Highly Efficient Video Codec for Entertainment-Quality

Seyoon Jeong, Sung-Chang Lim, Hahyun Lee, Jongho Kim, Jin Soo Choi, and Haechul Choi

We present a novel video codec for supporting entertainment-quality video. It has new coding tools such as an intra prediction with offset, integer sine transform, and enhanced block-based adaptive loop filter. These tools are used adaptively in the processing of intra prediction, transform, and loop filtering. In our experiments, the proposed codec achieved an average reduction of 13.35% in BD-rate relative to H.264/AVC for 720p sequences.

Keywords: Video coding, H.264/AVC, coding efficiency.

Manuscript received Mar. 6, 2010; revised Oct. 11, 2010; accepted Oct. 25, 2010.

The work was supported by the IT R&D program (KI001932, Development of Next Generation DTV Core Technology) of KEIT&KCC&MKE, Rep. of Korea.

Seyoon Jeong (phone: +82 42 860 5724, email: jsy@etri.re.kr), Sung-Chang Lim (email: sclim@etri.re.kr), Hahyun Lee (email: hanilee@etri.re.kr), Jongho Kim (email: pooney@etri.re.kr), and Jin Soo Choi (email: jschoi@etri.re.kr) are with the Broadcasting & Telecommunications Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Haechul Choi (corresponding author, email: choihc@hanbat.ac.kr) is with the Division of Information Communication and Computer Engineering, Hanbat National University, Daejeon, Rep. of Korea.

doi:10.4218/etrij.11.0110.0126

I. Introduction

A large quantity of video material is already being distributed digitally over broadcast channels, digital networks, and packaged media. More and more of this material will be distributed with increased resolution and quality. Recently, 4k×2k video (3840×2160) digital cameras have already shown up in the market, and display devices supporting 4k×2k spatial resolution are also appearing on the horizon. In addition, digital cinema is now capturing 4k×2k video to provide a captivating entertainment-quality experience. Evolution in technology will soon make possible the capture and display of video material with a quantum leap in quality, whereas networks are already finding it difficult to carry a large number of data rates for HDTV resolution to the end user. Moreover, further data-rate increases resulting from 4k×2k video will put additional pressure on the networks. Therefore, a new video compression technology that has sufficiently higher compression capability than the existing H.264/Advanced Video Coding (AVC) [1] standard is needed. The ISO/IEC JTC1/SC29 WG11 Moving Picture Experts Group (MPEG) and ITU-T Q.6/16 Video Coding Experts Group (VCEG) have jointly started a new video coding standard that is tentatively named high efficiency video coding (HEVC), and they publically issued a call for proposals on HEVC in January, 2010 [2], [3]. These standard groups urgently encourage new video coding algorithms for their new video coding standards.

In accordance with the status of such a new standard, we propose an enhanced video codec for entertainment-quality applications, such as DVD-video systems, HDTV, and 4k×2k video systems. In such applications, video sequences have

720×480 resolution and beyond, and those bitrates are larger than 3 Mb/s. For high coding efficiency, delay can be allowed. The proposed codec has novel video coding tools, including an intra prediction with offset (IPO), integer sine transform (IST), and enhanced block-based adaptive loop filter (E-BALF). These tools are used adaptively in the processing of intra prediction, transforms, and loop filtering. Moreover, by combining these tools on the top of H.264/AVC, we accomplish a video codec that can provide high-performance coding efficiency for entertainment-quality video.

This paper is structured as follows. Section II describes the proposed video codec including new coding tools. Section III shows experimental results, followed by a conclusion in section IV.

II. High Coding Efficient Video Codec

1. Codec Overview

The encoder structure of H.264/AVC is illustrated in Fig. 1. It also includes our proposed coding tools, which are presented with gray boxes. As shown in Fig. 1, a typical block-based hybrid video codec is composed of many processes, including intra prediction and interprediction, transforms, quantization, entropy coding, and filtering. Video coding technologies have been maturing through intensive research and development for a long time. To achieve significantly higher coding efficiency than current mature video codecs, various coding tools covering many processes must be developed in an efficiently combined way.

We have thoroughly studied H.264/AVC, which is the state-of-the-art video coding standard, to improve its coding performance. To obtain more attractive quality than the best one supported by H.264/AVC at the same bitrate, we have developed various normative algorithms that change both the decoding and encoding processes. The proposed video codec has three novel coding tools including the IPO, IST, and E-BALF. These proposed tools are switchable, and thus each of them is selectively used in the sense of rate-distortion optimization (RDO).

An IPO is an intra predictive coding tool that estimates an original signal by referring to reconstructed signals within a current slice. An accurate prediction can reduce the quantity of the signal to be coded. This is because only a residual signal, which is the difference between the original and predictive signals, is transmitted. An IPO compensates for the DC difference between the original and reference signals and can produce a more accurate prediction signal, particularly in cases where there is an illumination change across spatial regions.

