ANALYSIS OF CODING TOOLS AND IMPROVEMENT OF TEXT READABILITY FOR SCREEN CONTENT

Holger Meuel  Julia Schmidt  Marco Munderloh  Jörn Ostermann

Coding of Text in Video Coding

- Neglection of high frequencies in state-of-the-art video coding
- Lots of high frequencies contained in text
- Therefore degradation of text

Properties of Screen Content (SC)

Application scenarios
- Office applications (e.g. text/spreadsheet processing)
- Text insertions into natural video (news tickers etc.)
- Streaming services, online gaming, video conferencing

Properties of letters and symbols
- Sharp edges introduced by letters and symbols
- Translational movement during scrolling and window movement

Existing Coding Tools in AVC & their Appropriateness for Screen Content Coding (SCC)

- Data rate ratios I/P & I/B:
  - for natural video: 20−1000 (I/P), 50−2000 (I/B)
  - for screen content: 500−100000 (I/P), 500−1000000 (I/B)
- Smaller movement in sequence ⇒ higher ratio

- Distance of Ref. Frames:
  - SC sequences typically contain very slow movements
  - Spreading reference slices over time as wide as possible is beneficial for high coding efficiency

- Hierarch. B slices:
  - Little difference between frames for slowly changing content
  - No additional information in hierarchical B slices
  - Recommendation to dismiss reference B Slices for SCC

- Number of B slices:
  - Larger temporal distance between reference slices
  - Increasing of residuals of P slices
  - Optimal results with 3 B Slices
  - Disable B slices completely for small movements

- Adaptive Quantisation Parameters (QPs):
  - QP changes are expensive
  - Fixed QP coding often better

- Resolution of Motion Vectors (MV):
  - Same as for natural video

- Spatial and Temporal Direct Mode (DM):
  - Use of Spatial/Temporal Direct Mode stays same in SC/camera captured sequences
  - 95−98% of DM coded blocks are better coded spatially

Improvement of Text Readability

- Requirements for text detection in SCC:
  - Runtime efficient separation of text/background
  - High detection rate ⇒ Canny Edge Detector
  - No adequate quality evaluation possible for SCC with PSNR (>45dB) ⇒ Subjective evaluation

- Advantage: Usage of standard coder with externally provided QP map
- Experimental results on the right

Data rate ratios I/P & I/B:

<table>
<thead>
<tr>
<th>Data rate ratios I/P &amp; I/B</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>20−1000 (I/P)</td>
<td>500−100000</td>
<td></td>
</tr>
<tr>
<td>50−2000 (I/B)</td>
<td>500−1000000</td>
<td></td>
</tr>
</tbody>
</table>

Distance of Ref. Frames:

<table>
<thead>
<tr>
<th>Distance of Ref. Frames</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typically contain very slow movements</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

Hierarch. B slices:

<table>
<thead>
<tr>
<th>Hierarch. B slices</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little difference between frames for slowly changing content</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

Number of B slices:

<table>
<thead>
<tr>
<th>Number of B slices</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larger temporal distance between reference slices</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

Adaptive Quantisation Parameters (QPs):

<table>
<thead>
<tr>
<th>Adaptive Quantisation Parameters (QPs)</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>QP changes are expensive</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

Resolution of Motion Vectors (MV):

<table>
<thead>
<tr>
<th>Resolution of Motion Vectors (MV)</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same as for natural video</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

Spatial and Temporal Direct Mode (DM):

<table>
<thead>
<tr>
<th>Spatial and Temporal Direct Mode (DM)</th>
<th>Natural Video</th>
<th>Screen Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Spatial/Temporal Direct Mode stays same</td>
<td>50−2000 (I/B)</td>
<td>500−100000 (I/P)</td>
</tr>
</tbody>
</table>

95−98% of DM coded blocks are better coded spatially.