A Unified Estimation-Theoretic Framework For Error-Resilient Scalable Video Coding

Jingning Han[1], Vinay Melkote[1,2], and Kenneth Rose[1]

[1]Signal Compression Lab
Department of Electrical and Computer Engineering
University of California, Santa Barbara

[2] Dolby Labs Inc., San Francisco

July 10, 2012

A Sneak Peek

- Estimation-theoretic scalable video coding (ET-SVC) - a transform domain approach to optimal enhancement layer prediction
  - Optimally utilizes all available information including base-layer quantization intervals accessible only in the transform domain
- Robustness of ET-SVC to packet-losses requires choosing coding modes that minimize End-to-End Distortion (EED)
  - Conventionally calculated in the pixel domain, accounts for effects of quantization as well as packet losses
  - A well established approach for accurate EED estimation - Recursive Optimal Per-Pixel Estimate (ROPE)
- Achieve optimality on both fronts?

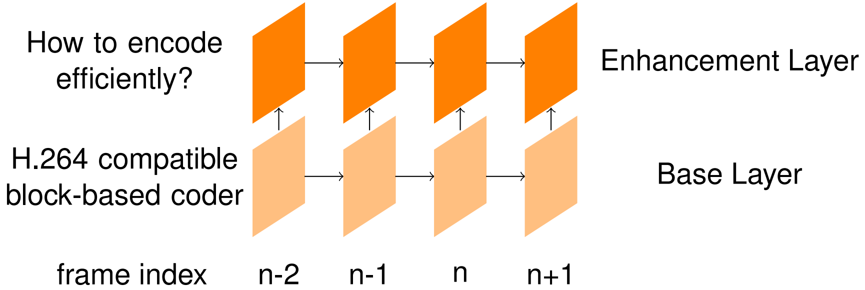
  A longstanding difficulty due to the fundamental conflict of operating space!

A Sneak Peek

- Proposed solution: A unified framework complementing ET-SVC with Spectral Coefficient-wise Optimal Recursive Estimate (SCORE) - EED estimation that operates directly in the transform domain
- Added bonus: enables estimation-theoretic (optimal) enhancement layer concealment at the decoder, fully accounted for by encoder EED estimation
- Overall system provides significant performance gains over competing optimized H.264/SVC solution

Scalable Video Coding

- Encode a video sequence into two layers of fidelity scalability.

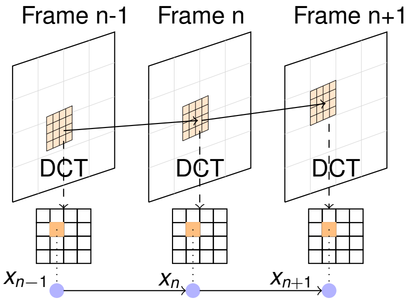


Enhancement Layer Prediction in SVC

- Information accessible for prediction at the enhancement layer:
  - High quality (enhancement layer) reconstructions of prior samples - inter frame prediction
  - Coarsely quantized (base layer) reconstructions of current samples - inter layer prediction
- Conventional solutions work in pixel domain
  - Weighted sum of the enhancement-layer motion compensation and base-layer reconstructed pixels
  - Adaptively choose the mode that minimizes rate-distortion cost

A Transform Domain Model

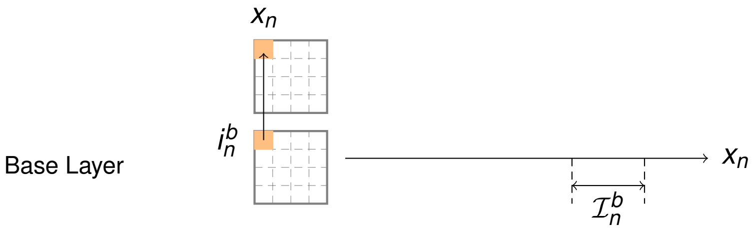
- DCT blocks along a motion trajectory form an AR process per frequency



- Specifically. $x_n = \rho x_{n-1} + z_n$, where $\{z_n\}$ are the i.i.d innovations with pdf $p_Z(z_n)$
- Advantage: largely eliminates spatial correlation before applying a temporal evolution model to individual frequency components

