Kobe University at TRECVID 2009 Search Task

Topic Retrieval based on Rough Set Theory and Partially Supervised Learning

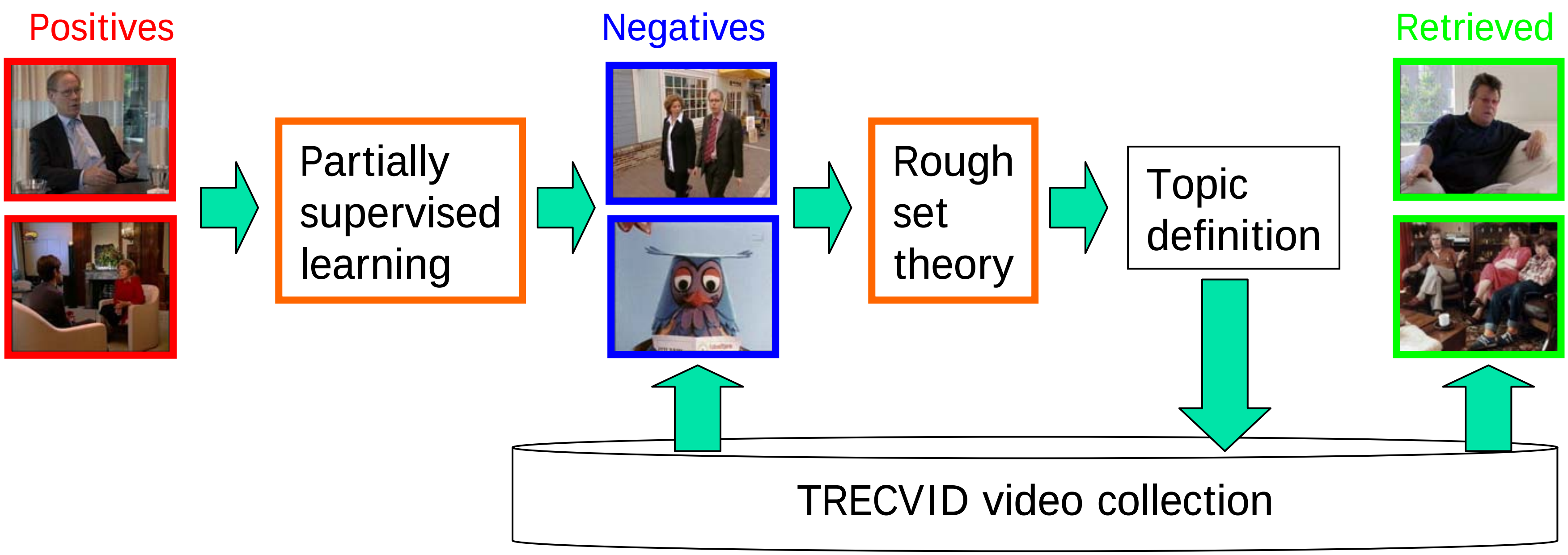
Kimiaki Shirahama, Chieri Sugihara,
Yuta Matsuoka and Kuniaki Uehara

System Overview

Difficulty of preparing indexing and retrieval models for all possible topics

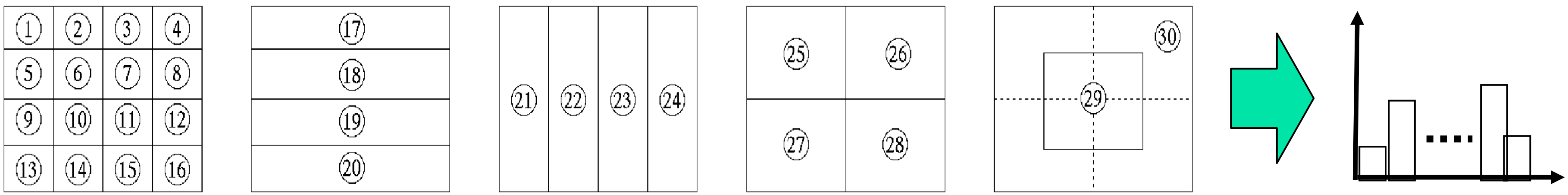
→ Define a topic based on examples provided by a user

