

TRECVID 2005 Experiments at MediaTeam Oulu

Mika Rautiainen, mika.rautiainen@ee.oulu.fi

**Matti Varanka, Ilkka Hanski, Matti Hosio,
Anu Pramila, Jialin Liu, Timo Ojala and Tapio Seppänen**

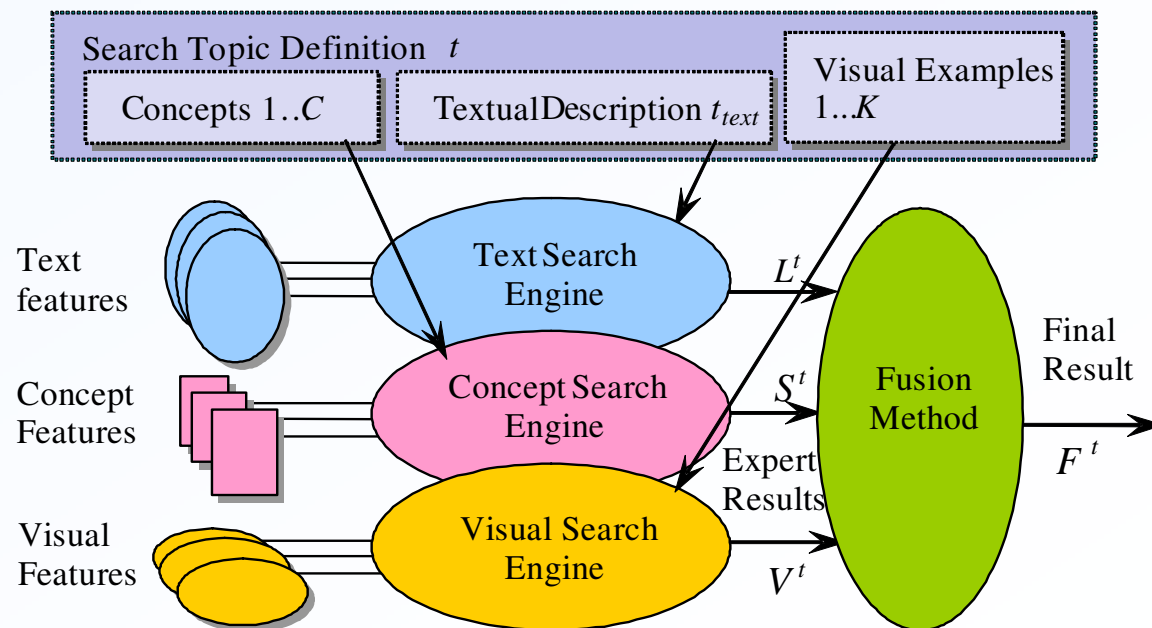
MediaTeam, Departm. of Electrical and Information Engineering
University of Oulu, Erkki Koiso-Kanttilankatu 3,
4SOINFO, 90014 University of Oulu, Finland

1. System Overview
2. Experimental Setup
3. 2005 Results
4. Conclusions

The Prototype Search System

Three search paradigms for retrieval with our video retrieval and browsing system (VIRE):

| | |
|----------------------------|---|
| I Text | Find named people, locations or events. Example: Find shots about the inauguration of Bill Clinton in front of the White House |
| II Concepts | Find common concept objects, events or scenes. Example: Find shots about birds flying in the sky |
| III Visual Examples | Find other video clips that look similar to this clip. Example: Find all occurrences of this analgesic advertisement in a month's recordings |

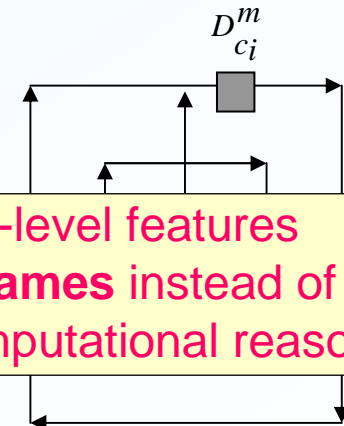


- **Color**

- **Temporal** Color Correlogram (TCC),
spatial color occurrences, 432 values



This year, we computed low-level features from **single subshot key frames** instead of temporal domain due to computational reasons



