

Technology Reports

International Standardization of Next Generation Video Coding Scheme Realizing High-quality, High-efficiency Video Transmission and Outline of Technologies Proposed by NTT DOCOMO

Along with the rise in the use of smartphones and the launch of LTE services, the conditions for viewing high-resolution and high frame-rate videos are improving and NTT DOCOMO is actively participating in the standardization activities for the next generation video coding schemes being managed by the ITU-T and the ISO/IEC. This aims at decreasing bitrates necessary for transmitting high-quality video by half. In the assessment held by both organizations in April 2010, the video coding scheme proposed by NTT DOCOMO, which features a simple configuration and is implementable with less hardware compared to other proposals, was evaluated to be one of the top five proposals, that all had almost the same performance, out of 27 proposals.

Research Laboratories

*Yoshinori Suzuki
Boon Choong Seng*

1. Introduction

NTT DOCOMO is currently participating in the standardization activities for the next generation video coding schemes promoted by ISO^{*1}/IEC^{*2} Moving Picture Experts Group (MPEG) and ITU-T WP 3/16^{*3}. Both standardization organizations aim to standardize a new video compression scheme by 2013 which will realize the transmission of high-resolution and high frame-rate videos more efficiently. The requirement set for this scheme is to be able to reduce bitrates to half of

that necessary for H.264/Advanced Video Coding (AVC)^{*4} (hereinafter referred to as “H.264”) [1][2], the latest international standard, which is used by One Seg and i-motion. This new coding scheme will be given the nickname of High Efficiency Video Coding (HEVC).

In recent years, the conditions for viewing High-Definition (HD) TV^{*5} videos over various services including mobile are getting better and better. In the case of mobile terminals, full HD (1,920 × 1,080 pix) video filming capability has begun to be supported starting with the summer 2010 models in the

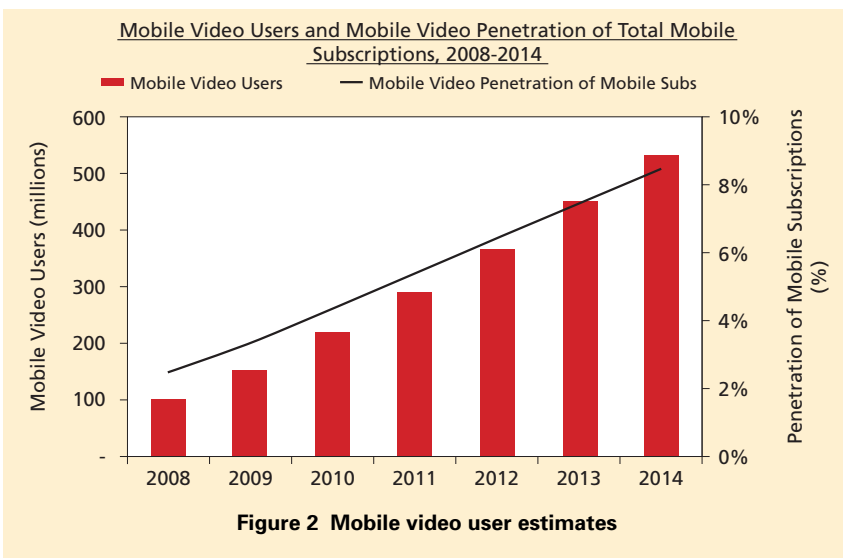
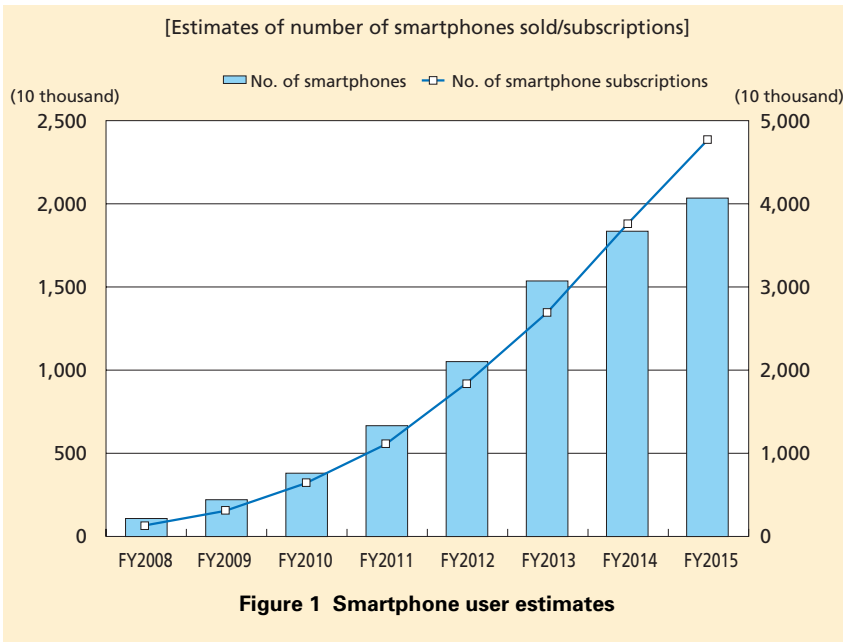
FOMA PRIME series. Also, as regards smartphones (**Figure 1**) [3], whose market is expected to grow from now on, there are already several models that can display HD (1,280 × 720 pix) video.

