INTRA-FRAME DYADIC SPATIAL SCALABLE CODING BASED ON A SUBBAND/WAVELET FRAMEWORK FOR MPEG-4 AVC/H.264 SCALABLE VIDEO CODING

Shih-Ta Hsiang
ARC-Multimedia Research Lab, Motorola Labs.
Schaumburg, IL 60173, USA
hsiang@motorola.com

ABSTRACT

This paper develops a new intra-frame dyadic spatial scalable coding framework based on a subband/wavelet coding approach for MPEG-4 AVC/H.264 scalable video coding (SVC). It is the first attempt in the literature to join the subband filter banks with the traditional macroblock and DCT based video coding system. We demonstrate that the current H.264 coding tools and syntax set can efficiently work together with the traditional subband filter banks for providing the improved efficiency for intra-frame dyadic spatial scalable coding. More importantly, unlike the classical wavelet coding, the proposed framework still allows the down-sampling filter to be flexibly designed to generate the ideal low resolution video for target applications. The proposed wavelet tool has been adopted by the JSVM for the next-phase SVC standard.

Index Terms— AVC, H.264, scalable coding, wavelet

1. INTRODUCTION

In recent years, subband/wavelet coding has been demonstrated to be one of the most efficient methods for image coding in the literature [1][2][3]. It has also been utilized in the international standard JPEG 2000 [4] for image and video (in the format of Motion JPEG 2000) coding applications in industry. Thanks to high energy compaction of subband/wavelet transform, these state-of-the-art coders are capable of achieving excellent compression performance without traditional blocky artifacts associated with the block transform. More importantly, they can easily accommodate the desirable spatial scalable coding functionality with almost no penalty in compression efficiency because the subband/wavelet transform is resolution scalable by nature.

On the other hand, the former video coding standards such as MPEG-2/4 and H.263+ and the emerging MPEG-4 AVC/H.264 scalable video coding (SVC) amendment [5] adopts a pyramidal approach to spatial scalable coding. This method utilizes the interpolated frame from the recovered base-layer video to predict the related high-resolution frame at the enhancement layer and the resulting residual signal is coded by the enhancement-layer bitstream. Unlike wavelet/subband coding with the low resolution signal determined by the lowpass filter of the selected analysis filter banks, the pyramidal coding scheme allows great flexibility for image down-sampler design. However, the number of source pixel samples is increased by 33.3% for building a complete image pyramidal representation in the resulting coding system, which can inherently reduce compression efficiency. The simulation results from the JVT core experiment also show that the current MPEG-4 AVC/H.264 joint scalable video model (JSVM) suffers from substantial efficiency loss for intra dyadic spatial scalable coding, particularly toward the high bitrate range [6].

This paper develops a new intra-frame dyadic spatial scalable coding framework based on a subband/wavelet coding approach. In the proposed framework the employed down-sampling filter for generating low resolution video at the base layer is not particularly tied to the specific subband/wavelet filter selection for signal representation, in a clear contrast to the traditional wavelet coding system. In addition, our research efforts have been further aimed at efficiently exploiting the subband/wavelet techniques within the traditional macroblock and DCT based video coding system for improved efficiency of intra-frame spatial scalable coding. Unlike the former MPEG-4 visual texture coding (VTC) [7] which is practically built upon a separate zero-tree based system for coding wavelet coefficients, the proposed subband coding framework has been integrated with the H.264 JSVM reference software with little modifications to the current standard. As such, the modified H.264 coding system can take advantage of the benefits of wavelet coding without much increase in implementation complexity.

The remaining paper is organized as follows. Section 2 presents the proposed intra-frame scalable coding framework. Section 3 describes how the proposed system can be implemented by efficient re-use of the existing MPEG-4 AVC/H.264 coding tools. The simulation results are provided in Section 4, followed by the summary and conclusion of this paper.

2. PROPOSED FRAMEWORK

The decoding system for the proposed subband/wavelet intra-frame coding framework is illustrated in Figure 1 for spatial scalable coding in two layers. At the lowest bitstream layer, the two-dimensional (2-D) dyadic down-sampled input frame is just represented by a traditional single-layer bitstream for scalable decoding at reduced spatial resolution. On the other hand, the individual video frame is represented in the subband/wavelet domain at the enhancement layer, as indicated in Figure 1.

