New Techniques for Next Generation Video Coding

Li Liu, Robert Cohen, Huifang Sun, Anthony Vetro, Xinhua Zhuang

TR2010-058 April 2010

Abstract

Video coding standards are rapidly evolving with the advance of the video compression techniques. As the state of the art video coding standard, H.264/AVC exhibits superior coding performance improvement over its predecessors. Currently, both VCEG and MPEG are launching their next-generation video coding project, which aims to meet the new requirements future applications may impose on the video coding standard. In this paper, we summarize the progress of those next generation video coding standard projects and existing new video coding techniques. Moreover, we also present details on our implementation of second-order prediction and reduced resolution update, followed by some experimental results.

IEEE International Symposium on Broadband Multimedia Systems

New Techniques for Next Generation Video Coding

Li Liu*[a], Robert Cohen[b], Huifang Sun[b], Anthony Vetro[b], Xinhua Zhuang[a]
[a]Dept. of Computer Science, University of Missouri, Columbia MO, 65211
[b]Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139

***Abstract*—Video coding standards are rapidly evolving with the advance of the video compression techniques. As the state of the art video coding standard, H.264/AVC exhibits superior coding performance improvement over its predecessors. Currently, both VCEG and MPEG are launching their next-generation video coding project, which aims to meet the new requirements future applications may impose on the video coding standard. In this paper, we summarize the progress of those next generation video coding standard projects and existing new video coding techniques. Moreover, we also present details on our implementation of second-order prediction and reduced resolution update, followed by some experimental results.**

***Index Terms*— Video Coding, H.264/AVC, HVC, NGVC, H.265, KTA**

1. INTRODUCTION

Video coding techniques provide efficient solutions to represent video data in a more compact and robust way so that the storage and transmission of video can be realized in less cost in terms of size, bandwidth and power consumption. To meet the industry requirement of standardizing existing video techniques, video coding standards were developed by two international organizations, ITU-T and ISO/IEC. The family of ISO/IEC MPEG standards includes MPEG-1, MPEG-2 MPEG-4, and MPEG-4 Part 10 (AVC). ITU-T H.26x series standards consist of H.261, H.263, and H.264. The evolution of video coding standards reflects the technological progress toward improving the coding efficiency of video compression technologies. For example, the state of art video coding standard H.264/AVC [1], jointly developed by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), is reported to achieve gains in compression efficiency up to 50% to its predecessor MPEG-2 [1].

However, the increasing popularity of high definition TV, video delivery on mobile devices, and other multimedia applications creates new demands for video coding standards. To face new challenges that new applications may impose, both MPEG and VCEG launched their next-generation video coding project, which potentially could be either an extension of H.264/AVC or a brand new standard.

In the remainder of this paper, current standard activities are summarized in Section 2. The new coding techniques proposed to VCEG and MPEG and their performances are analyzed in Section 3. Section 4 presents our implementation of two techniques: second-order prediction and reduced resolution update, along with related experimental results. The last part draws some conclusions.

2. CURRENT STANDARD ACTIVITIES

At the 86th MPEG meeting in Busan 2008, MPEG determined the need for a next-generation video compression technology called High-performance Video Coding (HVC). HVC would be intended mainly for high quality applications, by providing performance in terms of coding efficiency at higher resolutions, with applicability for entertainment-quality services such as HD mobile, home cinema and Ultra High Definition (UHD) TV [2]. A Call for Evidence was issued that allowed proponents to report about the existence of such technologies. The response to this Call for Evidence was evaluated at the 89th MPEG meeting in July 2009. It was found that there was sufficient evidence that compression technology had advanced enough to commence work on a new standard. In January 2010, a Call for Proposals (CfP) on video compression technology was issued. Responses to this CfP are currently being evaluated.

