Standardization Trends for the HEVC Next-generation Video Coding Standard

Shohei Matsuo and Seishi Takamura

Abstract

Ten years have passed since the H.264/MPEG-4 (Motion Picture Experts Group) AVC (Advanced Video Coding) standard was established in 2003. With the recent explosive increase in high-resolution video such as 8K Super Hi-Vision (SHV) video, there is worldwide demand for a video coding standard with higher quality compression. The international standards organizations, ITU-T (International Telecommunications Union, Telecommunication Standardization Sector) and ISO/IEC (International Organization for Standardization/International Electrotechnical Commission), have been collaborating on a next-generation video coding standard called High Efficiency Video Coding (HEVC). The basic standard has been completed, and work is proceeding on an extended standard. In this article, we describe the state of HEVC standardization and give a simplified technical description.

Keywords: HEVC, ITU-T, ISO/IEC

1. Introduction

In multimedia, video contains more information than other media, so efficient video compression methods are vital. Existing international video coding standards were created as a result of collaboration between two standardization organizations: the International Telecommunication Union, Telecommunication Standardization Sector (ITU-T), and the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC). The MPEG-2 (Motion Picture Experts Group) standard is the most well-known and widespread standard among them. MPEG-2 was completed in 1995 and is used in a wide range of fields including DVDs (digital versatile disks), hard-disk recorders, and digital broadcasting. MPEG-2 brought efficient video compression, enabling video services to spread around the world, which brought recognition to the importance of video coding standards. After the MPEG-2 standard was completed, work proceeded on the newer H.264/MPEG-4 AVC (Advanced Video Coding) standard (H.264/AVC), which achieves approximately double the coding efficiency of MPEG-2, and this standard was completed in 2003. After that, the H.264/AVC FRExt extension (Fidelity Range Extension) was standardized in 2007, which included support for various input video formats and a High Profile, or HP, with higher compression rates. H.264/AVC has received much industry attention and has been adopted in a wide range of video-related applications such as Blu-ray Discs, One-Seg and HDTV (high-definition television) broadcasting, digital cameras and handy cams, video capture devices, video conferencing, and video upload sites. It is widely used in the communications, broadcasting, and storage fields, and has become one of the foundational technologies used for video services.

2. Standardization activities for next-generation video coding

2.1 Trends with the HEVC basic standard

The Joint Collaborative Team on Video Coding (JCT-VC) was inaugurated in January 2010 with members from the ISO/IEC MPEG and the ITU-T

Table 1. HEVC standardization schedule and JCT-VC meetings.

Meeting date and location			Topic	No. of participants	Approx. no. of docs contributed
1	2010/04	Dresden (Germany)	TMuC decided	188	40
2	2010/07	Geneva (Switzerland)	–	221	120
3	2010/10	Guangzhou (China)	HM1 decided	244	300
4	2011/01	Daegu (South Korea)	HM2 decided	248	400
5	2011/03	Geneva (Switzerland)	HM3 decided	226	500
6	2011/07	Torino (Italy)	HM4 decided	254	700
7	2011/11-12	Geneva (Switzerland)	HM5 decided	284	1000
8	2012/02	San Jose (USA)	HM6 decided, CD issued	255	700
9	2012/04-05	Geneva (Switzerland)	HM7 decided	241	550
10	2012/07	Stockholm (Sweden)	HM8 decided, DIS issued	214	550
11	2012/10	Shanghai (China)	HM9 decided	235	350
12	2013/01	Geneva (Switzerland)	HM10 decided, FDIS/PDAM issued	262	450
13	2013/04	Incheon (South Korea)	HM11 decided, IS issued	183	450
14	2013/07	Vienna (Austria)	HM12 decided, DAM issued	161	350

