ABSTRACT

This article provides an overview of H.263, the new ITU-T recommendation for low-bit-rate video communication [1]. H.263 specifies a coded representation for compressing the moving picture component of audio-visual signals at low bit rates. The basic structure of the video source coding algorithm is taken from ITU-T Recommendation H.261 [2] and is a hybrid of interpicture prediction to reduce temporal redundancy and transform coding of the prediction residual to reduce spatial redundancy. The source coder can operate on five standardized picture formats: sub-QCIF, QCIF, CIF, 4CIF, and 16CIF. The decoder has motion compensation capability with half-pixel precision, in contrast to H.261 which uses full-pixel precision and employs a loop filter. H.263 includes four negotiable coding options which provide improved coding efficiency: unrestricted motion vectors, syntax-based arithmetic coding, advanced prediction, and PB-frames.

H.263: Video Coding for Low-Bit-Rate Communication

Karel Rijkse, KPN Research

There is growing interest in video coding technology and its applications over networks: video telephony and videoconferencing, security monitoring, interactive games, teleshopping, and other value-added services. The limited transmission rate available on the general switched telephone network (GSTN) and on wireless networks presents a significant challenge to digital video communications. With state-of-the-art V.34 modem technology the bit rate achievable on the GSTN has increased, but currently is still limited to 33.6 kb/s. Digital wireless communication, which has gained widespread acceptance recently, is also limited in available transmission rate to a few kilobits per second. Therefore, there is an increasing interest in video coding at such low bit rates.

Several international standards have recently been adopted for video compression, each serving a different type of application: Joint Photographic Experts Group (JPEG), International Telecommunications Union — Telecommunications Standards Sector (ITU-T) H.261, Motion Picture Experts Group type 1 (MPEG1) and MPEG2. Although the general source model used in these standardized coding algorithms provides only a basic and incomplete description of video scenes in general, very good picture quality is obtained at several megabits per second; picture quality acceptable for some applications is reached at 64 kb/s. However, below 64 kb/s these algorithms lead to annoying blocking artifacts or require operation at low frame rates, resulting in low temporal resolution and long end-to-end delay. Therefore, further coding improvements are required to reach an acceptable picture quality at these low bit rates.

The purpose of current research on low-bit-rate video coding is to find new coding techniques that improve video quality considerably. Future multimedia services may require new functionalities for coding algorithms in future standards.

This article first describes the scope and objectives of H.263 and describes its relation to International Organization for Standardization (ISO)-MPEG4. The most novel elements of H.263 are then discussed in some detail, and the reasoning behind some of the decisions made during the standardization process is given.

SCOPE AND OBJECTIVES

ITU-T PROJECT FOR LOW-BIT-RATE VIDEO COMMUNICATION

In the context of the development of a complete set of ITU-T recommendations for a very-low-bit-rate multimedia terminals, it was decided to develop two low-bit-rate video coding algorithms:

- H.263, based on existing technology, to be developed by 1995 (same time schedule as for the recommendations for the H.324 terminal description, multiplexing, control, and speech)
- H.263/L, the long-term algorithm, including technology with more advanced performance, to be developed by 1998

The objective for H.263 was to provide significantly better picture quality than the existing ITU-T algorithm for video compression, H.261. Due to the short schedule, H.263 is based on existing technology. The objective for H.263/L is to provide considerably better picture quality than H.263 and improved error resilience.

The relationship between H.263 and H.263/L needs explicit attention. Even when H.263/L replaces H.263 in the future, the latter may still function as a fallback mode. In addition to new techniques like shape coding, known techniques such as discrete cosine transform (DCT) and block motion compensation may still be used in H.263/L. Furthermore, the picture formats for both standards may be the same.

When the development of the very-low-bit-rate multimedia terminal was underway, it was felt that an earlier adaptation of H.324 for mobile networks was needed. Therefore, a parallel work item on a mobile version was established early in 1995. Recommendations for mobile extensions to the very-low-bit-rate multimedia terminal are now expected by 1996–1997. This work may also result in proposals for extensions to H.263 or new proposals for H.263/L.

DESIGN CONSIDERATIONS FOR H.263

At the start of the H.263 work in ITU, there was no clear consensus among the experts about the scope and objectives. Several complete algorithms were proposed, some based on products already on the market. Important discussion items were the trade-off between complexity and performance and, related to that, picture sizes. Some suggested that H.263 should use the compression algorithm of H.261 as is since they felt a significant breakthrough in picture quality was unlikely in two years, especially with existing technology.

