Non-TE 3: On estimation theoretic prediction for enhancement layer residual in scalable video coding

JCTVC-L0412

Ankur Saxena & Felix C. Fernandes

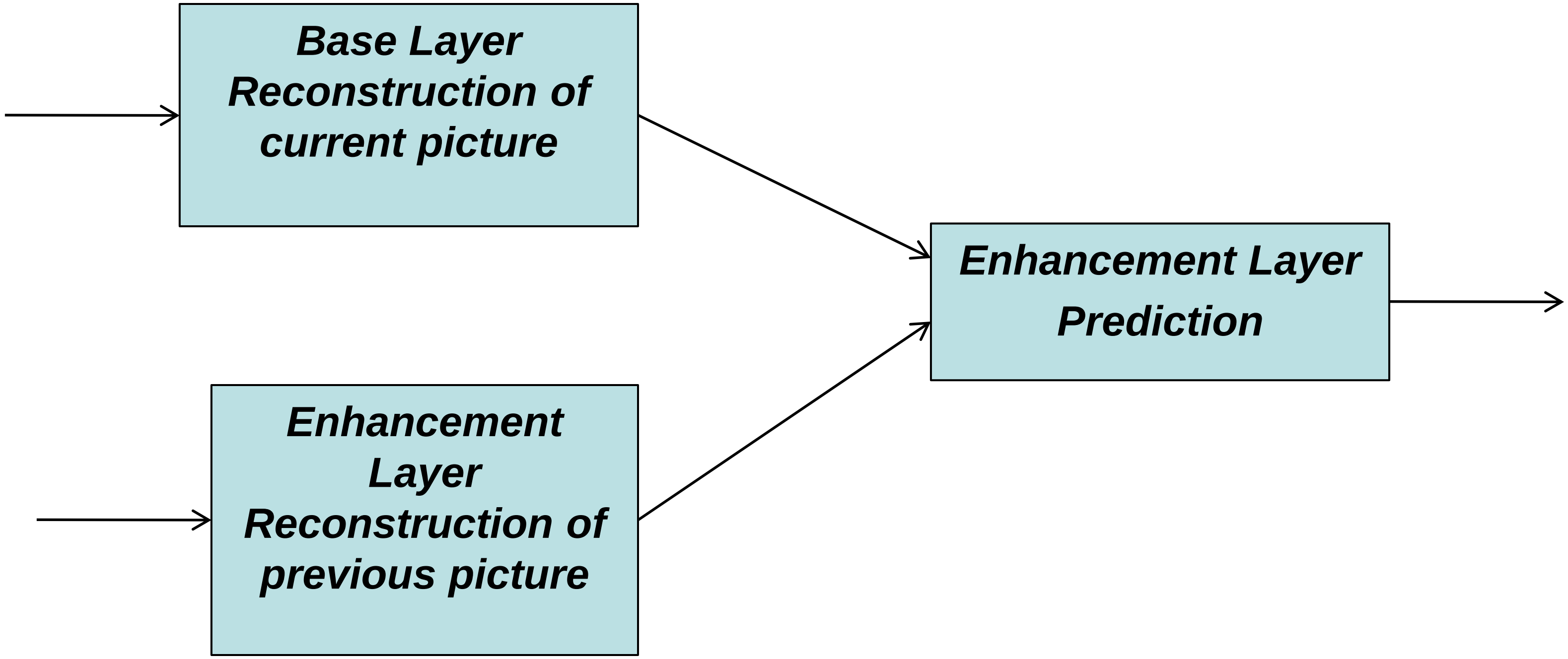
Samsung Electronics, Co. Ltd.

Dallas R&D Center

Problem Statement

- Enhancement layer prediction: How to ***efficiently*** use information from
  (a) base-layer reconstruction of current picture; and
  (b) enhancement layer reconstruction of previous picture

- Related to Tool Experiment 3.



