ON GENERALIZING THE ESTIMATION-THEORETIC FRAMEWORK TO SCALABLE VIDEO CODING WITH QUAD TREE STRUCTURED BLOCK PARTITIONS

Shunyao Li, Tejaswi Nanjundaswamy, Bohan Li, Kenneth Rose
Signal Compression Lab, UC Santa Barbara
We improved the enhancement layer prediction in SNR-scalable coding using the generalized ET framework, where the base layer quantizer interval information is utilized in predicting enhancement layer.

It accounts for the new coding tools in SHVC, and reaches 3.3% bitrate reduction in enhancement layer compared to standard full SHVC.
Scalable Video Coding (SVC): Encode the same video in different settings (SNR, resolution, frame rate)
Scalable Video Coding (SVC): Encode the same video in different settings (SNR, resolution, frame rate)
EL has access to both BL and EL. When predicting a block in EL, it can reference from both the same block in BL (exactly matched but lower quality) and the blocks in reference frames in EL (not exactly matched but higher quality). Choose the best (SHVC)
BACKGROUND

Other information we can get from base layer:

Motion vectors
Partitions
**Quantization intervals** (in transform domain)
ESTIMATION-THEORETIC (ET) APPROACH

\[ \varepsilon^b = x^T - \tilde{x}^b T \in (a, b) \]
\[ x^T \in (\tilde{x}^b T + a, \tilde{x}^b T + b) \]

Interval information from BL
ESTIMATION-THEORETIC (ET) APPROACH

\[ \epsilon^b = x^T - \tilde{x}^{bT} \in (a, b) \]
\[ x^T \in (\tilde{x}^{bT} + a, \tilde{x}^{bT} + b) \]
\[ f(x^T | \tilde{x}^T) = \frac{\lambda}{2} \exp(-\lambda|x^T - \tilde{x}^T|) \]
ESTIMATION-THEORETIC (ET) APPROACH

\[ \epsilon^b = x^T - \tilde{x}^{bT} \in (a, b), \]
\[ x^T \in (\tilde{x}^{bT} + a, \tilde{x}^{bT} + b). \]
\[ f(x^T \mid \tilde{x}^T) = \frac{\lambda}{2} \exp(-\lambda |x^T - \tilde{x}^T|) \]

Distribution info from EL

ET prediction:
\[ \tilde{x}_{ET}^e = E(x^T \mid x^T \in (\tilde{x}^{bT} + a, \tilde{x}^{bT} + b), \tilde{x}^e) \]

Interval information from BL
CHALLENGES FROM NEW CODING TOOLS

1) Quadtree partition:
EL partition and BL partition are not aligned
CHALLENGES FROM NEW CODING TOOLS

1) Quadtree partition:
EL partition and BL partition are not aligned

2) Rate distortion optimized quantizer (RDOQ):
Quantized indices are adjusted for better RD performance
-> erroneous interval information
CHALLENGES FROM NEW CODING TOOLS

1) Quadtree partition:
EL partition and BL partition are not aligned

2) Rate distortion optimized quantizer (RDOQ):
Quantized indices are adjusted for better RD performance
-> erroneous interval information

3) Hybrid Transform (DCT, ADST, Transform skip, etc.)
More variations in transform kernels leads to more variations in transform coefficients distribution
PROPOSED: GENERALIZED ET APPROACH

- Align the information from BL and EL
- Interval Adjustment for RDOQ
- Distribution Parameter Training for Hybrid Transform
PROPOSED: GENERALIZED ET APPROACH

- Align the information from BL and EL
- Interval Adjustment for RDOQ
- Distribution Parameter Training for Hybrid Transform
BL & EL ALIGNMENT

Intuitively, it only takes a linear transformation from one transform size to another transform size.

However:
1) The linear transform of interval information might result in very large interval or meaningless interval \((-\infty, \infty)\)
2) The linear transform of distribution information involves complicated convolution
BL & EL ALIGNMENT

Instead, we focus directly on the Expectation value

\[
E(Y) = E(A^T X A) = A^T E(X) A
\]

\[
E(X) = \tilde{x}_{ET}^{eT} = E(x^T | x^T \in (\tilde{x}^{bT} + a, \tilde{x}^{bT} + b), \tilde{x}^{eT})
\]

X: Transform domain coeff in BL TU size
Y: Pixel domain value in BL TU size
A: Inverse transform kernel
\(E(X)\): The optimal ET prediction (need to transform the EL prediction in the BL TU size to get \(\tilde{x}^{eT}\))
BL & EL ALIGNMENT