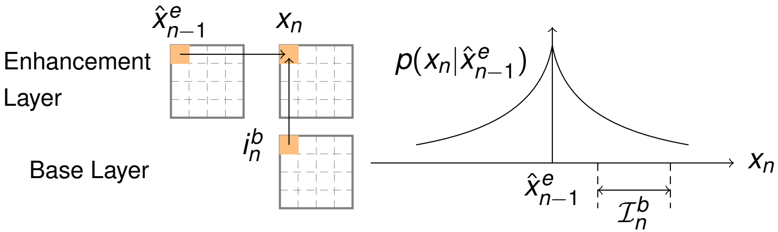
Estimation-Theoretic SVC

- All the relevant information provided by the base layer: $x_n \in \mathcal{I}_n^b$



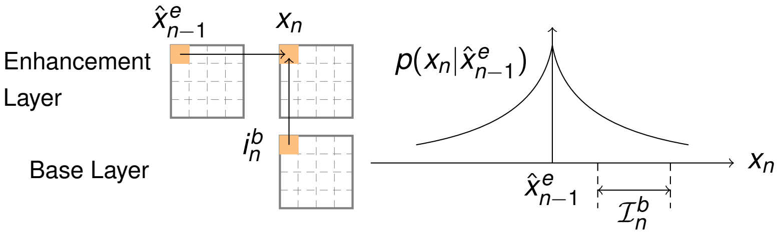
Estimation-Theoretic SVC

- All the relevant information provided by the base layer: $x_n \in \mathcal{I}_n^b$
- The information provided by prior enhancement layer: $p(x_n|\hat{x}_{n-1}^e)$



Estimation-Theoretic SVC

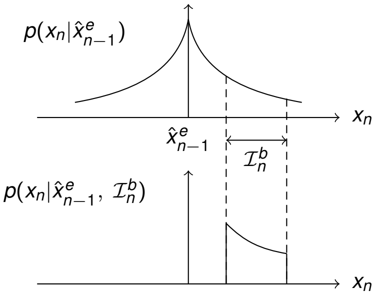
- All the relevant information provided by the base layer: $x_n \in \mathcal{I}_n^b$
- The information provided by prior enhancement layer: $p(x_n|\hat{x}_{n-1}^e)$
- How to optimally combine the two types of information?



Estimation-Theoretic SVC

- The conditional pdf of x_n hence can be expressed as:

$$p(x_n|\hat{x}^e_{n-1}, \mathcal{I}^b_n) \approx \begin{cases} \frac{p_Z(x_n - \hat{x}^e_{n-1})}{\int_{\mathcal{I}^b_n} p_Z(x_n - \hat{x}^e_{n-1}) dx_n} & x_n \in \mathcal{I}^b_n \, , \\ 0 & \text{otherwise} \, . \end{cases}$$



- The optimal enhancement layer prediction of x_n given all the available information is the non-linear estimate

$$f(\mathcal{I}^b_n, \hat{x}^e_{n-1}) = E(x_n|\hat{x}^e_{n-1}, \mathcal{I}^b_n)$$

- The prediction residue $x_n - f(\mathcal{I}^b_n, \hat{x}^e_{n-1})$ is quantized and coded into the enhancement layer

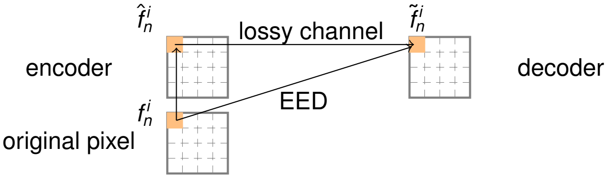
ET-SVC over Lossy Networks

- ET-SVC provides significant compression gains when the base layer interval and enhancement layer motion compensated reference are guaranteed
- What if the channel is lossy? Amongst other effects, the calculation of the base layer interval at the decoder would itself be subject to drift
- Drift due to packet loss can be mitigated via judicious choice of per-macroblock coding modes, partitions and QPs:
  - Intra mode vs Inter mode at the base layer
  - Inter-layer prediction mode vs ET prediction-mode at the enhancement layer
- Optimize coding decisions to minimize End-to-End Distortion (EED)
  - EED includes the effect of quantization as well as packet losses: can only be *estimated* at the encoder
- Efficient utility of the ET-SVC framework over lossy networks mandates an EED estimation mechanism that accommodates its transform domain operation