DAM: Draft Amendment
IS: international standard
PDAM: Proposed Draft Amendment

VCEG (Video Coding Experts Group), to begin serious work on the High Efficiency Video Coding (HEVC) standard as the next-generation video coding standard. First, a call for proposals of new coding methods satisfying specified requirements (bit rates, latency, etc.) was issued [1], and 27 proposals were received by the February deadline. In March, a focused, quantitative evaluation was done on the subjective quality of decoded video achieved using these methods [2]. In April 2010, the first meeting of the JCT-VC was held in Dresden, and a Test Model under Consideration (TMuC) was settled on, with algorithms in the proposals that ranked highest in the results of the evaluations described above considered as provisional HEVC reference software. Improvements were subsequently added, and official reference software—HEVC Test Model 1 (HM1)—was decided in October of that year. Ongoing technical improvements were made in subsequent JCT-VC meetings every three to four months. Specifically, coding methods proposed within particular domains were gathered, and core testing groups were organized to deliberate on them. Each proposed method was selected or rejected by evaluating them overall using various indices such as: (1) the BD-rate (Bjontegaard Distortion rate), which indicates the amount of improvement in coding efficiency; (2) the processing time for coding and decoding; (3) the memory bandwidth for hardware implementation; and (4) changes in the software source code. As the core test groups were newly established, succeeded in carrying out their tasks, and were disbanded with each meeting, technologies were selected from the various proposals to improve the performance of HEVC. The HEVC standardization process and the topics focused on at each meeting are listed in **Table 1**.

There are three main steps in establishing a standard: developing the Committee Draft (CD), the Draft International Standard (DIS), and the Final Draft of International Standard (FDIS), in that order. The CD of the HEVC standardization process was issued at the San Jose meeting in February 2012, where most of the technical specifications were decided. The DIS was issued at the Stockholm meeting in July 2012, where most of the specification details were completed. The FDIS was issued at the Geneva meeting in January 2013; it included some minor bug fixes and the completed final basic HEVC specification. The FDIS was approved through a ballot distributed to participating countries and was issued as an international standard in April 2013. As can be seen from Table 1, most of the technical specifications were decided with the CD, so just prior to this at the 7th meeting in Geneva in November/December 2011, there was very high participation from members, and many contributions were added. It is quite rare to have over 1000 contributions for a single meeting, and this indicates the unprecedented scale and amount of activity related to HEVC standardization.

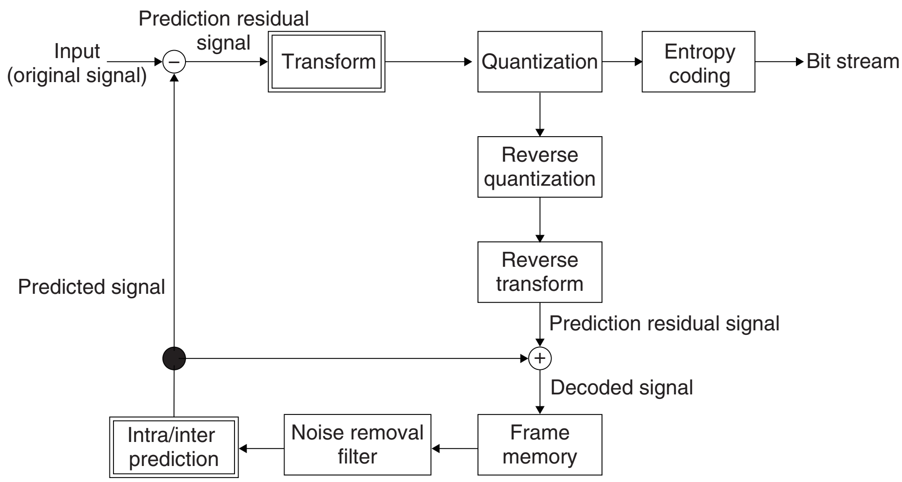

Fig. 1. Basic video coding process.

