Predictive Video Retrieval
A Matter of Trust

Bouke Huurnink

MediaMill
The Team

Bouke Huurnink
Jiyin He
Koen van de Sande
Ork de Rooij
Cees Snoek
Maarten de Rijke
Jan van Gemert
Jasper Uijlings
Xirong Li
Ivo Everts
Vladimir Nedovic

Michiel van Liempt
Richard van Balen
FeiYan
Muhammad Tahir
Krystian Mikolajczyk
Josef Kittler
Jan-Mark Geusebroek
Theo Gevers
Marcel Worring
Arnold Smeulders
Dennis Koelma
Come see our interactive search demo

Now with (inter)active learning!
Presented by Ork de Rooij
Why *predictive* video retrieval?

- Video retrieval is a multichannel problem:
  - Speech
  - Detectors
  - Examples

- Observations
  - Speech works for named entity topics
  - Detectors work when closely related to topic
  - Examples can also work pretty well

- We want to exploit this knowledge
Idea

• Predict which type of search - retrieval channel - we can trust for a topic

• Rerank results from this channel with secondary result information
Outline

• System description
• Result overview
• Analysis
• Conclusion
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Predict Trusted Channel

Retrieval Channels
Speech Search
Detector Search
Example Search

Result Lists
- Trusted Results
  - Detector results
- Secondary Results
  - Example results
  - Speech results

Reranking

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Retrieval Channels
- Speech Search
- Detector Search
- Example Search

Predict Trusted Channel

Result Lists
- Trusted Results
  - Detector results
- Secondary Results
  - Example results
  - Speech results

Reranking

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Retrieval Channels
- Speech Search
- Detector Search
- Example Search

Predict Trusted Channel

Result Lists
- Trusted Results
  - Detector results
- Secondary Results
  - Example results
  - Speech results

Distribute ASR and MT over shot neighbourhood, then retrieval using language modelling approach.

Content based selection from 57 learned concepts, followed by unweighted score-based fusion.

Pseudo active-learning, with positive examples from topic and 100 random negative examples from collection.

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Retrieve Channels
- Speech Search
- Detector Search
- Example Search

Predict Trusted Channel

Result Lists
- Trusted Results
  - Detector results
- Secondary Results
  - Example results
  - Speech results

Reranking

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Named entity? Trust speech results
Detector match? Trust detector results
Else...trust example results

Retrieval Channels
- Speech Search
- Detector Search
- Example Search

Results
- Detector results
- Secondary Results
  - Example results
  - Speech results

Reranking

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Retrieval Channels
- Speech Search
- Detector Search
- Example Search

Predict Trusted Channel

Result Lists
- Trusted Results
  - Detector results
- Secondary Results
  - Example results
- Secondary Results
  - Speech results

Reranking

Final Results
Our predictive system

Information Need
Find shots of pieces of paper with writing, typing, or printing, filling more than half of the frame area.

Retrieval Channels
- Speech Search
- Detector Search
- Example Search

Result Lists
- Trusted Results - Detector results
- Secondary Results - Example results

Combine results with (weighted) Borda fusion

Reranking
Truncate result lists to top 1000
Eliminate all results not in trusted list

Final Results
## Query-class vs Prediction

<table>
<thead>
<tr>
<th>Query-class</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query class determines retrieval strategy</td>
<td>Query features determine retrieval strategy</td>
</tr>
<tr>
<td>Focus on assigning query-class dependent weights</td>
<td>Focus on identifying trusted retrieval channel</td>
</tr>
</tbody>
</table>
Runs

• Speech channel only  UvA-MM-6
• Detector channel only  UvA-MM-5
• Example channel only  supplementary
• Predictive reranking  UvA-MM-4
• Predictive weighted reranking  UvA-MM-3
Overall Automatic Search Performance

Predictive weighted reranking
Predictive reranking

mean inferred average precision

Detector channel
Example channel
Speech channel

All runs
Overall Automatic Search Performance

Predictive weighted reranking
Predictive reranking

mean inferred average precision

Detector channel
Example channel
Speech channel

Predictive reranking outperforms individual channels
Overall Automatic Search Performance

- Predictive reranking
- Predictive weighted reranking

Detector channel
Example channel
Speech channel

Predictive reranking outperforms individual channels
Weighting did not have big influence
General findings

- 20 topics > 0.05 inferred average precision
  - 1 speech topic
  - 11 detector topics
  - 8 example topics
- Accurately predicted 15 of 20 topics
A closer look
A closer look

A lot of variance between channels
When prediction worked

- person opening door
- a bridge
- people with trees and plants
- paper with writing
- a map
- people looking in microscope
- people in a kitchen
- a crowd of people outdoors
- a classroom scene
- an airplane exterior
- a plant that is the main object
- a street scene at night
- people at table with computer
- people in white lab coats
- man talking to camera indoors

Inferred average precision

- Predictive w. reranking
- Detector channel
- Example channel
- Speech channel

Only trusted channel and reranked performance shown
When prediction worked

Predictive reranking often close to or better than trusted channel

Predictive w. reranking
Detector channel
Example channel
Speech channel
When prediction didn’t work

- Face filling over half the frame
- People with a body of water
- Vehicle moving away
- Person watching television
- Ships or boats in the water

Inferred average precision

Only trusted channel and reranked performance shown
When prediction didn’t work

Predictive reranking boosts trusted channel results

Only trusted channel and reranked performance shown
Conclusions
Conclusions

Predictive retrieval works, even with simple reranking
Conclusions

Predictive retrieval works, even with simple reranking

Incorrect predictions have limited impact
Conclusions

Predictive retrieval works, even with simple reranking

Incorrect predictions have limited impact

Good ingredients are crucial: garbage in garbage out!
Pondering

• What if we had more variety in query types?

Beeld en Geluid Searches

General
object queries

Named entity
queries