Comparison of Content Selection Methods for Skimming Rushes Video

Werner Bailer, Georg Thallinger
TRECVID Video Summarization Workshop @ ACM MM, 2008-10-31
Outline

- Summary creation process
- Content selection
  - a closer look to the problem
  - rule-based approach
  - HMM based approach
- Results and comparison
- Conclusion
Summary Creation Process

1. Content Analysis
2. Remove Unusable Content
3. Retake Detection
4. Representative Clip Selection
5. Editing

Legend:
- Content selection
- Content clustering
- Pre/Post processing
Summary Creation Process

Content Analysis -> Remove Unusable Content -> Retake Detection -> Representative Clip Selection -> Editing

- Content selection
- Content clustering
- Pre/Post processing

our focus in TRECVID 2007
Summary Creation Process

Content Analysis

Remove Unusable Content

Representative Clip Selection

Retake Detection

Editing

Content selection
Content clustering
Pre/Post processing

our focus in TRECVID 2008
Content Analysis & Junk Content Removal

- Shot boundary detection (hard cuts)
  - frame differences, SVM classifier trained on TRECVID 2006 data
- MPEG-7 Color Layout and EdgeHistogram
  - descriptors extracted from every 10th frame
- Visual activity
  - averaged over 10 frames
- Face detection
  - Viola/Jones, OpenCV implementation
- Junk content removal
  - skip short shots: duration < 10 frames
  - remove color bars and monochrome frames: standard deviation in columns < 15 intensity levels in each channel
Repeated Take Detection

- Take of same scene, from same camera
- Split shots into parts (subshots)
- Pair-wise matching of parts
  - match extracted colour, texture and visual activity descriptor sequences of the parts (temporally sub-sampled by 10)
  - modified Longest Common Subsequence (LCSS) algorithm
  - remove contained and overlapping matches
  - result is a similarity matrix of the take candidates
- Cluster take candidates
- Determine relevance
  - based on overlap with takes in the same cluster
Representative Clip Selection

- Content selection problem for BBC rushes 2007 test data
  - values based on ground truth provided by NHK

- Relevant content
  - mean 38.02% (min. 11.13%, max. 87.75%)
  - all "meaningful" content

- Non-redundant content
  - use longest take of all takes of a scene
  - mean 15.20%

- Summarization goal of 2% requires
  - discarding ~87% of non-redundant content
  - or using 7.6x acceleration
Input to Content Selection

- List of arbitrary segments
  - relevance value
  - redundancy information
    - absolute: probability that this segment is useless
    - relative: list of segments w.r.t. which the current segment is redundant, and a similarity value for each of these segments

- In our experiments
  - retakes: relative redundancy information + similarity values
  - junk content: absolute redundancy information
  - motion activity: selected segments with relevance
  - presence of faces: selected segments with relevance
Two Approaches to Content Selection

- **Rule-based approach**
  - merge relevant and redundant segment lists into one relevance function over time
  - adaptive thresholding yields list of segments (takes length constraint into account)
  - optimize by removing/adding parts of segments

- **HMM based**
  - vector of relevance/redundancy values for each time instant
  - selected/not-selected etc. are hidden states
  - training
    - extract relevance/redundancy vector sequences from test set
    - create state sequence from ground truth
  - content selection
    - find ML path for given sequence of relevance/redundancy vectors
HMM Based Approach

- 6 states
  - non-relevant (Npre)
  - relevant (Rpre)
  - selected (S)
  - scene boundary (B)
  - non-relevant (Npost)
  - relevant (Rpost)

- Parameter $\lambda$ in state transition matrix
  - control number and length of selected segments

- Limitations
  - not possible to enforce length constraint
  - junk content not deterministically excluded
Approaches to Content Selection - Overview

- **Selection Method**
  - Rule-based
  - HMM based

- **Redundant and Selected Segments**

- **Relative to Absolute Redundancy (if necessary)**
- **Weighted Sum of Relevance Redundancy Time Lines**

- **Estimate Hidden State Sequence**
  - HMM Parameters
  - One Selected Segments List

