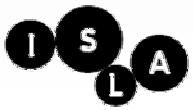
To Fuse or Not to Fuse: That is 101 Questions

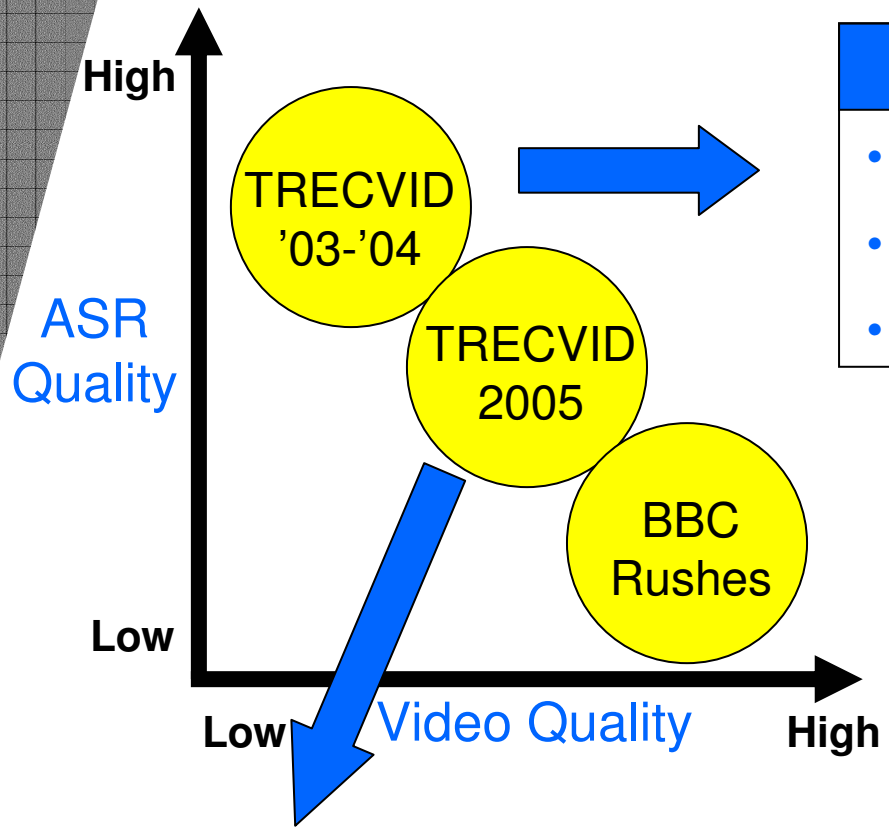
in semantic video analysis.

Cees Snoek, Jan van Gemert, Jan-Mark Geusebroek, Dennis Koelma,
Frank Seinstra, Arnold Smeulders, Cor Veenman, & Marcel Worring

Intelligent Systems Lab Amsterdam,
University of Amsterdam, The Netherlands



Introduction



Lessons Learned

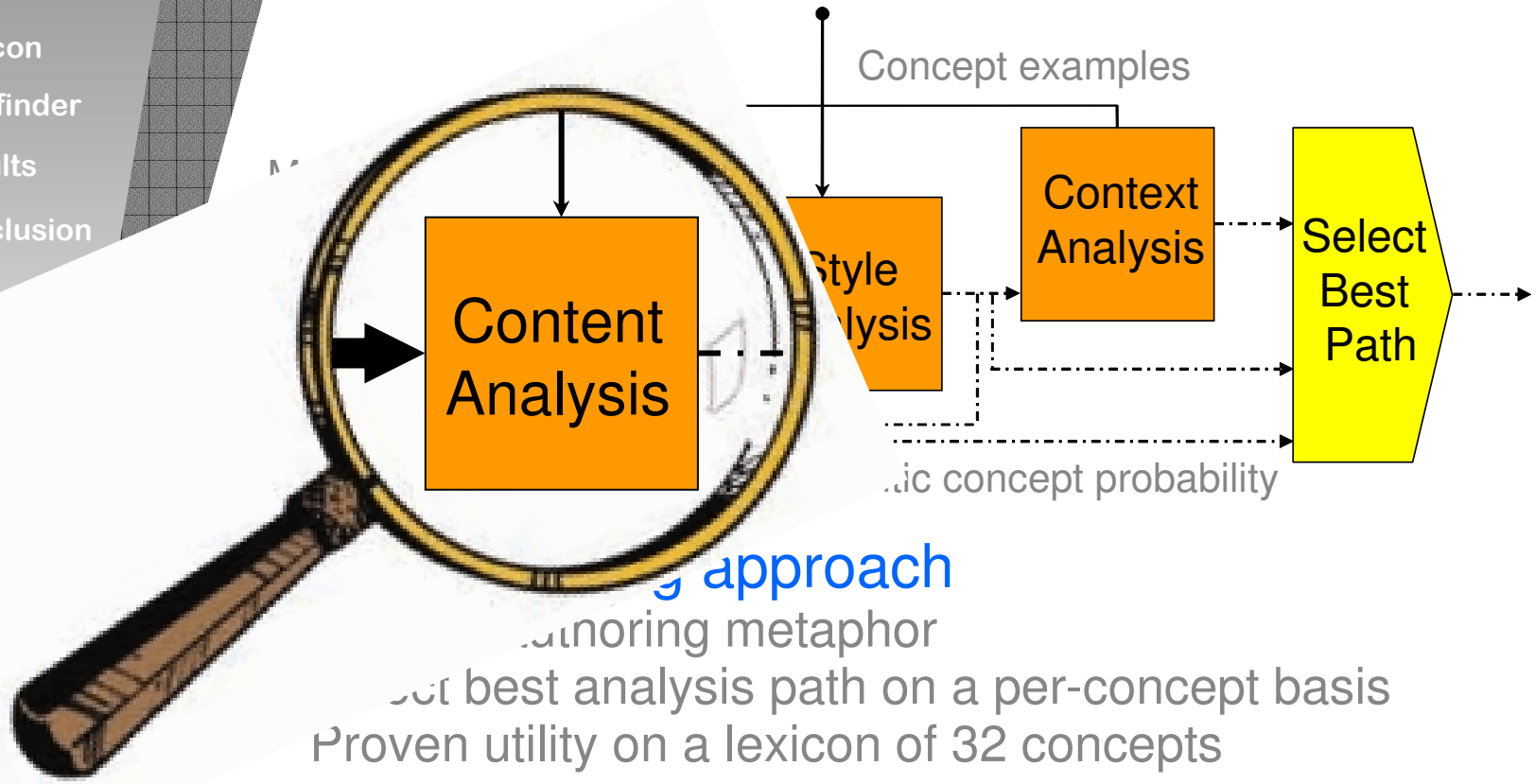
- Generic indexing is possible
- Fusion improves performance
- Text is the decisive modality

