APPROACHING OPTIMALITY IN SPATIALLY SCALABLE VIDEO CODING: FROM RESAMPLING AND PREDICTION TO QUANTIZATION AND ENTROPY CODING

Jingning Han and Kenneth Rose

Department of Electrical and Computer Engineering, University of California Santa Barbara, CA 93106
Email:{jingning,rose}@ece.ucsb.edu

ABSTRACT

This paper builds on our recent work on optimal prediction in spatially scalable video coding, and is inspired by earlier work in our lab on optimal approaches for quality (or SNR) scalability. The approach we propose herein complements the optimal enhancement-layer prediction, enabled by transform domain resampling that ensures the base layer information is maximally accessible and usable at the enhancement layer despite their differing signal resolutions, with an optimal approach to quantization and entropy coding that exploits all available information, encapsulated in the appropriate conditional distribution for transform coefficients, to yield a unified coding engine for spatial scalability. For such quantizers to fully exploit base layer information, the enhancement layer transform block size must proportionally match the signal block transformed at the base layer. The overall system incorporates switching that applies the full estimation-theoretic quantizer and entropy coder at the right block size, but may optionally employ other block sizes where it defaults to optimal prediction followed by standard quantization. It is experimentally shown that the proposed scheme provides considerable performance gains over conventional codec and other leading competitors.

***Index Terms*—** Scalable video coding, spatial scalability, estimation theory, entropy-constrained quantizer, arithmetic coding

1. INTRODUCTION

Scalable video coding (SVC) consists of encoding a video sequence into a single bit-stream comprising multiple layers with progressively higher spatial, temporal, or quantization resolutions [1, 2]. Of the various features of SVC, this work is focused on spatial scalability. For simplicity of exposition, we restrict our discussion throughout the text to a two-layered codec, while emphasizing that the proposed approach is extensible to more layers.

A spatial SVC scheme downsamples a high resolution video sequence to lower resolutions, and encodes them into separate layers. The lowest resolution signal is essentially coded by a single-layer coder, while the enhancement layers encode information necessary to reconstruct the sequence at progressively higher spatial resolution. Conventional designs of the enhancement layer coder typically inherit the base layer coder structure, while allowing utilization of additional base layer information to improve the *prediction* quality, and hence the coding performance. For instance, standard approaches perform the enhancement layer prediction in the pixel domain by selecting amongst the inter-layer and inter-frame references the one that minimizes rate-distortion cost (see Sec. 2 for detailed discussion). Significant earlier research has focused on prediction accuracy, e.g., [2, 3]. A notable approach was proposed in [4], where an additional prediction mode that is formed as a linear combination of inter-layer and motion compensated predictions is introduced, which substantially improves the enhancement layer coding performance.

The inherent limitation of the above schemes that do not fully exploit all available information, motivates the search for a truly optimal approach to spatial SVC. Inspiration was drawn from an estimation-theoretic (ET) approach earlier developed by our group [5] for the special SVC setting of *quality* (SNR) scalability, where the *same* original sequence is coded by all the layers but at different quantization resolutions. Thus, the true value of a transform coefficient must lie in the interval determined by base layer quantization. This observation effectively captures all the information provided by the base layer, and is the central postulate of the ET approach in [5], which employs a conditional probability density function (pdf), truncated by the base layer quantization interval, and computes the exact conditional expectation that forms the optimal prediction for the transform coefficient. The ET approach was later enhanced by allowing delayed prediction [6], and extended to incorporate resilience to packet loss [7]. It was also applied in Wyner-Ziv scalable coding [8]. Our recent work [9] further expands this approach to optimize entropy-constrained quantization. All these advances were in the setting of *quality scalability*.

In the setting of spatial scalability, however, the base layer encodes a *downsampled* version of the sequence encoded by an enhancement layer, i.e., different layers quantize different transform coefficients. This poses a major challenge in that the precise quantization intervals and other related base layer information are not directly usable at the enhancement layer. To overcome this obstacle, we developed in [10] a paradigm tailored to enable full exploitation of base layer information, which in conjunction with inter-frame motion compensation provides optimal enhancement layer *prediction*. In [10], a transform domain resampling technique was employed to render base layer quantization intervals accessible and relevant to the enhancement layer codec. It discards high frequency transform coefficients of the original signal and rebuilds the downsampled version from the remaining low frequency coefficients, thereby ensuring a direct mapping between coefficients of the two layers.

