Fast Prediction Mode Decision in HEVC Using a Pseudo Rate-Distortion Based on Separated Encoding Structure

Jinwuk Seok, Younhee Kim, Myungseok Ki, Hui Yong Kim, and Jin Soo Choi

A novel fast algorithm is suggested for a coding unit (CU) mode decision using pseudo rate-distortion based on a separated encoding structure in High Efficiency Video Coding (HEVC). A conventional HEVC encoder requires a large computational time for a CU mode prediction because prediction and transformation procedures are applied to obtain a rate-distortion cost. Hence, for the practical application of HEVC encoding, it is necessary to significantly reduce the computational time of CU mode prediction. As described in this paper, under the proposed separated encoder structure, it is possible to decide the CU prediction mode without a full processing of the prediction and transformation to obtain a rate-distortion cost based on a suitable condition. Furthermore, to construct a suitable condition to improve the encoding speed, we employ a pseudo rate-distortion estimation based on a Hadamard transformation and a simple quantization. The experimental results show that the proposed method achieves a 38.68% reduction in the total encoding time with a similar coding performance to that of the HEVC reference model.

Keywords: HEVC, Video encoding, pseudo, rate-distortion.

I. Introduction

As video resolutions increase and the available bandwidth to transfer high-resolution video remains limited, a new video compression standard with a high coding performance is required. ISO-IEC/MPEG and ITU-T/VCEG have recently formed the Joint Collaborative Team on Video Coding (JCT-VC) [1], aiming to develop the next-generation video coding standard called High Efficiency Video Coding (HEVC). Because HEVC is focusing on achieving a high coding efficiency [2], [3], its computational complexity is dramatically increased. For practical applications such as high-resolution video services, HEVC requires a significant reduction in computational complexity while maintaining a high coding performance.

Including the previous coding standard, AVC/H.264, several methods [4]–[6] have been proposed to reduce the encoding complexity, which are classified into two approaches. The first type of approach is fast intra-prediction methods [7]–[9]. For intra prediction, fast intra-mode decision methods and fast block-size decision methods [10]–[15] have mainly been proposed. Conventional studies into fast intra prediction for HEVC have shown a dramatic reduction in encoding time with a feasible amount of video degradation [16]–[20]. However, although conventional research shows that a reduction in the encoding time is about 30% to 40%, such reduction is limited to only intra frame or intra mode. Because the encoding time of intra frame or intra mode is a small part of an IDR period, the overall encoding time is not effectively reduced for practical applications.

Manuscript received Dec. 26, 2015; revised May 24, 2016; accepted June 13, 2016.

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-16-0295, Development of UHD Realistic Broadcasting, Digital Cinema, and Digital Signage Convergence Service Technology).

Jinwuk Seok (corresponding author, jnwseok@etri.re.kr), Younhee Kim (kimyounhee@etri.re.kr), Myungseok Ki (serdong@etri.re.kr), Hui Yong Kim (hykim5@etri.re.kr), and Jin Soo Choi (jschoi@etri.re.kr) are with the Broadcasting Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.16.2615.0031

The other approach is the use of fast inter-prediction methods [21]–[27]. The referenced studies also show that the reduction in the encoding time is about 20% with a small degradation of the video quality. However, because conventional studies have focused on the case of a zero bit occurrence or a specific case of the decision process, the practical encoding time is not effectively reduced when all default fast algorithms such as ESD, ECU, CFM, and FEN are active. In addition, some conventional studies have shown a lesser encoding performance in comparison to a combination of default fast algorithms provided by reference software, even though the reduction of the encoding time is beyond 20% to 30%. Therefore, it is crucial to develop a fast algorithm under the condition in which all default fast schemes are active.

In the HEVC reference software, called HM, rate-distortion optimization (RDO) is used for almost all procedures, including the Coding Unit (CU) mode decision, in the video encoding to improve the compression performance. However, because the RDO technique contains core procedures such as a transformation, quantization, inverse transformation, and de-quantization, the computation time for video encoding in HEVC is increased dramatically.

Accordingly, instead of a conventional rate-distortion method, which consumes a lot of computing time, it is necessary to develop a faster computational method to define a CU mode. In particular, because the Rate Distortion Optimized Quantization (RDOQ) in HEVC requires a remarkable computation burden for the quantization of Discrete Cosine Transformation (DCT) coefficients, an effective scheme to decide the CU mode without an RDOQ procedure is required.

In this paper, to avoid the use of complex compression tools, such as the RDOQ, which require a heavy computational burden, we propose a novel fast algorithm for a CU mode decision with a pseudo rate-distortion based on the Hadamard transformation. In addition, we propose a separated encoder structure in which the suggested compression method is effectively conducted during the HEVC encoding procedure. For the purpose of an effective implantation of the suggested algorithm, the encoder structure is separated into two parts: a prediction part using pseudo rate-distortion based on a Hadamard transformation, and a conventional transformation including DCT/quantization based on the RDOQ. Because it is possible for the pseudo rate-distortion to decide the CU mode, with a relatively small number of computations in comparison to a conventional rate-distortion, the transformation procedure, which consumes a significant amount of computing time, is conducted within minimal timeframes. It is thereby possible to improve the computational speed with a slight degradation in compressed video quality.

The remainder of this paper is organized as follows: Section II illustrates a conventional procedure for a CU mode decision implemented in the HEVC reference software. Section III describes a separated encoder structure, which is divided into a prediction part and a transformation part. In Section IV, we elaborate on the proposed algorithm for a CU mode decision based on the pseudo-rate distortion. Section V describes the main experimental results for a validity of the proposed scheme, and Section VI provides some concluding remarks regarding this research.

II. Overview of CU Mode Decision in HEVC

The fundamental encoding unit in HEVC is not a 16 × 16 macroblock, which is employed as a fundamental unit in previous video codecs such as MPEG-2 or H.264/AVC, but is a coding tree unit (CTU). The maximum size of a CTU is 64 × 64, 32 × 32, or 16 × 16, with 64 × 64 being generally used. Based on such a CTU, the HEVC encodes a picture of CTU size as a quad-tree structure. For the purpose of improving the encoding performance, the conventional HEVC encoder evaluates the rate-distortion cost and decides the prediction mode with respect to the quad tree coding of the CTU, the size of the CU, and the CU modes.

Furthermore, the prediction of each CU mode is conducted naturally prior to evaluating the rate-distortion value. However, because an HEVC encoder evaluates the rate-distortion value through encoding and decoding procedures, the encoding procedure consumes a heavy amount of computation time. Accordingly, when the encoding procedure in the HEVC reference software is maintained, it is almost impossible to implement a real-time encoder for 4 K UHD video based on software. Figure 1 shows a flow chart used to decide the CU mode in the HEVC reference software.

To solve the problem of a heavy computational burden in the mode-decision, the HEVC reference software provides some fast algorithms, such as a fast mode decision for skipping and merging, as well as the omission of intra prediction in the case of a small residual [20], [25]. Although such algorithms decrease the computation time of an additional rate-distortion optimization, they have a limitation in reducing the fundamental encoding time, owing to an evaluation of the rate-distortion cost.

