VideoCLEF 2009

Martha Larson
Delft University of Technology

Gareth Jones
Dublin City University

CLEF2009 Workshop, Corfu, Greece, October 1, 2009

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, Germay (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)
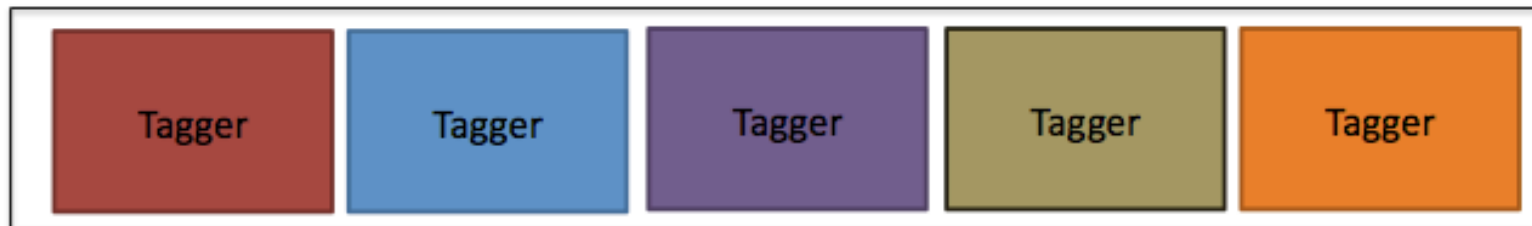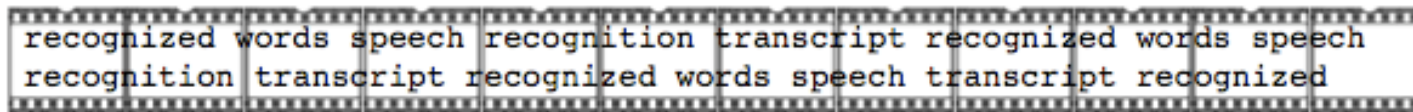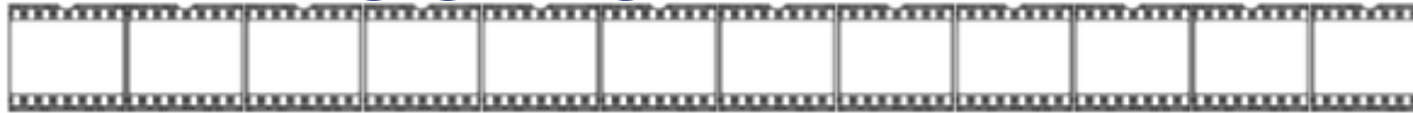gehandicapten (disabled)
geschiedenis (history)
havens (harbors)

Tagging Task Data

- Videos shows from Dutch language television series, mostly documentaries and talk shows

- Recycling the collections used by the TRECVid 2007 and 2008 benchmarks for a new and different task

- Videos supplied by the Netherlands Institute for Sound and Vision.

BEELD EN GELUID

Tagging Task Flow

recognized words speech recognition transcript recognized words speech recognition transcript recognized words speech transcript recognized

Automatic Speech Recognition Module

| Tagger | Tagger | Tagger | Tagger | Tagger |

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

run ID	MAP
cut1_sc_asr_baseline	0.0067
cut2_sc_asr_expanded	0.0842
cut3_sc_asr_meta_baseline	0.2586
cut4_sc_asr_meta_expanded	0.2531
cut5_sc_asr_meta_expanded	**0.3813**

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
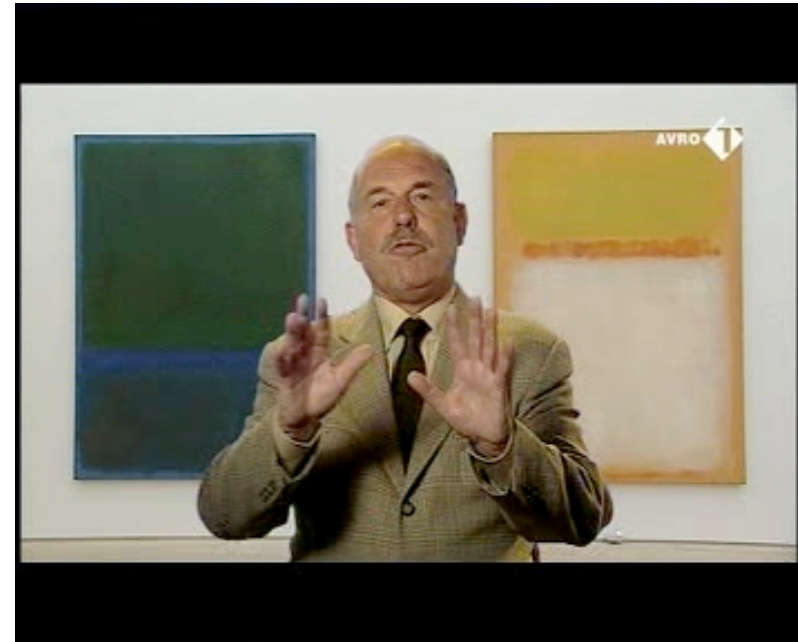
Describing the death of Marc Rothko