EED Estimation via ROPE

- ROPE: an established approach to recursively calculate EED *per pixel* while accounting for encoder and decoder operations, and channel stochasticity
- The decoder reconstruction $\tilde{f}_n^i$ is *a random variable* w.r.t the encoder. Expected EED for this pixel is:

$$E\{(f_n^i - \tilde{f}_n^i)^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\}.$$



- ROPE update recursions compute up to second moments of reconstructed pixels
- The pixel-domain framework of ROPE is incompatible with the non-linear transform domain operations of ET-SVC

Proposed Approach for EED Estimation

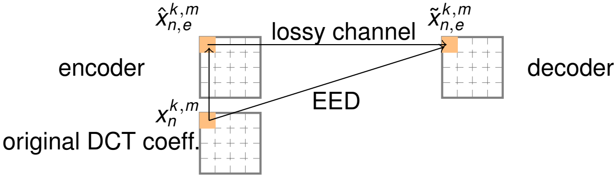
- The obvious: calculate EED in the transform domain - mean squared error is preserved under unitary transformation
- The not so obvious: complications arise due to interaction with motion compensation

Proposed Approach for EED Estimation

- The obvious: calculate EED in the transform domain - mean squared error is preserved under unitary transformation
- The not so obvious: complications arise due to interaction with motion compensation
- Proposed Solution: Spectral Coefficient-wise Optimal Recursive Estimate(SCORE)
  - SCORE provides a near-accurate per-transform coefficient estimate of EED
  - Recursively computes first and second moments of reconstructions of transform coefficients of on-grid blocks in a frame
  - Overcomes intricacies due to off-grid motion compensation references
  - Explicitly accounts for ET prediction in its update recursions

SCORE: Expected Distortion

- Specific focus on SCORE recursions at the enhancement layer
  - $x_n^{k,m}$: unquantized value of transform coefficient m in block k of frame n.
  - $\hat{x}_{n,e}^{k,m}$: enhancement layer encoder reconstruction of this coefficient.
  - $\tilde{x}_{n,e}^{k,m}$: enhancement decoder reconstruction, possibly after concealment. A random variable w.r.t the encoder.



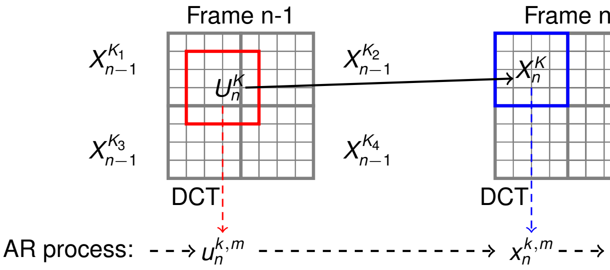
- The enhancement layer EED of coefficient $x_n^{k,m}$ is

$$E\{(x_n^{k,m} - \tilde{x}_{n,e}^{k,m})^2\} = (x_n^{k,m})^2 - 2x_n^{k,m} E\{\tilde{x}_{n,e}^{k,m}\} + E\{(\tilde{x}_{n,e}^{k,m})^2\}.$$

- SCORE recursively computes $E\{\tilde{x}_{n,e}^{k,m}\}$ and $E\{(\tilde{x}_{n,e}^{k,m})^2\}$

SCORE: Off-Grid Reference Challenge

- SCORE computes and retains first and second moments of transform coefficients of *on-grid blocks* of a frame
- However, an on-grid block in the current frame can have an *off-grid* motion compensation reference, whose moments will feature in the recursions



- Can we calculate first and second moments of off-grid transform coefficients from those of on-grid transform coefficients?

Solution to the Off-Grid Reference Challenge

- DCT is a linear transformation: there exist constants $a_{i,m}$ such that,

$$\tilde{u}_{n,e}^{k,m} = \sum_{i=1}^{4} \sum_{m=0}^{15} a_{i,m} \tilde{x}_{n-1,e}^{k_i,m} .$$

- The required first and second moments of off-grid blocks:

$$\begin{aligned} E\{\tilde{u}_{n,e}^{k,m}\} &= \sum_{i=1}^{4} \sum_{m=0}^{15} a_{i,m} E\{\tilde{x}_{n-1,e}^{k_i,m}\} , \\ E\{(\tilde{u}_{n,e}^{k,m})^2\} &= \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{m=0}^{15} \sum_{l=0}^{15} a_{i,m} a_{j,l} E\{\tilde{x}_{n-1,e}^{k_i,m} \tilde{x}_{n-1,e}^{k_j,l}\} . \end{aligned}$$