The inter-layer prediction in the proposed system utilizes the recovered lower-layer frame $x'^{(n-1)}$ for prediction of the low-pass subband $x^{(n)}_{\mathrm{LL}}$ at the current layer n, as shown in Figure 1. It is contrasted with the traditional pyramidal coding approach that employs the interpolated low-resolution frame for predicting the entire frame at the next higher resolution layer in the spatial domain. The resulting enhancement bitstream at layer n thus consists of the coded information for the three high-pass subbands $x^{(n)}_{\mathrm{HL}}$, $x^{(n)}_{\mathrm{LH}}$, $x^{(n)}_{\mathrm{HH}}$ and the residual signal, $e^{(n)}_{LL}$, from predicting the lowpass subband $x^{(n)}_{\mathrm{LL}}$. The reconstructed wavelet coefficients are processed by the synthesis filter banks to generate the final recovered frame at full resolution.

The block diagram for the corresponding encoding system is provided in Figure 2 for spatial scalable encoding in two layers. Only the dotted region in the figure is considered as the mandatory system part for creating a compatible bitstream to the proposed decoding system in Figure 1. It should be noted that the mechanism for creating the input video at the reduced dyadic resolution ratio is not mandatory. In our example system in Figure 2, the low resolution input frame $x^{(n-1)}$ is generated by the cascaded spatial low-pass filtering and 2x2 down-sampling operation on the related input frame, $x^{(n)}$, from the next higher resolution layer. The input frame at the enhancement layer is decomposed by the subband analysis filter banks before it is encoded by the texture coder. The inter-layer prediction at the enhancement layer *n* is performed on the low-pass subband $x^{(n)}_{\mathrm{LL}}$ and the prediction residual signal $e^{(n)}_{\mathrm{LL}}$ can be represented by

$$\begin{aligned} e^{(n)}_{LL} &= x^{(n)}_{\mathrm{LL}} - 2 \cdot x'^{(n\text{-}1)} \qquad (1) \\ &= x^{(n)}_{\mathrm{LL}} - 2 \cdot x^{(n\text{-}1)} - 2 \cdot q^{(n\text{-}1)} \\ &= d^{(n)}_{\mathrm{LL}} - 2 \cdot q^{(n\text{-}1)} \end{aligned}$$

where $q^{(n\text{-}1)}$ is the quantization error from the lower layer *(n-1)* and $d^{(n)}_{\mathrm{LL}}$ is the difference between the low-pass subband signal $x^{(n)}_{\mathrm{LL}}$ and the scaled base-layer input frame $2 \cdot x^{(n\text{-}1)}$.

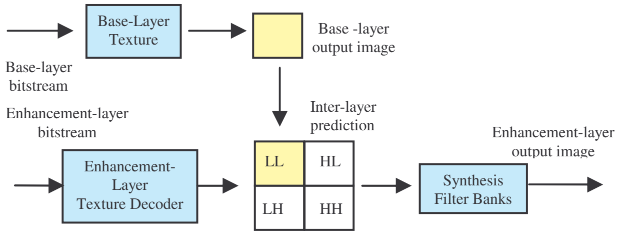

Figure 1: Illustration of the proposed intra-frame dyadic spatial scalable decoding system in the two layers.

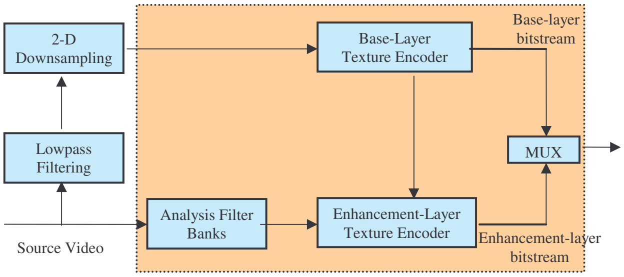

Figure 2: Block diagram of the proposed dyadic spatial scalable intra-frame encoding system in the two layers.

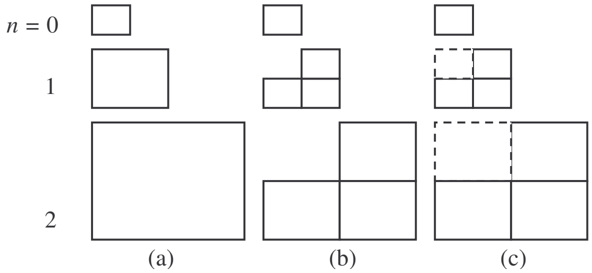

Figure 3: The coded signal representations by (a) pyramidal coding, (b) wavelet coding, and (c) the proposed approach.

