ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Theoretic Approach for Spatially Scalable Video Coding

Snehal N. Chaudhari [1], Prof. T. D. Shep[2]

PG Student, Dept. of Communication Engineering, MIT Aurangabad, Maharashtra, India[1]

Assistant Professor, Dept. of Electronics & Communication Engineering, MIT Aurangabad, Maharashtra, India [2]

ABSTRACT: A scalable extension to the H.264/AVC video coding standard has been developed within the Joint Video Team (JVT), a joint organization of the ITU-T Video Coding Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). The extension allows multiple resolutions of an image sequence to be contained in a single bit stream. In this paper, we introduce the spatially scalable extension within the resulting Scalable Video Coding standard. The high-level design is described and individual coding tools are explained. Additionally, encoder issues are identified. Finally, the performance of the design is reported.

KEYWORDS: H.264/AVC, Scalable Video Coding (SVC), spatial scalability.

I.INTRODUCTION

With the expectation that future applications will support a diverse range of display resolutions and transmission channel capacities, the Joint Video Team (JVT) has developed a scalable extension [1], [2] to the state-of-the-art H.264/AVC video coding standard [3]–[6]. This extension is commonly known as Scalable Video Coding (SVC) and it provides support for multiple display resolutions within a single compressed bit stream (or in hierarchically related bit streams), which is referred to here as spatial scalability. Additionally, the SVC extensions support combinations of temporal scalability (frame rate enhancement) and quality scalability (fidelity enhancement for pictures of the same resolution) with the spatial scalability feature [2]. This is achieved while balancing both decoder complexity and coding efficiency.

The resolution diversity of current display devices motivates the need for spatial scalability. Specifically, larger format, high definition displays are becoming common in consumer applications, with displays containing over two million pixels readily available. By contrast, lower resolution displays with between ten thousand and one hundred thousand pixels are also popular in applications constrained by size, power and weight. Unfortunately, transmitting a single representation of a video sequence to the range of display resolutions available in the market is impractical.
For example, it is rarely justifiable to design a device with low display resolution with the capacity for decoding and down-sampling high-resolution video material. Such a requirement could increase the cost and power of the device to the point of exceeding the very constraints that determined its display resolution. In addition, sending the high-resolution details that are ultimately not shown on the display for such a device is a waste of its receiving channel bit rate. Diverse, limited, and time-varying channel capacity provides a second motivation for spatial scalability. Here, the concern is that channel capacity may preclude the reliable transmission of high-resolution video to specific devices or at specific time instances. Spatial scalability allows for the rapid bit rate adaptation that can be a necessity in such scenarios. This bit rate adaptation is achieved without trans-coding operations or feedback to a complex real-time encoding process, both of which can introduce unacceptable complexity and delay.

The purpose of this paper is to discuss key concepts of spatial scalability within the SVC extension. This project is the fourth in a historical series of efforts to standardize spatially SVC schemes after prior efforts in MPEG-2 , H.264 , and MPEG-4 although the prior designs were basically not successful in terms of industry adoption. This paper points out several ways in which the new design addresses the problems of those prior approaches.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

II. SVC OVERVIEW

Spatial scalability is achieved by pyramid approach. The pictures of different spatial layers are independently coded with layer specific motion parameters. In order to improve the coding efficiency of the enhancement layers in comparison to simulcast, additional inter-layer prediction mechanisms have been introduced to remove the redundancies among layers. These prediction mechanisms are switchable so that an encoder can freely choose a reference layer for an enhancement layer to remove the redundancy between them. Since the incorporated inter-layer prediction concepts include techniques for motion parameter and residual prediction, the temporal prediction structures of the spatial layers should be temporally aligned for an efficient use of the inter-layer prediction. Three inter-layer prediction techniques, included in the scalable video coding as shown in fig.1 are:

• Inter-layer motion prediction: In order to remove the redundancy among layers, additional MB modes have been introduced in spatial enhancement layers. The MB partitioning is obtained by up-sampling the partitioning of the co-located 8x8 block in the lower resolution layer. The reference picture indices are copied from the co-located base layer blocks, and the associated motion vectors are scaled by a factor of 2. These scaled motion vectors are either directly used or refined by an additional quarter-sample motion vector refinement. Additionally, a scaled motion vector of the lower resolution can be used as motion vector predictor for the conventional MB modes.

