TRECVID 2005 Workshop

Columbia University High-Level Feature Detection: Parts-based Concept Detectors

Dong-Qing Zhang, Shih-Fu Chang,
Winston Hsu, Lexin Xie, Eric Zavesky

Digital Video and Multimedia Lab
Columbia University

(In collaboration with IBM Research in ARDA VACE II Project)

Columbia University in the City of New York

dvmm DIGITAL VIDEO • MULTIMEDIA LAB

data source and design principle

- Multi-lingual multi-channel video data
  - 277 videos, 3 languages (ARB, CHN, and ENG)
  - 7 channels, 10+ different programs
  - Poor or missing ASR/MT transcripts
- A very broad concept space over diverse content
  - object, site, people, program, etc
  - TV05 (10), LSCOM-Lite (39), LSCOM (449)
- Concept detection in such a huge space is challenging
  - Need a principled approach
  - Take advantage of the extremely valuable annotation set
  - Data-driven learning based approach offers potential for scalability

Insights from Samples: Object - flag




- Unique object appearance and structure
  - Some even fool the annotator
- Variations in scale, view, appearance, number
- Noisy labels
- Sometimes contextual, spatial cues are helpful for detection
  - Speaker, stage, sky, crowd

Site/location

- Again visual appearance and spatial structures very useful

Activity/Event

- Visual appearances capture the after effects of some events – smoke, fire
- Sufficient cues for detecting occurrences of events
- Other events (e.g., people running) need object tracking and recognition

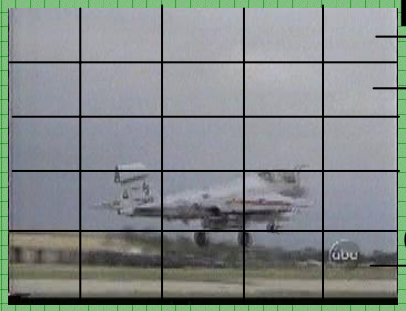
Motivation for Spatio-Appearance Models

- Many visual concepts characterized by
  - Unique spatial structures and visual appearances of the objects and sites

  - joint occurrences of accompanying entities with spatial constraints
- Motivate the deeper analysis of spatio-appearance models

Spatio-Features: How to sample local features?

traditional



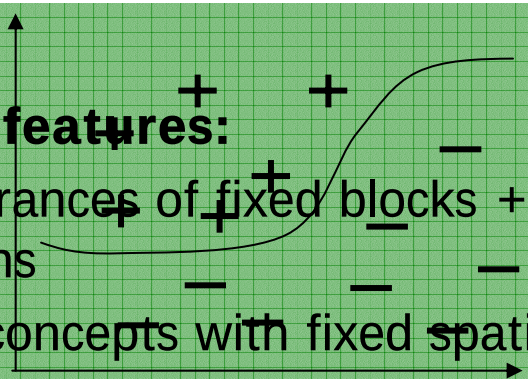
Color Moment

...

Color Moment

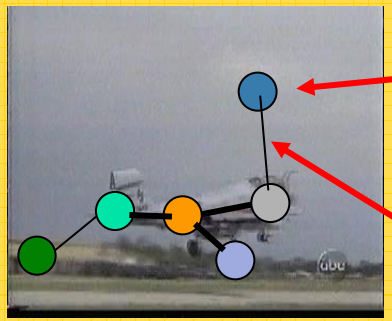
Block-based features:

- visual appearances of fixed blocks + block locations
- suitable for concepts with fixed spatial patterns



Support Vector Machine (SVM)

Adaptive Sampling: Object Parts



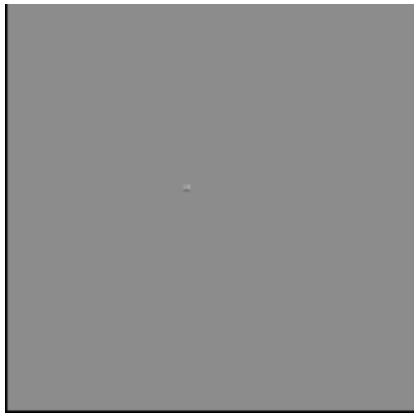
Part

Part relation

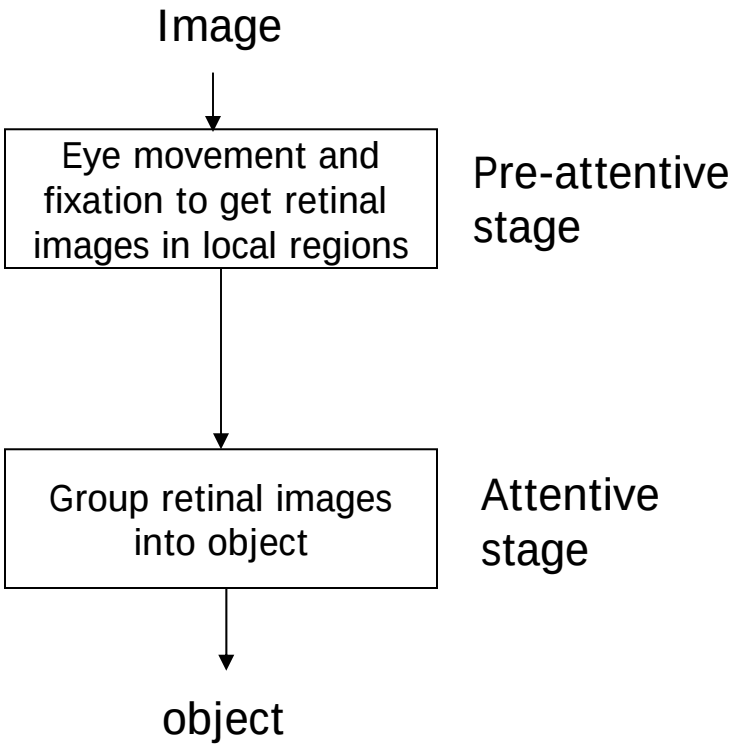
Part-based model:

- Model appearance at salient points
- Model part relations
- Robust against occlusion, background, location change

- Parts-based object detection paradigm also related to Human Vision System (HVS)



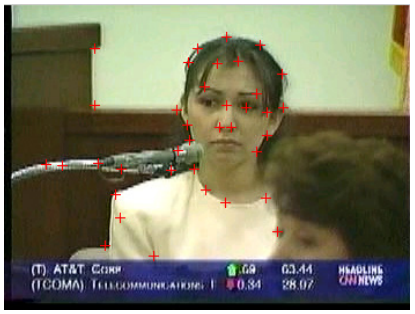
[Rybak et al. 98']



Our TRECVID 2005 Objectives

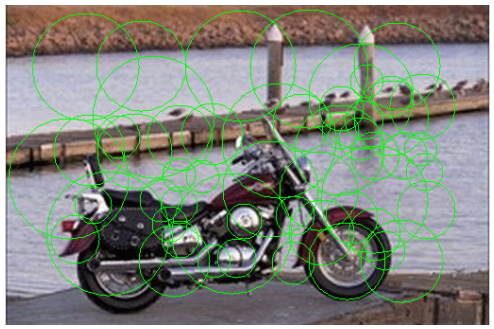
- Explore the potential strengths of parts-based models in
  - detecting spatio-dominant concepts
  - fusing with traditional fixed features models
  - detecting other interesting patterns such as *Near-Duplicates* in broadcast news

How do we extract and represent parts?



Interest points

Segmented Regions



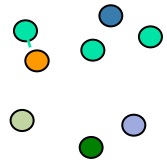
Maximum Entropy Regions

Part detection

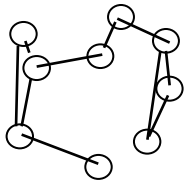
Gabor filter,
PCA projection,
Color histogram,
Moments ...

Feature Extraction within local parts

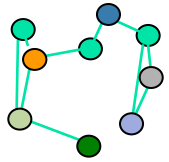
Part-based representation



Bag

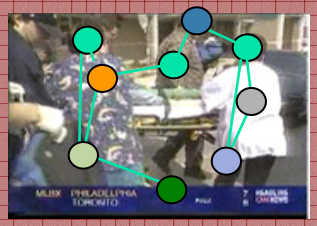


Structural Graph

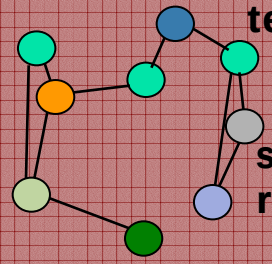


Attributed Relational Graph

Representation and Learning



Individual images
→ Salient points, high entropy regions

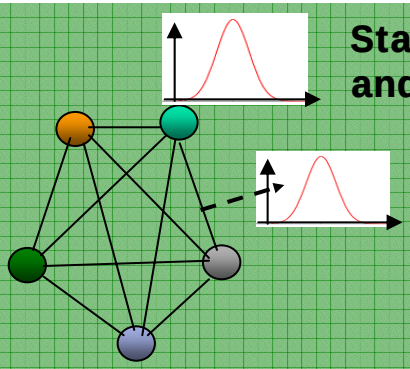


Attributed Relational Graph (ARG)

Graph Representation of Visual Content

Collection of training images

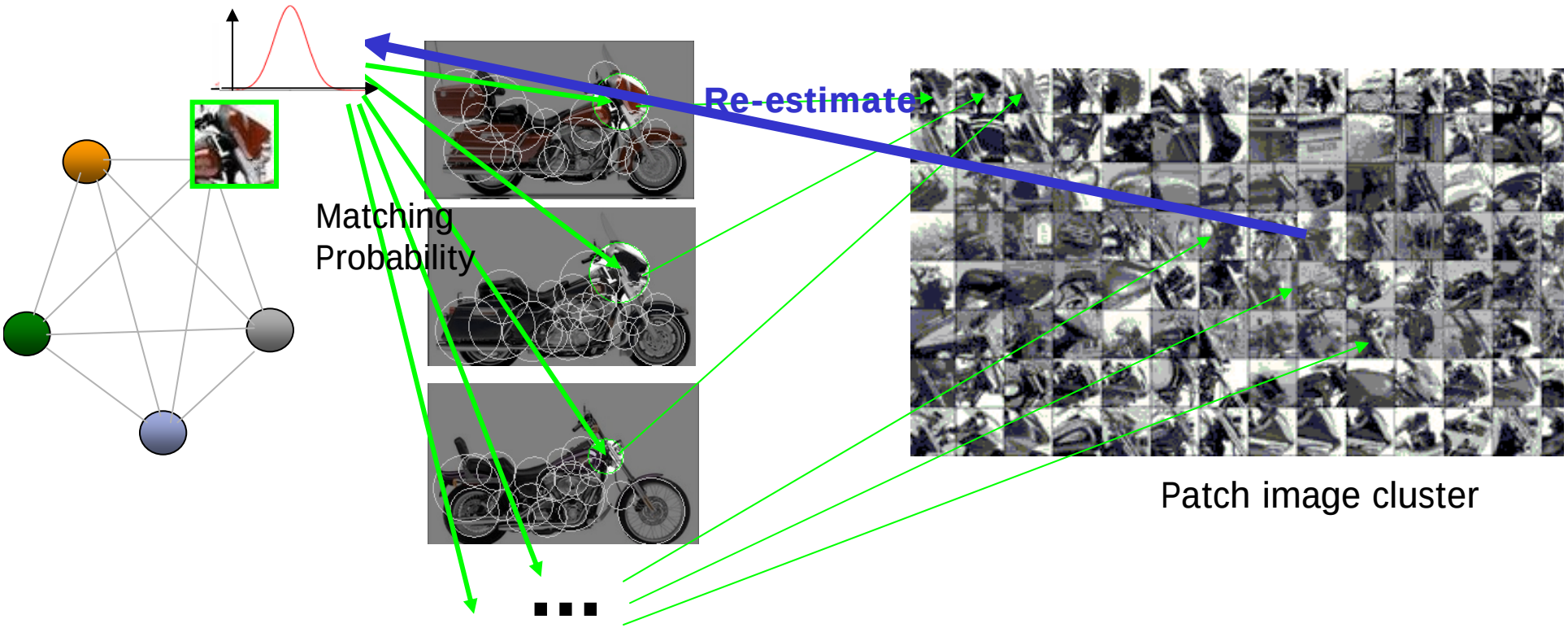
machine learning



Random Attributed Relational Graph (R-ARG)