2.2 Standardization activity for HEVC extensions

The basic HEVC standard discussed in the previous section is limited to input video with a 4:2:0 color format* and does not support 4:2:2 or 4:4:4 color formats. It also supports bit depths of eight and ten bits to express each pixel, but higher bit depths are not anticipated. There is a need to handle video formats that are not supported by the basic HEVC standard, particularly in professional video recording and playback devices and in medical imaging devices, so the standardization of extended standards to support such color formats and bit depths is in progress. These extended standards go through stages corresponding to CD, DIS, and FDIS, which are called the international standard Proposed Draft Amendment (PDAM), Draft Amendment (DAM), and Final Draft Amendment (FDAM), respectively. This standardization is proceeding on a schedule delayed six months from the basic standard, as shown in Table 1. At the Vienna meeting in July 2013, the extended standard DAM was issued, and the FDAM is expected to be issued at the 16th meeting in San Jose, in January 2014.

3. Technical description of HEVC

3.1 Basic video coding framework

The basic process of video coding for H.264/AVC and HEVC is shown in **Fig. 1**. The input video is partitioned into square blocks of $n \times n$ pixels, and coding is performed in units of these blocks. These blocks are the input in Fig. 1, and are encoded into a bit stream—a binary signal array of 0's and 1's—through processes including prediction and orthogonal transformation.

Prediction is a core technique in coding. Consider the example in **Fig. 2**. In the process of encoding the frame at time t, if the frame at time t-1 has a similar pattern, then only the difference (the prediction residual) between that pattern (the prediction signal) and the source signal, is sent. This concept of using information that is already at the decoder to efficiently compress information is called prediction. Between t-1 and t, the character shown in Fig. 2 moved one block position to the left. When an object moves during the time between one image and the next, a motion vector indicating the amount of movement is sent to the decoder, indicating its predicted position. For objects that do not move, like the cloud shown in the figure, no motion vector needs to be sent. Prediction that spans images in this way is called *inter prediction* (prediction between images). In contrast, prediction done within an image is called *intra prediction* (within an image). For example, as shown by the red box enclosing the cloud in Fig. 2, the block to the left is already present at the decoder, so coding is performed using a copy of the pixels on the left as the prediction signal and taking the difference between those values and the source signal. The

* Color format: Video can be separated into a luminance signal component and chrominance signal components. If the image size for the luminance signal is $m \times n$ (where m is the number of pixels horizontally and n vertically), then the format in which the chrominance signals are half the size in the horizontal and vertical directions, or $m/2 \times n/2$, is referred to as 4:2:0; that with only the horizontal direction halved, or $m/2 \times n$, is 4:2:2, and that with neither halved, or $m \times n$ chrominance signals, is 4:4:4.

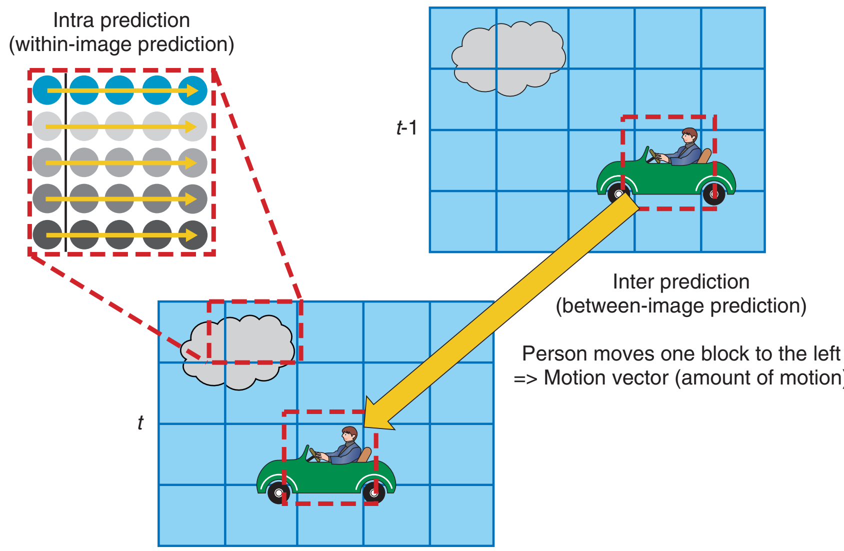

Fig. 2. Prediction overview.