An IST is a sine transform that can compact a low-correlated

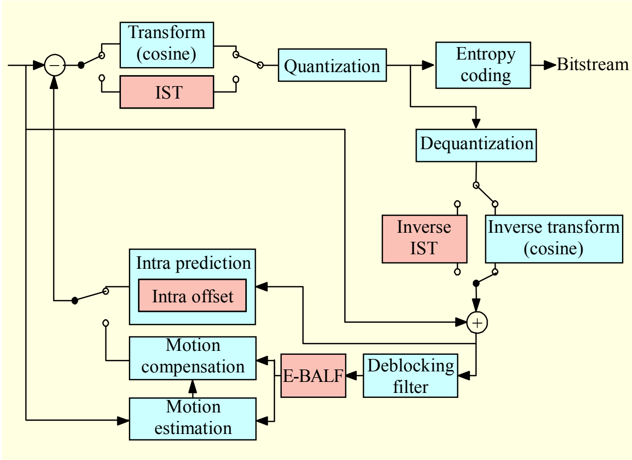


Fig. 1. Encoder block diagram of proposed video codec. Gray boxes are the proposed tools, and white boxes are H.264/AVC tools.

signal more highly than the integer transform of H.264 based on the cosine transform [4]. The higher compaction can lead to higher compression with the help of an appropriate quantization method, such as a nonlinear quantizer arranging larger step sizes at higher frequency. An IST can be applied to all signals regardless of the prediction method whether it be the intra prediction, inter prediction, or differential pulse-code modulation.

An E-BALF is an adaptive loop filter used to enhance the subjective quality of video as well as its objective quality. An adaptive loop filter is applied to a completely reconstructed signal, and the filtered signal is then used as a reference signal for subsequent pictures. An E-BALF makes a reconstructed signal more similar to a corresponding original signal, which mitigates information losses caused by coding processes such as quantization and deblocking filters. Filter coefficients of the E-BALF are determined on a slice-by-slice basis in the sense of minimization of the mean square error between the original and reconstructed signals. Note that the optimal filter coefficients should be transmitted. To reduce the quantity of bits for the filter coefficients, adaptive loop filter methods use a small number of unique filter coefficients by assuming symmetries across the horizontal, vertical, or centroid axes. It is a fact that the assumption of filter coefficients affects the performance of the filter. Since the optimum assumption to achieve high coding efficiency depends on picture contents, a strict and constant assumption across all pictures may degrade the performance of the adaptive loop filter. The proposed E-BALF uses various symmetric assumptions and makes a decision on which symmetric assumption is applied to reduce the number of filter coefficients. The decision is conducted slice-by-slice, and a flag indicating the determined symmetric assumption is transmitted at every slice.

2. Intra Prediction with Offset

In H.264/AVC, intra coding based on various directional predictions improves the coding efficiency by removing spatial redundancy across neighboring blocks. In detail, the current block to be coded can be predicted by using neighboring reconstructed pixels as a reference signal. If intra prediction mode is selected, the error between original and predictive signals is coded. To further reduce the prediction error, we introduce an IPO [5]. The IPO can contribute toward obtaining a more accurate prediction signal, for which the offset value should be determined through an RDO process. In the proposed IPO, each intra-coded macroblock can have a particular offset value, which is transmitted to a decoder. The following simple equation describes the IPO scheme:

$$pred\_block_{\text{offset}}(x, y) = pred\_block(x, y) + \alpha, \qquad (1)$$

where *pred_block*$_{\text{offset}}$ indicates an offset-compensated prediction block, and *pred_block* represents a prediction block made by the intra prediction process of H.264. The value of α is the integer offset. In point of complexity, as in (1), the operation for the proposed method at the encoder side is very simple, and the decoder also needs only 256 additions per macroblock when the offset value is not equal to zero. Moreover, the proposed method can be used for any type of intra prediction mode such as Intra_16×16, Intra_8×8, and Intra 4×4.

The optimum offset value is determined at the macroblock layer and is sent to a decoder. Thus, all pixels within one macroblock are compensated with one offset value. Basically, an IPO in the spatial domain has the same concept with a DC offset in the frequency domain. In other words, the offset plays a role as DC compensation in the frequency domain and is added to the current block. However, the dynamic range of the offset in the spatial domain is smaller than that of the DC value in the frequency domain. Therefore, it is beneficial to use an IPO scheme in the spatial domain.

3. Integer Sine Transform

In a predictive coding method, a residual signal, which is the difference between original and predictive signals, is coded. When an original signal is well predicted, the correlation of the residual signal is subject to a substantial decrease. For this kind of low correlated signal, a discrete cosine transform/integer cosine transform (ICT) may not appropriate. On the other hand, the sine transform is known as a sub-optimal substitute for the Karhunen-Loève transform for low correlated signals [6]. Thus, if the transform can be switchable according to the signal correlation, gain in coding efficiency can be achieved. We derived the IST from the discrete sine transform. In the proposed codec, the IST is alternatively used with the ICT as shown in Fig. 1 [4]. The derived 4×4 forward IST is

$$Y = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 1 & 1 & -1 & -1 \\ 2 & -1 & -1 & 2 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} X \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & -1 & -1 \\ 2 & -1 & -1 & 1 \\ 1 & -1 & 2 & -1 \end{bmatrix}, \qquad (2)$$

where X is the residual signal, and Y represents transformed coefficients. After performing the forward IST, the quantization for the transformed coefficients of the 4×4 IST is given by

$$Z_{(i,j)} = \operatorname{sgn}\left(Y_{(i,j)}\right)\cdot\left(\left|Y_{(i,j)}\right|\cdot MF_{(i,j)} + DZ\right) >> \left(15 + Q_D\right), \qquad (3)$$

where $MF_{(i,j)}$ represents the multiplication factor, and DZ controls the dead zone. The sign function is represented by sgn(·), and Q_D represents the greatest integer smaller than or equal to QP/6. The corresponding dequantization is given by

$$Y'_{(i,j)} = \left(Z_{(i,j)} \cdot SF_{(i,j)}\right) << Q_D, \qquad (4)$$

where $SF_{(i,j)}$ is the scaling factor. The following equation represents the inverse transform of the 4×4 IST.