Limitations of Current Schemes in TE 3

- Information from base and enhancement layers are combined in a ***linear*** manner, which is sub-optimal
- Different modes (e.g., for combining BL and EL information) are defined in an ad-hoc manner, and a Rate-Distortion search is performed to choose the best mode.
- Flag is explicitly signaled to the decoder.
- Additional motion compensation
- Additional Memory Accesses

Proposed Solution

- **Basis of the Proposal:** Toward Optimality in Scalable Predictive Coding, Kenneth Rose and Shankar Regunathan, IEEE Transactions on Image Processing, 2001.

- The enhancement layer prediction is an optimal combination of base and enhancement layer in the **transform domain**

Enhancement Layer Model

- At **Enhancement layer**, DCT coefficients for inter-frame (motion compensated) coefficients can be modeled as a Laplace-Markov process with $\rho \rightarrow 1$.

$$X_n = \rho X_{n-1} + Z_n$$
$$\approx \hat{X}_{EL,n-1} + Z_n$$

X_n : DCT Coefficient for current picture

Z_n : Innovation process with Laplacian density

$\hat{X}_{EL,n-1}$: Reconstruced DCT coefficient for previous picture n-1

Information from Base Layer

- In transform domain,

$$\hat{X}_{BL,n} = \tilde{X}_{BL,n} + \hat{E}_{BL,n} \quad ;$$

Reconstruction = Prediction + Error Residual

$$\hat{E}_{BL,n} \in Q^b : \textit{some quantization interval } (a,b)$$

- Therefore $X_n = \tilde{X}_{BL,n} + \hat{E}_{BL,n} \in (\tilde{X}_{BL,n} + a, \tilde{X}_{BL,n} + b)$

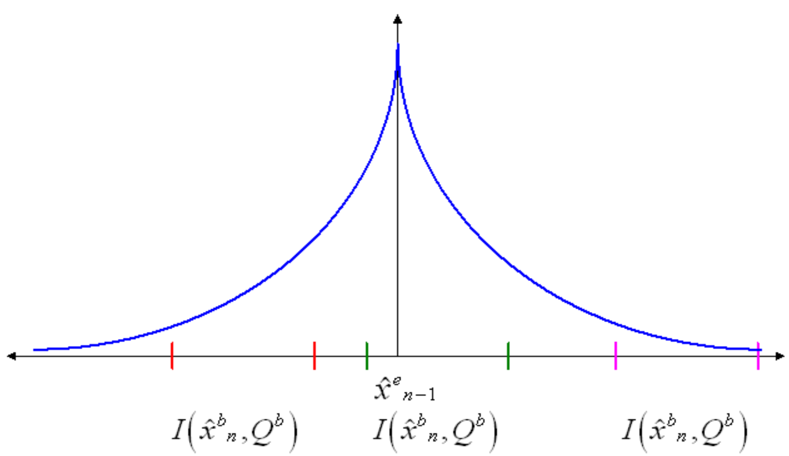
is all the information that can be utilized from Base Layer.

Optimal Prediction

- Optimal **prediction** at Enhancement Layer is:

$$
\begin{aligned}
\tilde{X}_{EL,n} &= E\ [X_n\ \ |\hat{X}_{EL,n-1}\,,\hat{X}_{BL,n}\ ] \\
&= E\ [X_n\ \ |\hat{X}_{EL,n-1}\,, X_n \epsilon\ (\tilde{X}_{BL,n} +\ a,\ \tilde{X}_{BL,n} + b\ )] \\
&\approx \hat{X}_{EL,n-1}\ \ + E\ [Z_n\ \mid\ \ X \epsilon\ (\tilde{X}_{BL,n} +\ a,\ \tilde{X}_{BL,n} + b\ )] \\
&= \hat{X}_{EL,n-1}\ \ + E\ [Z_n\ \mid\ \ Z_n \epsilon\ (\tilde{X}_{BL,n} - \hat{X}_{EL,n-1}\ \ + a,\ \tilde{X}_{BL,n} - \hat{X}_{EL,n-1}\ \ + b\ )]
\end{aligned}
$$