• Inter-layer residual prediction: The usage of inter-layer residual prediction is signal by a flag that is transmitted for all inter-coded MBs. When this flag is true, the base layer signal of the co-located block is block-wise up-sampled and used as prediction for the residual signal of the current MB, so that only the corresponding difference signal is coded.
• Inter-layer intra prediction: Furthermore, an additional intra MB mode is introduced, in which the prediction signal is generated by up-sampling the co-located reconstruction signal of the lower layer. For this prediction it is generally required that the lower layer is completely decoded including the computationally complex operations of motion-compensated prediction and de blocking. However, this problem can be circumvented when the inter-layer intra prediction is restricted to those parts of the lower layer picture that are intra-coded. With this restriction, each supported target layer can be decoded with a single motion compensation loop

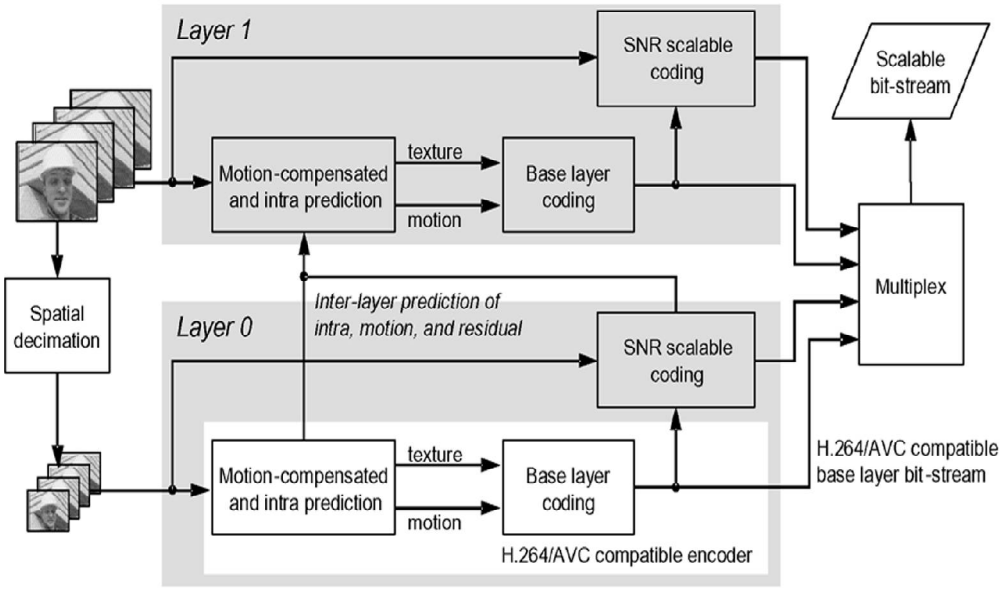

Fig.1 Svc Structure

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

III. THE UNIFIED ESTIMATION-THEORETIC FRAMEWORK FOR SVC

A. Transform Domain Re-sampling-

We assume separability of the 2-D transform, i.e., it is accomplished by applying 1-D operations sequentially along the vertical and horizontal directions. Hence, for clarity of exposition, we first present the main ideas in the framework of a 1-D transform. Consider a vector of pixels a = [a0, a1, …….., aN−1]T , with inter-pixel correlation ≈ 1. Here the superscript T denotes matrix transposition. The optimal approach to compress a into a vector of dimension M(< N) is to apply the Karhunen-Loeve transform (KLT) to fully de correlate the samples and discard the lower energy N − M coefficients. It is well known that the DCT exhibits de correlation and energy compaction properties approaching that of the KLT, and is commonly adopted as a substitute due to its low implementation complexity. Let TN denote the N-point DCT matrix, and α N = T N (a) is the DCT of vector a. Define:

$$f0(t) = \sqrt{\frac{1}{N}}; \quad fj(t) = \sqrt{\frac{2}{N}} \cos(j\pi t),\ j = 1, \ldots\ldots, N-1, \quad (1)$$

analog cosine functions with a period that is a sub-multiple of the time interval [0, 1]. Thus, the j th basis function (row) of TN can be generated by sampling fj(t) at time instances t = 1\2N , 3\2N , ……. , 2N−1\2N . Consequently, the continuous time signal $a(t) = \sum_{j=0}^{N-1} \alpha j f j(t)$ where α j is the j th transform coefficient in α N, when sampled at the rate 1\ 2N yields exactly the vector a. Now define,

