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

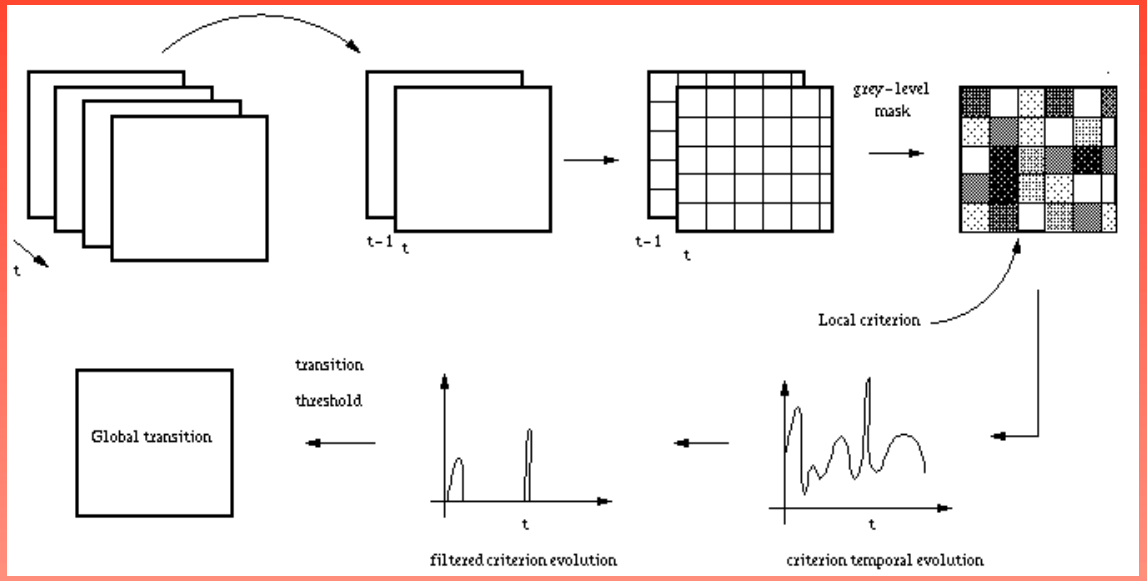
  based on ***morphological filters***

- Input:

  all kinds of color, non encoded sequences

  as video documents

Temporal splitting: geometrical transitions



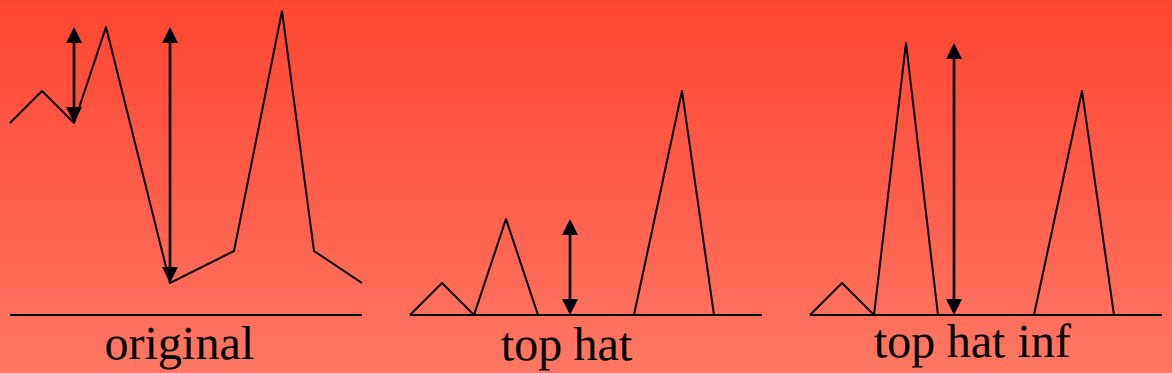
- Local similarity criterion (color distance)

  $\Rightarrow$ *no loss of spatial information*

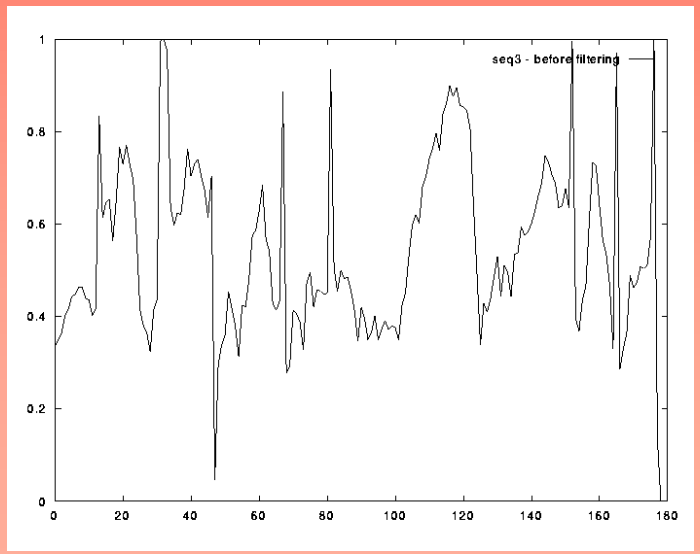
- Morphological filtering

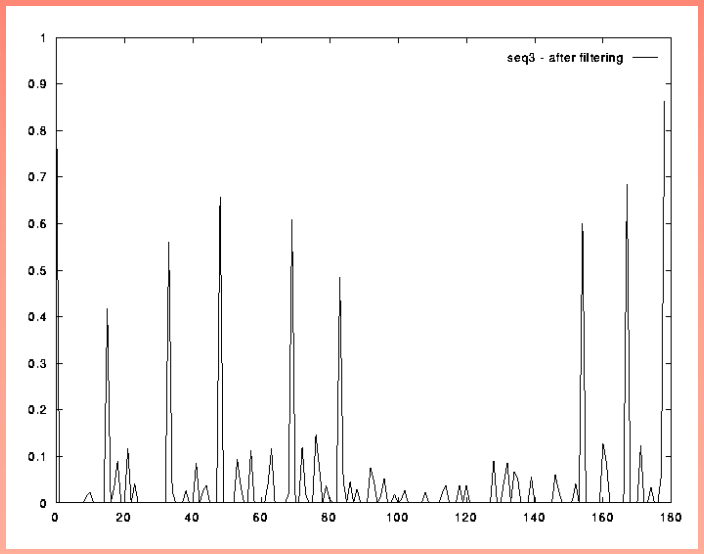
  $\Rightarrow$ *efficient extraction of peaks*

