Video Delivery | Video Coding | International Standardization

Technology Reports

Standardization of HEVC Video Coding Scheme Reducing Video Traffic by Half and Perspective for Mobile Services

The growing popularity of smartphones in society is rapidly increasing the use of mobile video. NTT DOCOMO has been participating in the creation of the HEVC standard, which is the latest video coding standard halving the volume of video traffic in the network, and has contributed to achieving a high compression ratio through its technology proposals and standardization support. Subjective evaluations using smartphones and tablets have shown that HEVC can achieve the same video quality as existing schemes even at half the amount of data, which demonstrates the feasibility of using HEVC in mobile terminals.

Research Laboratories
Yoshinori Suzuki
Akira Fujibayashi
Junya Takiue
DOCOMO Innovations Inc.
Frank Bossen

1. Introduction

High Efficiency Video Coding (HEVC) is the latest video coding standard jointly developed by ISO[*1]/IEC[*2] Moving Picture Experts Group (MPEG)[*3] and International Telecommunication Union-Telecommunication Standardization sector Working Party (ITU-T WP) 3/16[*4] and finalized in January 2013. The HEVC standard is supported by the technologies of institutions from more than 40 countries around the world participating in the Joint Collaborative Team on Video Coding (JCT-VC)[*5]. It achieves twice the compression efficiency of the video coding standard currently used for One Seg broadcasts, NTT DOCOMO's "dmarket" VIDEO/animation store services, and NOTTV™[*6], namely, H.264/Advanced Video Coding (AVC)[*7] [1][2] (hereinafter referred to as "H.264") [3].

The forecast is for mobile video traffic to continue growing as high-definition video becomes increasingly popular with the spread of large-screen smartphones and as Long Term Evolution (LTE) terminals began to penetrate the market. As shown in **Figure 1** [4], mobile data traffic is predicted to continue increasing driven mostly by mobile video. In this regard, the HEVC standard is capable of compressing video with twice the compression efficiency of existing schemes. It is a high-compression scheme covering a wide range of video from popular VGA[*8] mobile video to high-definition video exceeding even Full HD[*9], which is expected to become popular from here

*1 **ISO**: International Organization for Standardization; an organization for standardization in the information technology. Sets international standards for all industrial fields except electrical and telecommunication fields.
*2 **IEC**: International Electrotechnical Commission; an organization for standardization in the information technology. Sets standards in the electrical and telecommunication field.

on. Thus, in addition to halving the volume of video traffic, we can expect the widespread use of HEVC to contribute to the expansion of video services using high-definition video.

Since the beginning of HEVC standardization activities, NTT DOCOMO has been a guiding force in the creation of the HEVC standard by providing video materials and proposing requirements for standardization [5]. It has contributed to the realization of a high compression ratio through the adoption of many NTT DOCOMO technologies in HEVC. It has led HEVC standardization by serving as the chief coordinator of reference software, which is essential to achieving high compression in HEVC. NTT DOCOMO has also proposed the mobile use of HEVC at the 3rd Generation Partnership Project (3GPP) [6].

In this article, we first provide an overview of the HEVC standard and describe NTT DOCOMO's contribution to its development. We then report on the performance of HEVC anticipating its use in mobile devices and on the performance of a software decoder.

2. Features of the HEVC Standard

2.1 Basic Structure of Video Coding

Video sources have a huge amount of data but adjacent pixels and texture patterns of successive pictures are similar. A video coding scheme uses such similarities to compress the original amount of data.

The basic structure of the video coding scheme common to HEVC and H.264 is shown in **Figure 2.** In the scheme, a video source is divided into individual pictures and "block partitioning" divides a picture into small sections called a block (black-framed boxes shown in the picture in Fig. 2). Blocks are the units for coding and are input into "prediction" and "transform coding[*10]," one by one from left to right

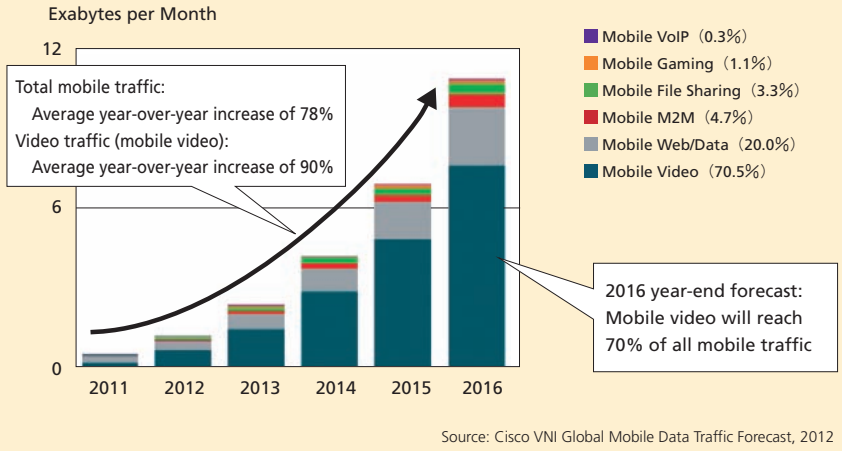

Figure 1 Global mobile data traffic forecast (2011—2016)