Solution to the Off-Grid Reference Challenge

- DCT is a linear transformation: there exist constants $a_{i,m}$ such that,

$$\tilde{u}_{n,e}^{k,m} = \sum_{i=1}^{4} \sum_{m=0}^{15} a_{i,m} \tilde{x}_{n-1,e}^{k_i,m} .$$

- The required first and second moments of off-grid blocks:

$$\begin{aligned} E\{\tilde{u}_{n,e}^{k,m}\} &= \sum_{i=1}^{4} \sum_{m=0}^{15} a_{i,m} E\{\tilde{x}_{n-1,e}^{k_i,m}\} , \\ E\{(\tilde{u}_{n,e}^{k,m})^2\} &= \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{m=0}^{15} \sum_{l=0}^{15} a_{i,m} a_{j,l} E\{\tilde{x}_{n-1,e}^{k_i,m} \tilde{x}_{n-1,e}^{k_j,l}\} . \end{aligned}$$

- Uncorrelatedness: a very good approximation in the transform domain.

$$E\{\tilde{x}_{n-1,e}^{k_i,m} \tilde{x}_{n-1,e}^{k_j,l}\} \approx E\{\tilde{x}_{n-1,e}^{k_i,m}\} E\{\tilde{x}_{n-1,e}^{k_j,l}\}, \; k_j \neq k_i \text{ or } m \neq l$$

SCORE: Enhancement Layer Update Recursions

- **Case 1:** Coding modes: Base layer - Intra, Enhancement layer - ET Prediction
- Current base layer packet lost with probability p_b, enhancement layer PLR p_e

Events: Base Layer	Events: Enhancement Layer	Probability	Enhancement Layer Decoder Reconstruction of $x_n^{k,m}$
received	received	$(1-p_b)(1-p_e)$	$\hat{r}_{n,e}^{k,m} + f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$
received	lost	$(1-p_b)p_e$	$\tilde{x}_{n,b}^{k,m}$
lost	received	$p_b(1-p_e)$	$\hat{r}_{n,e}^{k,m} + \tilde{u}_{n,e}^{k,m}$
lost	lost	$p_b p_e$	$\tilde{x}_{n,b}^{k,m}$

- SCORE update recursion:

$$
\begin{aligned}
E\{\tilde{x}_{n,e}^{k,m}\} = {} & (1-p_b)(1-p_e)(\hat{r}_{n,e}^{k,m} + E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})\}) \\
& + (1-p_b)p_e E\{\tilde{x}_{n,b}^{k,m}\} \\
& + p_b(1-p_e)(\hat{r}_{n,e}^{k,m} + E\{\tilde{u}_{n,e}^{k,m}\}) \\
& + p_b p_e E\{\tilde{x}_{n,b}^{k,m}\}
\end{aligned}
$$

SCORE: Enhancement Layer Update Recursions

- **Case 1:** Coding modes: Base layer - Intra, Enhancement layer - ET Prediction
- Current base layer packet lost with probability p_b, enhancement layer PLR p_e

Events		Probability	Enhancement Layer Decoder Reconstruction
Base Layer	**Enhancement Layer**		**of $x_n^{k,m}$**
received	received	$(1-p_b)(1-p_e)$	$\hat{r}_{n,e}^{k,m} + f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$
received	lost	$(1-p_b)p_e$	$\tilde{x}_{n,b}^{k,m}$
lost	received	$p_b(1-p_e)$	$\hat{r}_{n,e}^{k,m} + \tilde{u}_{n,e}^{k,m}$
lost	lost	$p_b p_e$	$\tilde{x}_{n,b}^{k,m}$

- SCORE update recursion:

$$E\{\tilde{x}_{n,e}^{k,m}\} = (1-p_e)(\hat{r}_{n,e}^{k,m} + (1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})\} + p_b E\{\tilde{u}_{n,e}^{k,m}\}) + p_e E\{\tilde{x}_{n,b}^{k,m}\}$$

$$\begin{aligned} E\{(\tilde{x}_{n,e}^{k,m})^2\} &= (1-p_e)((\hat{r}_{n,e}^{k,m})^2 + 2\hat{r}_{n,e}^{k,m}((1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})\} + p_b E\{\tilde{u}_{n,e}^{k,m}\}) \\ &+(1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})^2\} + p_b E\{(\tilde{u}_{n,e}^{k,m})^2\}) + p_e E\{(\tilde{x}_{n,b}^{k,m})^2\} \end{aligned}$$

