TRECVID-2006: Search Task

Alan Smeaton
Dublin City University
&
Tzveta Ianeva
NIST
Search Task Definition

- Goal: promote progress in content-based retrieval from digital video via open, metrics-based evaluation;
- Given a test collection, a topic and a common shot boundary reference, return a ranked list of at most 1,000 shots which best satisfy the need;
- NIST created more topics asking for general (vs. specific)
- NIST created 10 of 24 topics to ask for video of events – encouraging exploration beyond one-keyframe-per-shot
- Videos were viewed by NIST personnel, notes taken on content, and candidates emerging were chosen;
Search Task Definition

- **Per-search** measures: average precision, elapsed time
- **Per-run** measure: mean average precision (MAP)
- Interactive search participants were asked to have their subjects complete pre, post-topic and post-search questionnaires;
- Each result for a topic can come from only 1 user search; same searcher does not need to be used for all topics.
Search Task Definition

- Bing Xiang, John Makhoul, and Ralph Weischedel at BBN for providing MT/ASR
- Christian Petersohn (Fraunhofer Institute) for master shot reference
- DCU team for formatting and selecting keyframes
- MediaMill team for 101 features baseline results donation
- CMU and IBM for 449 LSCOM features annotations
Data characteristics

- TRECVID 2006 data is again (deliberately) text-noisy with video from English language, Arabic & Chinese broadcasts;
- 32.2% of the test video comes from programs not represented in the development data;
- Text is derived from speech recognition and then machine translation, thus poorer quality than with English-only sources but ASR/MT from “state-of-the-art” GALE system.
## 2006: Search task participants (26, up from 20)

<table>
<thead>
<tr>
<th>Institution</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T Labs – Research</td>
<td>USA</td>
</tr>
<tr>
<td>Beijing Jiaotong U.</td>
<td>China</td>
</tr>
<tr>
<td>Bilkent U.</td>
<td>Turkey</td>
</tr>
<tr>
<td>Carnegie Mellon U.</td>
<td>USA</td>
</tr>
<tr>
<td>Chinese U. of Hong Kong</td>
<td>China</td>
</tr>
<tr>
<td>City University of Hong Kong</td>
<td>China</td>
</tr>
<tr>
<td>CLIPS–IMAG</td>
<td>France</td>
</tr>
<tr>
<td>Columbia U.</td>
<td>USA</td>
</tr>
<tr>
<td>Dublin City U.</td>
<td>Ireland</td>
</tr>
<tr>
<td>Fudan U.</td>
<td>China</td>
</tr>
<tr>
<td>FX Palo Alto Laboratory Inc</td>
<td>USA</td>
</tr>
<tr>
<td>Helsinki U. of Technology</td>
<td>Finland</td>
</tr>
<tr>
<td>IBM T. J. Watson Research Center</td>
<td>USA</td>
</tr>
<tr>
<td>Imperial College London / Johns Hopkins U.</td>
<td>UK, USA</td>
</tr>
</tbody>
</table>
## 2006: Search task participants (continued)

<table>
<thead>
<tr>
<th>Institution</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS / I2R</td>
<td>Singapore</td>
</tr>
<tr>
<td>Mediamill / U. of Amsterdam</td>
<td>Netherlands</td>
</tr>
<tr>
<td>RMIT U. School of CS&amp;IT</td>
<td>Australia</td>
</tr>
<tr>
<td>Tsinghua U.</td>
<td>China</td>
</tr>
<tr>
<td>U. of Central Florida</td>
<td>USA</td>
</tr>
<tr>
<td>U. of Glasgow / U. of Sheffield</td>
<td>UK</td>
</tr>
<tr>
<td>U. of Iowa</td>
<td>USA</td>
</tr>
<tr>
<td>U. of Oxford</td>
<td>UK</td>
</tr>
<tr>
<td>U. Rey Juan Carlos</td>
<td>Spain</td>
</tr>
<tr>
<td>Zhejiang U.</td>
<td>China</td>
</tr>
</tbody>
</table>

**COST292** ([www.cost292.org](http://www.cost292.org))  France, Netherlands, UK, Ireland, Greece, Turkey, Serbia and Montenegro, Slovakia