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
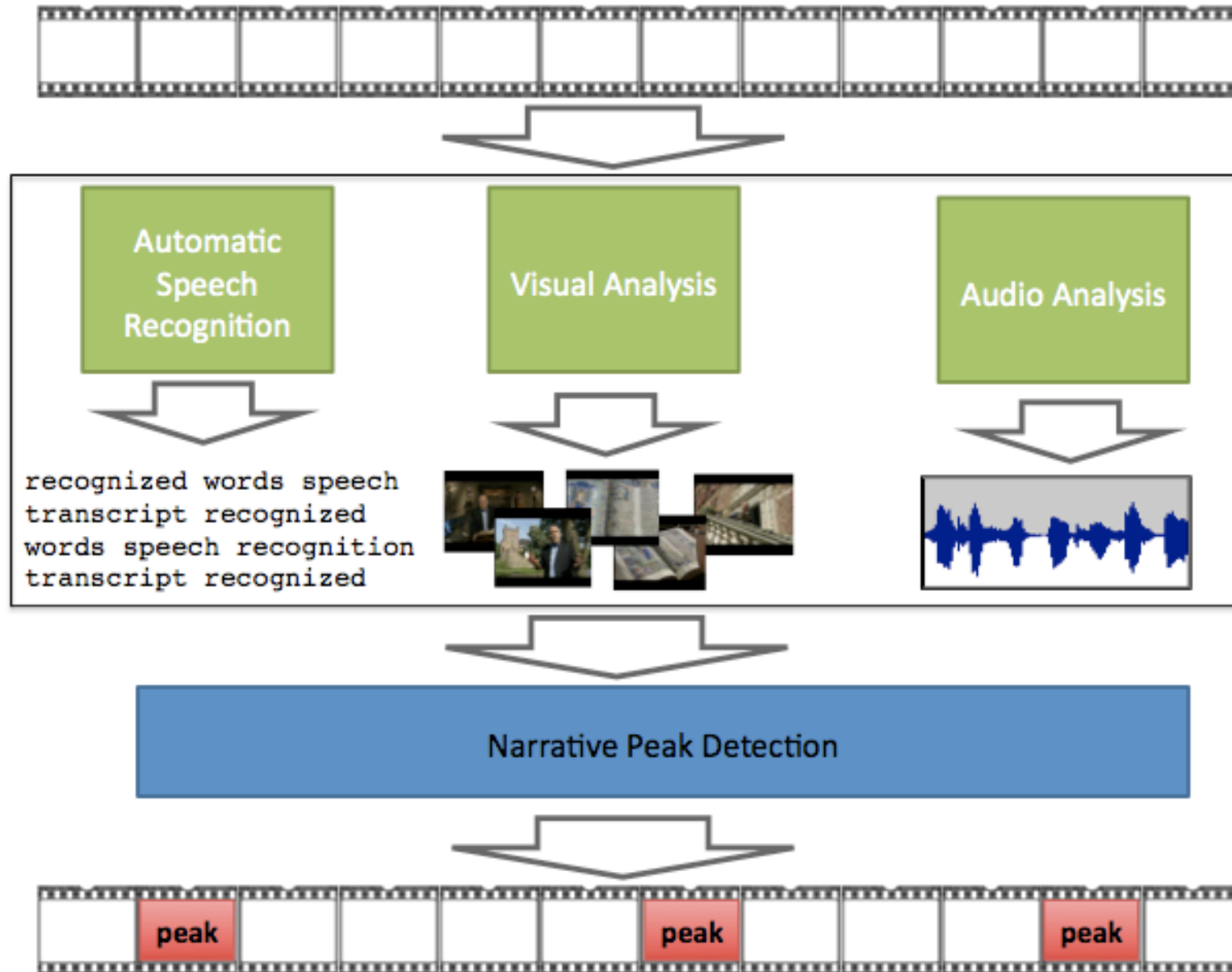
Narrative Peak Example
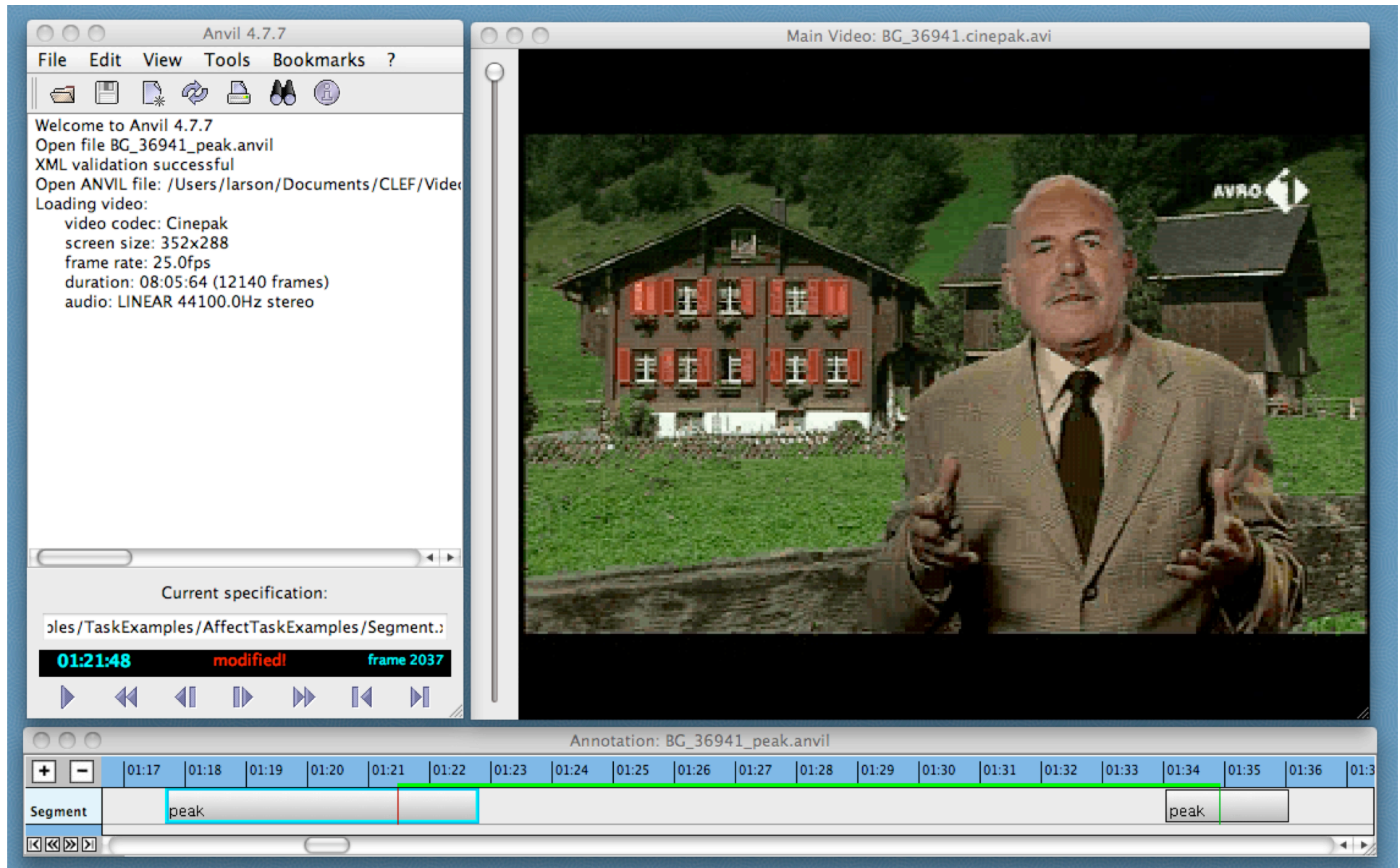
Narrative Peak Results

