Thor

High Efficiency, Moderate Complexity Video Codec using only RF IPR

draft-fuldseth-netvc-thor-00

Arild Fuldseth, Gisle Bjontegaard (Cisco)

IETF 93 – Prague, CZ – July 2015
Design principles

- Moderate complexity to allow real-time implementation in SW on common HW, as well as new HW designs
- Basic building blocks from well-known hybrid approach (motion compensated prediction and transform coding)
- Common design elements in modern codecs
  - Larger block sizes and transforms, up to 64x64
  - Quarter pixel interpolation, motion vector prediction, etc.
- Cisco RF IPR (note well: declaration filed on draft)
  - Deblocking, transforms, etc. (some also essential in H.265/4)
- Avoid non-RF IPR
  - If/when others offer RF IPR, design/performance will improve
Encoder Architecture

- **Input**: Input video
- **Transform**: Transform
- **Quantizer**: Quantizer
- **Entropy Coding**: Entropy Coding
- **Output**: Output bitstream
- **Loop filters**: Loop filters
- **Reconstructed Frame Memory**: Reconstructed Frame Memory
- **Motion Estimation**: Motion Estimation
- **Inter Frame Prediction**: Inter Frame Prediction
- **Intra Frame Prediction**: Intra Frame Prediction
- **Inverse Transform**: Inverse Transform
Decoder Architecture

Input Bitstream → Entropy Decoding → Inverse Transform

Intra Frame Prediction

Inter Frame Prediction

Loop filters

Reconstructed Frame Memory

Output video
Block Structure

- Super block (SB) 64x64
- Quad-tree split into coding blocks (CB) >= 8x8
- Multiple prediction blocks (PB) per CB
  - Intra: 1 PB per CB
  - Inter: 1, 2 (rectangular) or 4 (square) PBs per CB
- 1 or 4 transform blocks (TB) per CB
Coding-block modes

- Intra
- Inter0  MV index, no residual information
- Inter1  MV index, residual information
- Inter2  Explicit motion vector information, residual information
- Bipred  Explicit motion vector information (x2), residual information
Intra prediction

- DC
- Vertical
- Horizontal
- 5 angular modes, using filtered neighbor pixels
  - UUR: Up-Up-Right, arctan (1/2) degrees
  - UUL: Up-Up-Left, arctan (1/2) degrees
  - UL: Up-Left, 45 degrees
  - ULL: Up-Left-Left, arctan (1/2) degrees
  - DLL: Down-Left-Left, arctan (1/2) degrees
Inter prediction

- **Luma:**
  - 1/4 pixel resolution
  - 6-tap separable interpolation filter
  - Non-separable low-pass filter for (1/2,1/2) position

- **Chroma:**
  - 1/8 pixel resolution
  - 4-tap separable interpolation filter for chroma

- Multiple reference frames
Transforms

- Transforms are identical to H.265/HEVC (Cisco IPR)
  - Exception is 64x64 transform unique in Thor

- Integer-approximation to DCT

- 4x4, 8x8, 16x16, 32x32, and 64x64

- Embedded structure (inherited from exact DCT)
Deblocking

- Only 8x8 block edges
- On/off decisions per block edge and per pixel line
- No strong filter
- Input: 2 pixels on each side
- Output: 2 pixels on each side
Constrained Low-pass Filter

- In-loop filter applied after deblocking
- Fixed coefficients
- On/off switch signalled per super-block
- RDO-based on/off decision
Entropy coding

- VLC-based (non-arithmetic)

- Block-level parameters:
  Combined coding of multiple parameters

- Transform coefficient coding:
  Improvement of CAVLC scheme from HM1.0
Encoder Optimizations

• Motion estimation:
  – Fast search
  – SAD-based

• Intra mode selection:
  – SAD-based or RDO-based

• Choice of CB size and CB mode:
  – RDO-based

• Operation points:
  – high complexity, medium complexity, low complexity

• SIMD optimization for some low level functions
Planned extensions

- Re-ordering of frames
- Parallel processing tools
- Slices, tiles
Compression Performance

• Metric:
  – Bjøntegaard Delta Rate (BDR)
  – PSNR-based

• Test sequences:
  – HD format only
  – JCT-VC class B&E + 4 internal 1080p60 sequences
Compression Performance

• Anchor:
  – HM13.0 (HEVC reference software)
  – Low-delay B configuration
    (no look-ahead, no frame reordering, systematic QP variations)

• Thor:
  – Same constraints as the anchor

• VP9:
  -p 1 --cpu-used=$c --end-usage=q --cq-level=$q --lag-in-frames=0 --disable-kf

• X265:
  -l -1 --bframes 0 -F 1 --no-wpp --psnr --tune psnr -p $p --qp $q

Note: It was not possible to specify the exact same GOP structure for all encoders
## Compression Performance

### BDR at high complexity operating point

<table>
<thead>
<tr>
<th>Class</th>
<th>Sequence</th>
<th>Thor</th>
<th>VP9</th>
<th>x265</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class B</td>
<td>Kimono</td>
<td>20,3</td>
<td>21,7</td>
<td>16,2</td>
</tr>
<tr>
<td></td>
<td>ParkScene</td>
<td>27,0</td>
<td>31,4</td>
<td>21,1</td>
</tr>
<tr>
<td></td>
<td>Cactus</td>
<td>21,2</td>
<td>26,6</td>
<td>21,5</td>
</tr>
<tr>
<td></td>
<td>BasketballDrive</td>
<td>36,4</td>
<td>32,9</td>
<td>16,8</td>
</tr>
<tr>
<td></td>
<td>BQTerrace</td>
<td>48,4</td>
<td>84,1</td>
<td>52,9</td>
</tr>
<tr>
<td>Class E</td>
<td>FourPeople</td>
<td>11,5</td>
<td>35,5</td>
<td>21,5</td>
</tr>
<tr>
<td></td>
<td>Johnny</td>
<td>20,0</td>
<td>66,9</td>
<td>30,3</td>
</tr>
<tr>
<td></td>
<td>KristenAndSara</td>
<td>10,1</td>
<td>36,9</td>
<td>18,0</td>
</tr>
<tr>
<td>Internal</td>
<td>ChangeSeats</td>
<td>21,2</td>
<td>20,5</td>
<td>13,6</td>
</tr>
<tr>
<td></td>
<td>HeaAndShoulder</td>
<td>9,1</td>
<td>59,8</td>
<td>33,1</td>
</tr>
<tr>
<td></td>
<td>TelePresence</td>
<td>20,4</td>
<td>25,3</td>
<td>14,2</td>
</tr>
<tr>
<td></td>
<td>WhiteBoard</td>
<td>17,9</td>
<td>43,8</td>
<td>26,3</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>22,0</strong></td>
<td><strong>40,5</strong></td>
<td><strong>23,8</strong></td>
</tr>
</tbody>
</table>
Complexity Measurements

- Sequence: FourPeople
- QP: 32
- Single-core
Frame Rate vs. Compression

Frame rate vs. bandwidth

- Thor
- VP9
- X265

Bandwidth

Frame rate

0.0 % 10.0 % 20.0 % 30.0 % 40.0 % 50.0 % 60.0 % 70.0 % 80.0 % 90.0 % 100.0 %

0.01 0.10 1.00 10.00 100.00
Source Code

• Available at: github.com/cisco/thor