Such correspondence opens the door to achieving optimality of additional coder components. Consider the quantization of a random variable given its probability density function (pdf). It is effectively a partition of the support into several mutually exclusive cells, each represented by the corresponding centroid and associated with a probability of containing the source sample. The fundamental design problem can be formulated as a tradeoff between the expected reconstruction distortion (typically measured as the mean squared error), and the rate cost for specifying the cell, which is approximated by the entropy, as is justified in the case of arithmetic coding [11]. The design of the optimal quantizer that minimizes the rate-distortion cost has been intensively studied over decades [12]. In particular, it was shown that for a Laplacian process, a common model for a video signal's temporal innovations in the transform domain, the deadzone quantizer can achieve coding performance fairly close to the optimum [13]. The deadzone quantizer and its variants are widely adopted in single-layer video encoders that employ motion-compensated prediction, and were also "inherited" by the SVC codecs. However, for low frequency transform coefficients the enhancement layer has additional access to base layer information, conditioned on which, the effective pdf may differ significantly from the Laplacian distribution, thereby casting doubt about the efficacy of deadzone quantizers. In this paper, we approach this problem by first deriving the conditional probability distribution, given information from both the base and enhancement layers, based on which an optimal entropy-constrained quantizer can be selected for lossy compression, per transform coefficient of low frequencies.

Further, observe that the above derivation also provides both encoder and decoder with the probability of each cell (or quantization index), which is critical to the efficacy of arithmetic coding. We hence develop a quantizer-adaptive m-ary arithmetic coding (QA-MAC) for 2-D blocks of quantization indices at the enhancement layer, to replace the context-based adaptive binary arithmetic coding (CABAC) inherited from single-layer coding [14]. We note that such coding engine implicitly requires a proportionately larger transform block used at the enhancement layer. Practical hybrid transform coders consider transform blocks of various sizes to optimize trade-offs between coding performance on stationary signals and adaptivity to changes in statistics. To preserve such flexibility here, whenever a smaller block is needed, the ET prediction approach of [10] is employed followed by standard residual coding. The overall proposed coding scheme hence switches between the two spatial scalability modes so as to minimize the rate-distortion cost.

Related prior work in adaptive quantization includes a scalar quantizer design approach that exploits previously encoded local texture information for image coding [15]; and in [16], where a coding scheme switches quantizers depending on base layer information in scalable audio coding, a setting that does not exploit inter frame correlation. Related work offering improvement in entropy coding for SVC includes [17], where better probability estimation was devised for the CABAC context models for successive bit-plane coding, precluding temporal prediction from the enhancement layers.

The proposed approach is implemented in H.264/AVC Scalable Video Coding Extension reference framework to demonstrate its efficacy, but its principles are generally applicable to achieve optimal spatially scalable extensions of other predictive codecs, e.g., HEVC [18] and VPNext [19].

2. BASELINE SPATIALLY SCALABLE CODEC

The standard SVC coder spatially downsamples the original input sequence, and the resultant lower resolution signal is encoded by a standard single-layer codec into the base layer. The choice of downsampler is not standardized by H.264/AVC SVC, and commonly employed strategies include the windowed sinc filter, pixel decimation, etc. The enhancement layer predictor switches between the motion compensated reference from prior frames at the same layer, and the current base layer reconstruction (upsampled via pixel filtering), for the minimal rate-distortion cost. A significant amount of study has been devoted to designing the interpolation filter, and to determine whether supporting additional filters would be beneficial. However, no clear winner was identified [2]. A notable method was proposed in [3] where the upsampling filter is derived to match the downsampling operation while accounting for the quantization noise in the base layer reconstructed pixels. In [4], an additional mode that generates the prediction as a linear combination of inter-layer and inter-frame predictions is proposed for more efficient enhancement layer coding, where the weight coefficients are derived as a function of the resampling operations.

