TRECVID-2007:
Search Task

Alan Smeaton
Dublin City University
&
Paul Over
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
- Test and training videos were viewed by NIST personnel, notes taken on content, topic candidates chosen, examples added from development set and Web
- Different in 2007
  - more topics asking for generic (vs. specific, named) targets
  - Almost 1/2 topics ask for video of an event – encouraging exploration beyond one-keyframe-per-shot
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.

■ EXCEPT: experimental **collaborative runs** from FX Palo Alto Labs.
2007 data different from 2003-6 data

- Educational, cultural, youth-oriented programming, news magazines, historical footage, etc.
- Primarily in Dutch
- Much less repetition
  - No commercials
  - No repeated stock news footage
  - Greater variety of subject matter
## 2007: Search task participants

<table>
<thead>
<tr>
<th>Institution</th>
<th>FE</th>
<th>SE</th>
<th>SU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing Jiaotong University</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
<tr>
<td>Bilkent University</td>
<td>**</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>City University of Hong Kong (CityU)</td>
<td>--</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>COST292 Team</td>
<td>SB</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>CWI-CTIT-U.Twente</td>
<td>--</td>
<td>**</td>
<td>SE</td>
</tr>
<tr>
<td>Dublin City University</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
<tr>
<td>Etter Solutions Research Group</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
<tr>
<td>Fudan University</td>
<td>--</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>FX Palo Alto Laboratory Inc.</td>
<td>**</td>
<td>**</td>
<td>SE</td>
</tr>
<tr>
<td>Helsinki University of Technology</td>
<td>**</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>IBM T. J. Watson Research Center</td>
<td>**</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>INESC-Porto</td>
<td>--</td>
<td>**</td>
<td>SE</td>
</tr>
<tr>
<td>K-Space</td>
<td>--</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>Microsoft Research Asia</td>
<td>--</td>
<td>FE</td>
<td>**</td>
</tr>
<tr>
<td>Multimedia Computing Group (CAS) / Natl. U. of Singapore</td>
<td>--</td>
<td>FE</td>
<td>**</td>
</tr>
<tr>
<td>Oxford University</td>
<td>--</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>Tsinghua University / Intel China Research Center</td>
<td>SB</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>Universidad de Jaén (SINAI)</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
<tr>
<td>University of Amsterdam (MediaMill Team)</td>
<td>--</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>University of California, Santa Barbara</td>
<td>--</td>
<td>FE</td>
<td>SE</td>
</tr>
<tr>
<td>University of Central Florida</td>
<td>--</td>
<td>FE</td>
<td>**</td>
</tr>
<tr>
<td>University of Glasgow</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
<tr>
<td>University of Iowa</td>
<td>**</td>
<td>FE</td>
<td>--</td>
</tr>
<tr>
<td>University of Queensland</td>
<td>--</td>
<td>--</td>
<td>SE</td>
</tr>
</tbody>
</table>
Search Types: Automatic, Manual and Interactive

Automatically: System takes topic as input and produces result without any human intervention.

Manually: 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.

Interactively: System takes query as input and produces result without human intervention on the location.

Number of runs: 81 (76) automatic, 4 (11) manually assisted, 33 (36) interactive.
Trends continue

- Interactive
- Manual
- Fully automatic
24 Topics (events)

- Find shots of one or more people walking up stairs.
- Find shots of a door being opened.
- Find shots of a person walking or riding a bicycle.
- Find shots of hands at a keyboard typing or using a mouse.
- Find shots of a canal, river, or stream with some of both banks visible.
- Find shots of a person talking on a telephone.
- Find shots of a street market scene.
- Find shots of a street protest or parade.
- Find shots of a train in motion.
- Find shots with hills or mountains visible.
- Find shots of waterfront with water and buildings.
- Find shots of a street at night.

- Find shots with 3 or more people sitting at a table.
- Find shots with one or more people walking with one or more dogs.
- Find shots with sheep or goats.
- Find shots in which a boat moves past.
- Find shots of a woman talking toward the camera in an interview - no other people visible.
- Find shots of a very large crowd of people (fills more than half of field of view).
- Find shots of a classroom scene with one or more students.
- Find shots of a bridge.
- Find shots of a road taken from a moving vehicle through the front windshield.
- Find shots of one or more people playing musical instruments such as drums, guitar, flute, keyboard, piano, etc.
- Find shots that contain the Cook character in the Klokhuis series.
- Find grayscale shots of a street with one or more buildings and one or more people.
Distribution of hits for each topic

199 was intended as “person walking a bicycle or riding a bicycle” but was formulated as “Find shots of a person walking or riding a bicycle.” Doh!
Frequency of target topic-shots

2007
- Test shots * topics: 435,408
- Relevant topic-shots: 4,704 (1.1%)

2006:
- Test shots * topics: 1,907,616
- Relevant topic-shots: 7,225 (0.4%)