Although H.263 is in principle network-independent and can be used for a very large range of applications, its target application is visual telephony, and its target networks are low-bit-rate networks like the GSTN, integrated services digital network (ISDN), and wireless networks. Coding techniques that introduce a high delay and techniques that are only use-

0163-6804/96/$05.00 © 1996 IEEE

ful at very high bit rates (e.g., coding of interlaced TV signals) were not considered. During the development of H.263, the target bit rate was determined by the maximum bit rate achievable on the GSTN (at that time 28.8 kb/s). At these very low bit rates, it is important to keep the amount of transmitted overhead information very small. Other requirements used in the standardization of H.263 were:

- Use of available technology
- Low complexity (low cost)
- Interoperability and/or coexistence with other video communication standards (e.g., H.320/H.261)
- Robust operation in the presence of channel errors
- Flexibility to allow for future extensions (e.g., higher transmission speeds)
- Quality-of-service parameters, such as resolution, delay, frame rate, color performance/rendition
- Subjective quality measurements

The relative importance of these requirements may differ depending on the products or applications. The range of products using H.263 may be quite diverse, depending on the following:

- Implementation: hardware vs. software
- Network: GSTN vs. mobile networks
- Functionality: videophone or multimedia terminal
- User interface: PC-based, TV-based, or standalone videophone
- Market: consumer vs. business
- Time: 1996 or 2000

H.263 had to be as generic as possible in order to meet as many requirements as possible for all target applications and all possible types of products.

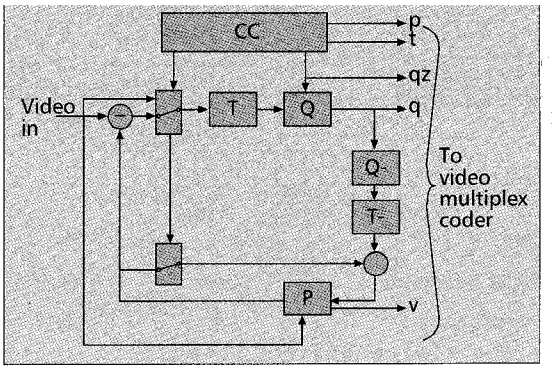
Based on all these design considerations, an efficient coding scheme was developed which gives flexibility to manufacturers to make a trade-off between picture quality and complexity. Although H.263 was optimized for bit rates below 28.8 kb/s, it was later discovered that it clearly outperformed H.261 at much higher bit rates as well (comparisons were made up to 600 kb/s). Thus, at the final stage of the H.263 standardization process, the use of H.263 at ISDN rates was also taken into account, and a few late changes were made to provide the necessary functionality at these rates (larger picture formats, considerations for multipoint, etc.).

Relation to MPEG4

MPEG4, a project of the ISO, is aimed at generic audio-visual coding systems for transmission or storage, such as multimedia communications, computers, and smart cards. The goal is a generic audio-visual coding system with acceptable consumer quality and enhanced functionality, markedly better than existing standards and currently available products. Additionally, MPEG4 is aimed at significantly enhanced flexibility and extensibility.

Generic coding will allow for easy interchange of audio-visual information between application domains, and will reduce costs because of the commonality of electronics for different applications. Therefore, it is the intention of the ITU to develop the ITU long-term video coding standard H.263/L in close cooperation with MPEG4. It is hoped that the ITU program will benefit from the generic coding algorithms of MPEG4, while MPEG4 will benefit by being compatible with the complete terminal interface standards of the ITU [3].

It is expected that MPEG4 will focus especially on new requirements, such as content-based scalability and hybrid natural and synthetic data coding. These new requirements may also be important for applications over networks and therefore for the ITU. However, as a standardization body for telecommunications, ITU considers network-related matters (error control, graceful degradation) and specific requirements for the videotelephony application (visual quality, low coding delay, interworking) as its main responsibility.



■ **Figure 1.** *Outline block diagram of the H.263 video encoder.*

At the end of the divergence phase of MPEG4, H.263 was used as a benchmark in the MPEG4 subjective tests. This allowed for a comparison of new proposals for the long-term standard with the current state-of-the-art video coding standard. For H.263 these subjective tests were seen as a performance verification, and the results of the tests indicated that H.263 performs very well. MPEG adopted H.263 as the basis for MPEG4; this can be seen as a major success for H.263.