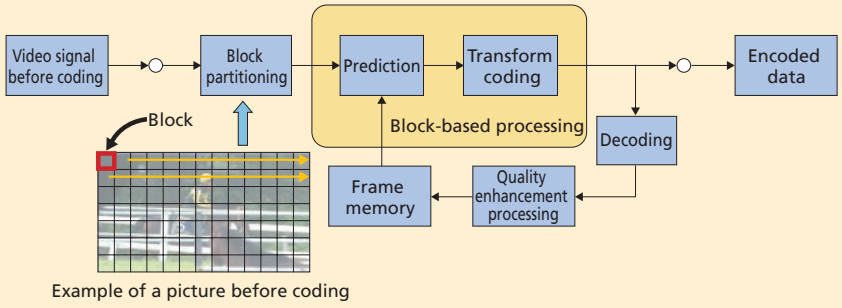

Figure 2 Basic structure of video coding

*3 **MPEG**: Technical standards for coding and transmission of digital audio and video. Standards developed by a working group under a Joint Technical Committee of ISO and IEC. MPEG-2 is used for digital TV and DVD while MPEG-4 is a coding scheme with extended application areas including mobile terminals operating at low bitrates.

*4 **ITU-T WP 3/16**: One of the Working Groups in charge of media coding schemes for video and audio in the Telecommunication Standardization Sector of the ITU which is a specialized organization of the United Nations in the field of telecommunications.

*5 **JCT-VC**: A joint team set up by ITU-T WP3/16 and ISO/IEC MPEG to study the next generation video coding scheme. Its participants are the members of the video coding expert groups of the two bodies.

*6 **NOTTV™**: A trademark or registered trademark of mmbi, Inc.

*7 **H.264/AVC**: A video encoding method standardized by the Joint Video Team (JVT) - a joint team between ITU-T WP3/SG16 and ISO/IEC MPEG. It achieves approximately twice the compression efficiency of earlier compression methods such as MPEG-2, and is used as the video compression format in services such as One Seg broadcasting.

*8 **VGA**: Picture format having a display resolution of 640×480 pix.

*9 **HD**: Picture format having a display resolution of 1,280×720 pix. A picture format having a display resolution of 1,920×1,080 pix is called Full HD.

and top to bottom.

In the "prediction" step, a signal similar to the pattern of the target block to be encoded (predicted signal) is generated from an already reconstructed signal. In the "transform coding" step the data in the residual signal, which is obtained by subtracting the predicted signal from the source signal of the target block, is compressed by exploiting the fact that the power in image information tends to concentrate in lower frequency components. In more detail, it transforms the residual signal into the frequency domain and then suppresses the amount of information in the residual signal by quantization[*11] that assigns more bits to subjectively important lower frequency components and less bits to the more numerous higher frequency components.

In the above way, "prediction" and "transform coding" steps can reduce the amount of information in the source signal by concentrating the signal power of the video source in the lower frequency components of the residual signal. On the other hand, the mosaic-like block borders may be visible on a reconstructed signal restored from encoded data in "decoding" since they are performed in units of blocks. To remove such block-shaped noise, a smoothing filter[*12] is applied to the reconstructed signal in "quality enhancement processing." The reconstructed signal after quality enhancement processing is then stored and managed in "frame memory" for prediction purposes.

The "prediction" step is classified into two types: intra-frame prediction and inter-frame prediction. These two methods are explained below.

1) Intra-frame Prediction

First, given a picture to be encoded, a predicted signal of the target block is produced by intra-frame prediction utilizing already reconstructed pixels surrounding the target block. An example of intra-frame prediction (for a block size of 8 × 8 pix) is shown in **Figure 3**. Here, a predicted signal for the pixels indicated by the white boxes and blue-framed box in the figure is produced using the already reconstructed pixels indicated by the light blue and greenish yellow boxes. For example, the predicted signal for blue-framed pixel P(x, y) is computed by extending greenish yellow pixel T. At the same time, multiple candidates for pixel T besides the greenish yellow box are prepared and one with less error is selected, since the similarity between already reconstructed pixels and pixel P(x, y) can vary in a complicated way according to the tex-

