Video Summarization
Preserving Dynamic Content

Francine Chen  Matthew Cooper  John Adcock

Approach

- Minimize task-specific components
- Handle camera-motion segments separately
- Filter non-interesting segments
- Keep non-redundant segments with motion
- Computationally expensive features need to add significant value
  - Face recognition (OpenCV) was slow, low-accuracy
- Presentation easily understandable
Outline

• Approach
• System
  – Static and dynamic camera
  – Audio clapboard detection
• Evaluation
• Summary and Conclusions

System Overview
Color-based Segmentation

- Inter-frame similarity: chi-squared histogram distance

Motion Features

- OpenCV
  - cvGoodFeaturesToTrack
  - Lucas-Kanade point tracking
- Features
  - histogram of motion magnitude
  - x, y, radial component of low-magnitude vectors
    - mean and variance
Motion Segmentation

- Colorbars
  - few points found
  - little global motion
- Pans and zooms

Segment Classification

- Too short
- Too fast

Finding Redundancies

- Cluster dynamic segments and static segments separately
  - similarity values have different ranges
  - can use different features and similarity measures
  - different weighting for each type of segment during segment selection

Clustering Dynamic Segments

- Color, motion, temporal distance
- PLSA-based dimension reduction
  - color, motion
- Similarity matrix

\[ S(i, j) = \exp\left(-\frac{(d_c(s_i, s_j))^2}{2\sigma_c^2} - \frac{(d_m(s_i, s_j))^2}{2\sigma_m^2} - \frac{(d_t(s_i, s_j))^2}{2\sigma_t^2}\right) \]

Odobez et al., 2003
- Spectral clustering
  - Adaptive threshold to determine number of clusters
Clustering Static Segments

- Features
  - color only
    - temporal info did not make a significant difference
  - average block histogram values
- Agglomerative clustering
- Semi-adaptive threshold to determine number of clusters

System Overview

(input video) -> decode video
  - compute color features
  - compute motion features
  - compute audio features

(color-based segmentation) -> motion-based segmentation

classify dynamic and static segments

cluster static segments

select summary segments

present summary segments

(summary video)
Audio Clapboard Detection

- Loudness features to detect clapboard sounds
  - Simpler than visual detector

Segment Selection

- Too few segments: default model
- Dynamic segments selected first
  - remove segments containing an audio clap
  - for each cluster, select segment representative of cluster and different than previously selected segments
  - max duration of 6 seconds
- Static segments selection
  - remove audio clapboards and singleton clusters
  - order clusters by total duration
  - for each cluster, select segment representative of cluster and different than previously selected segments
  - compute activity score of each segment
  - excerpt 3 second portion with highest activity score
- Stop when maximum summary duration reached or all segments included
Summary Presentation

- Did not use picture-in-picture or tiled display
  - too small, busy
- Rapid playback
  - static scenes played at 1.5x
  - dynamic scenes played at a rate dependent on amount of activity
  - pitch-preserving rate transformation for more natural audio
- Add cues to indicate clip properties and context

Outline

- Approach
- System
  - Static and dynamic camera
  - Audio clapboard detection
- Evaluation
- Summary and Conclusions
Performance

- Performance in middle by inclusion and easiness measures
- Fewer segments than baseline systems
  - longer segment durations
  - lenient filtering of non-interesting segments
- Default model too simple

Camera Motion Performance

- Evaluate on subset of labeled inclusions
  - kept inclusions that mention "pan", "zoom", "tilt"
- Relative performance better
  - almost all other groups perform worse on this subset
Judging Time

- Can judges rapidly understand a summary?
  - total judging time
  - total judging time / summary duration
  - time paused during judging / summary duration

Conclusions

- Handling dynamic segments separately from static segments was a reasonable approach
- Audio clapboard detector worked relatively well and was easy to compute
- Future
  - better default model
  - more detailed analysis of local motion within a segment
  - shorter segments if judges are allowed to pause