$$X' = \begin{bmatrix} \frac{1}{2} & 1 & 1 & 1 \\ 1 & 1 & -\frac{1}{2} & -1 \\ 1 & -1 & -\frac{1}{2} & 1 \\ \frac{1}{2} & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} Y' \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ 1 & 1 & -1 & -1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}. \qquad (5)$$

The 4×4 IST components are derived from the 4×4 DST-II in a similar way to the 4×4 ICT of H.264/AVC. The 8×8 IST components are also derived from the 8×8 DST-II. The details are found in [4].

The multiplication and scaling factors used in quantization and dequantization for the 4×4 IST are tabulated in Table 1, where Q_M indicates QP mod 6. Since the same quantization method as in H.264 is applied to the 4×4 IST, the post-scaling factor of the IST consists of the same values as the post-scaling factor of the ICT except for the positions of the values.

The proposed transform method utilizing the RDO process selects an optimal transform between the ICT and IST by introducing a flag for signaling the identification of a selected transform. That is, an encoder sends an additional flag per macroblock to the decoder. In principle, the proposed method can be applied to every 4×4 block or 8×8 block in a macroblock. However, 16 flag bits or 4 flag bits per macroblock may be a burden for the coding efficiency. Therefore, we designed the process in such a way that only one transform between the IST and ICT is used consistently in one macroblock unit. When a macroblock for either the P-frame or B-frame is coded as "SKIP" mode, or the coded block pattern

Table 1. Multiplication and scaling factors for 4×4 IST. P4×4_1 = positions for (0, 0), (2, 0), (2, 2), and (0, 2) in 4×4 matrix, P4×4_2 = positions for (1, 1), (1, 3), (3, 1), and (3, 3) in 4×4 matrix, and P4×4_3 = other positions except P4×4_1 and P4×4_2.

Q_M	Multiplication factor			Scaling factor		
	4×4 IST			4×4 inverse IST		
	P4×4_1	P4×4_2	P4×4_3	P4×4_1	P4×4_2	P4×4_3
0	5243	13107	8066	16	10	13
1	4660	11916	7490	18	11	14
2	4194	10082	6554	20	13	16
3	3647	9362	5825	23	14	18
4	3355	8192	5243	25	16	20
5	2893	7282	4559	29	18	23

(CBP) of its luminance component is equal to zero, the encoder does not send a flag for the indication of ICT/IST. The reason for no flag is because a residual signal within the macroblock does not exist. Therefore, there is no transform coefficient within the macroblock, and the decoder does not conduct the inverse transform and dequantization process. At the macroblock layer, the maximum number of bits for the indication flag is 4. The 4 bits have to be transmitted in the case where a macroblock is partitioned in "sub-macroblock" mode, and the CBPs of all the sub-macroblocks are not zero (1 bit per 8×8 block).

4. Enhanced Block-Based Adaptive Loop Filter

Chujoh and others [7], [8] proposed a block-based adaptive loop filter (BALF) to improve the coding efficiency of H.264/AVC. The BALF applies a frame-wise adaptive filter to some blocks of a reconstructed frame and signals filter coefficients and information for indicating the filtered blocks per frame. To reduce the number of bits used to transmit the filter coefficients, it is assumed that the statistical properties of an image signal are symmetric about its center as shown in Fig. 2. By this assumption, only 13 unique filter coefficients are transmitted to a decoder side even though a 5×5 Wiener filter is used.

We note that the assumption of symmetry can provide a good trade-off between the accuracy of the loop filter and the overhead bits used to transmit the filter coefficients. However, since the statistical properties of the video sequence can vary spatially and temporally, a fixed single symmetry assumption would not be appropriate for every frame in a whole video sequence. For example, some frames in a video sequence may contain relatively complex scenes that hold neither vertical nor

C0	C1	C2	C3	C4
C5	C6	C7	C8	C9
C10	C11	C12	C11	C10
C9	C8	C7	C6	C5
C4	C3	C2	C1	C0

Fig. 2. A 5×5 filter with central symmetric structure, where only 13 unique filter coefficients are needed.

Table 2. Four filter symmetric structures and associated filter modes used in proposed method.

Symmetric structure	Mode
Central	0
Vertical	1
Horizontal	2
Top-left diagonal	3

(a) Vertical

C0	C1	C2	C3	C4
C5	C6	C7	C8	C9
C10	C11	C12	C11	C10
C5	C6	C7	C8	C9
C0	C1	C2	C3	C4

(b) Horizontal

C0	C5	C10	C5	C0
C1	C6	C11	C6	C1
C2	C7	C12	C7	C2
C3	C8	C11	C8	C3
C4	C9	C10	C9	C4

(c) Top-left diagonal

C0	C1	C2	C3	C4
C5	C6	C7	C8	C3
C9	C10	C12	C7	C2
C11	C8	C10	C6	C1
C4	C11	C9	C5	C0

Fig. 3. Examples of 5×5 Wiener filters each with a vertical, horizontal, or top-left diagonal symmetric structure.

horizontal symmetry, whereas the scenes in other frames may be well characterized by either symmetric structure. For this reason, in addition to the central symmetric structure described in Fig. 2, we define three more filters with different symmetric structures to reflect the varying statistical properties of a video sequence as shown in Table 2 and Fig. 3. To have a decoder know which of the symmetric structures is used, an indicator is also transmitted along with the filter coefficients.

