SESAME MER

2013 TRECVID Meeting

Bob Bolles
November 21, 2013
Outline

• MER Demonstration
• MED Analysis
• MER Analysis
• Observations and Future Work
MER Demonstration – An Example

Event Search Results

Searching for event E006 (Birthday Party) in video data set MEDITest from event kit 100ex

New search

<table>
<thead>
<tr>
<th>Rank</th>
<th>Video</th>
<th>Observations</th>
<th>Importance</th>
<th>Confidence</th>
<th>Type</th>
<th>Video ID</th>
<th>Event Detection Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0006</td>
<td>Birthday Party for Linda (0:07-0:08)</td>
<td>0.31</td>
<td>0.97</td>
<td>Video, OCR</td>
<td>192865</td>
<td>0.759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female_Person, Person_clapping, Boy, Male_Person (0:23-0:32)</td>
<td>0.58</td>
<td>0.68</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0031</td>
<td>Birthday Party in PJ (0:01-0:05)</td>
<td>0.31</td>
<td>0.88</td>
<td>Video, OCR</td>
<td>736791</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female_Person, Person_clapping, Person_blowing_candles, Male_Person (2:54-3:04)</td>
<td>0.67</td>
<td>0.67</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Person_blowing_candles, Boy, Male_Person (2:13-2:17)</td>
<td>0.78</td>
<td>0.78</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Person_clapping, Male_Person, Boy (2:20-2:23)</td>
<td>0.8</td>
<td>0.8</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Person_clapping, Boy, Male_Person (2:55-2:58)</td>
<td>0.81</td>
<td>0.81</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Person_blowing_candles, Male_Person, Female_Person (3:48-3:52)</td>
<td>0.82</td>
<td>0.82</td>
<td>Visual Concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Mum’s 80th, birthday dinner 27</td>
<td>0.29</td>
<td>0.89</td>
<td>Video</td>
<td>759100</td>
<td>0.626</td>
</tr>
</tbody>
</table>
MED Analysis

Eight Feature- and Concept-based Classifiers
- Visual: 3 classifiers using 1,346 semantic concepts
  - Concepts-HIK (color histogram analysis)
  - Concepts-DC (static image Difference Coding)
  - SIFT-Fisher (Fisher encoding of differences)
- Motion: 2 classifiers
  - DTFV (Dense Trajectory Fisher Vectors) and MoSIFT
  - Action Concept HMMFV (96 Sarnoff/UCF actions and UCF 101 actions)
- Audio: 2 classifiers
  - MFCCs (low-level audio features)
  - ASR (Automatic Speech Recognition)
- Optical Character Recognition (OCR): 1 classifier

Fusion
- Late fusion of the eight results, based on arithmetic mean

Threshold Selection
- Threshold picked to maximize $R_0$ on a held-out set of data
# 2013 MED Results

## Pre-specified Event Performance

<table>
<thead>
<tr>
<th></th>
<th>Visual + Motion</th>
<th>Audio</th>
<th>ASR</th>
<th>OCR</th>
<th>FullSys</th>
</tr>
</thead>
<tbody>
<tr>
<td>100Ex</td>
<td>26.1%</td>
<td>5.9%</td>
<td>4.0%</td>
<td>0.2%</td>
<td>27.6%</td>
</tr>
<tr>
<td>10Ex</td>
<td>11.6%</td>
<td>2.6%</td>
<td>1.4%</td>
<td>0.2%</td>
<td>10.3%</td>
</tr>
<tr>
<td>0Ex</td>
<td>1.3%</td>
<td>1.7%</td>
<td>2.3%</td>
<td>2.4%</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

## Ad-hoc Event Performance

<table>
<thead>
<tr>
<th></th>
<th>Visual + Motion</th>
<th>Audio</th>
<th>ASR</th>
<th>OCR</th>
<th>FullSys</th>
</tr>
</thead>
<tbody>
<tr>
<td>100Ex</td>
<td>23.2%</td>
<td>5.6%</td>
<td>3.9%</td>
<td>0.2%</td>
<td>25.7%</td>
</tr>
<tr>
<td>10Ex</td>
<td>12.9%</td>
<td>2.7%</td>
<td>1.4%</td>
<td>0.2%</td>
<td>12.2%</td>
</tr>
<tr>
<td>0Ex</td>
<td>1.3%</td>
<td>2.2%</td>
<td>2.2%</td>
<td>2.8%</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

1. Our ad hoc performance is essentially the same as pre-specified
2. The visual and motion concepts dominate
3. Our OCR approach for 0Ex was better than our training-based technique
MER Analysis

High-level approach
- Each modality (visual, ASR, and OCR) generates a list of their top candidates
- Visual concepts: learn to detect the most discriminative video segments, and then select the most relevant concepts for the event in those segments
- Select a small set of concepts to include in the final list
- Sort (and present) the final list according to their times of occurrence in the video

Used the following to make the final selections
- “Importance” scores, set at training time
- “Confidences” produced by each detector at run time
- Keyword matching of extracted ASR & OCR text to event-specific lists
MER Analysis

Detect Event Positives

Extract Video Content

- Visual/Motion Concepts
- Spoken Lang. Text (ASR)
- Visual Text (VOCR)
- Low-level Features

Video

Select relevant video intervals

Select best concepts per interval

Select best spoken words

Select best visual words

Filter by keyword matching

Event ("Making a Sandwich," "Parkour," ...)

Event

Select maximum of N elements per modality

Sort the elements by time of occurrence

1. Select maximum of N elements per modality

2. Sort the elements by time of occurrence

Fuse the Recounting Results

MER
MER Results

Accuracy of Judge’s final decision: 64.1%
Judge’s evaluation of tag quality: 2.53
Percent recounting review time: 41.83%

SESAME achieved the best tag quality
Observations About Our MER Analysis

• Strategy of identifying key video segments, and then identifying key event-related concepts in those segments worked well

• MER contents
  – Visual concepts in 94% of the videos
  – ASR in 15%
  – OCR in 4%.

• Our filters on ASR and OCR were too strong (They eliminated ASR results from 50% of the videos and OCR results from 35%).

• For 10Ex and 0Ex, we relied more on substring matching to keyword lists than on importance scores for ASR & OCR
Future Work

• Merge overlapping and/or adjacent intervals
• Enhance the process that computes the importance of extracted concepts at training time
• Develop better normalization of importance scores across visual, action, ASR, and OCR
• Enhance the algorithm for automatically generating event-related keywords and their importance scores
Acknowledgement

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC0067. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes nonwithstanding any copyright annotation thereon.

Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.