VK Multimedia Information Systems

Mathias Lux, mlux@itec.uni-klu.ac.at

Dienstags, 16.oo Uhr c.t., V.1.08

Video Retrieval

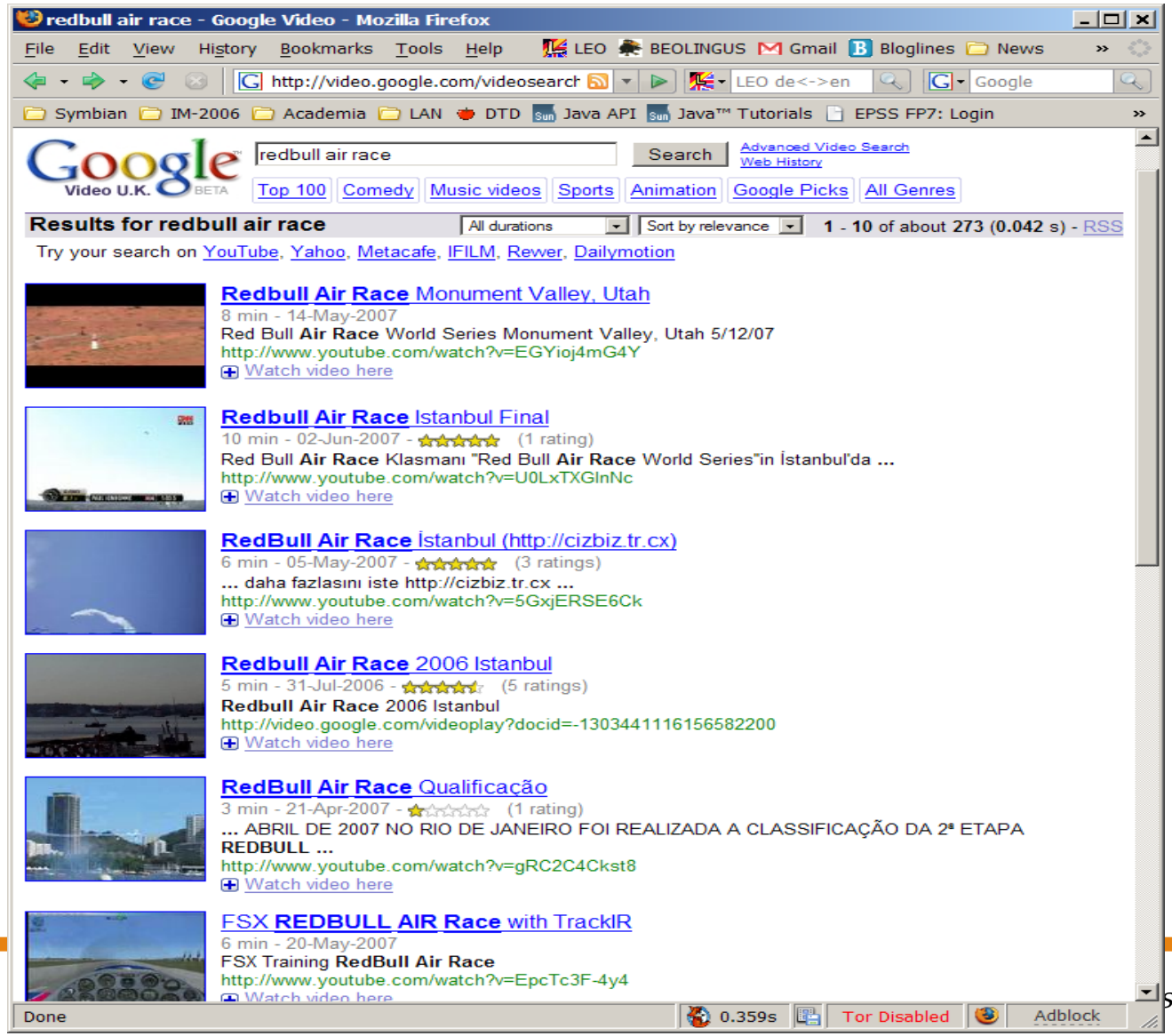
- Motivation & Problems
- Features & Descriptors
- Some Methods
  - Text Based
  - Shot Detection
- Video Retrieval Evaluation
- Applications
  - Video Summaries