From the viewpoint of computation time, in particular, the detailed procedure of CU mode prediction is illustrated as follows. In the first stage, for CU mode prediction, the encoder initially generates the prediction picture data for a CU, and a DCT transformation is conducted for the residual data between the prediction and original data. In the next stage, through the quantization of rate-distortion optimization, the encoder codes the quantized data using context-adaptive binary arithmetic

http://dx.doi.org/10.4218/etrij.16.2615.0031

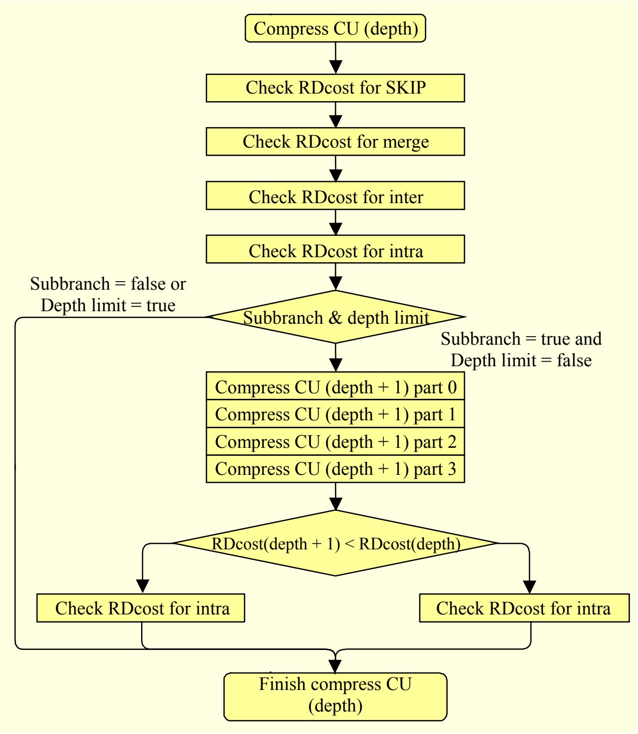


Fig. 1. Flow chart of CU mode decision in the HEVC reference software.

coding (CABAC). Following this, after the inverse quantization and inverse DCT, the encoder evaluates the residual cost through the sum of the squared difference between the reconstructed and original pictures. In addition, the encoder evaluates the equality constraint term, which is derived by multiplying the Lagrangian multiplier with the encoded bits described in the reconstructed picture in the CU. Finally, the encoder evaluates the rate-distortion cost by adding the residual cost and equality constraints for all CU modes, and then selects the best prediction modes having the minimum rate-distortion cost. As a result, the computation time of the conventional CU mode prediction is made up of the generation of the prediction data as well as the transformation and inverse transformation, including an evaluation of the rate-distortion cost. For the generation of the prediction data, it is possible to reduce the computation time through data parallelism, such as the single instruction multiple data (SIMD) technique. On the other hand, because the quantization procedure has to be computed serially, an evaluation of the rate-distortion cost requires a significant amount of computation time in comparison to generating the prediction data. In particular, when the quantization is accomplished based on the rate-distortion optimization to obtain better-quality video compression, the computation time from the rate-distortion cost is increased tremendously.

In attempt to reduce the computation time of a CU mode decision, the parallel computing of each CU mode prediction is a potent solution. However, although it is possible to reduce the total computation time, two main problems still remain. One problem is hardware resources for parallel computing. For the purpose of real-time HEVC encoding, because the parallel processing of a Group of Pictures (GOP), frame, slice, or tile, as well as the CTU, is necessary, the operating system assigns the parallel computing resources to the computational units. If parallel computing resources of the CU prediction modes are added, the HEVC encoder requires about a four-times larger number of resources during parallel computing. Consequently, it is difficult to reduce the computational time by increasing the parallel computing resources. The other problem is the computing time of parallel CU mode prediction, in that the number of bits generated in optimal prediction mode is much less than in other prediction modes. Because many bits for describing a reconstruction picture results in computational burden in the DCT-transformation and quantization, the computing time for the CU mode decision is increased. If the number of computations for each CU mode prediction is equal, the reduction in encoding time achieved through parallel computing could be significant. However, in many cases, because the optimal prediction mode requires less computation time in comparison to other prediction modes, the problems of load balancing and synchronization occur in parallel computing. In particular, the parallel computation time is the same as the lowest computation time of the CU prediction mode owing to the synchronization for the next encoding procedure such as a de-blocking filter or sample adaptive offset. As a result, the performance of parallel computing is dramatically decreased. The detailed procedure of CU mode decision is represented in Fig. 2.

Furthermore, the rate-distortion cost $h(x_{ji}, M)$ for a CU mode decision is as follows:

$$h(\bar{x}_{ji}, M) = D(\bar{x}_{ji}, M) + \lambda \cdot R(\bar{x}_{ji}, M), \tag{1}$$

where $\bar{x}_{ji} \in \mathbb{R}^2$ is the residual data between the reconstruction data and the original data in a CU with a row index of j and column index of i, and M is the index of CU prediction modes such that

$$M \in \{\text{Skip, Merge, Inter, Intra}\} = M_{Id}, \tag{2}$$

where λ is the Largragian multiplier depending on the quantization parameter for HEVC picture encoding; $R(\bar{x}_{ij}, M)$ is the generated bits in CU prediction mode M with header bits for the mode and context, which describes the residual $\bar{x}_{ji}$ with DCT/Q; and $D(\bar{x}_{ji}, M)$ is a distortion function depending on the residual and CU prediction mode, which is evaluated by the sum of the squared difference.

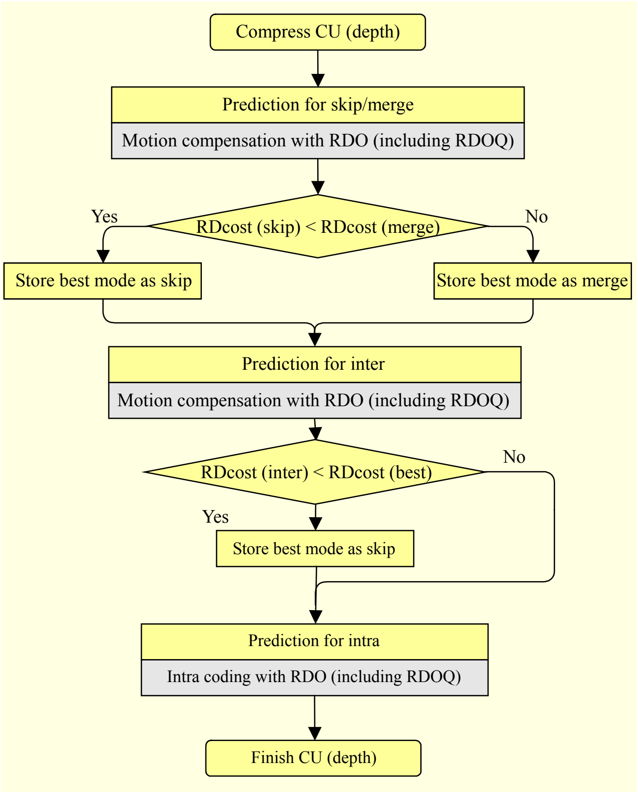


Fig. 2. Detailed flow chart of CU mode decision in HEVC reference software.