When the normalized subband low-pass analysis filter is adopted as the lowpass filter for image down-sampling at the base layer *(n-1),* the lowpass residual signal $e^{(n)}_{\mathrm{LL}}$ given by Eq. (1) is reduced to $2 \cdot q^{(n\text{-}1)}$. We can then simply skip the texture coding of the residual signal over the lowpass subband region LL in Figure 1 if the average scaled distortion $2 \cdot q^{(n\text{-}1)}$ from the base layer is near or below the optimal distortion level for the assigned bitrate or QP parameters at the current enhancement layer *n*. The critical sampling feature of subband/wavelet coding is thus retained for achieving best compression efficiency and reduced complexity overhead. Nevertheless, unlike the classical subband/wavelet image coding system, the proposed intra-frame scalable coding framework still possesses the freedom for designing the optimal down sampling filter at the encoder to generate the desirable source video of the reduced resolution for target applications. The resulting difference, $d^{(n)}_{\mathrm{LL}}$, between the original low-pass subband signal $x^{(n)}_{\mathrm{LL}}$ and the scaled base-layer frame $2{\cdot}x^{(n\text{-}1)}$ can be compensated by the coded lowpass subband residual signal $e'^{(n)}_{\mathrm{LL}}$.

Figure 3 compares the coded signal representations employed by pyramidal coding, subband/wavelet coding, and the proposed scalable coding approach, respectively. The residual coding of the lowpass subbands, as indicated by the dashed regions in the figure, is only optional in the proposed system. It can be utilized to compensate for the filter difference, $\mathrm{d}^{(n)}_{\mathrm{LL}}$, and/or to further reduce the quantization error $2 \cdot q^{(n\text{-}1)}$ fed back from the base layer.

3. MPEG-4 AVC/H.264 SBC

This subband coding framework for intra-frame spatial scalable coding can be efficiently implemented within the current MPEG-4 AVC/H.264 system by intelligent re-use of the existing standard coding tools. In our implementation of H.264 subband coding (SBC), the input base-layer video is just encoded as a H.264 compatible bitstream, in accordance with the current SVC draft.

At the enhancement layer, the existing SVC intra-slice macroblock coding tool is still employed for encoding a subband decomposed frame on a macroblock-by-macroblock basis. Specifically, we still employ the new SVC intra macroblock mode I_BL for encoding the residual signal from the inter-layer prediction. The inter-layer prediction signal is just set to 0 over all the highest frequency subband regions HL, LH, and HH.

The DCT transform has been utilized to encode the residual signal resulted from spatial prediction of the pixel sample values by the current H.264 intra macroblock coding tool. Exploiting the same texture coding tool for encoding the high-pass subband coefficients in H.264 SBC is motivated by the fact that the high-pass subband actually just corresponds to the error signal from the prediction stage of the lifting operations when the wavelet transform is implemented by the lifting approach.

A syntax element has been added to the slice header in H.264 SBC to indicate the number of the subband decomposition levels being employed for representing the current picture. The SVC pyramidal approach can still be applied when the number of decomposition level is set equal to 0. A new slice group map conforming to the dyadic subband partition has been introduced for grouping wavelet coefficients into subbands such that the slices within the LL subband region can be efficiently dropped. The constructed samples by the macro-block decoder then represent the recovered subband coefficients, instead of the recovered pixel samples, when the slice syntax element indicates a wavelet frame.

In summary, the only major modification to the current draft SVC standard is to replace the image upsampling stage for inter-layer prediction with a subband synthesis stage when it is signaled that the decoded frame is represented by wavelet coefficients.

4. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented based on the H.264 JSVM reference software version 7_12. The JVT CE test condition on intra dyadic coding [9] using the CABAC entropy coding mode was adopted for simulation. The CIF sequences BUS, FOOTBALL, FOREMAN, and MOBILE and 4CIF sequences CITY, CREW, HARBOUR and SOCCER were used for simulation. The CIF and 4CIF videos were encoded into two and three spatial scalable layers, respectively, using a variety of base- and enhancement-layer QP combinations, as listed in Table 1.

