H.261 and MPEG1 - A Comparison

Tino von Roden

University of Mannheim[1]
Praktische Informatik IV, D-68131 Mannheim, Germany
E-mail: roden@rummelplatz.uni-mannheim.de

ABSTRACT This paper is motivated by the fact that the USA and Europe have chosen different approaches for implementing video data compression for applications such as video conferencing. Primarily, the former uses JPEG and MPEG (VIC, see [CaJa95], [LAKL95]), the latter uses H.261, also known as p*64 (IVS, see [Turl93]).
Within the framework of this paper, a thorough introduction into both techniques is provided. This paper focuses on the video coding part. The starting point will be the analysis of the Baseline method of JPEG on which the intra-frame coding mode for both H.261 and MPEG1 (first phase of MPEG) is based. By studying the CCITT (now, ITU-TS - International Telecommunications Union, Teleconferencing Sector) recommendation H.261, motion-compensated prediction will be presented. Motion-compensated interpolation will be explained discussing the ISO standard MPEG1. Preliminary results of intermediate tests are presented. For running tests, public domain codecs for H.261 and MPEG1 provided by Stanford University are used.

1 Introduction

Image and video (sequences of images in time) applications emerged during the last decade. These require a lot of storage capacity. Improving network techniques with higher bit rates, in particular in the area of Wide Area Networks (WAN), encouraged new applications such as video telephony and video conferencing, to name a few. The solution for these applications is data compression before storing or transmitting the data, and decompression when needed. Data compression is the reduction or elimination of redundancy in data representation in order to achieve savings in storage and communication costs [KoRa95].
The coding techniques can be categorized into entropy, source and hybrid coding [Stei94]. Entropy coding is a lossless technique (e.g., runlength coding, Huffman coding). The data are processed just as a sequence of digital values. Source coding considers the semantics of the data to be encoded, separating relevant and irrelevant data. It can be lossless (e.g., DCT, motion compensation) or lossy (e.g., quantization). Hybrid coding techniques apply source coding and entropy coding. Characteristics of compression methods are depicted in Table 1 [Mild95].

Table 1 Characteristics of Compression Methods

Characteristic	Comment
lossless	original data can be recovered precisely
lossy	not lossless
intra-frame	frames are coded independently
inter-frame	frames are coded with references to previous and/or future frames
symmetrical	encoding and decoding time comparable
asymmetric	coding time much longer than decoding time
real-time	encoding and decoding delay under 150 ms
scaleable	multiple resolutions of an image

JPEG, H.261 and MPEG are hybrid coding techniques. Using quantization (except for the JPEG lossless mode), they are lossy compression techniques. Only those result in high compression ratios. They take advantage of the fact that the human visual system is not capable of sensing every detail in a frame (picture). Picture information that is not important for the perceived picture quality is discarded.
Video coding techniques exploit two fundamental properties of video sequences - spatial redundancy within a frame and temporal redundancy among consecutive frames. Spatial redundancy means that, locally, the sample values of images typically vary slowly from point to point. This property is utilized by applying domain-based transform compression; JPEG, H.261 and MPEG1 use DCT. Temporal redundancy reflects the fact that there is motion in video sequences leading to the assumption that locally (not the same everywhere in the picture) the current picture can be modeled as a translation from a picture at some previous and/or future time. The amplitude and the direction can vary across the picture. The translations are the motion vectors.
The following chapter, which is dedicated to JPEG, will introduce the basic compression concepts that can be found both in H.261 and MPEG1. While discussing H.261 in chapter 3, one type of inter-frame coding, namely motion-compensated prediction, will be introduced. In chapter 4, this will be supplemented by the motion-compensated interpolation of MPEG1 that it uses

[1] This paper has its origin in an exchange program between the University of Mannheim, Germany, and the Arizona State University in Tempe, USA. It was finished in cooperation with the department of Praktische Informatik IV, Mannheim, and the department of Computer Science and Engineering, Tempe.

0-7803-3255-5/96 $5.00 © 1996 IEEE

TCL-1029
TTE Technology Inc. d/b/a TCL North America v. InterDigital, Inc. IPR2026-00417
AMAZON-1029
Amazon v. InterDigital, IPR2026-00195

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

beyond motion-compensated prediction. An evaluation of both standards based on test results follows (chapter 5).

2 JPEG

JPEG stands for Joint Pictures Experts Group. Joint because the development of this standard was a joint effort of CCITT and ISO [Wall92]. JPEG is a compression standard for continuous-tone still images (greyscale and color). JPEG is supposed to be a generic standard for many image applications. It offers one lossless and three lossy encoding modes: sequential, progressive, hierarchical [PeMi93]. Of particular interest to this paper is a subset of the sequential mode, the Baseline method which is by far the most implemented JPEG method, and it builds the basis for the intra-frame coding mode for H.261 and MPEG1.