**K-Space** ([kspace.qmul.net](http://kspace.qmul.net))  UK, Germany, Austria, Greece, Ireland, Netherlands, France, Switzerland, Czechia
Search Types: Automatic, Manual and Interactive

**AUTOMATIC:**
- **TOPIC** → **SYSTEM** → **RESULT**
- System takes topic as input and produces result without any human intervention

**MANUAL:**
- **TOPIC** → **HUMAN** → **QUERY** → **SYSTEM** → **RESULT**
- Human formulates query based on topic and query interface, not on knowledge of collection or search results
- System takes query as input and produces result without further human intervention

**INTERACTIVE:**
- **TOPIC** → **HUMAN** → **QUERY** → **SYSTEM** → **RESULT**
- Human (re)formulates query based on topic, query, and/or results
- System takes query as input and produces result without further human intervention on this invocation

**Number of runs:**
- 76 automatic
- 11 manually assisted
- 36 interactive
Everybody likes to search automatically, dislikes manually
24 Topics [ number of image, video examples and relevant found]

173. Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible [3, 4, 142]

174. Find shots with one or more people leaving or entering a vehicle [0, 10, 675]

175. Find shots with one or more soldiers, police, or guards escorting a prisoner [0, 4, 204]

176. Find shots of a daytime demonstration or protest with at least part of one building visible [4, 4, 111]

177. Find shots of US Vice President Dick Cheney [3, 3, 393]

178. Find shots of Saddam Hussein with at least one other person's face at least partially visible [8, 0, 99]

179. Find shots of multiple people in uniform and in formation [3, 5, 191]

180. Find shots of US President George W. Bush, Jr. walking [0, 5, 197]
24 Topics  [number of image, video examples and relevant found]

181. Find shots of one or more soldiers or police with one or more weapons and military vehicles  [2, 6, 128]
182. Find shots of water with one or more boats or ships [3, 5, 307]
183. Find shots with one or more emergency vehicles in motion (e.g., ambulance, police car, fire truck, etc.) [0, 4, 299]
184. Find shots of one or more people seated at a computer with display visible [3, 4, 440]
185. Find shots of one or more people reading a newspaper [3, 4, 201]
186. Find shots of a natural scene - with, for example, fields, trees, sky, lake, mountain, rocks, rivers, beach, ocean, grass, sunset, waterfall, animals, or people; but no buildings, no roads, no vehicles [2, 4, 523]
187. Find shots of one or more helicopters in flight [0, 6, 119]
24 Topics [ number of image, video examples and relevant found]

188. Find shots of something burning with flames visible [3, 5, 375]

189. Find shots of a group including at least four people dressed in suits, seated, and with at least one flag [3, 5, 446]

190. Find shots of at least one person and at least 10 books [3, 5, 295]

191. Find shots containing at least one adult person and at least one child [3, 6, 775]

192. Find shots of a greeting by at least one kiss on the cheek [0, 5, 98]

193. Find shots of one or more smokestacks, chimneys, or cooling towers with smoke or vapor coming out [3, 2, 60]

194. Find shots of Condoleezza Rice [3, 7, 122]

195. Find shots of one or more soccer goalposts [3, 4, 333]

196. Find shots of scenes with snow [3, 6, 692]
Some statistics

- **2006:**
  - Number of shots in test collection: 79,484
  - ~9.1% relevant shots found: 7,225

- **2005**
  - Number of shots in test collection: 45,765
  - ~18.3% relevant shots found: 8,395

- **2004**
  - Number of shots in test collection: 33,367
  - ~5.4% relevant shots found: 1,800

- **2003**
  - Number of shots in test collection: 32,318
  - ~6.5% relevant shots found: 2,114
2006: 20 sites contributed one or more unique, relevant shots
2006: Rel shots contrib. uniquely per topic by team

186, 191, 196 have 500+
2006: Most rel shots uniquely returned by topic & team

186, 191, 196 have 500+
2006: Most rel shots uniquely returned by topic & team

186 have 500+
2006: Most rel shots uniquely returned by topic & team

191, 196 have 500+
Unique relevant shots return by Oxford U. for Topic 191 ("adult and child")
2006: Automatic runs - top 10 MAP (of 76)