Comparing with HVC, the goals of ITU-T's Next Generation Video Coding (NGVC) project are similar but more detailed. In 2005 to 2008, ITU-T VCEG studied the requirement definition for NGVC, and some agreements about the goals of the NGVC project were reached during the 37th VCEG meeting [3], with primary emphasis on computational efficiency and high compression performance. For instance, in terms of coding efficiency, NGVC should be capable of providing 50% bit rate savings over H.264/AVC at the same video quality representation. To address the concern of complexity, NGVC should be capable of operating with a complexity ranging from 50% to 3 times H.264/AVC High Profile. More specifically, when operated at a complexity of 50% compared to H.264/AVC High Profile, NGVC should provide a 25% bit rate savings compared to H.264/MPEG-4 AVC High Profile at equivalent subjective quality [3].

Though MPEG and VCEG could independently create separate next generation video coding standards, two new standards of similar functionalities might not be welcomed by industry. Based on the previous success in jointly creating H.264/AVC, future collaboration on NGVC and HVC, similar to the Joint Video Team effort, was considered during a joint meeting of MPEG and VCEG in July 2009. In January 2010, MPEG and VCEG agreed to work together on the joint CfP, establishing a Joint Collaborative Team (JCT) on video coding.

3. EXISTING NEW TECHNIQUES

The advances of video coding techniques were contributed by various parties. To provide a software platform to gather and evaluate these new techniques, a Key Technical Area (KTA) [4] platform was developed based on JM11 reference software, where the new coding tools are continuously added. So far, the major new coding tools added to KTA platform can be summarized as follows:

1. Intra Prediction: In [5], H.264 intra prediction is enhanced with additional Bi-directional Intra Prediction (BIP) modes, where BIP combines prediction blocks from two prediction modes using a weighting matrix. Furthermore, Mode-Dependent Directional Transform (MDDT) using transforms derived from KLT is applied to capture the remaining energy in the residual block.

2. Inter Prediction: To further improve inter prediction efficiency, finer fractional motion prediction and better motion vector prediction were proposed. Increasing the resolution of the displacement vector from 1/4-pel to 1/8-pel to obtain higher efficiency of the motion compensated prediction is suggested in [6]. In [7], a competing framework for better motion vector coding and SKIP mode is proposed, where both spatial and temporal redundancies in motion vector fields are captured. Moreover, [8] suggests extending the macroblock size up to 64x64 so that new partition sizes 64x64, 64x32, 32x64, 32x32, 32x16, and 16x32 can be used. Instead of using the fixed interpolation filter from H.264/AVC, Adaptive Interpolation Filters (AIF) are proposed, such as 2D AIF [9], Separable AIF [10], Directional AIF [11], Enhanced AIF [12], and Enhanced Directional AIF [13].

3. Quantization: To achieve better quantization, optimized quantization decision at the macroblock level and at different coefficient positions are proposed. Rate Distortion Optimized Quantization (RDOQ), which performs optimal quantization on a macroblock, was added to the JM reference software. RDOQ does not require a change of H.264/AVC decoder syntax. More recently, [14] gives an improved, more efficient RDOQ implementation. In [15], Adaptive Quantization Matrix Selection (AQMS), a method deciding the best quantization matrix index, where different coefficient positions can have different quantization steps, is proposed to optimize the quantization matrix at a macroblock level.

4. Transform: For motion partitions bigger than 16x16, a 16x16 transform is suggested in addition to 4x4 and 8x8 transforms [8]. Moreover, transform coding is not always a must. In [16], it is proposed that for each block of the prediction error, either standardized transform coding or spatial domain coding can be adaptively chosen.

5. In-loop Filter: In KTA, besides the deblocking filter, an additional Adaptive Loop Filter (ALF) is added to improve coding efficiency by applying filters to the deblocking-filtered picture. Two different ALF techniques are adopted so far: Quadtree-based Adaptive Loop Filter (QALF) [17] and Block-based Adaptive Loop Filter (BALF) [18].

6. Internal bit-depth increase: By using 12 bits of internal bit depth for 8-bit sources, so that the internal bit-depth is greater than the external bit-depth of the video codec, the coding efficiency can be further improved [19].