- TRECVID '05: A new hope for the visual modality?
  - ✓ Poor ASR because of non-English broadcasts
  - ✓ Best quality of video data so far
- Do the lessons learned still hold?

Semantic Pathfinder

TRECVID 2004

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion



g approach
noring metaphor
best analysis path on a per-concept basis
Proven utility on a lexicon of 32 concepts

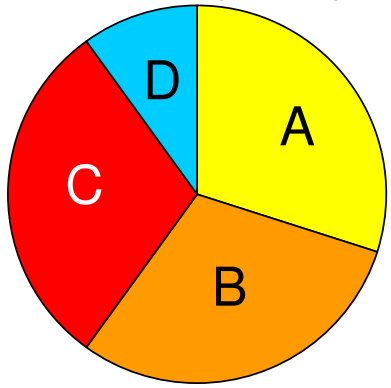
➢Our 2005 experiments focus on content

Preliminaries

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Data preparation

- ✓ We randomly split training set a priori into 4 sets
- ✓ Three sets for training (30%), 1 set for validation (10%)

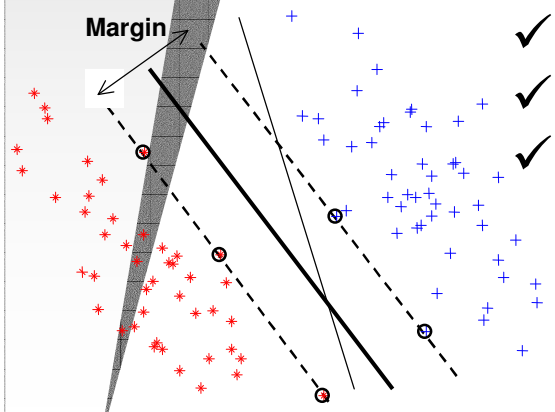


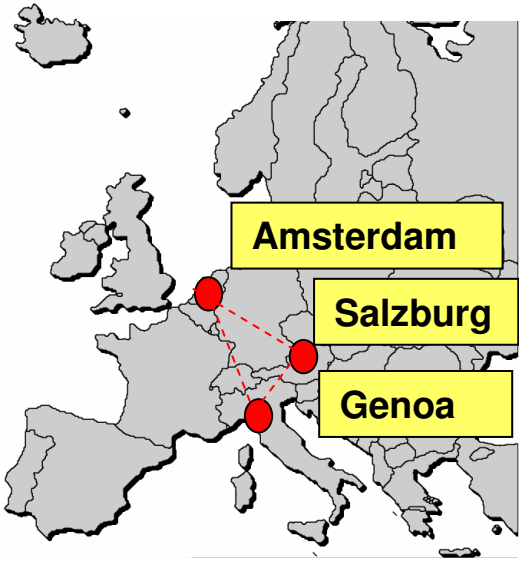
Concept annotation

- ✓ Common annotation effort as basis
- ✓ Extended manually to 101 concepts
- ✓ Incomplete, but reliable

Machine learning architecture

- ✓ Support Vector Machine
- ✓ Learn optimal parameters
- ✓ Using 3 x 3-fold cross validation
- ✓ Or grid-search on a 'grid'





Learning 101 Concepts


Aircraft


Allawi


Anchor


Animal

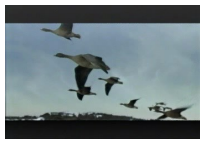
Arrafat


Baseball


Basketball


Beach

Bicycle


Bird

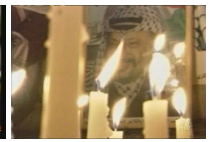

Boat


Building


Bus


Bush jr.

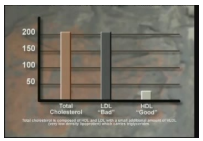

Bush sr.


