MPEG: A Video Compression Standard for Multimedia Applications

Didier Le Gall

AMAZON-1019
6,856,701

The development of digital video technology in the 1980s has made it possible to use digital video compression for a variety of telecommunication applications: teleconferencing, digital broadcast codec and video telephony.

Standardization of video compression techniques has become a high priority because only a standard can reduce the high cost of video compression codecs and resolve the critical problem of interoperability of equipment from different manufacturers. The existence of a standard is often the trigger to the volume production of integrated circuits (VLSI) necessary for significant cost reductions. An example of such a phenomenon—where a standard has stimulated the growth of an industry—is the spectacular growth of the facsimile market in the wake of the standardization of the Group 3 facsimile compression algorithm by the CCITT. Standardization of compression algorithms for video was first initiated by the CCITT for teleconferencing and videotelephony [7]. Standardization of video compression techniques for transmission of contribution-quality television signals has been addressed in the CCIR[1] (more precisely in CMTT/2, a joint committee between the CCIR and the CCITT).

Digital transmission is of prime importance for telecommunication, particularly in the telephone network, but there is a lot more to digital video than teleconferencing and visual telephony. The computer industry, the telecommunications industry and the consumer electronics industry are increasingly sharing the same technology—there is much talk of a convergence, which does not mean that a computer workstation and a television receiver are about to become the same thing, but certainly, the technology is converging and includes digital video compression. In the view of shared technology between different segments of the information processing industry, the International Organization for Standardization (ISO) has undertaken an effort to develop a standard for video and associated audio on digital storage media, where the concept of digital storage medium includes conventional storage devices CD-ROM, DAT, tape drives, winchesters, writable optical drives, as well as telecommunication channels such as ISDNs, and local area networks.

This effort is known by the name of the expert group that started it: MPEG—Moving Picture Experts Group—and is currently part of the ISO-IEC/JTC1/SC2/WG11. The MPEG activities cover more than video compression, since the compression of the associated audio and the issue of audio-visual synchronization cannot be worked independently of the video compression: MPEG-Video is addressing the compression of video signals at about 1.5 Mbits, MPEG-Audio is addressing the compression of a digital audio signal at the rates of 64, 128 and 192 kbits/s per channel, MPEG-System is addressing the issue of synchronization and multiplexing of multiple compressed audio and video bit streams. This article focuses on the activities of MPEG-Video. The premise of MPEG is that a video signal and its associated audio can be compressed to a bit rate of about 1.5 Mbits/s with an acceptable quality.

Two very important consequences follow: Full-motion video becomes a form of computer data, i.e., a data type to be integrated with text and graphics; Motion video and its associated audio can be delivered over existing computer and telecommunication networks.

Precompetitive Research

The growing importance of digital video is reflected in the participation of more and more companies in standards activities dealing with digital video; MPEG is a standard that responds to a need. In this situation a standards committee is a forum where precompetitive research can take place, where manufacturers meet researchers, where industry meets academia. By and large, because the problem to be solved was perceived as important, the technology developed within MPEG is at the forefront of both research and industry. Now that the work of the MPEG committee has reached maturity (a "Committee Draft" was produced in September 1990), the VLSI industry is ready and waiting to implement MPEG's solution.

MPEG Standard Activities

The activity of the MPEG committee was started in 1988 with the goal of achieving a draft of the standard by 1990. In the two years of MPEG activity, participation has increased tenfold from 15 to 150 participants. The MPEG activity was not started without due consideration to the related activities of other standard organizations. These considerations are of interest, not only because it is important to avoid duplication of work between standards committees but most of all, because these activities provided a very important background and technical input to the work of the MPEG committee.

Background: Relevant Standards

The JPEG Standard. The activities of JPEG (Joint Photographic Experts Group) [10] played a considerable role in the beginning of MPEG, since both groups were originally in the same working group of ISO and there has been considerable overlap in membership. Although the objectives of JPEG are focused exclusively on still-image compression, the distinction between still and moving image is thin; a video sequence can be

[1]CCIR is the International Consultative Committee on Broadcasting: CCITT is the International Committee on Telegraph and Telephones. CMTT is a joint committee of the CCITT and the CCIR working on issues relevant to television and telephony.

thought of as a sequence of still images to be coded individually, but displayed sequentially at video rate. However, the "sequence of still images" approach has the disadvantage that it fails to take into consideration the extensive frame-to-frame redundancy present in all video sequences. Indeed, because there is a potential for an additional factor of three in compression exploiting the temporal redundancy, and because this potential has very significant implications for many applications relying on storage media with limited bandwidth, extending the activity of the ISO committee to moving pictures was a natural next step.

CCITT Expert Group on Visual Telephony. As previously mentioned, most of the pioneering activities in video compression were triggered by teleconferencing and videotelephony applications. The definition and planned deployment of ISDN (Integrated Service Digital Network) was the motivation for the standardization of compression techniques at the rate of px64 kbits/s where p takes values from one (one B channel of ISDN) to more than 20 (Primary rate ISDN is 23 or 30 B channels). The Experts Group on visual telephony in the CCITT Study Group XV addressed the problem and produced CCITT Recommendation H.261: "Video Codec for Audiovisual Services at px64 kbits" [7, 9]. The focus of the CCITT expert group is a real-time encoding-decoding system, exhibiting less than 150 ms delay. In addition, because of the importance of very low bit-rate operation (around 64 kbits/s), the overhead information is very tightly managed.

After careful consideration by the MPEG committee, it was perceived that while the work of the CCITT expert group was of very high quality, relaxing the constraint on very low delay and the focus on extremely low bit rates could lead to a solution with increased visual quality in the range of 1 to 1.5 Mbits/s. On the other hand, the contribution of the CCITT expert group has been extremely relevant and the members of MPEG have strived to maintain compatibility, introducing changes only to improve quality or to satisfy the need of applications. Consequently, the emerging MPEG standard, while not strictly a superset of CCITT Recommendation H.261, has much commonality with that standard so that implementations supporting both standards are quite plausible.