Besides the techniques listed above, there are some noticeable contributions not added to KTA yet. For example, [20-22] proposed three methods, respectively, to use Decoder Side Motion Estimation (DSME) for B-picture motion vector decision, which improves coding efficiency by saving bits on B-picture motion vector coding. Also, some new techniques are under investigation and will be presented in the responses for call for proposals.

4. PROPOSED TECHNIQUES AND PERFORMANCE ANALYSIS

In this section, we present our detailed implementation of two techniques that can be applied to further improve H.264/AVC coding performance. The first one is called second-order prediction, which applies intra prediction on inter prediction residues, and the second one is called reduced resolution update, which reduces the bit rate by coding the residue at downsampled spatial resolution. Though similar ideas can be found in the literature, we implemented both techniques on the latest H.264/AVC reference JM15.1 so that their performance comparisons can be analyzed based on the state of the art video coding platform.

4.1 Second-Order Prediction for Inter Coding

Inter prediction explores temporal redundancy between frames to save coding bits. By using motion compensated prediction, the best matching position of current block is found within the reference picture so that only prediction difference needs to be coded.

H.264/AVC achieves a remarkable improvement on inter prediction coding performance over previous video coding standards. The reference picture is to be decided from a reference list containing multiple candidate reference pictures, and the reference list can be managed by marking or reordering reference pictures. As more sub-block sizes are supported, an inter predicted block is coded using a hierarchical partition structure. Each partition block can have its own reference picture and motion vector. The best partition structure is usually decided via rate-distortion optimization.

Though H.264/AVC inter prediction is considered to be very efficient, correlation can still be found within the residue. By applying additional round of prediction in the domain of inter prediction residues, second-order prediction may lead to improved prediction residues and thus fewer encoded bits. With second-order prediction, the residue reconstruction can be rewritten as:

Reconstructed pixel value
= Prediction from motion compensation (First-order)
+ Prediction of first order residue (Second-order)
+ Residue after second order prediction (To be coded)

An early form of second-order prediction can be found in H.264/AVC weighted prediction technique. In brightness variation condition, in addition to inter prediction, weighted prediction is supported by H.264/AVC, where a weighting factor a and an additive weighting offset b are used to compensate the lighting difference between the current picture and the reference picture. Here, the weighted prediction can be seen a form of DC prediction to reduce intra-picture redundancy when inter prediction alone is not capable of handling the brightness condition variation between frames very well.

As lighting conditions may vary not only between frames but also within a frame, to handle local lighting variation, [23] proposes a localized weighted prediction, where offset b is estimated based on the reconstructed neighboring samples of current block and its associated motion compensated samples in the reference picture. With estimated localized offset b, current block is compensated for lighting condition change by applying a DC offset to inter prediction residue. Comparing to H.264/AVC weighted prediction, the advantage of block-based localization is obvious: small additional cost, but more adaptive to uniform local changes. Similar ideas of block based second-order prediction can be found in [24] and [25] as well.

To utilize second-order prediction to improve coding performance not just in a lighting variation condition, several issues were addressed in our implementation: the second-order predictor, the decision of second-order prediction mode, and the choice of the second-order prediction block size. Here, we first review the second-order prediction procedure in addition to H.264/AVC inter coding, and then we describe the mode decision process.

1. For each macroblock partition, its best reference pictures and motion vectors are decided through a motion search step.

2. As seen in Fig. 1, for neighboring reconstructed pixels $a_{x,y}$ of a block in a partition, their associated pixels $b_{x,y}$ in reference picture are located through motion compensated prediction using the vector from step. 1.