It is also foreseen that the demand for video content on mobile networks will grow further (**Figure 2**) [4] triggered by the launch of LTE^{*6} services, and thus HD video transmission to mobile terminals and delivery services of video with more than 60 fps (frames/second) are under study in the 3GPP [5].

*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 Commis-

sion; an organization for standardization in the information technology. Sets standards in the electrical and telecommunication field.



Under such circumstances, in order to efficiently utilize mobile frequency bandwidth, it is necessary to have a video coding scheme that is capable of compressing video signals with higher picture resolution and frame-rates. NTT DOCOMO has been working on its own to develop a video

coding scheme in collaboration with DOCOMO USA Labs so as to establish an international coding scheme for such rich video formats with large data volumes. In addition, in view of the fact that coding schemes have been getting more and more complex in recent years, we aimed at achieving a higher

compression ratio using a simplified coding scheme. As a result, we succeeded in January 2010 in reducing bitrates to approximately one third of that for H.264.

NTT DOCOMO responded to the joint call for proposals [6] issued by ISO/IEC MPEG and ITU-T WP3/16 at the start of the HEVC standardization process and our proposal [7] was evaluated as one of the top five (these five were not ranked in any order) out of the 27 that were submitted [8].

This article describes the current status of HEVC standardization and presents the advantages of the simplified video coding scheme developed by NTT DOCOMO (hereinafter referred to as “DOCOMO Scheme”).

2. HEVC Standardization Overview

2.1 Standardization Objectives

HEVC aims at halving the transmission bitrates for video compared to that required for H.264 in the following conditions:

- Image resolution: QVGA^{*7} - 8 k × 4 k^{*8}
- Frame-rate: additional 60 - 120 fps
- Scanning: progressive scanning^{*9} only

Because the transmission video bitrates are halved, the time required for sending video with the same quality can be halved. It is also possible to improve

*3 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.

*4 H.264/AVC: One of video coding standards specified by the Joint Video Team (JVT) which is a joint team between ITU-T WP3/16 and MPEG of ISO/IEC.

*5 HD TV: The name of a high definition television format whose number of scanning lines is

twice that of standard television.

*6 LTE: Extended standard for the 3G mobile communication system studied by 3GPP. Achieves faster speeds and lower delay.

*7 QVGA: Picture format whose size is 320 × 240 pix.

the video quality from 15 fps to 30 fps and send it with the same bitrates. This means that video that looks “not-so-smooth” at a frame-rate such as that used in the case of One Seg will be able to be shown smoothly by utilizing HEVC.

Standards in the past assumed that video signals consisted of smaller pictures than analog TV broadcasting signals (720×480 pix), and these standards were applied to HD video without changes. Therefore, it may be safe to say that there have been no standards optimized for HD video. In HEVC a basic policy has been established, in the light of the fact that video content quality will be further enhanced taking advantage of improved shooting/display technologies, so that the new standard is developed utilizing a totally different scheme from the existing standards, targeting progressive scan HD videos. As regards the test sequences, sequences with 60 fps and the sizes of QVGA to full HD have been prepared and used for technical evaluation. NTT DOCOMO has been actively participating in the standardization activities from the initial stage of setting the goals of the new standard, and it has provided test sequences [9][10] and has made proposals on the requirements [11].