Candle


Car


Cartoon


Chair



Chart

Clinton


Cloud


Corp. leader

Court


Crowd


Cycling


Desert

Dog

Drawing


Drawing/ cartoon


Duo-anchor


Entertainment


Explosion


Face


Female

Learning 101 Concepts

36 - 70

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion


Fire weapon

Fish


Flag

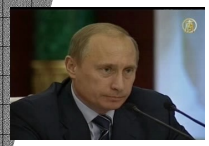
Flag USA


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

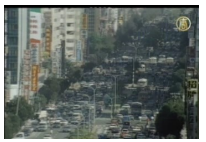

Mountain


Natural disaster

News paper

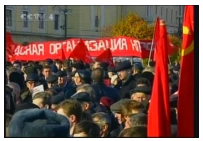

Night fire


Office


Outdoor


Overlayed text


People


People marching


Police / security

Learning 101 Concepts

71 - 101

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion


Powell

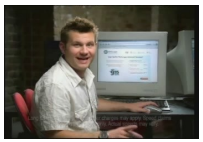

Prisoner


Racing


Religious leader


River


Road


Screen


Sharon


Sky


Smoke


Snow


Soccer


Split screen


Sports


Studio


Swimming pool


Table


Tank

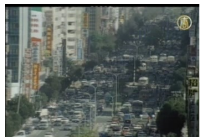

Tennis


Tony Blair


Tower


Tree


Truck


Urban

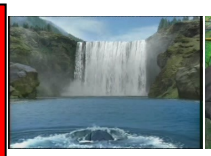

Vegetation


Vehicle


Violence


Walking


Water body

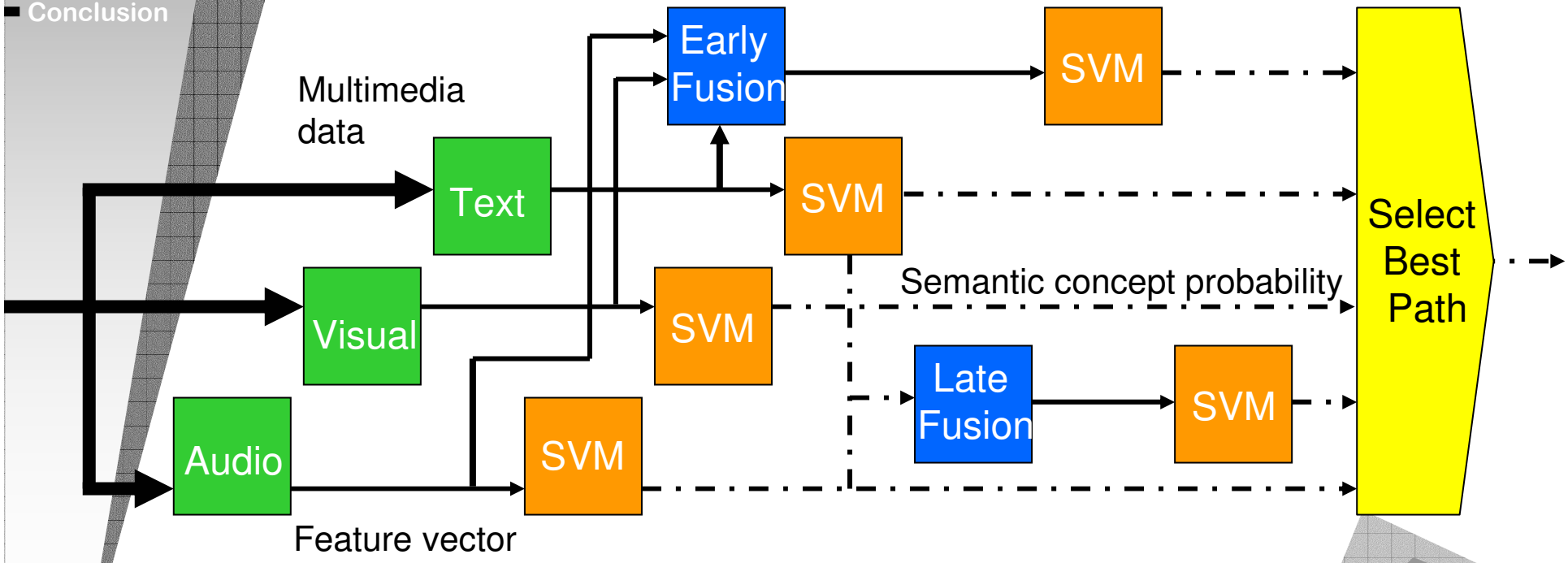

Waterfall


Weather news

Content Analysis Pathfinder
TRECVID 2005

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

- Determine best content analysis path for all 101 concepts
  - ✓ Unimodal analysis
  - ✓ Multimodal analysis
  - ✓ Machine learning



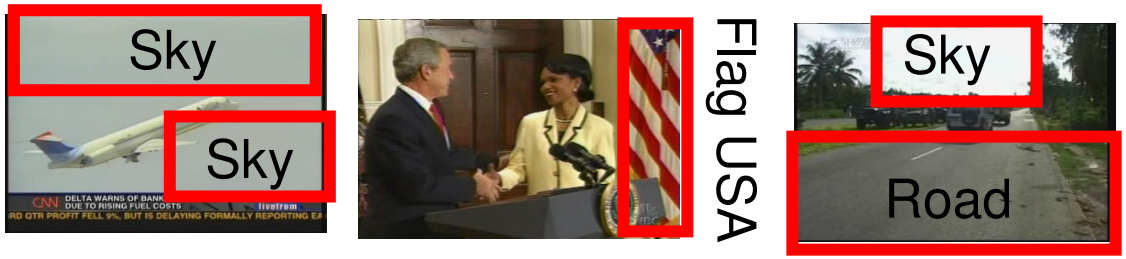
References:
J.C. van Gemert et al, submitted