Figure 3 illustrates examples of 5×5 Wiener filters with vertical, horizontal, and top-left symmetric structures. In the figure, the letter on each position represents a filter coefficient index. The indices with the same letter share the same filter coefficient. The proposed method selects the symmetry structure of filter coefficients per frame in order to capture the characteristics of each frame in a video sequence so that the difference between the original and filtered frames can be further minimized.

To determine the optimal filter symmetry structure for a frame among multiple filters, the RDO is used.

$$J = D_F + \lambda \times R_F, \quad (6)$$

where D_F is the distortion measured by the mean square error between the original and filtered frames, λ is the Lagrange multiplier, and R_F denotes generated bits for filter coefficients, the filter symmetry structure indicator, and control flags for block-based filtering. The filter coefficients and filter symmetry structure resulting in minimum rate-distortion cost (RD-cost) J are selected as the optimal filter coefficients and filter symmetry structure.

The proposed method consists of the following four steps.

Step 1. The filter coefficients for each filter symmetry structure are obtained by solving the Wiener-Hopf equations [9].

Step 2. The block-based filtering process using the filter coefficients for each symmetry structure obtained in step 1 is performed. The filtering process is based on a conventional BALF [7].

Step 3. The RD-cost is calculated for each filter symmetry structure.

Step 4. The filter symmetry structure resulting in the minimum RD-cost is selected as the optimal one. Then, the optimal filter symmetry structure and its coding results are coded.

III. Experiments

The proposed video coding tools were implemented on JM 11.0 of H.264/AVC reference S/W [10]. H.264/AVC High Profile is used as an anchor with which the proposed method is evaluated since it is the-state-of-the-art video coding standard. The test sequences were a set of various public sequences that have been used in standardization. The IPO is mainly related with spatial prediction coding, and thus its performance evaluation is conducted under the I-frame-only prediction structure. On the other hand, the IST, E-BALF, and a combination of our tools are conducted under both the IPPP and hierarchical B-picture prediction structures [11], [12]. One hundred frames of each test sequence are coded with the IPPP prediction structure and the I-frame-only prediction structure, and 98 frames are coded with the hierarchical B-picture prediction structure. The BD-rate and BD-PSNR [13], which provide the relative gain between the two methods by measuring the average difference between the two RD-curves, are used as coding performance measurements. To calculate the BD-PSNRs and BD-rates, quantization parameters of 22, 27, 32, and 37 are commonly used for all experiments in this paper. For the entropy coding, the context-adaptive binary arithmetic coding is employed. The test conditions including encoding parameters are the same as the recommended simulation common conditions of VCEG Key Technology Area development [14] except that RDO-Q is disabled.

For the complexity comparison, encoding and decoding runtime ratios between the H.264/AVC and the proposed tools were measured. Consequently, the encoding runtime is relatively increased for most of the proposed tools, whereas the decoding runtime is not increased except the E-BALF. The additional computational efforts at the encoder are because additional modes are introduced into the conventional method. The E-BALF needs an additional decoding computation for a decoder side filtering. Note that particular efforts to optimize algorithm complexity were not made.

As the first evaluation, we checked out the performance of each proposed coding tool. Table 3 shows the performance of the IPO compared with H.264/AVC High Profile. The IPO achieves an average –1.58% BD-rate gain over all test sequences, and the BD-rate ranges from –0.96% to –2.81%. The averages of the BD-rate are –1.40%, –1.37%, and –1.96% for CIF (352×288), 4CIF (704×576), and 720p (1280×720), respectively. A BD-rate value of –x% means that the proposed method can reduce x% of the total bits of the anchor. As listed in Table 3, the performance of the IPO is consistently better than H.264/AVC over all test sequences. At the encoder side, the IPO finds a best offset value from a predefined candidate set in a brute force way, where the IPO calculates RD-cost for

Table 3. Coding performance comparison between IPO vs. H.264/AVC High Profile for I-frame-only prediction structure.