CMTT/2 Activities. If digital video compression can be used for videoconferencing or videotelephony applications, it also can be used for transmission of compressed television signals for use by broadcasters. In this context the transmission channels are either the high levels of the digital hierarchy, H21 (34 Mbits/s) and H22 (45 Mbits/s) or digital satellite channels. The CMTT/2 addressed the compression of television signals at 34 and 45 Mbits/s [4]. This work was focused on contribution quality codecs, which means that the decompressed signal should be of high enough quality to be suitable for further processing (such as chromakeying). While the technology used might have some commonalities with the solutions considered by MPEG, the problem and the target bandwidth are very different.

MPEG Standardization Effort

The MPEG effort started with a tight schedule, due to the realization that failure to get significant results fast enough would result in potentially disastrous consequences such as the establishment of multiple, incompatible *de facto* standards. With a tight schedule came the need for a tight methodology, so the committee could concentrate on technical matters, rather than waste time in dealing with controversial issues.

Methodology. The MPEG methodology was divided in three phases: Requirements, Competition and Convergence:

Requirements. The purpose of the requirement phase was twofold: first, precisely determine the focus of the effort; then determine the rules of the game for the competitive phase. At the time MPEG began its effort, the requirements for the integration of digital video and computing were not clearly understood, and the MPEG approach was to provide enough system design freedom and enough quality to address many applications. The outcome of the requirement phase was a document "Proposal Package Description" [8] and a test methodology [5].

Competition. When developing an international standard, it is very important to make sure the tradeoffs are made on the basis of maximum information so that the life of the standard will be long: there is nothing worse than a standard that is obsolete at the time of publication. This means the technology behind the standard must be state of the art, and the standard must bring together the best of academic and industrial research. In order to achieve this goal, a competitive phase followed by extensive testing is necessary, so that new ideas are considered solely on the basis of their technical merits and the trade-off between quality and cost of implementation.

In the MPEG-Video competition, 17 companies or institutions contributed or sponsored a proposal, and 14 different proposals were presented and subjected to analysis and subjective testing (see Table 1). Each proposal consisted of a documentation part, explaining the algorithm and documenting the system claims, a video part for input to the subjective test [5], and a collection of computer files (program and data) so the compression claim could be verified by an impartial evaluator.

Convergence. The convergence phase is a collaborative process where the ideas and techniques identified as promising at the end

of the competitive phase are to be integrated into one solution. The convergence process is not always painless; ideas of considerable merit frequently have to be abandoned in favor of slightly better or slightly simpler ones. The methodology for convergence took the form of an evolving document called a simulation model and a series of fully documented experiments (called core experiments). The experiments were used to resolve which of two or three alternatives gave the best quality subject to a reasonable implementation cost.

TABLE 1.

Participation: Companies and Institutions having contributed an MPEG Video Proposal

Company	Country	Proposer
AT&T	USA	AT&T
Bellcore	USA	Bellcore
Intel	USA	Bellcore
GCT	Japan	Bellcore
C-Cube Micro	USA	C-Cube Micro.
DEC	USA	DEC
France Telecom	France	France Telecom
Cost 211 Bis	EUR	France Telecom
IBM	USA	IBM
JVC Corp	Japan	JVC Corp
Matsushita EIC	Japan	Matsushita EIC
Mitsubishi EC	Japan	Mitsubishi EC
NEC Corp.	Japan	NEC Corp.
NTT	Japan	NTT
Philips CE	Netherlands	Philips CE
Sony Corp.	Japan	Sony Corp.
Telenorma/U. Hannover	Germany	Telenorma/U. Hannover

TABLE 2.

Storage Media and Channels where MPEG could have Applications

CD-ROM
DAT
Winchester Disk
Writable Optical Disks
ISDN
LAN
Other Communication Channels

Schedule. The schedule of MPEG was derived with the goal of obtaining a draft of the standard (Committee Draft) by the end of 1990. Although the amount of work was considerable, and staying on schedule meant many meetings, the members of MPEG-Video were able to reach an agreement on a Draft in September 1990. The content of the draft has been "frozen" since then, indicating that only minor changes will be accepted, i.e., editorial changes and changes only meant to correct demonstrated inaccuracies. Figure 1 illustrates the MPEG schedule for the competitive and convergence phases.

MPEG-Video Requirements

A Generic Standard

Because of the various segments of the information processing industry represented in the ISO committee, a representation for video on digital storage media has to support many applications. This is expressed by saying that the MPEG standard is a *generic* standard. Generic means that the standard is independent of a particular application; it does not mean however, that it ignores the requirements of the applications. A generic standard possesses features that make it somewhat universal—e.g., it follows the toolkit approach; it does not mean that all the features are used all the time for all applications, which would result in dramatic inefficiency. In MPEG, the requirements on the video compression algorithm have been derived directly from the likely applications of the standard.

Many applications have been proposed based on the assumption that an acceptable quality of video

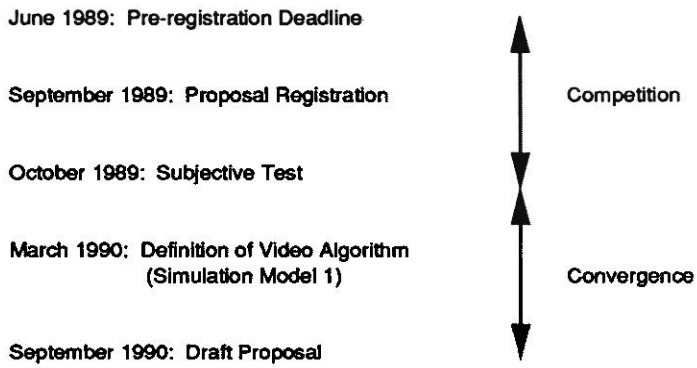


FIGURE 1.
MPEG Schedule for the Competitive and Convergence Phases

can be obtained for a bandwidth of about 1.5 Mbits/second (including audio). We shall review some of these applications because they put constraints on the compression technique that go beyond those required of a videotelephone or a videocassette recorder (VCR). The challenge of MPEG was to identify those constraints and to design an algorithm that can flexibly accommodate them.