Visual Feature Analysis

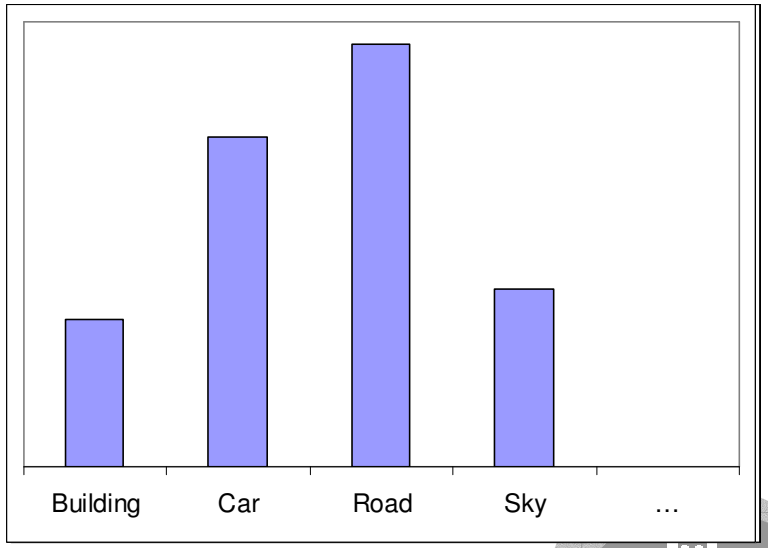
- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Proto-Concepts: semantic image region captured by

✓ Combination of color invariance and natural image statistics



Contexture: Occurrence Histogram of Proto-Concepts

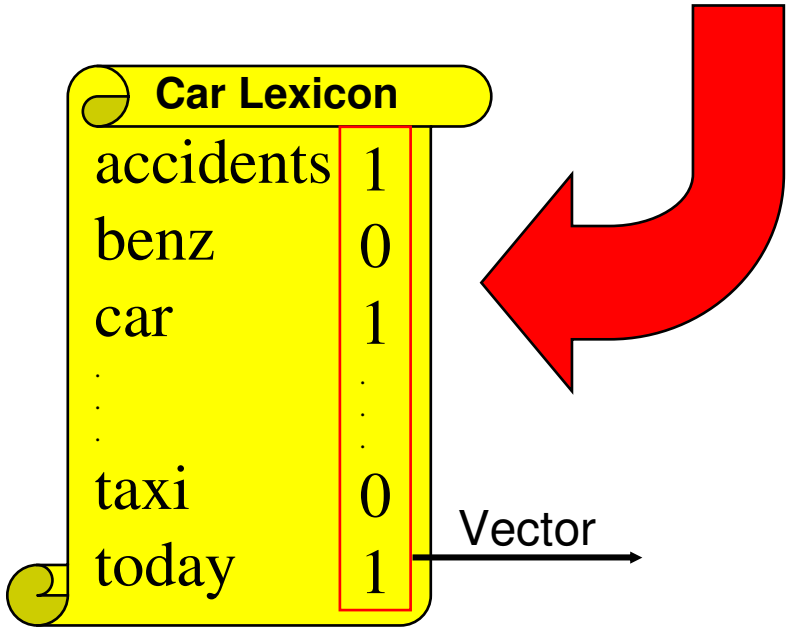


More visual analysis
in our BBC Rushes talk

Textual Feature Analysis

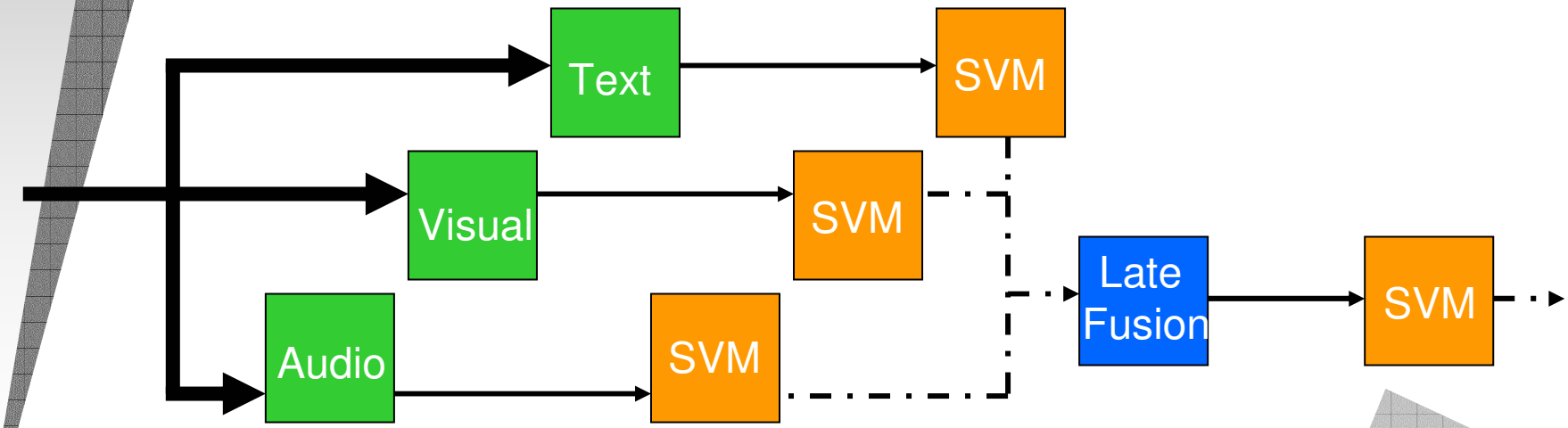
- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

"Car buckles help reduce deadly accidents, state officials announced today."