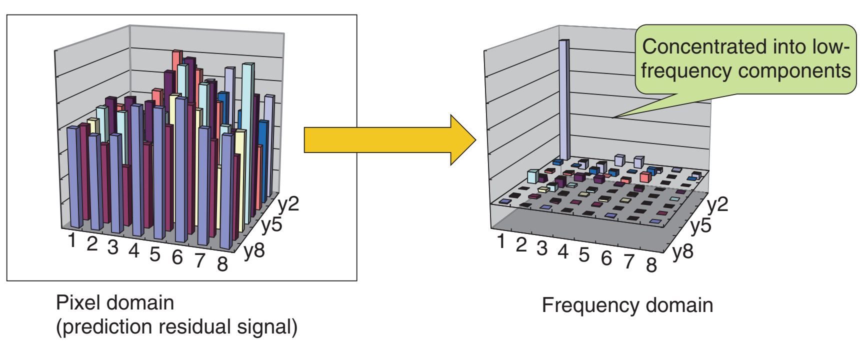

Fig. 3. Orthogonal transform overview.

intra/inter prediction in Fig. 1 is able to significantly reduce the amount of information using these processes, so it has become a core coding technology.

After the prediction process, an orthogonal transform is performed on the prediction residual signal, which converts it to a signal that can be compressed still further. When the signal is transformed into a different signal space called the frequency domain, efficient compression can be achieved by preserving the low frequency components, which express the basic structure of the pattern, and dropping the high-frequency components (**Fig. 3**).

3.2 New features of HEVC

The differences between HEVC and conventional H.264/AVC are summarized in **Table 2** [3]. H.264/AVC performs coding in units of 16 × 16 pixel blocks called macro-blocks, while HEVC does so in blocks of up to 64 × 64 pixels called coding units (CUs). CUs are partitioned into a quad-tree structure, with the leaf elements independently partitioned into prediction units (PUs), which are the unit for prediction, and transform units (TUs), which are the unit for orthogonal transforms. Processing in large units such as the 32 × 32 transforms or 64 × 64 predictions was not possible with H.264/AVC. These large units enable efficient compression, particularly for high-resolution images ((1)–(5) in Table 2 and **Fig. 4**).

For inter prediction, a scheme has been introduced that enables more accurate prediction signals to be generated. The process of creating signals for objects whose movement cannot be expressed in integer pixel

Table 2. Comparison of H.264/AVC and HEVC coding technologies.

	H.264/AVC	HEVC
(1) Coding units (CU)	16 × 16 only	8 × 8 to 64 × 64, tree structure
(2) Prediction units (PU)	4 × 4 to 16 × 16 (7 types, including rectangular)	4 × 4 to 64 × 64 (28 types, including rectangular)
(3) Transform units (TU)	4 × 4 or 8 × 8	4 × 4, 8 × 8, 16 × 16, 32 × 32
(4) Transform unit tree structure	No	Yes
(5) Transform types	DCT	DCT, DST (4 × 4, only intra), transform skipping
(6) Fractional pixel precision interpolation filter	2 or 6-tap	4-, 7-, or 8-tap
(7) Motion vector estimation	Spatial median estimation	Improved version, including time-space median
(8) Intra prediction	4 or 9 modes	35 modes, adaptive reference pixel smoothing
(9) Entropy coding	CABAC or CAVLC	CABAC
(10) Bit-depth extension	No	Internal bit-depth extension (+2 bits)
(11) Block noise removal filter	Yes	Yes (parallelizable simplified version)
(12) Coding noise removal filter	No	Yes (SAO)
(13) Parallel processing structure	Slice	Slice, WPP, tile

CABAC: Context-based Adaptive Binary Arithmetic Coding
CAVLC: Context Adaptive Variable Length Coding
DCT: discrete cosine transform
DST: discrete sine transform
SAO: sample adaptive offset
WPP: wavefront parallel processing