- $\bar{\gamma}_{c_i, c_j}^{(d)}(S) \equiv \Pr_{p_1 \in D_{c_i}^m, p_2 \in D_{c_j}^m} \left[p_2 \in D_{c_j}^m \mid |p_1 - p_2| = d \right]$

- Dissimilarity by color or structure is defined as a Manhattan distance between the feature vector values
- Fusion of low level similarities for one example query

- $$r^t(k, n) = \text{sum}\left(\frac{d_1^t(k, n)}{D_{1\max}^t(k)}, \dots, \frac{d_L^t(k, n)}{D_{L\max}^t(k)}\right)$$

Combining features using SUM of ranks works well for features having different dimensionalities [10]

- Combining results from K examples

- $$v^t(n) = \min\left(\frac{r^t(1, n)}{R_{\max}^t(1)}, \dots, \frac{r^t(K, n)}{R_{\max}^t(K)}\right)$$

Using MIN of ranks is more flexible than average when heterogeneous query example sets are provided.

Semantic Concept Features

- Semantic Concept Detectors:
Three different approaches were used in detectors
 1. SVM:
 - Entertainment(af+linr.), Outdoor(vf+linr.), Newsroom(vf+linr.), Desert(vf+linr.), Snow(vf+linr.), Natural disaster(vat+2poly)
 2. Propagated labelling with selected example queries [6]:
 - Fire-explosion-smoke, Maps-charts, Meeting-footage, Nature-footage, Weather, Sports, Water
 3. Cascade learning algorithm (Adaboost) [15]: **Faces**
- Concept confidences were based on the shot's relative rank given by the detectors
 - SVM: sigmoid-based probabilistic estimate
 - Labelling: nearest neighbours (ranks)
 - Cascade learning: number of detected faces

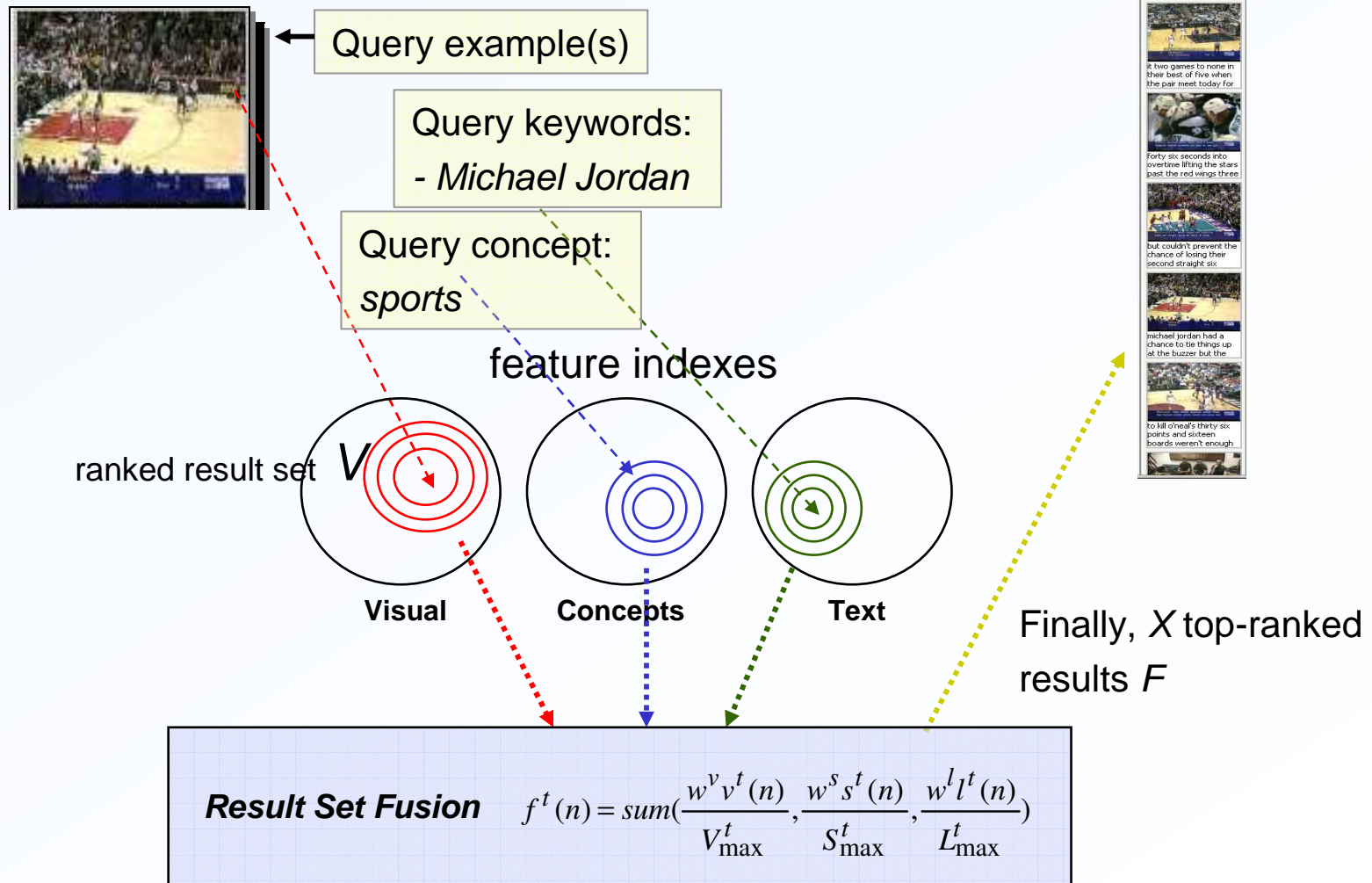
- Text index from ASR and MT transcripts (NIST & CMU)
 - Indexes created from the transcripts w/pre-processing
 - Re-formatting the source transcripts for our system
 - Stop word removal and Porter stemming
 - Inverted document indexes that are expanded using speaker segmentation boundaries and prioritization
 - ASR texts were patched with closed captions text