III. Proposed Method for CU Mode Prediction

1. Separated Structure of Encoder

As we discussed in the previous section, the computation time is in accordance with the rate-distortion cost, which is difficult to optimize through parallel computing. Additionally, it is impossible to remove the computations according to the rate-distortion, in that this generates a reconstruction picture for the CU. Consequently, we propose a novel encoder structure that is able to reduce the computation time by avoiding as many computations of the rate-distortion as possible. We present a flow chart of the proposed encoding structure in Fig. 3.

As shown in Fig. 3, the conventional computation of the rate-distortion is replaced with the pseudo rate-distortion. In addition, the rate-distortion is evaluated for one of two cases: inter prediction of the motion compensation or intra prediction mode. Consequently, if we evaluate a pseudo rate-distortion based on such a small amount of computations in comparison to a conventional rate-distortion, it is possible to reduce the computation time for a CU mode decision in that the computations for a rate-distortion are conducted only once during the CU mode decision.

For this purpose, the proposed structure is constructed with a separated prediction part, including pseudo rate-distortion, and a transformation part, including the DCT and quantization. Conversely, in the conventional encoding structure, both the prediction and transformation parts are tightly combined. Henceforth, if we construct a suitable pseudo rate-distortion function, it is possible for the proposed encoder structure to provide faster HEVC encoding with a small amount of degradation in the quality of the compressed video.

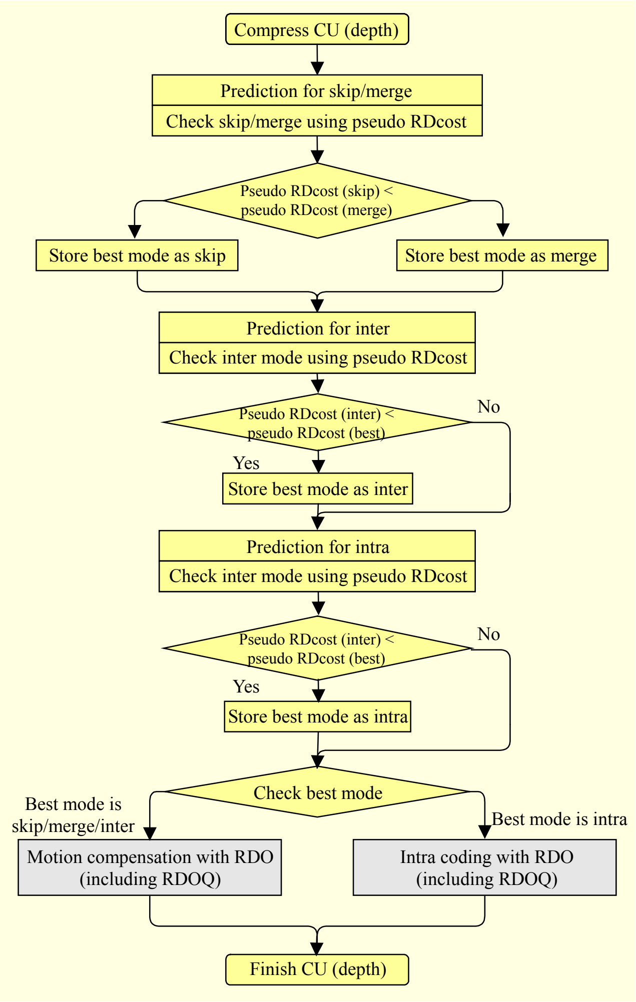


Fig. 3. Proposed encoding structure for CU mode decision.

2. Proposed Pseudo Rate Distortion Function for CU Mode Decision

In the HEVC reference software, for candidates of the intra prediction modes, some data are derived from a Hadamard transformation. Additionally, because a Hadamard transformation is based on simple addition and subtraction, it is possible to obtain a faster result through simple computation properties; regardless, a Hadamard transformation is implemented with a fast algorithm in comparison to a DCT transformation.

As a result, because the computational burden of the Hadamard transformation is less than that of a conventional DCT transformation in spite of providing an additional Hadamard transformation, the number of computations for the rate-distortion cost is less than that for a conventional rate-distortion cost. Furthermore, because it is possible to obtain prediction data for a CU prediction mode using a Hadamard transformation in the prediction candidates derived through a conventional method, the implementation of a pseudo rate-distortion function based on a Hadamard transformation is an alternative method for a conventional rate-distortion function. As a consequence, if we can obtain significant data from another well-defined alternative quantization, it will be possible to define a faster pseudo rate-distortion cost in comparison to conventional rate-distortion.

Based on the idea mentioned above, we define a pseudo rate-distortion function, $f(x_{ji}, M)$, as follows:

$$f(x_{ji}, M) = H(x_{ji}, M) + \sqrt{\lambda} \cdot \bar{R}(x_{ji}, M). \tag{3}$$

In (3), $x_{ji} \in \mathbb{R}^2$ is the residual data between the prediction and original data, but is not the residual data between the reconstruction and original data for a conventional rate-distortion cost. In addition, $H(x_{ji}, M) \in \mathbb{R}$ is the value of the Hadamard transformation depending on the residual x_{ji} and CU prediction mode M for the Hadamard transformation matrix $A = [a_{k,m}] \in \mathbb{R}^{n \times n}$, such that

$$H(x_{ji}, M) = \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{m=0}^{n} \sum_{i=0}^{n} | x_{j,k} a_{k,m} x_{m,i} |, \tag{4}$$

where $\sqrt{\lambda}$ is the square root of the Lagrangian multiplier defined in (2), and $\bar{R}(x_{ji}, M)$ is the estimated bits for describing a picture of the CU using prediction mode M. We define the estimated bit function, $\bar{R}(x_{ji}, M)$, which is a linear combination function with header bits $H_b(M)$ for a CU prediction mode and estimated bits depending on the quantized Hadamard level function $\mathcal{E}(H(x_{ji}, M), qp)$ for the quantization parameter, qp, such that

$$\bar{R}(x_{ji}, M) = H_b(M) + \mathcal{E}\left(H(x_{ji}, M), qp\right). \tag{5}$$

