A Relevance Feedback Approach to Video Genre Retrieval

Ionut MIRONICA[1]	Constantin VERTAN[1]	Bogdan IONESCU[1,2]

[1]LAPI – University Politehnica of Bucharest, Romania
[52]LISTIC – Université de Savoie, Annecy, 74944 France

IEEE International Conference on Intelligent Computer Communication and Processing - Cluj Napoca – ICCP 2011

Agenda

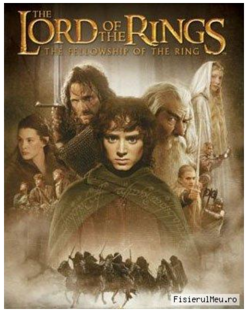
- Introduction
- Relevance Feedback Algorithms
- Hierarchical Clustering Relevance Feedback
- Implementation and Evaluation
- Conclusion

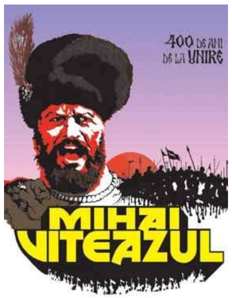
I. Introduction

Concepts

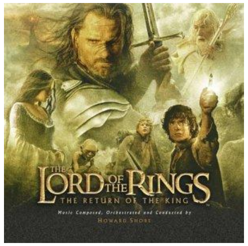
- **Content Based Video Retrieval**
- **Query by Example**

