|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11**  12th Meeting: Geneva, CH, 14–23 Jan. 2013 | Document: JCTVC-L0412  WG 11Number :\_ m28051 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Non-TE 3**: On estimation theoretic prediction for enhancement layer residual in scalable video coding | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Ankur Saxena  Felix C. Fernandes  1301 E. Lookout Drive,  Richardson, TX – 75082, USA. | Tel: Email: | 1-9727617761 [asaxena@sta.samsung.com](mailto:asaxena@sta.samsung.com),  [felix.f@sta.samsung.com](mailto:felix.f@sta.samsung.com) |
| *Source:* | Samsung Electronics Co., Ltd | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

In this contribution, an overview of estimation-theoretic prediction for enhancement layer residual in scalable video coding is provided. Coding of the enhancement layer residual prediction in scalable video coding can be enhanced by using information jointly from the previously reconstructed enhancement layer pictures, and the reconstructed base layer information. Traditionally, the base layer and previous enhancement layer information is combined by choosing amongst various linear combinations of base and enhancement layer information in the pixel domain via a Rate-Distortion search, and the criteria to choose a particular linear combination is signaled to the decoder, incurring some transmission overhead. However, these pixel-domain schemes are sub-optimal, and motivated by these facts, an overview of an estimation-theoretic scheme is provided for utilizing the base and enhancement layer information where the information is combined in the *transform domain*. The presented scheme does not require any additional signaling as well. Past research shows that such a scheme provides substantial gains without much increase in complexity.

# Introduction

In the current scalable extensions of HEVC standardization, various prediction schemes for the enhancement layer residual [1, 2, 3] are being evaluated in Tool Experiment 3 [4]. All these schemes are implemented in the pixel domain, and define various modes for adding a *linear* refinement from the base layer to the enhancement layer information to obtain the enhancement layer prediction residue. Then a Rate-Distortion search is performed to choose amongst the various modes, and this information is signaled to the decoder with prediction obtained from the best mode.

However, the above schemes are ad-hoc, and do not capture the base layer information efficiently. Further, they assume that the optimal estimate is a linear combination of the base- and enhancement-layer estimates. To overcome these shortcomings, a ***transform-domain*** Estimation Theoretic Enhancement-layer Residual Prediction **(ETERP)** was proposed in [5]. ETERP uses a Laplace-Markov model [7] to model the DCT coefficients of the motion-compensated enhancement layer residue. In [5], the information from the reconstructed base-layer error residue is used to compute the enhancement-layer residue, in a non-linear estimation theoretic manner. Specifically, the quantization interval for the base-layer residue (in the transform domain) is denoted as (a,b) and can be obtained from the quantized error residuals at both the encoder and decoder for each transform coefficient. The midpoint of the quantization interval is the centroid of that interval. The prediction error at the enhancement layer can then be calculated via the following equation from [5]:

where is a DCT transform coefficient in current frame , is the reconstruction (after appropriate motion compensation) for the transform coefficient in the previous frame for the enhancement layer; is the corresponding transform coefficient for the base layer prediction for the current frame ; and (a,b) denotes the quantization interval in which the quantized base layer residual lies. Note that, for simplicity we omit the index for the various frequencies when a transform such as the DCT is taken, i.e., denote as etc. More details on how to compute such an expectation are in [5].

By using the above expression, we note that no additional signaling information or Rate-distortion search at the encoder is required to communicate the base layer refinement to the decoder, and this improves the compression efficiency. However, the above scheme was presented a decade ago, and the scheme in [5] is incompatible with recent advances in video coding. We have refined the scheme appropriately for compatibility with various other new video coding tools in SMuC 0.1.1. However, at the time of submission of this contribution, our implementation in SMuC 0.1.1 was incomplete. Hence we do not present SMuC simulation results in this contribution.

# Discussion on Simulation Results

As mentioned above, our implementation in SMuC 0.1.1 was incomplete. However, to show the potential of the scheme, we present some results from the literature in [5] and [6], where the performance of **ETERP** is shown in H.263+, and H.264/AVC.

The table 1 from [5] is presented first. P1 and P2 denote the situations when only base layer, or enhancement layer respectively was used for enhancement layer prediction, H.263+ denotes the scheme used in H.263 codec where the codec can choose between (a) the prediction from current base-layer, (b) the prediction from the previous enhancement layer, and (c) the prediction from the weighted sum of different sources; and ET denotes the Estimation-Theoretic scheme which forms the basis of this contribution. The last column “Non-scalable” refers to the situation when a non-scalable coder with same total rate is used.

******

From the above contribution, PSNR gains of upto 0.52 dB were obtained for the ET scheme over H.263+.

Next, we present a recent result from [6] for spatial scalability where the ET implementation was performed in H.264/AVC codec. More details are in [6]. From the figure, we can see that gains upto almost 1 dB are obtained via the ET technique.



From the above results, it can be easily seen that substantial gains are possible by using **ETERP** scheme in H.263+ and H.264/AVC. We plan to continue working on ETERP, and present results in the SMuC at the next JCTVC meeting.

# Conclusions

In this contribution, we provide an overview of an estimation-theoretic scheme for optimally combining the base and enhancement layer information in the transform domain to obtain a prediction for the enhancement layer pixels. The presented scheme has been shown to provide substantial gains in coding efficiency when implemented in H.263+ and H.264/AVC. We intend to continue further study of the presented scheme for a possible eventual adoption in the scalability extensions of the HEVC standard.

# References

1. S. Lasserre, F. Le Léannec, E. Nassor, J. Taquet, N. Ouedraogo, S. Pautet, C. Gisquet, G. Laroche, T. Poirier, Y. Verdavaine, and E. François, “Description of the scalable video coding technology proposal by Canon Research Centre France”, JCTVC-K0041, Shanghai, China, Oct 2012.

2. J. Chen, K. Rapaka, X. Li, V. Seregin, L. Guo, M. Karczewicz, G. Van der Auwera, J. Sole, X. Wang, C. J. Tu, and Y. Chen, “Description of scalable video coding technology proposal by Qualcomm (configuration 2)”, JCTVC-K0036, Shanghai, China, Oct 2012.

3. O. Bici, D. Bugdayci, A. Aminlou, A. Hallapuro, M. Hannuksela, J. Lainema, and K. Ugur, “Description of scalable video coding technology proposal by Nokia (encoder configuration 1)”, JCTVC-K0040, Shanghai, China, Oct 2012.

4. X. Li, E. Francois, P. Lai, D. Kwon, and A. Saxena, “Description of Tool Experiment B3: Combined Prediction in SHVC,” JCTVC-K1103, Shanghai, China, Oct 2012.

5. K. Rose and S. Regunathan , “Toward Optimality in Scalable Predictive Coding,” IEEE Transactions on Image Processing, vol. 10, number 7, pages 965-976, July 2001.

6. J. Han, V. Melkote, and K. Rose, “An estimation-theoretic approach to spatially scalable video coding”, Proc. IEEE ICASSP, March 2012.

7. R. C. Reininger and J. D. Gibson, “Distributions of the two-dimensional DCT coefficients for images,” IEEE Trans. Comm., pp. 835–839, June 1983.

# Patent rights declaration(s)

**Samsung Electronics Co., Ltd., may have IPR relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**