SCORE: Enhancement Layer Update Recursions

- **Case 1:** Coding modes: Base layer - Intra, Enhancement layer - ET Prediction
- Current base layer packet lost with probability p_b, enhancement layer PLR p_e

Events: Base Layer	Events: Enhancement Layer	Probability	Enhancement Layer Decoder Reconstruction of $x_n^{k,m}$
received	received	$(1-p_b)(1-p_e)$	$\hat{r}_{n,e}^{k,m} + f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$
received	lost	$(1-p_b)p_e$	$\tilde{x}_{n,b}^{k,m}$
lost	received	$p_b(1-p_e)$	$\hat{r}_{n,e}^{k,m} + \tilde{u}_{n,e}^{k,m}$
lost	lost	$p_b p_e$	$\tilde{x}_{n,b}^{k,m}$

- SCORE update recursion:

$$E\{\tilde{x}_{n,e}^{k,m}\} = (1-p_e)(\hat{r}_{n,e}^{k,m} + (1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})\} + p_b E\{\tilde{u}_{n,e}^{k,m}\}) + p_e E\{\tilde{x}_{n,b}^{k,m}\}$$

$$\begin{aligned} E\{(\tilde{x}_{n,e}^{k,m})^2\} &= (1-p_e)((\hat{r}_{n,e}^{k,m})^2 + 2\hat{r}_{n,e}^{k,m}((1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})\} + p_b E\{\tilde{u}_{n,e}^{k,m}\}) \\ &+(1-p_b)E\{f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})^2\} + p_b E\{(\tilde{u}_{n,e}^{k,m})^2\}) + p_e E\{(\tilde{x}_{n,b}^{k,m})^2\} \end{aligned}$$

- **Non-linearity problem:** How to compute first and second moments of the non-linear ET prediction $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$?
  - Note: $\tilde{\mathcal{I}}_n^b$, calculated at the decoder, is itself impacted by packet loss

Solution to the Non-linearity Problem

- The base layer interval $\tilde{\mathcal{I}}_n^b$ can be decomposed into random and deterministic parts:
  - $\tilde{\mathcal{I}}_n^b = \tilde{x}_{n,b}^{k,m} + [-\delta_1, \delta_2]$, where $[-\delta_1, \delta_2]$ is completely determined by the base layer quantization index i_n^b

Solution to the Non-linearity Problem

- The base layer interval $\tilde{\mathcal{I}}_n^b$ can be decomposed into random and deterministic parts:
  - $\tilde{\mathcal{I}}_n^b = \tilde{x}_{n,b}^{k,m} + [-\delta_1, \delta_2]$, where $[-\delta_1, \delta_2]$ is completely determined by the base layer quantization index i_n^b
- $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$ can be represented as $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$

Solution to the Non-linearity Problem

- The base layer interval $\tilde{\mathcal{I}}_n^b$ can be decomposed into random and deterministic parts:
    - $\tilde{\mathcal{I}}_n^b = \tilde{x}_{n,b}^{k,m} + [-\delta_1, \delta_2]$, where $[-\delta_1, \delta_2]$ is completely determined by the base layer quantization index i_n^b
- $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$ can be represented as $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$
- $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$ approximated by Taylor series expansion of $f_{i_n^b}(x, u)$ about $(E\{\tilde{x}_{n,b}^{k,m}\}, E\{\tilde{u}_{n,e}^{k,m}\})$, retaining only up to the second order terms

Solution to the Non-linearity Problem

- The base layer interval $\tilde{\mathcal{I}}_n^b$ can be decomposed into random and deterministic parts:
    - $\tilde{\mathcal{I}}_n^b = \tilde{x}_{n,b}^{k,m} + [-\delta_1, \delta_2]$, where $[-\delta_1, \delta_2]$ is completely determined by the base layer quantization index i_n^b
- $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$ can be represented as $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$
- $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$ approximated by Taylor series expansion of $f_{i_n^b}(x, u)$ about $(E\{\tilde{x}_{n,b}^{k,m}\}, E\{\tilde{u}_{n,e}^{k,m}\})$, retaining only up to the second order terms
- Expectations of $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$ and $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})^2$ are evaluated in terms of known moments of the arguments
    - Note: SCORE should be run in the base layer as well