- Textual similarity between query text and a video shot

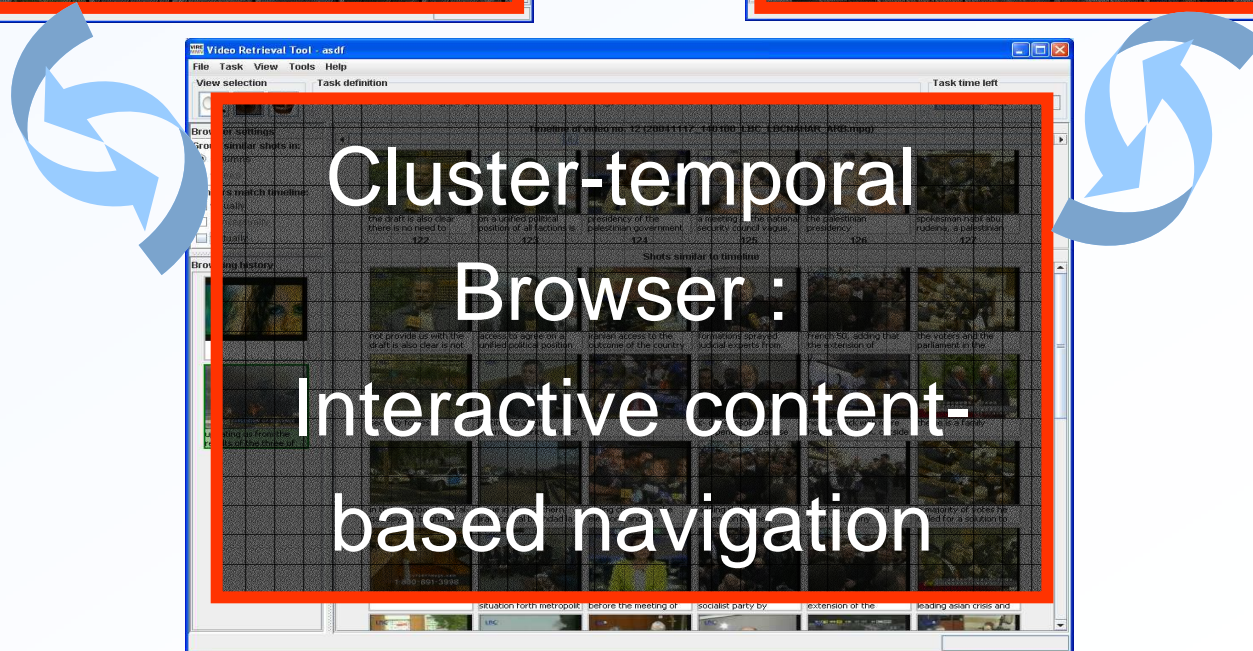
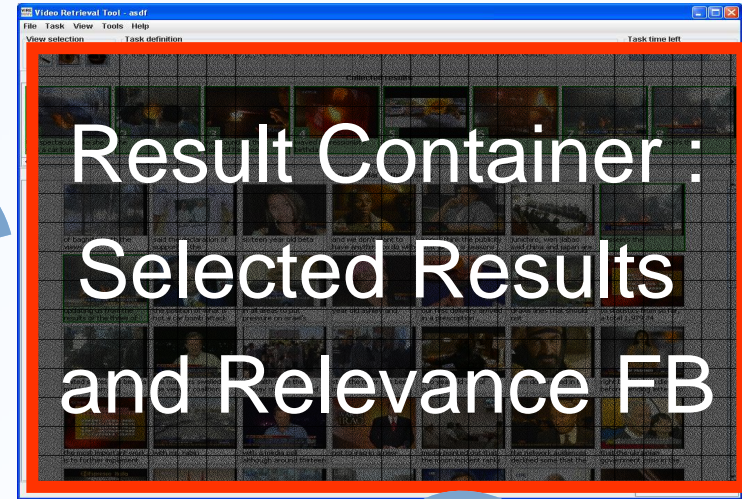
- Value is calculated using the following terms:
 - Ratio of matching words in a shot
 - Inverse freq. of the matching shots
 - Temporal weighting based on prioritization

