Efficient morphological algorithms for video structuring and indexing

Claire-Hélène Demarty & Serge Beucher
Centre de Morphologie Mathématique
Ecole des Mines de Paris
Outline

• Context
• Temporal splitting
  – geometrical transitions
  – chromatic transitions
• Key frame extraction
• Inner shot change detection
• Related shot detection
• Application: newscaster detection
Context

• Wanted: a first *structure* of a video document
• By the use of: *automatic, simple, fast* and *efficient* tools based on *morphological filters*
• Input: all kinds of color, non encoded sequences as video documents
Temporal splitting: geometrical transitions

- Local similarity criterion (color distance) ⇒ no loss of spatial information
- Morphological filtering ⇒ efficient extraction of peaks
Temporal splitting: geometrical transitions

Extracts particular shape of 1D signals
Results

• Parameters:
  – a transition threshold (same value 0.2)
  – block sizes

• less than real time

On 22 video documents (274 cuts):
  – 99.5% of correct detections
  – 3.7% of false alarms

false alarms by relation detection
Temporal splitting: 
transition mask

• Use of a local criterion to keep track of the transition geometry
• Study of the union of the binary transition masks:
  – Morphological filtering
  – Computation of the temporal evolution curves of simple measures on the mask
  – Correlation between these and precomputed curves of ideal transition models
Temporal splitting:

transition mask

⇒ Wipe from left to right
Temporal splitting: chromatic transitions

- Pixel-to-pixel criterion (number of pixels with a non-zero color difference)
- Hierarchical morphological filtering by successive erosions and top hat
  ⇒ duration, beginning, end
Results

• Parameters:
  – a transition threshold (same value 0.2)

• On 22 video documents (23 chromatic transitions):
  – 78.3% of correct detection
  – 65.4% of false alarms

• less than real time

▷ false alarms by relation detection

Side effect: detection of inner camera motion
Results
Key frame extraction

- First and last images
- Second hierarchy of peaks
- Information redundancy? (~35%)
**Inner shot change detection**

- Similarity between selected key frames of each shot
- Comparison to a change threshold
  - (1.5 times the transition threshold)
- ➔ 9% of redundancy
- ➔ From 2.1% to 1.4% of key frames for a given sequence
Related shot detection

- Similarity between key frames of different shots
- Comparison to a relation threshold (1.5 times the transition threshold)
- False alarm correction
- Flash detection
- Establishment of a relation graph
- Extraction of interview sequence
Related shot detection: correction of false transitions

- $S_i \not{\xrightarrow{}} S_{i+2}$
- no change in $S_{i+1}$
- dissolve: 65.4% to 33.3%
- cut: 3.7% to 3%
Related shot detection: flash extraction

- $S_i \Rightarrow S_{i+2}$
- only one frame
- increase of luminance
Related shot detection: relation graph and extraction of interview

Document
Interview

C1
S0
S2
C3
C5
S4
S6
C7
S8

25/10/99
Demarty - CMM - ARMINES - CCETT

copyright©A2/CMM/ENSMP
Application to newscaster detection

• Computation of groups of related shots
• Four criteria:
  – one connected component with the specific skin color, of a certain size and more or less in the middle of the frame
  – non moving background
  – “maximal” group
  – shots at the beginning and at the end of the document
• Fusion of the criteria by a simple mean
Application to newscaster detection

Final probability = 88.1%
Conclusion

• Two parameters only (block sizes and transition threshold)
• Automatic, simple, fast (less than real time) and efficient
• First efficient structure of a video document
• Already gives access to high level information
• Starting point for more sophisticated indexing