Statistical Graph Representation of Model

Learning Object Model

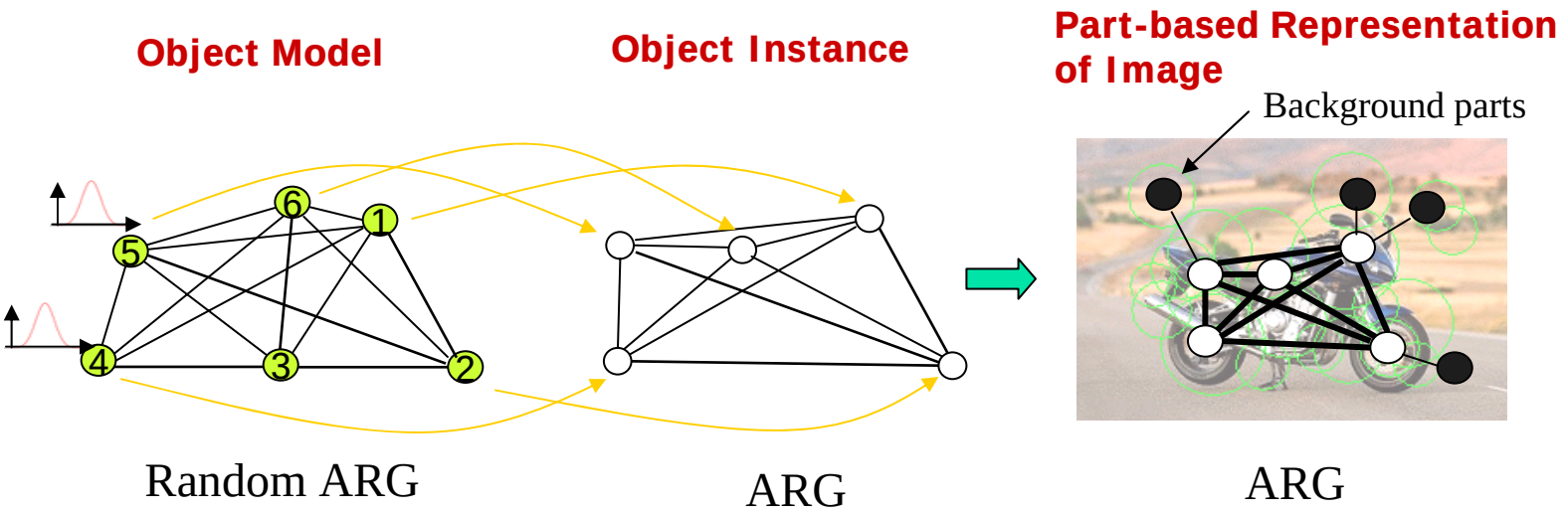


- **Challenge : Finding the correspondence of parts and computing matching probability are NP-complete**
- **Solution :**
  - **Apply and develop advanced machine learning techniques – Loopy Belief Propagation (LBP), and Gibbs Sampling plus Belief Optimization (GS+BO)**

(demo)

Role of RARG Model: Explain object generation process

- Generative Process : From object model to image



Sampling node occurrence
And node/edge features

Sampling background pdf and add background parts

Object Detection by Random AG

Binary detection problem : contain or not contain an object ?

H=1 **H=0**

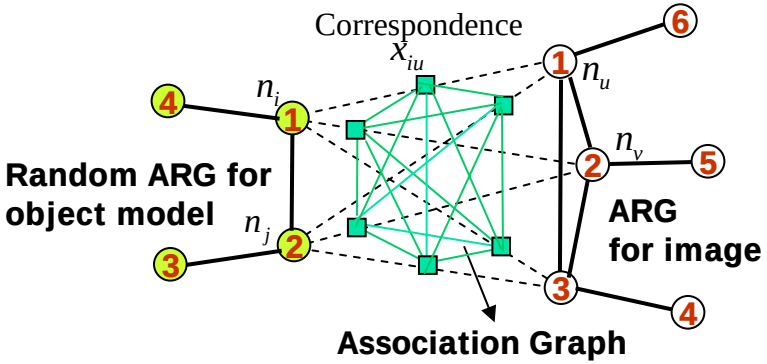
Likelihood ratio test :

$$\frac{p(O|H=1)}{p(O|H=0)} > \frac{p(H=0)}{p(H=1)} = \lambda$$, O: input ARG

Object likelihood :

$$P(O \mid H=1) = P(X \mid H=1)P(O \mid X, H=1)$$

X modeled by Association Graph



- Probabilities computed by MRF
- Likelihood ratio can be computed by variational methods (LPB, MC)

Columbia University in the City of New York

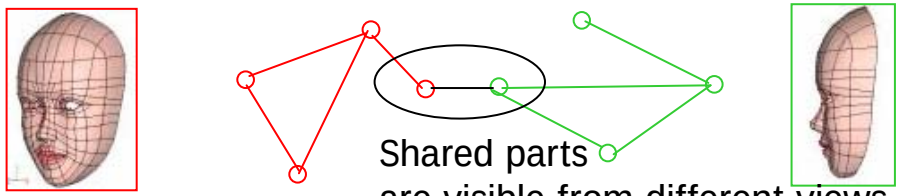
dvmm DIGITAL VIDEO • MULTIMEDIA LAB

Extension to Multi-view Object Detection

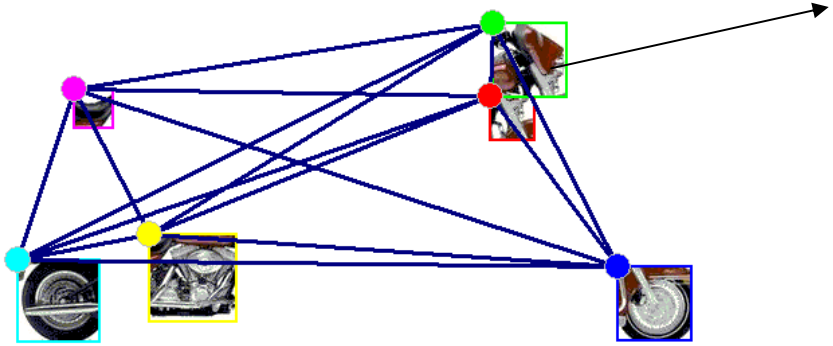
Challenge of multi-view object/scene detection