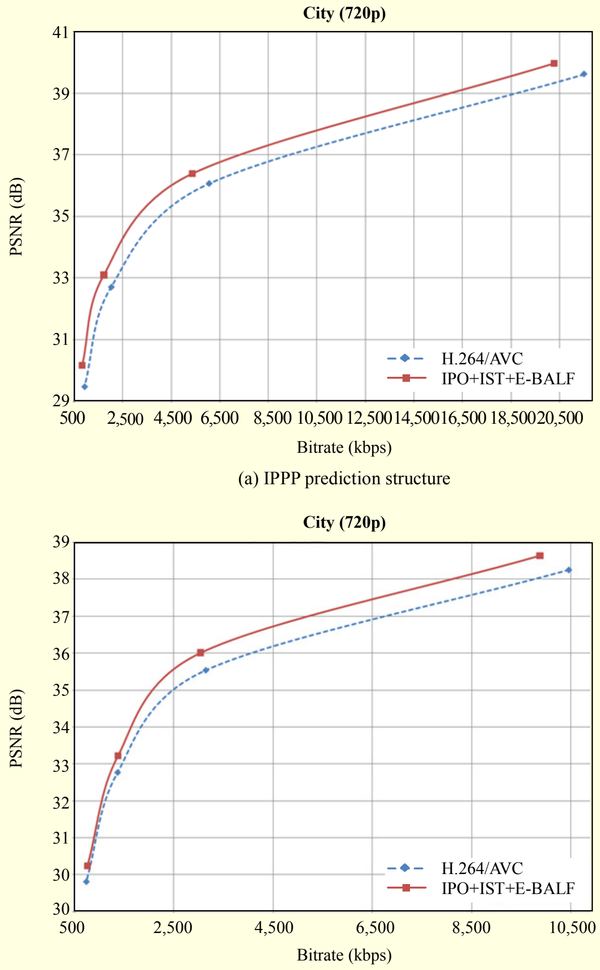
Sequence		I-frame only			
		BD-rate (%)	BD-PSNR (dB)	Time ratio	
				Encoding	Decoding
CIF	Bus	–1.27	0.10	5.42	1.02
	City	–1.33	0.09	4.26	1.01
	Mobile& Calendar	–1.23	0.13	5.40	1.01
	Soccer	–1.40	0.08	5.43	1.00
	Tempete	–1.75	0.15	5.44	1.00
	Average	–1.40	0.11	5.19	1.01
4CIF	City	–1.02	0.07	5.40	1.01
	Crew	–2.06	0.09	5.38	1.00
	Soccer	–1.02	0.06	5.39	1.01
	Average	–1.37	0.07	5.39	1.01
720p	Bigship	–2.06	0.10	5.39	1.01
	City	–0.96	0.07	5.39	1.01
	Night	–2.00	0.14	5.39	0.99
	ShuttleStart	–2.81	0.10	5.46	0.99
	Average	–1.96	0.10	5.41	1.00
Total average		–1.58	0.10	5.33	1.01

Table 4. Coding performance comparison between IST vs. H.264/AVC High Profile for IPPP and hierarchical B-picture prediction structure.

Sequence		IPPP				Hierarchical B-picture			
		BD-rate (%)	BD-PSNR (dB)	Time ratio		BD-rate (%)	BD-PSNR (dB)	Time ratio	
				Enc.	Dec.			Enc.	Dec.
CIF	Bus	–1.25	0.06	1.15	0.99	–1.63	0.08	1.27	0.97
	City	–1.40	0.06	1.14	1.02	–1.99	0.08	1.25	0.89
	Mobile& Calendar	–0.90	0.04	1.16	1.01	–1.83	0.09	1.28	0.86
	Soccer	–0.89	0.04	1.14	0.99	–1.69	0.07	1.27	1.35
	Tempete	–0.92	0.05	1.14	0.99	–1.39	0.07	1.26	0.80
	Average	–1.07	0.05	1.15	1.00	–1.70	0.08	1.26	0.97
4CIF	City	–1.31	0.04	1.15	1.02	–1.23	0.04	1.27	1.08
	Crew	–0.21	0.01	1.16	0.99	–0.35	0.01	1.29	0.99
	Soccer	–0.72	0.03	1.15	0.91	–1.68	0.07	1.28	1.16
	Average	–0.74	0.03	1.15	0.97	–1.09	0.04	1.28	1.08
720p	Bigship	–0.57	0.02	1.16	0.98	–1.34	0.03	1.31	1.05
	City	–1.29	0.04	1.15	0.96	–1.17	0.04	1.27	0.91
	Night	–0.30	0.01	1.16	1.00	–0.70	0.02	1.29	1.01
	Shuttle Start	–0.36	0.01	1.21	0.99	–1.22	0.03	1.37	0.99
	Average	–0.63	0.02	1.17	0.98	–1.11	0.03	1.31	0.99
Total average		–0.84	0.03	1.16	0.99	–1.35	0.05	1.29	1.01

each offset value. Thus, the IPO is on average 5.33 times slower than H.264/AVC. Consider that it is not optimized to computational complexity yet. The IPO is an intra prediction tool, and the number of intra-coded blocks is typically quite lower than inter-coded blocks. Therefore, if an early decision algorithm between inter or intra coding is adopted and the RD-cost for prediction modes are calculated in parallel, the encoding efforts would be significantly lightened without a lot of coding efficiency loss.

The performance of the IST is shown in Table 4. The values of the BD-rate are –0.21% to –1.4% for the IPPP prediction structure and –0.35% to –1.99% for the hierarchical B-picture prediction structure. The encoding runtime of the IST has an average of 1.16 times for the IPPP prediction structure and 1.29 times for the Hierarchical B-picture prediction structure, whereas the decoding runtime increase of the IST is negligible. As described in section II.3, the usefulness of IST is based on the fact that the sine transform is more suitable than the cosine transform for low-correlated signals. Typically, the hierarchical B-picture prediction structure may entail a more accurate prediction than the IPPP prediction structure due to the bi-predictive coding and multihypothesis prediction scheme in H.264/AVC. Therefore, the hierarchical B-picture prediction structure may generate a residual signal with a smaller correlation than the IPPP prediction structure. The IST, thereby, works better under the condition of the hierarchical B-picture prediction structure. Corresponding to this expectation, as shown in Table 4, it is proved that the proposed IST has better performance in the hierarchical B-picture prediction structure.



(a) IPPP prediction structure

(b) Hierarchical B-picture prediction structure

Fig. 4. RD-curves for E-BALF.