Applications of Compressed Video on Digital Storage Media

Digital Storage Media. Many storage media and telecommunication channels are perfectly suited to a video compression technique targeted at the rate of 1 to 1.5 Mbits/s (see Table 2). CD-ROM is a very important storage medium because of its large capacity and low cost. Digital audio tape (DAT) is also perfectly suitable to compressed video; the recordability of the medium is a plus, but its sequential nature is a major drawback when random access is required. Winchester-type computer disks provide a maximum of flexibility (recordability, random access) but at a significantly higher cost and limited portability. Writable optical disks are expected to play a significant role in the future because they have the potential to combine the advantages of the other media (recordability, random accessability, portability and low cost).

The compressed bit rate of 1.5 Mbits is also perfectly suitable to computer and telecommunication networks and the combination of digital storage and networking can be at the origin of many new applications from video on Local area networks (LANs) to distribution of video over telephone lines [1].

Asymmetric Applications. In order to find a taxonomy of applications of digital video compression, the distinction between symmetric and asymmetric applications is most useful. Asymmetric applications are those that require frequent use of the decompression process, but for which the compression process is performed once and for all at the production of the program. Among asymmetric applications, one could find an additional subdivision into electronic publishing, video games and delivery of movies. Table 3 shows the asymmetric applications of digital video.

Symmetric Applications. Symmetric applications require essentially equal use of the compression and the decompression process. In symmetric applications there is always production of video information either via a camera (video mail, videotelephone) or by editing prerecorded material. One major class of symmetric application is the generation of material for playback-only applications; (desktop video publishing); another class involves the use of telecommunication either in the form of electronic mail or in the form of interactive face-to-face applications. Table 4 shows the symmetric applications of digital video.

TABLE 3.

Asymmetric Applications of Digital Video

- Electronic Publishing
  - Education and Training
  - Travel Guidance
  - Videotext
  - Point of Sale
- Games
- Entertainment (movies)

TABLE 4.

Symmetric Applications of Digital Video

- Electronic Publishing (production)
- Video Mail
- Videotelephone
- Video Conferencing

Features of the Video Compression Algorithm

The requirements for compressed video on digital storage media (DSM) have a natural impact on the solution. The compression algorithm must have features that make it possible to fulfill all the requirements. The following features have been identified as important in order to meet the need of the applications of MPEG.

Random Access. Random access is an essential feature for video on a storage medium whether or not the medium is a random access medium such as a CD or a magnetic disk, or a sequential medium such as a magnetic tape. Random access requires that a compressed video bit stream be accessible in its middle and any frame of video be

decodable in a limited amount of time. Random access implies the existence of access points, i.e., segments of information coded only with reference to themselves. A random access time of about 1/2 second should be achievable without significant quality degradation.

Fast Forward/Reverse Searches. Depending on the storage media, it should be possible to scan a compressed bit stream (possibly with the help of an application-specific directory structure) and, using the appropriate access points, display selected pictures to obtain a fast forward or a fast reverse effect. This feature is essentially a more demanding form of random accessibility.

Reverse Playback. Interactive applications might require the video signal to play in reverse. While it is not necessary for all applications to maintain full quality in reverse mode or even to have a reverse mode at all, it was perceived that this feature should be possible without an extreme additional cost in memory.

Audio-Visual Synchronization. The video signal should be accurately synchronizable to an associated audio source. A mechanism should be provided to permanently resynchronize the audio and the video should the two signals be derived from slightly different clocks. This feature is addressed by the MPEG-System group whose task is to define the tools for synchronization as well as integration of multiple audio and video signals.

Robustness to Errors. Most digital storage media and communication channels are not error-free, and while it is expected that an appropriate channel coding scheme will be used by many applications, the source coding scheme should be robust to any remaining uncorrected errors; thus catastrophic behavior in the presence of errors should be avoidable.

Coding/Decoding Delay. As mentioned previously, applications such as videotelephony need to maintain the total system delay under 150 ms in order to maintain the conversational, "face-to-face" nature of the application. On the other hand, publishing applications could content themselves with fairly long encoding delays and strive to maintain the total decoding delay below the "interactive threshold" of about one second. Since quality and delay can be traded-off to a certain extent, the algorithm should perform well over the range of acceptable delays and the delay is to be considered a parameter.

Editability. While it is understood that all pictures will not be compressed independently (i.e., as still images), it is desirable to be able to construct editing units of a short time duration and coded only with reference to themselves so that an acceptable level of editability in compressed form is obtained.

Format Flexibility. The computer paradigm of "video in a window" supposes a large flexibility of formats in terms of raster size (width, height) and frame rate.

Cost Tradeoffs. All the proposed algorithmic solutions were evaluated in order to verify that a decoder is implementable in a small number of chips, given the technology of 1990. The proposed algorithm also had to meet the constraint that the encoding process could be performed in real time.

Overview of the MPEG Compression Algorithm

The difficult challenge in the design of the MPEG algorithm is the following: on one hand the quality requirements demand a very high compression not achievable with intraframe coding alone; on the other hand, the random access requirement is best satisfied with pure intraframe coding. The algorithm can satisfy all the requirements only insofar as it achieves the high compression associated with interframe coding, while not compromising random access for those applications that demand it. This requires a delicate balance between intra- and interframe coding, and between recursive and nonrecursive temporal redundancy reduction. In order to answer this challenge, the members of MPEG have resorted to using two interframe coding techniques: predictive and interpolative.

The requirements on the MPEG video compression algorithm have been derived directly from the likely applications of the standard.