- Aggregated with a variation of TFIDF measure

$$L(queryterm, s) = 0.2 \cdot \frac{\log(t + 1)}{\log(dl + 1)} * \log\left(\frac{N}{m}\right) + e^{-B \frac{f}{J}}$$



The Search System Interfaces



The screenshot displays the 'Video Retrieval Tool - asdf' window. The interface is divided into two main sections: 'Query Definition' on the left and 'Retrieved results are here' on the right.

Query Definition Section:

- File Task View Tools Help** menu bar.
- View selection** tab with icons for search, eye, and basket.
- Task definition** tab.
- Task time left** field.
- Semantic concepts** list:
 - Enabled: Entertainment, Faces, Fire-explosion-smoke, Maps-charts, Meeting-footage, Nature-footage, Newsroom, Outdoor.
 - Disabled: Forest, Wood, Tree, Tropic, Jungle, Plant, Green, Desert.
- 200 results** dropdown and **Submit** button.

Retrieved results are here Section:

- A grid of 20 video thumbnails, each with a title and a duration (e.g., 1:00, 1:05, 1:10).
- Titles include phrases like 'director of the contract is already minor official', 'in fact the director of secondary zook inkeyil', 'of the separation of planting particularly', 'be held assessing sino-us youth study institute', 'seoul term present planted central', 'stress in the country was chairman of', 'large member season potato as fadel flood', 'today not organized demonstrators and the', 'entrance to the city-street jubran khail', 'faltering serious damage, especially karmi', 'of the farm-river floods caused by', 'martyrs planted a tree of', 'venue of these developments waiting', 'of rapid winds', 'roofts of', 'strus house justified', 'coastal trees were counting on the', 'of the collapse of the party at the', 'areas symbols 10 years imprisonment for coming', 'the roots of the earth, temperature fawcett's four bulldozers worked', 'is temporary, but after that briefed', 'school for a month after the determination to', 'the fishermen', 'in the day the green line between israel and', 'tried default leaflets and a call after three years', 'the ocean and stop', 'he gave a tree, the tripoli-cars to'.