Motivation

Szenario A: Ad Hoc Search - Pull Information

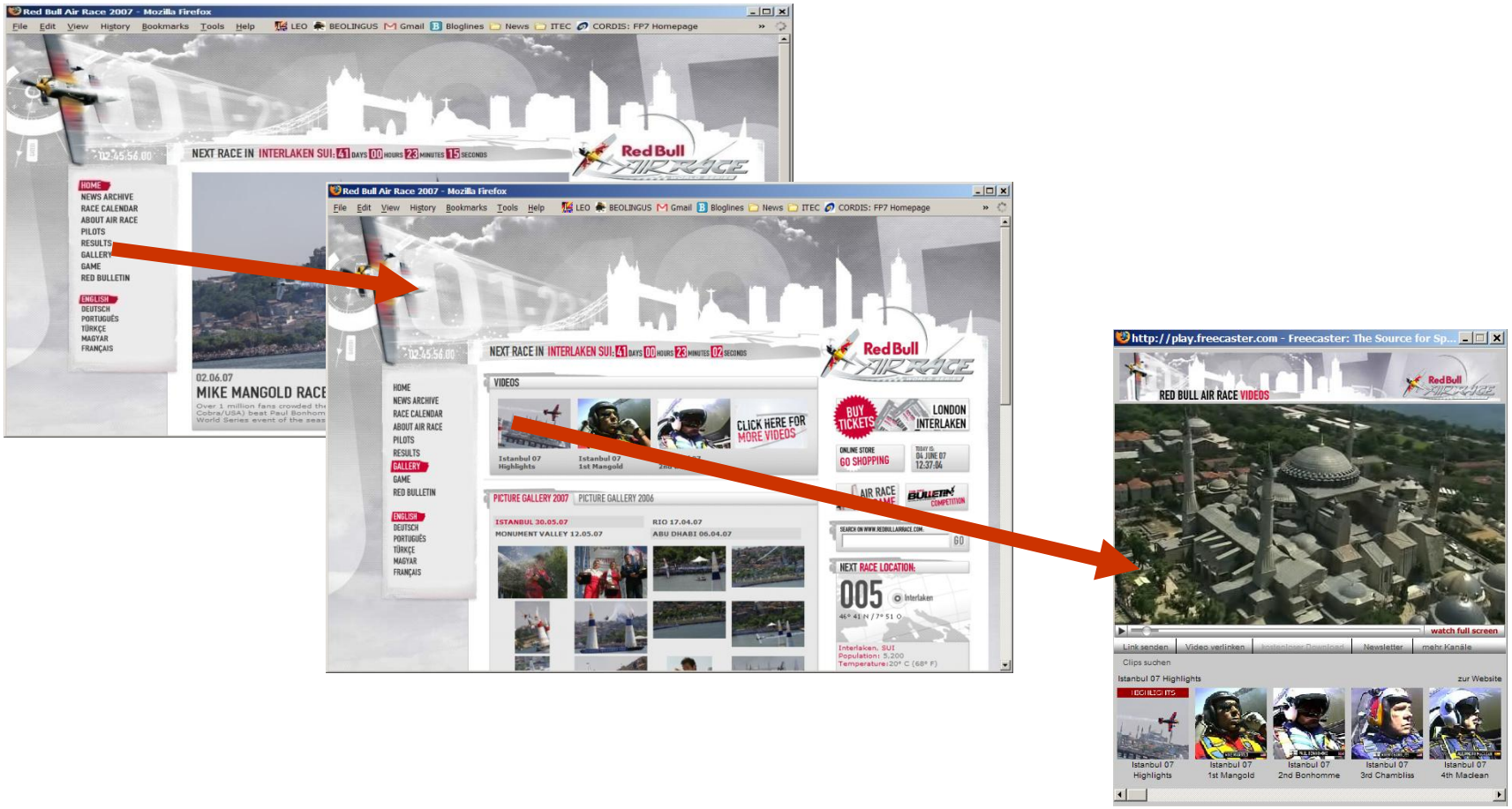
- Alice has heard about a recent event
  - Examples: Red Bull Air Race, etc.
- She wants to get an overview on
  1. Overview on **context**
  2. Coverage on the **outcomes & highlights**

Szenario A: Google Video



Szenario A: Web Site

http://www.uni-klu.ac.at



Szenario A: Analysis

Google Video	***Air Race Web Site***
Simple (Term) Search	Navigation (Gallery -> Video)
Short and ambiguous descriptions	Clear and intuitive meta information (thumbnails)
No additional information / interlinking	Further information provided
Fast, clean and efficient interface	Frisky and colorful interface
Legal issues ...	No legal issues

Szenario B:

Szenario B: Media Obervation

- George B. wants to find everything
  - Concerning certain Persons / Communities
  - Capturing the mood of media
- This includes
  - News broadcasts (language independent)
  - YouTube, MyVideo, etc.