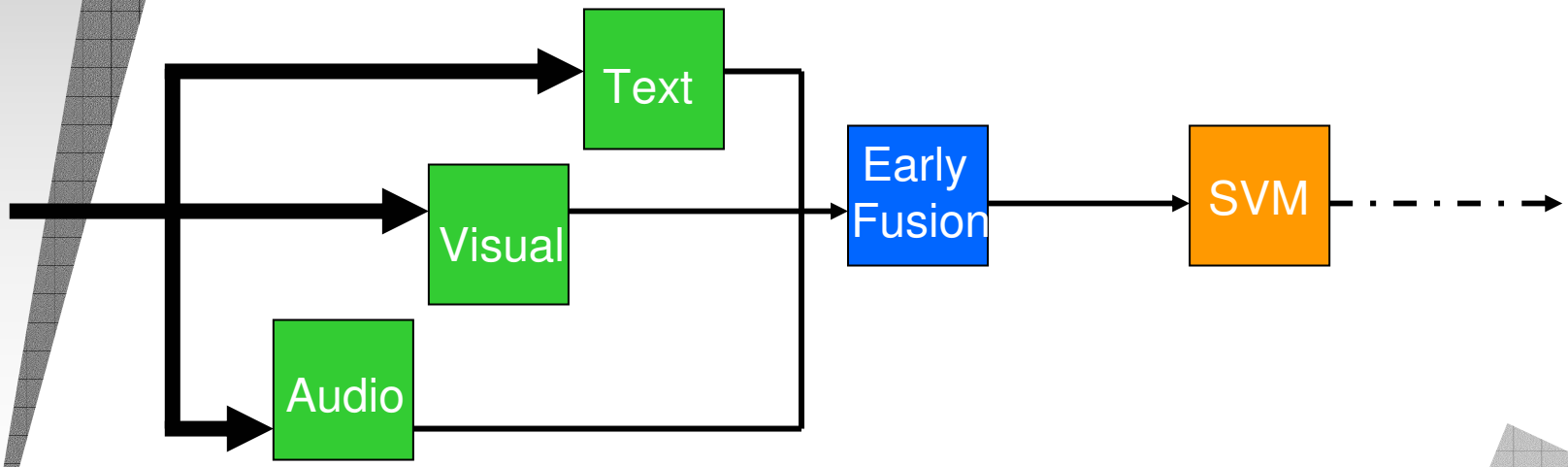
Late Fusion

- Pro's
  - ✓Focus on individual strength of modalities
  - ✓Fusion in semantic space
- Con's
  - ✓Expensive in terms of learning effort
  - ✓Possible correlation in mixed feature space is lost



Early Fusion

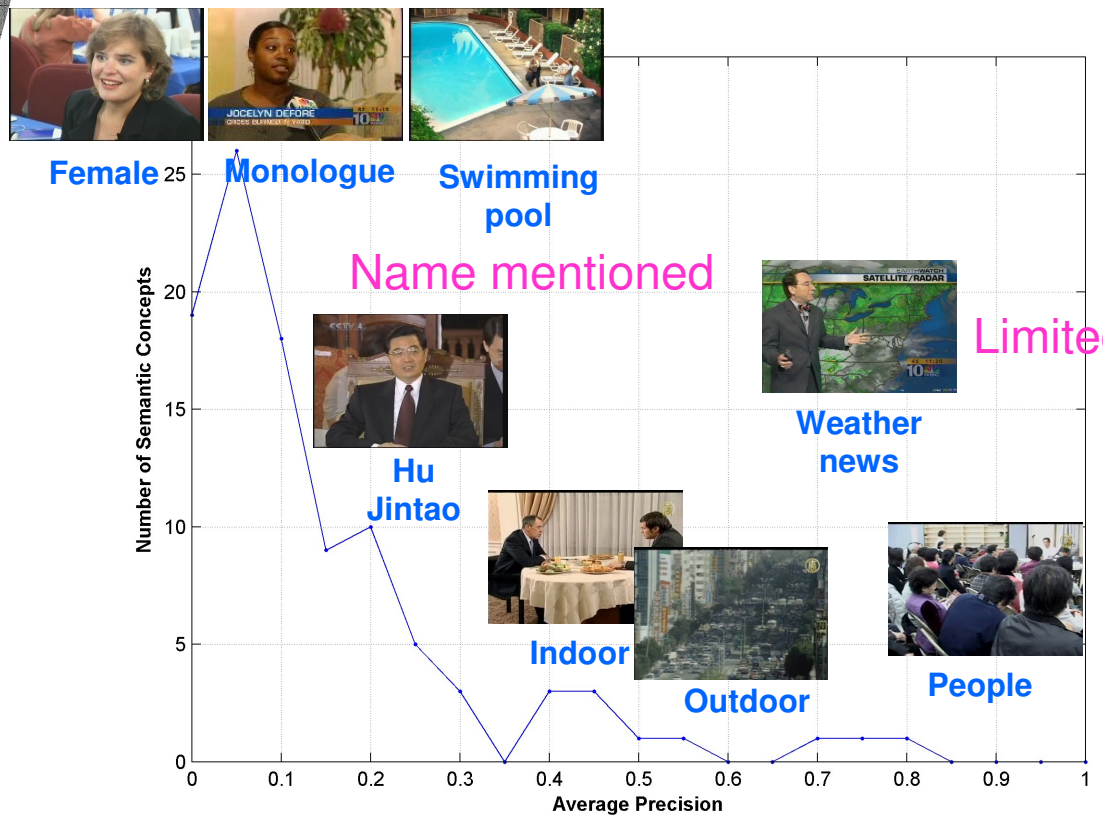
- Pro's
  - ✓Requires only one learning phase
  - ✓Truly multimedia representation used for learning semantics
- Con's
  - ✓Multimodal features combination rather ad hoc
  - ✓One modality is likely to dominate representation