2.2 Standardization Bodies

ISO/IEC MPEG and ITU-T WP 3/16 had been independently studying the new video coding scheme up to

2009. However, following suggestions from some companies including NTT DOCOMO, the Joint Collaborative Team on Video Coding (JCT-VC)^{*10} was founded in January 2010 as a joint project to study HEVC. After the JCT-VC finalizes the specifications, HEVC will become an international standard following approval by the two standardization bodies.

2.3 Call for Proposals

In order to collect technologies that will be incorporated into the standard, the JCT-VC made a Call for Proposals (CfP) [5] in January 2010, and 27 organizations responded. A test to evaluate the videos encoded by each of the 27 proposed schemes was carried out based on the subjective video quality assessment methods [12] specified by the ITU-R^{*11}, which concluded that the top five proposals have the same level of quality [8][13]. Four schemes out of the top five, including that from DOCOMO Scheme, further could achieve a compression ratio of approximately one third against H.264.

It was observed that many of the proposals adopted technologies that can efficiently compress high-resolution video and/or high-quality video with detailed textures and less-noise taking advantage of the basic coding structure.

2.4 Timeline

Figure 3 shows the timeline HEVC standardization. Based on the

technologies contained in the top five proposals in the CfP assessment, the JCT-VC decided a test model (HEVC Test Model: HM) in October 2010 which will serve as the preliminary version of HEVC. Specifications for HEVC will be finalized by adding technologies to the HM that prove to be effective.

The drafting of specifications stipulating the compressed data format and its decoding procedures will commence with the Committee Draft (CD) in February 2012 and the Final Draft International Standard (FDIS), which will be the official standard, will be drafted in January 2013 following the Draft International Standard (DIS) phase in July 2012.

3. Video Coding Technologies and Features of DOCOMO Scheme

3.1 Basic Structure of Video Coding Scheme

Video sources have a large amount of data but adjacent pixels and successive frames are quite similar. In the video coding schemes, data volumes are compressed by getting rid of such redundancy in the video sources according to certain rules.

Figure 4 shows the basic structure of a video coding scheme. In a video coding scheme, a video source is separated into pictures and coding is performed on the basis of “blocks” which

*8 **8k × 4k**: Picture format whose size is $7,689 \times 4,320$ pix. Also called “Ultra HD” because it was developed as a next generation HD TV broadcast format.

*9 **Progressive scanning**: A way to display the screen in which all the lines of each frame are

drawn in sequence. In contrast, with interlace scanning, only the odd lines are drawn in the first scanning and the even lines are drawn in the second scanning. If the number of frames is the same, progressive scanning will have less flickering and smearing.

*10 **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.