Problems

- Video Retrieval is a very broad field
  - Demands differ from professionals to hobbyists
- Videos are commonly rather ‘big’
  - Sighting of raw footage and search results is time consuming
  - Extraction, analysis and indexing of descriptors are challenging
- Indexing is rather complicated
  - Videos are multimodal

Example Problem: Size

- 15 minute video -> 25 fps, 720x576
  - # frames = 15 * 60 * 25 = 22,500
  - With 65k colors
    - Raw size = 22,500 * 720 * 576 * 2 ~ 17.4 GB
  - Indexed by color histogram
    - 256 colors with 256 levels each -> 16 Bit / frame
    - Size = 22.500 * 2 ~ 43.95 kB
  - In a video database
    - 1,000 videos -> ~ 44 MB descriptor data
    - 1,000,000 videos -> ~ 44 GB descriptor data

Video Retrieval

- Motivation & Problems
- Features & Descriptors
- Methods
  - Text Based
  - Shot Detection
- Video Retrieval Evaluation
- Applications
  - Video Summaries

Features and Descriptors

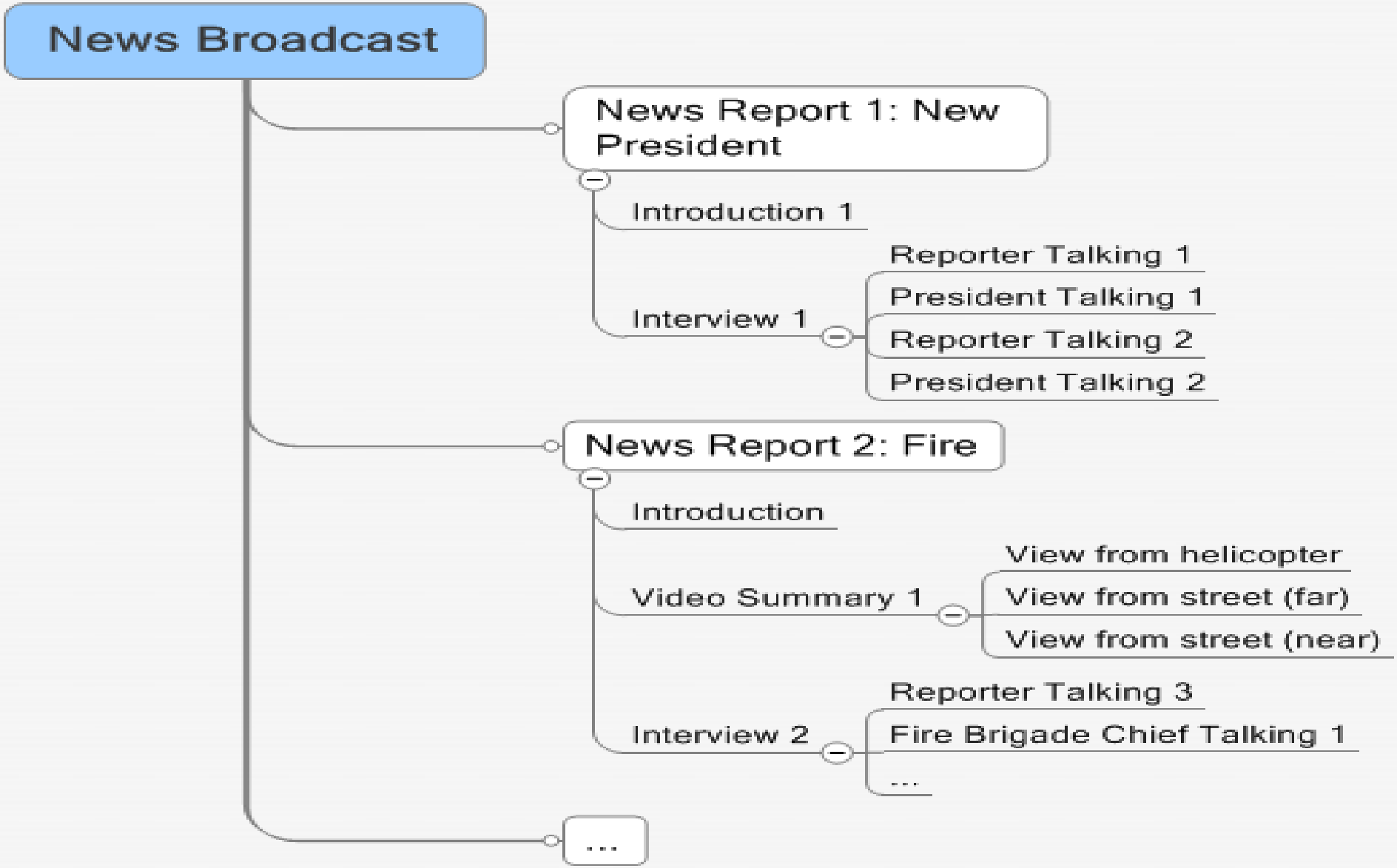
- Visual Descriptors:
  - Additional dimension: **Time**
  - Related to audio information
- Audio Descriptors
  - Related to visual information
- Multiple Streams
  - Different languages, comments
  - Different angles / viewpoints