The MPEG video compression algorithm [3] relies on two basic techniques: block-based motion compensation for the reduction of the temporal redundancy and transform domain-(DCT) based compression for the reduction of spatial redundancy. Motion-compensated techniques are applied with both causal (pure predictive coding) and noncausal predictors (interpolative coding). The remaining signal (prediction error) is further compressed with spatial redundancy reduction (DCT). The information relative to motion is based on 16 × 16 blocks and is transmitted together with the spatial information. The motion information is compressed using vari-

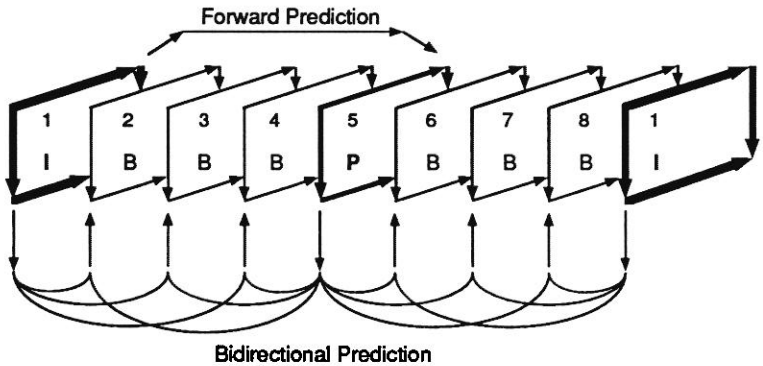


FIGURE 2.
Interframe Coding

able-length codes to achieve maximum efficiency.

Temporal Redundancy Reduction

Because of the importance of random access for stored video and the significant bit-rate reduction afforded by motion-compensated interpolation, three types of pictures are considered in MPEG.[2] Intrapictures (I), Predicted pictures (P) and Interpolated pictures (B—for bidirectional prediction). Intrapictures provide access points for random access but only with moderate compression; predicted pictures are coded with reference to a past picture (Intra- or Predicted) and will in general be used as a reference for future predicted pictures; bidirectional pictures provide the highest amount of compression but require both a past and a future reference for prediction; in addition, bidirectional pictures are never used as reference. In all cases when a picture is coded with respect to a reference, motion compensation is used to improve the coding efficiency. The relationship between the three picture types is illustrated in Figure 2. The organization of the pictures in MPEG is quite flexible and will depend on application-specific parameters such as random accessibility and coding delay. As an example in Figure 2, an intracoded picture is inserted every 8 frames, and the ratio of interpolated pictures to intra- or predicted pictures is three out of four.

[2]In addition to the three picture types mentioned in the text, an additional type "DC-picture" has been defined. The DC-picture type is used to make fast searches possible on sequential DSMs such as tape recorders with a fast search mechanism. The DC-picture type is never used in conjunction with the other picture types.

Motion Compensation.

Prediction. Among the techniques that exploit the temporal redundancy of video signals, the most widely used is motion-compensated prediction. It is the basis of most compression algorithms for visual telephony such as the CCITT standard H.261. Motion-compensated prediction assumes that "locally" the current picture can be modeled as a translation of the picture at some previous time. Locally means that the amplitude and the direction of the displacement need not be the same everywhere in the picture. The motion information is part of the necessary information to recover the picture and has to be coded appropriately.

Interpolation. Motion-compensated interpolation is a key feature of MPEG. It is a technique that helps satisfy some of the application-dependent requirements since it improves random access and reduces the effect of errors while at the same time contributing significantly to the image quality.

In the temporal dimension, motion-compensated interpolation is a multiresolution technique: a subsignal with low temporal resolution (typically 1/2 or 1/3 of the frame rate) is coded and the full-resolution signal is obtained by interpolation of the low-resolution signal and addition of a correction term. The signal to be reconstructed by interpolation is obtained by adding a correction term to a combination of a past and a future reference.

Motion-compensated interpolation (also called bidirectional prediction in MPEG terminology) presents a series of advantages, not the least of which is that the compression obtained by interpolative coding is very high. The other advantages of bidirectional prediction (temporal interpolation) are:

- It deals properly with uncovered areas, since an area just uncovered is not predictable from the past reference, but can be properly predicted from the "future" reference.
- It has better statistical properties since more information is available: in particular, the effect of noise can be decreased by averaging between the past and the future reference pictures.
- It allows decoupling between prediction and coding (no error propagation).
- The trade-off associated with the frequency of bidirectional pictures is the following: increasing the number of B-pictures between references decreases the correlation of B-pictures with the references as well as the correlation between the references themselves. Although this trade-off varies with the nature of the video scene, for a large class of scenes it appears reasonable to space references at about 1/10th second interval resulting in a combination of the type I B B P B B P B B .. I B B P B B.

Motion Representation, Macroblock. There is a trade-off between the coding gain provided by the motion information and the cost associated with coding the motion informa-

tion. The choice of 16 × 16 blocks for the motion-compensation unit is the result of such a trade-off, such motion-compensation units are called Macroblocks. In the more general case of a bidirectionally coded picture, each 16 × 16 macroblock can be of type Intra, Forward-Predicted, Backward-Predicted or Average. As expressed in Table 5, the expression for the predictor for a given macroblock depends on reference pictures (past and future) as well as the motion vectors: $\bar{x}$ is the coordinate of the picture element, $\overline{mv}_{01}$ the motion vector relative to the reference picture I_0, $\overline{mv}_{21}$ the motion vector relative to the reference picture I_1.

The motion information consists of one vector for forward-predicted macroblocks and backward-predicted macroblocks, and of two vectors for bidirectionally predicted macroblocks. The motion information associated with each 16 × 16 block is coded differentially with respect to the motion information present in the previous adjacent block. The range of the differential motion vector can be selected on a picture-by-picture basis, to match the spatial resolution, the temporal resolution and the nature of the motion in a particular sequence—the maximal allowable range has been chosen large enough to accommodate even the most demanding situations. The differential motion information is further coded by means of a variable-length code to provide greater efficiency by taking advantage of the strong spatial correlation of the motion vector field (the differential motion vector is likely to be very small except at object boundaries).