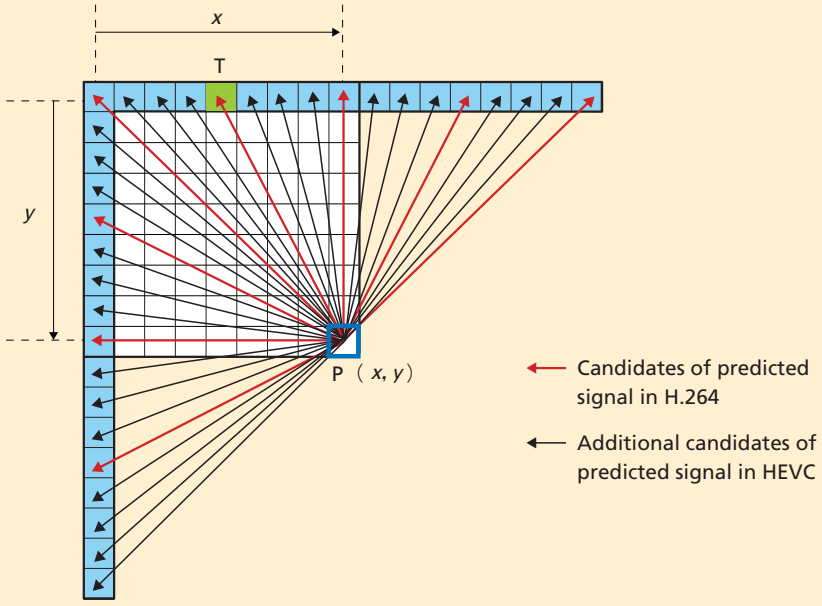

Figure 3 Example of intra-frame prediction

*10 **Transform coding**: To convert sequential data such as moving pictures into discrete data with only particular components on the frequency domain by mathematical processing. It allows compressing the amount of information needed to represent the images.

*11 **Quantization**: A process of assigning the values of discrete data generated by transform coding to values representing coarse intervals of scattered values. While resulting in some distortion, quantization can significantly reduce the amount of information.

*12 **Smoothing filter**: A filter to remove noise by cutting higher frequency components of the signal.

ture pattern within the picture. Specifically, various prediction directions are prepared as shown by the arrows in the figure and one direction will be chosen for each block.

2) Inter-frame Prediction

For the subsequent and later pictures, inter-frame prediction can be used to search for a block that has a similar pattern to the target block from previously reconstructed pictures (reference pictures) (**Figure 4**). Then, the distance and the direction from the original position of the target block to this similar block is detected as a "motion vector" and the reconstructed signal in the similar block is produced as the predicted signal. Moreover, the vector information different from the neighboring block motion vector is encoded as additional information because the motion vectors of succeeding blocks are similar.

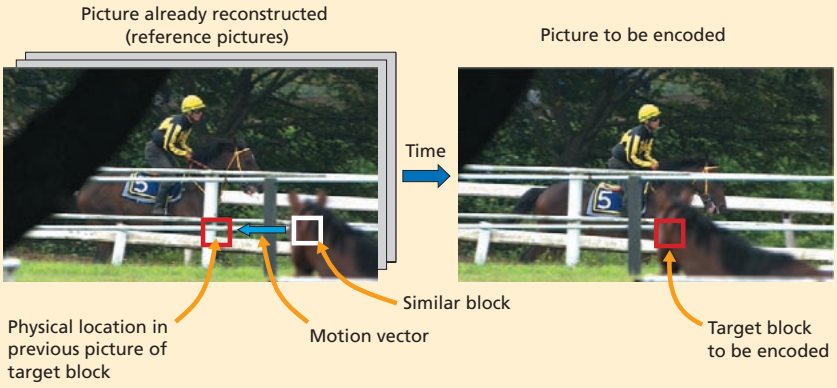

Figure 4 Inter-frame prediction

2.2 Differences between HEVC and Existing Technologies

1) Enlargement of Block Size

The H.264 and previous standards have targeted video with a picture resolution lower than that of analog television broadcasts (720×480 pix), and these standards have been applied without modification to even HD and higher resolution video. Consequently, granted that the combination of picture resolution and block size shown in **Figure 5**(a) would obtain optimal compression performance in existing schemes, there was no other option but to apply the same block size as that shown in Fig. 5(a) to even a video with four times the picture resolution as shown in Fig. 5(b). Since this was essentially the same as