GOP & GOF

Video stream is sequence of still images

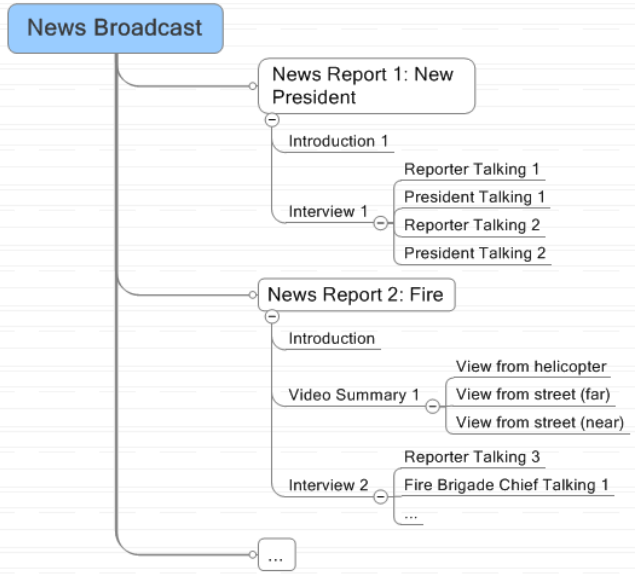
- Instead of single picture
  - Group of Frames (short: GOF)
  - Group of Pictures (short: GOP)
- Color description of multiple frames
  - e.g. averaged

Temporal Segmentation



Temporal Segmentation

- A single decomposition
  - Three different levels
  - Non-overlapping segments
- Visual and audio descriptors
  - Attached to nodes
  - Describing frames of GOF



MPEG-7

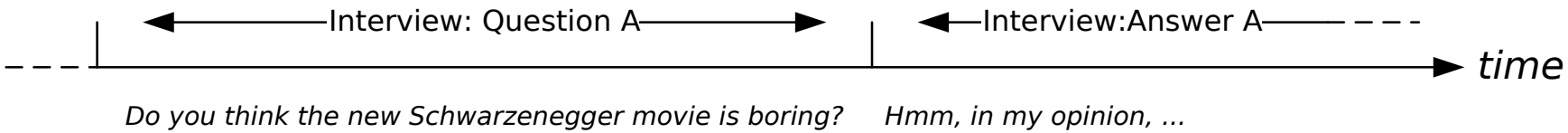
- Multiple segmentation trees possible
- Different stream combined
- No “general description format”
  - How many segmentations / levels
  - Selection of descriptors at nodes
  - Interconnection of streams

Video Retrieval

- Motivation & Problems
- Features & Descriptors
- Some Methods
  - Text Based
  - Shot Detection
- Video Retrieval Evaluation
- Applications
  - Video Summaries

Text Based Retrieval

- Text annotations assigned to segments
  - Transcriptions, metadata, etc.
- Retrieval is based on text
  - Inverted lists
  - Retrieval of relevant parts/documents