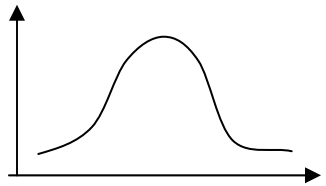
- Objects under different views have different structures
- Part appearances are more diverse

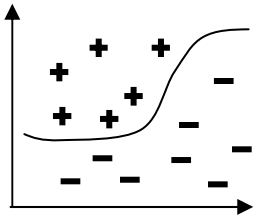
Structure variation could be handled by Random ARG model (each view covered by a sub-graph)



Adding Discriminative Model for Multi-view Concept Detection



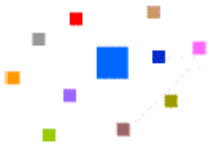




Previous :
Part appearance modeling by Gaussian distribution

Now :
Part appearance modeling by Support Vector Machine

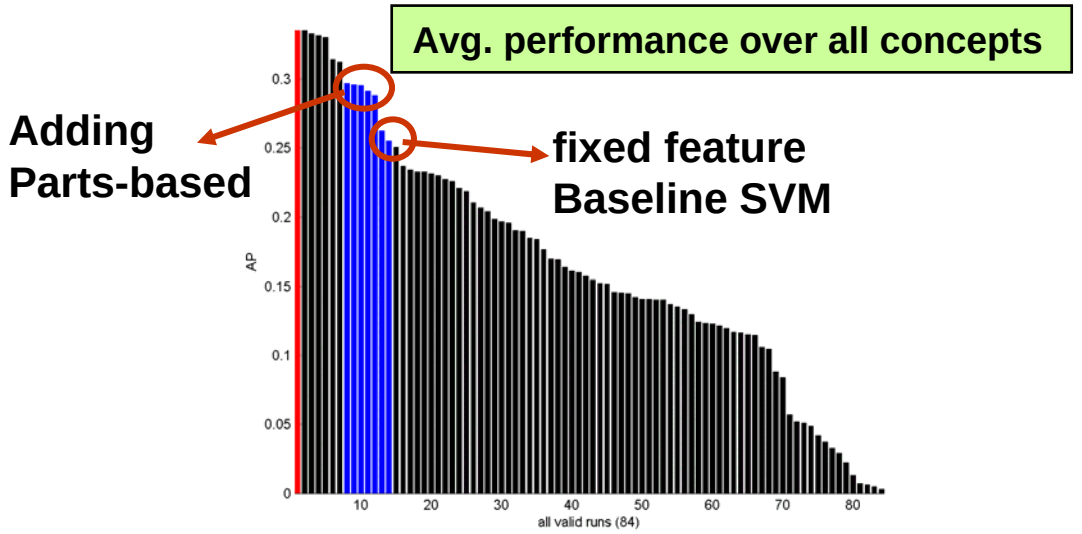
- Use SVM plus non-linear kernels to model diverse part appearance in multiple views
- principle similar to boosting