The JPEG Baseline method achieves compression by first applying DCT, then quantizing the corresponding DCT coefficients, and finally the quantized DCT coefficients are further compressed by Entropy encoding (that is, more frequent events are encoded with fewer bits). The basic coding units are 8x8 pixel blocks into which the whole image is divided.

2.1 Discrete Cosine Transform

The DCT exploits the spatial redundancy property of an image based on non-overlapping 8x8 pixel blocks where the pixel intensity values are transformed into frequency values, the DCT-coefficients. These DCT coefficients are determined by decomposing the sample values into a set of waveforms of different spatial frequencies (cosine basis functions, see [Mild95]). Before applying DCT, the sample values are shifted from unsigned integers with range $[0, 2^P-1]$ to signed integers with range $[-2^{P-1}, 2^{P-1}-1]$, thus reducing the internal precision requirements in the DCT calculations. For JPEG, the shift is from [0,255] to [-128,127].

The DCT coefficients are split into 63 AC coefficients and one DC coefficient. The DC coefficient represents the average value of the 64 image samples. Since usually there exists a high correlation between the DC coefficients of adjacent blocks, the quantized DC_i coefficient is encoded as the difference from the DC_{i-1} term of the previous block, thus improving the compression. The remaining 63 AC coefficients are encoded as the difference from the DC value. Typically, the whole process results in many zero or near-zero amplitudes, and hence, in conjunction with quantization which transforms near-zero values to zero amplitudes, coding is simple and economical. The DCT operation itself is lossless, but the limited precision of physical implementations induces slight losses.

2.2 Quantization

Quantization is the second step of achieving further compression. Corresponding to the 64 DCT coefficients, a 64-element quantization table is used. The DC coefficient and each AC coefficient are divided by the corresponding quantization table entry. Then, the values are rounded. For JPEG, there can be more than one quantization table specified by the application. Quantization matrices can vary depending on the characteristics of the intended display, the viewing distance, and the amount of noise in the source.

2.3 Zig-Zag Reordering

After performing DCT and quantization, most of the block energy is stored in the top-left area of the block. Therefore, the frequency values are reordered (Figure 1) before being forwarded to the Entropy encoder. Thus, low-frequency coefficients (which are more likely to be non-zero) are placed before high-frequency coefficients (more likely to be zero).

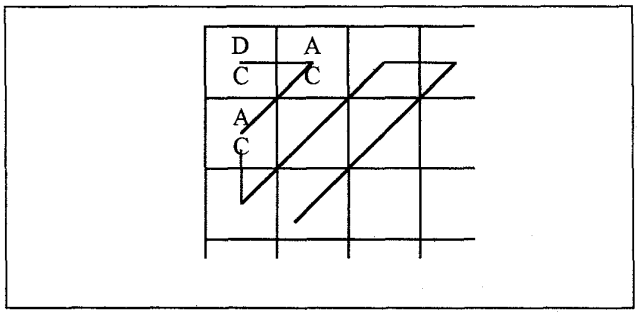


Figure 1 Zig-Zag Reordering

2.4 Entropy Encoding

Entropy encoding leads to additional lossless compression. Hereby, the quantized DCT coefficients are coded more compactly based on their statistical characteristics. For the Baseline method of JPEG, Huffman coding is employed. Usually, the input to the Entropy encoder are a few non-zero and many zero-valued coefficients. The zig-zag-reordered data stream of quantized DCT coefficients is processed in two steps. The first step transforms this data stream into an Intermediate Entropy Coding Representation [Wall92] where each non-zero AC coefficient is represented by a pair of symbols, symbol-1 (runlength, size) and symbol-2 (amplitude). Symbol-1 stores the 'runlength' (number) of preceding zero values, and the number of bits needed to encode its non-zero amplitude. Symbol-2 represents the non-zero amplitude. For the differentially coded DC coefficient just the size information of symbol-1 is needed. After that, variable-length codes are assigned. Symbol-1 is entropy encoded using Huffman tables, and symbol-2 is represented by variable-length integer value.

3 H.261

H.261 was motivated by the definition and planned deployment of ISDN (Integrated Services Digital Network). It describes the video coding and decoding methods for the moving picture component of audio visual services at the rate of p*64 kbit/s, where p takes values from 1 up to 30 (number of B channels of ISDN). A Basic Rate ISDN consisting of two B-channels and a signalling channel (2B+D) is provided by many countries, enabling video coding at 112 kbits/s and audio at 16 kbits/s [NeHa95]. A real-time encoding-decoding with less than 150 ms delay was the focus of the CCITT experts group. H.261 just specifies the decoder, the syntax and semantic of the coded bitstream. While JPEG treats each picture independently (intra-frame coding), H.261 tries to predict the current picture from the previous one (inter-frame coding) to reduce the picture information to be transmitted (motion compensation).