Topic 289: one or more people, each sitting in a chair, talking



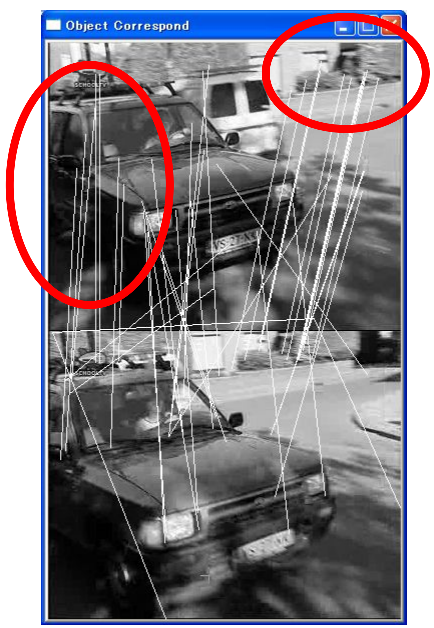
Features

1. Grid-based color, edge and visual word histograms



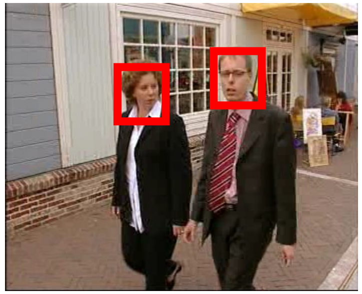
2. Moving regions

R = (x, y, size,
h_move, v_move)



3. # of faces with a certain size

One large-size face



Two small-size faces

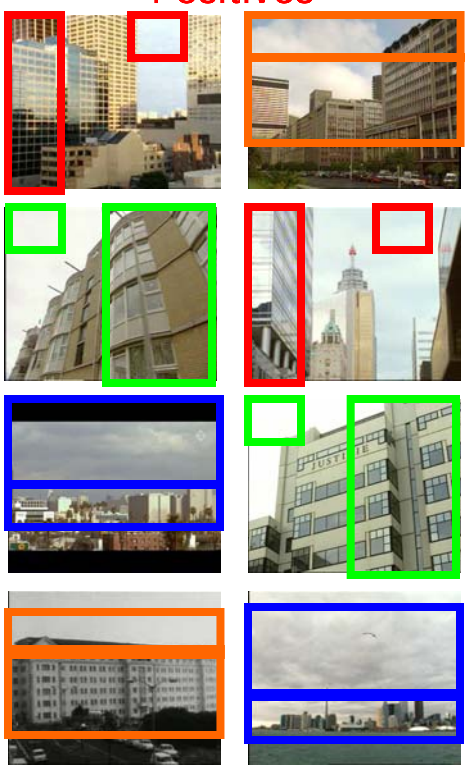
One shot is represented by the Total 94 features!

Rough Set Theory

Large variation of features in the same topic
→ Extract **subsets** where positives can be correctly discriminated from all negatives

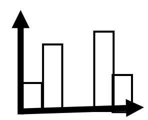
Topic 271: A view of one or more tall buildings …

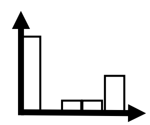
Positives



Negatives

Subsets are computed by boolean algebra of features and described by ***decision rules***.

IF (Color hist. is similar to )

∧ (Edge hist. is similar to )

, THEN *Positive*

Difficulty of Selecting Negative Examples

A great variety of shots can be negatives

Topic 271: A view of one or more tall buildings (more than 4 stories) and the top story visible

Positive

Too much dissimilar
→ Many irrelevant features are included in decision rules

Neither similar nor dissimilar
Many relevant features are included in decision rules,
e.g. long vertical edges, few edges in the upper part, etc.

(with two stories)

Too much similar
→ Many relevant features are ignored

How to select effective negatives for defining a topic?