$$g0(t) = \sqrt{\frac{1}{M}}, \qquad gj(t) = \sqrt{\frac{2}{M}} \cos(j\pi t), \quad j = 1, \ldots\ldots\ldots, M-1, \quad (2)$$

the analog cosine functions which when sampled at rate 1\2M yield the basis functions for a DCT of dimension M. The best approximation (in mean squared error sense) for the signal a(t) using only M of the N transform coefficients in α N is that provided by choosing the M coefficients of lowest frequency

$$a(t) \approx \sum_{j=0}^{M-1} \alpha j f j(t) \;=\; \sum_{j=0}^{M-1} \left(\sqrt{\frac{M}{N}}\, \alpha j\right) gj(t). \qquad (3)$$

This implies that the N-point pixel vector a can be down sampled by a factor M\ N to b as:

$$b = \sqrt{\frac{M}{N}}\, T\, (\ IM \quad 0M\ )\, T\, N\, a, \qquad (4)$$

where IM and 0M denote the identity and null matrices, respectively, of dimension M × M. Conversely, the up-sampling from the M point pixel vector b to an N- tuple can be accomplished by inserting zeros as high frequency coefficients:

$$a = \sqrt{\frac{N}{M}}\ \ T\ \ (IM \backslash 0M)\ T\ (\ b) \qquad (5)$$

This transform domain re-sampling approach can in general serve as an alternative to the pixel domain down sampling and interpolation traditionally employed in spatial SVC. However, as discussed next, this re-sampling method is of particular advantage to the proposed ET spatial SVC paradigm.

B. The Optimal Enhancement Layer Prediction-

Consider encoding the enhancement layer blocks {Ai, i = 0, …….. , 3} in frame n . The entire region R is mapped into block B in the base layer frame via the transform domain down sampling previously described. Let x(i, j), where i, j ∈{0,…… ,N − 1}, denote the value of the transform coefficient at frequency (i, j) obtained by applying a DCT of size N × N to R. the first M × M transform coefficients of the resultant DCT are scaled appropriately to yield x(i, j), i, j ∈{0,…… ,M − 1}, the transform coefficients of the base layer:

$$x_n^b(i, j) \;=\; M\backslash N\, x_n^e\,(i, j),\ \ i, j \in \{0, \ldots\ldots\ldots, M-1\}. \qquad (6)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

However, such an approach that combines reconstructions in the pixel domain suffers from significant under-utilization of the information provided by the base layer. In particular, note that on account of the transform domain re sampling the following relation holds:

$$x_n^b(i, j) \in I_n^b(i, j) = N\backslash M I_n^b(i, j) \ (i, j), \ i, \ j \in \{0, \ldots\ldots\ldots, M-1\}, \qquad (7)$$

which implies that the information in the base layer quantization intervals directly translates into information about transform coefficients at the enhancement layer. This information cannot be utilized in the pixel domain. The ET prediction approach we now describe improves coding performance by specifically utilizing this interval information.

IV. RESULT AND DISCUSSION

The proposed unified ET approach in the JSVM reference framework. The competing codec was created by modifying standard H.264/SVC to support multi-loop inter-layer prediction, using the 4-tap polyphase filter and de blocking operations for up sampling, in addition to the inter-frame prediction, which is hence forth referred to as H.264/SVC ML. The matched up sampling filter proposed in was further tested, which is denoted by H.264/SVC MF. The scheme that allows an additional mode, where the prediction is formed as a linear combination of inter-layer and inter-frame predictions was also implemented in the modified H.264/SVC framework, and is referred to as H.264/SVC LC. Regular pixel domain motion estimation is enabled at quarter-pixel resolution for all four coded.

Hence, for fair comparison, the transform domain down sampled sequence is used in all SVC code. The enhancement layer coding performance of the four coded for the sequence foreman at CIF resolution is shown in Fig 2.