Cluster-temporal Browser

The screenshot displays the 'Video Retrieval Tool - asdf' window. The interface includes a menu bar (File, Task, View, Tools, Help), a 'View selection' panel with icons for search, zoom, and pan, and a 'Task definition' field containing the text: 'Find shots of something (e.g., vehicle, aircraft, building, etc) on fire with flames and smoke visible'. A 'Task time left' indicator shows 4:21. The main area is divided into two sections: 'Timeline of video no. 12 (20041117_140100_LBC_LBCNAHAR_ARB.mpg)' and 'Shots similar to timeline'. The timeline section shows a grid of video frames with time markers 122, 123, 124, 125, 126, and 127. The 'Shots similar to timeline' section displays a grid of video frames with captions such as 'not provide us with the', 'access to agree on a', 'iranian access to the', 'formations sprayed', 'french 50, adding that', 'the voters and the', 'the draft is also clear', 'on a unified political', 'presidency of the', 'a meeting of the national', 'the palestinian', 'spokesman nabi abu', 'security forces', 'internal security minister', 'independent lebanese', 'helping them to', 'outside', 'in the neighbourhood of', 'issue in the southern', 'ceiling change to the', 'adding that the', 'the constitution and', 'a majority of votes he', 'called for a solution to', 'in your house', 'round about the', 'situation forth metropolit', 'you, ended shortly', 'before the meeting of', 'of the progressive', 'socialist party by', 'adding that the', 'extension of the', 'points of the votes', 'leading asian crisis and'.

Settings
for
Browsing

Browsing
History

Selected video broadcast timeline

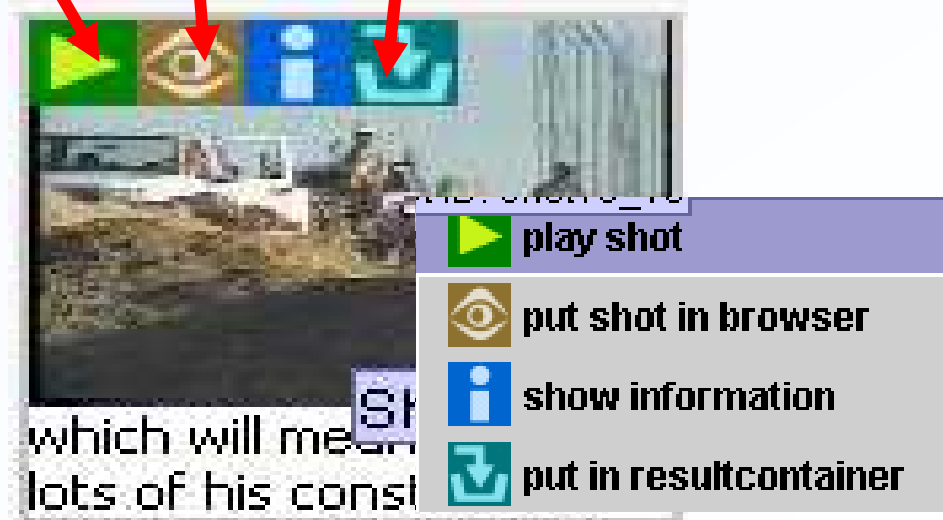
Automatically generated
view of similar video segments
in the 60 hour video database

