John R. Smith
IBM

The H.264 Video Coding Standard

Hari Kalva
Florida Atlantic University

The H.264 video coding standard, also known as MPEG AVC, represents the state of the art in video compression. The standard was developed recently through the joint work of the International Organization for Standardization's MPEG group (ISO/IEC/SG29/WG11) and the International Telecommunication Union's video coding experts group (VCEG, ITU-T/SG16/Q.6). H.264 is a general-purpose codec intended for applications ranging from low-bitrate mobile video applications to high-definition TV. The broad range of applications as well as a significant improvement in compression efficiency has resulted in a strong interest from the industry.

Currently, virtually all high-quality video applications (such as digital TV, digital cable, DVD, video on demand, direct satellite TV, and HDTV) use the MPEG-2 video coding standard. The use of this technology represents billions of dollars of investment in the MPEG-2 infrastructure that's already deployed and continuing to be deployed. Although MPEG-2 technology has worked well in its applications, the codec was standardized more than 10 years ago with the available algorithmic tools and hardware development constraints, such as processor speeds and memory costs. There have been significant improvements in video coding algorithms, lower memory costs, and faster processors in the past 10 years. These advances have made it possible to develop H.264 as a computationally complex, but highly efficient, video coding standard. However, H.264's deployment must overcome the MPEG-2 inertia, which will impact how the H.264 deployment progresses.

This article provides an overview of the H.264 standard. To help the reader understand the similarities and differences between it and MPEG-2 (and other hybrid video coding standards), I describe the H.264 features and compare them to functionally similar MPEG-2 features. (A complete overview of the H.264 video standard is available elsewhere.[1,2])

Editor's Note

The H.264 video coding standard was created to support the next generation of multimedia applications. H.264 improves performance over previous video coding standards, such as MPEG-2, H.263, and MPEG-4 part 2, by applying more sophisticated techniques for intraframe and interframe prediction, transform coding, entropy coding, and so on. The H.264 standard is unique in its broad applicability across a range of bit rates and video resolutions and is gaining momentum in its adoption by industry. Hari Kalva's article reviews H.264, highlights its unique features, and describes its applicability to emerging applications such as IPTV.

—John R. Smith

Hybrid video coding

Most of the video coding standards today, including H.264 and MPEG-2, are based on *hybrid video coding*—video is compressed using a hybrid of motion compensation and transform coding. These video coding algorithms compress the video data by reducing the redundancies inherent in video, which fall into four classes:

- spatial,
- temporal,
- perceptual, and
- statistical.

Figure 1 shows the various tools video coding algorithms use to reduce these redundancies. The video compression algorithms differ in which tools are used for reducing the redundancies and in the specific ways these tools are applied. (I'll highlight the differences between MPEG-2 and H.264 later.)

The video coding standards standardize the decoding process, but not the encoding process. This approach leaves room for innovation at the encoder side, which is much more complex than the decoder. The H.264 and MPEG-2 standards

Published by the IEEE Computer Society

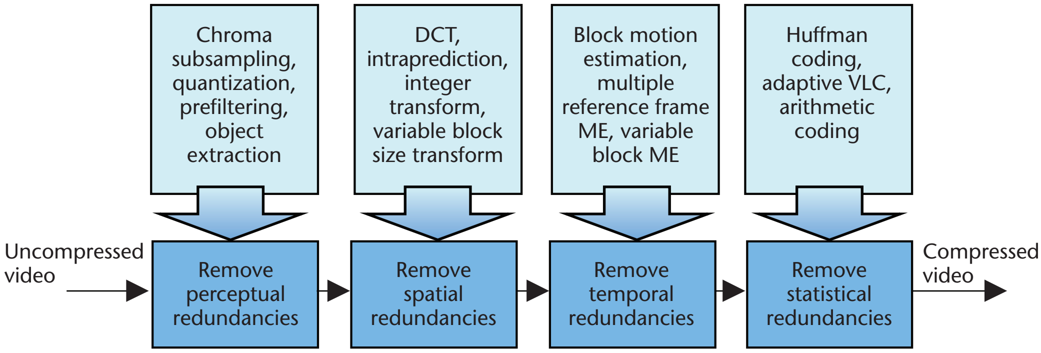

Figure 1. Functional components in video compression. ME = motion estimation and VLC = variable-length coding.