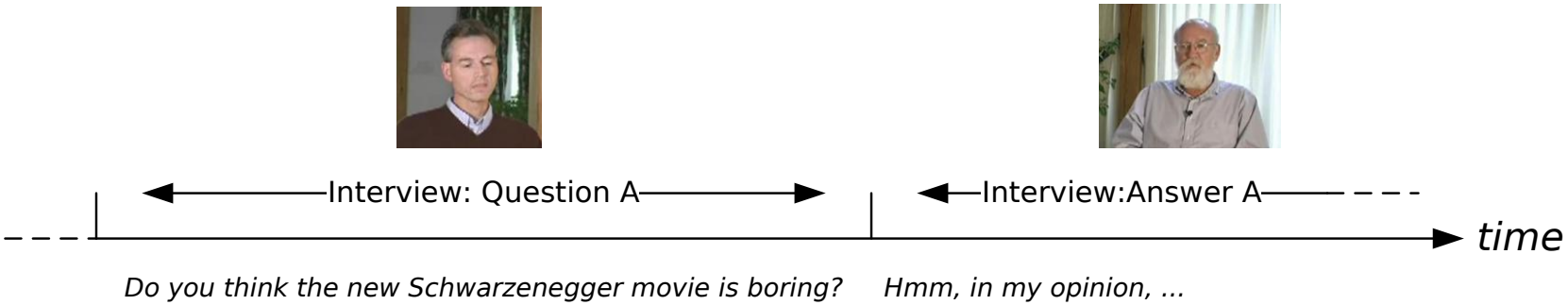


Text Based Retrieval: Applications

- Speech oriented videos
  - Speech recognition & manually
  - Transcription available for disabled people
  - Examples: News, Cartoons
- Metadata of videos
  - Tagging and descriptions like in YouTube
  - Manual annotations (e.g. sports videos)
  - Spotted keywords

Shot Detection

- Automatic Segmentation of video stream
  - Find frame where new shot starts
  - Find frame describing the shot best



Different Cuts

- Simple Cuts (elephantsdream)

- Transitions & combinations (casino royale)

Shot Detection: Methods

- Uncompressed Domain
  - Video is decoded
  - RGB or YUV values are used for computation
- Compressed Domain
  - Characteristics of the codec are exploited

Shot Detection: Uncompressed Domain

- Rather good methods already available
  - Detection up to 95%
  - Depends on domain
- Ad detection
  - Logo tracking in the corner of the frame
- News Broadcasts
  - Background tracking (studio environment)
- General approaches
  - Grey values / Color Histogram

Shot Detection: Uncompressed Domain

Common Algorithm

- For each frame n
  - Extract *histogram(n)*
  - Compute distance to *histogram(n-1)*: *d(n-1, n)*
  - If (*d(n-1, n)* > *threshold*) report shot boundary
- Problems
  - Each frame has to be decompressed
  - Threshold is domain dependent.

Shot Detection: Compressed Domain

- Motion Vectors
  - Investigate major direction / amount changes
- Bit Rate
  - VBR: Higher amount -> shot boundary
- Number Macro Blocks / Type
  - More I-Blocks -> shot boundary
- Position of I-Frames
  - Actually a shot detection in encoding

Video Indexing based on Shots

- Indexing Shots instead of frames
  - Number of shots depends on the domain
  - Considerably smaller than number of frames
- What to index about a shot?
  - Identify one or more “key frames”
  - Index the key frames
- Retrieval based on shots
  - Result is “part of the video”
  - Grouping possible, weighting neccessary

Video Retrieval

- Motivation & Problems
- Features & Descriptors
- Some Methods
  - Text Based
  - Shot Detection
- Video Retrieval Evaluation
- Applications
  - Video Summaries

Retrieval Evaluation

- Similar to IR Evaluation
- Several different tasks
  - Depending on the forum

Retrieval Evaluation Forums

- TRECVID
  - Indexing and searching in video DBs
- VideoCLEF
  - Video content in multilingual environments
- INEX Multimedia
  - XML (Fragments) based multimedia retrieval

