COST 292

- A research network of European institutions on video retrieval
- Bringing the scientists in the field together and setting a common framework for video retrieval
- Working towards standardisation
- TRECVID: A target initiative for all COST 292 participants
- Participation in all tasks with joint effort of the participants
Rushes Task

• Three main participants:
  – University of Bristol: Key frame layout
  – TU Delft: Shot boundary detector and Affective analysis
  – LABRI: Camera motion detector

• Main idea:
  – Keeping the approach generic
  – Interpreting the most important/representative scenes/frames
  – Presenting the material efficiently.
Introduction

- Domain switching: comic-like video summary
- Automatically generate comic based on a video
- Intuitive video summarisation
- Uses narrative structure of comics
- Limitation: Fixed aspect ratio of a frame
- More important frames should be bigger
Scheme
Key-frame extraction

- Shot boundary information
- Analyses camera work
- Complexity reduction: DC sequence, temporal
- Skims – few frames conveying content of a shot
Camera work

- Essential info for production
- Pan, zoom, tilt
- Location, duration, speed
- Different types of motion depicted on the interface
Summary with Camera Work
Key-frame grouping

- Most representative frames for page/tape
- Highlight unexpected content
- Similar to the scene based grouping of shots
- Layout \textbf{cost function} derived from clusters
- A number of clustering techniques tested
Cost function $C(i)$

- Expresses desired size of a frame on a page $S(i)$
- Layout algorithm minimises overall approximation error $C(i) - S(i)$
- Should set the higher importance to the frames closer to the centre of the cluster and outliers, and lower importance of ones already represented
Panels

- Follow comic narrative structure
- Intuitive and generic
- Fixed aspect ratio – no image cropping
- Panel generator:
  - Left-to-right orientation
  - Top-to-bottom orientation
- Creates a set of panel templates
- For a given normalised row height - XML file
Panel Templates (row heights 1-4)
Layout

• Initially: Full search on combinatorial compositions/partitions
  – Slow, optimal
• Cutting/packing (Knapsack) problems
  – Heuristics, not optimal, unsuitable
• Dynamic programming – Knuth’s line breaking
• Finds sub-optimal solution in linear time
• Generic layout algorithm for 2D displays
I hour tape summary on 1 page
Adaptive formats
TRECVID rushes
Affective Analysis

- Interpreting exciting (important, significant etc...) parts in the video
- Clues: psychologically motivated audio and visual signal futures
  - Motion intensity
  - Sound level
  - Pitch
  - ...
- Tracking these features and guessing which parts are more significant
Affective Analysis

Feature vector \( \vec{f} = [f_1, f_2, \ldots, f_N] \)

Smoothing features
\[
\tilde{f}_i(t) = \frac{1}{T} \cdot \sum_{i=t}^{t+T} f_i(t)
\]

Enhancing peaks using nonlinear transformation
\[
E(t) = \frac{1}{N} \cdot \sum_{i=1}^{N} w_i(t) \cdot \tilde{f}_i(t)
\]

Smoothing excitement curve
\[
\overline{E}(t) = \frac{1}{K} \cdot \sum_{i=t}^{t+K} E(t)
\]
Affective Analysis

• An extra module to the key-frame based interface
• More accessibility and more intelligent browsing
Summary

- System for intuitive summary of large video collections
- Panel generation based on comic book style
- Efficient layout algorithm using dynamic programming
- Generic algorithm for any 2D layout: print, screen, etc.

**TESTING**
- Modules tested individually
- What is the best way to test the system as “editing system”
- Using already edited version!