EXPLOITING MPEG-7 TEXTURE DESCRIPTORS FOR FAST H.264 MODE DECISION

N. Kamnoonwatana, D. Agrafiotis, C. N. Canagarajah
Talk Overview

- MPEG-7 and Video Coding
- Mode Observations
- Proposed Fast mode Algorithm
- Experimental results
- Conclusions
MPEG-7 and Video Coding

• MPEG-7
  • Describes audio-visual content.
  • Applications: storage, browsing, etc.
• MPEG-7 and Video Coding?
  • What if these indexing data are available alongside the video content?
  • Previous work: improve coding efficiency.
  • Proposed: reduce coding time (fast mode decision)
Mode Decision Observations (1)

- Macroblocks in the same homogeneous region are likely to share the same or similar modes.
- This behaviour persists for a period of time (e.g. one shot).
- Therefore:
  - If these homogenous regions are known…
  - the range of possible coding modes can be reduced by making use of prior information regarding the coding mode of already coded MBs of the same region in the current and previous frame.
Mode Decision Observations (2)

Flower Garden Sequence

Stefan Sequence

- Skip
- Inter
- Intra4x4
- Intra16x16
Mode Decision Observations (3)

Coding modes for MBs in the sky region of *Flower Garden* sequence
Homogeneous Region Clustering (1)

- To form homogeneous clusters a criterion of homogeneity must be used (e.g. intensity, colour, texture)
- We use the MPEG-7 Homogeneous Texture Descriptor (HTD).
- In each frame the HTD is calculated for every MB.
- These HTD vectors are then used to form homogeneous texture clusters by a clustering algorithm.
Homogeneous Regions Clustering (2)

Note that for both algorithms the number of clusters was limited to 20.
Proposed Fast Mode Decision (1)

• The aim:
  “To narrow down the choices of coding modes for each MB based on the information coming from other coded MBs that belong to the same homogeneous cluster”

• The information used is:
  • Coded modes of the MBs that belong to the same homogeneous cluster in the same and previous frames.
Proposed Fast Mode Decision (2)

Current frame

\( \text{Mode}_{\text{cluster}} : \text{Coding mode of 1st MB in the cluster} \)

Available modes for current MB

Texture similarity criterion

Information from previous frame

Start new
Proposed Fast Mode Decision (3)

**Probability Decision Block**

IN

Mode probability based on previous frames

Available modes for current MB
Proposed Fast Mode Decision (4)

• The proposed algorithm can be integrated with the existing fast mode decision methods used in JM12.2 namely Early SKIP detection and Selective Intra mode decision.

• It can be placed right after the early SKIP mode testing. If the conditions for the SKIP mode are met then this algorithm will not be executed.
Experimental Results (1)

- JM12.2 reference H.264 software
- The MV search range was set to 16 with one reference frame
- GOP structure of IPPP...I, with one I frame every 30 frames.
- The motion estimation scheme was set to fast full search.
- Four quantisation parameters: 25, 30, 35, and 40.
- The hierarchical clustering method was used with a Euclidean distance metric and with the number of clusters set to 40.
- Results are presented in the form of bit rate, PSNR, and encoding time differences ($\Delta$Rate, $\Delta$Y-PSNR, and $\Delta$Time respectively)

$$\Delta Time(\%) = \left( \frac{Time_{proposed} - Time_{reference}}{Time_{reference}} \right) \times 100$$
### Experimental Results (2)

**Comparison with exhaustive mode decision**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$\Delta$Rate(%)</th>
<th>$\Delta$Y-PSNR(dB)</th>
<th>$\Delta$Time(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar</td>
<td>1.118</td>
<td>-0.051</td>
<td>-46.5042</td>
</tr>
<tr>
<td>Silent</td>
<td>5.616</td>
<td>-0.219</td>
<td>-49.5397</td>
</tr>
<tr>
<td>Stefan</td>
<td>3.501</td>
<td>-0.161</td>
<td>-44.3363</td>
</tr>
<tr>
<td>Container</td>
<td>0.843</td>
<td>-0.029</td>
<td>-47.8207</td>
</tr>
<tr>
<td>Table tennis</td>
<td>3.019</td>
<td>-0.104</td>
<td>-46.3773</td>
</tr>
<tr>
<td>Flowergarden</td>
<td>1.388</td>
<td>-0.076</td>
<td>-50.2867</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>2.58</strong></td>
<td><strong>-0.11</strong></td>
<td><strong>-47.48</strong></td>
</tr>
</tbody>
</table>
Experimental Results (3)
Comparison with the JM12.2 fast mode decision

<table>
<thead>
<tr>
<th>Sequence</th>
<th>ΔRate(%)</th>
<th>ΔY-PSNR(dB)</th>
<th>ΔTime(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar</td>
<td>1.049</td>
<td>-0.05</td>
<td>-33.5376</td>
</tr>
<tr>
<td>Silent</td>
<td>4.594</td>
<td>-0.189</td>
<td>-14.5673</td>
</tr>
<tr>
<td>Stefan</td>
<td>2.383</td>
<td>-0.11</td>
<td>-20.5754</td>
</tr>
<tr>
<td>Container</td>
<td>0.605</td>
<td>-0.021</td>
<td>-9.4811</td>
</tr>
<tr>
<td>Table tennis</td>
<td>2.027</td>
<td>-0.071</td>
<td>-19.1694</td>
</tr>
<tr>
<td>Flowergarden</td>
<td>1.174</td>
<td>-0.066</td>
<td>-31.0334</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>1.97</strong></td>
<td><strong>-0.08</strong></td>
<td><strong>-21.39</strong></td>
</tr>
</tbody>
</table>
Experimental Results (4)

- Works very well in sequences where the assumptions are true – “there are well defined homogeneous texture regions with MBs that undergo similar motion”

- Different number of clusters?
  - Larger number of clusters can lead to a smaller reduction in coding time with a lower penalty in rate-distortion performance.
  - Vice versa for smaller number of clusters.
  - For example, considering comparison with the exhaustive mode decision,
    - Number of clusters is increased from 40 to 80
    - An average of 36.82% decrease in coding time is achieved (instead of 47.48%)
    - The average increase in bit rate reduced to 1.8% and the PSNR penalty reduced to 0.07 dB
Conclusions

• We have presented a fast mode decision algorithm that successfully exploits MPEG-7 metadata in the context of H.264 video coding.

• The MPEG-7 Homogenous Texture descriptor is used to form homogeneous clusters in every frame.

• The algorithm exploits the coding mode decision similarities that can be observed in these clusters so that the possible encoding modes for each MB are limited.

• As a result the proposed method reduces the encoding time significantly without introducing any noticeable rate-distortion penalty.