

## USER EVALUATION OF MULTI-EPIISODES VIDEO SUMMARIES

Itheri Yahiaoui, Bernard Merialdo, Benoit Huet  
*Multimedia Communication Department  
Institut Eurécom  
2229 Route des Crêtes,  
06904 Sophia-Antipolis, France  
{itheri.yahiaoui, bernard.merialdo, benoit.huet}@eurecom.fr*

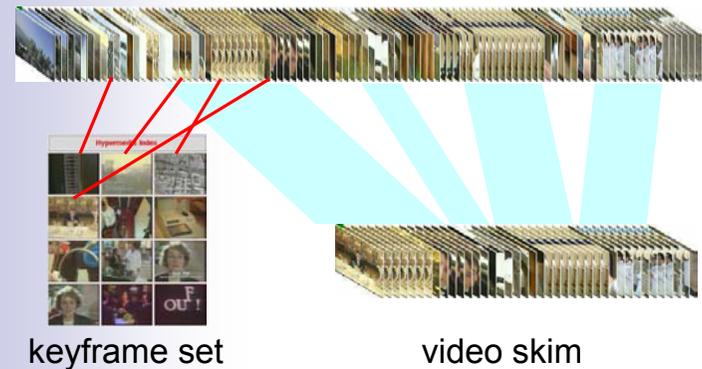
## Content

- Video summaries
- Multi-Episode Video Summaries
- Optimal summaries
- Maximal Recall Automatic Summarization
- Experiments
- User Evaluation
- Conclusion and Future work.

## Video Summaries

- A summary is a subset of the video
  - Identify important information
  - Constrained duration
- A summary can be good or bad
  - Depends on task
    - Movie Trailer, Informative or Descriptive, etc...
  - Quality is generally difficult to evaluate

## Video Summary format



## Multi-Episode Summaries

- Independently created summaries may contain redundant information
- Specific requirements to construct multi-episode summaries
- Identification of :
  - What is common to several episodes
  - What is specific (unique) to each episode
- Typical applications
  - TV series, Set-Top-Box, etc...

## Optimal Summaries

- What is the best summary for a video?
- Many proposals, two basic approaches:
  - User-based evaluation (qualitative)
    - Smith and Kanade [CBAIVL 1998]  
*Infomedia Project: video skims.*
  - Mathematical criterion (quantitative)
    - Gong and Liu [ICME 2000]  
*Use of SVD over a feature frame matrix.*
    - Uchihashi and Foote [ICASSP 1999]  
*Definition of a shot importance measure.*

## Ideal summary evaluation

- User  $u$  without summary performs task  $T$ :
  - performance  $p_T(u)$
- User  $u$  with summary  $S$  performs task  $T$ :
  - performance  $p_T(u | S)$
- Ideal summary efficiency:
  - $\text{average}( p_T(u | S) - p_T(u) )$
- But:
  - users are different (many users required)
  - users learn (cannot compute  $p_T(u | S)$  after  $p_T(u)$ )
  - evaluation is very expensive (often not feasible)

## Maximal Recall Task

- Idea: Identify a movie from a picture from a magazine
- Formalization:
  - User  $u$  knows summaries  $S_i$  of video  $V_i$
  - User  $u$  is shown an excerpt  $E$  (from video  $V_j$ )
  - User  $u$  is asked to guess  $j$
- Optimal summaries:
  - Should maximize the performance over all  $E$
  - Evaluation can be automated if the behavior of  $u$  can be reasonably simulated

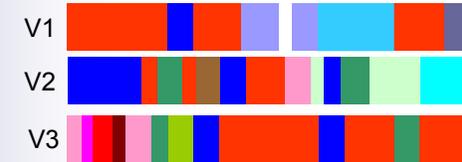
## Maximal Visual Recall

- User chooses video  $j$  if (s)he recognizes similar images in both excerpt  $E$  and summary  $S_j$
- In case of ambiguity: no decision
- This process can be automated based on similarity measure
- Similarity based on color histograms



## Intuitive Idea

- Consider videos:



- red frame is good for  $V_1$ , but will generate ambiguities with  $V_2$  and  $V_3$
- Summary should contain frames:
  - frequent in one video
  - unfrequent in others

## Evaluation Criterion

### ■ User Performance

- Number of excerpts with correct unambiguous answers

$$\text{Card} \left\{ (i, v) : \exists j \quad f_j \in E_i^v \exists f_m \text{ similar to } f_j \text{ and } f_m \in S_v, \right. \\ \left. \forall v' \neq v \forall f_j \in E_i^{v'} \forall f_m \text{ similar to } f_j \quad f_m \notin S_{v'} \right\}$$

- Computed using all excerpts of fixed duration  $d$  from all the videos

- Note: performance vary with  $d$ .

## Summary construction

### ■ Iterative process

- Greedy algorithm
- Selection based on frame coverage

### ■ In-place refinement

- Try to replace each frame individually to improve quality
- Repeat until no change

				V1
				V2
				V3
				V4
				V5
				V6

## Experiments

- Six episodes from the TV serie «Friends»
- Total videos duration 83150 frames ( $\approx$  99 min)
- Summary of **six key-frames per video**
- Key-frames are selected according to method described earlier
- Video processing
  - Elimination of jingle and credits
  - Feature Vectors construction

## Video Summary Evaluation

- Issue: Evaluation
  - Two opposite approaches
    - User based evaluation: difficult to set-up, possible bias, ...)
    - Mathematical criterion: (easy to set-up, difficult to interpret)
- Simulation of user behavior based on Maximal Recall
- Real experimentation
  - User simulated performance measure
  - Limitation of image similarity measure
  - Single and Multi-episode videos