Partially Supervised Learning

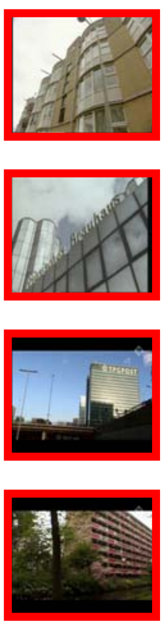
Build a classifier only from positives by selecting negatives from ***unlabeled*** examples

- Web document classification → Documents on the Web as unlabeled examples
- Our topic retrieval → Shots except for positives as unlabeled examples

Similarity-based method (Fung et al. TKDE 2006)

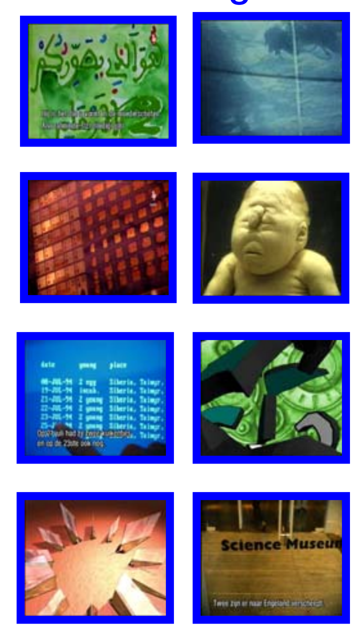
→ Effective in the case where only a small number of positives are available

Positives



1. Reliable negative selection
2. Clustering-based additional negative selection

Reliable negative



Partially Supervised Learning

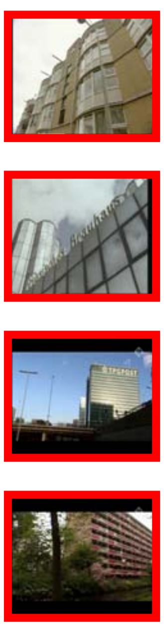
Build a classifier only from positives by selecting negatives from ***unlabeled*** examples

- Web document classification → Documents on the Web as unlabeled examples
- Our topic retrieval → Shots except for positives as unlabeled examples

Similarity-based approach (Fung et al. TKDE 2006)

→ Effective in the case where only a small number of positives are available

Positives

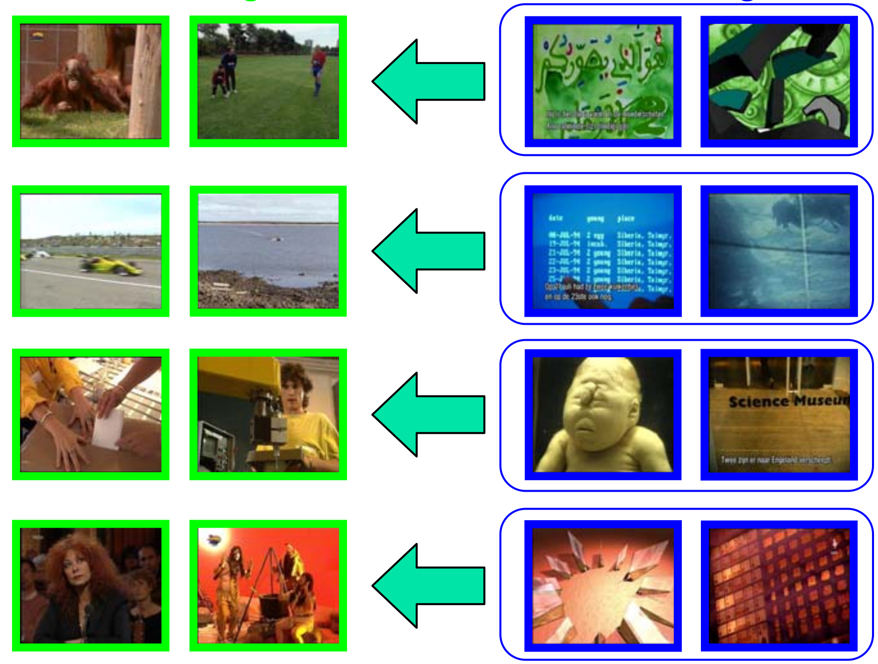


1. Reliable negative selection
2. Clustering-based additional negative selection

How to calculate similarities in a high-dimensional feature space?