Query Results



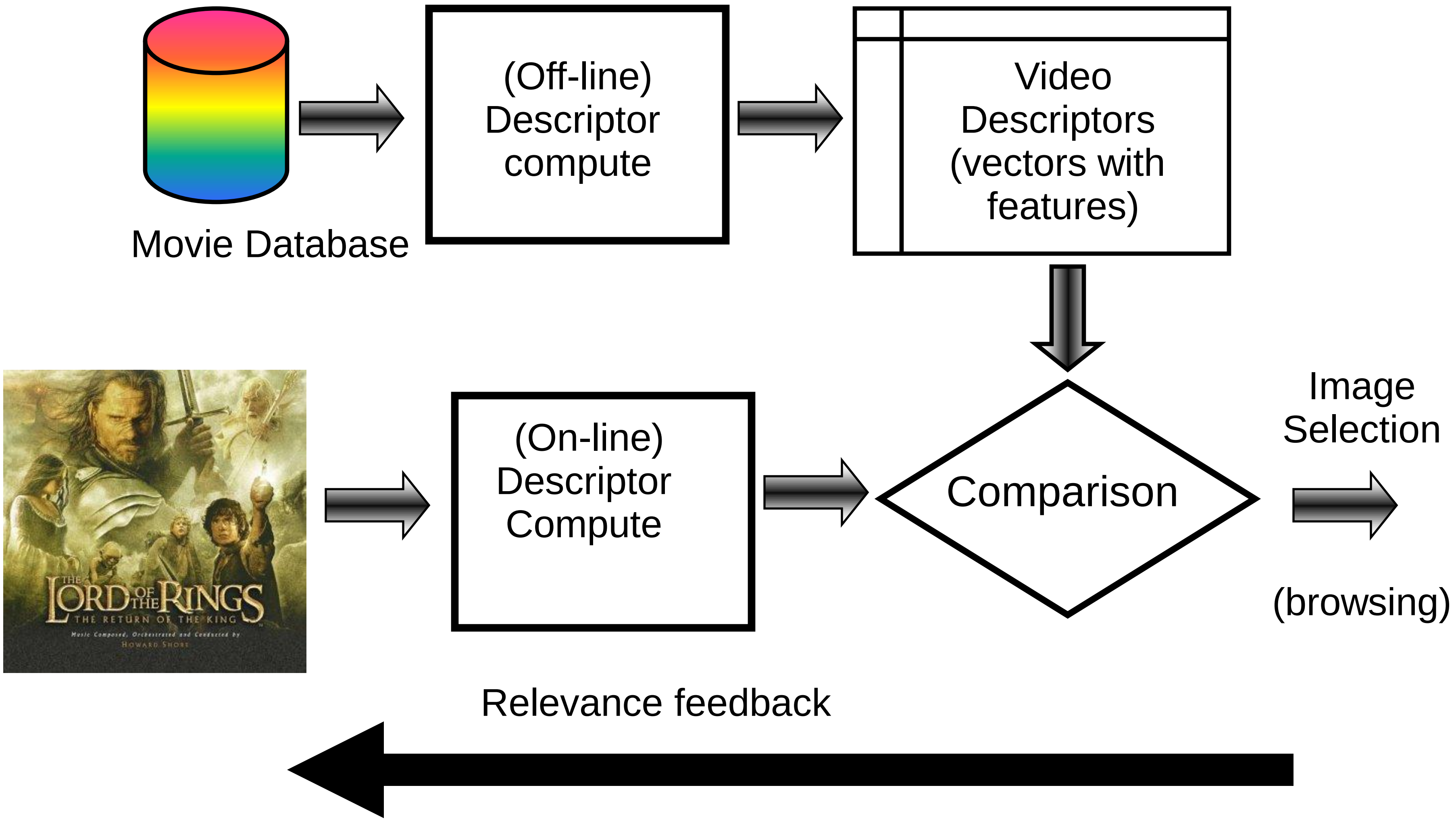


Query Database



Movie Sample

II. CBVR System



II. Color Descriptors

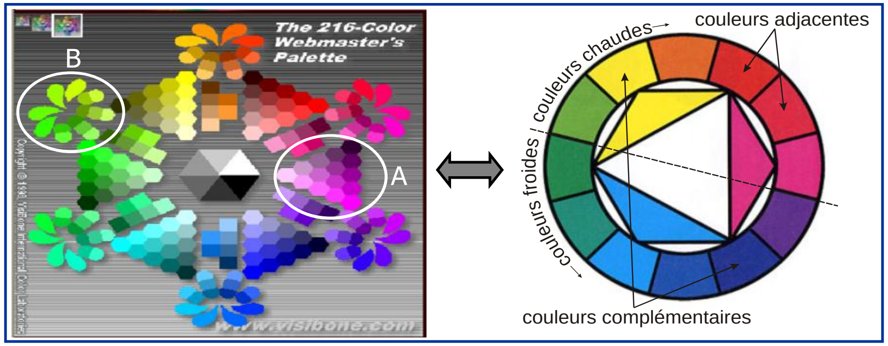
Objective: capture movie's global color contents in terms of *color distribution*, *elementary hues*, *color properties* and *color relationship*;

Global weighted histogram:

$$h_{GW}(c) = \sum_{i=0}^{M} \left[\frac{1}{N_i} \cdot \sum_{j=0}^{N_i} h_{shot_i}^{j}(c) \right] \cdot \omega_i$$

Elementary color histogram:

$$h_E(c_e) = \sum_{c=0}^{215} h_{GW}(c) \Big|_{Name(c_e) \subset Name(c)}$$



Webmaster 216 colors ⟺ Itten's color wheel

[B. Ionescu, D. Coquin, P. Lambert, V. Buzuloiu'08]

II. Color Descriptors

Color properties

$$P_{light} = \sum_{c=0}^{215} h_{GW}(c)\Big|_{W_{light} \subset Name(c)}$$: amount of bright colors in the movie, $W_{light} \in \{light, pale, white\}$;

P_{dark} : amount of dark colors in the movie, $W_{dark} \in \{dark, obscure, black\}$;

P_{hard} : amount of saturated colors, $W_{hard} \in \{hard, faded\} \cup$ elem.;

P_{weak} : amount of low saturated colors, $W_{weak} \in \{weak, dull\}$;

P_{warm} : amount of warm colors, $W_{warm} \in$ {Yellow, Orange, Red, Yellow-Orange, Red-Orange, Red-Violet, Magenta, Pink, Spring};

P_{cold} : amount of cold colors, $W_{cold} \in$ {Green, Blue, Violet, Yellow-Green, Blue-Green, Blue-Violet, Teal, Cyan, Azure};

II. Action Descriptors

Objective: capture movie's temporal structure in terms of *visual rhythm*, *action* and *gradual transition* %;

Gradual transition %: $GT = \frac{T_{dissolve} + T_{fade-in} + T_{fade-out}}{T_{total}}$

Rhythm: capture the movie's changing tempo

$\xi_{T=5s}(i)$: relative number of shot changes within time window *T* starting from frame at time index *i*;

$\bar{v}_{T=5s} = E\{\xi_{T=5s}(i)\}$: movie's average shot change speed;

Action: in general related to a high frequency of shot changes;

→ "hot action" → "low action"

[B. Ionescu, L. Ott, P. Lambert, D. Coquin, A. Pacureanu, V. Buzuloiu'10]

II. Contour Descriptors

Objective: describe structural information in terms of contours and their relations;

Contour properties:

b : degree of curvature (proportional to the maximum amplitude of the bowness space); – straight vs. bow

ζ : degree of circularity; – ½ circle vs. full circle

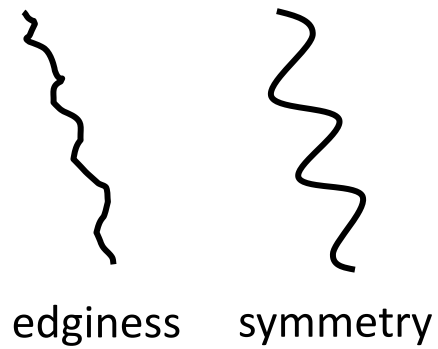
e : edginess parameter – zig-zag vs. sinusoid;

y : symmetry parameter – irregular vs. "even"

+ Appearance parameters:

c_m, c_s: mean, std.dev. of intensity along the contour;