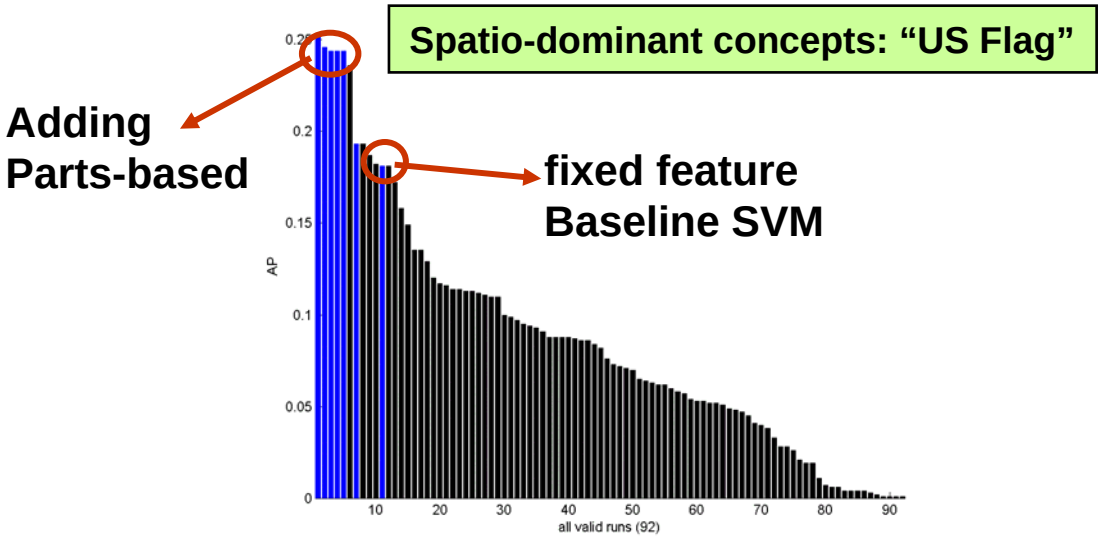


Evaluation in TRECVID 2005

Parts-based detector performance in TRECVID 2005

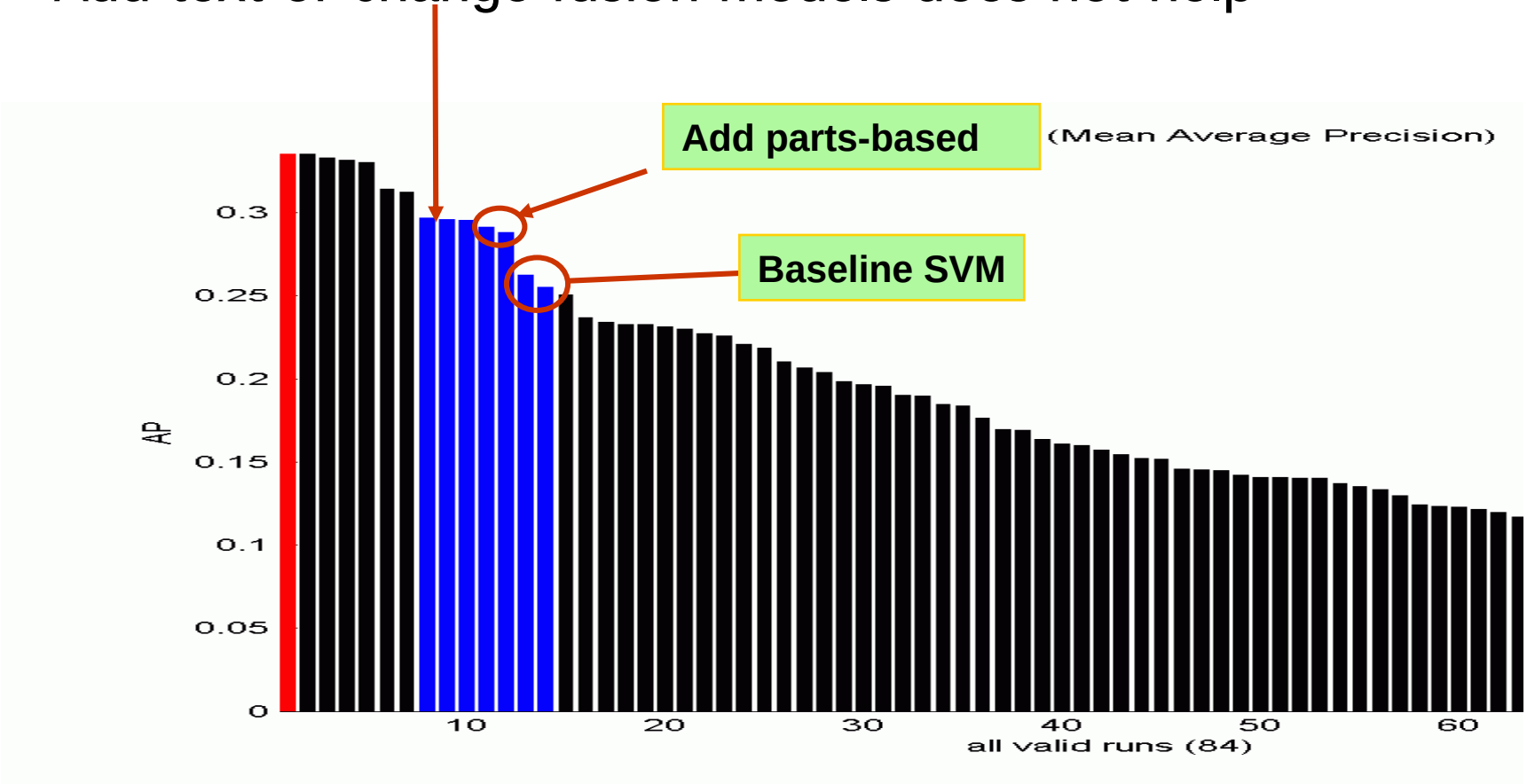
- Parts-based detector consistently improves by more than 10% for all concepts
- It performs best for spatio-dominant concepts such as “US flag”.
- It complements nicely with the discriminant classifiers using fixed features.



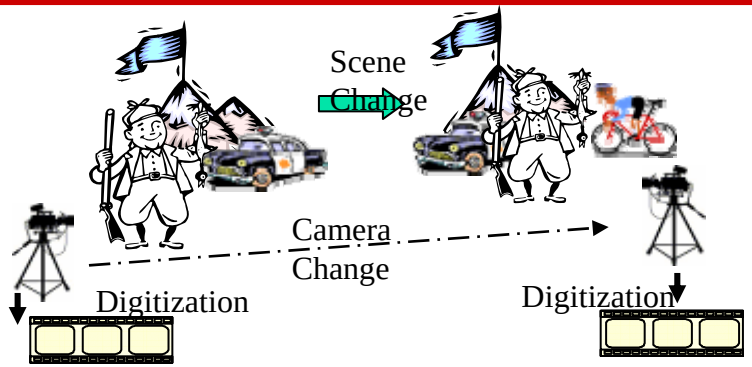


Relative contributions

- Add text or change fusion models does not help



Other Applications of Parts-Based Model: Detecting Image Near Duplicates (IND)

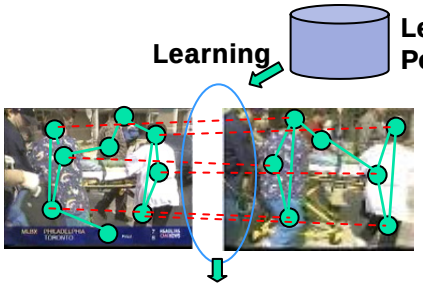


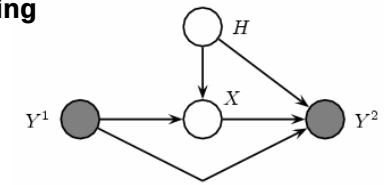
Many Near-Duplicates in TRECVD 05

- **Near duplicates occur frequently in multi-channel broadcast**
- **But difficult to detect due to diverse variations**
- **Problem Complexity**