Quick Buttons for Streamlined Interaction

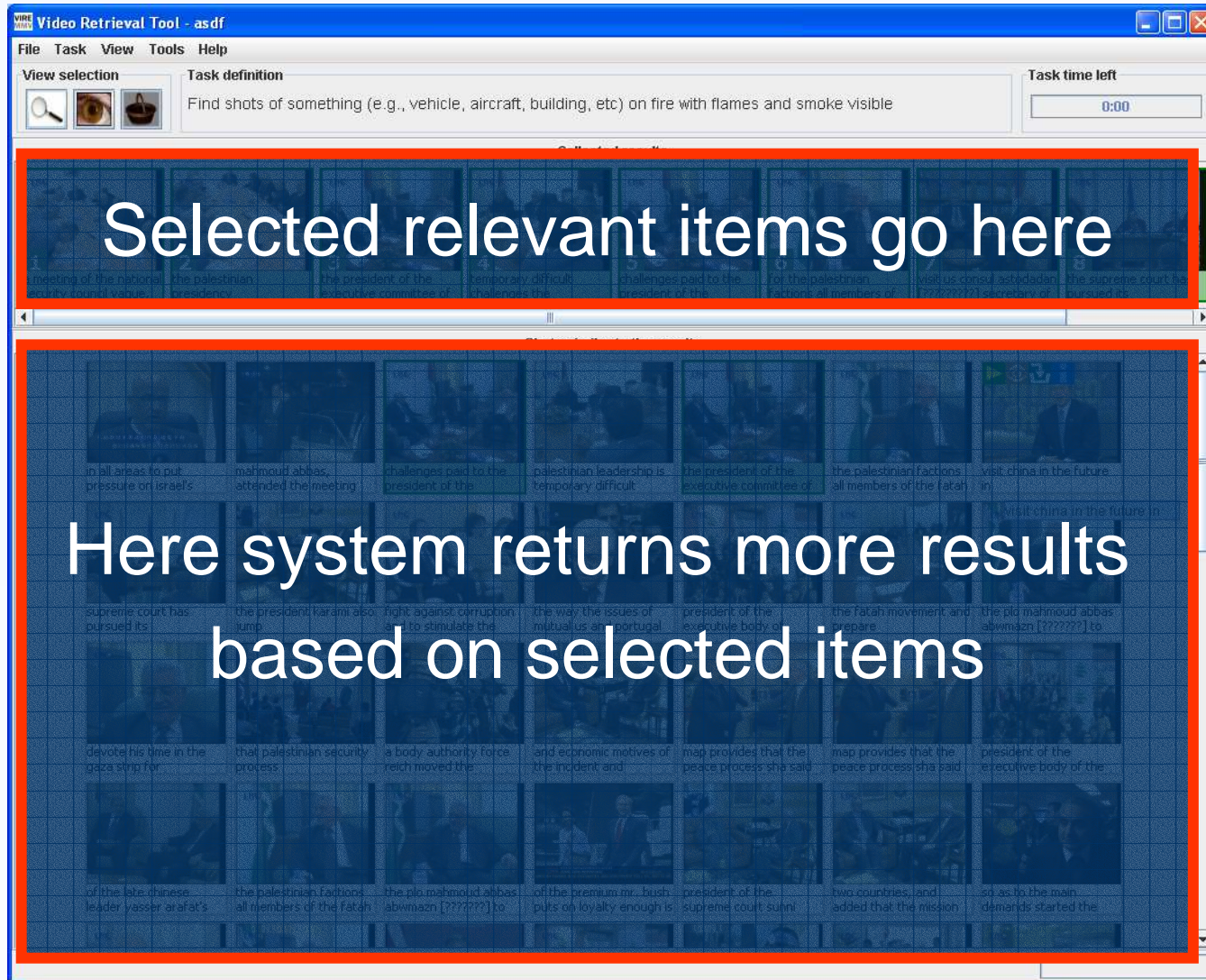
Play Shot

Browse News Video

Select as a result and
move to Result Container



Result Container: Relevance Feedback based on selected results



- MediaTeam participated in manual and interactive search tasks with following 7 runs:
 - **OUMT_I1Q_1:** interactive with **browsing disabled, expert** users
 - **OUMT_I2B_2:** interactive with **browsing enabled, expert** users
 - **OUMT_I3Q_3:** interactive with **browsing disabled, novice** users
 - **OUMT_I4B_4:** interactive with **browsing enabled, novice** users
 - **OUMT_M5T_5:** manual text search with official text transcripts
 - **OUMT_M6TS_6:** manual text search + semantic concepts
 - **OUMT_M7TE_7:** manual text search + visual examples

Total of eight test users did

- **12 test topics** using **two** different **system configurations**
- enjoyed break and refreshment after six topics and spent about three hours in total for this experiment
- four users were experts
 - very knowledgeable with the system, but had not seen the given search topics or any content from the test database.
- four users were novices
 - mainly information engineering undergraduate or post-graduate students, having good skills in using computers but little experience in searching video databases.