Table 1: The QP values for simulation

QP_L0	QP1(L1, L2)	QP2(L1, L2)	QP3(L1, L2)	QP4(L1,L2)
16 (R6)	(16,16)	(19,22)	(22,28)	(25,34)
20 (R5)	(20,20)	(23,26)	(26,32)	(29,38)
24 (R4)	(24,24)	(27,30)	(30,36)	(33,42)
28 (R3)	(28,28)	(31,34)	(34,40)	(37,46)
32 (R2)	(32,32)	(35,38)	(38,44)	(41,50)
36 (R1)	(36,x)	(39,x)	(42,x)	(45,x)

The first set of our simulation results are based on a wavelet critical sampling setting. We employ the Daubechies 9/7 wavelet filter set for the subband synthesis of the individual frame at an enhancement layer. The corresponding low-pass analysis filter is utilized for 2-D dyadic down-sampling to generate the input video at a low-resolution layer. The residual coding of the low-pass subband is skipped for each enhancement layer. Thus, the total number of coded samples is the same as that of the source samples.

Figure 4 provides the example Y-PSNR results for decoding the CIF video BUS at full resolution from the QCIF-to-CIF scalable bits-stream in two layers. In this figure, each curve segment shows the coding results generated by using the same base-layer QP value (listed in the column QP_L0 in Table 1) and four different enhancement-layer QP values, respectively. The second test point (starting from the lowest bitrate) in each segment effectively leads to uniform quantization across all the subbands at an enhancement layer and is approximate to the optimal base and enhancement QP combination for decoding video at full resolution in a rate-distortion sense. Our results and reference results are indicated by "JSVM-SBC" and "JSVM_7_12", respectively. As we can see, the results by the proposed system significantly outperform the related JSVM results, particularly around the optimal rate-distortion operation points. The improvements increase with encoding bitrate toward the high bitrate range, the typical rate range adopted for high quality intra-frame coding applications. For reference, we also provide the related PSNR results from using H.264 single-layer (non-scalable) coding.

Figure 5 summarizes the average bitrate improvements by the proposed system over the related JSVM results for decoding our CIF test sequences at full resolution. Each point in the figure represents the average bitrate increase calculated by the method specified in [11] using four sets of enhancement–layer PSNR results associated with the same base QP, as listed in Table 1. The equivalent average PSNR gains [11] are listed in Figure 6. In Table 2 and Table 3, we further provide the average improvements that are calculated over the results from four optimal test points, associated with uniform subband quantization using enhancement-layer QP values 26, 30, 34, and 38, respectively.

In the second setting of our encoding system, we use the same subband setting as the previous experiment but the low-pass subband region at the enhancement layer is still encoded to further refine the coded lowpass subband coefficients. As we can see in Figure 7, the results by the proposed method lead to a smooth rate-distortion curve and consistently outperform the related JSVM results. More importantly, the enhancement-layer performance by the proposed system does not vary significantly with the adopted base QP value, in a clear contrast to the JSVM reference results.

In the third setting of our encoding system, we employ the AVC down-sampler to generate the dyadic down-sampled input video at the base layer. It is useful for the applications that prefer the relatively smooth low-resolution video with reduced aliasing artifacts. As we can see in Figure 8, the proposed algorithm can perform as well as the current JSVM when the conventional smooth down-sampler is employed at the base layer.

The software and more experimental results, including the related CAVLC entropy coding results showing similar performance gains, are available at [10] for further study.

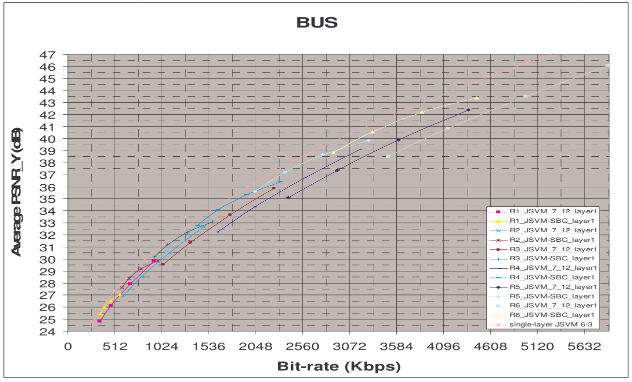

Figure 4: The PSNR results at layer 1 using the subband lowpass filter as down-sampler versus the anchor results.