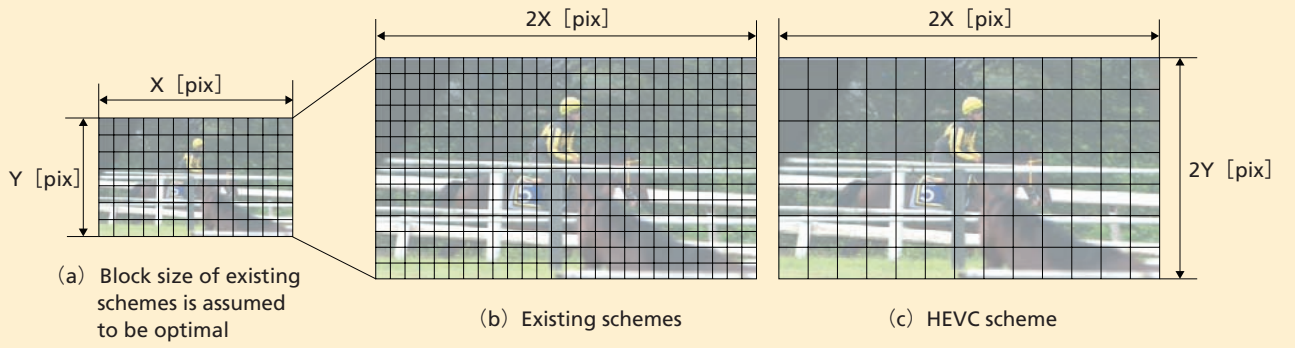

Figure 5 Block size comparison

dividing each block in Fig. 5(a) into four blocks, the optimal compression performance would not be obtained and the efficiency of prediction and transform coding would drop. To cope with this problem, the basic block size of 16 × 16 pix in existing schemes has been expanded to 64 × 64 pix in the HEVC standard. Since the interior of a basic block can be further partitioned into blocks as small as 4 × 4 pix, the HEVC standard enables a block size appropriate to picture resolution to be selected as shown in Fig. 5(c).

2) Improvement of Intra-frame Prediction

Intra-frame prediction in H.264 limits the number of prediction directions that can be selected to only the eight shown by the red arrows in Fig. 3. In contrast, the HEVC standard has added the directions shown by the black arrows in the figure for a total of 33 prediction directions to achieve efficient prediction for high-definition video with detailed textures made possible by advances in camera technology. This enhancement has improved prediction performance in high-definition video.

3) Further Improvement of Video Quality

With the aim of removing ringing[*13] or salt and pepper noise, the mechanism to bring the reconstructed signal after the removal of block-noise closer to the original signal is added to the quality enhancement processing step in HEVC (Fig. 2). Since this improved signal is used for prediction, applying this technique helps to improve prediction performance.

2.3 NTT DOCOMO's Contributions

With the aim of achieving high compression in HEVC, NTT DOCOMO devised more than 10 technologies for improving intra-frame/inter-frame prediction and transform coding and succeeded in having them adopted in the HEVC standard.

1) Technology Adoption Contributions

The following describes two key NTT DOCOMO technologies adopted by HEVC.

(1) Method for managing reference pictures

This method improves the performance of inter-frame prediction when the playback of video can be started from a midway point. It is common for a user enjoying video programs to play back a video from a midway point, such as when performing a play-forward operation or changing broadcast channels. To enable video playback from a midway point, the video coding scheme needs to compress consecutive pictures so that encoded data can be decoded from any picture. Encoding/display order of consecutive pictures and a procedure for managing reference pictures used in inter-frame prediction are shown in **Figure 6**. The squares at the top of the figure represent each picture while the middle and bottom parts of the figure show the state of frame memory (Fig. 2) storing reconstructed pictures as reference pictures.

To enable video playback from a midway point, the random-access point (I) which is encoded using only intra-frame prediction is arranged in encoded data every 1-2 seconds. Furthermore, to enable fast forward playback, pictures (P), which must be reconstructed and stored in frame memory as reference pictures, and pictures (B), whose playback can be skipped, are arranged in encoded data. Since the performance of inter-frame prediction can be improved by including future pictures in the display order of reference pictures, the coding order of B picture is usually delayed for one picture (see top of Fig. 6). This means, for example, that future picture P_3 in the display order can be used for the prediction of picture B_2 in addition to picture P_1 ((A) in

*13 **Ringing**: Artificial noise generated at the edge of an image appearing as an overlay of edges that are actually non-existent. Caused by unnecessary high-frequency components.