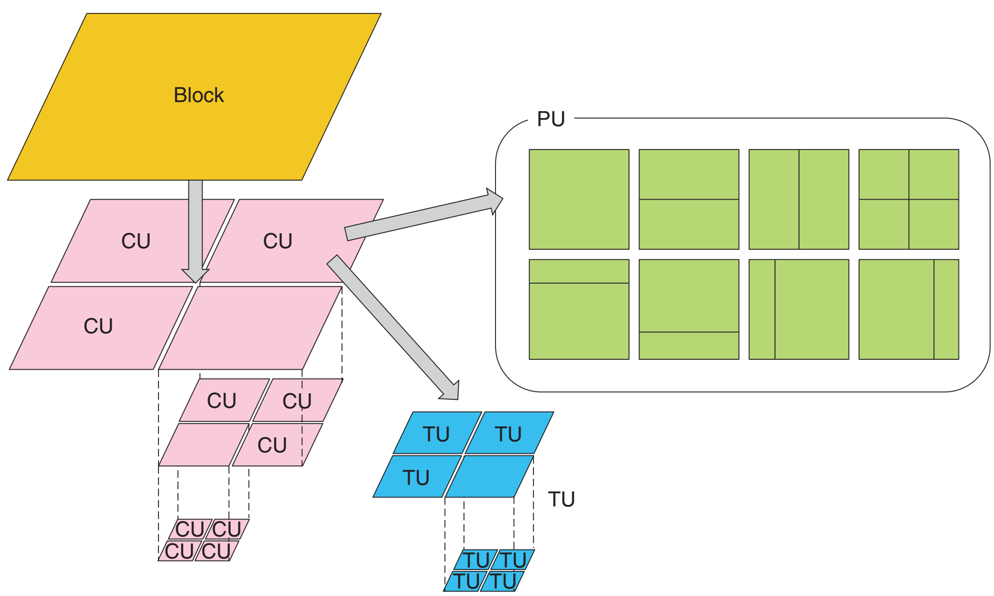

Fig. 4. HEVC flexible processing structure.

motions is called interpolation, and prediction performance was improved by increasing the number of taps on the interpolation filter to express the motion more accurately ((6) in Table 2), and by improving the prediction methods for the motion vectors themselves ((7) in Table 2). For intra prediction, performance was improved by using approaches such as greatly increasing the number of prediction directions and by applying smoothing filters adaptively when copying the source pixels, depending on the block size or direction ((8) in Table 2).

Context-based Adaptive Binary Arithmetic Coding (CABAC) is used uniformly ((9) in Table 2) as entropy coding to binary-encode the information. Also, the bit depth of the input video is increased by bit-shifting it to the left, and this value is used in the coding

Table 3. Coding performance improvement of HEVC compared to H.264/AVC, BD-rate.

	Y	U	V	YUV
Intra prediction set	−23.0%	−22.0%	−22.0%	−23.0%
Inter prediction set (for random access)	−33.5%	−34.1%	−35.0%	−33.8%
Inter prediction set (for low-delay communications)	−36.4%	−34.3%	−35.4%	−36.3%

- H.264/AVC reference software: JM18.4
- HEVC reference software: HM10.0,
  Results of performance comparison are given. (e.g.: A result of −23% indicates HEVC provided a 23% improvement in coding efficiency over H.264/AVC.)

loop and in orthogonal transform processing. This reduces rounding errors in orthogonal transformation and increases coding efficiency for intra and inter prediction ((10) in Table 2).

The source for the predicted signal is called the reference signal. This refers to the *t*-1 frame in Fig. 2 for inter prediction, and the set of pixels at the left-most edge of the enlarged red frame at time *t* for intra prediction. These reference signals are from the decoded signal, so they also have distortion known as block noise superimposed on them. With HEVC, the filters for removing this distortion have been improved ((11) in Table 2). New filters that are not in H.264/AVC have also been introduced. These filters bring the reference signal closer to the original signal, which reduces the prediction residual signal components that need to be transmitted and increases the coding efficiently significantly ((12) in Table 2). A mechanism for efficiently controlling partitioning losses and supporting parallel processing was also introduced ((13) in Table 2), in response to the need to support recent multi-core environments.