Temporal splitting: geometrical transitions

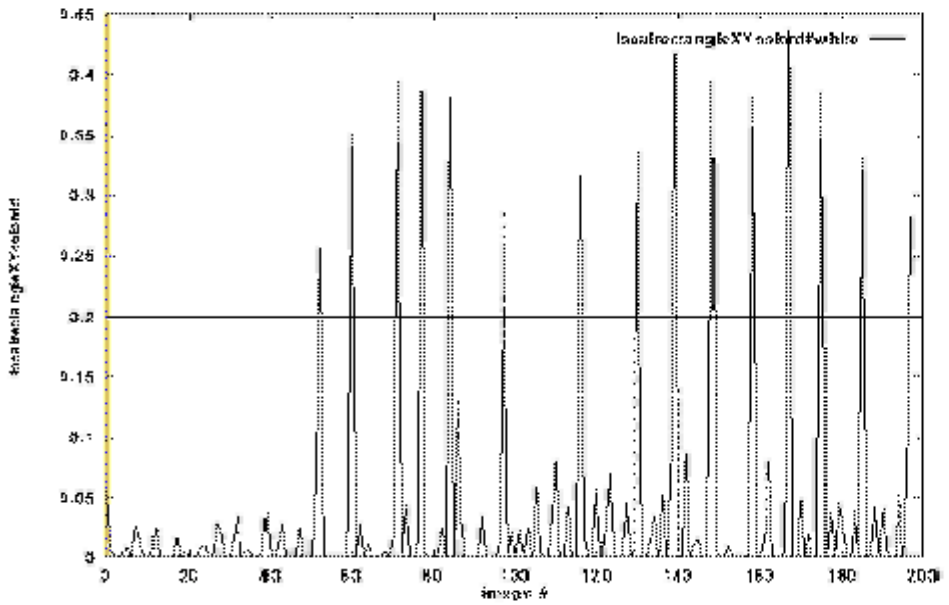


➔Extracts particular shape of 1D signals





Results



- **Parameters:**
  - a transition threshold (same value 0.2)
  - block sizes
- **less than real time**

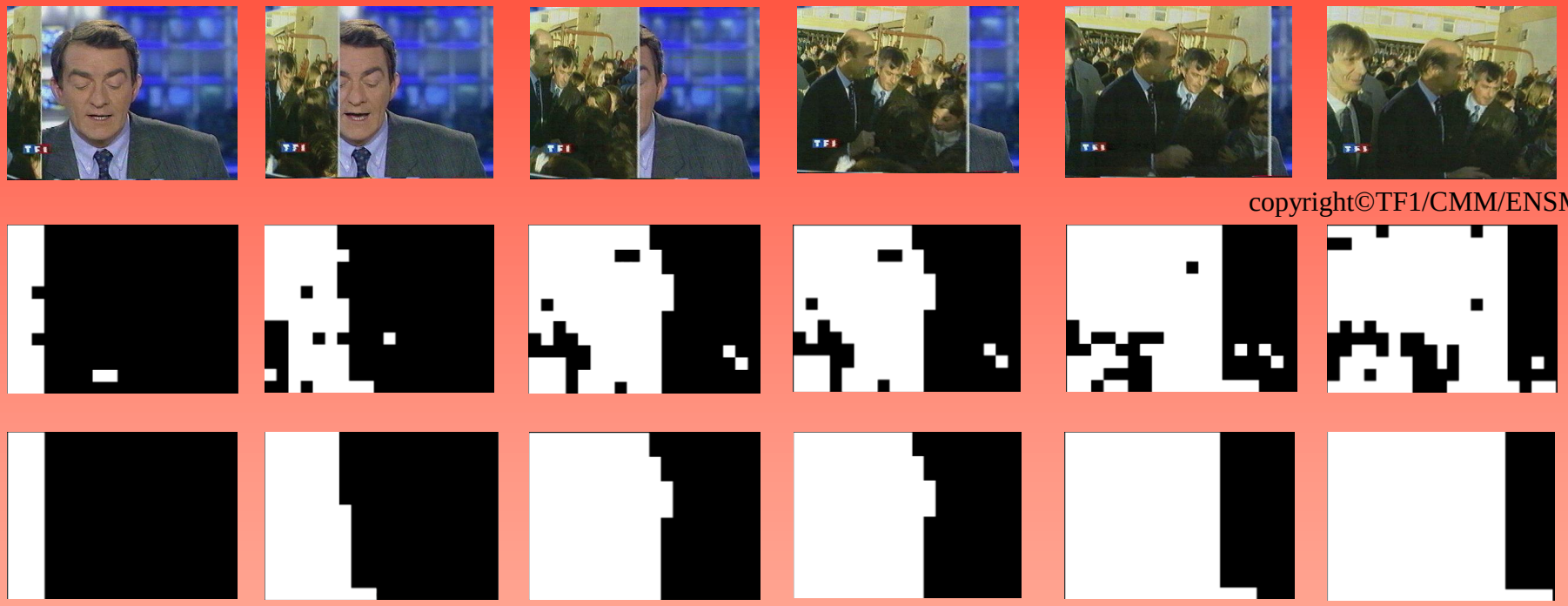
- **On 22 video documents (274 cuts):**
  - 99.5% of correct detections
  - 3.7% of false alarms