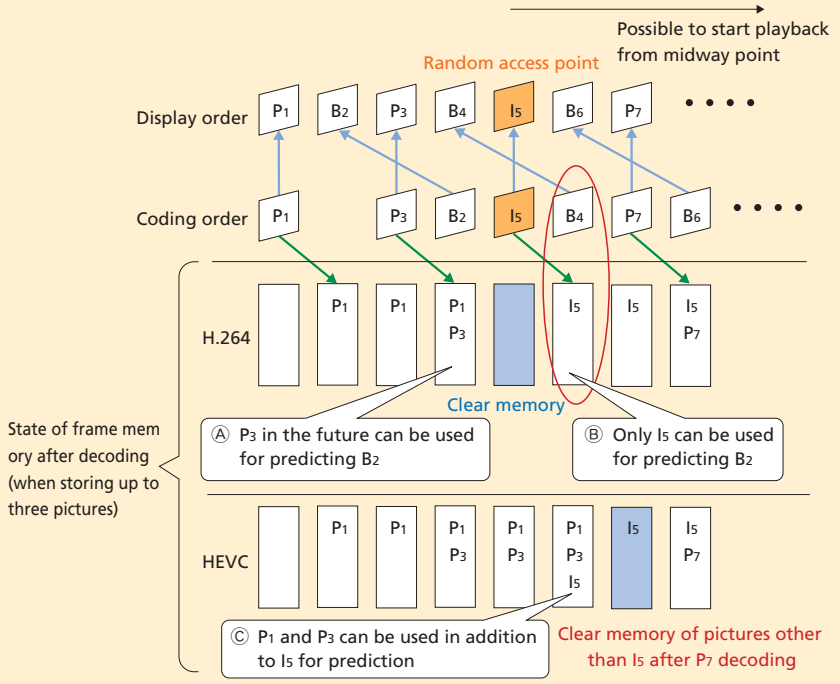

Figure 6 Example of a procedure for managing reference pictures

the figure).

Now, when the playback from the midway point starts from picture I_5, reconstructed pictures prior to I_5 cannot be used in prediction processing for picture P_7. In H.264, reference pictures that had been stored in frame memory are made unusable just after decoding of I_5 (see blue rectangle in the middle of Fig. 6). As a result, only I_5 can be used for the prediction of B_4 even when all P_1-P_7 pictures are decoded ((B) in the figure).

Focusing attention on the fact that the performance of inter-frame prediction improves when using more reference pictures, we proposed a technique that enables multiple reference pictures to be used for the prediction of B_4 ((C) in Fig. 6) by delaying the timing for making reconstructed pictures prior to I_5 unusable to after the decoding of P_7 (see blue rectangle at the bottom of Fig. 6). This technique improves the compression ratio by 4% on average and up to 8% at maximum (doubling the compression ratio corresponds to a compression-ratio improvement of 50%).

(2) Motion information sharing

Motion information sharing is a technique to achieve the efficient coding of motion between two pictures. Referring to **Figure 7**, it happens that there is a block that contains two areas with different types of motion as shown by the red box. In this example, the left half of the red box and all of the white box have the same motion as shown in the blue box. The past approach was to divide the red box into two and encode the motion of each section separately.

Thus, we proposed a technique for sharing the motion information of the white box with the left half of the red box so that the blue box can be predicted using a single motion. This technique enables the left half of the red box to be predicted by referencing the motion of the white box. Sharing the motion of two adjacent blocks in the coding process improves the compression ratio by 3% on average and up to 5% at maximum.

2) Standardization Support Contributions

NTT DOCOMO contributions to the standardization of HEVC are summarized below.

To begin with, we have been a

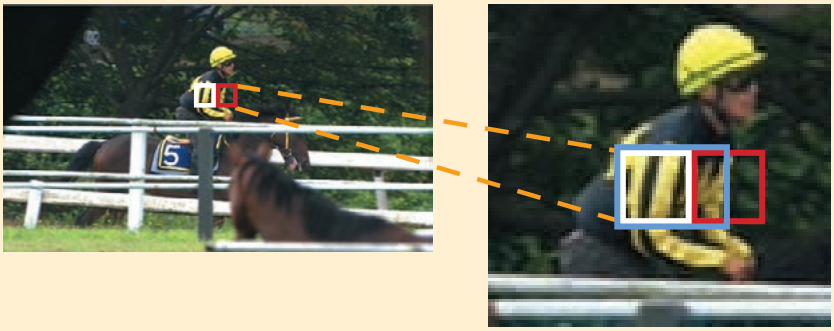

Figure 7 Motion sharing prediction

guiding force since the beginning of HEVC standardization activities by providing video materials used in standardization work and proposing requirements related to the target coding performance of HEVC performance. In the call for proposals conducted at the start of the HEVC standardization process, NTT DOCOMO's proposal was evaluated as one of the top five—these five were not ranked in any order—out of 27 that were submitted [5].

After the start of standardization activities, NTT DOCOMO supported the standardization process as the chief coordinator of reference software essential to achieving high compression in the HEVC standard. It played an important role in improving stability in the HEVC standard by reviewing proposals for which a software implementation was difficult and by encouraging the elimination of conflicts between software and draft texts for the standard.

NTT DOCOMO also investigated the effectiveness of HEVC tools deemed essential to mobile video services by conducting subjective evaluations and proposed profiles[*14] that specify the tool set when using HEVC [7]. The current HEVC profiles were defined with reference to our proposal [8].