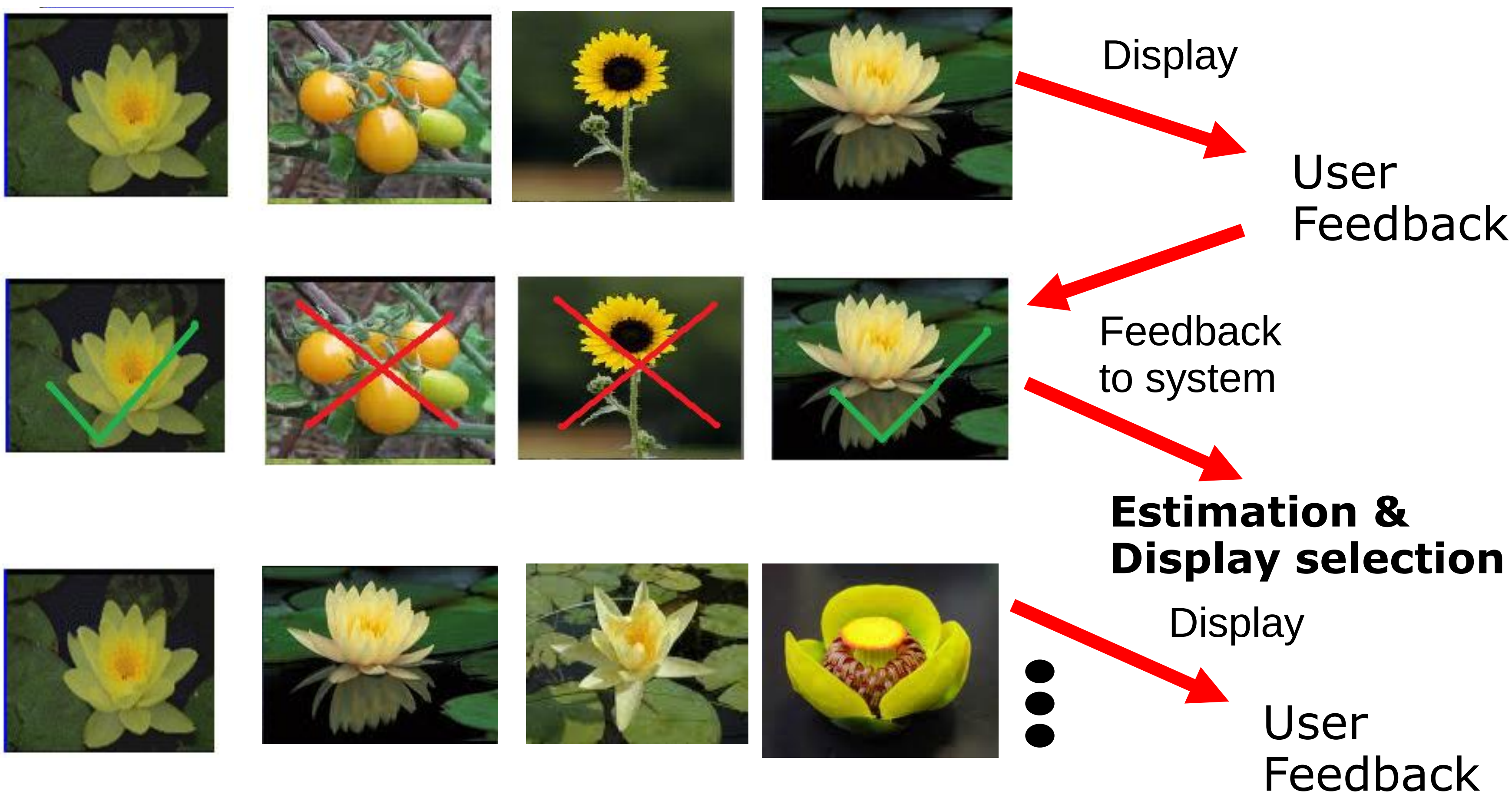
f_m, f_s: fuzziness, obtained from a blob (DOG) filter: I * DOG



[IJCV, C. Rasche'10]

III. Semantic Gap and Relevance Feedback

- **Relevance feedback** uses positive and negative examples provided by the user to improve the system's performance.



III. Relevance Feedback

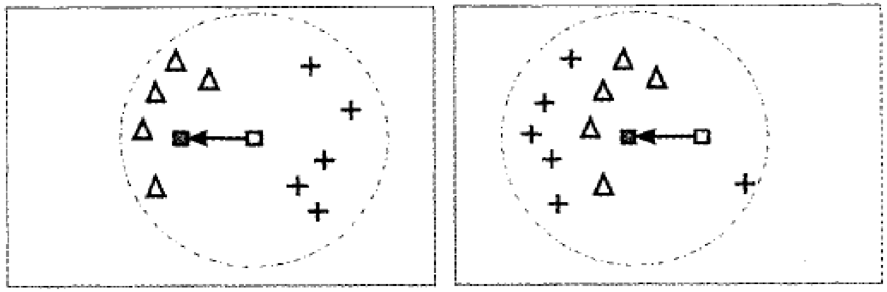
Algorithms:

- Rocchio Algorithm
- Feature Relevance Estimation (FRE)
- SVM – Relevance Feedback
- Hierarchical Clustering Relevance Feedback

III. Rocchio Algorithm

- Uses a set R of relevant documents and a set N of non-relevant documents, selected in the user relevance feedback phase, and updates the query feature.

$$Q' = \alpha Q + \frac{\beta}{|R|}\sum_{R_i \in R} R_i - \frac{\gamma}{|N|}\sum_{N_i \in N} N_i$$