According to (5), if we can evaluate the suitable estimation bits for a specific CU prediction mode, it is possible to obtain a feasible pseudo rate-distortion cost without a DCT transformation or quantization procedure, which consume a significant amount of computation time. Because we do not have any information on the bits for CU mode except for the Hadamard transformation value, $H(x_{ji}, M) \in \mathbb{R}$, the quantization for a pseudo rate-distortion is accomplished for the Hadamard transformation value. Additionally, to maintain the consistency of the quantization, we apply the quantization rule for the DC component in the DCT transformation to the quantization of the Hadamard value. Subsequently, for a CU size of $N \times N$, we can define the following equation as the quantization of the pseudo rate-distortion for a luma component of the Hadamard value $H^Y(x_{ij}, M)$ such that

$$L_Y\left(H(x_{ij}, M), qp\right) = \left(H^Y(x_{ij}, M) \cdot Q_\mathrm{s}(qp\%6)\right) \cdot 2^{-(14 + qp/6 + \log_2 N)}. \tag{6}$$

For the chroma component $H^{Cx}(x_{ij}, M)$, the quantization rule of the pseudo rate-distortion is as follows:

$$L_{Cx}\left(H(x_{ij}, M), qp\right) = \left(H^{Cx}(x_{ij}, M) \cdot Q_s(qp\%6)\right) \cdot 2^{-(13 + qp/6 + \log_2 N)}. \tag{7}$$

In (6) and (7), the % operation denotes the remainder when a numerator is divided by a denominator, and $Q_\mathrm{s}(x)$, which is the scale value for the remainder divided by the denominator, is defined as follows:

$$\begin{gathered} \forall k \in [0, 6) \subset \mathbb{Z}, \\ Q_\mathrm{s}(k) = \{26214, 23303, 20560, 18396, 16384, 14564\}, \end{gathered} \tag{8}$$

where $\mathbb{Z}$ is a set of integers. Subsequently, we define the level of the Hadamard value by the weighted average of the luma and chroma Hadamard values, such that

$$\begin{aligned} \bar{L}\left(H(x_{ij}, M), qp\right) = \frac{1}{8}\Big\{ & L_Y\left(H^Y(x_{ij}, M), qp\right) \cdot 7 \\ & + L_C b\left(H^{Cb}(x_{ij}, M), qp\right) \\ & + L_C r\left(H^{Cb}(x_{ij}, M), qp\right)\Big\}. \end{aligned} \tag{9}$$

As a result, it is possible to define the estimation bits according to the Hadamard-quantization level, which is denoted as (9). For instance, one of the candidates for the estimation bits, evaluated using second-order functions, is as follows:

$$\begin{aligned} \mathcal{E}\left(\bar{L}(x_{ij}, M), qp\right) = & \bar{L}^2\left(H(x_{ij}, M), qp\right) \\ & + \frac{1}{2} \cdot \bar{L}\left(H(x_{ij}, M), qp\right) + B_h. \end{aligned} \tag{10}$$

In (10), B_h is a proportional constant and is set to 2.

The proposed equation of the pseudo rate-distortion is evaluated from the prediction picture and the original picture during the prediction stage of a CU mode decision. Therefore, because it is not necessary to conduct the transformation procedure to obtain the rate-distortion cost based on the reconstruction picture, fast HEVC encoding is enabled in comparison to a conventional method. However, the proposed CU mode prediction based on the pseudo rate-distortion degrades the video quality, which is attributed to the employment of prediction picture data instead of reconstruction picture data. In addition, for fast video encoding, it should be possible to decide whether the encoder accomplishes the CU mode decision or not before entering the decision procedure. In an attempt to solve such problems, we discuss the implementation of a CU mode decision using the proposed pseudo rate-distortion function in the next section.

3. Fundamental Property of Pseudo Rate-Distortion

Suppose that the Hadamard-quantization level of A-mode is less than that of B-mode such that

$$\bar{L}\left(H(x_{ij},A),qp\right)<\bar{L}\left(H(x_{ij},B),qp\right). \tag{11}$$

Additionally, assume that the Hadamard transformation of the chroma component corresponds to the that of the luma component with a small deviation of $\varepsilon \geq 0$ for all modes, such that

$$\begin{gathered} H^{Y}(x_{ij},M_{A})\geq\frac{1}{4}H^{C_x}(x_{ij},M_{B})+\varepsilon, \\ \forall M_{A},M_{B}\in M_{Id},\ M_{A}\neq M_{B}. \end{gathered} \tag{12}$$

Theorem 1. If assumptions (11) and (12) hold and the chroma Hadamard value is the same for each mode, the Hadamard value of each mode satisfies the following inequality under the same quantization parameter.

$$H^{Y}(x_{ij},A)\leq H^{Y}(x_{ij},B). \tag{13}$$

Proof: From the definitions in (6), (7), (9), and (12), we obtain the following relation.

$$\frac{Q_{\mathrm{s}}}{N}\left(7H^{Y}(M_{A})+4H^{C_x}(M_{A})\right)\leq\frac{Q_{\mathrm{s}}}{N}\left(7H^{Y}(M_{B})+4H^{C_x}(M_{B})\right). \tag{14}$$

In (14), N is a proportional constant such that

$$N=2^{(14+qp/6+log_2 N)}. \tag{15}$$

Applying (12) through (14) and expanding the inequality, we obtain

$$\begin{aligned} &\frac{Q_{\mathrm{s}}}{N}\left(7H^{Y}(M_{A})+4H^{C_x}(M_{A})\right)\leq\frac{Q_{\mathrm{s}}}{N}\left(7H^{Y}(M_{B})+4H^{C_x}(M_{B})\right), \\ &\Rightarrow 7H^{Y}(M_{A})+H^{Y}(M_{A})\leq 7H^{Y}(M_{B})+H^{Y}(M_{B}), \\ &\Rightarrow H^{Y}(M_{A})\leq H^{Y}(M_{B}). \end{aligned} \tag{16}$$

■

As a result of theorem 1, the comparison among the pseudo rate-distortion of each mode becomes simple. For simplicity of this discussion, we suppose that the estimated rate function depending on the Hadamard value satisfies the following condition.

$$\forall M_{A},M_{B}\in M_{id},\xi\in\mathbb{Z}^{+},\frac{\partial\mathcal{E}}{\partial\bar{L}}(\bar{L})\cdot\xi>\|H_{b}(M_{A})-H_{b}(M_{B})\|. \tag{17}$$

Theorem 2. When the assumption of theorem 1 holds, and it is supposed that the estimation bit function is analytic, monotone increasing, and convex, the comparison of the pseudo rate-distortion is the same as the comparison of the Hadamard-quantization level under the condition of $H_b(M_A) < H_b(M_B)$ and the assumption of (17), such that

$$\begin{aligned} &\bar{L}\left(H(x_{ij},M_{A}),qp\right)<\bar{L}\left(H(x_{ij},M_{B}),qp\right), \\ &\Leftrightarrow f(x_{ji},M_{A})<f(x_{ji},M_{B}). \end{aligned} \tag{18}$$

Proof:

Sufficiency:

$$\begin{aligned} &\bar{L}\left(H(x_{ij},M_{A}),qp\right)<\bar{L}\left(H(x_{ij},M_{B}),qp\right), \\ &\Rightarrow \varepsilon\left(\bar{L}(M_{A})\right)<\varepsilon\left(\bar{L}(M_{B})\right)\text{by monotone increasing} \\ &\Rightarrow H_{b}(M_{A})+\varepsilon\left(\bar{L}(M_{A})\right)<H_{b}(M_{B})+\varepsilon\left(\bar{L}(M_{B})\right), \\ &\Leftrightarrow \bar{R}(M_{A})<\bar{R}(M_{B}), \\ &\Leftrightarrow \lambda\cdot\bar{R}(M_{A})<\lambda\cdot\bar{R}(M_{B}), \\ &\Rightarrow H(M_{A})+\lambda\cdot\bar{R}(M_{A})<H(M_{B})+\lambda\cdot\bar{R}(M_{B})\text{ by Theorem 1} \\ &\Leftrightarrow f(M_{A})<f(M_{B}). \end{aligned} \tag{19}$$

Necessity:

$$\begin{aligned} &f(M_{A})<f(M_{B}), \\ &\Leftrightarrow H(M_{A})+\lambda\cdot\bar{R}(M_{A})<H(M_{B})+\lambda\cdot\bar{R}(M_{B}), \\ &\Rightarrow H_{b}(M_{A})+\varepsilon(\bar{L}(M_{A}))<H_{b}(M_{B})+\varepsilon\left(\bar{L}(M_{B})\right), \\ &\Rightarrow \varepsilon\left(\bar{L}(M_{A})\right)<H_{b}(M_{B})-H_{b}(M_{A})+\varepsilon\left(\bar{L}(M_{B})\right), \\ &\Rightarrow \varepsilon\left(\bar{L}(M_{A})\right)<\frac{\partial\mathcal{E}}{\partial\bar{L}}(M_{B})\cdot\xi+\varepsilon\left(\bar{L}(M_{B})\right), \\ &\Rightarrow \varepsilon\left(\bar{L}(M_{A})\right)<\varepsilon\left(\bar{L}(M_{B})+\xi\right). \end{aligned} \tag{20}$$

■

Theorem 2 means that if the first-order differential of the Hadamard quantization is larger than the difference among the header bits of the CU modes, it is possible to decide the CU mode using the level with an integer unit.

IV. Implementation of Proposed Method for CU Mode Prediction

1. Prediction of Skip and Merge Modes

Instead of a method that searches for the optimal prediction vector using a conventional rate-distortion cost, we use a scheme in which the optimal prediction vector is obtained using the Hadamard cost for the skip/merge mode decision based on the proposed pseudo rate-distortion [20]. Under the proposed encoder structure, the prediction of the skip and merge modes is always conducted. Hence, the Hadamard value, level, and prediction picture data of the skip/merge mode are also always generated.

In addition, according to (9), when the Hadamard-quantization level is larger than 1, such that

$$\bar{L}\left(H(x_{ij},M),qp\right)>1, \tag{21}$$

a transformation of the skip mode prediction is not conducted. Moreover, if it is larger than 5, the transformation of merge mode is also not conducted. In particular, if the Hadamard-quantization level is equal to zero, the prediction procedures of inter and intra mode are skipped, as are the transformation procedures of inter, intra, and merge modes. We present a flow chart of the proposed encoding procedure for skip/merge mode in Fig. 4.

2. Prediction of Inter and Intra Modes

Because the elision of computations during the prediction procedure includes the skipping of the transformation part of a CU mode decision, the encoding speed is elevated more effectively in comparison to other methods. Conversely, the compressed video is seriously degraded for the same reason. Hence, the selection of a suitable level for the elision of the prediction procedure in inter and intra modes under limited information is important.

The first part we consider is concerned with the elision of the inter prediction. Under the proposed encoding structure, in an attempt to decide whether the inter prediction is conducted or not, we have only information regarding the pseudo rate-distortion cost of skip mode. Consequently, we set the lowest and highest levels to accomplish the inter prediction. If the Hadamard-quantization level is under or above the lowest or highest level, respectively, we do not conduct the inter prediction procedure. The lowest level for the elision of inter prediction is zero, which is the threshold for conducting a skip mode operation. On the other hand, the highest level is defined according to the quantization parameter and the slice depth. For ten video sequences with a resolution of 3,840 by 2,160, we evaluate the highest level using a simple least-mean-square (LMS) method to minimize the BD-rate difference between two algorithms. One algorithm is applied at the highest level, as we mentioned above, and the other is not. The high level of inter mode elision is represented as Fig. 5.

The other part is the intra prediction elision. In comparison to the prediction mode using a motion vector, the property of the picture in the CU is generally quite different. Hence, when a comparative low Hadamard quantization level is presented, it is natural to infer that the intra prediction mode can be skipped. However, when the current CU is predicted as an intra mode, or all samples in the CU have almost the same value, the level of intra prediction is less than that of the other modes. Therefore, a high Hadamard-quantization level does not always indicate an intra mode.

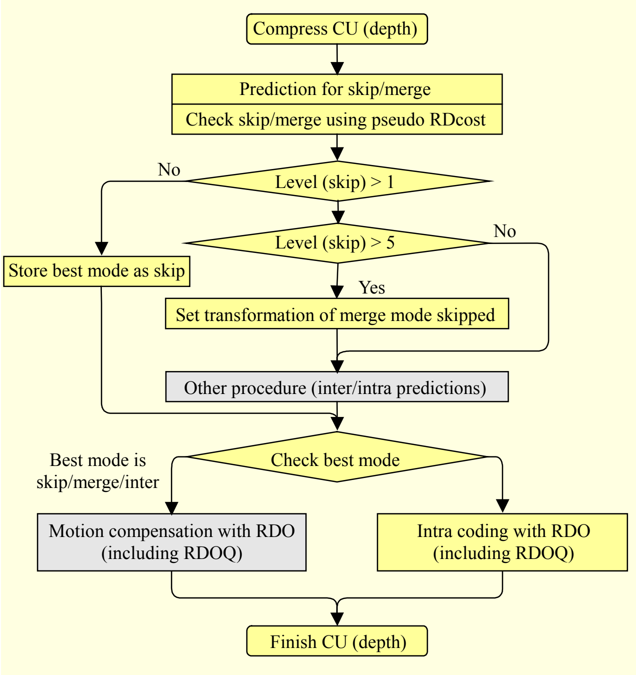
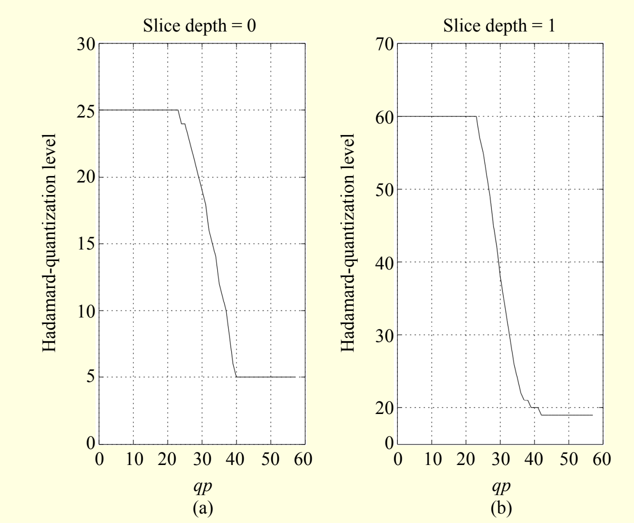


Fig. 4. Prediction procedure of skip/merge mode.