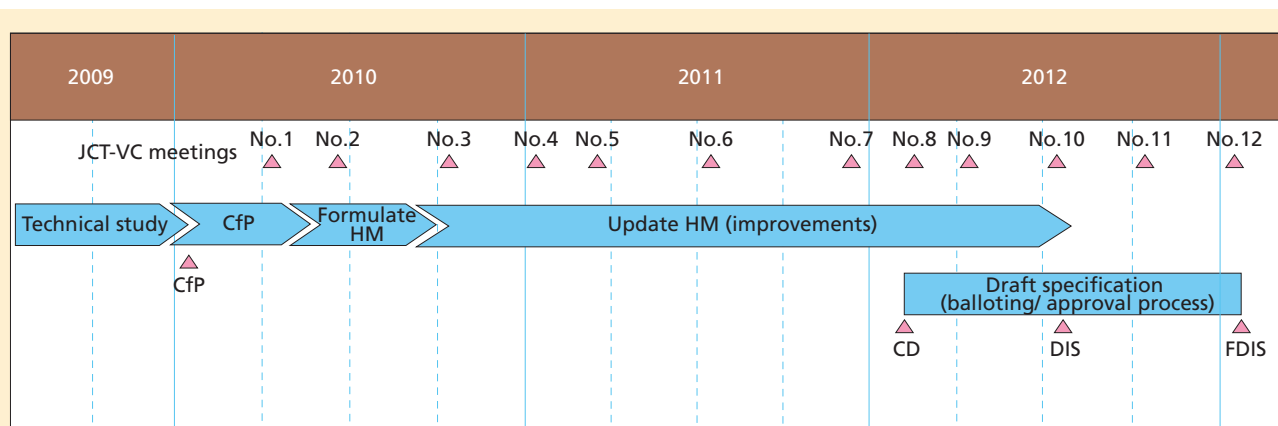


Figure 3 HEVC standardization timeline

are obtained by partitioning a picture into smaller sections (Figure 5). Blocks are the units for coding and are input into the “prediction” and “transform coding” processes one by one from left to right and from top to bottom.

In the case of prediction, at first the block that has a similar pattern to the block to be encoded (target block) will be searched for from the already reconstructed pictures (Figure 6). Then, the distance and the direction from the original position of the target block to this similar block is encoded as “motion vector,” and the difference between the source signal in the target block and the reconstructed signal in the similar block (“predicted signal”) is calculated as the “residual signal.” In such a way, by expressing the target block with “motion vector” and “residual signal” by utilizing previously reconstructed signals, it becomes possible to suppress redundancy in the video source. As

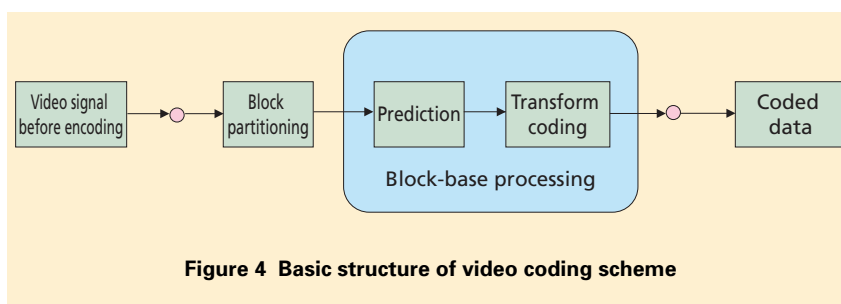


Figure 4 Basic structure of video coding scheme

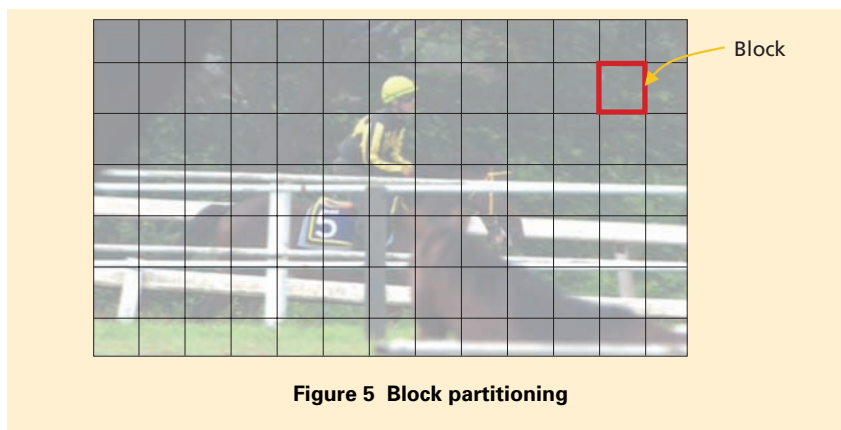


Figure 5 Block partitioning

regards “motion vectors,” usually the “differential vector” is encoded because the “motion vectors” of succeeding blocks are quite similar.

In the case of transform coding, the residual signal obtained in prediction is

transformed into the frequency domain in order to compress the data by taking advantage of the fact that the video signal power is concentrated in the lower frequency components. After the transformation, the amount of encoded data

*11 ITU-R: Radiocommunication Sector of the ITU which recommends methodologies for subjective video quality assessment in addition to administration and coordination activities related to radiocommunications.

is suppressed by quantizing the fewer lower frequency components with more bits and the more numerous higher frequency components with less bits.

It is also effective to partition a picture into blocks that are suitable for texture pattern changes and localized movements within the picture in order to more effectively concentrate the residual signal power into lower frequency components. In the case of H.264, the basic block size is 16×16

pix, but it can be further partitioned into 4×4 pix blocks which are the smallest possible blocks (**Figure 7**).

