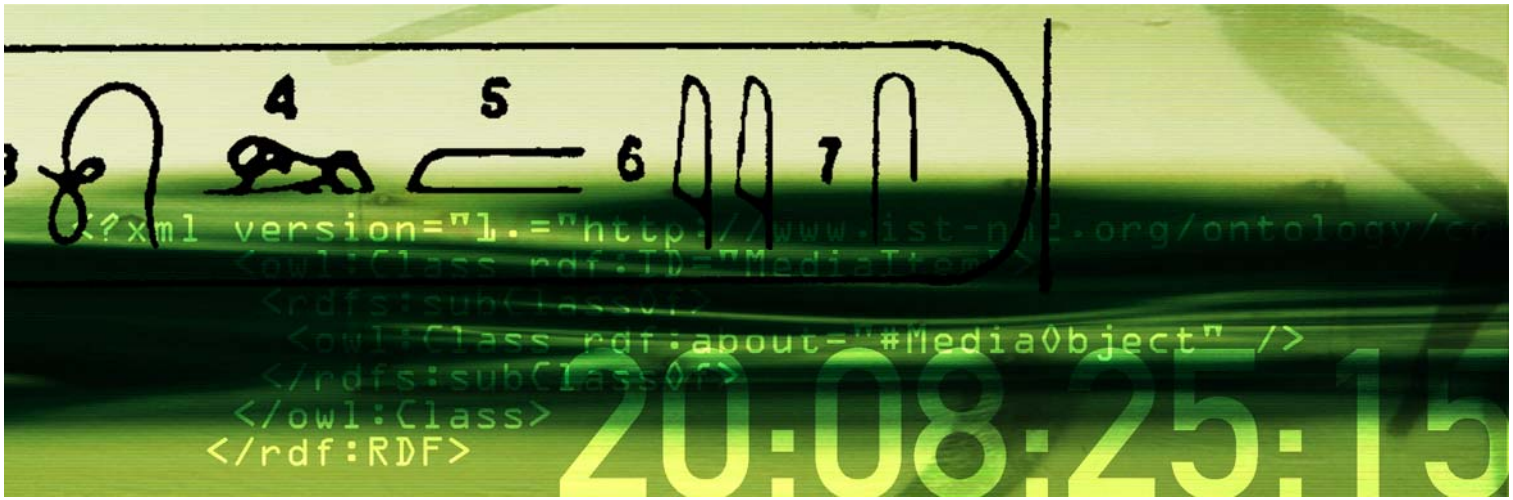
JOANNEUM RESEARCH

Institut für Informationssysteme & Informationsmanagement



Skimming Rushes Video Using Retake Detection

Werner Bailer, Felix Lee, Georg Thallinger

TRECVID Video Summarisation Workshop @ ACM MM, 2007-09-28

www.joanneum.at

JOANNEUM RESEARCH

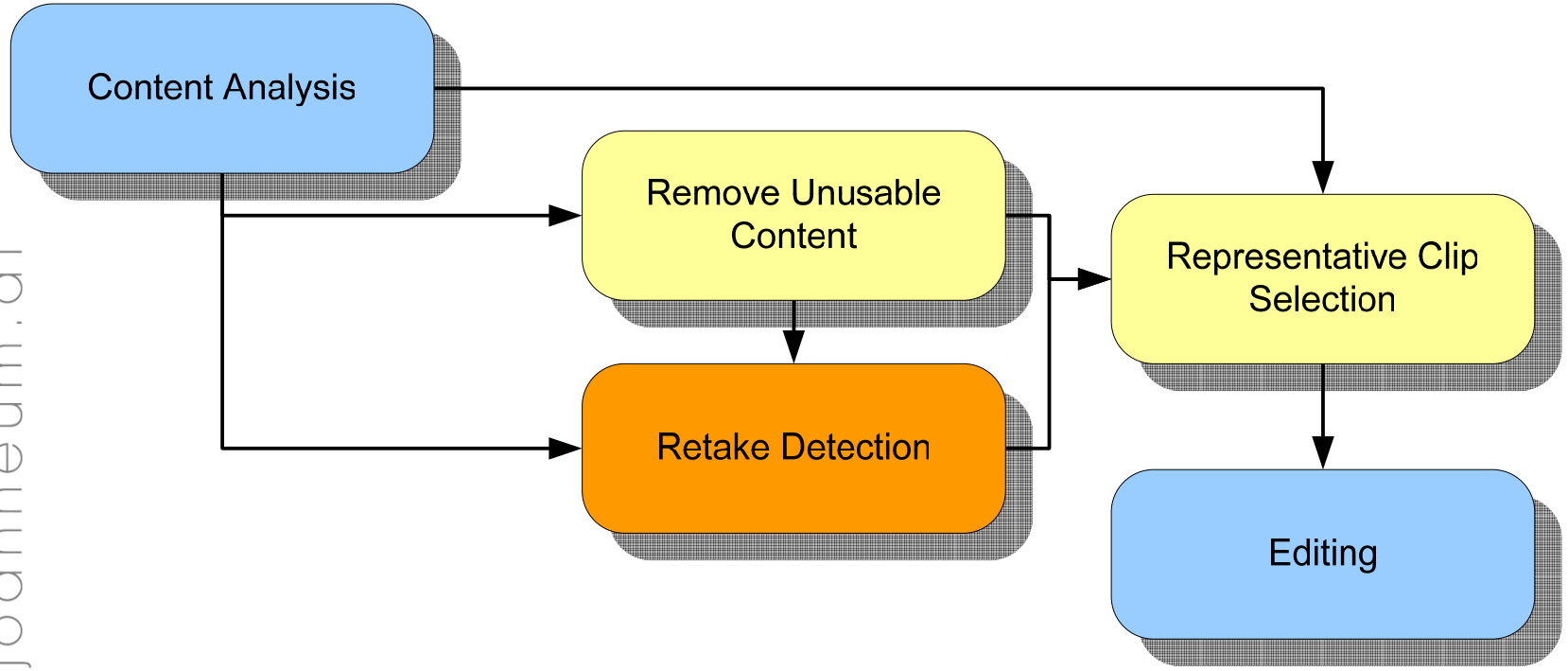
Overview

- Process Overview
- Retake Detection
- Content Selection
- Results

www.joanneum.at

JOANNEUM RESEARCH

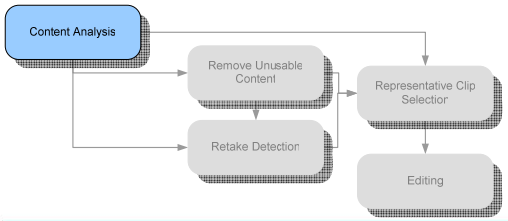
Skim Creation Process



Content selection
Content clustering
Pre/Post processing

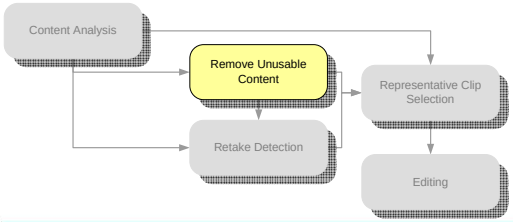
www.joanneum.at

JOANNEUM RESEARCH



Content Analysis

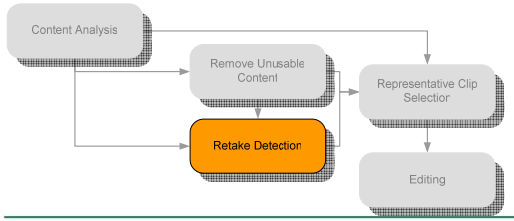
- Shot boundary detection
  - frame differences, SVM classifier trained on TRECVID 2006 data
- MPEG-7 Color Layout and EdgeHistogram
  - descriptors extracted from every 10th frame
- Visual activity
  - Averaged over 10 frames
- Face Detection
  - Viola/Jones, OpenCV implementation



Remove Unusable Content

- Skip short shots
  - duration < 10 seconds
- Remove color bars and monochrome frames
  - standard deviation in columns < 15 levels in each channel