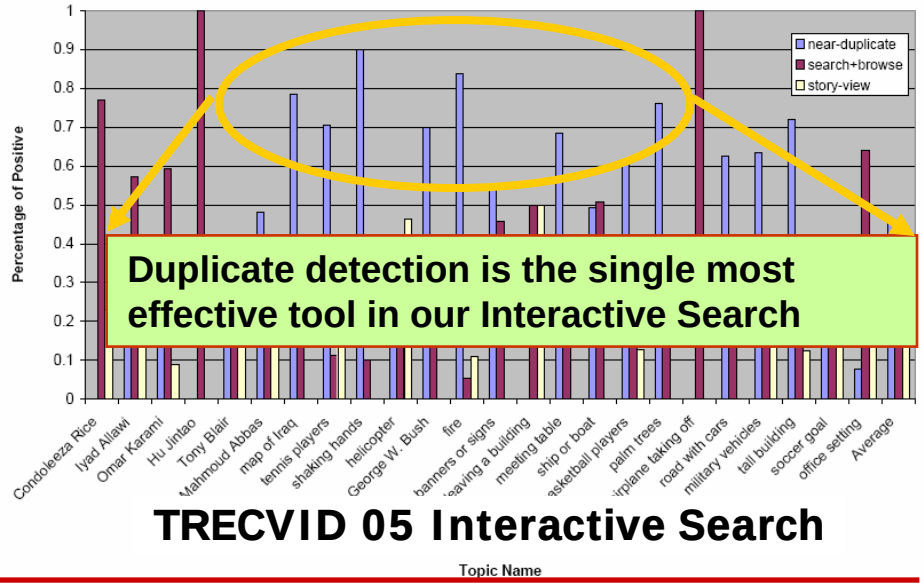
Similarity matching < IND detection < object recognition

