The Mediamill 2005 semantic video search engine

Marcel Worring, Richard van Balen, Jan-Mark Geusebroek, Jan van Gemert, Bouke Huurnink, Dennis Koelma, Giang Nguyen, Ork de Rooij, Arnold Smeulders, Cees Snoek

Intelligent Systems Lab Amsterdam
University of Amsterdam
TRECVID interactive search

- Task
  - Search the right shots for some given information need

- Data
  - Chinese, Arabic, American News data

- Constraint
  - Per topic time is 15 minutes maximum
The 101 Concepts

Fire weapon  Fish  Flag
Flag USA  Food  Football  Golf  Gov. building
Gov. leader  Graphics  Grass  Hassan Nasrallah  Horse
Horse racing  House  Hu Jintao
Indoor  Kerry  Lahoud  Male
Map  Meeting  Military  Monologue
Motorbike  Mountain  Natural disaster  Newspaper
Night fire  Office  Outdoor  Overlayed text
People  People marching  Police / security
The 101 Concepts

- Powell
- Prisoner
- Racing
- Religious leader
- River
- Road
- Screen
- Sharon
- Sky
- Smoke
- Snow
- Soccer
- Split screen
- Sports
- Studio
- Swimming pool
- Table
- Tank
- Tennis
- Tony Blair
- Tower
- Tree
- Truck
- Urban
- Vegetation
- Vehicle
- Violence
- Walking
- Water body
- Waterfall
- Weather news
Interactive retrieval

- If there is a simple mapping from topic to concepts
  - ...... and the concept is of sufficient quality
  - performing a concept query yielding a ranking is sufficient

- One additional observation
  - If one element of a program is relevant, other elements of the program are often also relevant
The Cross Browser
So what if the relation is less obvious

- Use the structure of the multimodal concept space
  - How are different information sources related?
  - Can we use combinations of different concepts?

- But the space is very complicated
  - So can we guide the user through concept space?
Finding semantic threads

Clustering algorithm

- Shots with cluster assignments

Threading algorithm

- Threads of shots

Buildings in back
timeline
anchors
The Sphere Browser
And if there are no concepts?

- Visual browsing
  - Exploring a large part of the whole dataset
  - Make use of clusters in visual appearance

- Also relevant if there are sub-concepts to be found
  - Especially if the instantiations of the sub concepts are visually similar
The offline stage

Image collection → High dimensional feature space → Projection to visualization space → Preserve neighbor relations → Clustering
On-line

Preserve neighbor relations and assure visibility

Show representative set

User selection of one positive example

Enter cluster

Drag rectangle for selection

Select border images

SVM based Active learning

Show final result
The Galaxy Browser

- The collection and positions of the current display set
- Navigation window
- Dragged rectangle
- Selected training examples
- Detailed view

User simulation studies show its effectiveness
An effective setup
Results

TRECVID 2005 Overall Search Results

- CrossBrowser
- SphereBrowser
- GalaxyBrowser
- MixedBrowser
- Automatic
- Manual visual-only
- Text-only
Conclusions

- A large multimodal lexicon
  - Is a great aid in all search tasks

- Visualization
  - Helps significantly when there is no 1-1 mapping from query to concept

- Future work
  - Employ semantic concept structure
  - Even more advanced visualizations