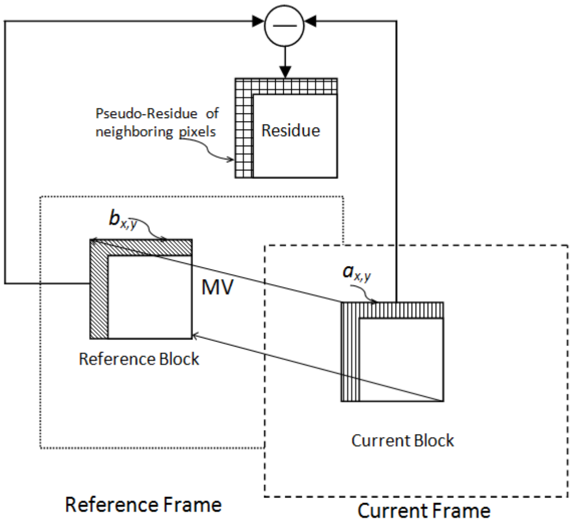

Fig. 1. Second-order prediction procedure

3. By taking the difference of $a_{x,y}$ and $b_{x,y}$, we get a reconstructed neighboring pixel difference between the current block and its associated block. Since this difference is not the real residue, we call it pseudo-residue.

4. Assume the pseudo-residue has similarities with the first-order inter prediction residue. This pseudo-residue then can be used for second-order intra prediction of the actual residue. Similar to 4x4 intra prediction in H.264/AVC, different second-order prediction modes are supported by adjusting the prediction direction, as shown in Fig. 2.

5. For each partition, its best second-order prediction modes associated with every blocks are decided using the rate distortion framework.

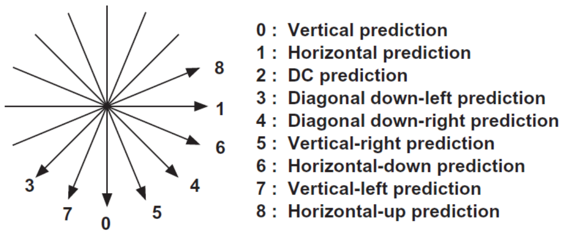

Fig. 2. Second-order prediction modes

Though it is possible to assign each block an individual second-order prediction mode, we find it is not cost efficient for blocks whose size is smaller than 8x8, as too much side information needs to be coded. Therefore, the second-order prediction decision is made on 8x8 blocks. For a partition size larger than 8x8, it will be divided into multiple 8x8 sub-blocks, each with its own second-order prediction mode.

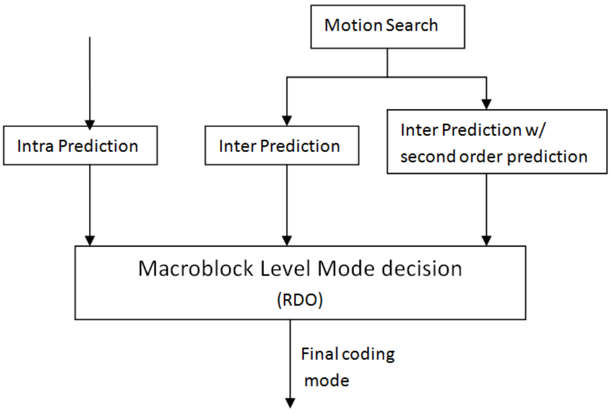

Fig. 3. Modified mode decision

After second-order prediction modes are added, they need to compete with existing H.264/AVC modes for best coding performance purposes. The final macroblock level coding mode is decided among intra coding, original inter coding, and inter coding with second-order prediction using the RDO rule as shown in Fig.3.

4.2 Reduced Resolution Update for Intra Coding

Reduced resolution update (RRU) is a technique that aims to save coding bits by resize image/prediction residues to a reduced spatial resolution.

It's known that down-sampling an image to a low resolution, then compressing the lower resolution, and subsequently interpolating the result to the original resolution can improve the overall PSNR performance of the compression process. In [26], authors present an analytical model and a numerical analysis of the down-sampling, JPEG compression and up-sampling process that makes explicit the possible quality/compression trade-offs.