⇘ 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

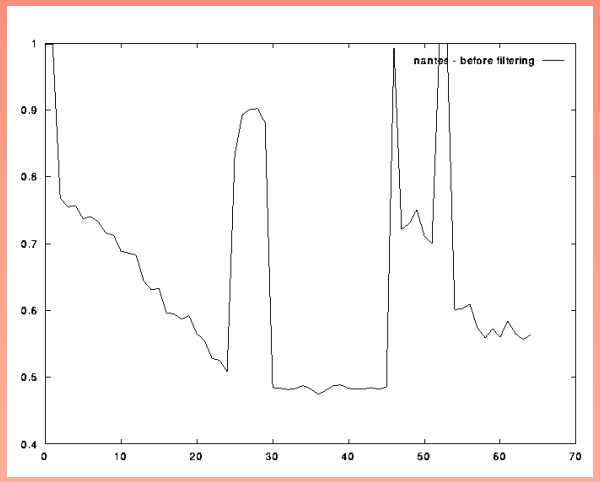


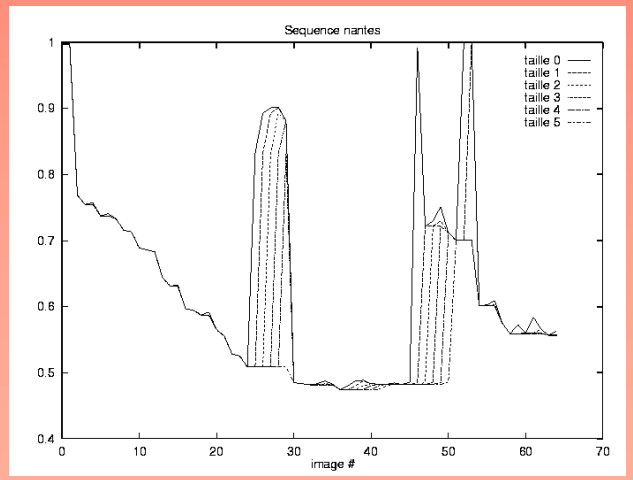
⇒ Wipe from left to right

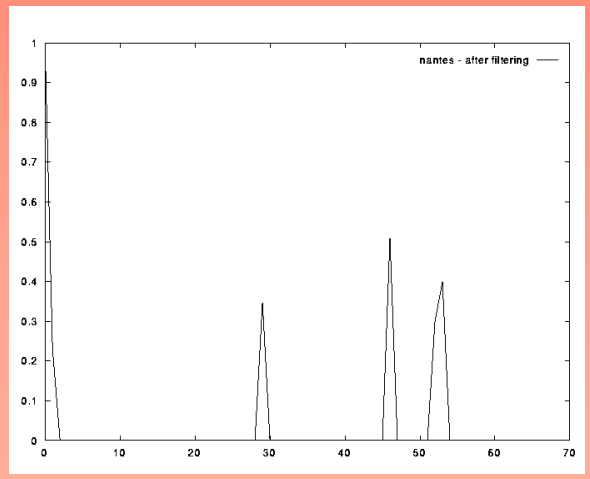
Temporal splitting: chromatic transitions

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

  $\Rightarrow$ duration, beginning, end







Results

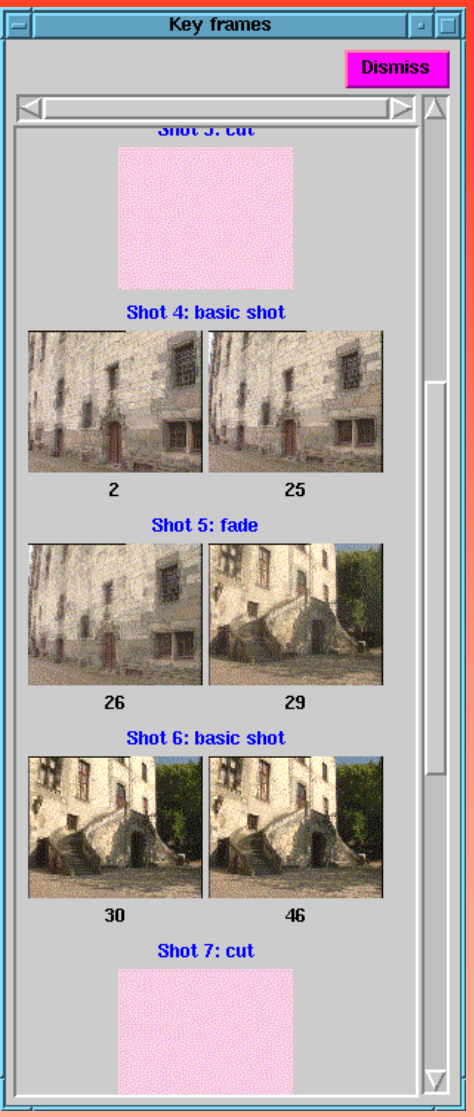
- **Parameters:**
  - a transition threshold (same value 0.2)