Motion Estimation. Motion estimation covers a set of techniques used to extract the motion information from a video sequence. The MPEG syntax specifies how to represent the motion information: one or two motion vectors per 16 × 16 subblock of the picture depending on the type of motion compensation: forward-predicted, backward-predicted, average. The MPEG draft does not specify how such vectors are to be computed, however. Because of the block-based motion representation however, block-matching techniques are likely to be used; in a block-matching technique, the motion vector is obtained by minimizing a cost function measuring the mismatch between a block and each predictor candidate. Let M_i be a macroblock in the current picture I_c, v the displacement with respect to the reference picture I_r, then the optimal displacement ("motion vector") is obtained by the formula:

$$v_i^* = \min_{\bar{x} \in V}{}^{-1} \sum_{\bar{x} \in M_i} D[I_c(\bar{x}) - I_r(\bar{x} + \bar{v})]$$

where the search range V of the possible motion vectors and the selection of the cost function D are left entirely to the implementation. Exhaustive searches where all the possible motion vectors are considered are known to give good results, but at the expense of a very large complexity for large ranges: the decision of tradeoff quality of the motion vector field versus complexity of the motion estimation process is for the implementer to make.

The freedom left to manufacturers... means the existence of a standard does not prevent creativity and inventive spirit.

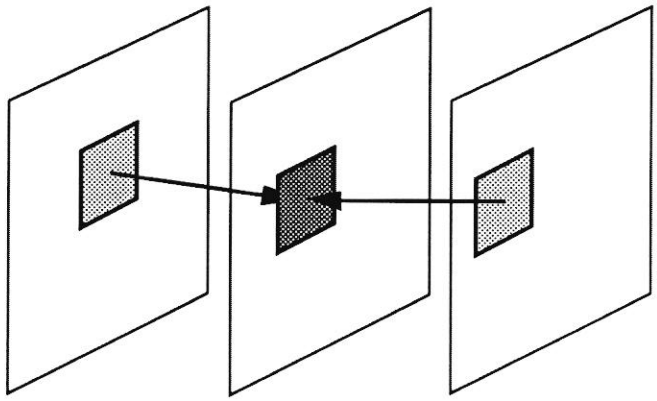
Spatial Redundancy Reduction

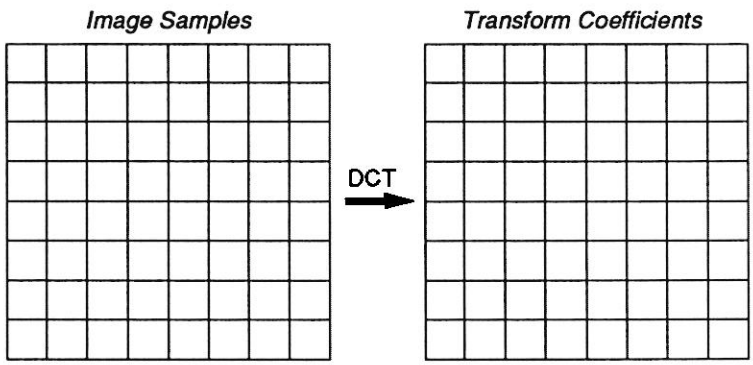
Both still-image and prediction-error signals have a very high spatial redundancy. The redundancy reduction techniques usable to this effect are many, but because of the block-based nature of the motion-compensation process, block-based techniques are preferred. In the

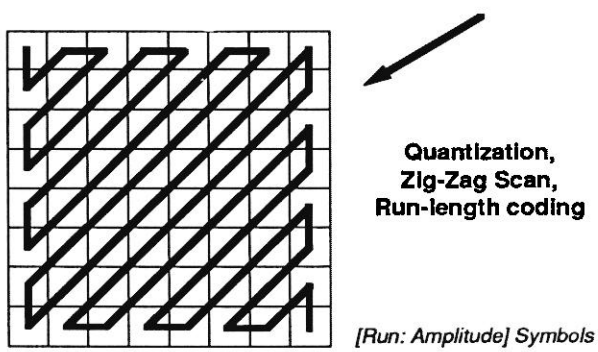
TABLE 5.

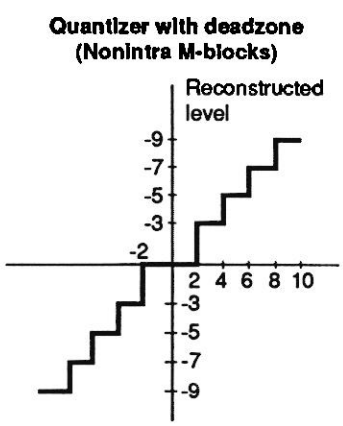
Prediction Modes for Macroblock in B-Picture

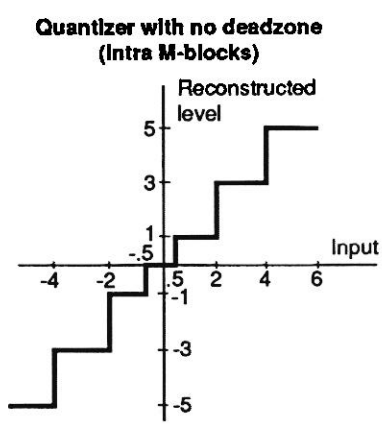
Macroblock Type	Predictor	Prediction Error
Intra	$\hat{I}_1(\bar{x}) = 128$	$I_1(\bar{x}) - \hat{I}_1(\bar{x})$
Forward Predicted	$\hat{I}_1(\bar{x}) = \hat{I}_0(\bar{x} + \overline{mv}_{01})$	$I_1(\bar{x}) - \hat{I}_1(\bar{x})$
Backward Predicted	$\hat{I}_1(\bar{x}) = \hat{I}_2(\bar{x} + \overline{mv}_{21})$	$I_1(\bar{x}) - \hat{I}_1(\bar{x})$
Average	$\hat{I}_1(\bar{x}) = \frac{1}{2}[\hat{I}_0(\bar{x} + \overline{mv}_{01}) + \hat{I}_2(\bar{x} + \overline{mv}_{21})]$	$I_1(\bar{x}) - \hat{I}_1(\bar{x})$











field of block-based spatial redundancy techniques, transform coding techniques and vector quantization coding are the two likely candidates. Transform coding techniques with a combination of visually weighted scalar quantization and run-length coding have been preferred because the DCT presents a certain number of definite advantages and has a relatively straightforward implementation; the advantages are the following:

- The DCT is an Orthogonal Transform:
  Orthogonal Transforms are filter-bank-oriented (i.e., have a frequency domain interpretation).
  Locality: the samples on a 8 × 8 spatial window are sufficient to compute 64 transform coefficients (or subbands).
  Orthogonality guarantees well-behaved quantization in subbands.
- The DCT is the best of the orthogonal transforms with a fast algorithm, and a very close approximation to the optimal for a large class of images.
- The DCT basis function (or subband decomposition) is sufficiently well-behaved to allow effective use of psychovisual criteria. (This is not the case with "simpler" transform such as Walsh-Hadamard.)

In the standards for still image coding (JPEG) and for visual telephony (CCITT H.261), the 8 × 8 DCT has also been chosen for similar reasons. The technique to perform intraframe compression with the DCT is essentially common in

FIGURE 3.
Motion-Compensated Interpolation

FIGURE 4.
Transform Coding, Quantization and Run-Length Coding

FIGURE 5.
Quantizer Characteristics for Intra- and Non-Intra Blocks (stepsize = 2)

the three standards and consists of three stages: computation of the transform coefficients; quantization of the transform coefficients; and conversion of the transform coefficients into {run-amplitude} pairs after reorganization of the data in a zigzag scanning order (see Figure 4).

Discrete Cosine Transform. The Discrete Cosine Transform has inputs in the range [−255, 255] and output signals in the range [−2048, 2047], providing enough accuracy even for the finest quantizer. In order to control the effect of rounding errors when different implementations of the inverse transform are in use, the accuracy of the inverse transform is determined according to the CCITT H.261 standard specification [9].

Quantization. Quantization of the DCT coefficients is a key operation, because the combination of quantization and run-length coding contributes to most of the compression; it is also through quantization that the encoder can match its output to a given bit rate. Finally, adaptive quantization is one of the key tools to achieve visual quality. Because the MPEG standard has both intracoded pictures as in the JPEG standard and differentially coded pictures (i.e., pictures coded by a combination of temporal prediction and DCT of the prediction error as in CCITT Recommendation H.261), it combines features of both standards to achieve a set of very accurate tools to deal with the quantization of DCT coefficients.

Visually weighted quantization. Subjective perception of quantization error greatly varies with the frequency and it is advantageous to use coarser quantizers for the higher frequencies. The exact "quantization matrix" depends on many external parameters such as the characteristics of the intended display, the viewing distance and the amount of noise in the source. It is therefore possible to design a particular quantization matrix for an application or even for an individual sequence. A customized matrix can be stored as context together with the compressed video.

Quantization of Intra vs. Nonintra Blocks. The signal from intracoded blocks should be quantized differently from the signal resulting from prediction or interpolation. Intracoded blocks contain energy in all frequencies and are very likely to produce "blocking effects" if too coarsely quantized; on the other hand, prediction error-type blocks contain predominantly high frequencies and can be subjected to much coarser quantization. It is assumed that the coding process is capable of accurately predicting low frequencies, so that the low frequency content of the prediction error signal is minimal; if it is not the case, the intracoded block type should be preferred at encoding. This difference between intracoded blocks and differentially coded blocks results in the use of two different quantizer structures: while both quantizers are near uniform (have a constant stepsize), their behavior around zero is different. Quantizer for intracoded blocks have no deadzone (i.e., the region that gets quantized to the level zero is smaller than a stepsize while quantizers for nonintrablocks have a large deadzone). Figure 5 illustrates the behavior of the two quantizers for the same stepsize of 2.

Modified Quantizers. Not all spatial information is perceived alike by the human visual system and some blocks need to be coded more accurately than others: this is particularly true of blocks corresponding to very smooth gradients where a very slight inaccuracy could be perceived as a visible block boundary (blocking effect). In order to deal with this inequality between blocks, the quantizer stepsize can be modified on a block-by-block basis if the image content makes it necessary. This mechanism can also be used to provide a very smooth adaptation to a particular bit rate (rate-control).

The flexibility of the video sequence parameters in MPEG is such that a wide range of spatial and temporal resolution is supported.

Entropy coding. In order to further increase the compression inherent in the DCT and to reduce the impact of the motion information on the total bit rate, variable-length coding is used. A Huffman-like table for the DCT coefficients is used to code events corresponding to a pair {run, amplitude}. Only those codes with a relatively high probability of occurrence are coded with a variable-length code. The less-likely events are coded with an escape symbol followed by fixed length codes, to avoid extremely long code words and reduce the cost of implementation. The variable-length code associated with DCT coefficient is a superset of the one used in CCITT recommendation H.261 to avoid unnecessary costs when implementing both standards on a single processor.

Layered Structure, Syntax and Bit Stream

Goals. The goal of a layered structure is to separate entities in the bitstream that are logically distinct, prevent ambiguity and facilitate the

decoding process. The separation in layers supports the claims of *genericity, flexibility* and *efficiency*.

Genericity. The generic aspect of the MPEG standard is nowhere better illustrated than by the MPEG bit stream. The syntax allows for provision of many application-specific features without penalizing applications that do not need those features. Two examples of such "bit-stream customization" illustrate the potential of the syntax:

Example 1: Random access and editability of video stored on a computer hard disk. Random accessibility and easy editability require many access points; groups of pictures are of short duration (e.g., 6 pictures, 1/5 second) and coded with a fixed amount of bits (to make editability possible). The granularity of the editing units (group of pictures only coded with reference to pictures within the group) allows editability to one-fifth of a second accuracy.