3. HEVC Performance

3.1 HEVC Basic Performance

We here report on the results of a comparison test to assess the compression performance of widely used H.264 versus HEVC. Specifically, we compared the amount of data that would be needed by H.264 and HEVC to obtain the same subjective quality.

In this verification test, we had 24 non-experts view video of various data amounts compressed by H.264 and HEVC and had them evaluate each video sample using a 5-level absolute grade (in which "1" is poor and "5" is excellent). This evaluation method is based on a standard methodology specified by ITU [9]. Test results are summarized in **Figure 8.** The horizontal axis of the chart represents video resolution and type of video and the vertical axis represents the Mean Opinion Score (MOS)[*15] of the evaluations made by the 24 subjects with respect to subjective quality. The navy blue and purple bars represent the evaluations of video compressed by H.264 and HEVC, respectively. The numerals within each bar indicate amount of data (bitrate) and the interval symbol at the top of each bar indicates reliability of that score. If the interval symbols of two adjacent bars should overlap, the subjective qualities of those two video samples are considered to be essentially no different from a statistical point of view. These test results demonstrate that HEVC can achieve essentially the same subjective quality at half the bitrate or less of H.264 regardless of the resolution or type of video. They also show that the same results can be obtained for PC displays in addition to handset- and tablet-type of mobile terminals.

3.2 Application of HEVC to NTT DOCOMO Services

As explained in the Introduction to

*14 **Profile**: A standard subset of all coding functions defined with the aim of ensuring interconnectivity between terminals for a given application.

*15 **MOS**: A widely used measure of subjective quality representing the average value of subjective evaluations given by multiple subjects.

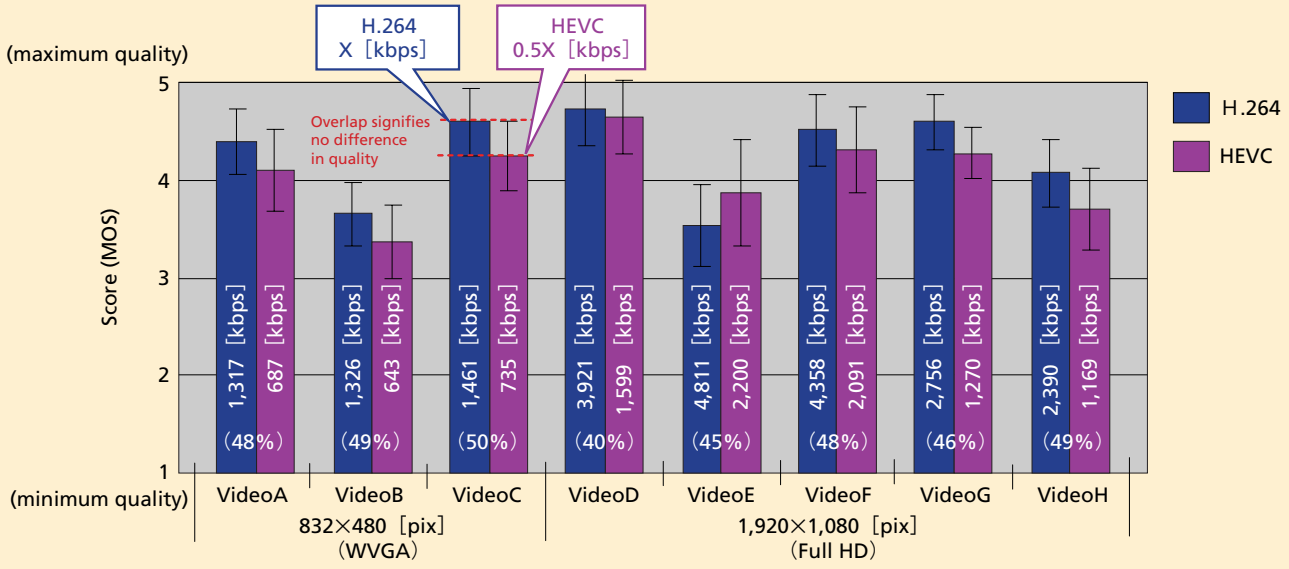

Figure 8 HEVC basic performance

this article, the H.264 video coding scheme has been used in NTT DOCOMO video services. Accordingly, the high compression performance of HEVC means that we can expect similar subjective quality as that presently provided at half the bitrate or less of current services. In other words, the amount of transmitted data can be reduced while maintaining existing video quality (**Table 1**), or from a different viewpoint, video of even higher quality can be delivered without changing the current amount of transmitted data. For example, HEVC would make it possible to provide high-quality HD video (1,280×720 pix, 30 fps) at the same bitrate (1.5 Mbps) of "very beautiful" content (Wide-VGA (WVGA)[*16], 30 fps) currently available in NTT DOCOMO's VIDEO store. Additional, HEVC would also exhibit significant compression efficiency for animation, enabling the provision of very high quality animation (WVGA, 24 fps) even at 200 kbps.