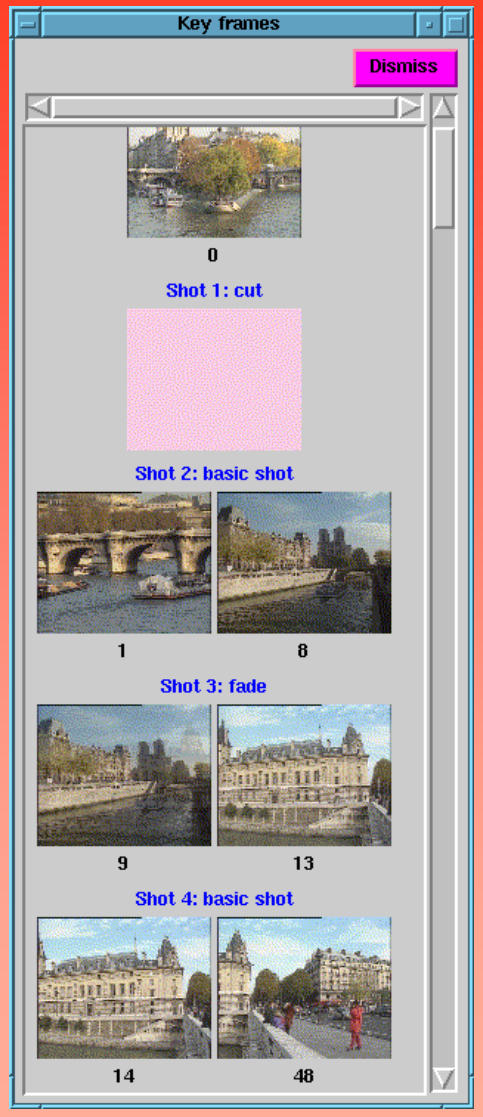
- **On 22 video documents (23 chromatic transitions):**
  - 78.3% of correct detection
  - **65.4%** of false alarms
- **less than real time**

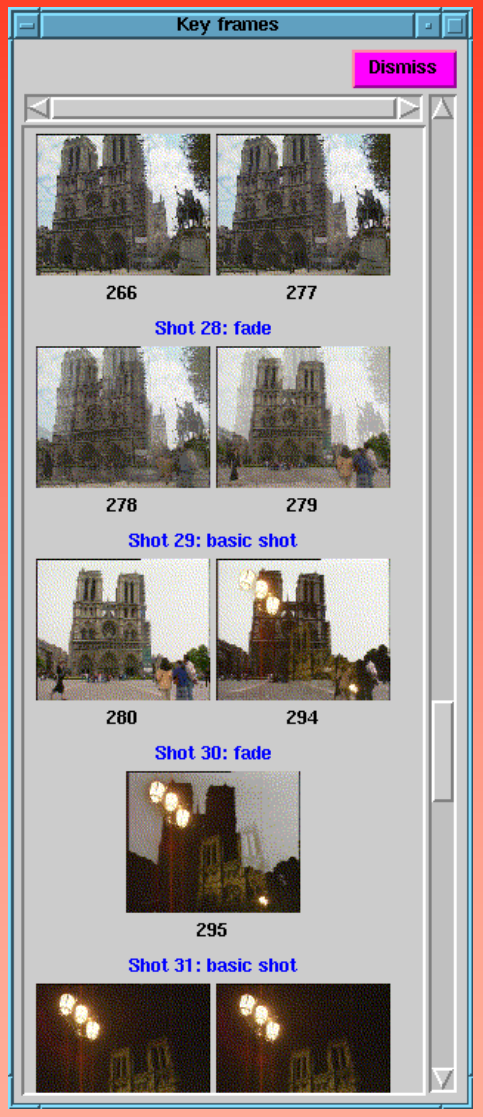
⇘ **false alarms by relation detection**

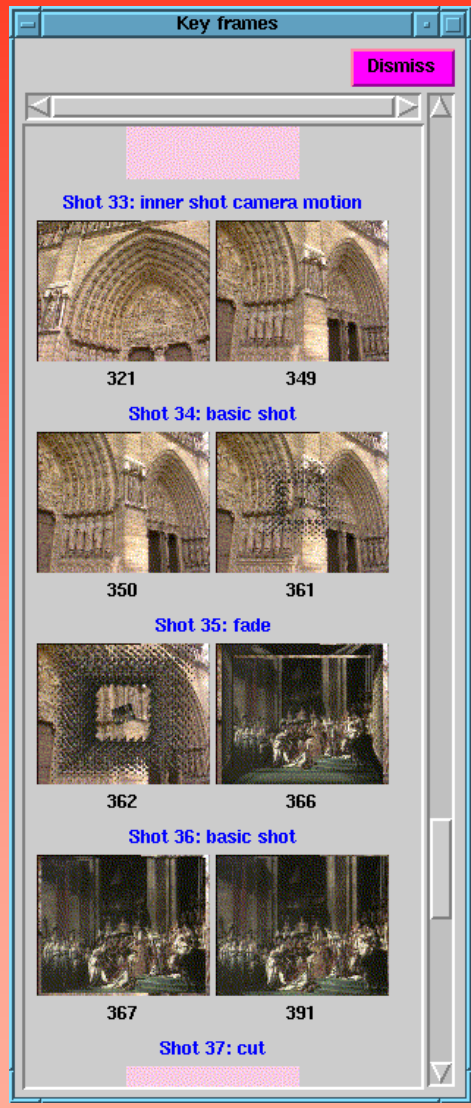
Side effect: detection of inner camera motion

Results


Key frames
Dismiss
Shot 4: basic shot
2
25
Shot 5: fade
26
29
Shot 6: basic shot
30
46
Shot 7: cut