The above scheme is commonly referred to as multi-loop design. The standard codec uses a variant called single-loop design, where the base layer reconstructed residuals are upsampled and optionally added to the inter-frame motion compensated reference. It is known that multi-loop provides slightly better coding performance than single-loop design. We hence modify the H.264/SVC framework to support multi-loop prediction and use it as reference.

3. THE UNIFIED ESTIMATION-THEORETIC FRAMEWORK FOR RESAMPLING AND QUANTIZATION

We devise a unified ET approach that incorporates transform domain resampling operations to enable optimum enhancement layer predictive quantization. In the discussion that follows, the base layer block is of dimension $M \times M$, and is obtained by downsampling a block of size $N \times N$ at the enhancement layer.

3.1. Transform Domain Resampling

We assume separability of the 2-D transform, and hence first present the basic principle in the framework of a 1-D transform. Consider a vector of pixels $\underline{a} = [a_0, a_1, \cdots, a_{N-1}]^T$, with inter-pixel correlation ≈ 1. Here the superscript T denotes transposition. The optimal approach to convert $\underline{a}$ into a vector of dimension $M(< N)$ is to apply the Karhunen-Loeve transform (KLT) to fully decorrelate the samples and discard the lower energy $N - M$ coefficients. It is well known that the DCT exhibits decorrelation and energy compaction properties approaching that of the KLT, and it is commonly adopted as a substitute due to its low implementation complexity. Let T_N denote the N-point DCT matrix, and $\underline{\alpha}_N = T_N\underline{a}$ is the DCT of vector $\underline{a}$. Define $f_0(t) = \sqrt{\frac{1}{N}}$; $f_j(t) = \sqrt{\frac{2}{N}}cos(j\pi t)$, $j = 1, \cdots, N-1$, analog cosine functions with a period that is a sub-multiple of the time interval $[0, 1]$. Thus, the basis functions (rows) of T_N can be generated by sampling $\{f_j(t)\}$ at time instances $t = \frac{1}{2N}, \frac{3}{2N}, \cdots, \frac{2N-1}{2N}$. Consequently, sampling at the rate $\frac{1}{N}$ the continuous-time signal $a(t) = \sum_{j=0}^{N-1} \alpha_j f_j(t)$, where α_j are the transform coefficients in $\underline{\alpha}_N$, yields exactly the discrete-time signal $\underline{a}$. Now define $g_0(t) = \sqrt{\frac{1}{M}}$, $g_j(t) = \sqrt{\frac{2}{M}}cos(j\pi t)$, $j = 1, \cdots, M-1$, the analog cosine functions to be sampled at rate $\frac{1}{M}$ to yield the basis functions for a DCT of dimension M. Approximating the signal $a(t)$ using only M of the N transform coefficients in $\underline{\alpha}_N$ is done by retaining the M lowest frequency coefficients:

$$\tilde{a}(t) \approx \sum_{j=0}^{M-1} \alpha_j f_j(t) = \sum_{j=0}^{M-1} \left(\sqrt{\frac{M}{N}} \alpha_j \right) g_j(t). \quad (1)$$

Hence we downsample from N-point pixel vector $\underline{a}$ to M-point vector $\underline{b} = \sqrt{\frac{M}{N}}\, T_M^T \begin{pmatrix} I_M & 0_M \end{pmatrix} T_N \underline{a}$, where I_M and 0_M denote identity and null matrices, respectively. Conversely, up-sampling from M-point pixel vector $\underline{b}$ to N-tuple is accomplished by zero-padding the high frequency coefficients:

$$\hat{\underline{a}} = \sqrt{\frac{N}{M}}\, T_N^T \begin{pmatrix} I_M \\ 0_M \end{pmatrix} T_M \underline{b}.$$

Under the assumption that the DCT possesses performance very close to the KLT, the resultant $\hat{\underline{a}}$ has minimum mean squared distance from the original vector $\underline{a}$, and downsampling to $\underline{b}$ maximally preserves the information in $\underline{a}$. Related material on DCT domain resampling can be found in, e.g., [10, 20, 21]. The extension to 2D blocks is straightforward by applying downsampling (or upsampling) sequentially to columns and rows. This transform domain resampling approach can in general serve as an alternative to the pixel-domain downsampling and interpolation traditionally employed in spatial SVC. However, as discussed next, this resampling method is of particular advantage to the proposed ET spatial SVC paradigm.