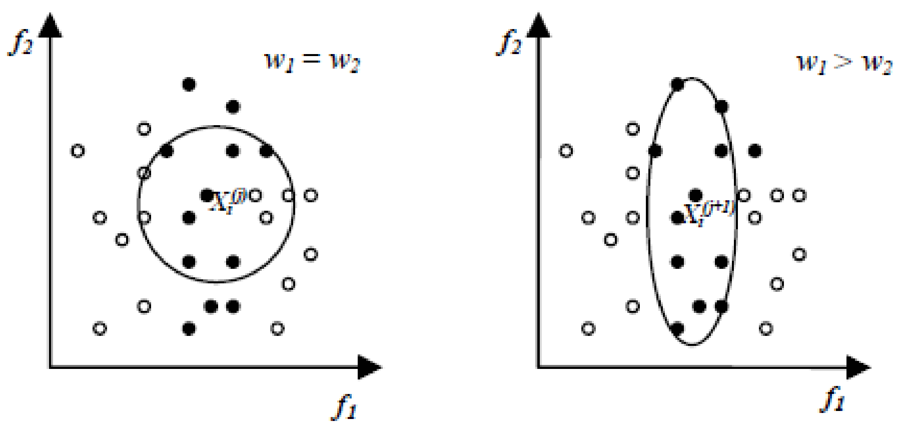


Δ *documents marked as relevant*
□ *query*
+ *other returned documents*
documents not returned

III. Feature Relevance Estimation (FRE)

- some specific features may be more important than other features
- every feature will have an importance weight that will be computed as Wi = 1/σ, where σ denotes the variance of video features.

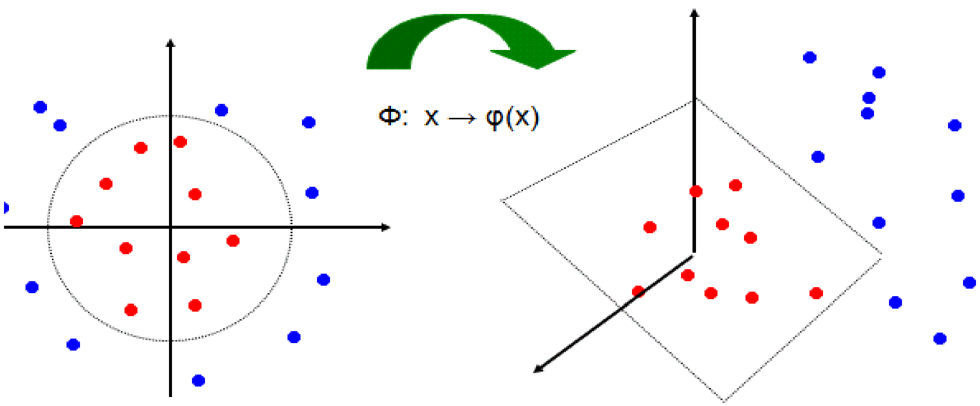
$$Dist(X,Y) = \sqrt{\sum_{i=1}^{d} W_i (X_i - Y_i)^2 \Big/ \sum_{i=1}^{d} W_i}$$



III. SVM – Relevance Feedback

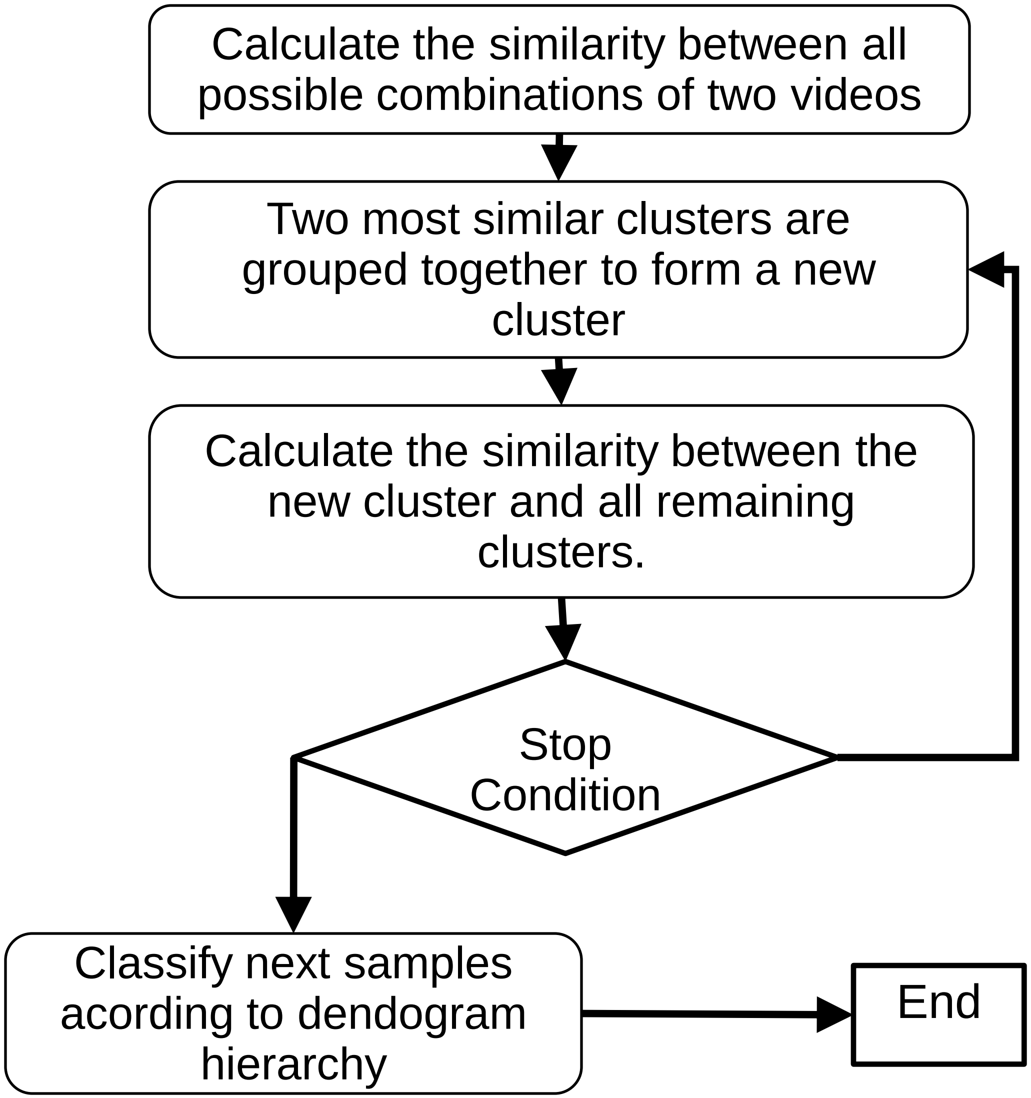
- Train SVM with two classes:
  - positive samples
  - negative samples