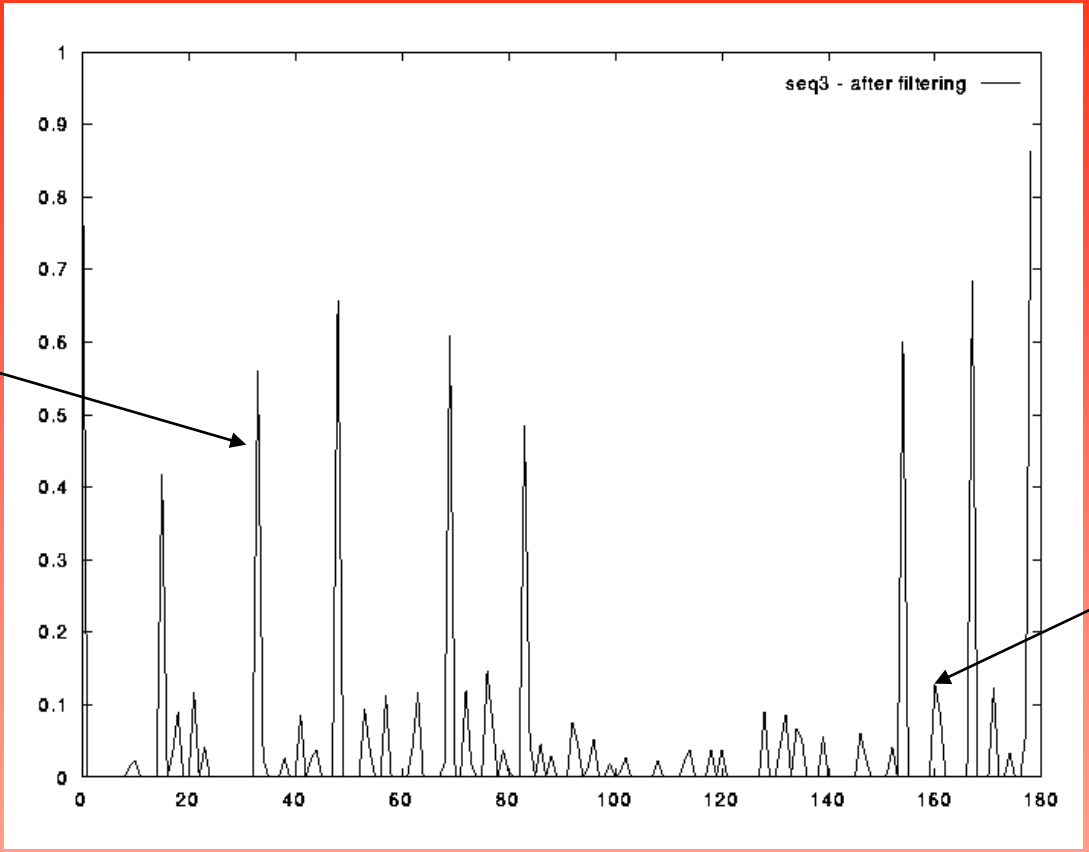

Key frames
Dismiss
0
Shot 1: cut
Shot 2: basic shot
1
8
Shot 3: fade
9
13
Shot 4: basic shot
14
48


Key frames
Dismiss
266
277
Shot 28: fade
278
279
Shot 29: basic shot
280
294
Shot 30: fade
295
Shot 31: basic shot


Key frames
Dismiss
Shot 33: inner shot camera motion
321
349
Shot 34: basic shot
350
361
Shot 35: fade
362
366
Shot 36: basic shot
367
391
Shot 37: cut

Key frame extraction

First hierarchy:
Transitions



Second hierarchy:
Key frames

- First and last images
- Second hierarchy of peaks
- Information redundancy ? (~35%)

Inner shot change detection

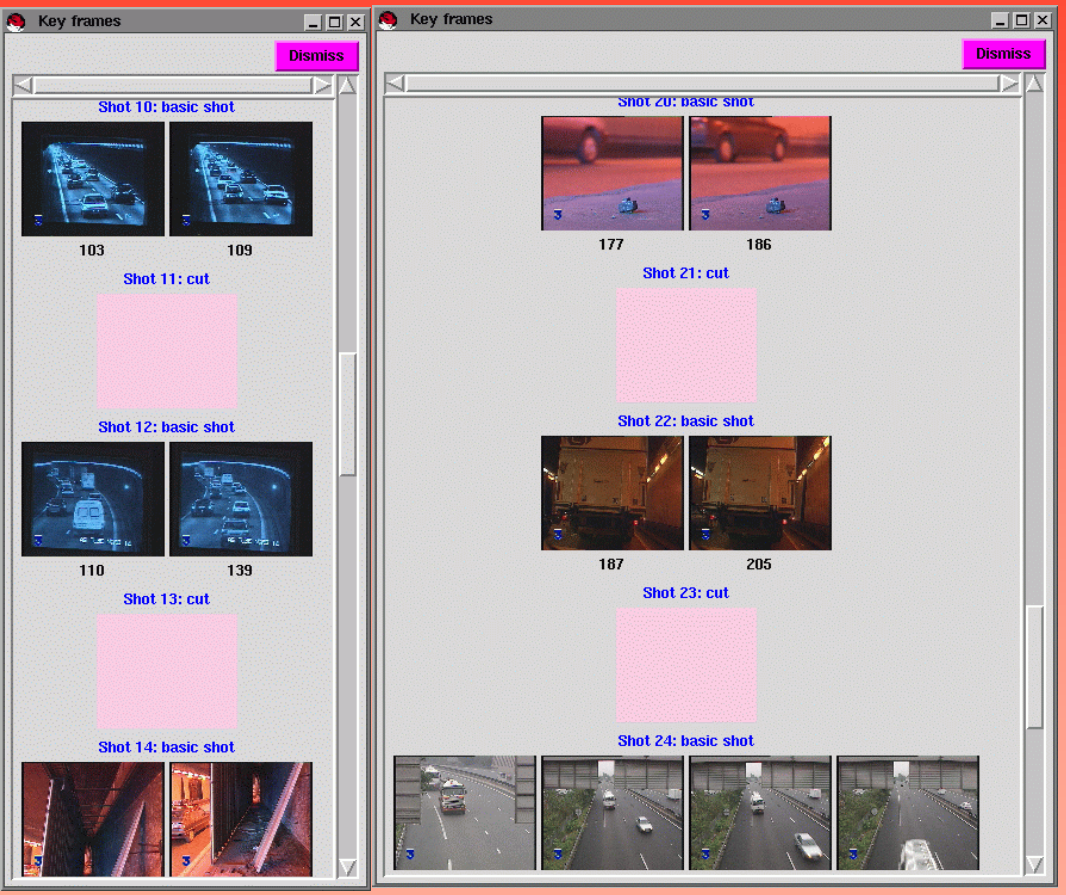
- Similarity between selected key frames of each shot
- Comparison to a change threshold