Overview of the Coding Algorithm

Introduction

The H.263 source encoder is shown in generalized form in Fig. 1. Like ITU-T Recommendation H.261, H.263 is a hybrid of interpicture prediction to reduce temporal redundancy and DCT coding of the residual prediction error signal to reduce spatial redundancy. After transformation T, the prediction error signal is quantized with a scalar quantizer Q, and the resulting symbols are variable-length-coded and transmitted. At the decoder, the prediction error signal is reconstructed and added to the prediction, thus creating the reconstructed picture which is further stored in a picture memory P so that it can serve as a reference for the prediction of the next picture. In the encoder the same decoding operation is performed so that the encoder and decoder have the same reconstructed picture in memory P [4].

For coding efficiency, each picture is divided into macroblocks, where each macroblock consists of four luminance blocks and two spatially aligned color difference blocks. Each block consists of 8 pixels x 8 lines of luminance or chrominance. One or more macroblock rows are combined into a group of blocks (GOB) to enable quick resynchronization after transmission errors. The H.263 GOB structure differs from that of H.261; it is simpler, and GOB headers are optional, their use dependent on the encoder's trade-off between error resilience and coding efficiency.

The H.263 decoder has block motion compensation capability for improved interpicture prediction. The use of motion compensation is optional in the encoder. The main principle of block motion compensation is that interpicture prediction can be improved when the prediction blocks can be taken from different positions in the previous picture. One translational vector is transmitted per macroblock; this way, simple translational motion can be compensated for. Half-pixel precision is used for motion compensation, in contrast to H.261, where full-pixel precision and a loop filter are used. Additionally, H.263 uses a more advanced approach for motion vector prediction: the motion vector symbols are transmitted to the decoder after variable-length coding.

Several parameters may be varied in the encoder to control the bit rate of coded video. These include processing of the video signal prior to source coding, scale of the quantizer, mode

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

■ **Table 1.** *Spatial resolutions for each of the H.263 picture formats.*

selections, and picture rate. An H.263 decoder can signal its preferences for certain trade-offs between spatial and temporal resolution by external means (e.g., the H.245 protocol) [5].

A negotiable continuous presence multipoint mode is provided in which up to four independent H.263 QCIF bitstreams can be multiplexed as independent "sub-bitstreams" in one new video bitstream. Also, an optional forward error correction method is included. Use of these options should be restricted to systems in which the systems layer does not provide the same functionality (e.g., H.320) [6].

In addition to the core coding algorithm described above, H.263 includes four negotiable advanced coding options: unrestricted motion vectors, advanced prediction, PB-frames, and syntax-based arithmetic coding. The first three options are used to improve interpicture prediction; the fourth is related to lossless coding of the symbols to be transmitted. When this option is used, arithmetic coding is used instead of variable-length coding. These coding options increase the complexity of the video codec, but also significantly improve picture quality. Selection of these options is achieved by negotiation between two terminals by external means (e.g., the H.245 protocol); in this way, H.263 allows manufacturers to trade off picture quality against complexity.

Picture Formats

The source coder can operate on one of five standardized picture formats: sub-QCIF, QCIF, CIF, 4CIF, and 16CIF (Table 1). This family of CIF-based formats covers a large range of spatial resolutions. The maximum picture rate of encoders can be controlled by not transmitting a minimum number of pictures between each pair of transmitted pictures. Selection of this minimum number is achieved by negotiation between two terminals by external means (e.g., the H.245 protocol).

The CIF, 4CIF, and 16CIF picture formats are optional for encoders as well as decoders. Support of both the sub-QCIF and QCIF picture formats is mandatory for decoders. For encoders, support of only one of these formats (sub-QCIF or QCIF) is mandatory. This requirement of two mandatory formats for the decoder and only one for the encoder is a compromise between high resolution and low cost. It allows for very inexpensive products, since the more complex and therefore more expensive encoder may support only sub-QCIF. However, remote terminals are not forced into this very-low-resolution mode, since QCIF is also mandatory for the decoder. This is especially useful in multipoint conferencing and video retrieval applications.

Unrestricted Motion Vector Mode

In the default prediction mode of H.263, motion vectors are restricted such that all pixels referenced by them lie within the coded picture area. A consequence of this restriction is that often all pixels of a macroblock at the picture border will have suboptimal prediction, even when for large parts of the macroblock a much better prediction is available within the coded picture area. In the unrestricted motion vector mode this restriction is removed, and motion vectors are allowed to point outside the picture. Thus, a much better prediction can be found when only a small part of the prediction is located outside the picture and therefore is not available. For unavailable prediction pixels, the edge pixels are used instead. With the unrestricted motion vector mode a significant gain is achieved, especially for the smaller picture formats if there is motion at near the picture boundaries.