(a) (b)

Fig. 5. High level for inter mode elision.

In a similar manner of inter prediction, we omit the intra prediction when the Hadamard-quantization level is equal to zero. Additionally, because the Hadamard-quantization level of inter mode is predicted before that of intra mode, we also omit the intra prediction when the level of inter mode is equal to zero. Furthermore, for the purpose of a more accurate elision of intra mode prediction, we evaluate the Hadamard-quantization level using a rescaled quantization parameter instead of the original quantization parameter in the current CU when the

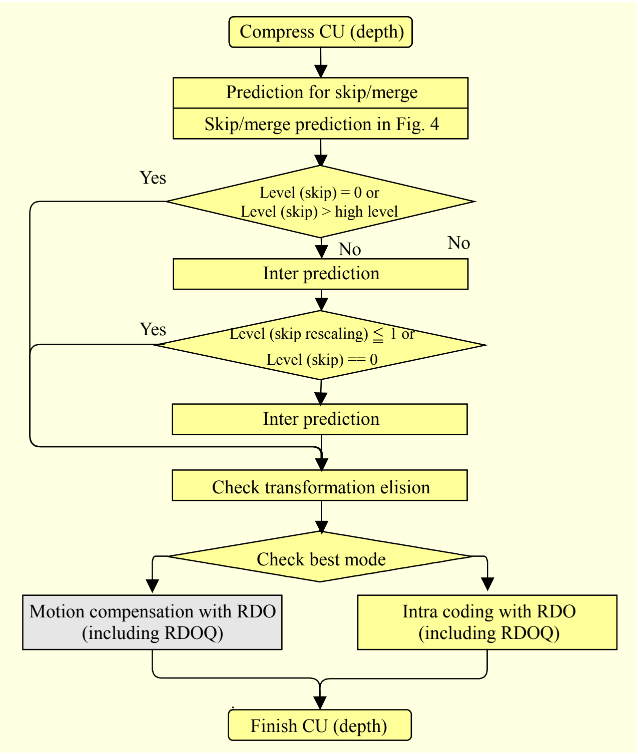


Fig. 6. Prediction procedure of inter/intra mode.

quantization parameter is larger than 29 and the level of skip mode is not zero. If we omit the intra prediction when the level is less than or equal to 1 when the quantization parameter is not taken into account, the degradation of video quality represented through the BD-rate is relatively enlarged when the quantization parameter is larger than 29.

Consequently, when it is larger than 29, we always set the quantization parameter for the rescaled Hadamard-quantization level to 29. As a result, it is possible to omit the intra prediction when the Hadamard-quantization level with the rescaled quantization parameter is larger than zero and less than or equal to 1. We present a flow chart of the proposed encoding procedure for skip/merge mode in Fig. 6.

3. Elision of Transformation Mode

Whereas the elision of the prediction procedure employs a strong condition, the elision of the transformation procedure uses an intuitive condition, in that it is possible to obtain the pseudo rate-distortions of each mode through prediction procedures.

At first, considering the elision of the transformation procedure for merge or intra mode, it is possible to accommodate the results of the theorem because the header of merge mode has 2 bits and the header of intra mode has 6 bits to 13 bits. Thus, in the decision of merge and intra modes, when the level is different, we can select the mode with a low level as the activation mode of transformation. When the level is the same, we can select the mode with a low pseudo rate-distortion as an activation of the transformation mode, and the other mode is not. This procedure is represented as follows:

Step 1: $M = \left(\bar{L}(\text{Merge}) < \bar{L}(\text{Intra})\right)?\,\text{Merge: Intra}.$

Step 2: If $\bar{L}(\text{Merge}) = \bar{L}(\text{Intra})$,

$$M = \bar{f}(x_{ji}, \text{Merge}) < \bar{f}(x_{ji}, \text{Intra})\,?\,\text{Merge: Intra}.$$

In step 2, because the difference with the same level has the same estimated bits, $\bar{f}$ is defined as follows:

$$\bar{f}(x_{ji}, M) = H(x_{ij}, M) + \sqrt{\lambda}\cdot H_b(M). \qquad (22)$$

Second, consider a case of merge or inter mode. In terms of which number of header bits in inter mode is larger than that of intra mode, it is possible to accommodate the above procedure to the elision of the transformation for merge and inter modes. Therefore, the decision procedure for the elision of merge and inter modes is represented as follows:

Step 1: $M = \left(\bar{L}(\text{Merge}) < \bar{L}(\text{Inter})\right)?\,\text{Merge: Inter}.$

Step 2: If $\bar{L}(\text{Merge}) = \bar{L}(\text{Inter})$,

$$M = \bar{f}(x_{ji}, \text{Merge}) < \bar{f}(x_{ji}, \text{Inter})\,?\,\text{Merge: Inter}.$$

However, the decision procedure used for inter and intra modes is more complex. Unlike the previous cases, because the number of header bits of both modes is not definitely smaller or larger than in the other mode, a direct comparison of the level in both modes is inappropriate. Consequently, by setting a range for the decision, we can avoid a wrong decision from a direct comparison of the level in both modes. According to (20) in theorem 2, it is sufficient for the size of the range to be 1, that is, $\xi = 1$. Thus, the decision procedure for inter and intra modes is represented as follows:

Step 1: $M = \left(\bar{L}(\text{Intra}+1) < \bar{L}(\text{Inter})\right)?\,\text{Intra} : \varnothing$

Step 2: $M = \left(\bar{L}(\text{Inter}+1) < \bar{L}(\text{Intra})\right)?\,\text{Inter} : \varnothing$

Step 3: If $M = \varnothing$

$$M = \bar{f}(x_{ji}, \text{Intra}) < \bar{f}(x_{ji}, \text{Inter})\,?\,\text{Intra: Inter}$$

V. Experimental Results

The proposed method was implemented based on the HM 14.0 encoder, which was used as an anchor in the experiments. The configuration of the HEVC encoder was set to the common test condition with 30 tiles, and for the purpose of a fair experiment, all prediction parts were optimized with SIMD as far a possible. The test video sequences used in the experiment were provided from TUT with 4 K resolution (3,840 × 2,160). Moreover, we conducted the experiment for HD (1,920 × 1,080/ 2,560 × 1,600) test sequences provided by JCT-VC as shown in Tables 1 and 2. The other conditions used in the experiment are

Table 1. Simulation results for 4 K test sequences.