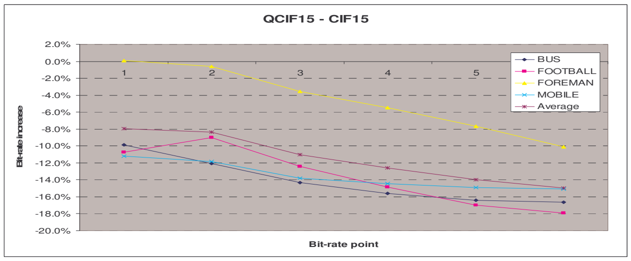

Figure 5: Average bitrate increases at the full resolution layer.

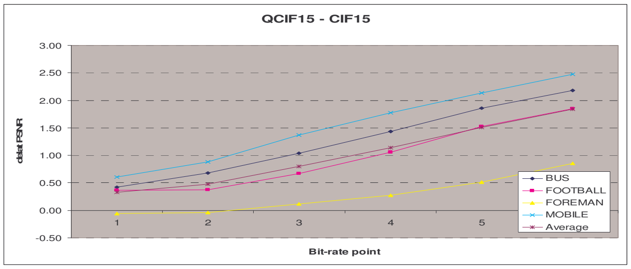

Figure 6: Average PSNR gains at the full resolution layer.

Table 2: Average bitrate increases with uniform quantization

CIF	BUS	FOOTBALL	FOREMAN	MOBILE	Average
%	-16.93	-15.96	-6.96	-15.16	-13.75
4CIF	CITY	CREW	HARBOUR	SOCCER	Average
%	-19.44	-14.60	-26.18	-20.83	-20.26

Table 3: Average PSNR gains with uniform quantization

CIF	BUS	FOOTBALL	FOREMAN	MOBILE	Average
dB	1.27	0.98	0.36	1.44	1.01
4CIF	CITY	CREW	HARBOUR	SOCCER	Average
dB	1.04	0.59	1.79	0.91	1.08

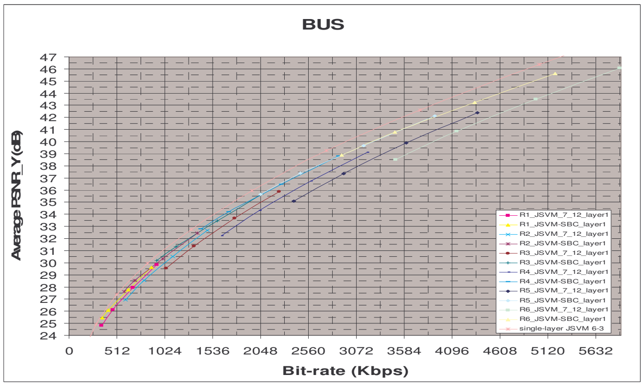

Figure 7: The PSNR results at layer 1 using the subband lowpass filter as down-sampler with the LL band refinement.

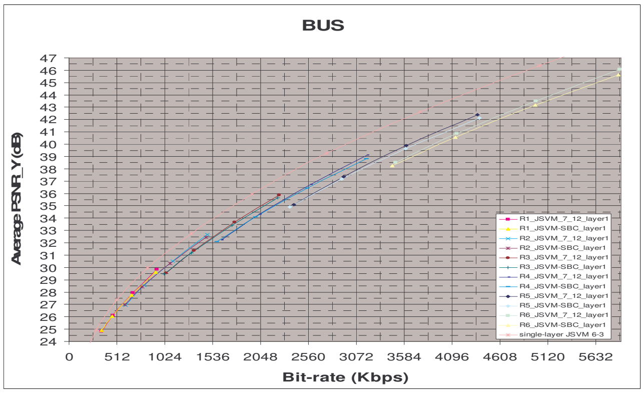

Figure 8: The PSNR results at layer 1 using the AVC down-sampler versus the anchor results.

5. SUMMARY AND CONCLUSION

This paper has presented a new subband coding framework for intra-frame dyadic spatial scalable coding. In addition, we have demonstrated how to efficiently integrate this subband framework with the conventional macroblock and DCT based video coding system by mostly re-using the existing H.264 coding tools.