Additional negative

Reliable negative



Subspace Clustering

Due to many irrelevant features, we cannot appropriately calculate similarities
→ Find specific features to each example

Subspace clustering (*PROCLUS* proposed by C. Aggarwal et al. SIGMOD 99)
→ Group examples into clusters in different subspaces of the high-dimensional space

Calculate similarities of an example to the other examples only by using the set of associated features!

Submitted Runs

1. **M_A_N_cs24_kobe1_1**
   - Positives by manual and negatives by random
2. **M_A_N_cs24_kobe2_2**
   - Positives by manual and negatives by Partially Supervised Learning
3. **I_A_N_cs24_kobeS_3 (supplemental)**
   - Positives by manual and negatives by random
   - Positives and negatives interactively selected from each retrieval result

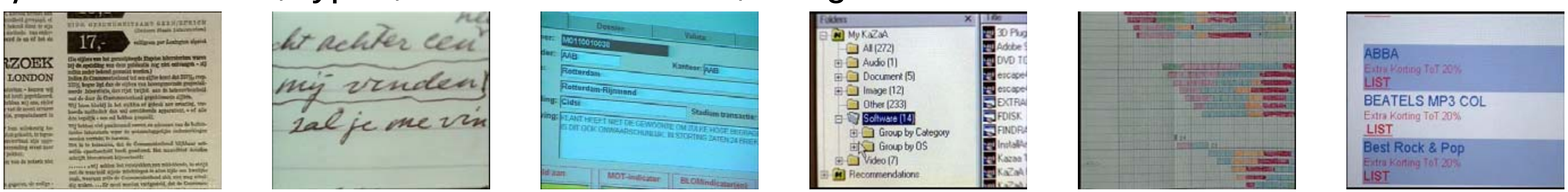
Experimental purposes

- Examine the effectiveness of rough set theory
- Examine the effectiveness of partially supervised learning
- Examine the Influence of positives and negatives on the performance