- Speech transcript-based approaches showed strongest performance

- Video and audio features not yet successfully exploited

- Challenging task!

run ID	point-based
duotu09fix	47
duotu09ind	55
duotu09rep	30
duotu09pro	**63**
duotu09rat	**63**
unige-cvml1	39
unige-cvml2	41
unige-cvml3	42
unige-cvml4	43
unige-cvml5	43
uaic-run1	33
uaic-run2	41
uaic-run3	33

Linking

- 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

run ID	MRR
dcu_run1	0.268
dcu_run2	0.275
dcu_run3	0.089
dcu_run4	0.190
tno_run1	0.460
tno_run2	0.428
tno_run3	**0.484**
tno_run4	0.392
tno_run5	0.368

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
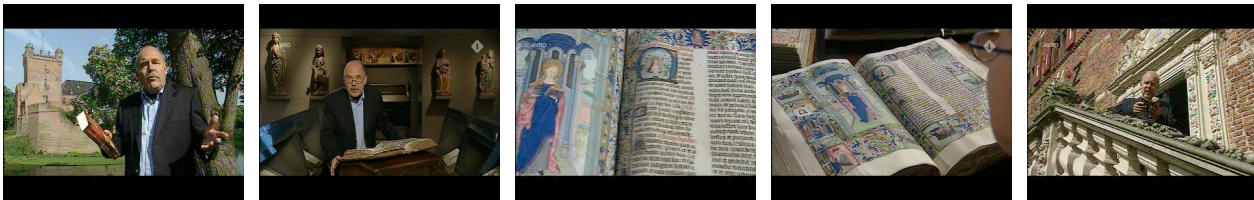
Appeal task

- 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