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

3.1 Quantization

Unlike JPEG and MPEG1, H.261 does not make use of quantization matrices. It just applies quantization factors. By changing the quantization stepsize with a common factor, the codec bit rate can be adapted to changeable bandwidths, or to meet the requirements of the Hypothetical Reference Decoder (HRD, see [NeHa95] and [H.261]).

In the case of the intra DC coefficient, there is one quantizer stepsize (value = 8). For AC coefficients, there are 31 quantizers providing even-valued stepsizes ranging from 2 to 62. There exists also a dead-zone around zero for the AC quantization, meaning that the area that gets quantized to zero is larger than the stepsize, thus decreasing sensitivity to noise in the input.

3.2 Entropy Coding

Unlike JPEG, H.261 represents a run of zeroes and the non-zero value just by a pair of values (Run, Amplitude). For the most likely occurring combinations a variable-length code entry exists. Otherwise, a fixed-length code (20 bits) is used consisting of a 6-bit Escape code, 6 bits Run, and 8 bits Amplitude (see [H.261] for more information).

3.3 Motion-Compensated Prediction

The basic concept is to look for a certain area (block) of a previous frame that matches very closely an area of the same size in the current frame (see also Figure 2). If successful, i.e., a best matching block is found, then the differences between the block intensity values (previous and current) and the motion vector (coordinate differences of the corresponding blocks in both x- and y- direction) are calculated. These differences are also termed prediction error, because they represent the deviation between the reference block and the predicted block. Usually, a search area - a fixed distance in x- and y-direction - can be defined to find a match for a block. This is done to reduce the costs of motion estimation (process of extracting motion information).

In H.261, the prediction is based on macroblocks. A macroblock consists of six 8x8 pixel blocks (four luminance, two chrominance blocks, see Table 2) whereby just the four luminance blocks are used specifying a 16x16 pixel area. H.261 takes only the closest previous frame into account for the motion compensation. The motion compensation algorithm tries to find a close match in the previous frame for the macroblock currently encoded. The search area for the motion vector is at most +/- 15 pixels in x- and y-direction. The motion compensation unit of H.261 tries to detect motion by checking macroblocks. If it cannot find a close match, it employs exactly the same coding for the macroblocks as in intra-frame coding. Otherwise, the motion vector is coded differentially. Usually, the motion vector of the macroblock to the left is used as a prediction. Then variable-length coding (VLC) is applied, and it is transmitted along with the DCT-coded prediction error (process as in intra-frame coding). The motion vector serves as the displacement information for each of the four luminance blocks of a macroblock. For the chrominance blocks, the value of the motion vector is halved in x- and y-direction. The only difference with the usual DCT-coding is that there is no DC coefficient. That is, no average value is calculated for the prediction error.

Table 2 Coding layers

Picture	one video picture (frame)
Group of Blocks	33 Macroblocks (1/12 CIF, 1/3 QCIF)
Macroblocks	16x16 Y, 8x8 Cb/Cr (motion compens.)
Blocks	8x8 pixels (coding unit for DCT)

3.4 Coding Modes

H.261 knows two types of video pictures, namely intra-pictures and inter- or predicted pictures. Intra-pictures are coded independently using DCT, quantization and run-length coding. Beyond this, H.261 tries to make use of temporal redundancy between frames by motion-compensated prediction. If the motion estimation is successful, and the prediction error is not too large, inter-frame coding is applied for the macroblocks.

3.5 Coding Algorithm for H.261

A H.261 encoder processes image files in CIF (Common Intermediate Format) or QCIF (Quarter CIF), see Table 3. The first frame to be transmitted is always an intra-coded frame. The image is coded block-wise. That is, the entire picture is divided

Table 3 Source Image Formats

	CIF	QCIF
Y	352x288	176x144
Cb, Cr	176x144	88x72

into non-overlapping 8x8 pixel blocks on which, first, the DCT (see 2.1) is applied. Then, the resulting 64 DCT coefficients are quantized and zig-zag-reordered. Finally, the Entropy coding is employed. In view of the inter-frame coding, the recently coded frame is decoded again within the encoder using Inverse Quantization and Inverse Discrete Cosine Transform (IDCT). This occurs in order to obtain exactly the same reference frame as the decoder.