Example of Good Retrieval

Topic 277: A person talking behind a microphone

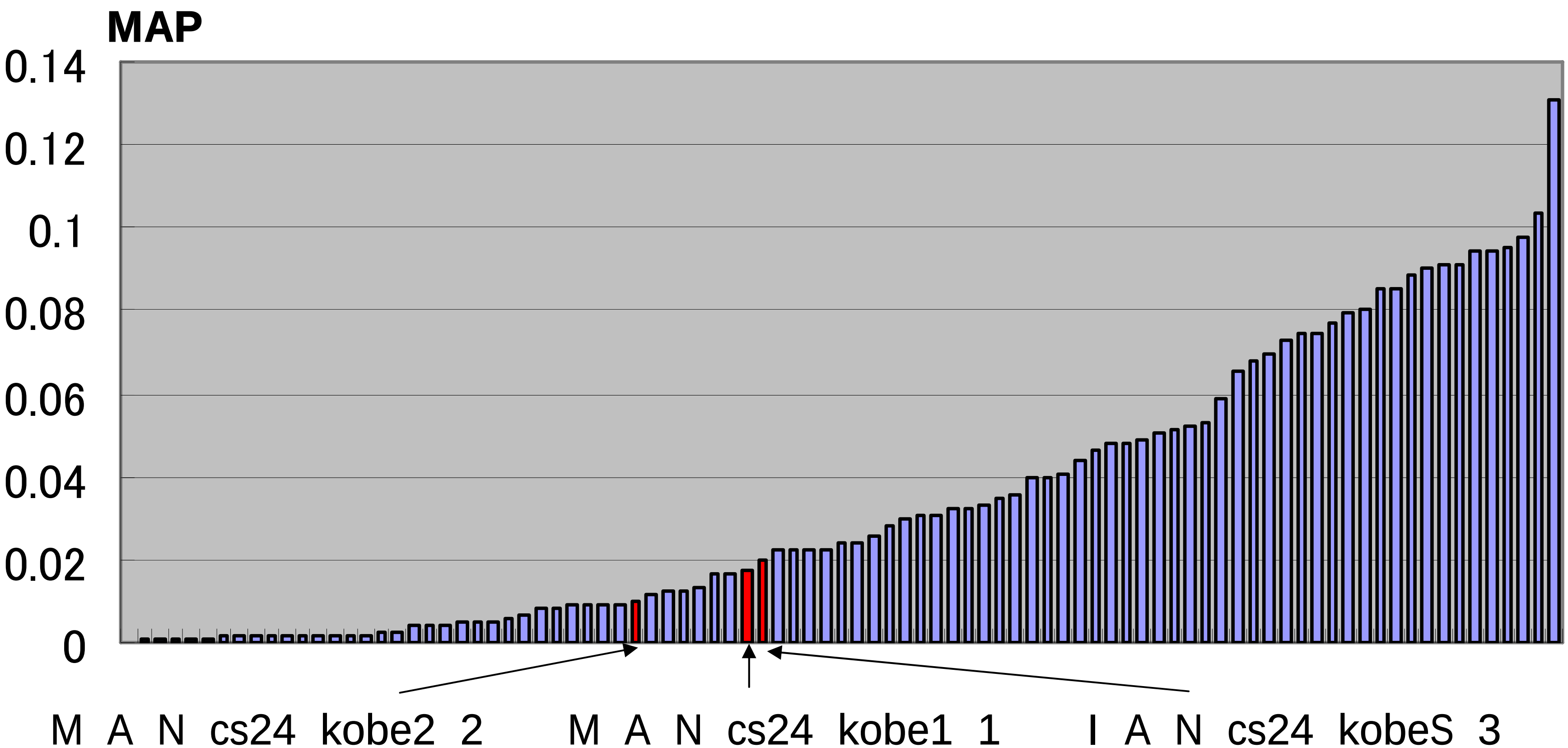
Topic 285: Printed, typed, or handwritten text, filling more than half of the frame area



Topic 289: One or more people, each sitting in a chair, talking

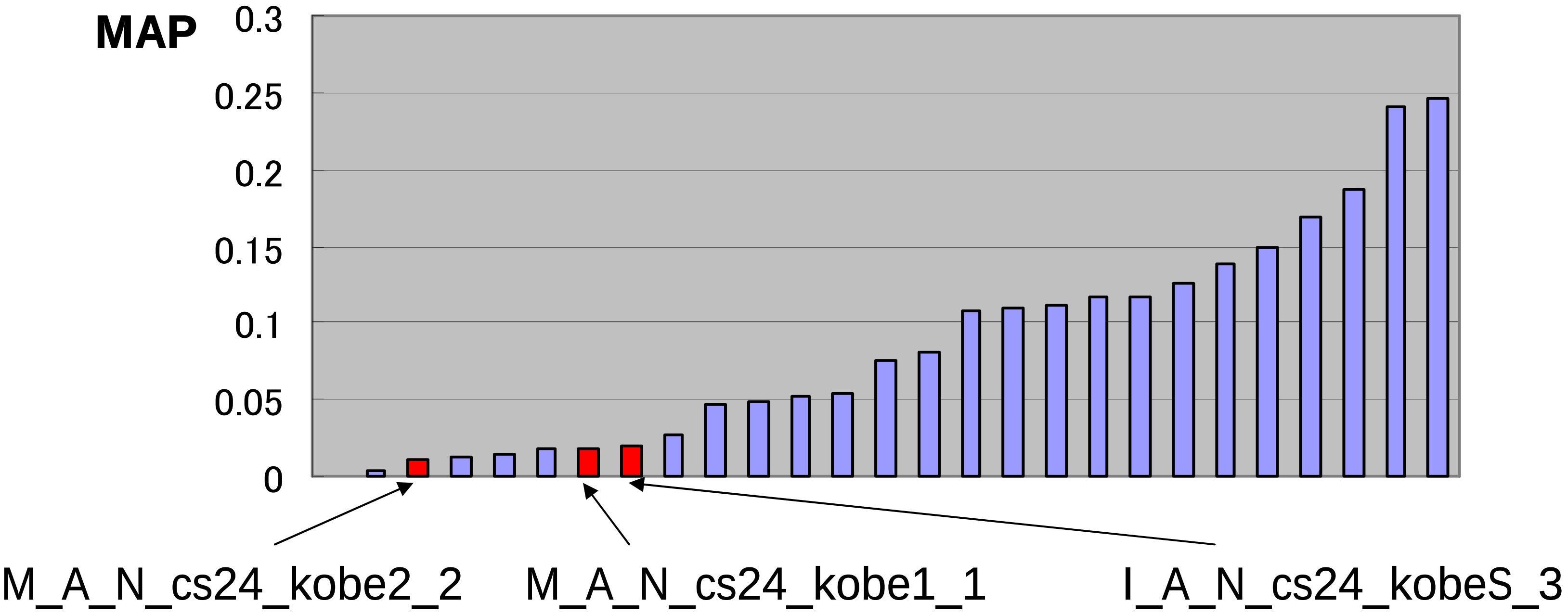
Rough set theory can cover a large variation of features in the same topic!