Sequence	QP	Δ Bit rate (%)	Δ PSNR (dB)	Δ Time (%)	BD rate Y (%)	BD rate U (%)	BD rate V (%)
Jockey (3,840 × 2,160, 50 Hz)	22	–0.4538	–0.0322	–36.38	2.3	–0.7	0.0
	27	17.3325	0.1196	–31.88			
	32	29.3219	0.4096	–27.90			
	37	24.9931	0.5650	–36.28			
YachtRide (3,840 × 2,160, 50 Hz)	22	–23.7451	–0.8999	–40.47	3.4	1.5	1.5
	27	–24.2894	–1.1538	–41.11			
	32	–23.1140	–1.1218	–44.12			
	37	–32.9509	–1.5629	–50.62			
ReadySetGo (3,840 × 2,160, 50 Hz)	22	1.3979	–0.0288	–36.16	4.0	2.7	2.8
	27	2.2003	–0.0334	–41.46			
	32	1.7347	–0.0659	–43.34			
	37	2.7730	–0.1199	–46.02			
ShakeNDry (3,840 × 2,160, 50 Hz)	22	–47.4686	–0.7667	–47.56	1.3	5.6	5.1
	27	–26.3769	–0.6968	–36.46			
	32	–14.8188	–0.4020	–31.23			
	37	–16.5656	–0.3900	–42.81			
Bosphorus (3,840 × 2,160, 50 Hz)	22	0.4393	–0.0094	–30.07	1.3	0.7	1.6
	27	0.8322	–0.0135	–30.28			
	32	0.0981	–0.0302	–34.68			
	37	–0.0946	–0.0382	–44.77			
Average	N/A	–6.4400	–0.3100	–38.68	2.5	2.0	2.2

Table 2. Simulation results for 2 K test sequences provided by JCT-VC.

Sequence	QP	Δ Bit rate (%)	Δ PSNR (dB)	Δ Time (%)	BD rate Y (%)	BD rate U (%)	BD rate V (%)
Kimono (1,920 × 1,080, 24 Hz)	22	0.1054	–0.0471	–36.17	1.7	3.7	5.2
	27	0.2608	–0.0445	–32.23			
	32	0.5334	–0.0253	–33.06			
	37	2.3617	–0.0319	–40.42			
ParkScene (1,920 × 1,080, 24 Hz)	22	–6.9629	–0.4291	–43.68	2.8	0.8	1.8
	27	–3.7985	–0.2063	–39.17			
	32	–2.5713	–0.0921	–41.90			
	37	0.2956	–0.1147	–47.34			
Cactus (1,920 × 1,080, 50 Hz)	22	–14.7401	–0.3278	–43.31	4.6	1.8	1.3
	27	–3.4594	–0.1207	–42.37			
	32	–1.1223	–0.1624	–37.06			
	37	0.8919	–0.3084	–37.83			
BasketballDrive (1,920 × 1,080, 50 Hz)	22	–5.9872	–0.4793	–31.45	5.3	2.8	2.5
	27	–1.9627	–0.1592	–37.67			
	32	–0.9571	–0.0462	–33.81			
	37	1.8106	–0.0738	–43.32			
Stream (2,560 × 16, 50 Hz, 10 bit)	22	–2.9499	–0.7703	–37.71	1.9	2.1	1.4
	27	–3.0487	–0.2681	–45.51			
	32	–4.5370	–0.2318	–45.91			
	37	–0.6337	–0.188	–42.67			
Average	N/A	–2.3200	–0.210	–39.63	3.3	2.2	2.5

Table 3. Experimental conditions.

Item	Contents
System	Intel Xeon 2.67 GHz 2 CPUs
OS	MicroSoft Windows 2008 Server
QP	22, 27, 32, 27
GOP structure	Random Access
Profile	Main10 (for 4 K)/ Main (for HD)
Entropy coding	CABAC
Maximum partitioning depth	4
Largest CU size	64 × 64

represented as Table 3.

In addition, we evaluated the compressional performance of the proposed method in terms of the change in average bit-rate, peak signal-to-noise ratio (PSNR), and total encoding time, which were reported based on the following equation.

$$\Delta BR(\%) = \frac{\text{Bit rate (Proposed)} - \text{Bit rate (Anchor)}}{\text{Bit rate (Anchor)}} \times 100, \quad (23)$$

$$\Delta PSNR = \text{Y-PSNR (Proposed)} - \text{Y-PSNR (Anchor)}, \quad (24)$$

$$\Delta Time = \frac{\text{Time (Proposed)} - \text{Time (Anchor)}}{\text{Time (Anchor)}} \times 100. \quad (25)$$

The experimental results are shown in Tables 1 and 2. Note that, with the proposed method, the PSNR decrease is 0.31, but that the bit-rate decrease is also 6.44%, which infers a negligible rate-distortion performance degradation for the proposed method. In terms of the overall encoding performance, an average saving in encoding time of 38.68% is achieved with a 2.5% BD-rate loss.

VI. Conclusions

In this paper, we proposed a fast algorithm for a CU mode decision based on a Hadamard transform. In addition, we proposed a separate encoding structure supporting easy implementation of a fast encoding algorithm. In contrast with a conventional HEVC encoder employing an encoding structure with combined prediction and transformation, we showed that the proposed separated encoding structure provides the implementation of many fast encoding algorithms with a low degradation in video quality. The experiment results show that the proposed algorithm encodes a video sequence of 4 K resolution with a time reduction of 38.68% and a quality loss of 2.5%.

References

[1] ITU-T Q6/16 and ISO/IEC JTC1/SC29/WG11, *Joint Call for Proposals on Video Compression Technology*, Kyoto, Jan. 2010.

[2] S.Y. Jeong et al., “Highly Efficient Video Codec for Entertainment-Quality,” *ETRI J.*, vol. 33, no. 2, Apr. 2011, pp. 145–154.

[3] H.H. Lee et al., “Enhanced Block-Based Adaptive Loop Filter with Multiple Symmetric Structures for Video coding,” *ETRI J.*, vol. 32, no. 4, Aug. 2010, pp. 626–629.

[4] J. Xin, A. Vetro, and H. Sun, “Efficient Macroblock Coding-Mode Decision for H.264/AVC Video Coding,” *Proc. Picture Coding Symp.*, Dec. 2004.

[5] J.C. Wang et al., “A Fast Mode Decision Algorithm and Its VLSI Design for H.264/AVC Intra-prediction,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 10, Oct. 2007, pp. 1414–1422.

[6] C.C. Lien and C.P. Yu, “A Fast Mode Decision Method for H.264/AVC Using the Spatial-Temporal Prediction Scheme,” *IEEE Int. Conf. Pattern Recogn.*, Hong Kong, Aug. 20–24, 2006, pp. 334–337.

[7] C.C. Cheng and T.S. Chang, “Fast Three Step Intra Prediction Algorithm for 4 × 4 blocks in H.264,” *IEEE Int. Symp. Circuits Syst.*, Kobe, Japan, May23–26. 2005, pp. 1509–1512.