2005
- Test shots * topics: 1,098,360
- Relevant topic-shots: 8,395 (0.8%)

2004
- Test shots * topics: 800,808
- Relevant topic-shots: 1,800 (0.2%)

2003
- Test shots * topics: 775,632
- Relevant topic-shots: 2,114 (0.3%)
Very few unique, relevant shots by group
Automatic runs - top 10 MAP (of 81)
(mean elapsed time (mins) / topic)

Another view: in highest scoring run, on average a little more than 2 of the top 10 shots returned contained the desired video
2006: Automatic runs - top 10 MAP (of 76)
(mean elapsed time (mins) / topic)

![Precision vs Recall Graph]

- F_A_2_TJW_Qclass_4 (15)
- F_A_2_TJW_Qcomp_2 (15)
- F_A_2_CMU_Taste_5 (15)
- F_A_2_TJW_Qind_5 (15)
- F_B_2_i2Rnus_1 (6)
- F_B_2_i2Rnus_2 (6)
- F_B_2_COLUMBIA_RR9_storyqibteviscon (15)
- F_B_2_COLUMBIA_RR8_textibviscon (15)
- F_B_2_THU03_3 (0.49)
- F_B_2_THU02_2 (0.5)
Significant differences among top 8 automatic runs (using randomization test, $10^{5}$ iterations, $p < 0.05$)

<table>
<thead>
<tr>
<th>Run name</th>
<th>(MAP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_2_MSRA-..._1</td>
<td>0.088</td>
</tr>
<tr>
<td>A_2_MSRA-..._2</td>
<td>0.065</td>
</tr>
<tr>
<td>A_2_MSRA-..._3</td>
<td>0.062</td>
</tr>
<tr>
<td>A_2_MSRA-..._4</td>
<td>0.066</td>
</tr>
<tr>
<td>A_2_NUSICT_2</td>
<td>0.061</td>
</tr>
<tr>
<td>A_2_NUSICT_3</td>
<td>0.044</td>
</tr>
<tr>
<td>A_2_NUSICT_4</td>
<td>0.044</td>
</tr>
<tr>
<td>A_2_tsinhua_2</td>
<td>0.043</td>
</tr>
<tr>
<td>A_2_tsinhua_4</td>
<td></td>
</tr>
</tbody>
</table>

* A_2_MSRA-..._1
  - A_2_MSRA-..._2
  - A_2_MSRA-..._3
  - A_2_MSRA-..._4

> A_2_MSRA-..._4
> A_2_MSRA-..._2
> A_2_MSRA-..._3

= A_2_NUSICT_3
  - A_2_NUSICT_2
    - A_2_NUSICT_4
    - A_2_tsinhua_2
    - A_2_tsinhua_4

= A_2_tsinhua_4
= A_2_tsinhua_2
Significant differences among top 8 automatic runs (using randomization test, 10**5 iterations, p < 0.05) without Topic 219

<table>
<thead>
<tr>
<th>Run name</th>
<th>(MAP)</th>
<th>A_2_NUSICT_3</th>
<th></th>
<th>A_2_MSRA..._1</th>
</tr>
</thead>
<tbody>
<tr>
<td>* A_2_NUSICT_3</td>
<td>0.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>= A_2_MSRA..._1</td>
<td>0.054</td>
<td>➤ A_2_tsinghua_2</td>
<td>0.043</td>
<td>➤ A_2_NUSICT_2</td>
</tr>
<tr>
<td>&gt; A_2_tsinghua_2</td>
<td>0.043</td>
<td>➤ A_2_NUSICT_4</td>
<td>0.041</td>
<td>➤ A_2_MSRA..._2</td>
</tr>
<tr>
<td>&gt; A_2_NUSICT_2</td>
<td>0.043</td>
<td>➤ A_2_MSRA..._5</td>
<td>0.040</td>
<td>...</td>
</tr>
<tr>
<td>&gt; A_2_NUSICT_4</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>= A_2_MSRA..._3</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; A_2_MSRA..._2</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; A_2_MSRA..._5</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TRECVID 2007
Full randomization test results - top 10 automatic search runs