thus standardize the decoding process. Figure 2 shows an H.264 decoder's key components. This decoding structure, without the Intraprediction and Deblocking components that are new in H.264, is common to most video codecs in use today, including MPEG-2. Figure 2 also shows that the H.264 and MPEG-2 are to a large extent functionally similar (without the red boxes, the block diagram would represent an MPEG-2 decoder).

Despite that similarity, H.264 and MPEG-2 aren't compatible. However, the similarities allow low-complexity transcoding between the video formats, which is necessary in environments that support both video standards.

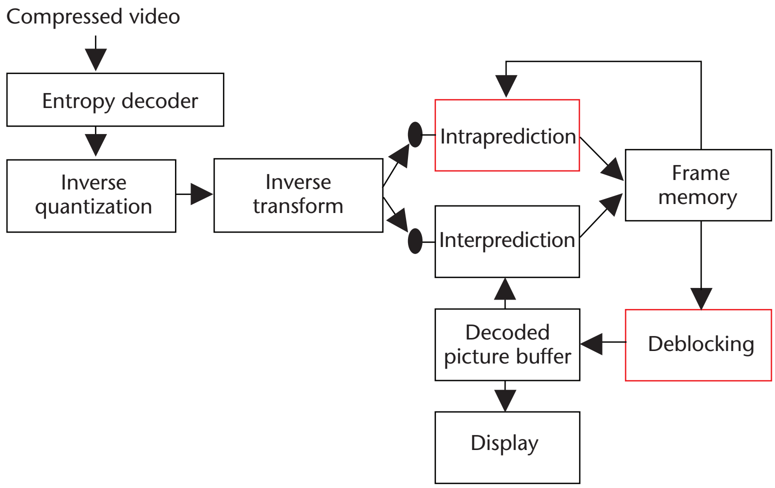

Figure 2. Components of an H.264 video decoder. The red boxes correspond to the functional components that are part of H.264 and not MPEG-2.

H.264's key features

The H.264 video coding standard is flexible and offers numerous tools to support a range of applications with low- and high-bitrate requirements. Compared with MPEG-2 video, the H.264 video gives perceptually equivalent video at one-third to one-half of the MPEG-2 bitrates. The bitrate gains are the result of a combination of encoding tools. These gains come with a significant increase in encoding and decoding complexity.[3] The H.264 video uses the same hybrid coding approach as other MPEG video standards: motion-compensated transform coding.

Even though H.264 is functionally similar to MPEG-2, H.264 introduces several new coding tools that significantly improve coding efficiency. Similar to other hybrid video coding standards, H.264 is a block-based video coding standard where the video is encoded and decoded one macroblock (MB)—a 16 × 16 block of video—at a time. The I, P, and B frames are similar to MPEG-2 frame types and two new frame types, switching-I (SI) and switching-P (SP) frames, are for streaming applications. (Intra [I] frames are coded without using other frames for prediction, predictive [P] coded frames form prediction using previously coded frames, and bipredictive [B] frames use two previously coded frames to form prediction.)

Entropy decoder

Entropy coding refers to the encoding of parameters, coefficients, or other numeric values using binary codes. Entropy encoders attempt to reduce the statistical redundancies in compressed video.

The first stage of the decoding process is when the numeric values are recovered from the binary codes. The entropy coding in H.264 uses universal variable-length coding (VLC) for all syntax elements except the quantized coefficients. We can code the coefficients using either context-adaptive VLC (CAVLC) or context-adaptive binary arithmetic coding (CABAC). The CABAC gives the most coding gains but is computationally expensive. MPEG-2 video uses fixed VLC tables for each syntactic element. These VLC tables are developed using statistical models to give a good performance for a general case. Fixed VLC tables are easy to implement but can't fully exploit statistical redundancies.