3.3 Comparison of coding performance

A comparison of HEVC and H.264/AVC performance using reference software [4], [5] and JCT-VC standard test images is presented in **Table 3** [6]. Common conditions were used for the tests, and the results are an average of BD-rates indicating the coding efficiency improvements for a total of 20 images ranging from a 2560 × 1600 pixel high-resolution image down to a 416 × 240 low-resolution image. Y is the luminance signal, U and V are chrominance signals, and YUV is a weighted average of the three. An improvement of approximately 23% was achieved with intra prediction, and approximately 35% with inter prediction (for both random access and low delay communications settings). Looking at the image quality objectively, we found that we did not double the coding efficiency, which would be an approximately 50% improvement comparing BD-rates. However, it was confirmed that when the subjective quality was the same, HEVC achieved approximately twice the coding efficiency of H.264/AVC.

4. Future developments

We have described standardization activities and given a simple technical explanation of HEVC, the next-generation video coding standard. Extensions to the HEVC standard will be completed in January 2014, and six months later, HEVC scalability extension standards should be completed. Work on extensions for 3D video is also proceeding steadily. There is strong demand for these technologies in society, so we can expect further developments in the future as well.

References

[1] ITU-T VCEG and ISO/IEC JTC 1, "Joint Call for Proposals on Video Compression Technology," ITU-T SG16/Q6 document VCEG-AM91 and ISO/IEC MPEG document N11113, Jan. 2010.

[2] V. Baroncini, J.-R. Ohm, and G. J. Sullivan, "Report of Subjective Test Results of Responses to the Joint Call for Proposals (CfP) on Video Coding Technology for High Efficiency Video Coding (HEVC)," ITU-T SG16 WP3 Q6 and ISO/IEC JTC1/SC29/WG11, JCTVC-A204, Apr. 2010.

[3] G. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649–1668, 2012.

[4] H.265/HEVC Reference Software (online). https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags

[5] H.264/AVC Reference Software (online). http://iphome.hhi.de/suehring/tml/

[6] B. Li, G. J. Sullivan, and J. Xu, "Comparison of Compression of HEVC Draft 10 with AVC High Profile," 13th JCTVC Meeting, JCTVC-M0329, Incheon, Korea, Apr. 2013.

Shohei Matsuo
Researcher, Video Coding Group, NTT Media Intelligence Laboratories.
He received the B.E. degree from Faculty of Science and Engineering and the M.E. degree from the graduate school of Global Information and Telecommunication Studies from Waseda University, Tokyo, in 2004 and 2006, respectively. He joined NTT in 2006 and engaged in R&D of video coding and its standardization activities. His current research interests include video coding technologies, image processing, and human visual systems. He has received four academic awards so far from the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Information and Television Engineers (ITE), and the Information Processing Society of Japan (IPSJ). He is a member of IEICE and ITE.

Seishi Takamura
Distinguished Technical Member, Senior Research Engineer, Supervisor in the Visual Media Coding Group of NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. degrees from the Department of Electronic Engineering, Faculty of Engineering, the University of Tokyo, in 1991, 1993, and 1996, respectively. His current research interests include efficient video coding and ultrahigh quality video coding. He has held numerous positions in the research and academic community, including serving as: Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology, and Council Member of the IEEE Tokyo Section and IEEE Japan Council. He is currently serving as IEICE Image Engineering SIG Vice Chair, Picture Coding Symposium of Japan (PCSJ) Chair, ISO/IEC JTC 1/SC 29 Japan Head of Delegates, and Director–Technical Relations of ITE. From 2005 to 2006, he was a visiting scientist at Stanford University, California, USA. He has received 26 academic awards including the ITE Niwa-Takayanagi Best Paper Award in 2002, IPSJ Nagao Special Researcher Award in 2006, and PCSJ Frontier Awards in 2004 and 2008. He is a senior member of IEEE and IEICE and a member of MENSA, ITE, IPSJ, and the Institute of Image Electronics Engineers of Japan.