Example 2: Broadcast over noisy channel. There are occasional remaining uncorrected errors. In order to provide robustness, the predictors are frequently reset and each intra and predicted picture is segmented in many slices. In addition, to support "tuning in" in the middle of the bit stream, frequent repetitions of the coding context (Video Sequence Layer) are provided.

Flexibility. The flexibility of the MPEG standard is illustrated by the large number of parameters defined in the Video Sequence Header. Table 6 shows the video sequence header. The range of those parameters is fairly large, and while the MPEG standard is focused at bit rates about 1.5 Mbits/s and resolutions of about 360 pels/line, higher resolution and higher bit rates are not precluded.

Efficiency. A compression scheme such as the MPEG algorithm needs to provide efficient management of the overhead information (displacement fields, quantizer stepsize, type of predictor or interpolator). The robustness of the compressed bit stream also depends to a large extent on the ability to quickly regenerate lost context after an error.

Layered Syntax. The syntax of a MPEG video bit stream contains six layers (see Table 7); each layer supports a definite function: either a signal-processing function (DCT, Motion Compensation) or a logical function (Resynchronization, Random access point).

Bit Stream. The MPEG syntax [3] defines a MPEG bit stream as any sequence of binary digits consistent with the syntax. In addition, the bit stream must satisfy particular constraints so that the bit stream is to be decodable with a buffer of an appropriate size. These additional constraints preclude coded video bit streams that have "unreasonable" buffering requirements. Every bit stream is characterized (at the sequence layer) by two fields: bit rate and buffer size. The buffer size specifies the minimum buffer size necessary to decode the bit stream within the context of the video buffer verifier.

TABLE 6.

Video Sequence Parameters
Picture Width
Picture Height
Pel Aspect Ratio
Frame Rate
Bit Rate
Buffer Size

Video Buffer Verifier. The video buffer verifier [3] is an abstract model of decoding used to verify that an MPEG bit stream is decodable with reasonable buffering and delay requirement—expressed in the sequence header in the fields bit rate and buffer size. The model of the video buffer verifier is that of a receiving buffer for the coded bit stream and an instantaneous decoder so that all the data for a picture is instantaneously removed from the buffer. Within the framework of this model, the MPEG Committee Draft establishes constraints on the bit stream—by way of the buffer occupancy—so that decoding can occur without buffer underflow or overflow.

Decoding Process. The MPEG draft standard defines the decoding process—not the decoder. There are many ways to implement a decoder and the standard does not recommend a particular way. The decoder structure of Figure 6 is a typical decoder structure with a buffer at the input of the decoder. The bit stream is demultiplexed into overhead information such as motion information, quantizer stepsize, macroblock type and quantized DCT coefficients. The quantized DCT coefficients are dequantized, and are input to the Inverse Cosine Transform (IDCT). The reconstructed waveform from the IDCT is added to the result of the prediction. Because of the particular nature of Bidirectional prediction, two reference pictures are used to form the predictor.

Standard and Quality

Conformance: Encoder and Decoders

Bit Stream and Decoding Process. The MPEG standard specifies a syntax for video on digital storage media and the meaning associated to this syntax: the decoding process. A decoder is an MPEG decoder if it decodes an MPEG bit stream to a result that is within acceptable bounds (still to be determined) of the one specified by the decoding process; an encoder is a MPEG encoder if it can produce a legal MPEG bit stream.

Encoders and Decoders. The stan-

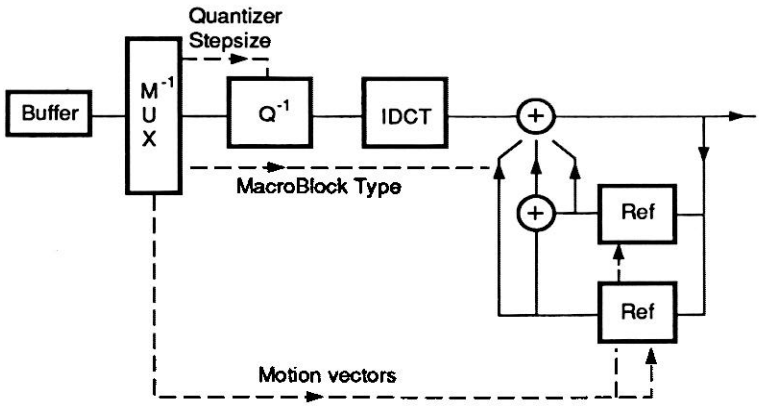


FIGURE 6.
Schematic Block Diagram of the Decoding Process

TABLE 7.

Six Layers of Syntax of the MPEG Video Bit Stream

Sequence Layer:	(Random Access Unit: Context)
Group of Pictures Layer:	(Random Access Unit: Video Coding)
Picture Layer:	(Primary Coding Unit)
Slice Layer:	(Resynchronization Unit)
Macroblock Layer:	(Motion Compensation Unit)
Block Layer:	(DCT Unit)

TABLE 8.

Parameters of the MPEG Constrained Parameter Set

Horizontal size <= 720 pels

Vertical size <= 576 pels

Total number of Macroblocks/picture <= 396

Total number of Macroblocks/second <= 396*25 = 330*30

Picture Rate <= 30 Frames/second

Bit Rate <= 1.86 Mbits/second

Decoder Buffer <= 376832 bits

TABLE 9.

Perspectives of Application of the MPEG Algorithm beyond the Constrained Parameter Set

Format	Video Parameters	Compressed Bit Rate
SIF	352 × 240 30Hz	1.2–3 Mbps
CCIR 601	720 × 486 30Hz	5–10 Mbps
EDTV	960 × 486 30Hz	7–15 Mbps
HDTV	1920 × 1080 30Hz	20–40 Mbps

dard defines only the bit-stream syntax and the decoding process; manufacturers are entirely free to make good use of the flexibility of the syntax to design very high-quality encoders and very low-cost decoders. The freedom left to manufacturers at the encoder covers such important quality factors as motion estimation, adaptive quantization and rate control. This means that the existence of a standard does not prevent creativity and inventive spirit in implementing encoders.