- Classify next samples as positive or negative using SVM network



- Use the confidence values of the classifiers to sort the images

IV. Hierarchical Clustering RF



IV. Hierarchical Clustering RF

Stop Condition

Variant 1:

- The number of clusters = a fixed number (e.q. quarter of the number of videos within a retrieved batch)

Variant 2:

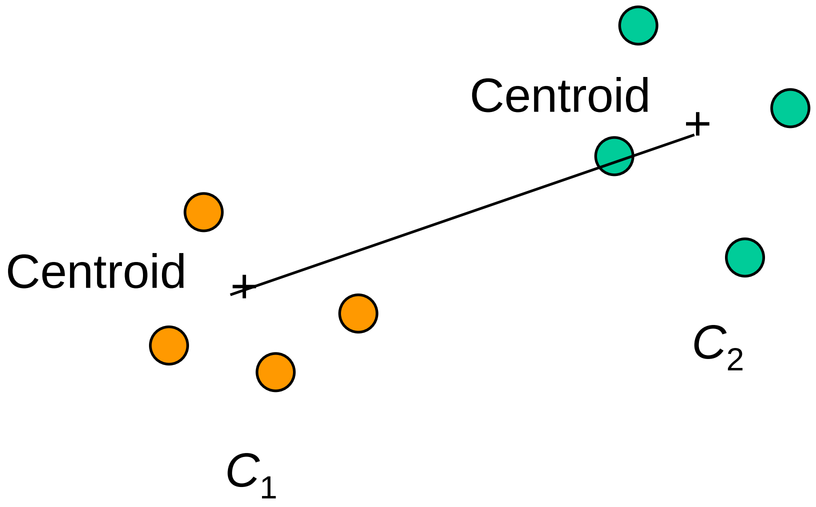
- The number of clusters = an adaptive number