Motivated by the above facts, we modified the framework of H.264/AVC so that residue after intra prediction can be optionally downsampled before the transform and quantization steps (Fig. 4). For instance, a 16x16 block can be downsampled by a factor of 2 so that only an 8x8 block needs to be coded (Fig. 5). If the corresponding block was downsampled at the encoder side, the decoder shall upsample the downsampled residues to reconstruct full resolution picture.

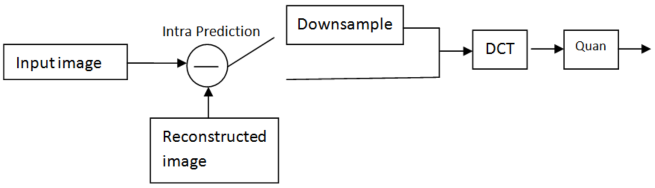

Fig. 4. Intra coding with optional RRU

Though RRU is capable of saving bits as fewer residue pixels need to be coded, it introduces more distortion at the same time. Therefore, the choice of RRU should be considered under a rate-distortion optimization framework.

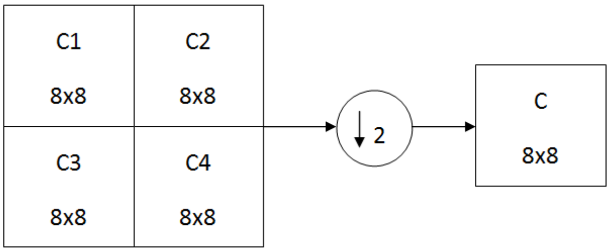

Fig. 5. Downsampling a 16x16 block

4.3 Experimental Results

We implemented techniques listed in Sections 4.1 and 4.2 using H.264/AVC reference software JM15.1 to verify their effectiveness in terms of improving coding performance based on common test conditions listed in [27].

In the first experiment, we compare second-order prediction with the original H.264/AVC inter coding. For each sequence, the first frame is coded as an I picture and the remaining frames are coded as P pictures. Only the 4x4 DCT is allowed. Four QP values are tested: 23, 28, 33, and 38. CABAC is used for entropy coding.

Table I Performance comparison (Experiment 1)

Sequence	Size	Bitrate Savings %	PSNR Gain (dB)
Keiba	832x480	0.97	0.041
Mobisode2	832x480	0.93	0.03
Tennis	1920x1080	1.18	0.04
kimonol	1920x1080	0.65	0.02

Though extensive experiments are performed, only partial results are presented in Table I. The bit saving ratio and PSNR gain are computed according to [28].

From the results, it can be seen that though consistent coding performance improvements are achieved for all test coding sequences, the gains are not significant. This can be explained as follows. For the current implementation, at encoder side, second-order prediction is performed sequentially after motion search for the first-order inter prediction mode. This indicates that the motion vector is not optimized for second-order prediction whose optimal motion vector position may be different from the first-order motion vector. Therefore, to further improve the efficiency of the second-order prediction, it is desirable to perform motion vector search for each individual second-order prediction mode. The advantage is that no syntax changes made at the decoder side, but the new implementation can impose increased computational complexity at the encoder side.

In the second experiment, we compare reduced resolution update method with the original H.264/AVC intra coding, for 16x16 blocks. The downsamping and upsampling filter used are as follows:

Downsampling: 5-tap filter [-1 2 6 2 -1]/8

Upsampling: 7-tap filter [-1 0 9 16 9 0 -1]/16

To compare only intra coding efficiency, all frames are coded as I frames. Using the same four QP values from the first experiment, each sequence is coded using RRU, and the performance comparisons are given in Table II.