The unrestricted motion vector mode also includes support for larger motion vectors (a motion vector range of [–31.5,31.5] instead of [–15.5,16]). This is especially useful in the case of camera movement or when very large picture formats are used (4CIF or 16CIF). The extended motion vector range was added to H.263 at a very late stage of the standardization process and without any change to the bitstream syntax; only the interpretation of the transmitted motion vector codes is different when this mode is used.

Advanced Prediction Mode

In this optional mode, overlapped block motion compensation is used for the luminance component of P-pictures, resulting in a considerable subjective improvement due to the reduction of blocking artifacts. Overlapping is performed on an 8 x 8 block basis; the overlapping is four pixels deep and is only performed for the surrounding 8 x 8 blocks that have 4-connectivity with the current 8 x 8 block. In this way, each pixel in an 8 x 8 luminance prediction block is a weighted average of three prediction values. One prediction value is obtained by applying the motion vector for the current block, while the two other prediction values are obtained by applying the motion vectors of the two closest surrounding blocks. Overlapping is performed for all predicted blocks, even for noncoded blocks (these are, in fact, predicted with a zero motion vector). For chrominance, however, no overlapped prediction is applied, since the gain in performance is negligible while the extra complexity is not.

Four 8 x 8 vectors instead of one 16 x 16 vector are used for some of the macroblocks in the picture. The encoder decides which type of vectors to use; four vectors use more bits, but give better prediction. It was found that an adaptive motion block size (8 x 8/16 x 16) yielded a considerable improvement only if combined with overlapped prediction. Therefore, the adaptive motion block size is only included in this optional mode.

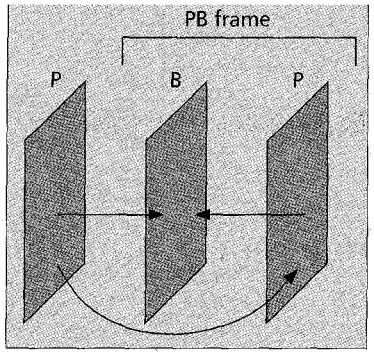
A problem with overlapped motion compensation in general is that a decoder cannot calculate the reconstruction of a macroblock before the motion vectors of the macroblocks to the right and below are known. For H.263, it was accepted that the decoder has to wait for the motion vectors of the macroblock at the right, since these vectors will arrive immediately after the information for the current macroblock. However, with the macroblock below the current one no overlapping is performed in order to reduce complexity and/or delay.

In the advanced prediction mode, motion vectors are allowed to cross picture boundaries, as is the case in the unrestricted motion vector mode. Thus, a complex situation with different kinds of motion vector restrictions at or near the borders of the picture is avoided. The extension of the motion vector range, however, is not automatically included, but is

only active if the unrestricted motion vector mode is explicitly used.

PB-Frames Mode

The main benefit of the PB-frames mode is to increase frame rate without significantly increasing the bit rate. A PB-frame consists of two pictures coded as one unit. The term "PB" comes from the names of picture types in H.262 [7], P-pictures and B-pictures. A PB-frame consists of one P-picture, which is predicted from the previous decoded P-picture, and one B-picture, which is predicted from both the previous decoded P-picture and the P-picture currently being decoded. The term "B-picture" was chosen because parts of B-pictures may be *bidirectionally* predicted from past and future pictures. The prediction process is illustrated in Fig. 2.



■ **Figure 2.** *Prediction in PB-frame mode.*

Information for the B- and P-picture is interleaved at the macroblock level in such a way that the information for each P-macroblock is directly followed by the information for the related B-macroblock. At the decoder the P-macroblock is reconstructed first; the B-macroblock is bidirectionally predicted from this reconstructed P-macroblock and the previous P-picture. When the reconstruction of the PB-frame is completed, the B-picture is displayed first and then the P-picture.

Both the forward and backward vectors for the bidirectional prediction of the B-macroblock are derived from the vector for the P-macroblock. In the calculation of these vectors, the temporal position of the B-picture relative to the previous and next P-pictures is taken into account. If these derived forward and backward motion vectors are not good enough, only one additional delta vector per B-macroblock may optionally be transmitted, which is used to adjust both the forward and backward motion vectors. This way, the additional overhead related to B-pictures is kept extremely low, while the B-pictures are normally very well predicted, especially for scenes with moderate motion.