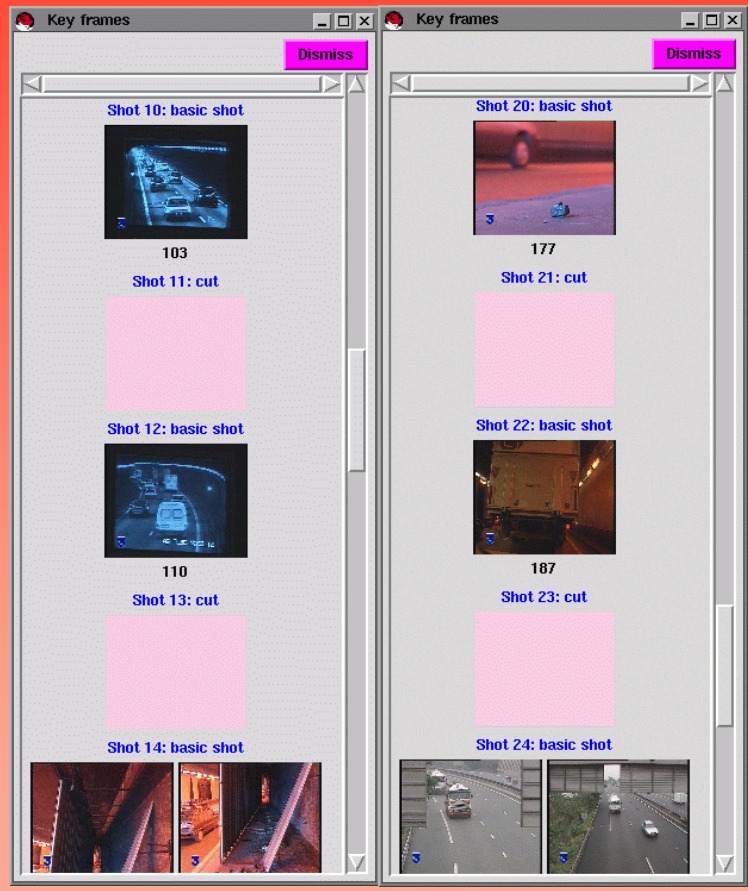
  (1.5 times the transition threshold)

➔ 9% of redundancy

➔ From 2.1% to 1.4% of key frames for a given sequence

Inner shot change detection



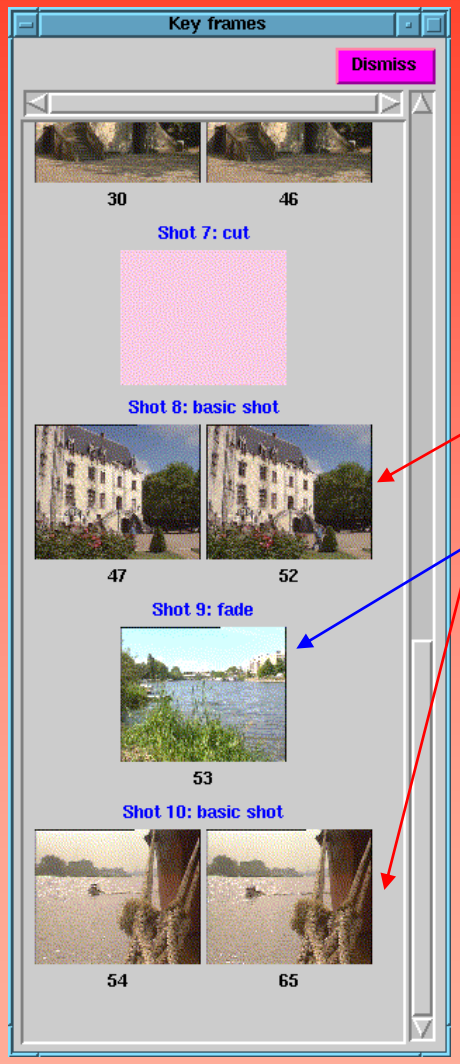


Related shot detection

- Similarity between key frames of different shots
- Comparison to a relation threshold
  (1.5 times the transition threshold)

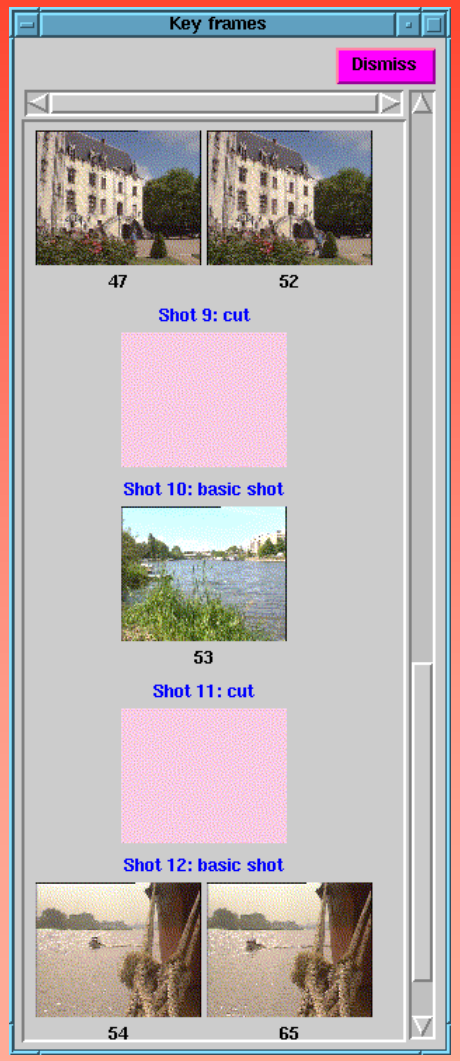
- False alarm correction
- Flash detection
- Establishment of a relation graph
- Extraction of interview sequence