Table II Performance comparisons (Experiment 2)

Sequence	Size	Bitrate Savings %	PSNR Gain (dB)
Foreman	352x288	4.82	0.27
Akiyo	352x288	5.45	0.36
Keiba	416x320	2.08	0.12
Mobisode2	832x480	4.21	0.08

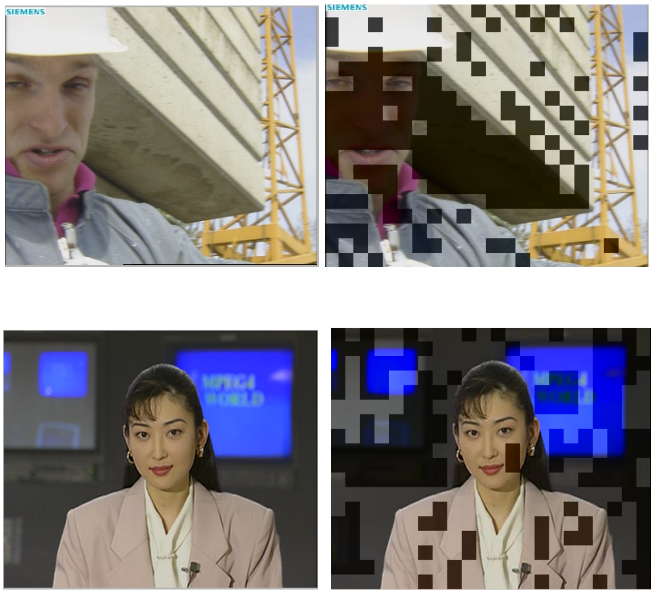

Fig. 6 The original image (left) and regions adopting RRU are marked with gray color (right)

In Fig. 6, we mark regions where blocks choose RRU over the original intra coding mode with the gray color. As one can see, RRU improves the coding efficiency for those areas of medium content complexity, as H.264/AVC intra prediction is very efficient for flat areas and RRU may bring too much loss for areas with strong high-frequency content.

In the second experiment, though RRU works well for 16x16 blocks, the contribution of this method to overall intra coding performance shall depend on the percentage 16x16 block size is used over 4x4 and 8x8 modes. Therefore, further research shall be carried to study the performance of RRU on smaller block sizes.

5. CONCLUSIONS

Both the new techniques listed in Section 3 and our experiments on second-order prediction and RRU prove that there is still room for performance improvement of current coding standard. The Call for Evidence for HVC provided results that averaged a 15-25% gain in coding efficiency. While not enough to constitute a new standard, the evidence was sufficient to warrant issuing a Call for Proposals. Once the proposals have been evaluated, it is expected that over the next several months, these new technologies will advance to the point where the ITU and MPEG can issue a new joint video coding standard.

REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, 2003, pp. 560-576.

[2] “MPEG Considers the Development of HVC – the Next Generation of Video,” *ISO/IEC JTC 1/SC 29/WG 11 N10117*, Busan, Korea, 2008.

[3] G.J. Sullivan, “Informal report of VCEG actions at SG 16 meeting Jan/Feb 2009,” *ITU-T Q.6/SG16 VCEG, VCEG-AK04*, Yokohama, Japan, 2009.

[4] http://iphome.hhi.de/suehring/tml/download/KTA/ .

[5] Y. ye and M. Karczewicz, “Improved Intra Coding,” *ITU-T Q.6/SG16 VCEG, VCEG-AG11*, Shenzhen, China, 2007.

[6] J. Ostermann and M. Narroschke, “Motion compensated prediction with 1/8-pel displacement vector resolution,” *ITU-T Q.6/SG16 VCEG, VCEG-AD09*, Hangzhou, China, 2006.

[7] G. Laroche, J. Jung, and B. Pesquet-Popescu, “RD Optimized Coding for Motion Vector Predictor Selection,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, 2008, pp. 1681-1691.

[8] P. Chen, Y. Ye, and M. Karczewicz, “Video Coding Using Extended Block Sizes,” *ITU-T SG16/Q6, doc. C-123*, 2009.

[9] Y. Vatis and J. Ostermann, “Prediction of P- and B-Frames Using a Two-dimensional Non-separable Adaptive Wiener Interpolation Filter for H.264/AVC,” *ITU-T Q.6/SG16 VCEG, VCEG-AD08*, Hangzhou, China, 2006.