<table>
<thead>
<tr>
<th>Run B</th>
<th>Run A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F_A_1_dcu-ImgBaseline_4</strong></td>
<td><strong>F_A_1_dcu-ImgBaseline_4</strong></td>
</tr>
<tr>
<td><strong>F_A_2_NUSICT_2</strong></td>
<td><strong>F_A_2_NUSICT_2</strong></td>
</tr>
<tr>
<td><strong>F_A_2_NUSICT_3</strong></td>
<td><strong>F_A_2_NUSICT_3</strong></td>
</tr>
<tr>
<td><strong>F_A_2_NUSICT_4</strong></td>
<td><strong>F_A_2_NUSICT_4</strong></td>
</tr>
<tr>
<td><strong>F_A_2_tsinghua_2</strong></td>
<td><strong>F_A_2_tsinghua_2</strong></td>
</tr>
<tr>
<td><strong>F_A_2_tsinghua_4</strong></td>
<td><strong>F_A_2_tsinghua_4</strong></td>
</tr>
<tr>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_1</strong></td>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_1</strong></td>
</tr>
<tr>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_2</strong></td>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_2</strong></td>
</tr>
<tr>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_3</strong></td>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_3</strong></td>
</tr>
<tr>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_4</strong></td>
<td><strong>F_A_2_MSRA-USTC-SJTU-SEARCH_4</strong></td>
</tr>
</tbody>
</table>

Probability that the difference in runs (A>B) is due to chance:
- **p<0.01**
- **p<0.05**
- **p<0.10**

TRECVID 2007 16
Manual runs – All 4
(mean human effort (mins) / topic)

Another view: in highest scoring run, on average not quite 1 of the top 10 shots returned contained the desired video
Interactive runs - top 10 MAP (of 33)
(mean elapsed time (mins) / topic)

Another view: in highest scoring run, on average 8 of the top 10 shots returned contained the desired video
2006: Interactive runs - top 10 MAP (of 36)
(mean elapsed time for all == ~15 mins/topic)
Top 10 Interactive Runs

MAP vs mean elapsed search time

Mean average precision vs mean elapsed search time (mins.) per 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_TJW_EER_1</td>
<td>0.362</td>
</tr>
<tr>
<td>A_2_FXPAL_CO15_4</td>
<td></td>
</tr>
<tr>
<td>A_2_NUSICT_1</td>
<td></td>
</tr>
<tr>
<td>B_2_UvA-MM_1</td>
<td>0.259</td>
</tr>
<tr>
<td>B_2_UvA-MM_2</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO15_4</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO11_5</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO15_4</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO_3</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_MMA_1</td>
<td></td>
</tr>
<tr>
<td>A_2_NUSICT_1</td>
<td>0.255</td>
</tr>
<tr>
<td>B_2_UvA-MM_1</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO15_4</td>
<td></td>
</tr>
<tr>
<td>B_2_UvA-MM_1</td>
<td></td>
</tr>
<tr>
<td>B_2_UvA-MM_2</td>
<td></td>
</tr>
<tr>
<td>A_2_FXPAL_CO15_4</td>
<td>0.238</td>
</tr>
<tr>
<td>A_2_tsinghua_6</td>
<td></td>
</tr>
</tbody>
</table>

TRECVID 2007
Average precision by topic
Interactive runs’ median average precision by topic

**Interactive median AP**

- 219: The Cook from the Klokhuis program
- 218: People playing musical instruments
- 214: A very large crowd of people
- 207: A waterfront with water and buildings
- 212: A boat moving past
- 206: Hills or mountains
- 208: A street at night
- 220: Grayscale shots of a street, people, building

TRECVID 2007
2006: Interactive runs’ median average precision by topic

- Interactive median AP
- 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
Automatic runs’ median average precision by topic
Who did what?

- Speaker slots to follow:
  - IBM
    - New interactive search system (tagging/browsing hybrid)
    - Expanded concept lexicon
  - FX Palo Alto Laboratories
    - Realtime search cooperation among 2 or more searchers
  - Tsinghua U. / Intel China
    - More work on concept-based search and example-to-concept mapping
  - University of Amsterdam
    - Crossbrowser meets Forkbrowser
  - National University of Singapore
    - Focus on low-level visual and motion features
    - User choice among 3 feedback strategies
Previous Observations

- Observations 2005 (new BN data, multi-lingual)
  - 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,
  - Some leveraged the structured nature of B/News;
  - Many did automatic search & fewer did interactive
  - Most common issue was combination of text/image search/concepts

- Observations 2006 (bigger collection)
  - Top performances on all 3 search types are down (collection x2, half as many relevant shots, harder topics)
  - Increase in automatic search & fewer interactive search
  - Manual runs no longer outperform automatic – so few manual, and does it make sense to keep this processing type?
Observations/Questions

- Still can’t easily compare performance across very different data/topics (unless same system run on both to estimate data effect).
- However, difference between mean of top-10 interactive and the mean of the top-10 automatic runs has increased over 2006 but not over 2005.
Top-10 interactive vs top-10 automatic runs (MAP)
Questions…

- Did systems adapt to new data/topic characteristics?
  - What old approaches stopped/continued working?
  - What new approaches were tried with(out) success?

- Did systems do anything special to support search for events?

- How did systems handle search for grayscale video?

- What is collaborative search all about?

- What experimental designs are being used to isolate the system effect from the search effect in interactive searches?
Thanks to all who participated, contributed and organised

For citing TREC Vid - see the website for preferred citation