$$d = 1 - \frac{\min D_{ij}}{\max D_{ij}}$$

where Dij represents the distance between two clusters

IV. Hierarchical Clustering RF

Similarity measures – Centroid Distance

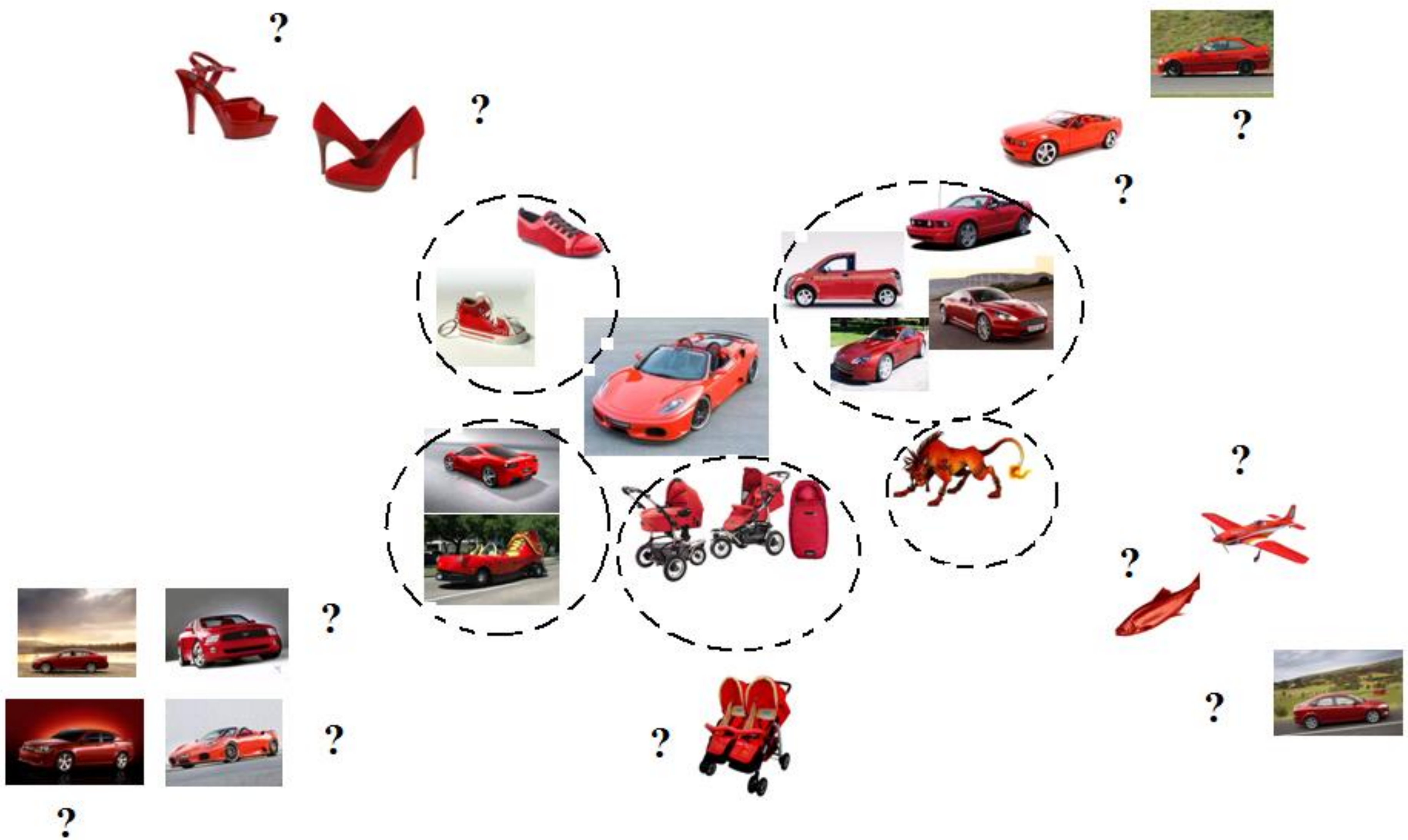


IV. Hierarchical Clustering RF

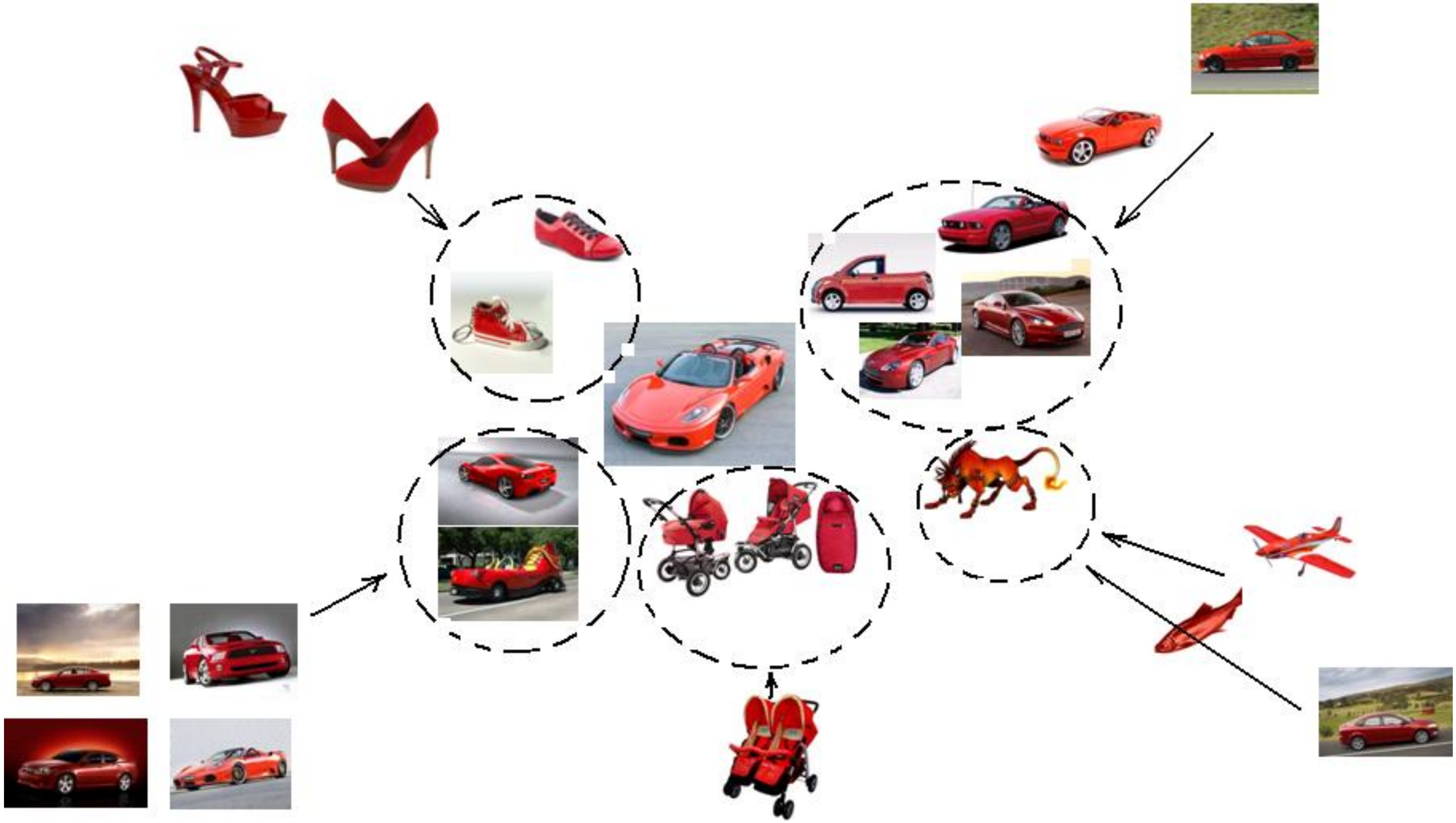
How will be clustered?