[10] S. Wittmann and T. Wedi, “Simulation results with separable adaptive interpolation filter,” *ITU-T Q.6/SG16 VCEG, VCEG-AG10*, Shenzhen, China, 2007.

[11] D. Rusanovskyy, K. Ugur, and J. Lainema, “Adaptive Interpolation with Directional Filters,” *ITU-T Q.6/SG16 VCEG,AG21*, Shenzhen, China, 2007.

[12] Y. Ye and M. Karczewicz, “Enhanced Adaptive Interpolation Filter,” *ITU-T SG16/Q.6 Doc. T05-SG16-C-0464*, Geneva, Switzerland, 2008.

[13] T. Arild Fuldseth, D. Rusanovskyy, K. Ugur, and J. Lainema, “Low Complexity Directional Interpolation Filter,” *ITU-T Q.6/SG16 VCEG, VCEG-AI12*, Berlin, Germany, 2008.

[14] M. Karczewicz, Y. Ye, and I. Chong, “Rate Distortion Optimized Quantization,” *ITU-T Q.6/SG16 VCEG, VCEG-AH21*, Antalya, Turkey, 2008.

[15] A. Tanizawa and T. Chujoh, “Adaptive Quantization Matrix Selection,” *ITU-T Q.6/SG16 VCEG, D.266*, 2006.

[16] M. Narroschke and H.G. Musmann, “Adaptive prediction error coding in spatial and frequency domain for H.264/AVC,” *ITU-T Q.6/SG16 VCEG, VCEG-AB06*, Bangkok, Thailand, 2006.

[17] T. Chujoh, N. Wada, and G. Yasuda, “Quadtree-based adaptive loop filter,” *ITU-T Q.6/SG16 VCEG,C.181*, 2009.

[18] G. Yasuda, N. Wada, T. Watanabe, and T. Yamakage, “Block-based Adaptive Loop Filter,” *ITU-T Q.6/SG16 VCEG, VCEG-AI18*, Berlin, Germany, 2008.

[19] T. Chujoh and R. Noda, “Internal bit depth increase for coding efficiency,” *ITU-T Q.6/SG16 VCEG, VCEG-AE13*, Marrakech, Morocco, 2007.

[20] S. Klomp and J. Ostermann, “Response to Call for Evidence in HVC: Decoder-side Motion Estimation for Improved Prediction,” *ISO/IEC JTC 1/SC 29/WG 11 M16570*, London, UK, 2009.

[21] T. Murakami, “Advanced B Skip Mode with Decoder-side Motion Estimation,” *ITU-T Q.6/SG16 VCEG, VCEG-AK12*, Yokohama, Japan, 2009.

[22] S. Kamp, M. Evertz, and M. Wien, “Decoder Side Motion Vector Derivation,” *ITU-T Q.6/SG16 VCEG, VCEG-AG16*, Shenzhen, China, 2007.

[23] Peng Yin, A. Tourapis, and J. Boyce, “Localized Weighted Prediction for Video Coding,” *2005 IEEE International Symposium on Circuits and Systems*, Kobe, Japan, , pp. 4365-4368.

[24] S. Chen, J. Wang, S. Li, and L. Yu, “Second Order Prediction (SOP) in P Slice,” *ITU-T Q.6/SG16 VCEG, VCEG-AI27*, Berlin, Germany, 2008.

[25] M. Yang, D.J. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, 2002, pp. 34-58.

[26] A.M. Bruckstein, M. Elad, and R. Kimmel, “Down-Scaling for Better Transform Compression,” *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 12, 2001, pp. 1132--1144.

[27] T.K. Tan, G. Sullivan, and T. Wedi, “Recommended Simulation Common Conditions for Coding Efficiency Experiments Revision,” *ITU-T Q.6/SG16 VCEG, VVCEG-AH10*, 2008.

[28] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves,” *ITU-T Q.6/SG16 VCEG, VCEG-M33*, Austin, Texas, 2001.