3.2. Estimation-Theoretic Coding Engine

We now consider encoding the enhancement layer blocks $\{A_i, i = 0, \cdots, 3\}$ in frame n (Fig.1). The entire region R is mapped into block B in the base layer frame via the *transform domain downsampling* previously described in Sec. 3.1. Let $x_n^e(i,j)$, where $i,j \in \{0, \cdots, N-1\}$, denote the value of the transform coefficient at frequency (i,j) obtained by applying a DCT of size $N \times N$ to R. Using (1), the first $M \times M$ transform coefficients of the resultant DCT are scaled appropriately to yield the transform coefficients of the base layer block B:

$$x_n^b(i,j) = \frac{M}{N} x_n^e(i,j), i,j \in \{0, \cdots, M-1\}. \quad (2)$$

The base layer coding process essentially prescribes a quantization interval $\mathcal{I}_n^b(i,j)$ that contains the true value of $x_n^b(i,j)$, which summarizes all the information provided by the base layer about the transform coefficient $x_n^b(i,j)$. Accordingly the interval that contains the true value of $x_n^e(i,j)$ is:

$$x_n^e(i,j) \in \mathcal{I}_n^e(i,j) = \frac{N}{M} \mathcal{I}_n^b(i,j) \quad (3)$$

3.2.1. *Optimal Entropy-Constrained Predictive Quantizer*

Having established the correspondence between $x_n^e(i,j)$ and $x_n^b(i,j)$, we are now able to derive the optimal quantizer and subsequent entropy coding, while fully accounting for all the available information. Let $\hat{x}_{n-1}^e(i,j)$ denote the transform coefficient of the same frequency as $x_n^e(i,j)$ of the motion-compensated reference block generated from the previously reconstructed frame. This enhancement layer reference can then be combined with the known interval $\mathcal{I}_n^e(i,j)$, in an estimation-theoretic framework to obtain the conditional pdf of coefficient $x_n^e(i,j)$.

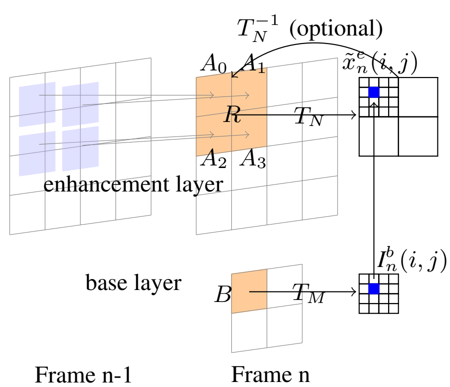

Fig. 1: Estimation-theoretic enhancement layer coding.

We model the DCT coefficients of blocks along a motion trajectory as an AR process per frequency coefficient [5]: $x_n = \rho x_{n-1} + z_n$, where $\{z_n\}$ are i.i.d innovations of the process with pdf $p_Z(z_n)$.

The implicit assumption in standard pixel domain motion compensated prediction that the temporal correlation coefficient $\rho \approx 1$, is retained here for simplicity at all frequencies. Assuming that $\hat{x}_{n-1}^e(i,j) \approx x_{n-1}^e(i,j)$, we obtain the temporally conditioned pdf $p(x_n^e(i,j)|\hat{x}_{n-1}^e(i,j)) \approx p_Z(x_n^e(i,j) - \hat{x}_{n-1}^e(i,j))$. The base layer further indicates that $x_n^e(i,j) \in \mathcal{I}_n^e(i,j)$, which refines the conditional pdf of x_n^e to[1]

$$p(x_n^e|\hat{x}_{n-1}^e, \mathcal{I}_n^e) \approx \begin{cases} \frac{p_Z(x_n - \hat{x}_{n-1}^e)}{\int_{\mathcal{I}_n^b} p_Z(x_n - \hat{x}_{n-1}^e) dx_n} & x_n \in \mathcal{I}_n^b \\ 0 & else \end{cases}. \quad (4)$$

We assume $\{z_n\}$ form Laplacian distribution, i.e., $p_Z(z_n) = \frac{1}{2}\lambda e^{-\lambda|z_n|}$, where λ is a frequency dependent factor [5]-[7]. The optimal predictor at the enhancement layer is hence

$$\tilde{x}_n^e = E[x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e], \quad (5)$$

the centroid of the above pdf over the *entire* interval $\mathcal{I}_n^e$ [5].