Additional memory is needed for this technique in both the encoder and decoder to store the B-picture, and there is an increase in computational complexity due to the increase in frame rate. However, B-pictures are computationally much less expensive than P-pictures because the motion estimation can be done in a much smaller area, fewer blocks are coded, and, on average, fewer coefficients are transmitted per block. In addition, if both advanced prediction and PB-frames are used, overlapped motion compensation is used only for P-pictures and not for B-pictures.

For a given P-picture rate, use of PB-frames causes almost no extra delay: the PB-frame is displayed in the same time interval as a normal P-picture when not in PB-frames mode. There is one difference, however: the P-picture is displayed only during the last part of the time interval, since the B-picture has to be displayed first. Thus, the initial delay (related to the time the P-picture appears on the screen) is larger when the PB-frames mode is used, but the more important maximum delay (related to the time when the P-picture disappears) is the same as when this mode is not used.

Syntax-Based Arithmetic Coding Mode

Because H.263 is optimized for very low bit rates, the amount of overhead bits is very low and the VLC tables are very efficient for these bit rates. If the transmission error probability is low, GOB headers may not be transmitted, which further reduces the amount of overhead bits. With the optional syntax-based arithmetic coding (SAC) mode, the number of bits to be transmitted can be even further reduced. In this mode, all the variable-length coding/decoding operations of H.263 are replaced with arithmetic coding/decoding operations. While in the normal VLC/VLD process each symbol must be encoded into a fixed integral number of bits; this restriction is removed when arithmetic coding is applied, resulting in a reduced bit rate. Since the models with probabilities for the symbols are nonadaptive, the decoder's models cannot be corrupted by transmission errors. In addition, the synchronization capability of SAC is approximately the same as for normal variable-length coding, since the picture and GOB headers, including the start codes, are passed through the SAC as normal fixed-length codes. This means that resynchronization at the GOB level in SAC bitstreams is as easy as in variable-length-coded bitstreams. For more detailed information about arithmetic coding, refer to [8].

Conclusion

Significantly better picture quality is achieved with H.263 than with H.261, and the level of improvement depends on the video scene and coding parameters. The implementation cost of the H.263 video codec can be kept low by implementing only the baseline functionality required for interoperability. ITU-T adopted H.263 as a mandatory video coding algorithm for use in H.324 (public switched telephone network, mobile) and as an option in H.320 (ISDN), H.323 [9] (LANs, Internet) and H.310 [10] (B-ISDN). Additionally, MPEG adopted H.263 as the basis for the coming MPEG4 standard. It is expected that H.263 will be very successful and implemented in a broad range of products in the near future.

References

[1] ITU-T Rec. H.263, "Video Codec for Low Bit Rate Communication," 1996.
[2] ITU-T Rec. H.261, "Video Codec for Audio-Visual Services at 64–1920 kbit/s," 1993.
[3] R. Schaphorst and C. Reader, "Status of ITU and ISO/MPEG4 Video Coding," SPIE paper, 1994.
[4] ITU-T SG 15 Experts Group for Very Low Bitrate Visual Telephony, "Video Codec Test Model for the Near Term 5 (TMN5)," Jan. 1995.
[5] ITU-T Rec. H.245, "Control Protocol for Multimedia Communication," 1996.
[6] ITU-T Rec. H.320, "Narrowband ISDN Visual Telephone Systems and Terminal Equipment," 1993.
[7] ITU-T Rec. H.262, "Generic Coding of Moving Pictures and Associated Audio: Video," ISO/IEC 13818-2, 1995.
[8] I. Witten, R. Neal, and J. Cleary, "Arithmetic Coding for Data Compression," *Commun. ACM*, vol. 30, no. 6, June 1987, pp. 520–40.
[9] ITU-T Rec. H.323, "Visual Telephone Systems and Equipment for Local Area Networks which Povide Non-guaranteed Quality of Service," 1996.
[10] ITU-T Rec. H.310, "Broadband Audiovisual Communication Systems and Terminals," 1996.

Biography

KAREL J. RIJKSE (M.S.E.E. 1989) joined KPN Research in 1990 and has since engaged in studies on video source coding, error robustness of video signals, and multimedia applications. He was one of the people behind MUPCOS, a proposal for MPEG2-video that was evaluated in MPEG2 subjective testing. In the context of the European research program RACE, he investigated the ATM error resilience of MPEG2 in the RACE HIVITS project, and in RACE MAVT he investigated new techniques for very-low-bit-rate video coding. In ITU-T he led the work on ITU-T Recommendation H.263 and was also its editor. Currently he is working on projects in mobile multimedia and MPEG4 standardization, and a nonconversational service platform based on Recommendation H.324.