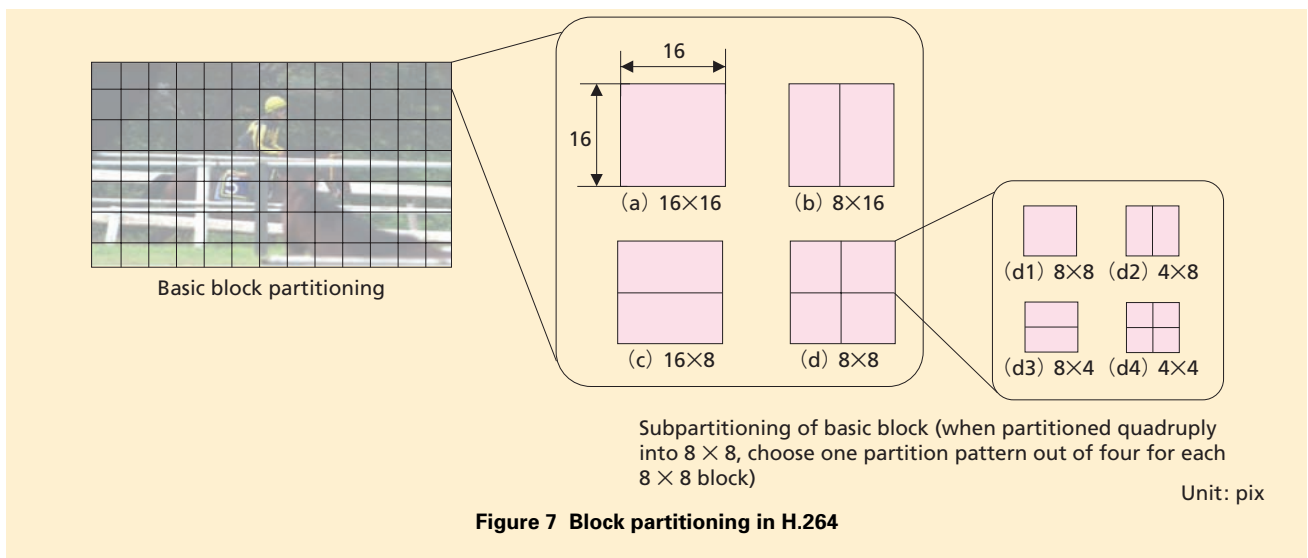
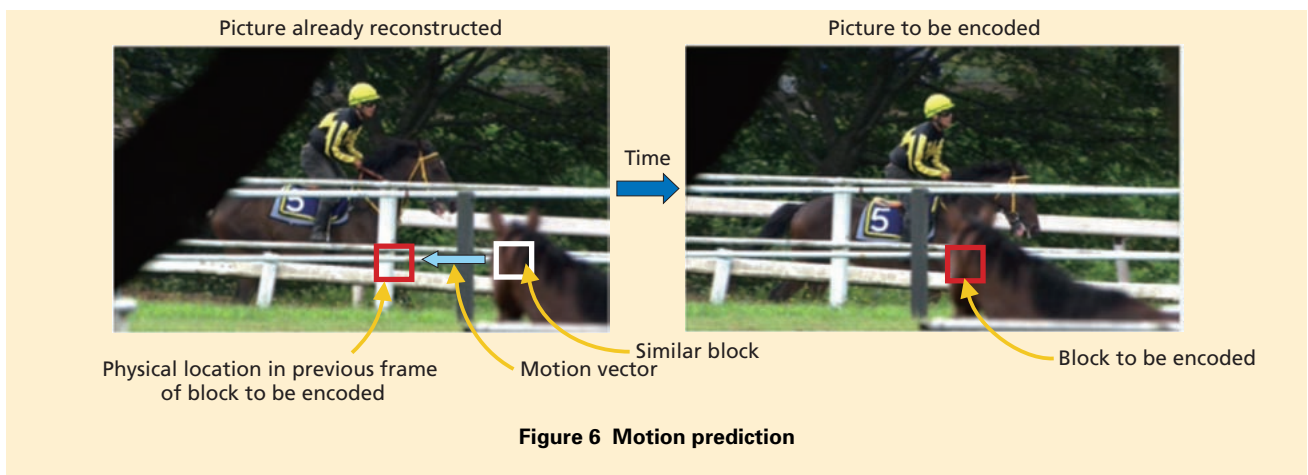
3.2 Technologies Proposed to HEVC