If the prediction and quantization operations are separately processed, then the traditional course of action would now be to quantize the residual $(x_n^e - \tilde{x}_n^e)$ via a deadzone quantizer and encode the index (typically) using CABAC. However, such separate treatment of the prediction and quantization suffers from significant under-utilization of the available information. In particular, in anticipation of the optimally matched entropy coder to be discussed next, the optimal entropy-constrained quantizer for x_n^e, given the conditional pdf $p(x_n^e|\hat{x}_{n-1}^e, \mathcal{I}_n^e)$ can be obtained via a variant of the Lloyd-Max algorithm [22]. Consider a scalar quantizer of N levels. Let the decision or boundary points of the partition be denoted by $\{t_i|i = 0, 1, \cdots, N\}$, and the reproduction levels by $\{r_i|i = 1, 2, \cdots, N\}$. The interval $\mathcal{I}_n^e$ bounds the support of the signal, t_0 and t_N. The necessary conditions for optimality of an entropy-constrained quantizer of N levels were specified in [12]:

$$\beta\, log_2(\frac{P_{i+1}}{P_i}) = (r_{i+1} - r_i)(r_{i+1} + r_i - 2t_i),\ \forall i = 1,\ 2, \cdots,\ N-1,$$

where P_i is the probability of the i^{th} region, and β is the Lagrangian multiplier whose value may be varied to obtain the desired point on the operational rate-distortion curve. This necessary condition leads to an entropy-constrained Lloyd-Max quantizer design. The variant of the design algorithm for the enhancement layer quantizer is derived in a straightforward manner and the pseudo-code is given in Fig. 2, where ϵ determines the convergence test. Upon convergence, the rate-distortion cost associated with this N-level scalar quantizer can be calculated. We then vary the value of positive integer N to find the one that provides overall minimum rate-distortion cost as the optimum quantizer for x_n, given conditional pdf $p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^b)$.

The Laplacian memoryless property allows for a set of generic quantizers to be pre-calculated and stored during the initial stage of coding process. The encoder can then simply fetch the needed quantizer, conditioned on the motion compensated reference and base layer information, eliminating the need to redesign quantizers. Hence the overall increment in computational complexity is modest.

3.2.2. *Quantizer-Adaptive M-ary Arithmetic Coding*

The H.264/AVC standard and the SVC extension employ CABAC for entropy coding, which adjusts the probability models according

[1] The frequency index (i,j) is omitted to streamline notation where there is no risk of confusion.

repeat

$r_N \leftarrow c$

for $i = 1$ to $(N-1)$ **do**

$$r_i \leftarrow \frac{\int_{t_{i-1}}^{t_i} x_n\, p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n}{\int_{t_{i-1}}^{t_i} p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n}$$

$$P_i \leftarrow \int_{t_{i-1}}^{t_i} p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n$$

$$P_{i+1} \leftarrow \int_{t_i}^{t_{i+1}} p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n$$

$$t_i \leftarrow \frac{1}{2}(r_i + r_{i+1}) - \frac{\beta}{2}\frac{log_2(P_{i+1}/P_i)}{r_{i+1}-r_i}$$

end for

$$c \leftarrow \frac{\int_{t_{N-1}}^{t_N} x_n\, p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n}{\int_{t_{i-1}}^{t_i} p(x_n|\hat{x}_{n-1}^e, \mathcal{I}_n^e)dx_n}$$

until $|c - r_N| < \epsilon$

Fig. 2: Pseudo code for the enhancement layer entropy-constrained predictive quantizer design.

to coding information of the neighboring blocks in a spatially adaptive manner, thereby achieving significant rate reduction over other variable length based methods [14].