Related shot detection: correction of false transitions

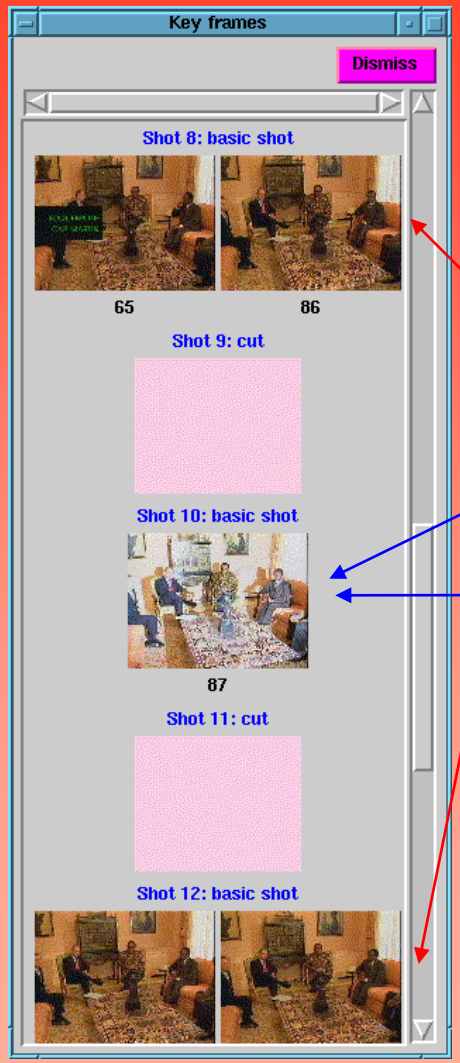


- $S_i \not\Re S_{i+2}$
- no change in S_{i+1}

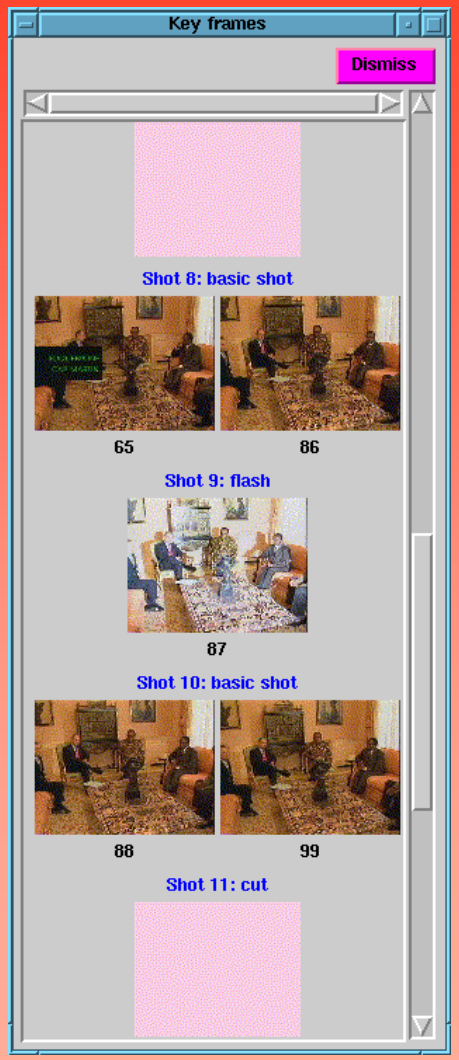
- dissolve: 65.4% to 33.3%
- cut: 3.7% to 3%



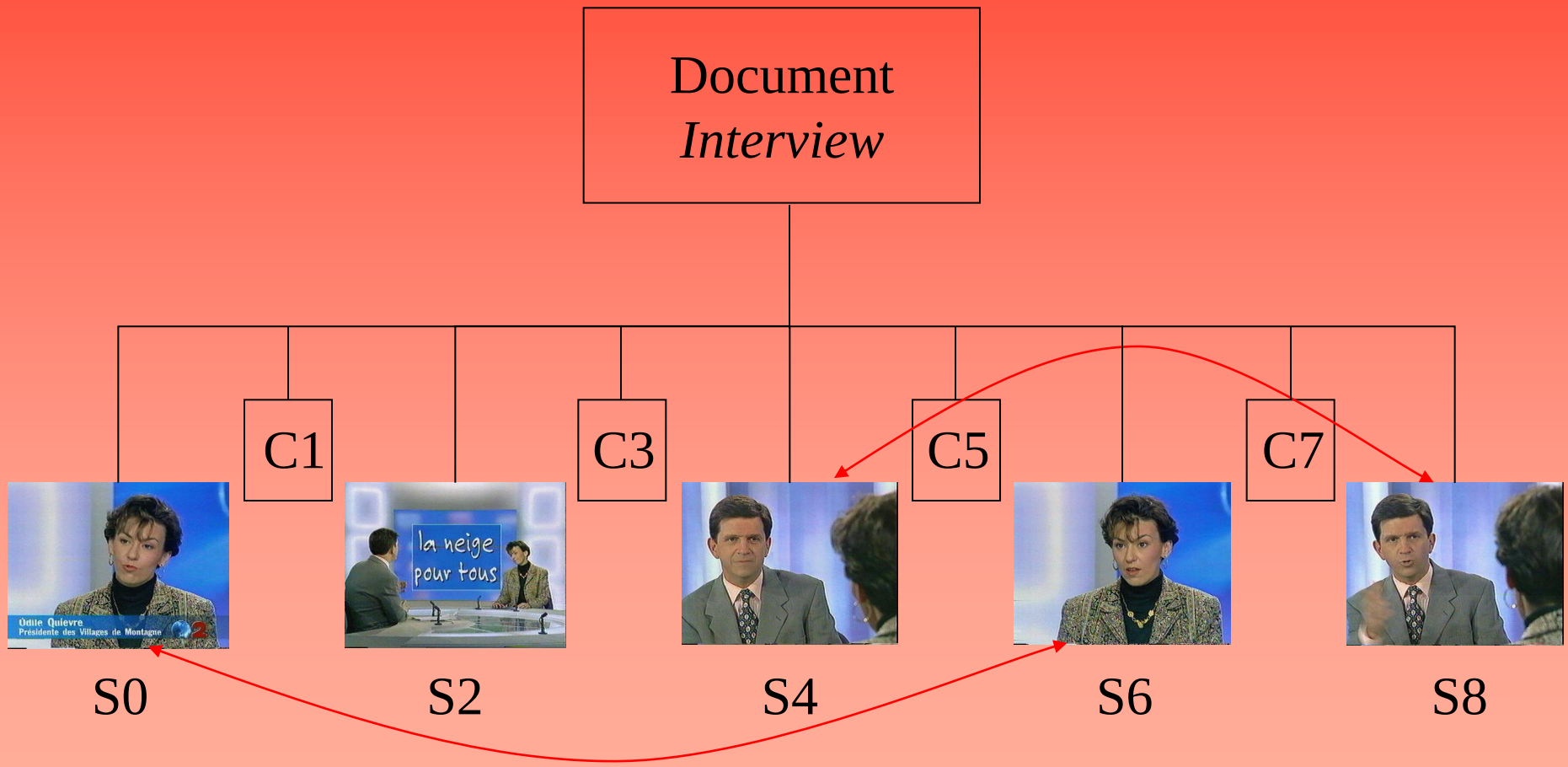
Related shot detection: flash extraction