Textual Analysis Results

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Obvious or Sparse



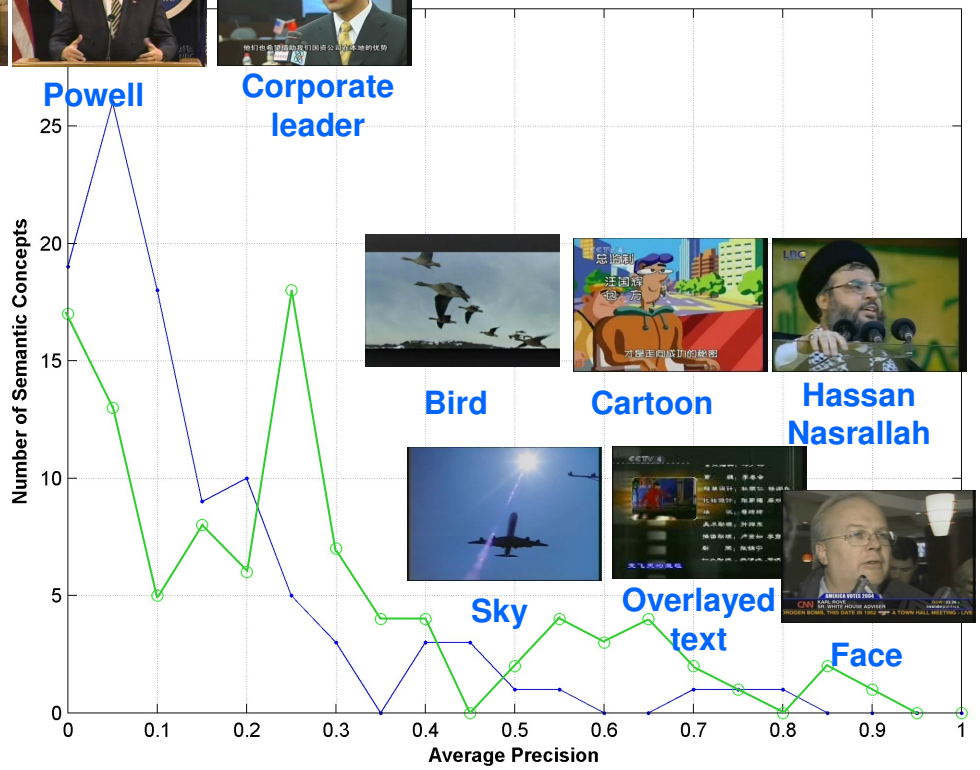
Validation set D MAP: 0.143

Visual Analysis Results

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Sparse, person x, or vague definition