Comparison to Automatic Runs



NOTE: Only three runs have been submitted for the manually-assisted category.

Comparison to Interactive Runs



Difficulty of deriving an accurate conclusion for partially supervised learning

Why our runs are so bad?

Additional Experiment

Our assumption: Features in submitted runs are ineffective

Additional Experiment

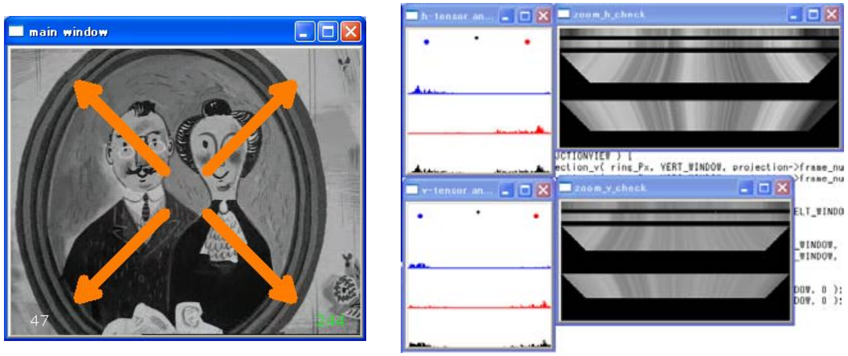
- Select 50 positives and 50 negatives from TRECVID 2008 test videos
- Use various combinations of features
- Features used in submitted runs:
  - Color, edge and visual word histograms,
  - Moving regions, # of faces with a certain size
- Additional features:
  - Grid-based color moment
  - Gabor texture
  - Concept detection scores (provided by MediaMill)
  - HOG
  - Camera work
- Retrieve shots of a topic in 200 of TRECVID 2009 test videos

Main reason for our bad runs

Topic ID	271	272	287	291	292
Same features	14	3	5	2	9
Effective features	**90**	**11**	**50**	**12**	**38**
Estimated best values	70	22	86	22	10
Best values in TRECVID ‘09	209	66	257	66	30

Using ineffective features is the main reason for our bad runs!

- Promising performance when effective features can be selected
- Effectiveness of camera work feature



Zoom in/out estimation by split tensor histogram method (Kumano et al. ITE (In Japanese))

Topic ID	271	272	[illegible]	[illegible]	292
Same features	14	3	5	2	9
Effective features	**90**	**11**	**50**	**12**	**38**
Estimated best values	70	22	86	22	10
Best values in TRECVID ‘09	209	66	257	66	30

Using ineffective features is the main reason for our bad runs!