The RD-curves of the E-BALF are shown in Fig. 4, and the BD-rate and BD-PSNR are listed in Table 5. The proposed E-BALF achieves enormous coding gain at a high bitrate, while the gain is slightly decreased at a low bitrate. One reason for the difference in performance across bitrate points is that a large quantity of bits for filter coefficients and filter information significantly degrades the coding efficiency at low bitrate points. When computational complexity of the E-BALF is compared with H.264/AVC, encoding runtime increases an average of 1.73 times for the IPPP prediction structure and 1.49

Table 5. Coding performance comparison between E-BALF vs. H.264/AVC High Profile for IPPP and hierarchical B-picture prediction structure.

Sequence		IPPP				Hierarchical B-picture			
		BD-rate (%)	BD-PSNR (dB)	Time ratio Enc.	Time ratio Dec.	BD-rate (%)	BD-PSNR (dB)	Time ratio Enc.	Time ratio Dec.
CIF	Bus	–9.22	0.43	1.61	1.55	–4.95	0.25	1.42	1.02
	City	–5.99	0.29	1.52	1.65	–9.73	0.42	1.39	1.22
	Mobile& Calendar	–6.58	0.32	1.66	1.84	–4.81	0.23	1.45	1.03
	Soccer	–7.54	0.34	1.53	1.44	–8.91	0.36	1.40	1.22
	Tempete	–4.97	0.26	1.57	1.68	–4.27	0.22	1.41	1.25
	Average	–6.86	0.33	1.58	1.63	–6.53	0.30	1.42	1.15
4CIF	City	–15.71	0.57	1.70	1.76	–11.68	0.43	1.49	1.72
	Crew	–10.47	0.35	1.77	1.57	–7.49	0.24	1.53	1.41
	Soccer	–17.53	0.77	1.70	1.63	–10.84	0.45	1.48	1.62
	Average	–13.03	0.43	1.73	1.65	–10.42	0.33	1.50	1.59
720p	Bigship	–11.43	0.34	1.82	1.53	–11.70	0.31	1.54	1.68
	City	–21.67	0.76	1.79	1.41	–13.89	0.49	1.51	1.38
	Night	–6.99	0.27	1.89	1.55	–6.12	0.22	1.54	1.63
	Shuttle Start	–12.01	0.34	2.05	1.27	–9.99	0.28	1.64	1.32
	Average	–13.03	0.43	1.89	1.44	–10.42	0.33	1.56	1.50
Total average		–10.84	0.42	1.73	1.57	–8.70	0.33	1.49	1.41

times for the hierarchical B-picture prediction structure, and decoding runtime increases an average of 40% to 60% because of a decoder side filtering. In comparison with the BALF, encoding runtime increases an average of 10% because of the added symmetric structures. However, since the data path of each symmetric structure is independent, the parallel implementation can be adopted to make the computational complexity level of the proposed method similar to the BALF. On the other hand, the decoder has almost the same computational complexity as the BALF. Figure 5 shows original and reconstructed images by using the E-BALF and H.264/AVC. In this figure, Bigship was coded at QP=32 by using IPPP prediction structure. It shows that the E-BALF makes a reconstructed image more similar to the corresponding original image. As for a filter selection ratio, when three newly added filters are applied, a large percentage of the central symmetric structure that is the only filter in the BALF is distributed over the proposed three filters. It is found that the percentage of each selected filter relies on characteristics of video sequences and quantization parameters. More information about the percentage of the selected filter and the experimental results for the comparison with BALF is found in [15].