- **Estimate Threshold**
  - Until Target Length Matched

- **Apply Threshold**
  - If below minimum length (optional check)
    - Yes: Add parts of the segments just above threshold
    - No: Final Selected Segments List
## Results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>JRS1 rule</th>
<th>JRS2 HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>min. segment length</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>max. segment length</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>min. segment distance</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>$w_{rel}=w_{red}$</td>
<td>0.5</td>
<td>n/a</td>
</tr>
<tr>
<td>max. total length</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>min. total length</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>rel. to abs. redundancy</td>
<td>longest</td>
<td>n/a</td>
</tr>
<tr>
<td>split long segments</td>
<td>true</td>
<td>true</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th></th>
<th>JRS1 median</th>
<th>(JRS1 median)</th>
<th>JRS2 median</th>
<th>(JRS2 median)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>18.50</td>
<td>(0.15)</td>
<td>14.00</td>
<td>(0.03)</td>
</tr>
<tr>
<td>XD</td>
<td>13.38</td>
<td>(0.24)</td>
<td>14.20</td>
<td>(0.22)</td>
</tr>
<tr>
<td>TT</td>
<td>25.33</td>
<td>(0.09)</td>
<td>26.67</td>
<td>(0.13)</td>
</tr>
<tr>
<td>VT</td>
<td>20.00</td>
<td>(0.05)</td>
<td>18.33</td>
<td>(0.00)</td>
</tr>
<tr>
<td>IN (median)</td>
<td>0.22</td>
<td>(0.19)</td>
<td>0.28</td>
<td>(0.27)</td>
</tr>
<tr>
<td>IN (min)</td>
<td>0.00</td>
<td></td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>IN (max)</td>
<td>0.53</td>
<td></td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>JU (median)</td>
<td>3.67</td>
<td>(1.00)</td>
<td>3.00</td>
<td>(0.50)</td>
</tr>
<tr>
<td>RE (median)</td>
<td>4.00</td>
<td>(1.00)</td>
<td>4.00</td>
<td>(1.00)</td>
</tr>
<tr>
<td>TE (median)</td>
<td>3.33</td>
<td>(1.00)</td>
<td>2.33</td>
<td>(0.50)</td>
</tr>
</tbody>
</table>
Results - MS221050

rule

HMM
Results - MS221050

<table>
<thead>
<tr>
<th></th>
<th>JRS1 (rule)</th>
<th>JRS2 (HMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>14.00</td>
<td>4.20</td>
</tr>
<tr>
<td>XD</td>
<td>6.19</td>
<td>15.99</td>
</tr>
<tr>
<td>TT</td>
<td>17.33</td>
<td>21.00</td>
</tr>
<tr>
<td>VT</td>
<td>16.33</td>
<td>7.67</td>
</tr>
<tr>
<td>IN</td>
<td>0.28</td>
<td>0.61</td>
</tr>
<tr>
<td>JU</td>
<td>4.33</td>
<td>3.33</td>
</tr>
<tr>
<td>RE</td>
<td>4.33</td>
<td>4.00</td>
</tr>
<tr>
<td>TE</td>
<td>2.67</td>
<td>1.67</td>
</tr>
</tbody>
</table>

below/exactly/above median of this run
# Results - MS221050

<table>
<thead>
<tr>
<th></th>
<th>JRS1 (rule)</th>
<th>JRS2 (HMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>14.00</td>
<td>4.20</td>
</tr>
<tr>
<td>XD</td>
<td>6.19</td>
<td>15.99</td>
</tr>
<tr>
<td>TT</td>
<td>17.33</td>
<td>21.00</td>
</tr>
<tr>
<td>VT</td>
<td>16.33</td>
<td>7.67</td>
</tr>
<tr>
<td>IN</td>
<td>0.28</td>
<td>0.61</td>
</tr>
<tr>
<td>JU</td>
<td>4.33</td>
<td>3.33</td>
</tr>
<tr>
<td>RE</td>
<td>4.33</td>
<td>4.00</td>
</tr>
<tr>
<td>TE</td>
<td>2.67</td>
<td>1.67</td>
</tr>
</tbody>
</table>

*below|exactly|above* median of all runs on this video
Results – Comparison

- Both runs yield short summaries, well below the 2% limit
  - the rule based: 58.00% of max. length, 1.20% of original content
  - HMM run: 49.65% of max. length, 0.99% of original content
- HMM based selected method
  - 6% higher inclusion (increase of 27%)
  - duration is 24% shorter
  - lower score for pleasant timing
  - lower score for junk (not causally related to shorter duration or higher inclusion)
  - 47% higher editing time (more and shorter segments)
  - estimation of ML state sequence takes on average 4.75 sec/video
  - evaluation against NHK ground truth supports the results (precision and recall in the range 0.3-0.35)
Conclusion

- Comparison of two methods for content selection
- Both parametrized to yield quite short summaries
  - high scores for pleasant tempo, repeated content and junk
  - low inclusion score
- Comparison
  - HMM slightly higher inclusion at shorter duration
  - HMM difficult to control (junk, length constraint)