Table 1 Transmitted video data by HEVC

Resolution [pix]	No. of pictures [fps]	Amount of data required for achieving sufficient service quality [kbps]
360p	30	300
WVGA		750
HD		1,000
Full HD		2,000
4K		10,000

From NTT DOCOMO's viewpoint, HEVC will enable more efficient usage of the transmission band and help mitigate the rapid increase of video traffic. From the user's viewpoint, HEVC can lead to more comfortable video viewing by shortening the wait time until pay-

*16 **WVGA**: Picture format having a display resolution of 832×480 pix.

back, reducing jerky video playback, etc. Content providers, meanwhile, can look forward to the mobile delivery of high-quality video. Finally, as shown in Table 1, the deployment of HEVC will enable the delivery—at actual LTE transmission rates—of smooth high-quality, high-frame-rate video such as Full HD or ultra-high-definition 4K[*17] video having four times the pixels as Full HD.

3.3 Further Application of HEVC

Apart from NTT DOCOMO video services, there has been increasing mobile use of video on-demand services like Hulu[*18] and video sharing sites like YouTube™[*19]. This means that there will still be much content compressed by existing schemes (H.264, MPEG2, etc.) even after the spread of HEVC on the Internet, which means that an immediate reduction in video traffic cannot be expected.

With this being the case, we studied the possibility of reducing video traffic by playing back Internet-based video content compressed by existing schemes and recompressing it using HEVC.

Specifically, we studied the extent to which recompression by HEVC could reduce the amount of data in YouTube video, which currently accounts for most of the video traffic in the mobile network. In this study, we targeted YouTube video belonging to several popular genres and used the evaluation method [9] of section 3.1 to see whether subjective quality could be maintained at reduced bit rates when recompressing various YouTube video samples.

Here, we had 20 non-experts view on a PC display original YouTube videos and the same videos recompressed with HEVC at various bit rates, and had them evaluate the quality of each video using a 5-level absolute grade as before.

Test results are shown in **Figure 9**. The horizontal axis of the chart represents content resolution and type of video and the vertical axis represents the MOS of the evaluations made with respect to video quality. The interval symbol has the same meaning as that explained in section 3.1. As can be seen from these results, HEVC can be used to recompress YouTube video data to approximately 45% the original amount while maintaining subjective quality. It was also found through this verification test that HEVC could recompress some YouTube content down to 30% of the original amount without incurring a subjective quality difference.

On the basis of these results, we can expect video traffic to be further reduced while maintaining subjective quality provided that the amount of video data flowing on the Internet can be reduced by subjecting it to HEVC recompression before allowing it to flow on the mobile network.

4. HEVC Software Decoder Development

As described above, HEVC in a mobile environment can achieve twice the compression efficiency of H.264 and reap a variety of advantages as a result. However, the fact that HEVC is being applied to a mobile environment means that it must be able to perform on a mobile terminal.

To address this issue, we undertook the development of an HEVC software decoder and a video player application for Android™[*20] terminals mounting the popular ARM processor[*21] to verify computational complexity and the amount of memory needed.

As a result of promoting this development, the HEVC decoder specified in the HEVC Draft International Standard (DIS) has been shown to be capable of real-time playback of HD-quality HEVC video (1,280×720 pix, 30 fps, 1 Mbps) on a Dual Core[*22] 1.5-GHz ARM processor even in software form.

It was also shown that the CPU utilization rate during HEVC video playback was approximately 60% and that

*17 **4K**: Picture format having a display resolution of 3,840×2,160 pix or 4,096×2,340 pix.
*18 **Hulu**: A trademark or registered trademark of Hulu, LLC.
*19 **YouTube™**: A trademark or registered trademark of Google, Inc.
*20 **Android™**: A software platform for smartphones and tablets consisting of an operating system, middleware and major applications. A trademark or registered trademark of Google Inc., United States.
*21 **ARM processor**: Generic name of processors adopting the ARM architecture developed by ARM Holdings in the United Kingdom. Has a high usage rate in smartphones.
*22 **Dual Core**: A processor integrating two CPU cores in one package.