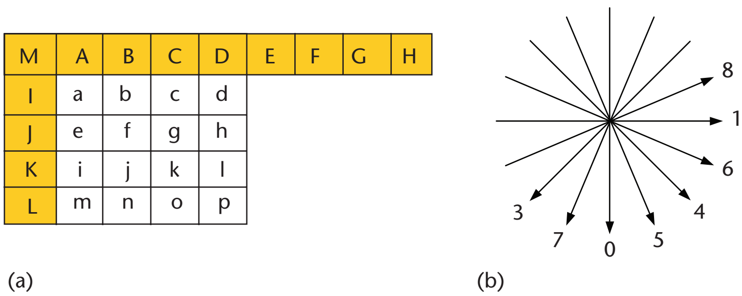

Figure 3. Intraprediction in H.264. (a) Prediction samples for a 4 x 4 block. (b) Directions for prediction modes for 4 x 4 and 8 x 8 blocks.

Intraframe prediction

H.264 exploits spatial redundancies better than MPEG-2 by allowing intraframe prediction. An MB is coded as an intra-MB when temporal prediction is impossible (for the first frame of video) or inefficient (at scene changes). The prediction for the intra-MB is determined using the neighboring pixels in the same frame because they're likely similar. An MB can be coded as one 16×16, four 8×8, or sixteen 4×4 blocks. The predictions for these blocks are determined using the weighted average of the pixels to the left of and above the current block.

Figure 3a shows the neighboring pixels used to predict a 4×4 block. The current block's pixels (lowercase alphabet in the figure) are predicted using a weighted average of the neighboring pixels (uppercase alphabet in the figure). The weights for the prediction are determined using the directional prediction modes that are specified for each block size. There are nine prediction modes for 4×4 and 8×8 block sizes and four prediction modes for the 16×16 block. Figure 3b shows the eight prediction directions allowed for 4×4 and 8×8 block sizes; the mode 2, DC mode, is the average of neighboring pixels and isn't shown in the figure.

These intraprediction modes perform well when a picture contains directional structures. The difference between the current and predicted blocks is then encoded. With the intraframe prediction, the intra-MBs can be encoded more efficiently compared to MPEG-2, which doesn't support intraframe prediction.

Interframe prediction

The interframe prediction is the traditional motion-compensated prediction supported by earlier MPEG video coding standards including MPEG-2. The H.264 standard extends this in several ways:

- variable block sizes for motion compensation,
- multiframe references for prediction,
- generalized B frame prediction,
- use of B frames as references,
- weighted prediction, and
- fractional pixel accuracy for motion vectors.

These extensions improve the coding performance and increase the complexity substantially. The motion compensation of MBs in H.264 uses variable block sizes and motion vectors with quarter-pixel resolution. An MB can be coded as one 16×16, two 16×8, two 8×16, or four 8×8 blocks. Each 8×8 block can in turn be coded as one 8×8, two 4×8, two 8×4, or four 4×4 sub-