Solution to the Non-linearity Problem

- The base layer interval $\tilde{\mathcal{I}}_n^b$ can be decomposed into random and deterministic parts:
  - $\tilde{\mathcal{I}}_n^b = \tilde{x}_{n,b}^{k,m} + [-\delta_1, \delta_2]$, where $[-\delta_1, \delta_2]$ is completely determined by the base layer quantization index i_n^b
- $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,e}^{k,m})$ can be represented as $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$
- $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$ approximated by Taylor series expansion of $f_{i_n^b}(x, u)$ about $(E\{\tilde{x}_{n,b}^{k,m}\}, E\{\tilde{u}_{n,e}^{k,m}\})$, retaining only up to the second order terms
- Expectations of $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})$ and $f_{i_n^b}(\tilde{x}_{n,b}^{k,m}, \tilde{u}_{n,e}^{k,m})^2$ are evaluated in terms of known moments of the arguments
  - Note: SCORE should be run in the base layer as well
- Recursions for the remaining coding modes are discussed in the paper

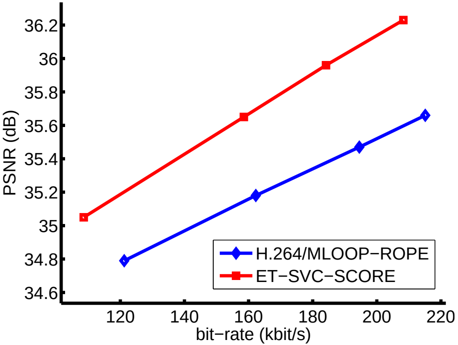
Estimation Theoretic Concealment

- Estimation theoretic prediction inspires an approach for optimal enhancement layer concealment at the decoder when the base layer is received
  - The base layer provides the interval $\tilde{\mathcal{I}}_n^b$
  - The base layer motion vector points to a motion reference in the prior enhancement layer reconstruction $\tilde{u}_{n,c}^{k,m}$
  - The optimal concealment of the transform coefficient at the enhancement layer is $f(\tilde{\mathcal{I}}_n^b, \tilde{u}_{n,c}^{k,m})$
- SCORE recursions at the encoder naturally account for usage of ET concealment at the decoder
  - Note: ET concealment is also not compatible with ROPE
- Provides an additional shot of performance

Results: The Competing Systems

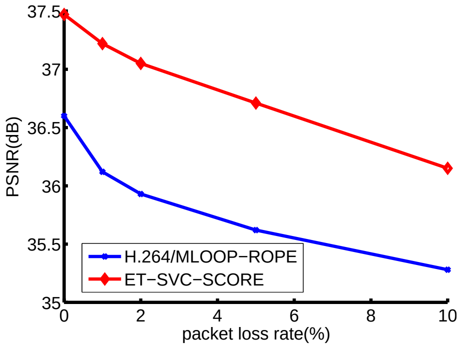
- State-of-the-art: H.264/SVC with multiloop prediction at enhancement layer, optimized via ROPE - H.264/MLOOP-ROPE
- Proposed system: ET-SVC optimized via SCORE -ET-SVC-SCORE
- Both competitors use the same base layer: H.264-ROPE
- Note: SCORE is run in parallel at the base layer but does not influence coding decisions

Enhancement Layer Decoding Quality Versus Bit-Rate



- Sequence *foreman* at *QCIF* resolution: the base layer is encoded at 128 *kbps*, and transmitted at packet loss rate 1% and the enhancement layer has a packet loss rate of 5%.
- Similar performance gains observed for other sequences

Enhancement Layer Decoding Quality Versus PLR



- Sequence *coastguard* at *QCIF* resolution: the base layer bit-rate is 170 *kbps*; the enhancement layer bit rate is 340 *kbps*
- The gain at 0% PLR is primarily due to ET-SVC
- This gain is maintained as the PLR increases due to the optimization of coding decisions via SCORE

Conclusions

- Proposed a transform-domain approach to efficient and robust scalable video coding that is a union of optimal compression via ET-SVC and accurate EED estimation via SCORE
- SCORE overcomes intricacies of transform domain EED estimation that arise due to motion compensation references frequently being off-grid
- SCORE naturally accommodates the non-linear transform-domain operations of ET-SVC via suitable approximation and the usage of ET concealment at the decoder
- The proposed unified system provides significant performance gains over a competing state-of-the-art pixel-domain SVC approach that is optimized via ROPE

Thanks