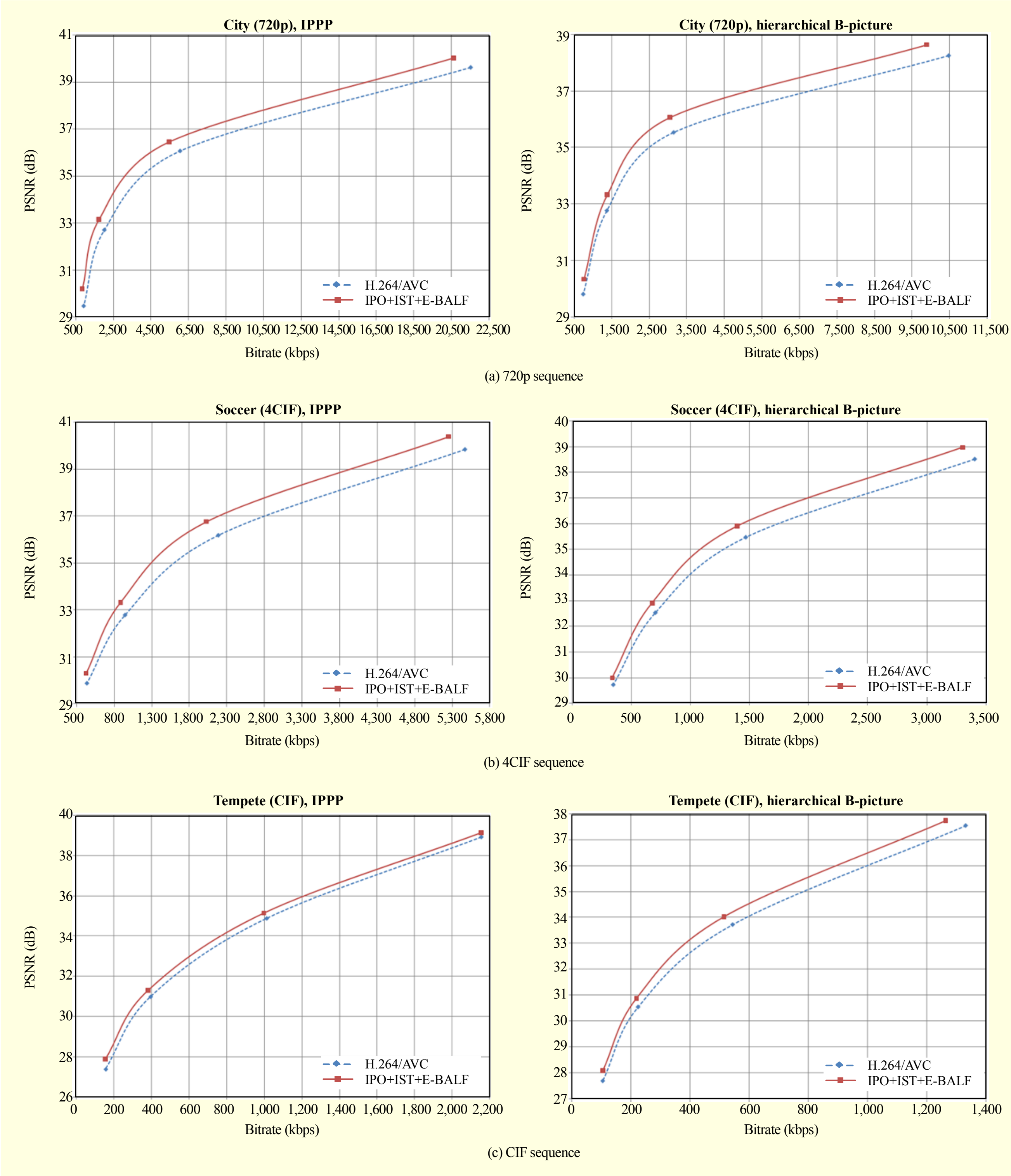
(a) Original (b) E-BALF (c) H.264/AVC

Fig. 5. Subjective quality comparison between E-BALF and H.264/AVC (QP=32, IPPP, 60th frame, cropped version).

Table 6. Performance comparison between the combination of the proposed tools vs. H.264/AVC High Profile for IPPP and hierarchical B-picture prediction structure.

Sequence		IPPP				Hierarchical B-picture			
		BD-rate (%)	BD-PSNR (dB)	Time ratio Enc.	Time ratio Dec.	BD-rate (%)	BD-PSNR (dB)	Time ratio Enc.	Time ratio Dec.
CIF	Bus	–10.20	0.48	4.07	1.56	–6.86	0.35	3.80	0.93
	City	–7.36	0.35	4.30	1.67	–12.48	0.55	3.71	1.02
	Mobile& Calendar	–7.46	0.37	3.99	1.64	–8.05	0.39	3.80	0.94
	Soccer	–8.56	0.39	4.17	1.24	–11.22	0.45	4.08	1.20
	Tempete	–5.88	0.31	4.09	1.66	–6.98	0.36	3.85	1.21
	Average	–7.89	0.38	4.13	1.55	–9.12	0.42	3.80	1.06
4CIF	City	–16.64	0.61	4.15	1.25	–13.63	0.50	3.65	1.48
	Crew	–11.00	0.36	3.98	1.19	–8.19	0.27	3.53	1.14
	Soccer	–18.04	0.79	4.05	1.21	–13.02	0.55	3.58	1.39
	Average	–15.23	0.59	4.06	1.22	–11.61	0.44	3.59	1.34
720p	Bigship	–11.55	0.34	4.24	1.19	–12.92	0.35	3.80	1.32
	City	–22.11	0.77	4.25	1.21	–15.43	0.55	3.71	1.38
	Night	–7.81	0.30	4.34	1.42	–7.71	0.28	3.80	1.61
	Shuttle Start	–11.93	0.34	4.65	1.21	–10.78	0.31	4.08	1.35
	Average	–13.35	0.44	4.37	1.26	–11.71	0.37	3.85	1.41
Total average		–11.54	0.45	4.18	1.34	–10.61	0.41	3.71	1.27

Table 6 shows the results of the combination of our tools, which is the overall performance of the proposed video codec. The BD-rates are –5.88% to –22.11% for the IPPP prediction structure and –6.86% to –15.43% for the hierarchical B-picture prediction structure. Figure 6 shows RD-curves of the combined tools. As shown in Table 6 and Fig. 6, the proposed codec significantly outperformed H.264/AVC High Profile. In particular, it has better performance as the bitrate increases. Therefore, we deduce that it will have a larger bit reduction for



(a) 720p sequence

(b) 4CIF sequence

(c) CIF sequence

Fig. 6. RD-curves for combination of proposed tools.

4k×2k video. The average encoding runtime ratio of the tool combination is 4.18 times for the IPPP prediction structure and 3.71 times for the hierarchical B-picture prediction structure relative to H.264/AVC. The additional computational efforts are mainly caused by the IPO and the E-BALF. However, as described above, the complexity efforts can be reduced if a fast intra offset value search is developed.

Various experimental results for other sequences under the condition of HVC CfP [2] are found in [16]. Coding efficiency performance is shown in [5], [16] for when the proposed tools are combined with mode-dependent directional transform and an extended macroblock.