Candle | Powell | Corporate leader



Near-copy / commercials

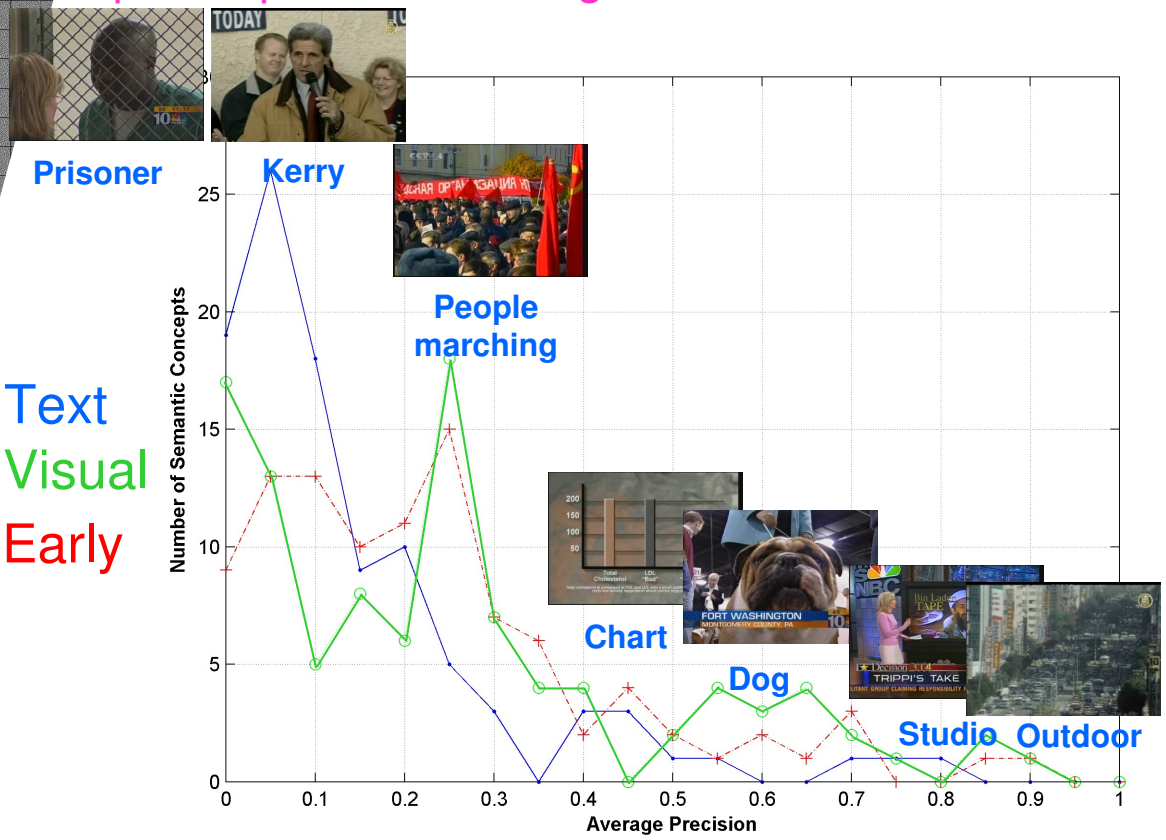
Common concepts

Validation set D MAP: 0.254

Early Fusion Results

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Sparse, person x, or vague definition



Less common / added description

Common concepts

Validation set D MAP: 0.231

Late Fusion Results

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Sparse, person x, or vague definition