For the next frame to be encoded, the last previously coded and stored frame is used for making the decision about intra- or inter-frame coding of the macroblocks (see 3.3). The motion estimation process can result in four possible decisions for the coding of a macroblock [ACDHHP93]. First, intra-coding where the original intensity values are transform coded, secondly, inter-coding without motion compensation. That is, the motion vector has the value zero. Therefore, just the prediction error (difference in intensity values) is transmitted. Thirdly, inter-coding with motion compensation and non-zero motion vector, and finally, beyond motion compensation a filter can be used to smooth the picture (to improve the image quality), which is especially useful for low bit rates. For more details, see [PVRG1]. Besides the coding of a macroblock, there is also the possibility of skipping a macroblock. Therefore, the six blocks of a macroblock are examined. If all or nearly all quantized sample values of a block are zero, it may not be encoded. If it is the same for all six blocks of an to be inter-coded macroblock without motion compensation, the whole macroblock may be skipped [NeHa95]. It is even possible to skip up to three frames between two coded frames [Mild95], but at least once every 132 transmitted frames a macroblock should be intra-coded to alleviate error-propagation (forced updating) [H.261].

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

4 MPEG1

MPEG (Moving Picture Experts Group) defines a bit-stream representation for synchronized digital video and audio, compressed to fit into a bandwidth of 1.5 Mbit/s [Gall91]. MPEG-1 consists of three parts, the MPEG video, MPEG audio and MPEG system that is responsible for multiplexing and synchronizing audio and video. About 1.1 Mbits/s are for video, 128 kbits/s for audio, and the remainder are for MPEG system. A major application of MPEG1 is the storage of audio visual information on digital storage media, such as CD ROM and DAT, which have the corresponding data retrieval speed. Developing the MPEG standard, H.261 and JPEG were taken into consideration. MPEG1 is a generic coding standard for many digital video implementations. It defines the syntax and the semantic of the video bit stream that must be decodable at the decoder end.

4.1 Quantization

Like H.261, there is one quantizer for the DC coefficient with a fixed stepsize of 8 and 31 even-valued quantizer stepsizes (2 to 62). Unlike H.261, there is no dead-zone for the AC coefficients of intra-coded blocks [Gall91]. Just the AC coefficients of inter-coded blocks are quantized with a dead-zone around zero. The quantized values range from -255 to 255. Luminance and chrominance components are quantized using the same quantization table. For more details, see [PVRG2] and [NeHa95].

4.2 Entropy Encoding

The Entropy encoding is mainly the same as in H.261, it uses runlength-level VLC. If there is no VLC entry for the run-length-level combination, 6-bit Escape, 6-bit-FLC for Runlength, an 8 bit FLC for Levels in the range from -127 to 127 and a 16-bit FLC for other Levels are used to encode the quantized DCT coefficients (see also [MPEG1]).

4.3 Motion-Compensated Prediction

The concept is the same as in H.261. MPEG1 has a (forward) prediction mode (P- or predicted picture) where a reference frame in the past is used to predict areas of the current frame based on macroblocks (see Table 4). The difference with H.261 is that the reference frame needs not to be the preceding frame, but the last I- (Intra-) or P-frame some time in the past. Furthermore, there is no limitation defined in the MPEG1 standard for the search area for possible motion vectors.

Table 4 Coding Layers

Video Sequence	independent video stream
Group of Pictures	random access unit (e.g., IBBPBB)
Picture	single frame in the video sequence
Slice	resynchronization unit, variable number of Macroblocks
Macroblock	16x16 Y, 8x8 Cb/Cr (motion comp.)
Block	8x8 pixels (coding unit for DCT)

4.4 Motion-Compensated Interpolation

Comparing H.261 and MPEG1, this is a major new function added to the video codec to achieve further (better) compression. Interpolation is also termed bidirectionally coding or coding using the average. Like motion-compensated prediction, it also utilizes the temporal redundancy among consecutive frames. But while motion-compensated prediction uses only one reference frame (in the past), motion-compensated interpolation makes use of two reference frames, one some time in the past, and one some time in the future (Figure 2). For a given target block of the frame currently encoded, the motion estimation algorithm looks for two matching macroblocks, one in the previous and the other one in the future reference frame. Taking the average of both reference blocks, the prediction error to the target block is calculated. Like motion-compensated prediction, the prediction error and the resulting difference signal are transmitted, applying first DCT, then quantization and Entropy encoding. By contrast, there are two motion vectors, both of which specify the coordinate differences to the matching block in the previous and in the future reference block.

