tan and S. C. Hall, "A low complex software video decoder", *Signal Processing: Image Communication, Special Issue on Software Decoding*, June issue 1994.
- [40] 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.

- [41] A. Puri and A. Wong, "Spatial Domain Resolution Scalable Video Coding", *Proc. SPIE Visual Communications and Image Processing*, Boston, MA, November 1993.
- [42] J. De Lameilieu and G. Schamel, "Hierarchical Coding of TV/HDTV within the German HDTV Project", *Proc. Int. Workshop on HDTV'93*, pp. 8A.1.1 - 8A.1.8, Ottawa, Canada, October 1993.
- [43] P. J. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Trans. COM*, Vol. COM-31, pp. 532-540, 1983.
- [44] A. K. Jain, "Image data compression: A review", *Proc. IEEE*, Vol. 69, pp. 349-384, 1981.
- [45] M. Kunt, A. Ikonomopoulos and M. Kocher, "Second generation image-coding techniques", *Proc. IEEE*, Vol. 73, No. 4, pp. 549-574, April 1985.
- [46] H. G. Musmann, M. Hötter and J. Ostermann, "Object-Oriented Analysis-Synthesis Coding of Moving Images", *Signal Processing: Image Communication*, Vol. 1, No. 2, pp. 117-138, 1989.
- [47] K. Aizawa, H. Harashima and T. Saito, "Model Based Analysis Synthesis Image Coding (MBASIC) System for a Persons Face", *Signal Processing: Image Communication*, Vol. 1, No. 2, pp. 138-152, 1989.
- [48] P. Salembier and J. Serra, "Morphological Multiscale Image Segmentation", *Trans. SPIE, Visual Communications and Image Processing*, Vol. 1818, pp. 620-631, 1992.
- [49] L. Cheveau and R. Miles, "Digitisation of the Eurovision network", *EBU Technical Review*, No. 258, pp. 14-24, Winter 1993.