Resolution, Bit Rates and Quality

The quality of video compressed with the MPEG algorithm at rates of about 1.2 Mbits/s has often been compared to VHS recording [1]. The qualificative VHS-like and better than VHS have been used. The spatial resolution is limited to 360 samples per video line and the video signal at the input of the source coder has 30 frames/s non-interlaced. For most source material, artifact-free renditions can be obtained, but for the most demanding material, it is at times necessary to trade resolution for impairments.

The flexibility of the video sequence parameters in MPEG is responsible for these characteristics: a wide range of spatial and temporal resolution is supported, and it has the capability of using a large range of bit rates. It is, however, important to guarantee interoperability of equipment using MPEG, without forcing the equipment manufacturers to build very overdesigned systems. For this reason a special subset of the parameter space has been defined that represents a reasonable compromise well within the prime target of MPEG of address-

ing video coded at about 1.5 Mbits/s. A "constrained parameter bit stream" was defined [3] with the parameters shown in Table 8.

It is expected that all "MPEG" decoders be capable of decoding a constrained parameter "Core" bit stream. Or beyond the "Core" bitstream parameters, the MPEG algorithm can be applied to a wide range of video formats. It can be argued, however, that at those higher resolutions and those higher bit rates, the MPEG algorithm is not necessarily optimal since the technical trade-offs have been widely discussed mostly within the range of the "Core" bit stream (see Table 9).

A new phase of activities of the MPEG committee (ISO-IEC/JTC1/SC2/WG11) has been started to study video compression algorithm of higher resolution signals (typically CCIR 601) at bit rates up to 10 Mbits/s.

Conclusion

It is anticipated that the work of the MPEG committee will have a very significant impact on the industry and that products based on MPEG are expected as early as 1992. Indeed, the concept that a video signal and its associated audio can be compressed to a bit rate of about 1.5 Mbits/s with an acceptable quality has been proven and the solution appears to be implementable at low cost with today's technology. The consequences for computer systems and computer and communication networks are likely to open the way to a wealth of new applications loosely labeled "multimedia," because they integrate text, graphics, video, and audio. The exact impact of "multimedia" is of course yet to be determined, but is likely to be very great.

MPEG has a Committee Draft; the path to an International Standard calls for an extensive review process by the National Member Bodies[3], followed bv an intermediate stage as a Draft International Standard (DIS) and a second review process. Prior to the review process itself, it is expected that a real-time MPEG decoder will be demonstrated.

In addition to the ongoing effort, the algorithmic and technical avenues opened by MPEG are making the concepts of digital videotape recorders and digital video broadcasting more likely to occur quite soon. A second phase of work has been started in the MPEG committee to address the compression of video for digital storage media in the range of 5 to 10 Mbits/s.

[3]The membership of the ISO committee consists of National Member Bodies (ANSI in the US, . . .) who send delegations to the International Standards committee.

Acknowledgments

Now that MPEG is widely recognized as an important milestone in the evolution of digital video, the author would like to acknowledge Hiroshi Yasuda, Convenor of WG8 under whose guidance both JPEG and MPEG were started and Leonardo Chiariglione, Convenor of WG11 without whose vision there would have been no MPEG. The author would also like to thank all the technical teams that contributed proposals to the MPEG-Video test, and most of all, the people that contributed to putting together the MPEG Simulation Models and Committee Drafts. ■

References

1. Anderson, M. VCR quality video at 1.5 Mbits/s. *National Communication Forum* (Chicago, Oct. 1990).
2. Chen, C.T. and Le Gall, D.J. A Kth order adaptive transform coding algorithm for high-fidelity reconstruction of still images. In *Proceedings of the SPIE* (San Diego, Aug. 1989).
3. Coding of moving pictures and associated audio. Committee Draft of Standard ISO11172: ISO/MPEG 90/176, Dec. 1990.
4. Digital transmission of component coded television signals at 30–34 Mbits/s and 45 Mbits/s using the discrete cosine transform. CCIR-CMTT/2. Document CMTT/2. July 1988.
5. Hidaka, T., Ozawa, K. Subjective assessment of redundancy-reduced moving images for interactive applications: Test methodology and report. *Signal Processing: Image Commun. 2*, 2 (Aug. 1990).
6. JPEG digital compression and coding of continuous-tone still images. Draft ISO 10918. 1991.
7. Liou, M.L. Overview of the px64 kbps video coding standard. *Commun. ACM 34*, 4 (Apr. 1991).
8. MPEG proposal package description. Document ISO/WG8/MPEG/89–128 (July 1989).
9. Video codec for audio visual services at px64 kbits/s. CCITT Recommendation H.261, 1990.
10. Wallace, G.K. The JPEG still-picture compression standard. *Commun. ACM 34*, 4 (Apr. 1991).

CR Categories and Subject Descriptors: C.2.0 [**Computer-Communication Networks**]: General—*Data communications*; I.4.2 [**Image Processing**]: Compression (coding)—*Approximate methods*

General Terms: Design, Standardization

Additional Key Words and Phrases: MPEG, multimedia, video compression

About the Author:

DIDIER LE GALL is Director of Research at C-Cube Microsystems. He has been involved with the MPEG standardization effort since its beginning and is currently serving as chairperson of the MPEG-Video group at C-Cube Microsystems. His current research interests include signal processing, video compression algorithms and architecture of digital video compression systems.

Author's Present Address: C-Cube Microsystems, 399-A W. Trimble Road, San Jose, CA 95131. email: djl@c3.pla.ca.us

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© ACM 0002-0782/91/0400-046 $1.50