Parts-based Stochastic Attribute Relational Graph Learning



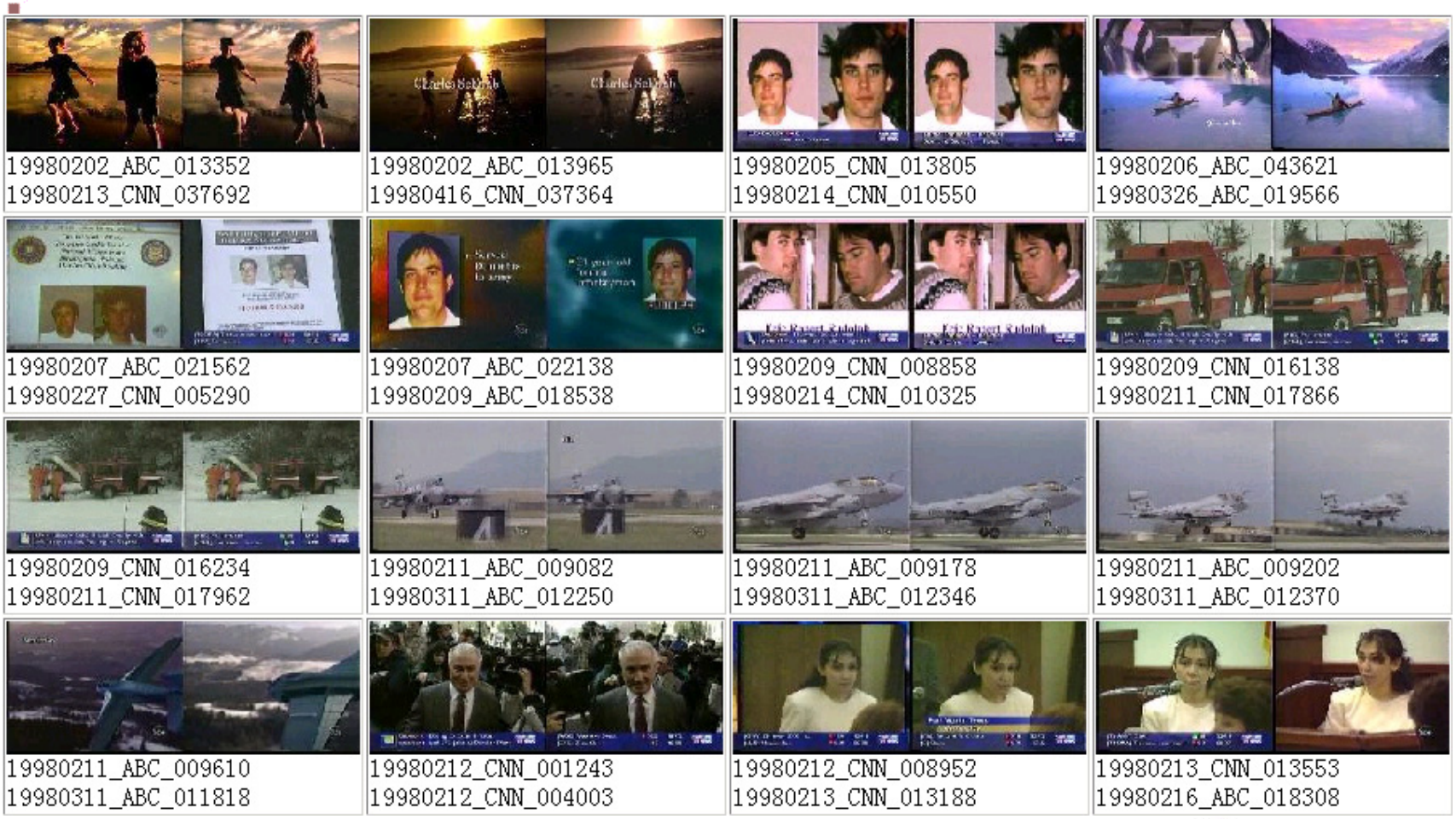


Stochastic graph models the physics of scene transformation

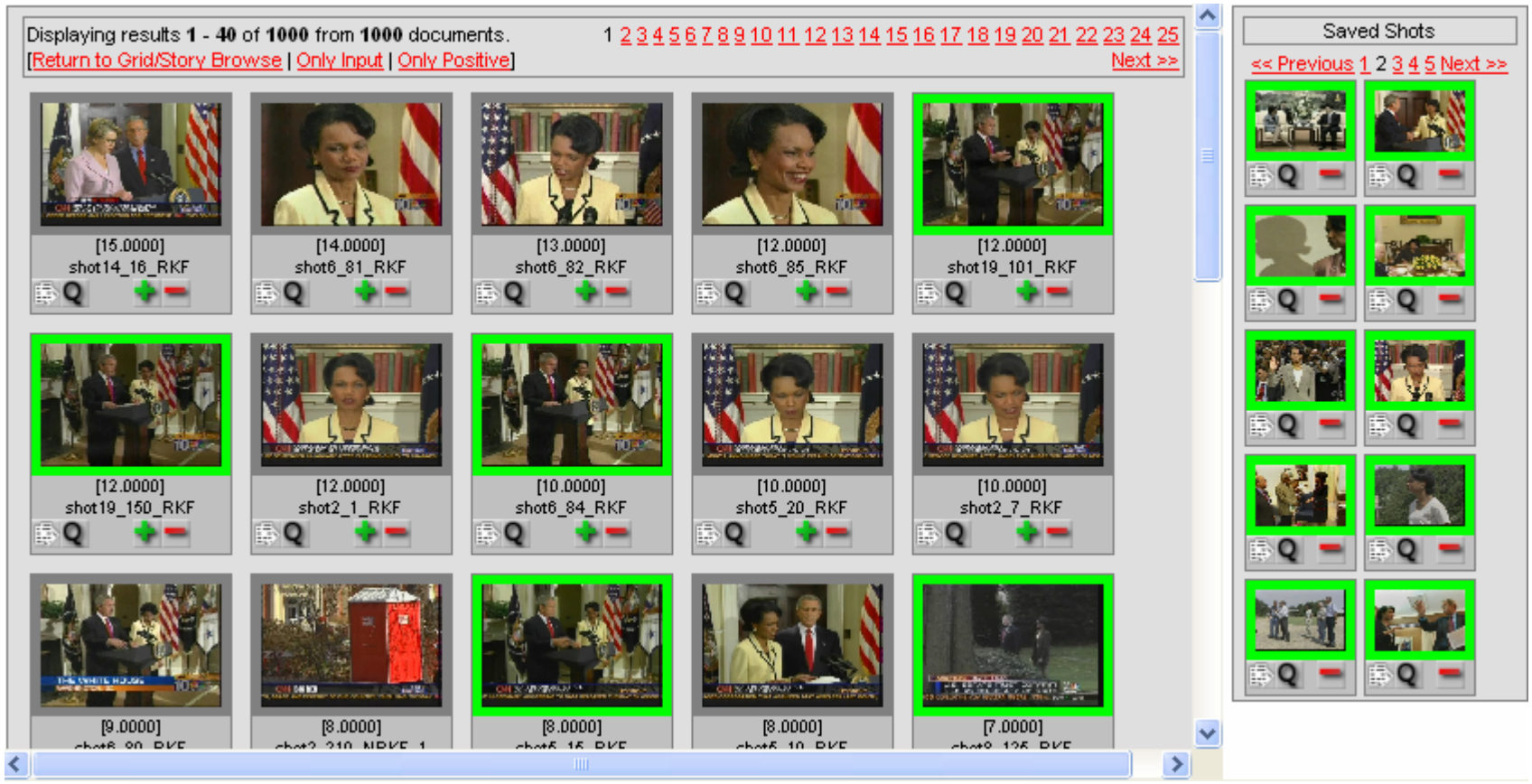


Near Duplicate Benchmark Set

(available for download at Columbia Web Site)



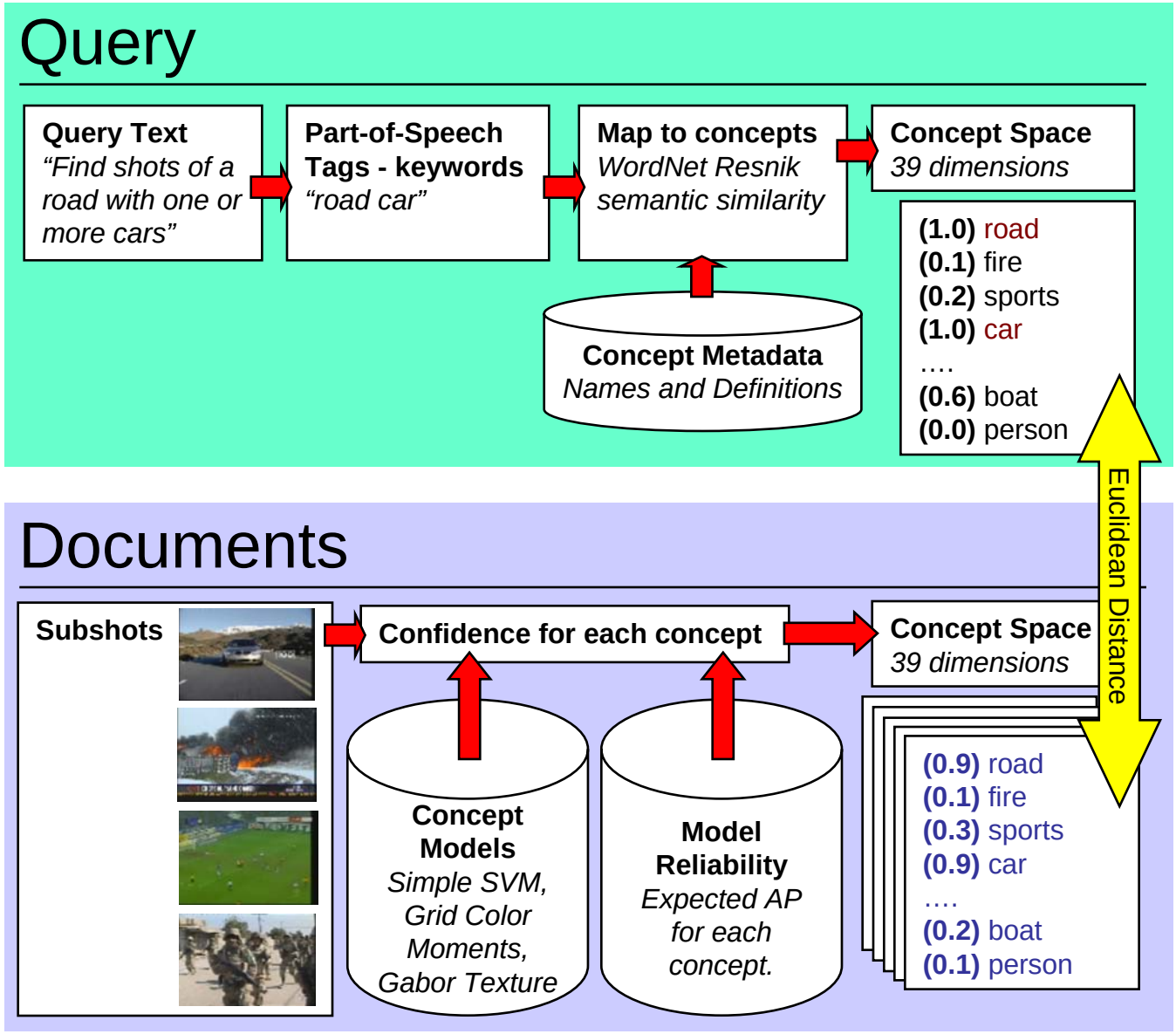
Examples of Near Duplicate Search in TRECVID 05