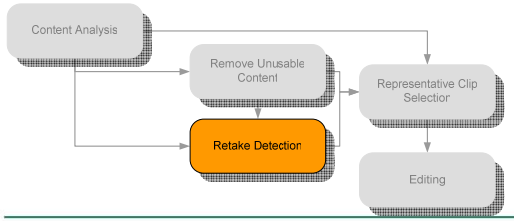
JOANNEUM RESEARCH



Retake Detection Overview

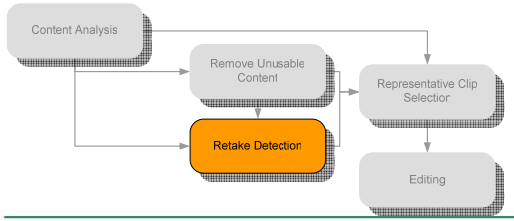
- retake = take of same scene, from same camera
- split shots into parts
  - split at short-term local maxima of visual activity (clapboard movements, production staff walking around)
- pair-wise matching of parts
  - match extracted colour, texture and visual activity descriptor sequences of the parts (temporally sub-sampled by 10)
  - modified Longest Common Subsequence (LCSS) algorithm
  - clean up matches (remove contained and largely overlapping matches): set of (partial) “take candidates”
  - result is a similarity matrix of the take candidates
- cluster take candidates
- determine relevance over time
  - based on overlap with takes in the same cluster

JOANNEUM RESEARCH



Retake Detection Matching

- Transform problem of matching parts to problem of matching sequences of feature vectors of these parts
- Requirements to matching algorithm
  - Match similar, but mostly not identical feature sequences
  - Enforce minimum length of matches
  - Accept gaps and insertions, but enforce maximum length of gap/insertion



