Outline

- Why VideoCLEF?
- Who were we this year?
- Tasks 2009
  - Tagging task
  - Affect task
  - Linking task
- Future Plans
Goals of VideoCLEF

- Achieve better access to video in a multilingual setting
- Promote the use of text, speech and language in multimedia retrieval
- Encourage combination of speech and visual features
- Develop and evaluate video analysis tasks
- Build on the rich research tradition in video retrieval (e.g., the TRECVID benchmark)
Participants
VideoCLEF 2009

- Alexandru Ioan Cuza University, Romania (uaic)
- Chemnitz University of Technology, Germany (cut)
- Delft University of Technology and University of Twente, Netherlands (duotu)
- Dublin City University, Ireland, (dcu)
- TNO, Netherlands (tno)
- University of Geneva, Switzerland (unige)
- University of Jaén (sinai)
VideoCLEF Tasks 2009

- **Tagging task** subject classification
  automatic tagging of videos with subject theme labels

- **Affect task** narrative peak detection
  finding points at which viewers perceived dramatic tension

- **Linking task** finding related resources across languages
  linking video to material on the same subject in a different language
Tagging Task

- **Task**: Participants must automatically assign subject labels to videos.
- **Ground truth**: subject labels from the archive
- Each episode (video file) comes with speech recognition transcripts and archival metadata (title and description).

Examples of the 46 subject labels used in 2009:

- geneeskunde (medicine)
- dieren (animals)
- aanslagen (attacks)
- verkiezingen (elections)
- armoede (poverty)
- genocide (genocide)
- burgeroorlogen (civil wars)
- criminaliteit (crime)
- dierentuinen (zoos)
- economie (economy)
- fabrieken (factories)
- gehandicapten (disabled)
- geschiedenis (history)
- havens (harbors)
Tagging Task Data

- Videos shows from Dutch language television series, mostly documentaries and talk shows
- Recycling the collections used by the TREC Vid 2007 and 2008 benchmarks for a new and different task
- Videos supplied by the Netherlands Institute for Sound and Vision.
Tagging Task Flow

Automatic Speech Recognition Module

Tagger  Tagger  Tagger  Tagger  Tagger

recognized words speech recognition transcript recognized words speech recognition transcript recognized words speech transcript recognized
Tagging Task Results

- Tagging can be approached as an ad hoc retrieval task
- Query expansion improves performance
- Best run made use of both metadata and speech recognition transcript

<table>
<thead>
<tr>
<th>run ID</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>cut1.sc.asr.baseline</td>
<td>0.0067</td>
</tr>
<tr>
<td>cut2.sc.asr.expanded</td>
<td>0.0842</td>
</tr>
<tr>
<td>cut3.sc.asr.meta.baseline</td>
<td>0.2586</td>
</tr>
<tr>
<td>cut4.sc.asr.meta.expanded</td>
<td>0.2531</td>
</tr>
<tr>
<td>cut5.sc.asr.meta.expanded</td>
<td>0.3813</td>
</tr>
</tbody>
</table>

Mean Average Precision Results
Chemnitz University of Technology
Affect Task

- New task this year!
- **Task**: Participants must automatically detect narrative peaks (dramatic moments)
- **Ground truth**: generated by human assessors

*Describing the death of Marc Rothko*
Affect Task Data

- 45 Episodes from “Beeldenstorm,” a short-form documentary series on the visual arts
- Why Beeldenstorm?
  - Combination of “Fact and Fun”
  - Henk van Os is known for his narrative ability
  - Each episode lasts 8 minutes
Affect Task

- New task this year!
- **Task**: Participants must automatically detect narrative peaks (dramatic moments)
- **Ground truth**: generated by human assessors

*Describing the death of Marc Rothko*
Affect Task Flow

Automatic Speech Recognition

Visual Analysis

Audio Analysis

recognized words speech transcript recognized
words speech recognition transcript recognized

Narrative Peak Detection
Narrative Peak Example
Narrative Peak Results

- Speech transcript-based approaches showed strongest performance
- Video and audio features not yet successfully exploited
- Challenging task!

<table>
<thead>
<tr>
<th>run ID</th>
<th>point-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>duotu09fix</td>
<td>47</td>
</tr>
<tr>
<td>duotu09ind</td>
<td>55</td>
</tr>
<tr>
<td>duotu09rep</td>
<td>30</td>
</tr>
<tr>
<td>duotu09pro</td>
<td>63</td>
</tr>
<tr>
<td>duotu09rat</td>
<td>63</td>
</tr>
<tr>
<td>unige-cvml1</td>
<td>39</td>
</tr>
<tr>
<td>unige-cvml2</td>
<td>41</td>
</tr>
<tr>
<td>unige-cvml3</td>
<td>42</td>
</tr>
<tr>
<td>unige-cvml4</td>
<td>43</td>
</tr>
<tr>
<td>unige-cvml5</td>
<td>43</td>
</tr>
<tr>
<td>uaic-run1</td>
<td>33</td>
</tr>
<tr>
<td>uaic-run2</td>
<td>41</td>
</tr>
<tr>
<td>uaic-run3</td>
<td>33</td>
</tr>
</tbody>
</table>
New task this year! “Finding Related Resources Across Languages”

Data: 45 Episodes from the short-form documentary series “Beeldenstorm”

Participants are supplied with 165 anchors (short video segments) that need to be linked

Task: Participants must find a target page on the topic that is being treated in the video at the point of the anchor

Ground truth: generated by human assessors
Linking Task Example

Identify articles in English-language Wikipedia that will support comprehension of Dutch-language videos
Linking Task Flow
Linking Results

- Information retrieval approach: transcript words used as query
- Good strategy: Query Dutch index and return the corresponding English page.
- Not a named-entity task, but treatment of named-entities is critical

<table>
<thead>
<tr>
<th>run ID</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcu_run1</td>
<td>0.268</td>
</tr>
<tr>
<td>dcu_run2</td>
<td>0.275</td>
</tr>
<tr>
<td>dcu_run3</td>
<td>0.089</td>
</tr>
<tr>
<td>dcu_run4</td>
<td>0.190</td>
</tr>
<tr>
<td>tno_run1</td>
<td>0.460</td>
</tr>
<tr>
<td>tno_run2</td>
<td>0.428</td>
</tr>
<tr>
<td>tno_run3</td>
<td>0.484</td>
</tr>
<tr>
<td>tno_run4</td>
<td>0.392</td>
</tr>
<tr>
<td>tno_run5</td>
<td>0.368</td>
</tr>
</tbody>
</table>
Future Plans

**Continue to promote multimodality**
- The continuing quest to integrate speech, audio, and visual information to improve multimedia access

**Expand to use a social video collection**
- Internet video = variability of production values
- User contributed information such as tags and ratings are an important information source.
- Relationships between users in a social network can be exploited
Exploratory tasks 2009–2010

Semantic keyframe selection
- Select a keyframe set to provide a semantic representation of thematic content of the entire video

Appeal task
- Predict ability of video to appeal to viewers (independently of its topic)
Acknowledgements

- University of Twente for supplying the speech recognition transcripts
- Netherlands Institute of Sound and Vision for supplying the video
- TrebleCLEF for annotation support
- Colleagues at DCU for supplying shot segmentation
- Colleagues at TU-Delft and in PetaMedia
- Anvil video annotation research tool
- Flickr images from mafleen & kappuru