- Promising performance when effective features can be selected
- Effectiveness of camera work feature

What is an Effective Feature?

Topic ID	271	272	287	291	292
Original result	72	8	34	9	24
Original features	Concept	Concept + Color mom.	Concept	Concept	Camera work + # of faces
Best result	90	11	50	12	38
Effective features	Color hist.	Color hist.	Camera work	Gabor tex.	Concept
Worst result	76	2	16	3	24
Ineffective features	Gabor tex.	Edge hist.	Edge hist.	Visual words	Gabor tex.
All features	66	7	19	1	7
Posteriori Comb.	80	4	36	4	37
Features	Color hist. + Edge hist. + Color mom. + Camera work	Color hist. + Gabor tex.	Color hist. + Moving reg. + Gabor tex. + Camera work	Edge hist. + Moving reg. + Gabor tex.	Concept + Color mom.

Rather than many features, using two or three features leads to the best performance!

Neither visual words nor HOG are effective features.

How Retrieved Shots Change Depending on Features?

Topic ID	271		272		292	
Original result	72		8		24	
Original Feature	Concept		Concept		Camera work + # of faces	
	Color hist. (Effective)	Gabor tex. (Ineffective)	Camera work (Effective)	Edge hist. (Ineffective)	Concept (Effective)	Gabor tex. (Ineffective)
Overlapping shots	**66**	**61**	**28**	**9**	**22**	**14**
Removed shots	6	11	6	25	2	10
Added shots	24	15	16	7	16	10

NOTE: Similar results are obtained for Topic 287 and 291

++ Effective features preserve many relevant shots retrieved by original features, and add more relevant shots.

-- Ineffective features remove many relevant shots retrieved by original features.

How Decision Rules Change Depending on Features?

Topic 271: Tall building

	Building	Sky	Urban
Concept (Original)	357	210	385
Concept + color hist.	361	204	342
Concept + Gabor tex.	**241**	**152**	**327**

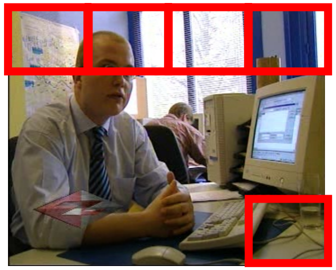
Topic 287: People, table and computer

	Face	Office	Computer or Television
Concept (Original)	177	284	235
Concept + Camera work	138	355	174
Concept + Edge hist.	**77**	303	**86**

++ Effective features preserve most of useful decision rules

-- Ineffective features substitute useful decision rules with inaccurate ones

Wrong match



Wrong match

How to Select Negatives?

Topic ID	271	272	287	291	292
Baseline	80 **(+8)**	3 (-5)	58 **(+24)**	12 **(+3)**	33 **(+9)**
Features	Concept	Concept + Color mom.	Concept	Concept	Camera work + # of faces
Best result	92 **(+2)**	8 (-3)	56 **(+6)**	15 **(+3)**	36 (-2)
Added feature	Color hist.	Color hist.	Moving Reg.	Camera work	Visual words

Topic 287: one or more people, each at a table or desk with computer visible

<u>Random</u>

- Many edges in the upper part
- Many shots where a person appears




<u>Partially supervised learning</u>

- Few edges in the upper part
- Small number of shots where a person appears




Near miss negatives are not useful for defining a topic in videos!

Conclusion and Future Works

Conclusion:

Example-based topic retrieval system

■ Rough set theory for covering a large variation of features in a topic
→ Relevant shots containing significantly different features can be retrieved.
■ Partially supervised learning for negative example selection
→ Selected negatives are more useful than negatives selected by random
But, much more improvement is needed for a satisfactory retrieval!

Future works:

■ Learning a similarity measure which is closely associated with human perception, by using training image pairs labeled as “similar” or “dissimilar”

■ Constructing an event ontology in order to retrieve an event by considering its relation to the other events

■ Developing a browser which enables users to easily collect a sufficient number of positives and negatives.

Thank you!