The following are some of the major technologies that were proposed for HEVC in the CfP:

- 1) Introduction of Basic Blocks Larger than 16×16 pix

The basic block size in the current

standard (16×16 pix) is too small for HD video. If an HD picture is partitioned into blocks of this size, many blocks with a uniform pattern are created. Therefore, if we conduct a motion search for blocks with such a uniform pattern, many candidates of similar block will be found. Since it is generally difficult to choose one optimum block out of many similar blocks, the motion vector tends to fluctuate from one block to the next and therefore the



coding bit of differential vectors will become large.

To cope with this issue, many of the proposed schemes use a basic block larger than 16×16 pix. **Figure 8** shows an example where the size of the basic block is set at 64×64 pix, and its interior is expressed by a combination of blocks with four sizes (64×64 , 32×32 , 16×16 , 8×8). By choosing the suitable block size fit for motion search based on local characteristics inside the picture, it is possible to suppress the fluctuation of motion vectors.

There are also proposals to utilize blocks larger than 16×16 pix for transform coding. Generally, the effect of large block sizes is bigger in the case where the texture pattern is complex and prediction is difficult, whereas the effect is smaller in areas where the residual signal can be sufficiently compressed by prediction. Supposing that there is a region that has significant information in a residual signal and that it is further partitioned it into smaller blocks which are transformed to the fre-

quency domain, the lower frequency components will be scattered among every block in this region and the amount of encoded data is increased. In such a case, significant information is concentrated in one place and it is coded efficiently by utilizing larger blocks.

Since the enlargement of the block size may result in an increase in the gate counts of hardware circuits, more studies are needed on its restrictions and adaptations before adopting it into the standard.

2) Picture Quality Improvement Filter

The mosaic like block borders are visible on a reconstructed signal when a picture is encoded using a block-base coding. Usually this noise is removed by using a smoothing filter^{*12}, but there can be a problem as the image may get blurred. Thus, there is a proposal to decrease errors between the original source signal and the reconstructed signal using an adaptive filter^{*13}. It has the effect of reducing the residual signal power and contributes to improving

coding efficiency since the reconstructed signal with less noise is used for prediction.

3) Improvement in Signal Processing Precision

It has been proposed that the rounding errors at the time of mathematical calculation are reduced by improving signal processing precision during prediction and transformation processes. This improves coding efficiency because the information which was lost during the coding process can be preserved.

4) Multi-directional Intra-frame Prediction

There are some proposals to enhance the directional characteristics for the intra prediction that predicts pixels in a target block using the reconstructed pixels surrounding it.

In the intra-frame coding, the 64 pixels shown in the white and yellow boxes in **Figure 9** are predicted using reconstructed pixels shown in the blue