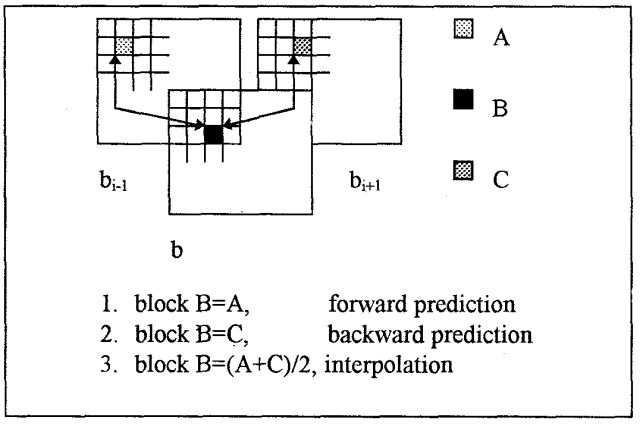


Figure 2 Motion Compensation

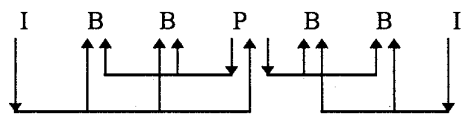
4.5 Coding Modes

There are four different picture types in the MPEG1 standard, namely, I-, P-, B- and DC-frames. The very first picture of the video stream must be an intra-coded picture (I-picture). Beyond the I-picture, there exist three prediction-based or inter-coded picture types. Forward-predicted pictures are also termed P-pictures. The reference herefore is the last I- or P-picture. Another type of prediction is backward prediction with a reference to a future I- or P-picture (backward prediction is a choice within the interpolation mode, hence assigned to B-pictures). These are completed by the bi-directionally-coded frame (B-picture) where the references are the previous and the next I- or P-picture. B-pictures themselves can never be used as a reference. Another picture type is just briefly mentioned, the D- or DC-picture that ignores AC coefficients. It can never be used with the other picture types.

The reference pictures must be transmitted first. Therefore, the transmission order and the display order are different as depicted in Figure 3. At the beginning, there is always an I-picture. For the first two B-pictures the first I- and the first P-picture serve as a reference. At the same time, the first I-picture is also the reference for the P-picture. Since the I-picture is coded independently without any references to other pictures, these pictures simply yield a moderate compression. P-pictures require about 1:3 of the data of an I-picture. B-pictures require

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

2:1 to 5:1 less data than a P-picture [Bufo94]. Some aspects for the picture sequence structure are time constraints, buffer constraints, necessity of random-access points, or the desired compression.



Transmission order: **IPBB**IBB
(highlighted frames are a Group of Pictures, from an I-picture on until the next but except the I-picture)

Figure 3 MPEG1-Transmission Order

4.6 Coding Algorithm for MPEG1

MPEG1 processes many different image formats. Two Standard Interchange Formats (SIF) are defined, namely PAL (Phase Alternation Line) and NTSC (National Television Standards Committee) (Table 5). Restrictions are set by the Constraint Parameter Bit-Stream (CPS) [NeHa95]. The first frame to be transmitted is always an intra-coded frame. The image is coded block-wise. That is, the entire picture is divided into non-overlapping 8x8 pixel blocks on which, the DCT (see 2.1) is first applied. Then, the resulting 64 DCT coefficients are quantized, and zig-zag-reordered. Finally, the Entropy coding is employed.

Table 5 Source Image Formats

	Y	Cb/Cr	Frames/s
SIF: PAL	352x288	176x144	25 (50 Hz)
NTSC	352x240	176x120	29,97 (60Hz)

Other frequencies, picture widths and heights are possible. See also [PVRG2].

Unlike H.261, the reference frame is not necessarily the last previously coded frame. The reference frame or frames are decoded again within the encoder using Inverse Quantization and Inverse Discrete Cosine Transform (IDCT). As with H.261, just the four luminance blocks of a macroblock are involved in the motion estimation process.

The use of the motion estimation process depends on the structure of the video sequence. It is up to the application (user) to determine a picture pattern. That is, the structure of a group of pictures defined by the appearance and order of B- and P-frames between two I-frames, for example, IBBPBBI. Motion estimation must fulfill its task when B- or P-pictures appear in the picture pattern. This is the most common case, since these achieve the tremendous compression. The MPEG1 standard does not specify any limitations for coordinate values of the motion vector. If a P-frame must be encoded, the motion estimation unit uses the last I- or P-frame as the reference frame, and it tries to find a good match to predict macroblock units of the current frame. If successful, a motion vector and the resulting prediction error are encoded for the macroblock. Otherwise, the macroblock is simply intra-coded. The decision process for B-frames is more complex. Four possibilities must be taken into account. Forward or backward prediction, interpolation. or, if these are not applicable, the macroblock is intra-coded. If interpolation is applied, two reference frames must be available, the closest I- or P-frame in the past and in the future, yielding two motion vectors and one prediction error block. The reference frames for P- and B-pictures must be transmitted first.