IV. Hierarchical Clustering RF

IV. Hierarchical Clustering RF



IV. Hierarchical Clustering RF



V. Implementation and Evaluation

The Test database

- 91 hours of video

- 20h30m of animated movies (long, short clips and series),
- 15m of TV commercials,
- 22h of documentaries (wildlife, ocean, cities and history),
- 21h57m of movies (long, episodes and sitcom),
- 2h30m of music (pop, rock and dance video clips),
- 22h of news broadcast
- 1h55min of sports (mainly soccer) (a total of 210 sequences, 30 per genre).

V. Implementation and Evaluation

- Precision – Recall Chart

$$\text{Precision} = \frac{\text{Number of returned relevant images}}{\text{Total number of returned images}}$$

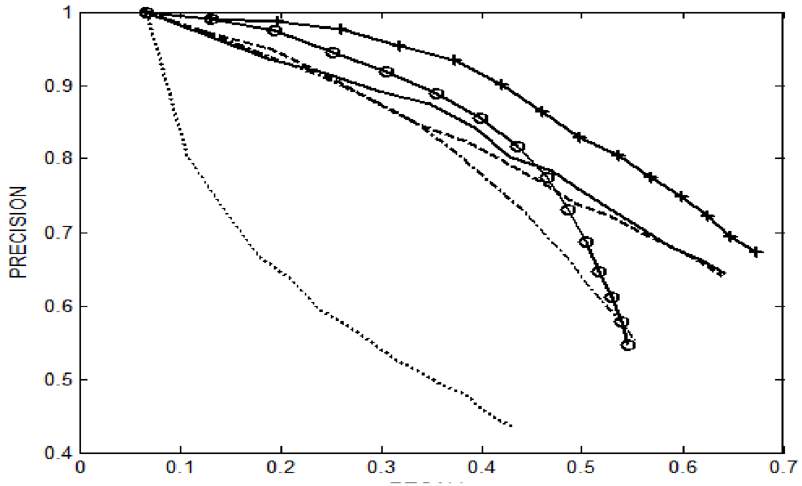
$$\text{Recall} = \frac{\text{Number of returned relevant images}}{\text{Total number of relevant images}}$$