Optimal Prediction



Results (1, From [5]) for SNR scalability

TABLE I

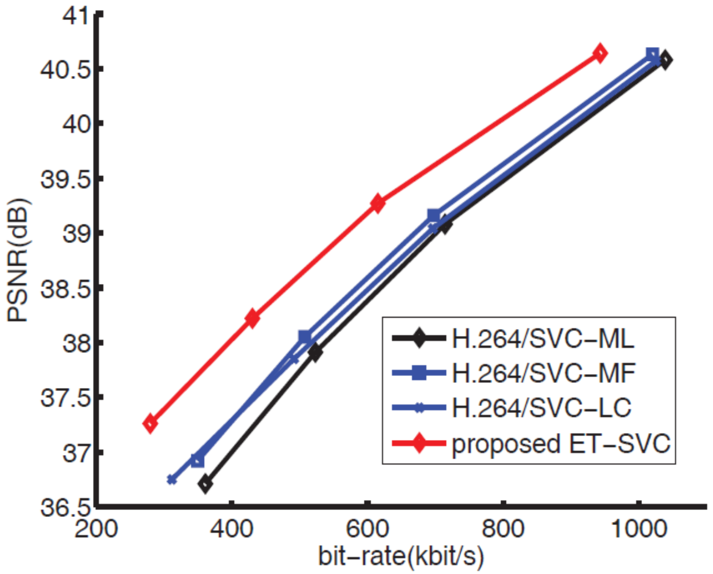
PERFORMANCE OF *TWO-LAYER* SCALABLE CODERS, WHICH DIFFER IN THEIR ENHANCEMENT-LAYER PREDICTION MODULE, AND NON-SCALABLE CODER. ENCODED SEQUENCE: *CARPHONE* AT QCIF RESOLUTION. THE ENTRIES PROVIDE THE AVERAGE PSNR (IN dB) OF RECONSTRUCTED FRAMES VERSUS TOTAL RATE OF BASE AND ENHANCEMENT LAYERS (Kbps). TOTAL NUMBER OF FRAMES WAS 267 AT FRAME SKIP OF 3. FOR ALL THE METHODS, THE BASE-LAYER RATE WAS FIXED AT 32 Kbps, AND THE CORRESPONDING PSNR WAS 31.52 dB

Rate	P1	P2	H.263+	ET	Non-scalable
64	32.80	31.99	33.26	33.70	34.46
80	33.43	33.41	34.27	34.79	35.43
96	34.03	34.50	35.13	35.65	36.28
128	35.08	36.17	36.62	37.13	37.68
160	35.98	38.54	38.85	39.20	39.57

P1: Using base layer information only
P2: Using enhancement layer information only
H.263+: P1+P2+ Weighted prediction mode
ET : Presented Estimation Theoretic scheme
Non-Scalable: Single codec at EL + BL rate

Roughly 0.5 dB gain for Estimation Theoretic Scheme

Results (2, From [6]) for Spatial Scalability



Upto 1 dB gain for Estimation Theoretic Scheme

Fig. 3: 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 $408kbit/s$ with reconstruction quality $39.7dB$ (with respect to the downsampled sequence).

Conclusions

- Presented an overview of Estimation Theoretic Enhancement Layer Prediction Residual (ETERP) scheme.
- Optimal prediction using all the base layer and enhancement layer information.
- Further study for the technique in S-HEVC Test Model

References

- 1. S. Lasserre et al., “Description of the scalable video coding technology proposal by Canon Research Centre France”, JCTVC-K0041, Shanghai, China, Oct 2012.
- 2. J. Chen et al., “Description of scalable video coding technology proposal by Qualcomm (configuration 2)”, JCTVC-K0036, Shanghai, China, Oct 2012.
- 3. O. Bici et al., “Description of scalable video coding technology proposal by Nokia (encoder configuration 1)”, JCTVC-K0040, Shanghai, China, Oct 2012.
- 4. X. Li et al., “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.

Thank You