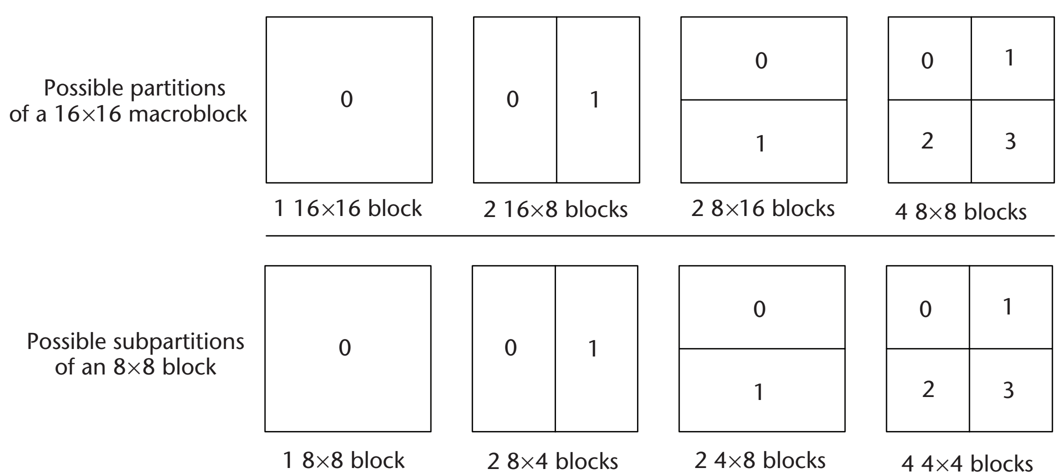

Figure 4. Subblocks of a macroblock in interprediction.

blocks (see Figure 4). An I frame can use as many as 16 different reference frames, and the actual number of reference frames is only constrained by the buffer size in a particular profile and level.

The algorithm predicts a bipredictive MB (B MB) in H.264 from any two predictions as opposed to the one forward and one backward prediction MPEG-2 video coding uses. Furthermore, a B picture can also be used as a reference picture. These motion compensation tools, though complex, contribute significant coding gains in H.264.

The H.264 standard is designed to support a range of applications, including networked and low-bitrate applications.

Transform coding

H.264 applies the transform coding to the prediction residual, as in MPEG-2. The transform it uses, however, is a 4×4 or 8×8 integer transform instead of the 8×8 discrete cosine transform (DCT) traditionally used in many video coding standards. The integer transform is designed such that the transformation involves only additions and shift operations, and no mismatch exists between the forward and inverse transforms.

This reduces the complexity significantly compared to the DCT MPEG-2 uses. The smaller the transform's size, compared to the 8×8 DCT in MPEG-2, the more it reduces the block noise in the decoded video. The transform size used in H.264 encoding is independent of the block size used for motion compensation or intraprediction.

Deblocking filter

Deblocking improves the perceptual quality and the prediction efficiency in interframes. H.264 applies a deblocking filter to the block edges, except picture boundary and block boundaries for which deblocking is turned off. The transform's size (4×4 or 8×8) determines the block size used for deblocking.

In each picture, H.264 applies the deblocking to MBs in the raster scan order: left to right and top to bottom. The vertical edges are deblocked first and then the horizontal edges. The deblocking process is applied after all of the MBs have been reconstructed. This implies that the candidate pixels used in intraprediction are the decoded pixels before the deblocking operation.

MPEG-2 doesn't support deblocking.

Network-friendly features

The H.264 standard is designed to support a range of applications, including networked and low-bitrate applications. The standard conceptually separates the video coding and networked delivery by providing a video coding layer (VCL) and a network abstraction layer (NAL). Although the VCL focuses on efficient video compression, the NAL formats the coded data for transport over various networks. The NAL layer uses a generic structure called an *NAL packet* with a syntax that identifies the VCL's higher-level syntactic elements. The network-specific applications adapt and format the NAL packets appropriately for each transport network. The standard also supports several error resilience features: parameter sets, data partitioning, flexible MB ordering, and flexible slice ordering.

MPEG-2 was designed for broadcast applications, assumes a reliable transport layer, and has limited support for error resilience.

Applications

The H.264 standard is beginning to see strong interest from industry. The standard has been adopted for HD-DVD along with MPEG-2 and VC-1. H.264's compression efficiency is particularly suitable for video services over IP networks. The lower bitrates required for H.264 enables video download services where video is distributed to the end users using simple byte serving, instead of expensive video servers, and can be played back at a later time over standard equipment.