By adapting the quantizer, the Coding Control is able to change the output bit rate to satisfy the requirements of the Video Buffer Verifier (VBV), and to improve the picture quality as much as possible. In case of a transmission error, the remaining macroblocks of a slice are replaced by skipped macroblocks until a new slice starts. Unlike H.261, no picture skipping is allowed [NeHa95].

5 Performance Tests of H.261 and MPEG1

There is a large variety of commercial products available for both video codec standards. Software implementations (as used for the tests) are usually more flexible than hardware implementations, since more parameters can be changed to influence the coding of the video stream. A major drawback is the lower encoding/decoding time. Regarding time measurements, relative interpretation is appropriate. Real-time considerations are not reasonable. For a comprehensive evaluation, both the motion estimation algorithm and the data rate control algorithm should be considered. They are just as little part of the considered standards as the applied DCT algorithm. Nevertheless, they do have impact on coding time, quality and compression. With respect to video conferencing, the goal is to elaborate the strengths and weaknesses that are inherent to both standards. The conceptions of the two standards are of interest. This performance comparison is based on tests run by Torsten Milde, University of Mannheim [Mild95]. Other codec implementations may yield different results.

5.1 Test Environment

As the basis for the tests served the software codecs PVRG- P64 1.1 [PVRG1] and PVRG-MPEG 1.1 [PVRG2] provided by the Stanford University. The aim of the codecs is to comply to the corresponding standards, but not optimized speed. Three videos, shown on cable TV in Germany, were used for the tests. 'Bunny' (PRO7), 'Clip' (MTV) and 'Kicker' (DSF). 'Bunny' (cartoon) and 'Clip' (music clip) have the complexity of video conferencing applications. There is motion only in a few areas of a frame simultaneously. Whereas 'Kicker' (soccer game) contains a lot of motion. Players move around, and the camera angles change. The color sample precision is 24 bit. Each clip consists of 30 frames (digitized with 15 frames/s). The common picture resolution is 352 by 288 pixels with 25 frames/s (PAL). A DEC/alpha workstation with 200 MHz CPU was used for the tests. A SPARCstation with a 24-bit display was used for the visual tests to avoid the impact of dithering on the quality.

5.2 Test Criteria

The criteria used for the evaluation of the codecs are the subjective and the quantitative picture quality, the encoding time as well as the compression rate and the data rate.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

For the quantitative picture quality, the Signal-To-Noise-Ratio (SNR) is used. The SNR represents the ratio weighted on the squared value of the original signal [PVRG1], [PVRG2].
There are four levels of subjective quality:

(Q1) no differences to the original can be determined,
(Q2) differences to the original can be determined by direct comparison,
(Q3) acceptable quality, but differences determinable,
(Q4) non-acceptable differences to the original.

The compression rate is $\frac{f_{coded}}{f_{orig}}$, where

$$f_{orig} := 30 * 1.5 * 352 * 288 = 4{,}561{,}920 \text{ bytes.}$$

(Author's note: 1.5 due to the usage of YCbCr, one Cb and one Cr sample value belong to four Y samples).

5.3 Test Results With Changing Quantization Stepsize

The data rate control was deactivated to prevent bit stuffing and skipped macroblocks. The quantization stepsizes used were 1, 5, 10, 15, 20, 25 and 31 For H.261, intra coding (INTRA), motion compensation with zero-valued motion vector (INTER_0), and motion compensation with the maximum search area of 30 pixels were chosen (INTER_30). For MPEG1, four different Group of Pictures consisting of nine pictures were used:

(GOP1) = IIIIIIIII, (GOP2) = IBBBBBBBB
(GOP3) = IPPPPPPPP, (GOP4) = IBBPBBPBB

The coding of a video using a GOP with different search areas did not result in considerable quality or compression differences. Yet, the encoding time increased significantly when a search area of 15 instead of 0 pixels was used, (GOP2) 190%, (GOP3) 70% and (GOP4) 140%. This also indicates the higher computational complexity of B-frames. Therefore, a common search area of seven pixels was used. Except for MPEG1-(Q4), both H.261 and MPEG1 achieved each subjective quality level. The (DCT-) typical blocking artifacts were observed, especially with increasing quantization stepsize.

5.3.1 H.261

Results of the tests are depicted in Table 6 and Table 7 (Qual.=Quality, Q.St.=Quantization Stepsize). If not the best quality is desirable, inter-coding is superior to intra-coding. No major quality differences between (INTER_0) and (INTER_30) were observed. (INTER_0) is a good compromise between encoding time needed and the obtainable compression ratio.