[8] B. Meng, and O.C. Au, “Fast Intra-prediction Mode Selection for 4 × 4 blocks in H.264,” *Int. Conf. Acoust., Speech, Signal Process.*, Hong Kong, Apr. 2003.

[9] Y.D. Zhang, F. Dai, and S.X. Lin, “Fast 4 × 4 Intra-prediction Mode Selection for H.264,” *Proc. IEEE Int. Conf. Multimedia Expo*, 2004.

[10] J.F. Wang et al., “A Novel Fast Algorithm for Intra Mode Decision in H.264/AVC Encoders,” *Proc. IEEE Int. Symp. Circuits Syst.*, Kos, Greece, May 21–24, 2006, pp. 3498–3501.

[11] T. Tsukuba et al., “H.264 Fast Intra-prediction Mode Decision Based on Frequency Characteristic,” *Proc. Eur. Signal Process. Conf.*, Antalya, Turkey, Sept. 4–8, 2005.

[12] T. Hattori and K. Ichige, “Intra-prediction Mode Decision in H.264/AVC using DCT Coefficients,” *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Tottori, Japan, Dec. 12–15, 2006, pp. 135–138.

[13] M.C. Hwang et al., “Fast Intra Prediction Mode Selection Scheme Using Temporal Correlation in H.264,” *Proc. IEEE TENCON 05*, Nov. 21–24, 2005, pp. 1–5.

[14] A. Elyousfi, A. Tamtaoui, and H. Bouyakhf, “Fast Mode Decision Algorithm for Intra prediction in H.264/AVC Video Coding,” *IJCSNS Int. J. Compu. Sci. Netw. Security*, vol. 7, no. 1, Jan. 2007, pp. 356–364.

[15] F. Pan et al., “Fast Mode Decision Algorithm for Intra Prediction in H.264/AVC Video Coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, July 2005, pp. 813–822.

[16] D. Jun and H. Park, "An Efficient Priority-Based Reference Frame Selection Method for Fast Motion Estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, Aug. 2010, pp. 1156–1161.

[17] B.G. Kim, J.H. Kim, and C.S. Cho, "Fast Inter Mode Decision Algorithm Based on Macroblock Tracking in H.264/AVC Video," *ETRI J.*, vol. 29, no. 6, Dec. 2007, pp. 736–744.

[18] B.G. Kim and J.H. Kim, "Efficient Intra-mode Decision Algorithm for Inter-frames in H.264/AVC Video Coding," *IET, Image Process.*, vol. 5, no. 3, Apr. 2011, pp. 286–295.

[19] X. Lu et al., "Fast Mode Decision and Motion Estimation for H.264 with a Focus on MPEG-2/H.264 Transcoding," *IEEE Int. Symp. Circuits Syst.*, Kobe, Japan, May 23–26, 2005, pp. 1246–1249.

[20] JCT-VC, "High Efficiency Video Coding (HEVC) Test Model 2 (HM 2) Encoder Description," JCTVC Document, JCTVC-D502, Jan. 2011.

[21] JCT-VC, "WD2: Working Draft 2 of High-Efficiency Video Coding," JCTVC Document, JCTVC-D503, Jan. 2011.

[22] L. Zhao et al., "Fast Mode Decision Algorithm for Intra Prediction in HEVC," *IEEE Visual Commun. Image Process.*, Tainan, Taiwna, Nov. 6–9, 2011, pp. 1–4.

[23] J.H. Lee et al., "Intra-mixture Prediction Mode and Enhanced Most Probable Mode Estimation for Intra Coding," *ETRI J.*, vol. 31, no. 5, Oct. 2009, pp. 610–612.

[24] Y. Kim et al., "A Fast Intra-prediction Method in HEVC Using Rate-Distortion Estimation Based on Hadamard Transform," *ETRI J.*, vol. 35, no. 2, Apr. 2013, pp. 270–280.

[25] D. Yoon and Y. Ho, "Fast Mode Decision Method for HEVC in Depth Video," *J. Korean Inst. Commun. Inf. Sci.*, vol. 37, no. 1, 2012, pp. 51–56.

[26] S.H. Jung, and H.W. Park, "A Fast Mode Decision Method in HEVC Using Adaptive Ordering of Modes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, 2015, pp. 1–13.

[27] S. Yang et al., "Fast Inter Mode Decision Process for HEVC Encoder," *IEEE IVMSP Workshop*, Seoul, Rep. of Korea, June 10–12, 2013, pp. 1–4.

Jinwuk Seok received his BS and MS degrees in electrical control engineering from Hong-Ik University, Seoul, Rep. of Korea in 1993 and 1995, respectively, and his PhD in electrical engineering from Hong-Ik University in 1998. He has been a principal member of the engineering staff at ETRI, Daejeon, Rep. of Korea since 2000, and an adjunct professor of the Computer Software Engineering Department at the University of Science and Technology, Daejeon, Rep. of Korea since 2009. His research interests include video compression, pattern recognition, neural networks, and stochastic nonlinear control.

Younhee Kim received her BS and MS degrees in computer science from Ajou University, Suwon, Rep. of Korea in 2000 and 2002, respectively, and her PhD in computer science from George Mason University, Fairfax, VA, USA in 2009. She has been a senior researcher with ETRI since 2009. She is currently involved in the development of an HEVC video real-time encoder. Her current research interests include video coding, image and video signal processing, optimized video encoding, and information hiding in the field of multimedia communication.

Myungseok Ki received his BS and MS degrees in computer engineering from Chonnam University, Gwangju, Rep. of Korea in 1999 and 2001, respectively. He is currently a senior member of the research staff at ETRI. His research interests include image processing, multimedia streaming, and interactive broadcasting systems.

Hui Yong Kim received his BS and MS degrees and his PhD from Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea in 1994, 1998, and 2004, respectively. From 2003 to 2005, he was the leader of the Multimedia Research Team of AddPac Technology Co., Seoul, Rep. of Korea, Ltd. In 2005, he joined the Broadcasting and Telecommunications Media Research Laboratory of ETRI, and currently serves as the director of the Visual Media Research Section. From 2006 to 2010, he was also an affiliate professor at UST. From 2013 to 2014, he was a visiting scholar at the Media Communications Lab of the University of Southern California, Los Angeles, USA. He has made many contributions to the development of international standards such as MPEG's Multimedia Application Format and JCT-VC's High Efficiency Video Coding as an active technology contributor, editor, and ad-hoc group co-chair. His current research interests are in image and video signal processing and compression for realistic media applications such as UHD, 3D, VR, and HDR.

Jin Soo Choi received his BE and ME degrees and his PhD in electronic engineering from Kyungpook National University, Daegu, Rep. of Korea in 1990, 1992, and 1996, respectively. Since 1996, he has been a principal member of the research staff in ETRI. He has been involved in developing the MPEG-4/HEVC codec system, data broadcasting system, and 3D/UHDTV broadcasting system. His research interests include visual signal processing, video coding, and realistic media technologies.