- $S_i \Re S_{i+2}$
- only one frame
- increase of luminance



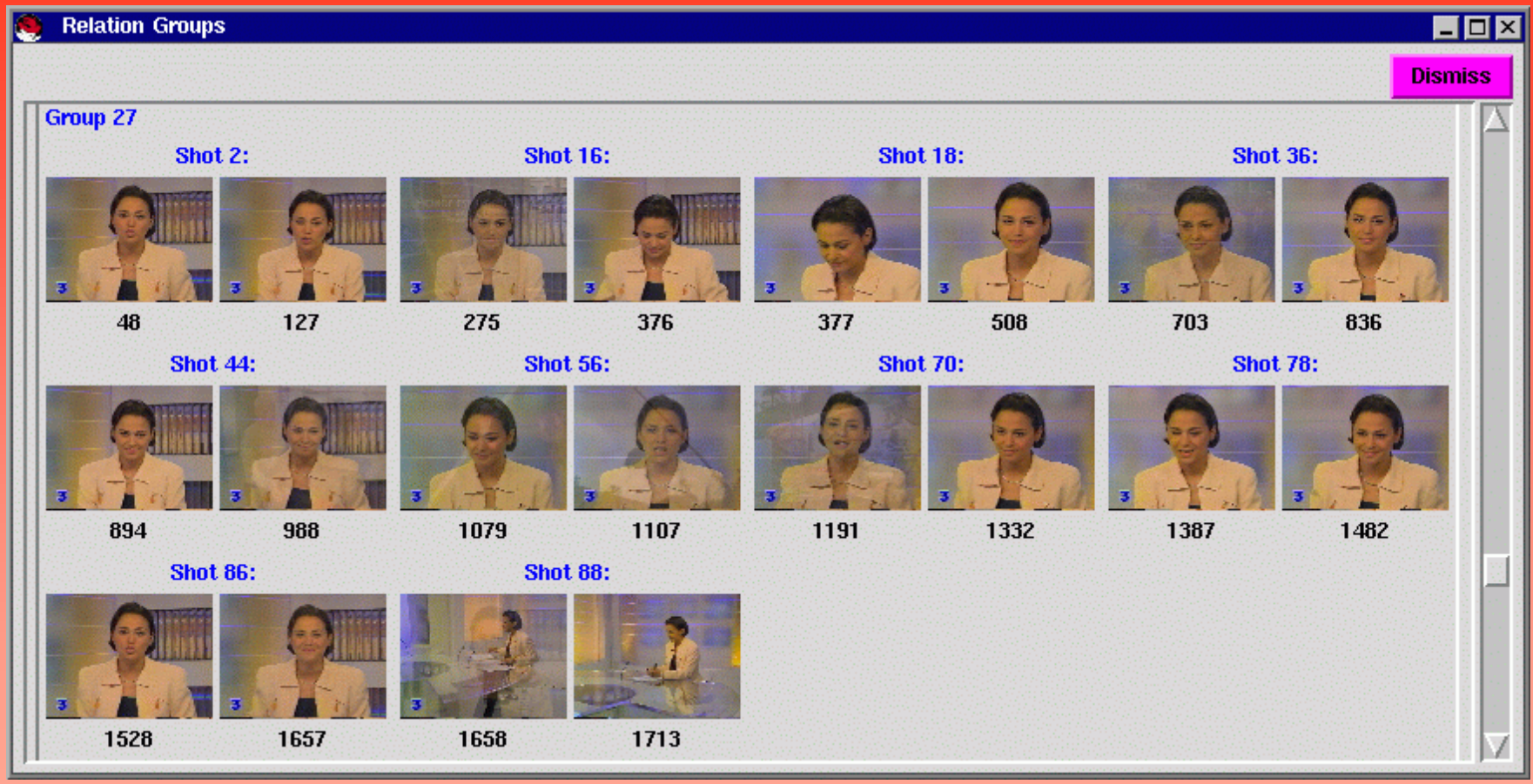
Related shot detection: relation graph and extraction of interview



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