The wide use of MPEG-2 today makes it unlikely that H.264 will replace MPEG-2 in the near term. However, we'll likely see H.264 video deployment in areas with limited or no existing infrastructure constraints. For example, telecommunications companies are using H.264 for testing and deploying video IP services because they don't have the constraints of an existing infrastructure. High-definition cameras and camcorders are another area that's likely to use H.264. With multistandard HD-DVDs expected, HD camcorders can use H.264 to allow longer recording times on the same media. Mobile and

What's Next?

While the H.264 standard is beginning to see adoption, a number of activities are already extending the core H.264 standard. Two key extensions to H.264 currently undergoing standardization are Scalable Video Coding (SVC) and Multiview Video Coding (MVC). The SVC standardization is mostly complete and will likely be final in late 2006. SVC is intended to support heterogeneous receivers by allowing spatial, temporal, and quality scalability from a single compressed bitstream.[1,2] The MVC standardization activity began in early 2006 and is specifying tools for jointly compressing the multiple views of the same scene captured using multiple cameras.[3] The MVC video has applications in Free Viewpoint TV and 3DTV.

The next leap in compression, labeled H.265, hopes to cut the H.264 bitrate in half, while maintaining the same quality, without a significant increase in complexity. The plans are preliminary and the H.265 standard is targeted for completion around 2010.

References

1. W.H. Peng et al., "Advances of MPEG Scalable Video Coding Standard," *Proc. 9th Int'l Conf. Knowledge-Based & Intelligent Information & Engineering Systems* (KES), *Int'l Workshop on Intelligent Information Hiding and Multimedia Signal Processing*, LNCS 3684, Springer, 2005, pp. 889-895.
2. S. Wenger, Y. Wang, and M. Hannuksela, "RTP Payload Format for H.264/SVC Scalable Video Coding," *J. Zhejiang Univ.—Science A*, vol. 7, no. 5, 2006, pp. 657-667.
3. ISO/IEC JTC1/SC29/WG11, "Introduction to Multi-View Video Coding," MPEG Document N7328, 2005.

PC-based IP video services can also quickly adopt H.264 because of ease of deployment.

Conclusion

The H.264 video coding standard is gaining momentum and the deployment pace will pick up as more systems are deployed. The continuing increases in processing power (Moore's law) have begun to mitigate H.264's complexity cost. The standard's broad applicability in broadcast and low-bitrate IP services makes it desirable to content providers.

It's likely that MPEG-2 and H.264 will coexist for some time, so efficient transcoding between the formats (for virtually all existing digital TVs and home receivers using MPEG-2) will be necessary. The key to transcoding is reusing the information from the decoding stage to reduce the complexity. Several researchers have already developed low-complexity algorithms to transcode MPEG-2 to H.264 and H.264 to MPEG-2.[4,5] More work in this area is still necessary (see the "What's Next?" sidebar).

Newer services that telecommunications companies are championing (such as IPTV) aren't limited by compatibility issues, so H.264 adoption will move forward there. Even if H.264 doesn't turn out to be the video format of choice for the full range of applications it promises to support, it's likely to have a significant impact on the digital video industry. **MM**

References

1. T. Wiegand et al., "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003, pp. 560-576.
2. I. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*, John Wiley & Sons, 2003.
3. H. Kalva and B. Furht, "Complexity Estimation of the H.264 Coded Video Bitstreams," *Computer J.*, vol. 48, no. 5, 2005, pp. 504-513.
4. G. Fernandez-Escribano et al., "Very Low Complexity MPEG-2 to H.264 Transcoding Using Machine Learning," to be published in *Proc. ACM Multimedia 2007*, ACM Press, 2006.
5. J. Chu et al., "H.264/MPEG-2 Transcoding Based on Personal Video Recorder Platform," *Proc. 9th Int'l Symp. Consumer Electronics*, 2005, pp. 438-440.

Readers may contact Hari Kalva at hari@cse.fau.edu.

Contact Standards editor John R. Smith at jsmith.us.ibm.com.