(mean elapsed time (mins) / topic)
2005: Automatic runs - top 10 MAP (of 42)
(mean elapsed time (mins) / topic)
Significant differences among top 8 automatic runs (using randomization test, p < 0.05)

<table>
<thead>
<tr>
<th>Run name</th>
<th>(MAP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>* A_2_TJW_Qclass_4</td>
<td>(0.087)</td>
</tr>
<tr>
<td>A_2_TJW_Qcomp_2</td>
<td>(0.086)</td>
</tr>
<tr>
<td>A_2_CMU_Taste_5</td>
<td>(0.079)</td>
</tr>
<tr>
<td>A_2_TJW_Qind_5</td>
<td>(0.076)</td>
</tr>
<tr>
<td>B_2_i2Rnus_1</td>
<td>(0.075)</td>
</tr>
<tr>
<td>&gt; B_2_i2Rnus_2</td>
<td>(0.067)</td>
</tr>
<tr>
<td>&gt; B_2_COLUMBIA_RR9_storyqibteviscon_1</td>
<td>(0.060)</td>
</tr>
<tr>
<td>&gt; B_2_COLUMBIA_RR8_textibviscon</td>
<td>(0.056)</td>
</tr>
<tr>
<td>A_2_TJW_Qclass_4</td>
<td>B_2_COLUMBIA_RR9_storyqibteviscon_1</td>
</tr>
<tr>
<td></td>
<td>B_2_COLUMBIA_RR8_textibviscon</td>
</tr>
<tr>
<td></td>
<td>B_2_i2Rnus_2</td>
</tr>
<tr>
<td>A_2_TJW_Qcomp_2</td>
<td>B_2_i2Rnus_2</td>
</tr>
<tr>
<td></td>
<td>B_2_COLUMBIA_RR9_storyqibteviscon_1</td>
</tr>
<tr>
<td></td>
<td>B_2_COLUMBIA_RR8_textibviscon</td>
</tr>
<tr>
<td>A_2_CMU_Taste_5</td>
<td>B_2_i2Rnus_1</td>
</tr>
<tr>
<td></td>
<td>B_2_COLUMBIA_RR9_storyqibteviscon_1</td>
</tr>
<tr>
<td></td>
<td>B_2_COLUMBIA_RR8_textibviscon</td>
</tr>
</tbody>
</table>
2006: Manual runs - top 10 MAP (of 11)
(mean human effort (mins) / topic)
2005: Manual runs - top 10 MAP (of 26)
(mean human effort (mins) / topic)

Recall

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Precision

M_A_2_CMU.Manu.ExpECA.QC04CR.PU_5 (15)
M_A_2_CMU.Manu.ExpE.QC05U_7 (15)
M_A_2_PicSOM-M3_2 (0.93)
M_A_2_FD_MM_BC_1 (11.1)
M_A_2_OUMT_M7TE_7 (5.06)
M_A_2_OUMT_M6TS_6 (5.02)
M_A_2_PicSOM-M2_4 (0.87)
M_A_2_FD_AOH_LR_ONLINE_3 (11.1)
M_A_1_OUMT_M5T_5 (5.01)
M_A_1_dcu_manual_text_img_6 (3)
2006: Interactive runs - top 10 MAP (of 36)

(mean elapsed time for all == ~15 mins/topic)
2005: Interactive runs - top 10 MAP (of 44)

(mean elapsed time for all == ~15 mins/topic)
Significant differences among top 8 interactive runs (using randomization test, p < 0.05)

<table>
<thead>
<tr>
<th>Run name</th>
<th>(MAP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_2_CMU_See_1</td>
<td>0.303</td>
</tr>
<tr>
<td>B_2_UvA-MM_1</td>
<td>0.267</td>
</tr>
<tr>
<td>A_2_CMU_Hear_2</td>
<td>0.226</td>
</tr>
<tr>
<td>A_2_UCFVISION_1</td>
<td>0.225</td>
</tr>
<tr>
<td>A_2_CMU_ESP_3</td>
<td>0.216</td>
</tr>
<tr>
<td>B_2_UvA-MM_2</td>
<td>0.212</td>
</tr>
<tr>
<td>B_1_FXPAL5LNP_5</td>
<td>0.210</td>
</tr>
<tr>
<td>B_1_FXPAL4UNC_4</td>
<td>0.210</td>
</tr>
</tbody>
</table>