Table 6 H.261-Quality and Compression

Qual.	Compression	Q.St.	Picture	Comments
(Q1)	< 1:5	5	intra	(Q1) just intra
(Q2)	< 1:9 < 1:12	10 10	intra inter	(*)
(Q3)	1:30, 'Kicker' 1:80, otherwise	20 15	inter inter	(*)

(*) (INTER_30) compression ratio 30% higher than (INTER_0)

Table 7 H.261-Encoding Time Compared to (INTER_0)

Encoding Time	INTRA	INTER_0	INTER_30
INTER_0=100%	2-3	1	6-7

5.3.2 MPEG1

Table 8 and Table 9 depict results regarding quality and encoding time, respectively. Without considering the quantization factor, the (GOP2) achieved the best compression for each quality level. The compression achieved was less than 50% of (GOP3) causing the double encoding time. The properties of (GOP3) are comparable with H.261 inter-coding (prediction). Due to the I-frames of (GOP3), the maximum compression rate of H.261 is superior to (GOP3). On the other hand, due to the I-frames, (GOP3) achieves better picture quality than H.261.

Table 8 MPEG1-Quality and Compression

Qual.	Compr.	Qu. St.	Picture	Comments
(Q1)	< 1:5	10	(GOP1)	(Q1) just intra
(Q2)	< 1:8 < 1:12	10 10	(GOP3) (GOP4)	(*)
(Q3)	1:26	20	(GOP2-4)	comparable to H.261

(*) Regarding encoding time and compression ratio, (GOP4) proved to be a good compromise between (GOP2) and (GOP3). If encoding time is unimportant, (GOP2) should be chosen.

Table 9 MPEG1-Encoding Time Compared to (GOP3)

Encoding Time	GOP1	GOP2	GOP3	GOP4
GOP3=100%	1,2	2	1	1,5

5.4 Test Results With Constant Data Rate

An adaptable quantizer stepsize provided by a buffer-proportional feedback [PVRG2], skipped macroblocks and bit stuffing enable a constant data rate. This is not specified within the standards, but it does have significant impact on picture quality and compression. The lower data rate limit is the data rate that does not cause any buffer overflow. According to the tests, the following subjective qualities can be guaranteed with the corresponding minimum SNR:

(Q3): > 27 dB, (Q2): > 33 dB, (Q1): > 37 dB

Table 10 shows the relationships between the subjective picture quality and the resulting data rates. Assuming an appropriate upper data rate limit of 1.9 Mbits/s, only the MPEG1 (GOP1) is able to obtain (Q1) quality.

Table 10 Qualities and Data Rates

Quality	Data Rate	Coding Mode	Video
(Q1)	30*64 kbits/s ≈ 1.9 Mbits/s	(GOP1)	'Bunny'
(Q2)	19*64 kbits/s ≈ 1.2 Mbits/s	H.261-inter	'Bunny', 'Clip'
	≈ 2 Mbits/s	(GOP3) (GOP4)	
	5 Mbits/s	(GOP1) H.261-inter	'Kicker'
	8 Mbits/s	(GOP2-4)	
(Q3)	6*64 kbits/s minimum	all	'Bunny', 'Clip'
	19*64 kbits/s ≈ 1.2 Mbit/s	(GOP3)(GOP4)	'Kicker'
	26*64 kbits/s ≈ 1.7 Mbits/s	H.261-inter	

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.

The results for ('Kicker'-Q2) for (GOP1) and (GOP2-4) are out of the ordinary (GOP1 is the best), because the data rate control algorithm had problems to deal with motion-intense frames.

6 Conclusion

Both standards, H.261 and MPEG1, apply motion compensation based on macroblocks. H.261 is limited to motion vectors ranging from -15 to +15 in both x- and y-directions. By contrast, for MPEG1 there is not such a limitation defined by the standard. For better compression results (see Table 11), MPEG1 uses motion-compensated interpolation besides motion-compensated prediction which is equally used by H.261 and MPEG1.