TRECVID 2007

- Shot boundary Detection
  - Automatic comparison to human annotation reference data.
- High Level Feature Extraction
  - Classification based on 39 concepts
- Search
  - Ranked list based on shots compared to test collection
  - automatic, manually assisted & interactive
- Rushes Summarization
  - Management of raw video material (near duplicate scenes, no audio etc.)
  - Evaluation by a single human judge

VideoCLEF 2008

- Classification Task: Vid2RSS
  - Dutch television footage
  - Dual language: English & Dutch
  - Both contribute, not translations
  - Transcriptions, keyframes, metadata provided
  - Task: RSS feed for each category
- ImageCLEF
  - Image retrieval tasks

INEX Multimedia

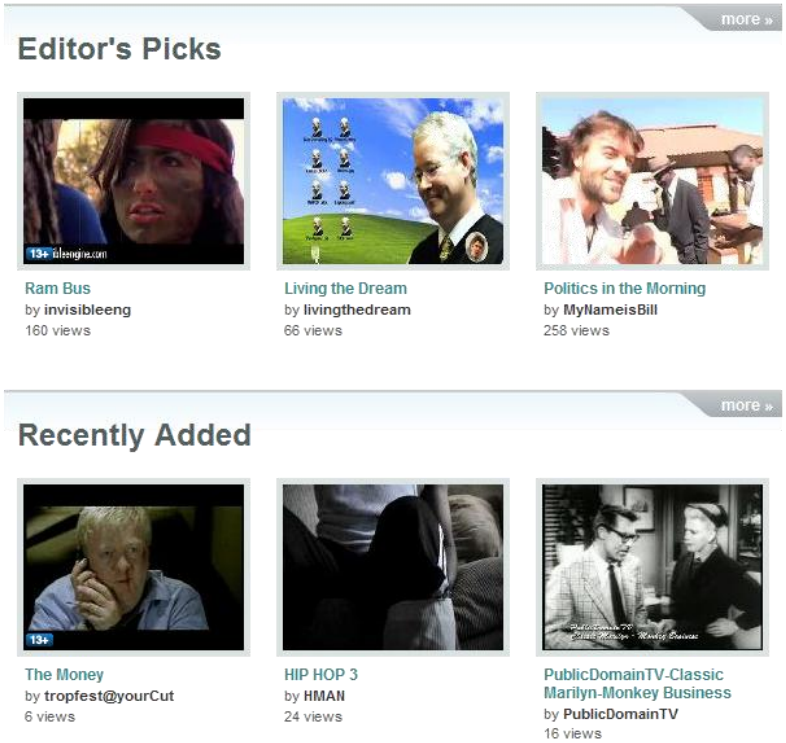
- Retrieving relevant document fragments with multimedia character
- Input (Query):
  - Either Text or Text & Image
- Output (Result):
  - Image or text or both
- Evaluation
  - Human assessment