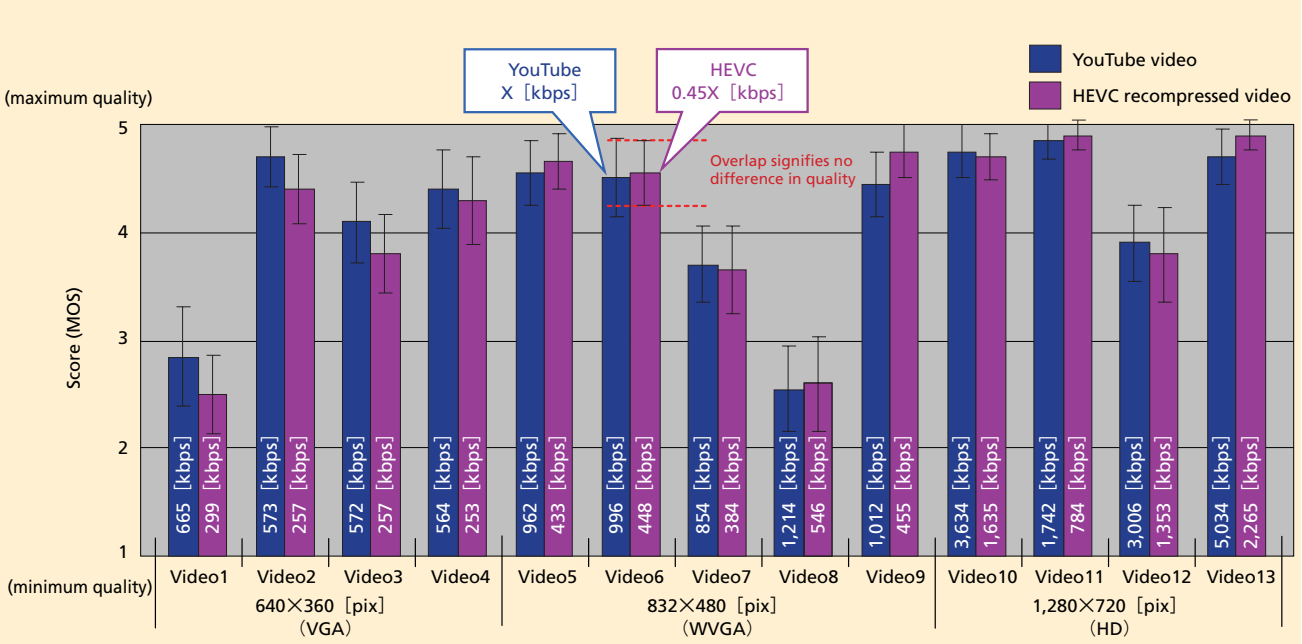

Figure 9 HEVC recompression performance of Internet video

continuous playback of HD video could be performed for more than eight hours. These results show that HEVC has reached a practical level also in terms of power consumption. In the future, we plan to further develop and study the HEVC video coding scheme for application to even more practical environments.

5. Conclusion

The features of the HEVC standard finalized in January 2013 and NTT DOCOMO's contribution to its development were described. It was verified by subjective evaluation that the HEVC standard can reduce the amount of transmitted data in mobile video by half compared to existing schemes without any degradation of visual quality through the subjective evaluation. In addition, the advantage of HEVC for NTT DOCOMO video services and the perspective for the HEVC standard in mobile video were discussed.

We plan to encourage the adoption of HEVC at 3GPP and promote its deployment in mobile services with the aim of achieving worldwide penetration.

References

[1] ISO/IEC 14496-10: 2009: "Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding," May 2009.
[2] ITU-T Recommendation H.264: "Advanced video coding for generic audiovisual services," Mar. 2005.
[3] V. Baroncini, J.-R. Ohm and G. Sullivan: "Report on preliminary subjective testing of HEVC compression capability," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVCH1004, 8th Meeting: San José, CA, USA, 1-10 Feb. 2012.
[4] Cisco Systems, Inc.: "Cisco Visual Networking Index (VNI): Global Mobile Data Traffic Forecast, 2011–2016," Feb. 2012.
[5] Y. Suzuki et al: "International Standardization of Next Generation Video Coding Scheme Realizing High-quality, High-efficiency Video Transmission and Outline of Technologies Proposed by NTT DOCO-

MO," NTT DOCOMO Technical Journal, Vol. 12, No. 4, pp. 38-45, Mar. 2011.

[6] Qualcomm, NTT DOCOMO, et al.: "New Work Item Description - High Efficiency Video Coding," 3GPP, S4-121208, TSG SA WG4 70th Meeting, Aug. 2012.

[7] T. K. Tan, A. Fujibayashi, Y. Suzuki and J. Takiue: "Objective and subjective evaluation of HM5.0.," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-H0116, 8th Meeting: San José, CA, USA, 1-10 Feb. 2012.

[8] M. Horowitz, A. Ichigaya, K. McCann, T. Nishi, S. Sekiguchi, T. Suzuki, T. K. Tan W. Wan and M. Zhou: "Straw man for Profiles and levels," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-H0734, 8th Meeting: San José, CA, USA, 1-10 Feb. 2012.

[9] Recommendation ITU-R BT.500-11: "Methodology for the subjective assessment of the quality of television pictures."