IV. Conclusion

A novel video codec for video content with increased resolution and quality was presented. It has newly developed coding tools: the IPO, IST, and E-BALF. These tools are used adaptively in the processing of intra prediction, transform, and loop filtering. Moreover, by combining these tools with H.264/AVC, we accomplished a video codec that can provide a significantly high performance of coding efficiency. Experimental results showed that the proposed codec achieved high bitrate reduction by an average of 13.35% in BD-rate relative to H.264/AVC for 720p sequences under the condition of IPPP prediction. The experimental results also confirm that the proposed codec has higher coding efficiency as the bitrate, and spatial resolution of the sequences increases. We can thereby conclude that the proposed codec will be appropriate for an entertainment-quality video service with ultra high definition video (4k×2k and 8k×4k) as well as with high definition video.

References

[1] ITU-T and ISO/IEC JTC 1, Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC), 4th ed., Sept. 2008.

[2] ISO/IEC JTC 1 SC29 WG11, "Joint Call for Proposals on Video Compression Technology," Doc. N11113, Jan. 2010.

[3] ISO/IEC JTC 1 SC29 WG11, "Vision, Applications and Requirements of High-Performance Video Coding," Doc. N11096, Jan. 2010.

[4] S.C. Lim et al., "Rate-Distortion Optimized Adaptive Transform Coding," *Optical Eng.*, vol. 48, Aug. 2009, 087004.

[5] S.C. Lim et al., "Intra Prediction with Offset," ITU-T SG16/Q.6 Doc. VCEG-AL29, July 2009.

[6] C.F. Chen and K.K. Pang, "The Optimal Transform of Motion-Compensated Frame Difference Images in a Hybrid Coder," *IEEE Trans. Circuits Syst. II: Analog Digital Signal Process.*, vol. 40, no. 6, June 1993, pp. 393-397.

[7] T. Chujoh et al., "Block-Based Adaptive Loop Filter," ITU-T SG16/Q.6, Doc. VCEG-AI18, July 2008.

[8] T. Chujoh et al., "Improvement of Block-Based Adaptive Loop Filter," ITU-T SG16/Q.6, Doc. VCEG-AJ13, Oct. 2008.

[9] Y.J. Chiu and L. Xu, "Adaptive (Wiener) Filter for Video Compression," ITU-T SG16 Contribution, C437, Geneva, Apr. 2008.

[10] H.264/AVC Reference Software Joint Model (JM) version 1x.0. http://iphome.hhi.de/suehring/tml/

[11] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B-Pictures" Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, JVT-P014, July 2005.

[12] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," *Proc. ICME*, Toronto, Canada, July 2006.

[13] G. Bjontgaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T SG16 Q.6 VCEG, Doc. VCEG-M33, 2001.

[14] T.K. Tan, G. Sullivan, and T. Wedi, "Recommended Simulation Common Conditions for Coding Efficiency Experiments Revision 4," VCEG-AJ10r1, July 2008.

[15] 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.

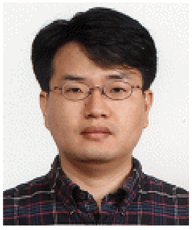
[16] H. Kim et al., "Description of Video Coding Technology Proposal by ETRI," Doc. JCTVC-A127, Apr. 2010.

Seyoon Jeong received the BS and MS in electronics engineering from Inha University, Korea, in 1995 and 1997, respectively. Since 1996, he has been a senior member of research staff with ETRI, Korea, and he is also working toward the PhD in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea. His current research interests include video coding, video transmission, and UHDTV.

Sung-Chang Lim received the BS (with highest honors) and MS in computer engineering from Sejong University, Korea, in 2006 and 2008, respectively. Since 2008, he has been a member of engineering staff in Broadcasting and Telecommunications Media Research Department of ETRI, Daejeon, Korea. His research interests include video coding, mobile video transmission, and image processing.

Hahyun Lee received the BS in electronics engineering from Korea Aerospace University, Korea, in 2002, and the MS in mobile communication and digital broadcasting engineering from the University of Science and Technology (UST), Daejeon, Korea, in 2007. Since 2008, he has been a member of engineering staff in Broadcasting and Telecommunications Media Research Department of ETRI, Daejeon, Korea. His research interests include video coding, image processing, and video transmission.

Jongho Kim received BS from Control and Computer Engineering Department, Korea Maritime University in 2005 and his MS from the University of Science and Technology (UST) in 2007. In September 2008, he joined Broadcasting and Telecommunications Media Research Department, ETRI, Korea, where he is currently a researcher. His research interests include video processing and video coding.



Jin Soo Choi received the BE, ME, and PhD in electronic engineering from Kyungpook National University, Korea, in 1990, 1992, and 1996, respectively. Since 1996, he has been a principal member of engineering staff in ETRI, Korea. He has been involved in developing the MPEG-4 codec system, data broadcasting system, and UDTV. His research interests include visual signal processing and interactive services in the field of the digital broadcasting technology.

Haechul Choi received the BS in electronics engineering from Kyungpook National University, Daegu, Korea, in 1997, and the MS and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1999 and 2004, respectively. He is an assistant professor of Division of Information Communication and Computer Engineering in Hanbat National University, Daejeon, Korea. From 2004 to 2010, he was a senior member of research staff in the Broadcasting Media Research Group of ETRI. His current research interest includes image processing, video coding, and video transmission.