

DCU at VideoClef 2008

Vid2RSS Pilot Task

Eamonn Newman
Gareth Jones

Contents

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Vid2RSS task and data
- DCU @ VideoCLEF
 - System description
 - Results and Conclusions
 - Future Work

Vid2RSS task

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Goals of VideoCLEF benchmark evaluation
 - Promote research on intelligent access to multilingual video collections
 - Encourage exploitation of speech transcripts
 - Encourage exploitation of video metadata
- Develop and evaluate multilingual video analysis tasks
- Extend the recent Cross-Language Speech Retrieval track with a new track for CLEF 2008

Data for Vid2RSS 2008

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- **50 dual language videos** (30 hours) from The Netherlands Institute for Sound and Vision (Beeld en Geluid)
- Videos are episodes of Dutch television shows, mostly documentaries
- Dutch is the main language; English is an embedded language.
- Embedded language is spoken mainly by interviewees
- Videos are accompanied by Dutch-language archival metadata records
- Metadata includes series title, episode title, description, date of broadcast and other production information
- Speech recognition transcripts and Shot-level keyframes

Task requirements

Classification Task (Main Task)

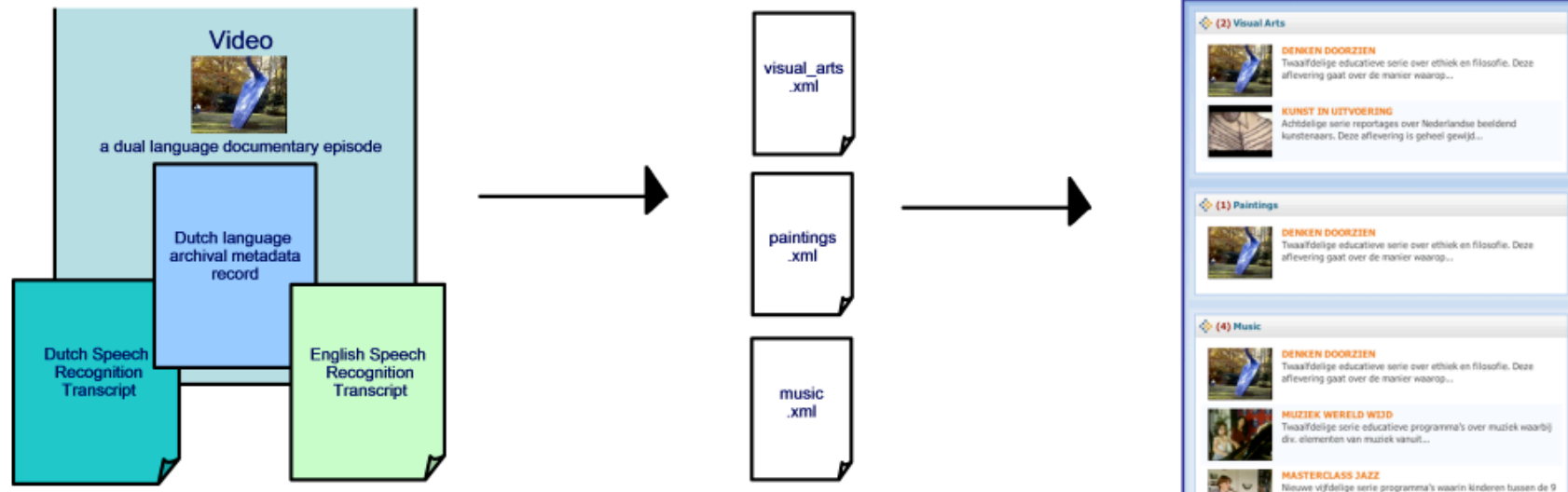
- Assign videos to thematic classes using speech recognition transcripts only (required)
- Use combination of metadata and speech recognition transcripts to perform classification

Translation Task

- Translate output RSS-feeds (e.g., into English)

Keyframe Extraction Task

- Select a keyframe to provide a semantic representation of the entire video to be used to depict the video in the feed.



System Overview

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Information Retrieval system classifies videos into particular topic-based RSS feeds
- Based on Lucene Search and Indexing technology
- Designed to provide a baseline for the Vid2RSS classification task

System Functionality

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Free text index built using ASR transcripts and video metadata
- Feed labels were used as queries (eg “Architecture”, “Visual Arts”)
- Retrieved items added to that feed
- Some of the runs allowed an item to appear in one feed only, while others allowed items to appear in multiple feeds

Submitted Runs

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- **Run 1** Dutch ASR transcripts; query only
- **Run 2** English ASR transcripts; query only
- **Run 3** Dutch ASR transcripts; relevance feedback on query
- **Run 4** English ASR transcripts; relevance feedback on query
- **Run 5** Catalogue metadata in Dutch; query only

Results

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

metric	Run 1	Run 2	Run 3	Run 4	Run 5
micro-average precision	0.50	0.32	0.16	0.17	0.83
micro-average recall	0.35	0.21	0.91	0.72	0.18
f-score micro-average	0.41	0.25	0.28	0.28	0.29
macro-average precision	0.54	0.62	0.42	0.50	0.93
macro-average recall	0.55	0.38	0.90	0.70	0.28
f-score macro-average	0.54	0.47	0.58	0.59	0.43

Results

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- **Runs 1 and 2** Dutch ASR transcripts are more useful than English.
- **Runs 3 and 4** Relevance Feedback improves recall but downgrades precision.
- **Run 5** The catalogue metadata achieves high precision scores.

Future Work

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Refine relevance feedback techniques to maximise precision and recall.
- Examine system performance at query-level
- Investigate combination of ASR and catalogue metadata.

Class Labels

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Archaeology
- Architecture
- Chemistry
- Dance
- Film
- History
- Music
- Paintings
- Scientific Research
- Visual Arts