* A_2_CMU_See_1

B_2_UvA-MM_1

A_2_UCFVISION_1

A_2_CMU_ESP_3

B_2_UvA-MM_2

B_1_FXPAL5LNP

B_1_FXPAL4UNC

A_2_CMU_Hear_2
2006: Average precision by topic

Interactive max
Manual max
Automatic max
Interactive median
Manual median
Automatic median

Events

People in uniform and in formation
Soldiers, police or guards escorting a prisoner
Soccer goalposts
Condoleezza Rice

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Mean average precision

Interactive max
Manual max
Automatic max
Interactive median
Manual median
Automatic median

Topic number
2005: Average precision by topic
2006: Interactive runs’ median average precision by topic

- 195: Soccer goalposts
- 196: Scenes with snow
- 179: People in uniform and in formation
- 194: Condoleezza Rice
- 178: Saddam Hussein with at least one other person’s face
- 173: Tall buildings (more than 4 stories)
- 175: Soldier/s, police, or guard/s escorting a prisoner
- 192: Greeting by at least one kiss
2005: Interactive runs’ median average precision by topic

- 156: Tennis players on the court – both players visible at the same time
- 153: Tony Blair
- 171: Goal being made in a soccer match
- 149: Condoleezza Rice
- 151: Omar Karami

Interactive median AP
2006: Manual runs’ median average precision by topic

178: Saddam Hussein with at least one other person's face
179: People in uniform and in formation
195: Soccer goalposts
181: One or more soldiers or police with one or more weapons and military vehicles
188: Something burning with flames visible
175: Soldier/s, police, or guard/s escorting a prisoner
192: Greeting by at least one kiss on the cheek
176: Daytime demonstration or protest with at least part of one building visible
2005: Manual runs’ median average precision by topic

- 151: Omar Karami, the former PM of Iraq
- 152: Hu Jintao, President of the People’s Republic of China
- 153: Tony Blair
- 171: tall building
- 164: ship or boat
2006: Automatic runs’ median average precision by topic

- 196: Scenes with snow
- 178: Saddam Hussein with at least one other person's face
- 195: Soccer goalposts
- 188: Something burning with flames visible
- 194: Condoleezza Rice
- **175: Soldier/s, police, or guard/s escorting a prisoner**
- 189: A group of at least 4 people dressed in suits, seated, and with at least one flag
- 180: US President George W. Bush Jr. walking
2005: Automatic runs’ median average precision by topic

- 171: Goal being made in a soccer match
- 151: Omar Karami, the former PM of Iraq
- 153: Tony Blair
- 152: Hu Jintao
- 164: ship or boat
2006: Mean average precision (interactive max) vs total number relevant
2005: Mean average precision (interactive max) vs total number relevant
Who did what?

- Speaker slots to follow:
  - Carnegie Mellon University
  - University of Amsterdam
  - Columbia University
  - IBM

- Demos?
- Posters?
Observations 2005!

- We’re still getting “Lots of variation, interesting shot browsing interfaces, mixture of interactive & manual”, and additionally automatic runs;
- Top performances on all 3 search types are up, even with more difficult data, but data is different, systems are different … anybody run 2004 system on 2005 data?
- Some leveraged the structured nature of B/News;
- Many did automatic search & fewer did interactive search - because it’s easier (no users)?
- Most common issue explored was the best combination of text vs. image search vs. concept/features;
- Search participants are the “regulars” plus new groups, some bigger, some smaller;
Observations 2006

- Top performances on all 3 search types are down
  - Test collection is twice as big
  - Half as many relevant shots
  - Harder topics? Data? ‘Events’ in topics?

- Again, increase in automatic search & fewer did interactive search, almost nobody manual
  - It’s easier (no users)?
  - Topic to query translation good enough?
  - ?

- Manual runs no longer outperform automatic – is this because so few manual, and does it make sense to keep this processing type?