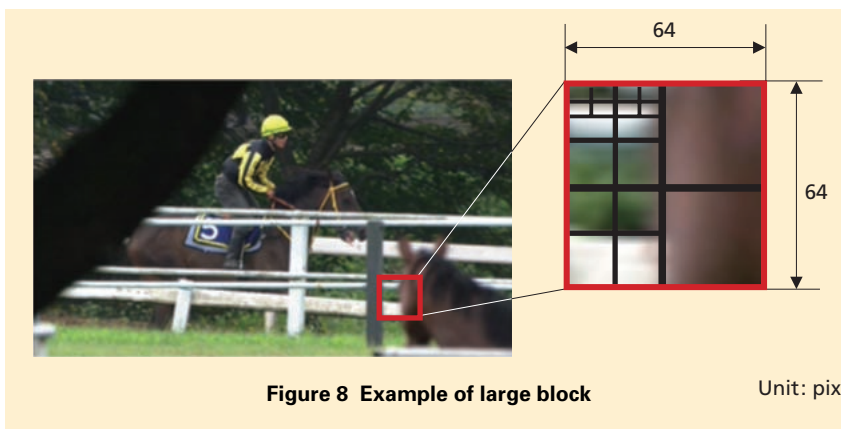


Figure 8 Example of large block

Unit: pix

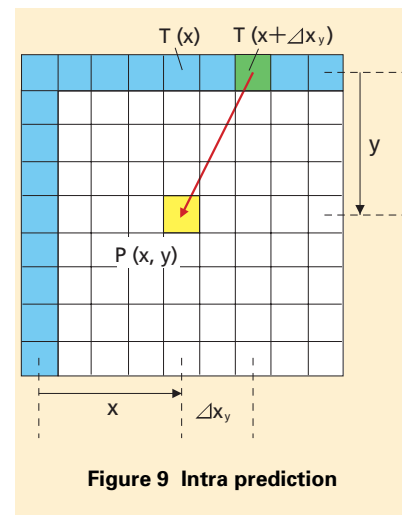


Figure 9 Intra prediction

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

*13 **Adaptive filter:** A filter that can adapt its filter coefficients so as to obtain an ideal output value for the input value.

and greenish yellow boxes. When the texture pattern is found in the direction indicated by the red arrow, then the yellow pixel $P(x, y)$ is predicted using the greenish yellow pixel $T(x + \angle x_y)$. By changing minutely the value of $\angle x_y$, precisely, it is possible to predict texture patterns with various directionalities with less error.

3.3 Technologies Proposed in the DOCOMO Scheme

We have developed the simplified video coding scheme without the use of the large basic block size or the adaptive block size so that it is easy to be handled and implemented with less gate counts of hardware circuit.

1) Fixed Block Size and Motion Vector Optimization

We have fixed the block size for inter-frame motion prediction to 8×8 pix and introduced ways to reduce the fluctuations of motion vectors even if the basic block size is small.

- Motion vector search

We have devised a way to reduce the fluctuations of motion vectors among adjacent blocks without using blocks larger than 16×16 pix. As described earlier, motion vectors usually fluctuate when the block size is small. However, because the actual movements of adjacent blocks are quite similar, the sum of the encoding cost for residual signal and motion vector is not so very different even if the

motion vector of the target block is replaced with that of the adjacent block. Thus we have introduced a method that optimizes the balance between motion vector and the residual signal by replacing motion vectors of the group of blocks with one of their motion vectors after a block-based motion search is performed. With this method, the number of coding bits for motion vectors can be reduced regardless of picture resolutions by suppressing the motion vector fluctuation among adjacent blocks.

- Shared motion prediction

On the boundaries of a moving object it happens that there is a block that contains two areas with different motions from each other as is shown in the box with red border in **Figure 10**. In this example, the left half of the box with red border and whole of the box with white border have the same motion as shown in the box with blue border. Thus we have introduced a method

to predict part of a block using the motion vector of the neighboring block so that the box with blue border can be predicted using a single motion vector. In this method, the left half of the box with red border is predicted using the motion vector of the box with white border.

2) Picture Quality Improvement Filter

In addition to the adaptive filter that reduces blurring in the reconstructed picture, we have introduced a technology that makes contours that existed in the original pictures clear, resulting in reduced distortion around the contours of human figures and alphanumeric characters [14]. By using a signal with reduced distortion, the number of coding bits of residual signal can be reduced.

3) Improvement in Signal Processing Precision

We have introduced a method to increase the processing precision in transform coding by extending the bit width^{*14} before the prediction. This enables a reduction in the number of the

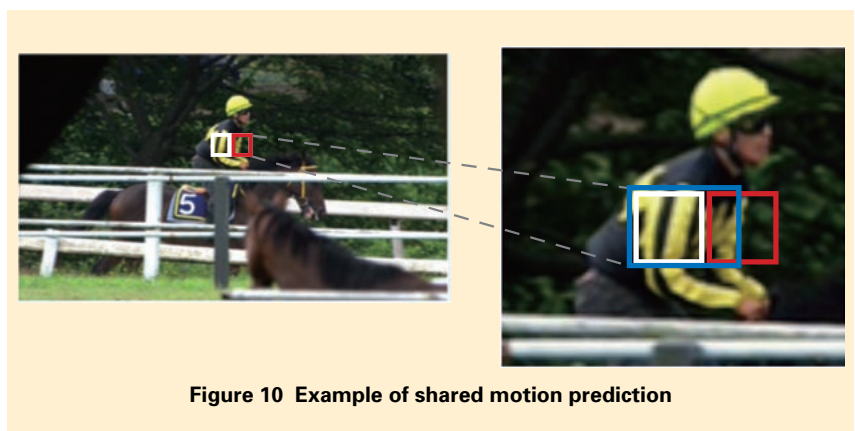


Figure 10 Example of shared motion prediction

*14 **Bit width:** Number of bits used by the hardware in its internal computation circuits and data bus.