Video Retrieval

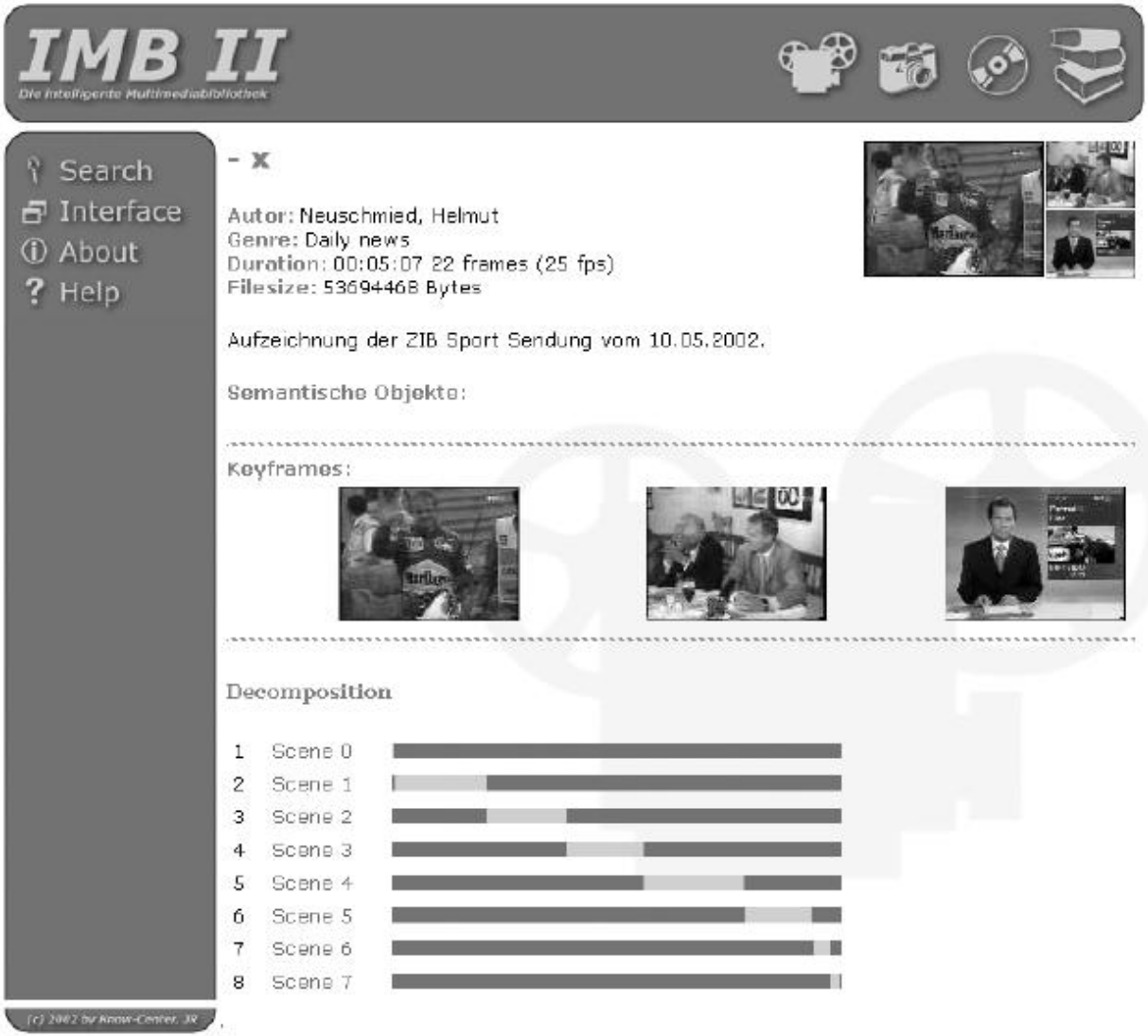
- Motivation & Problems
- Features & Descriptors
- Some Methods
  - Text Based
  - Shot Detection
- Video Retrieval Evaluation
- Applications
  - Video Summaries

Video Summaries

- Methods for getting the most out of a video in minimum time



Video Summary Example



Key Frames

Goals

- Select appropriate frames for a summary
- Weight frames according to relevance
- Visualize in an ‘optimal’ way

Problems

- Which are the most relevant frames?
  - Sort out transitions, motion blurred frames
- How many are there?

Video Summaries: Animations

- Selection of key frames
- Rotated in a loop

http://www.myvideo.de/watch/1544203

Video Summaries: Stripe Images

- Only one pixel column per frame
- Concatenate the pixel columns
  - frame height = stripe image heigth
  - frame number is stripe image width
- Visualization Benefits
  - Size of shots, Movement
- Visualization Disadvantages
  - No ‘big picture’

Video Summary Generation

- Approaches use most salient frames
  - Based on user attention models
    - Motion, static shots, faces, etc.
  - Clustering & SVD
    - Employ dimensionality reduction
    - Find groups and take representative group members
    - The bigger the group the more important
  - Optimization
    - Minimizes sum of distances to all other frames.
    - While maximizing the distances between key frames

Exercise 04

- Create a video summary
  - ... of *Chad Vader I – Day Shift Manager*
    - http://www.youtube.com/watch?v=4wGR4-SeuJ0
- Use *Video Downloader* to grab video
  - http://javimoya.com/blog/youtube_de.php
- Decide yourself which visualization you want to implement ...
  - Do not use frames displaying text
- Send me the resulting image / document

Exercise 04 Option: Stripe Image

- Use **FFMPEG** to grab frames
  - e.g. the windows binary
  - `ffmpeg -i [invideo] -f image2 -ss frame%6d.png`
  - see e.g. http://wiki.cs.sfu.ca/vml/DigitalVideoHowTo
- Use e.g. **Irfanview** to put them together
  - Batch Processing -> Crop images ...
  - Image -> Panorama image ...

Thank you …

… for your attention