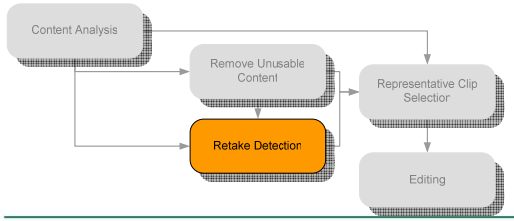
Retake Detection LCSS

- LCSS variant proposed by [Vlachos et al., 2002] for 2d trajectories

$$\begin{cases} 0 \quad if \ \text{A} \ or \ \text{B} \ is \ empty, \\ 1 + LCSS_{\delta,\epsilon}(Head(A), Head(B)), \\ \quad if \ \ |\text{a}_{x,n} - b_{x,m}| < \epsilon \text{ and } |a_{y,n} - b_{y,m}| < \epsilon \text{ and } |n - m| \leq \delta \\ \max(LCSS_{\delta,\epsilon}(Head(A), B), LCSS_{\delta,\epsilon}(A, Head(B))), \\ \quad otherwise \end{cases}$$

- our modifications:
  - replace ε by a vector of thresholds $\{\varepsilon_1,...,\varepsilon_m\}$ for m features, which are weighted by weights $\{w_1,...,w_m\}$
  - discard δ, absolute temporal distance of feature vectors is irrelevant
  - introduce maximum gap γ between two consecutive matching feature vectors
  - accept all matches longer than minimum length of a take

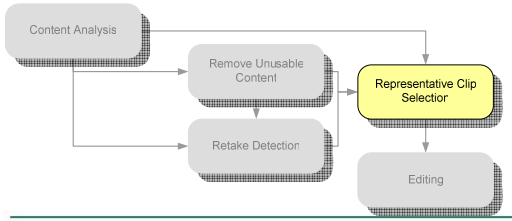
JOANNEUM RESEARCH



Retake Detection Clustering

- hierarchical single-linkage clustering of the take candidates
  - distance between clusters: 1 - minimum of normalised LCSS
  - constraint: assign single takes to cluster before merging clusters (avoid merging similar scenes before all takes of one scene are clustered)
  - clustering stops when distance reaches minimum length of match between takes

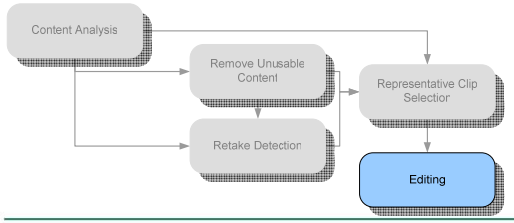
JOANNEUM RESEARCH



Representative Clip Selection

- assign weights to each frame
  - initialize with relevance from retake detection
  - weight with visual activity, reduce weight at beginning of frame to discard clapper board
  - reduce if no face present
- select a clip from each take cluster (or unclustered take)
  - usually only from the take with highest relevance rating (i.e. most overlap with other takes in the cluster)
  - if the cluster spans a long time and there are several equally relevant takes, more takes can be considered
  - extract clip around relvance maximum of take
  - minimum length of clip: 1 second

JOANNEUM RESEARCH



Editing

- Extract clips from source video
- Demultiplex
- Insert text overlays: “*n* more takes”
- Audio fade in/out (or suppress for short clips)
- Multiplex
- Concatenate
- Tools
  - ffmpeg, mencoder, SoX

JOANNEUM RESEARCH

Results (1)

- Evaluation (overall)
  - inclusion: mean 0.46, median 0.47
  - understandability: mean 3.57 (2nd), median 3.67 (best)
  - duplicates: 3.78 (6th), 3.67 (6th)
- MRS044500
  - inclusion: 0.31, understandability: 2.67, duplicates 3.33
  - one duplicate: men going into building – umbrella gets caught
    - wastes time that we would have needed for other scenes
  - missed items
    - Clips selected from clusters are too short and show part of the actions
    - Our selection prefers parts with much activity – not always a good choice
    - None of the missed items is outside of our take clusters
  - no clapboards in this skim

JOANNEUM RESEARCH

Results (2)

- How good does retake detection work?
  - Evaluated with manually created ground truth on 6 videos of test set

	Nr. of scenes		Nr. of takes		Precision	Recall
	Truth	Detected	Truth	Detected		
Mean	7.00	7.50	29.00	25.83	0.71	0.64

- Segmentation of shots into sub-shots fails
  - some takes lost, and maybe unique part that is only in one take
  - clapboards in the skim
- Little action, similar location
  - wrong assignment of takes

JOANNEUM RESEARCH

Conclusion

- Removal of redundant content and clustering of takes worked well
- Better segmentation into subshots (both with and without clapboard present)
- Strategy for selecting clips from take clusters needs improvement
- Interesting future improvements
  - use audio information
  - display information about differences between takes
  - Group takes of same action from different camera positions
  - better editing (observe visual grammar, separate editing of video and audio tracks)