Table 11 (GOP2)-Overall Compression Ratios at (Q1)

Quantization Stepsize	1	5
Compression Ratio	< 1:3	< 1:7

Associated with motion-compensated interpolation is more computational complexity, and more delay, since a reference frame in the past and in the future is needed (see Table 10). The picture patterns used by MPEG1 improve the picture quality. The usage of reference frames (I-, P-frames) limits the error propagation. In H.261, only every 132nd frame must be intra-coded (forced update). Besides the global quantization (quantization stepsize), MPEG1 also provides spectral quantization (quantization tables). For less complex pictures, the coding of references resulted in compression ratios worse than H.261 at comparable qualities. With respect to speed, H.261 is superior to MPEG1. If quality losses are acceptable (Q3), H.261 should be used for video sequences similar to the complexity of video telephony. At CIF resolution, 25Hz and a data rate of 6*64 kbits/s, compression ratios of less than 1:80 are possible. The optional filter smoothing the pictures is advantageous, especially at low bit rates. Moreover, the H.261 standard specifies some multi-point features such as the Freeze Picture signal [NeHa95], [H.261]. This forces the decoder to stop decoding, and continually repeating the recently received picture. These features support video conferencing applications. Regarding encoding time and compression, the (GOP4) of MPEG1 is a good compromise between (GOP2) and (GOP3) for different complexities of video sequences. It was able to achieve a ratio of 1:30 for 'Kicker' at 1.2 Mbits/s (Q3). Other strengths of MPEG1 should be emphasized, such as the simple and economical way of editing videos (I-frames!), and the possibility of fast forward and backward. Due to the conception of MPEG1, it is more appropriate for real video applications.

Acknowledgment

The author would like to thank Prof. Forouzan Golshani, Arizona State University, for the encouragement to write this paper, he also kindly undertook to review this paper before the first submission. The author is grateful to Torsten Milde, University of Mannheim, for providing test results published in [Mild95]. Prof. Wolfgang Effelsberg and Ralf Keller, University of Mannheim, deserve a special word of recognition. Their support proved invaluable to the completion of the final version of this paper. Ralf Keller also reread this paper.

References

[ACDHHP93] Rangarajan Aravind, Glenn L. Cash, Donald L. Duttweiler, Hsueh-Ming Hang, Barry G. Haskell, Atui Puri, „Image and Video Coding Standards", AT&T Technical Journal, Jan./Feb. 1993

[Bufo94] John F. Koegel Buford, Multimedia Systems, Addison Wesley, 1994, pp. 165-169

[CaJa95] Steven McCanne, Van Jacobsen, „VIC: A flexible Framework for Packet Video", ACMM95, 1995, pp. 511-522

[Gall91] Didier Le Gall, „MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, Vol.34, No.4, Apr. 1991

[H.261] „Video Codec for Audiovisual Services at px64 kbit/s", CCITT Recommendation H.261, International Telecommunication Union, 1993

[KoRa95] Mario Kovac, N. Ranganathan, „JAGUAR, A fully pipelined VLSI Architecture for JPEG Image Compression standard", Proceedings of the IEEE, Vol.83, No.2, Feb. 1995

[LAKL95] Bernd Lamparter, Andres Albanese, Malik Kalfane, Michael Luby, „PET - Priority Encoding Transmission: A new robust and efficient Video Broadcast Technology", International Computer Science Institute Berkeley, CA, Aug. 1995

[Mild95] Torsten Milde, Video Compression Methods compared: JPEG, MPEG, H.261, XCCC, Wavelets, Fraktale (in German), dpunkt, Verlag für digitale Technologie GmbH, Heidelberg, 1995

[MPEG1] „Information Technology -Coding of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 Mbit/s (MPEG) - Part 2: Video", International Standard ISO/IEC IS 11172-2, 1993

[NeHa95] Arun N. Netravali, Barry. G. Haskell, Digital Pictures: Representation, Compression, and Standards, Plenum Press, 1995, pp. 511-640

[PeMi93] William B. Pennebaker, Joan L. Mitchell, JPEG still Image Data Compression Standard, Van Nostrand Reinhold, 1993, pp. 1-45

[PVRG1] Andy C. Hung, PVRG-P64 CODEC 1.1, „Documentation of the PVRG-P64 Software Codec", Stanford University, 1993

[PVRG2] Andy C. Hung, PVRG-MPEG CODEC 1.1, „Documentation of the PVRG-MPEG1 Software Codec", Stanford University, 1993

[Stei94] Ralf Steinmetz, „Data Compression in Multimedia Computing - Principles and Techniques", Multimedia Systems, Vol.1, No.4, 1994, pp. 166-172

[Turl93] Thierry Turletti, „H.261 Software Codec for Video Conferencing over the Internet", Unité de Recherche INRIA-Sophia Antipolis, Jan. 1993

[Wall92] Gregory K. Wallace, „The JPEG Still Picture Compression Standard", IEEE Transactions on Consumer Electronics, Vol.38, No.1, Feb. 1992

Biography

Tino von Roden is a student at the University of Mannheim. He is currently working on a study thesis on H.261 and MPEG1.

Authorized licensed use limited to: University of Oklahoma Libraries. Downloaded on January 09,2026 at 19:38:16 UTC from IEEE Xplore. Restrictions apply.