Transforming the EL prediction in the size of the BL TU involves blocks merging, extending and cropping. Depending on the mismatch of partitions:

- **Divide EL PU into small blocks; Calculate the ET prediction for each small block, then merge together**

- **Extend the EL prediction to the BL TU size using the same MV; Calculate the ET prediction then crop and copy to each small block**

- **Combination of the previous two**
PROPOSED: GENERALIZED ET APPROACH

- Align the information from BL and EL
- Interval Adjustment for RDOQ
- Distribution Parameter Training for Hybrid Transform
INTERVAL ADJUSTMENT FOR RDOQ

In RDOQ, in addition to the correct quantizer magnitude $L$, the encoder also considers $L-1$ and $0$, and choose the one with the lowest RD cost.

The encoder also has the option to set a 4x4 coeff group to 0

In skip mode, the residual of the whole block will be set to 0

\[ I_{-2} \quad I_{-1} \quad I_0 \quad I_1 \quad I_2 \]
INTERVAL ADJUSTMENT FOR RDOQ

In the generalized ET approach, assume $i$ as the coded quantizer index for one transform coefficient.

If $i=0$, set the interval as $I_{-1} \cup I_0 \cup I_1$
If $i>0$, set the interval as $I_{i-1} \cup I_i$
If $i<0$, set the interval as $I_i \cup I_{i+1}$
If $i=0$ for all the transform coefficients in the block, assume it is in skip mode, set interval as $(-\infty, \infty)$
PROPOSED: GENERALIZED ET APPROACH

- Align the information from BL and EL
- Interval Adjustment for RDOQ
- Distribution Parameter Training for Hybrid Transform
DISTRIBUTION PARAMETER TRAINING FOR HYBRID TRANSFORM

We extend the laplacian distribution assumption for DCT transform coefficients to ADST and Transform Skip (TS), and train the $\lambda$ for different transform types.

Following the maximum-likelihood estimation, with $N$ number of samples,

$$
\lambda = \frac{N}{\sum_{i=0}^{N-1} |x_i^T - \tilde{x}_i^eT|}
$$

Adaptive to QP and block size.
THE GENERALIZED ET FRAMEWORK IN SHVC

Enhancement layer ET prediction for quadtree structure
EXPERIMENTS

SHM 8.0
Two-layer SNR scalability
Lowdelay P
QP offset: -3
Use a look-up table to store the centroid offset within the interval
EXPERIMENTS

Experiment 1: the prediction gain purely from the blocks that have valid ET prediction in EL

<table>
<thead>
<tr>
<th></th>
<th>Prediction Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQMall (480p)</td>
<td>2.67 dB</td>
</tr>
<tr>
<td>BasketballDrill (480p)</td>
<td>1.69 dB</td>
</tr>
<tr>
<td>Keiba (480p)</td>
<td>3.00 dB</td>
</tr>
<tr>
<td>FourPeople (720p)</td>
<td>2.07 dB</td>
</tr>
<tr>
<td>Johnny (720p)</td>
<td>0.75 dB</td>
</tr>
<tr>
<td>Vidiyol (720p)</td>
<td>1.29 dB</td>
</tr>
<tr>
<td>Cactus (1080p)</td>
<td>2.77 dB</td>
</tr>
<tr>
<td>BasketballDrive (1080p)</td>
<td>2.89 dB</td>
</tr>
<tr>
<td>BQTerrace (1080p)</td>
<td>1.78 dB</td>
</tr>
<tr>
<td>Kimono (1080p)</td>
<td>5.05 dB</td>
</tr>
<tr>
<td>ParkScene (1080p)</td>
<td>3.20 dB</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
<td><strong>2.47 dB</strong></td>
</tr>
</tbody>
</table>

However, only 3%-15% of the blocks have valid interval information.
EXPERIMENTS

Experiment 2: the RD performance of the generalized ET framework with the full standard SHVC

Experiment 3: the RD performance of the original ET framework with the constrained SHVC (where quad tree, RDOQ, Hybrid transform are disabled)

We have effectively tackled the challenges of the new coding tools by achieving a similar gain.
SUMMARY

The generalized ET framework combines the interval information in BL and the transform coeff distribution information in EL together.

To comply with the new coding tools in SHVC (quadtree partition, RDOQ, hybrid transform), it

- aligns the information from BL and EL by exploiting the linearity of expectation
- adjusts the interval information when RDOQ is used
- trains parameters in laplacian distribution for different types of transform, QP and block size

3.3% reduction in bitrate is achieved when comparing with the full standard SHVC