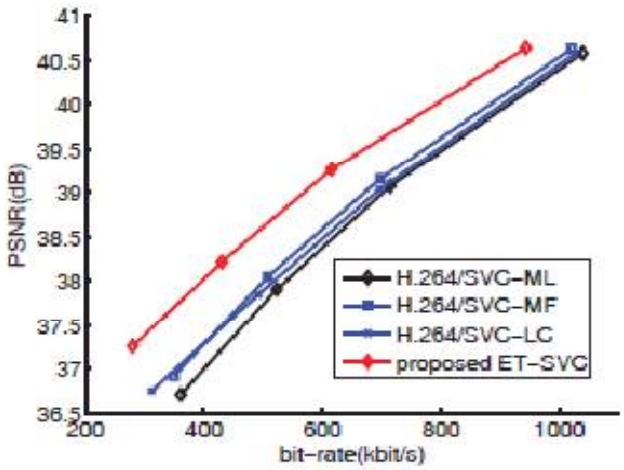

Fig 2 : Comparison of the coding performance of the competing spatial SVC approaches: The testing sequence is foreman at CIF resolution. The base layer is at QCIF resolution, and is coded at the 408kbit/s with reconstruction quality 39.7dB

The subjective test required the viewer to visually compare two down sampled and coded versions of the same original video sequence (one using the pixel domain down sampler and the other using the transform domain down sampler) against a reference un-coded version. Since pixel domain down sampling is generally accepted as the standard, we employed the un-coded pixel domain down sampled sequence as the reference. The two coded base layer sequences are obtained at similar bit-rates (subject to small differences due to encoder constraints) – Table I provides the bit-rates for the two coded versions of different test clips featured in the test.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

TABLE I
The Resolutions AND Bit-Rates Of Coded Sequences In The Test

Test Clip	Frame Size	Bit Rate (kbps)	
		pixel domain	transform domain
bus	176 × 144	758	738
foreman	176 × 144	207	218
mobile	176 × 144	1087	1115
city	352 × 288	1347	1328
harbour	352 × 288	1915	2103
soccer	352 × 288	1161	1181
oldtown	960 × 512	2355	2269
parkjoy	960 × 512	12844	12995

V.CONCLUSION

This paper proposes a novel unified framework for re sampling and estimation theoretic enhancement layer prediction in spatial SVC. Aided by unconventional transform domain re sampling, the ET prediction approach maximally utilizes information from the base layer and prior enhancement layer reconstructions, and combines them into an appropriate conditional pdf. The enhancement layer prediction is then obtained as the corresponding conditional expectation. Considerable and consistent coding gains are obtained by using the proposed unified framework, in comparison to standard H.264/SVC and one of its variants. The ET scheme is also devised that greatly reduces the codec complexity, while retaining major coding performance gains.

REFERENCES

1. H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 9, 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. 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.
4. X. Li, N. Oertel, A. Hutter, and A. Kaup, “Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding,” IEEE Trans. Circuits Syst., Video Technol., vol. 19, no. 2, pp. 193–205, Feb. 2009.
5. J. Han, V. Melkote, and K. Rose, “Estimation-theoretic approach to delayed prediction in scalable video coding,” IEEE Proc. ICIP, pp. 1289–1292, Sep 2010.
6. J. Han, V. Melkote, and K. Rose, “A unified framework for spectral domain prediction and end-to-end distortion estimation in scalable video coding,” IEEE Proc. ICIP, Sep 2011.
7. X. Li et al, “Rate-Complexity-Distortion evaluation for hybrid video coding”, IEEE Trans. on CSVT, vol. 21, pp. 957 - 970, July 2011
8. J. Han, V. Melkote, and K. Rose, “An estimation-theoretic framework for spatially scalable video coding with delayed prediction,” in Proc. 19th Int. Packet Video Workshop (PV), May 2012, pp. 167–172.
9. G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
10. J. Han, V. Melkote, and K. Rose, “An estimation-theoretic approach to delayed decoding of predictively encoded video sequences,” IEEE Trans. Image Process., vol. 22, no. 3, pp. 1175–1185, Mar. 2013
11. G.J. Sullivan et al, “Standardized extensions of High Efficiency Video Coding (HEVC)”, IEEE J-STSP, vol. 7, no. 6, pp. 1001 – 1016, Dec. 2013.
12. K. Shah, “Reducing the complexity of Inter-prediction mode decision for HEVC”, M.S. Thesis, University of Texas at Arlington, UMI Dissertation Publishing, April 2014.