Columbia University in the City of New York

dvmm DIGITAL VIDEO • MULTIMEDIA LAB

Application: Concept Search



- **Map text queries to concept detection**
- **Use human-defined keywords from concept definitions**
- **Measure semantic distance between query and concept**
- **Use detection and reliability for subshot documents**

Concept Search

Automatic - help queries with related concepts

“Find shots of boats.”

Method	AP
Story Text	.169
CBIR	.002
Concept	.115
Fused	**.195**

“Find shots of a road with one or more cars.”

Method	AP
Story Text	.053
CBIR	.009
Concept	.090
Fused	**.095**

Manual / Interactive

Manual keyword selection allows more relationships to be found.

Query Text	Concepts
“Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people”	*Office*
“Find shots of one or more people entering or leaving a building”	*Person, Building, Urban*
“Find shots of a graphic map of Iraq, location of Bagdhad marked - not a weather map”	*Map*
Find shots of people with banners or signs	*March or protest*

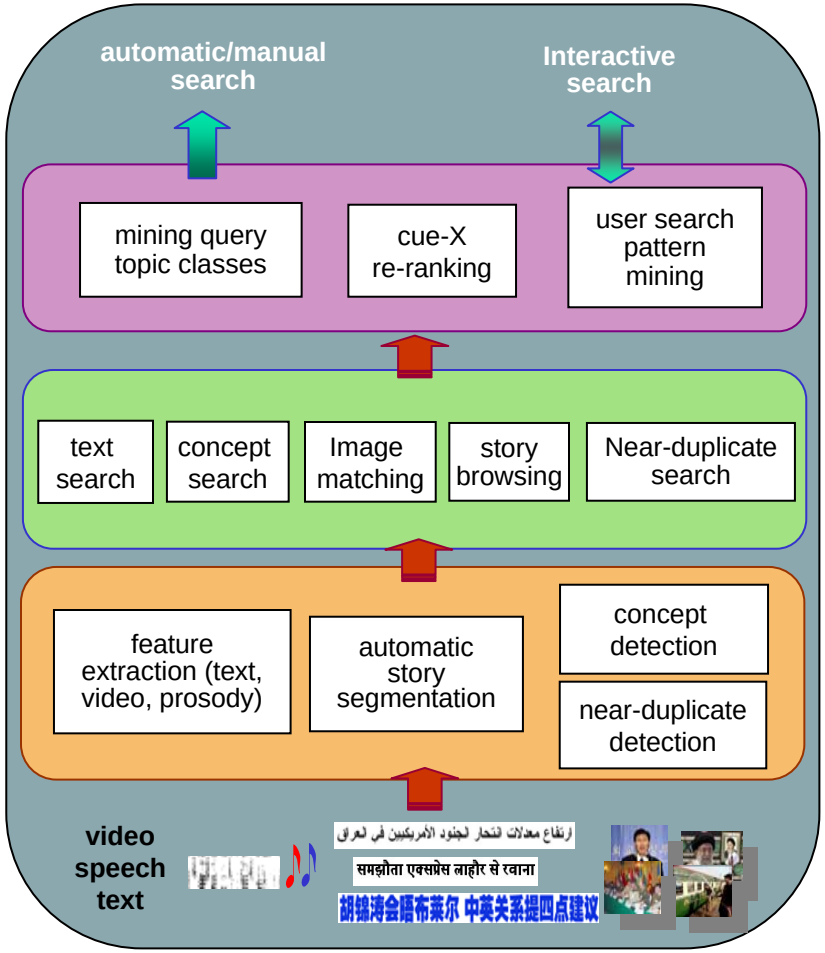
Columbia Video Search Engine System Overview

http://www.ee.columbia.edu/cuvidsearch

User Level Search Objects
- Query topic class mining
- Cue-X reranking
- Interactive activity log

Multi-modal Search Tools
- combined text-concept search
- story-based browsing
- near-duplicate browsing

Content Exploitation
- multi-modal feature extraction
- story segmentation
- semantic concept detection



Demo in the poster session

Search User Interface

Columbia University in the City of New York

dvmm DIGITAL VIDEO • MULTIMEDIA LAB

Conclusions

- Parts-based models are intuitive and general
  - Effective for concepts with strong spatio-appearance cues
  - Complementary with fixed feature classifiers (e.g., SVM)
  - Semi-supervised: the same image-level annotations sufficient, no need for part-level labels
- Parts models also useful for detecting near duplicates in multi-source news
  - Valuable for interactive search