Search configuration:

I1Q: Variant A: S1[149-154],S3[155-160],S2[161-166],S4[167-172]

I2B: Variant B: S2[149-154],S4[155-160],S1[161-166],S3[167-172]

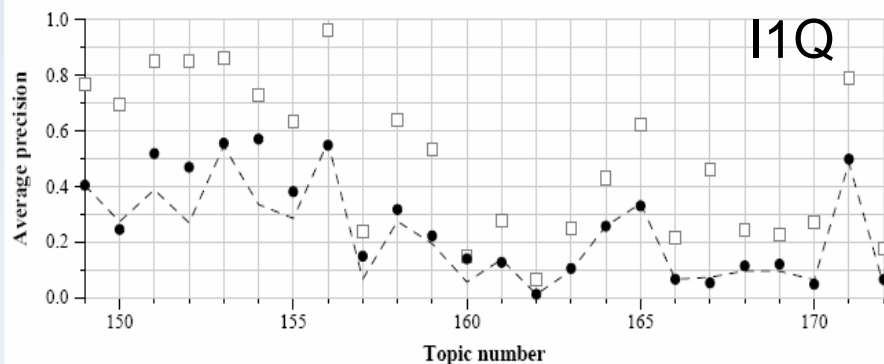
I3Q: Variant A: S7[149-154],S5[155-160],S6[161-166],S8[167-172]

I4B: Variant B: S8[149-154],S6[155-160],S5[161-166],S7[167-172]

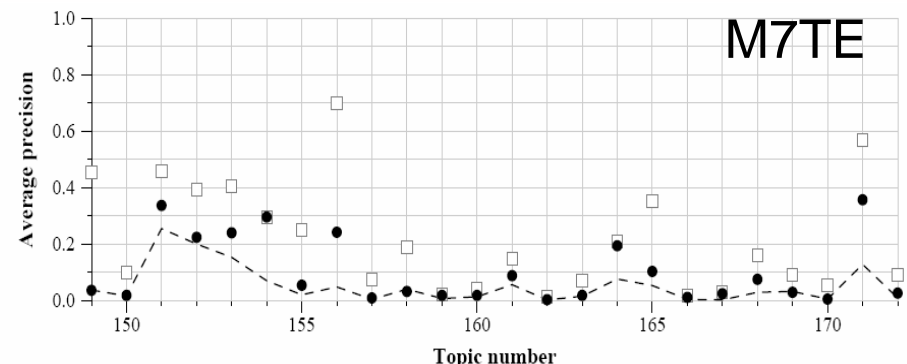
| Search Run ID | MAP | Total Relevant Shots Returned |
|--|-------|-------------------------------|
| I1Q (interactive, expert users) | 0.264 | 2284 |
| I2B (interactive, expert users) | 0.242 | 1916 |
| I3Q (interactive, novice users) | 0.202 | 1907 |
| I4B (interactive, novice users) | 0.226 | 1998 |
| Mean (interactive) | 0.218 | 1618 |
| Max (interactive) | 0.414 | 3044 |
| M5T (baseline text search) | 0.081 | 1836 |
| M6TS (txt search+semantic) | 0.097 | 2003 |
| M7TE (txt search+examples) | 0.102 | 1972 |
| Mean (manual) | 0.067 | 1510 |
| Max (manual) | 0.169 | 2278 |

- Interactive runs
 - **12% better** MAP-performance for **novice** users **using cluster-temporal browser than without it**
 - The result is in line with previous reported experiments with novice test users [5].
 - However, expert users had marginally better MAP (0.264 vs 0.242) without the Cluster-temporal Browser, why?
 - Expert knowledge about system capabilities and limitations makes them perform well with every configuration. Also personal skills vary depending on the role in development
 - on average expert users had **18% better search performance over novice users**
 - It shows that the test design has a significant effect to the outcome of the interactive test.

- Manual runs:
 - **text + semantic concept** search gives about **19% better performance than text baseline**
 - **text + example** based search gives approximately **25% performance gain over the baseline.**
 - The results show that specific visual search examples accumulate better overall precision than the queries defined with our detected set of semantic concepts.



Run score (dot) versus median (---) versus best (box) by topic



Run score (dot) versus median (---) versus best (box) by topic

- Main conclusions from this study:
 - **Cluster-temporal browsing improves search performance** over traditional query + relevance feedback paradigm for **novice** users
 - content-based example and concept search components **improve search performance** over straightforward text-based search
 - search examples seem to contribute more than concepts in our system
 - The setting for interactive experiment is an important factor in the overall search performance
 - The expert users are able to 'push' the system limits and obtain good performance in both configurations.

Thank you

- mika.rautiainen@ee.oulu.fi