Female

Prisoner

Police / security

Less common / added description

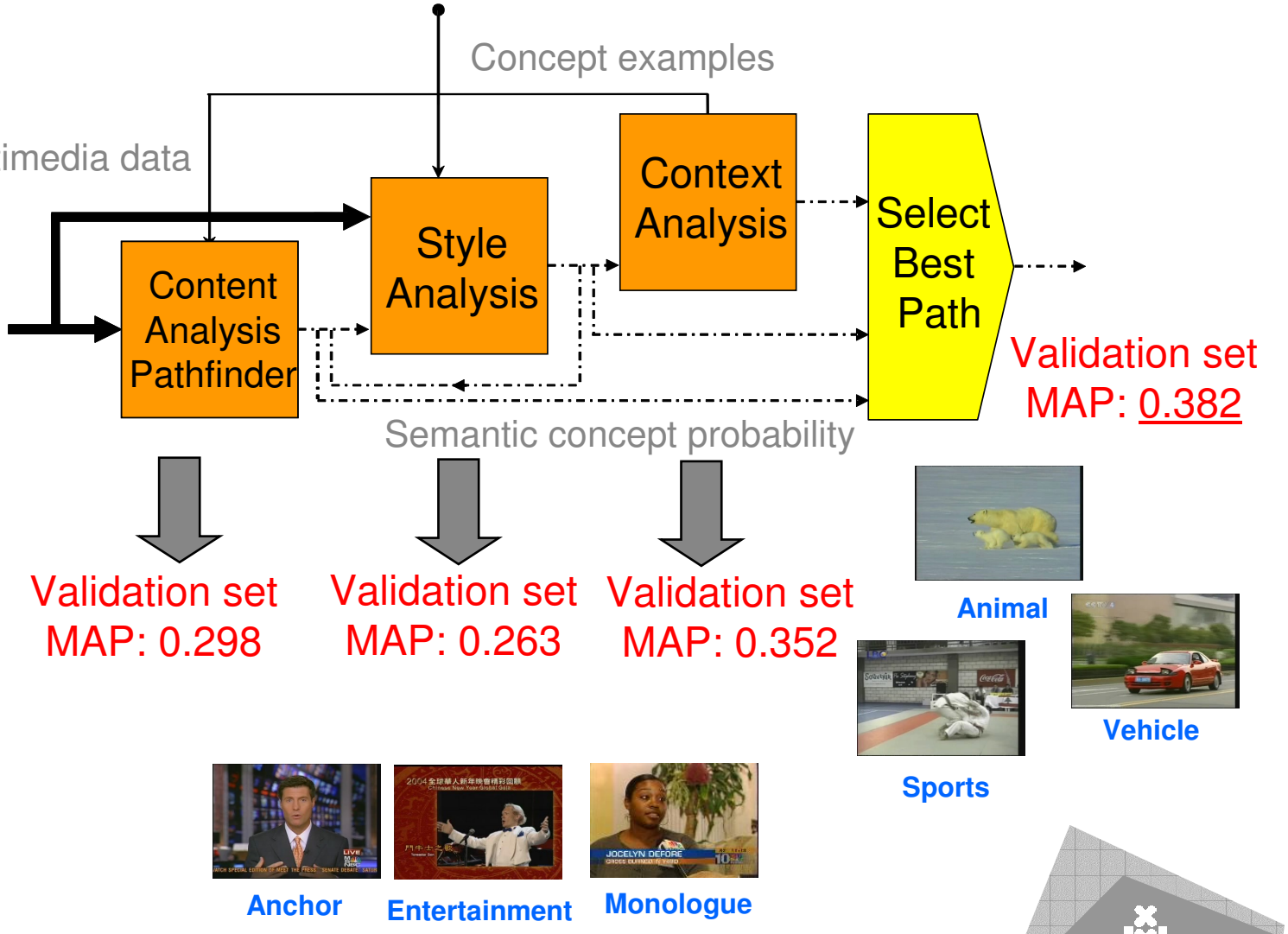
Common concepts

Validation set D MAP: 0.224

Semantic Pathfinder

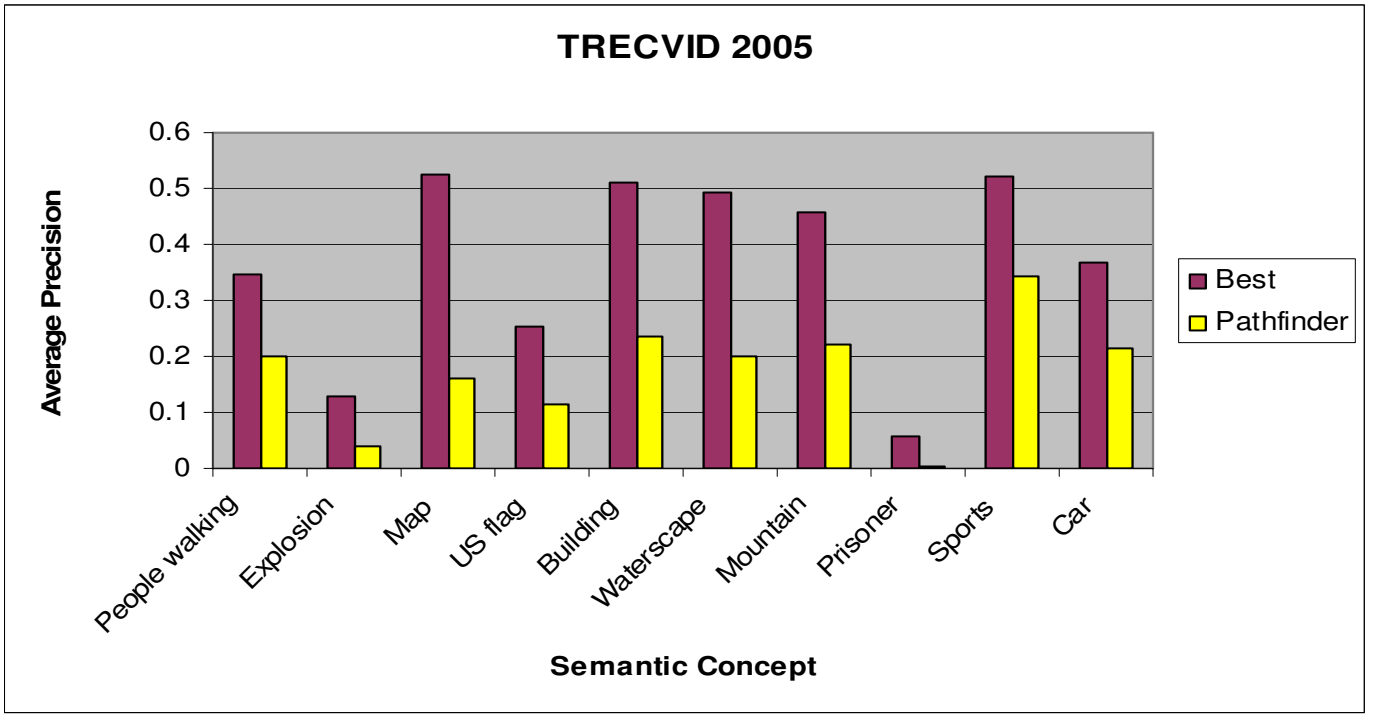
TRECVID 2005

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion



Benchmark Results

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion



- Benchmark performance
  - ✓ Completely generic approach,
  - ✓ No fine tuning on Chinese, Arabic, or English
  - ✓ Average precision only one side of the coin…

User Satisfaction?

Map

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Pathfinder

Best performer

User Satisfaction?

Mountain

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Pathfinder

Best performer

Conclusions & future work

- Introduction
- Lexicon
- Pathfinder
- Results
- Conclusion

Semantic pathfinder facilitates generic indexing

- ✓ Currently detects up to 101 concepts
- ✓ Some concepts are content, others are style, or context
- ✓ For content a separation between analysis steps exists also

To fuse, or not to fuse?

- ✓ No best method for all concepts exists,
- ✓ Best to learn optimal approach per concept
- ✓ Sparse, person x, and ill-defined concepts still problematic
- ✓ Power of visual modality underestimated

Input for discussion

- ✓ Focus on specific indexing methods is hampering progress? Focus on commercials or anchors in concept detection results is hampering progress?

Thanks for your attention

- More info on semantic video indexing:

  Cees Snoek

  - http://www.science.uva.nl/~cgmsnoek
  - cgmsnoek@science.uva.nl

Come see our demo's

- More info on visual features:
  - ✓ See BBC rushes talk by Jan van Gemert
- More info on video retrieval:
  - ✓ See search talk by Marcel Worring