## Video Summary Evaluation

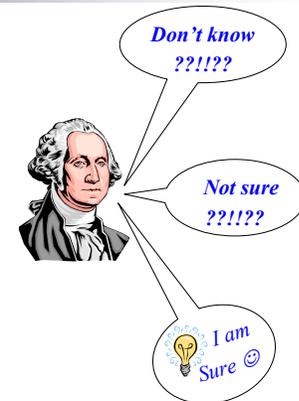
Experiment based on visual recall capabilities

- Show summaries  $S_1, \dots, S_k$  of videos  $V_1, \dots, V_k$  to the user
  - For example a grid of images, where each line represent a Video
- Show an excerpt  $E$  of a video  $V_i$  to the user, then ask the user to guess  $i$



## Video Summary Evaluation

- User answers:
  - Don't Know
    - Unknown case when no similar image between  $E$  and any summary  $S_i$
  - Confused
    - Ambiguous case when similar images between  $E$  and summaries  $S_i$  and  $S_j$
  - Sure
    - Unambiguous case when similar images between  $E$  and a single summary  $S_i$

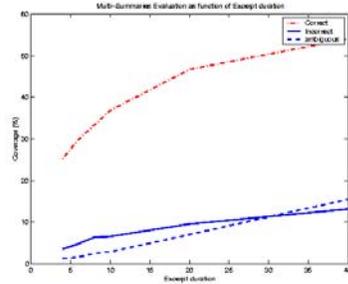


# Experimental results

## Coverage over the original videos

Excerpt duration	% correct	% ambiguous	% incorrect
4 sec	25.25	1.27	3.53
6 sec	29.87	1.61	4.79
8 sec	33.36	2.51	6.38
10 sec	36.82	2.86	6.54
20 sec	46.70	7.02	9.54
40 sec	54.06	15.47	13.14

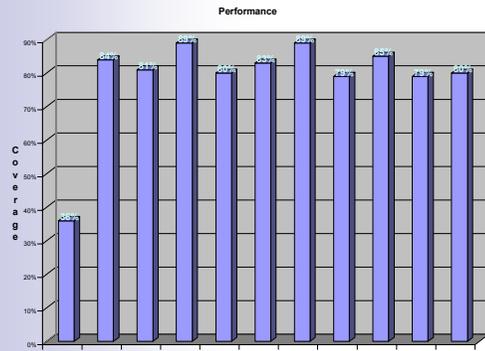
## Evaluation of summaries



# Evaluation Results Analysis

- Idea: Look precisely at the difference between the system's evaluation method and the user's answers.
  - Count the number of correct and wrong answers
  - Discuss the reason of the choice made by the users
  - Results based on 100 excerpts for 10 users

# Evaluation Results

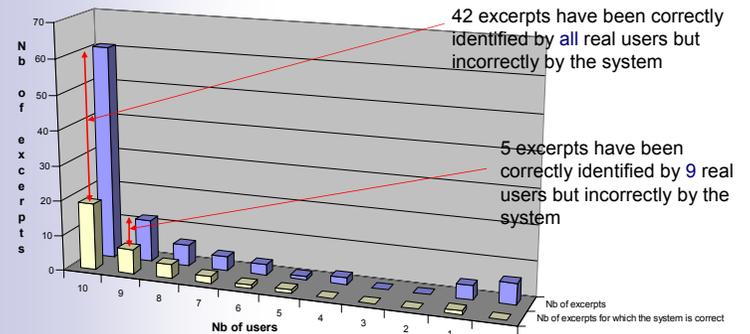


Average Real User Performance = 82.9%

People with high score (89%) are fan of the serie

What makes our Simulated User perform so poorly (36%) ?

# Evaluation Results



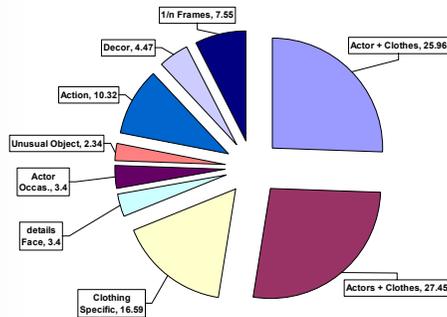
□ Nb of excerpts for which the system is correct	10	9	8	7	6	5	4	3	2	1	0
■ Nb of excerpts	19	7	4	2	1	1	0	0	0	1	0

## Results Analysis

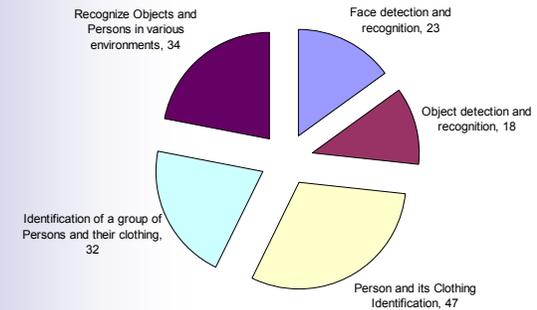
- Objective: Improve the performance of our automatic summarization scheme

- Major factors:

- Person,
- Object,
- Action,
- Location,
- Time



## Improvements



Nb of excerpts for which the system could be correct depending on methodology employed.  
(out of the 47 incorrectly identified excerpts)

## Conclusion

- Novel approach to automated video summary creation (inc. Multi-Video case)
- New method for evaluation
  - Use of Maximal Recall
  - Performance levels are easy to understand
- New method for summary creation
  - Suboptimal automatic construction
  - Summary duration is user definable
- Work on Region Matching/Recognition