The entropy coder proposed here exploits the fact that the optimal quantizer *explicitly* provides the probability mass function, which can be used to optimize the index coding, namely, the proposed quantizer-adaptive m-ary arithmetic coding (QAMAC). In particular, we assign to the most probable cell of each transform coefficient the index zero. The encoder then scans the 2-D block in descending order of coefficient total probability of significance (instead of the traditional zig-zag order), and generates a binary-valued significance map which is coded by the binary arithmetic coder. To encode the significant coefficients, the QAMAC employs an m-ary arithmetic coder, the recursive interval subdivisions of which are conditioned on the quantizers for the significant coefficients. Suppose a transform coefficient is coded by an N-level quantizer, where the k^{th} region is most probable. The significance map indicates that the true value of this coefficient does not fall into the most probable region, which eliminates r_k (indexed zero) from the sample space and refines the probability mass function of this significant coefficient as $\tilde{P}_i = \frac{P_i}{1-P_k}$, $\forall i \neq k$, $i \in \{1, 2, \cdots, N\}$. The internal range of the arithmetic coder is thus divided into $(N-1)$ subintervals, the i^{th} of which has a length proportional to $\tilde{P}_i$. Depending on the observed symbol value, the corresponding subinterval will be chosen as the new current interval. The binary expansion of the number pointing into this interval effectively represents the sequence of symbols coded so far, and hence forms the coded bit stream.

For coding the high frequency coefficients, a regular dead-zone quantizer followed by CABAC is employed as usual. It is important to note that the above unified coding engine requires that a fixed transform dimension is used by the enhancement layer, and is optimal for locally stationary signals. In practice, it is worthwhile to allow various transform block sizes, to optimize the tradeoff between coding performance of stationary signal and adaptivity to the changes in statistics. To maintain the flexibility in transform choices, whenever a smaller transform block is needed, the ET prediction of [10] is used followed by regular residual coding. The overall scheme hence switches between these two modes to minimize the rate-distortion cost.

4. SIMULATION RESULTS

The proposed ET coding engine is implemented in H.264/SVC framework, and is referred to as ETQ-SVC. The standard SVC codec was modified to support multi-loop design denoted by H.264/SVC-ML. The linear combination approach of [4] was also included and

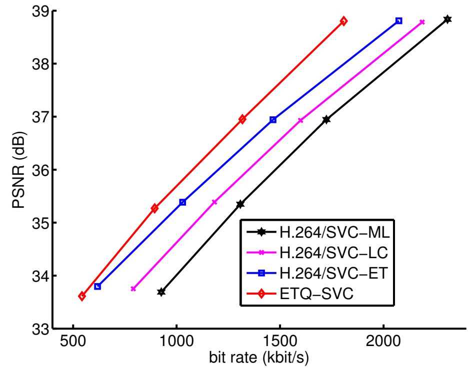

Fig. 3: Enhancement layer coding performance of $coastguard$ at CIF resolution. The base layer is coded at 700 $kbit/s$ and $QCIF$.

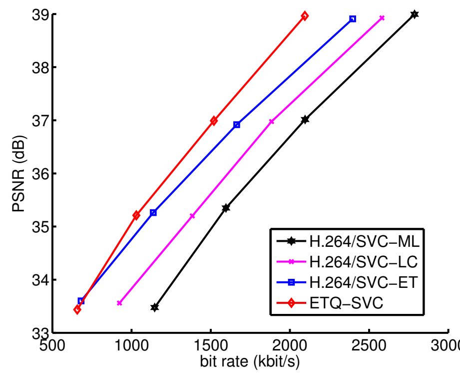

Fig. 4: Enhancement layer coding performance of $harbour$ at CIF resolution. The base layer is coded at 1000 $kbit/s$ and $QCIF$.

marked by H.264/SVC-LC. The ET prediction for spatial SVC [10] recently developed by us to improve prediction but not quantizer and entropy encoder, was also included as H.264/SVC-ET. All the competing SVC codecs use the same base layer coder, and employ regular quarter-pixel motion search and single reference frame. We emphasize that more sophisticated inter-frame motion compensated prediction methods can be directly incorporated in the proposed framework. The test sequences are coded in $IPPP$ format at frame rate of 30 f/s. The enhancement layer coding performance for sequence $coastguard$ at CIF resolution is presented in Fig.3, where we fix the base layer and vary the Lagrangian multiplier β to generate enhancement layer operational points. Clearly, the proposed scheme significantly outperforms other competing schemes. Similar coding performance gains are achieved for sequence $harbour$ at CIF resolution (Fig.4). Experiments with other test sequences yielded similar gains.