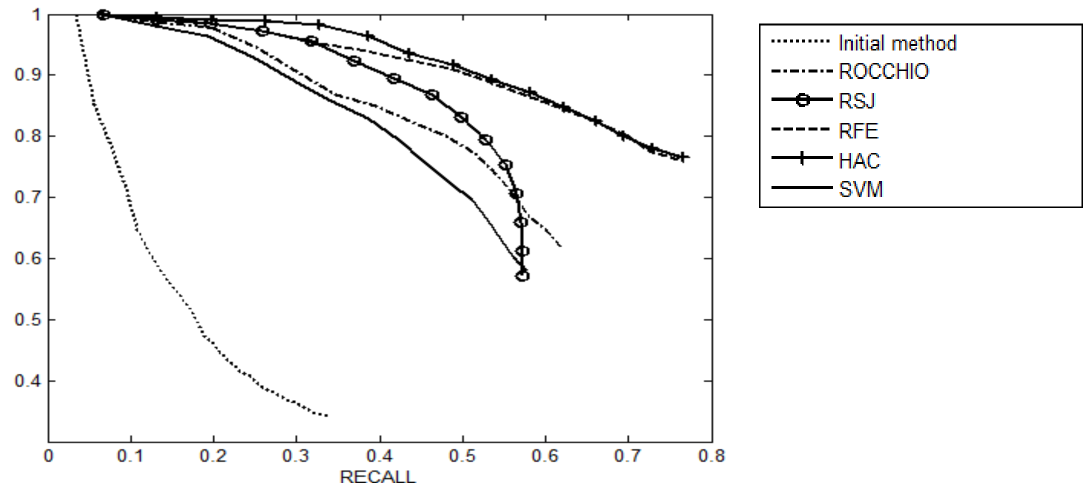
- Average Precision

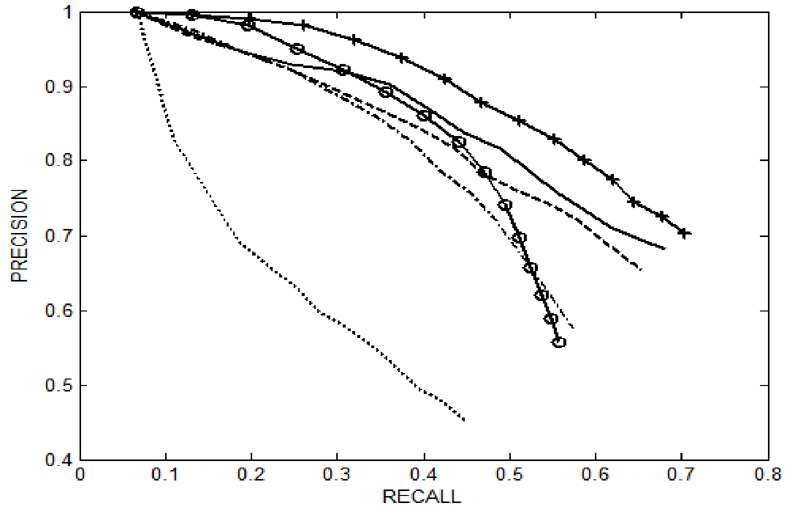
$$AP = 1/d \sum_{i=1}^{d} precision$$

V. Implementation and Evaluation (1)

Evaluation for the first feedback



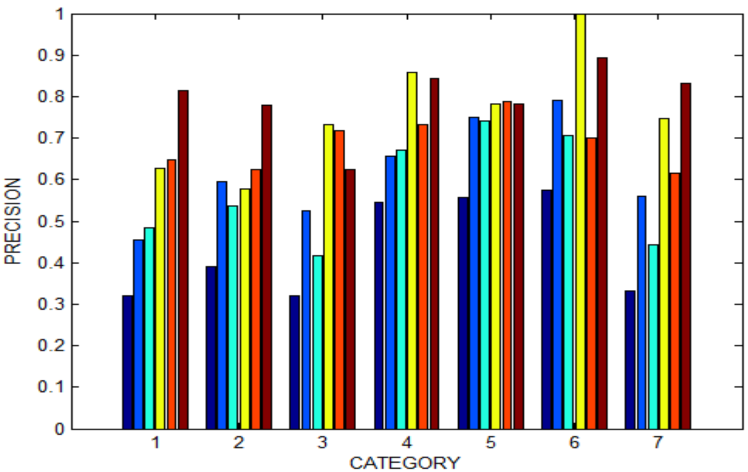


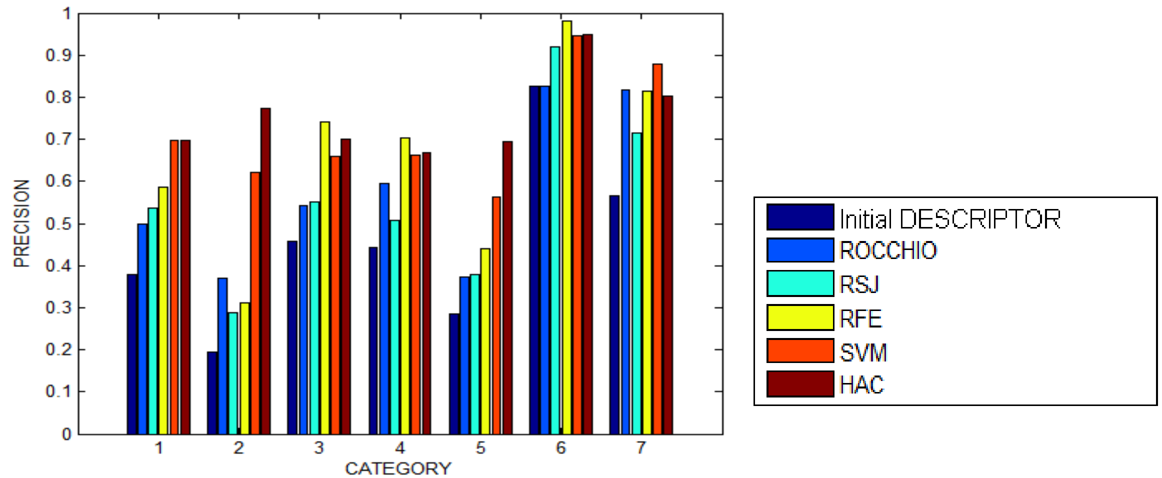


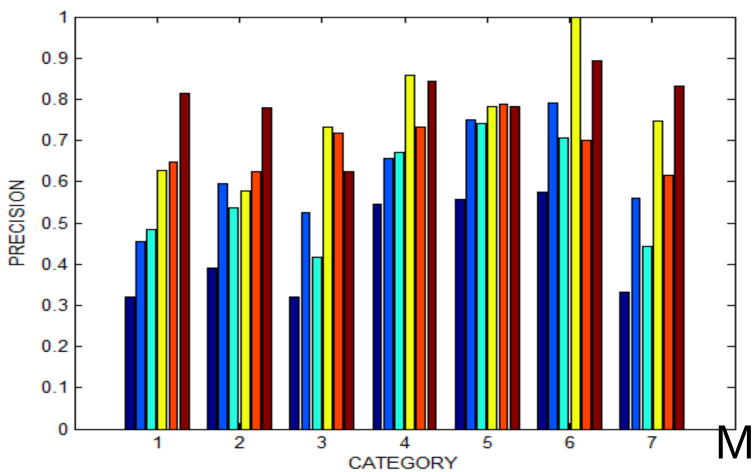
Precision-recall curves per genre category curves for different descriptors:
- Color & Action
- Contour
- Color & Action & Contour

V. Implementation and Evaluation (2)

Evaluation per video genre







Initial Descriptor	40.82%
Rocchio	58.20%
Robertson/Starck-Jones	55.83%
Feature Relevance Estimation	68.48%
Support Vector Machines	70.28%
Hierarchical Clustering RF	**76.61%**

Mean precision improvement with Relevance Feedback

VI. Conclusions

- Relevance feedback is a powerful tool for improving content based video retrieval systems

- The Hierarchical Clustering RF Approach outperforms classical RF algorithms (such as Rocchio or RFE and SVM) in terms of accuracy and computational effort.

Future Work

- Test the algorithm on bigger and more difficult databases with more classes, e.g. MediaEval 2011 – 26 genres, 2500 sequences from Blip.Tv (social media platform)

Thank you!

Questions?