rounding processes that take place several times during the prediction and transformation processes. This can improve coding efficiency of images with detailed texture patterns by avoiding the loss of subtle changes in the signal which had been lost when existing coding schemes were used.

4) Performance and Evaluation

Following tests using 18 different pictures and five coding bitrates, it was confirmed that the number of coding bits can be reduced by 31% on average and 46% at the maximum compared to H.264. Furthermore, in the subjective assessment conducted by the JCT-VC, it was confirmed that comparable subjective quality could be obtained without using large block size.

4. Conclusion

The standardization activities of HEVC that can compress data for high resolution and high frame-rate pictures and DOKOMO Scheme have been described.

As a result of the assessment conducted by the JCT-VC, a joint project between the ITU-T/SG16 and ISO/IEC MPEG, on the 27 encoding schemes submitted as proposals for HEVC, the DOKOMO Scheme which can improve compression efficiency by approximately 30% compared to conventional schemes was listed among the top five proposals which all had almost the

same performance. Many of the proposals for HEVC adopt large block sizes in an adaptive manner though they increase the complexity. However, with the DOKOMO Scheme we can obtain comparable data compression efficiency with a fixed block size.

We will strive for the completion of HEVC so as to reduce the number of coding bits of pictures to half of that generated by conventional schemes and make the processing structure of HEVC as simple as possible while utilizing developed technologies contained in the DOKOMO Scheme.

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] MM Research Institute: "Market forecasts for domestic mobile phones and smartphones," Aug. 2010.
- [4] Pyramid Research: "Mobile video market to grow five-fold by 2014," Jun. 2009.
- [5] 3GPP TR 26.903 V9.0.0: "Improved video support for Packet Switched Streaming (PSS) and Multimedia Broadcast/Multicast Service (MBMS) Services," Mar. 2010.
- [6] ITU-T Q6/16 Visual Coding and ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio: "Joint Call for Proposals on Video Compression Technology," ISO/IEC JTC1/SC29/WG11/N11113, Jan. 2010.
- [7] F. Bossen et al.: "Description of video coding technology proposal by France Telecom, NTT, NTT DoCoMo, Panasonic and Technicolor," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document JCTVC-A114, Dresden, DE, 15-23, Apr. 2010.
- [8] G. J. Sullivan and J.-R. Ohm.: "Meeting report of the first meeting of the Joint Collaborative Team on Video Coding (JCT-VC)," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document JCTVC-A200, Dresden, DE, 15-23, Apr. 2010.
- [9] Y. Suzuki and TK Tan: "Response to Call for Test Materials for High-Performance Video Coding Standards Development," ISO/IEC JTC1/SC29/WG11/ M16035, Lausanne, Feb. 2009.
- [10] A. Fujibayashi, Boon Choong Seng and TK Tan: "1080p, WVGA, WQVGA video coding test sequences," ISO/IEC JTC1/SC29/WG11/M16035, Lausanne, Feb. 2009.
- [11] Y. Suzuki, F. Bossen and TK Tan: "Comments on Draft Call for Proposals on High-Performance Video Coding (HVC)," ISO/IEC JTC1/SC29/WG11/ M17028, Xian, Oct. 2009.
- [12] Rec. ITU-R BT.500-11: "Methodology for the subjective assessment of the quality of television pictures," Jun. 2002.
- [13] V. Baroncini, J.-R. Ohm and Gary 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)," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, document JCTVC-A204, Apr. 2010.
- [14] O. G. Guleryuz: "A Nonlinear Loop Filter for Quantization Noise Removal in Hybrid Video Compression," Proc. of ICIP05, Vol.2, pp.336-340, Nov. 2005.