5. CONCLUSIONS

This paper proposes a novel enhancement layer coding engine for optimal compression in spatial SVC. The approach complements a transform domain resampling technique and efficiently combines all relevant information from both base and enhancement layers in an ET framework to derive the conditional pdf, which enables derivation of the optimal entropy-constrained predictive quantizer and conditional arithmetic coder. Considerable and consistent coding gains are obtained by using the proposed ET coding engine, in comparison to standard H.264/SVC as well as our earlier ET approach to spatial SVC, which focused on enhancement layer prediction.

6. REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 17, pp. 1103–1120, Sep. 2007.

[2] C. A. Segall and G. J. Sullivan, "Spatial scalability within the H.264/AVC scalable video coding extension," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 17, no. 9, pp. 1121–1135, Sep. 2007.

[3] C. A. Segall and A. Katsaggelos, "Resampling for spatial scalability," *Proc. IEEE ICIP*, pp. 181–184, Oct. 2006.

[4] R. Zhang and M. Comer, "Efficient inter-layer motion compensation for spatially scalable video coding," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 18, pp. 1325–1334, Oct. 2008.

[5] K. Rose and S. L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Img. Proc.*, vol. 10, no. 7, pp. 965–976, July 2001.

[6] J. Han, V. Melkote, and K. Rose, "Estimation-theoretic approach to delayed prediction in scalable video coding," *Proc. IEEE ICIP*, pp. 1289–1292, Sep. 2010.

[7] J. Han, V. Melkote, and K. Rose, "A unified framework for spectral domain prediction and end-to-end distortion estimation in scalable video coding," *Proc. IEEE ICIP*, pp. 3278–3281, Sep. 2011.

[8] H. Wang, N. M. Cheung, and A. Ortega, "A framework for adaptive scalable video coding using Wyner-Ziv techniques," *EURASIP Jrnl. App. Sig. Proc.*, Jan. 2006.

[9] J. Han and K. Rose, "Towards predictor, quantizer and entropy coder optimality in scalable video coding," *Proc. IEEE ICIP*, Oct. 2012.

[10] J. Han, V. Melkote, and K. Rose, "An estimation-theoretic approach to spatially scalable video coding," *Proc. IEEE ICASSP*, pp. 817–820, Mar. 2012.

[11] J. Rissanen and G. G. Langdon, "Universal modeling and coding," *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 12–23, Jan. 1981.

[12] T. Berger, "Minimum entropy quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 149–157, Mar. 1982.

[13] G. J. Sullivan, "Efficient scalar quantization of exponential and Laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1365–1374, Sep. 1996.

[14] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 13, no. 7, pp. 620–636, July 2003.

[15] A. Ortega and M. Vetterli, "Adaptive scalar quantization without side information," *IEEE Trans. Img. Proc.*, vol. 6, no. 5, pp. 665–676, May 1997.

[16] A. Aggarwal, S. L. Regunathan, and K. Rose, "Efficient bitrate scalability for weighted squared error optimization in audio coding," *IEEE Trans. Audio, Speech, and Language Proc.*, vol. 14, no. 4, pp. 1313–1327, July 2006.

[17] S.-H. Kim and Y.-S. Ho, "Fine granular scalable video coding using context-based binary arithmetic coding for bit-plane coding," *IEEE Trans. Circ. Sys. Video Tech.*, vol. 17, no. 10, pp. 1301–1310, Oct. 2007.

[18] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circ. Sys. Video Tech.*, 2012.

[19] open source, *http://www.webmproject.org/code/*.

[20] J. M. Adant, P. Delogne, E. Lasker, B. Macq, L. Stroobants, and L. Vandendorpe, "Block operations in digital signal processing with application to TV coding," *Signal Processing*, vol. 13, pp. 385–397, Dec. 1987.

[21] S. A. Martucci, "Image resizing in the discrete cosine transform domain," *Proc. IEEE ICIP*, vol. 13, pp. 244–247, Oct. 1995.

[22] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.