Different from the conventional wavelet coding, the proposed framework can support flexible spatial down-sampler design for the target applications. The three useful system settings have been presented and evaluated in our coding experiments. The proposed system can choose a relatively smooth lowpass filter for image down sampling to eliminate annoying aliasing artifacts. Alternatively, it can utilize a wavelet critical sampling setting for achieving best compression. Our extensive experimental results demonstrate that, built upon the same system components, the proposed H.264 SBC coding system can significantly improve the current JSVM intra-frame scalable coding system, which is based on the traditional pyramidal coding approach. For example, our results show that for the decoded video at full resolution the proposed H.264 SBC system, using a wavelet critical sampling setting and uniform subband quantization, can achieve average bitrate savings of 13.75% and 20.26% or average PSNR gains of 1.02 dB and 1.08 dB for coding CIF and 4CIF sequences, respectively, compared with the related JSVM results.

Note that under the wavelet critical sampling setting the proposed coding system still has the same total number of the samples for processing/coding as that of the source samples. In this way, the proposed system just works like a conventional single-layer coder without any compression and complexity overhead. It is also worth mentioning that, except for the lowpass subband at the AVC compatible base layer, all the subbands in the proposed H.264 SBC system are encoded by the I_BL macroblock mode for generating our simulation results. However, we should note that these I_BL macroblocks are actually encoded by a transform coding approach because the prediction signal is set equal to 0 over the high-pass subband regions. Therefore, by just re-use of the existing H.264 coding tools to work with the added subband filter banks, we can provide the H.264 standard with an alternative intra-frame coding scheme that is primarily based on the transform coding approach with the embedded spatial scalable functionality. In addition, this new intra-coding scheme does not incur the costs of conventional spatial scalable coding and is free from drift, error propagation, and the complex predictor mode selection process associated with the current H.264 intra-frame hybrid coding.

The development of the MPEG-4 AVC/H.264 standard has been targeted at supporting a wide variety of the video coding application areas, ranging from low bitrate mobile video to digital cinema. The prior experimental results already demonstrated that the current H.264 intra coding could outperform the state-of-the-art image coding standard JPEG2000 in compression efficiency for non-scalable coding [13]. Given the significant gains demonstrated in the SVC core experiments, the proposed H.264 SBC tool has been adopted by the H.264 JSVM for next-phase SVC standard adoption [10][12]. With limited modifications to the standard, our proposed algorithm can add the useful wavelet tool to the H.264 standard for providing improved scalable coding functionality to further accommodate the broad areas of scalable image and intra-frame video coding applications.

6. REFERENCES

[1] S.-T. Hsiang and J. W. Woods “Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling,” in *IEEE ISCAS*, vol. 3, pp. 662–665, May 2000.

[2] A. Said and W. Pearlman, “A new fast and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Trans. on Circuits and Syst. for Video Tech.,* vol. 6, pp. 243-250, June 1996.

[3] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. Image Proc*., pp. 1158-1170, July 2000.

[4] ISO/IEC FCD 15444-1: Information Technology — JPEG 2000 image coding system: Core coding system, Mar. 2000.

[5] T. Wiegand, G. Sullivan, J. Richel, H. Schwartz, M. Wien, eds., “Joint draft 10 of SVC amendment,” JVT-W201, April 2007.

[6] S. Sun, V. Bottreau, “CE4: Texture upsampling results,” JVT-U065, Hangzhou, October 2006.

[7] ISO/IEC 14496-2:1999: Information technology — Coding of audio-visual objects — Part 2: Visual, December 1999.

[8] S.-T. Hsiang, “Preliminary results for intra-frame dyadic spatial scalable coding based on a subband/wavelet filter banks framework,” Doc. JVT-U133, Hangzhou, Oct. 2006.

[9] J.-Z. Xu and S.H. Hsiang, “CE 5: Subband technologies,” Joint Video Team, Doc. JVT-U305, Hangzhou, Oct. 2006.

[10] S.-T. Hsiang, “CE3: Intra-frame dyadic spatial scalable coding based on a subband/wavelet filter banks framework,” Joint Video Team, Doc. JVT-W097, San Jose, April 2007.

[11] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", Doc. VCEG-M33, April 2001.

[12] T. Wiegand, G. Sullivan, J. Richel, H. Schwartz, M. Wien, eds., “Joint scalable video model (JSVM) 10,” JVT-W202, April 2007.

[13] D. Marpe, et. al. “Performance Evaluation of Motion-JPEG2000 in Comparison with H.264 / AVC Operated in Intra Coding Mode”, Proc. SPIE, Vol. 5266, pp. 129-137, Feb. 2004.