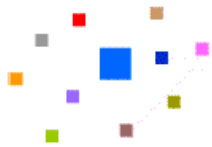


TRECVID 2005 Workshop



# 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)

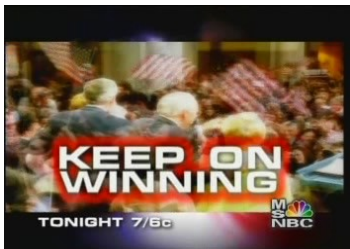


# data source and design principle

- 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

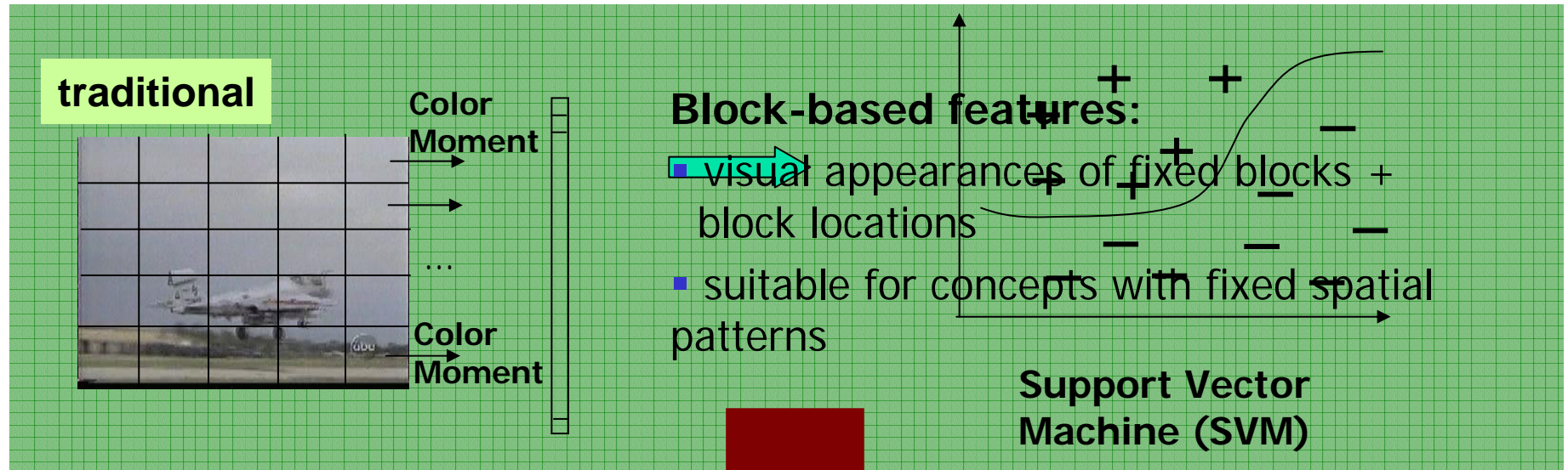
- Many visual concepts characterized by
  - Unique spatial structures and visual appearances of the objects and sites



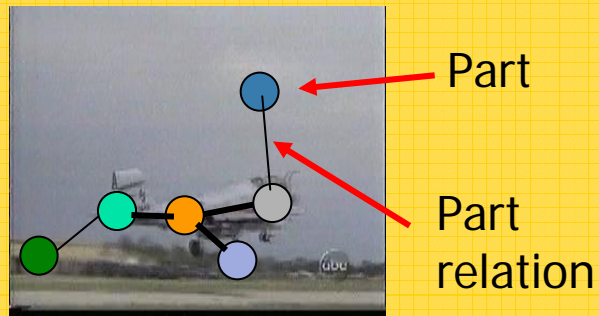
- joint occurrences of accompanying entities with spatial constraints
- Motivate the deeper analysis of spatio-appearance models



# Spatio-Features: How to sample local features?



## Adaptive Sampling: Object Parts



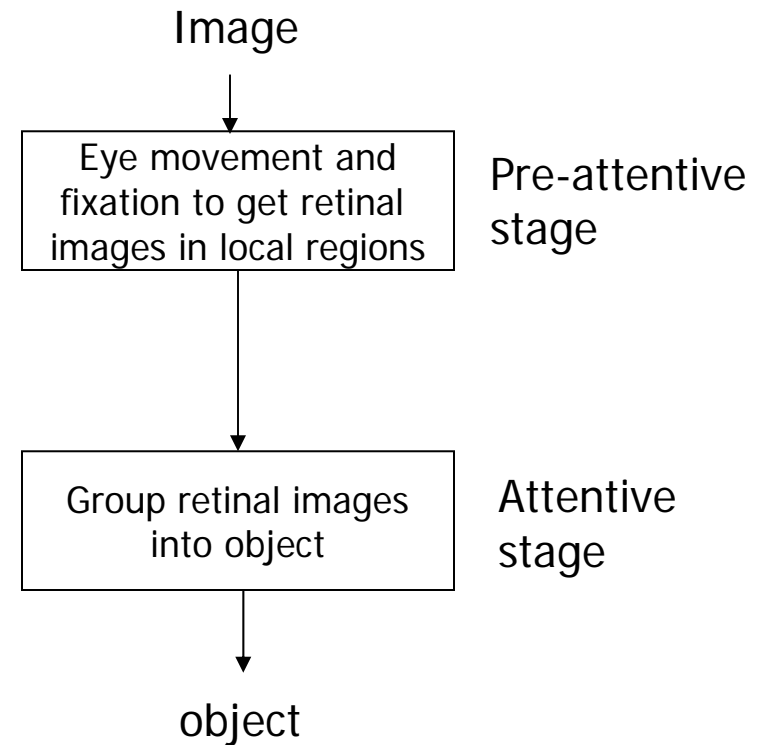
### Part-based model:

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

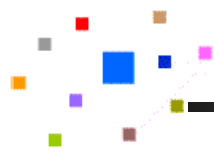
- Parts-based object detection paradigm also related to Human Vision System (HVS)



[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

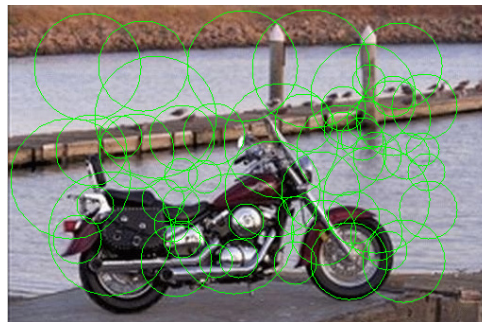
# How do we extract and represent parts?



Interest points



Segmented Regions



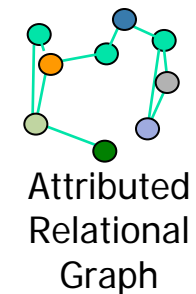
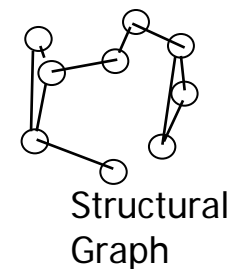
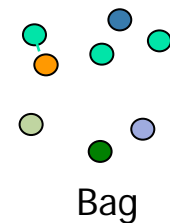
Maximum Entropy Regions

Part detection

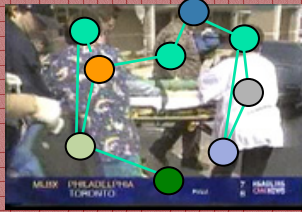
Gabor filter,  
PCA projection,  
Color histogram,  
Moments ...

Feature Extraction  
within local parts

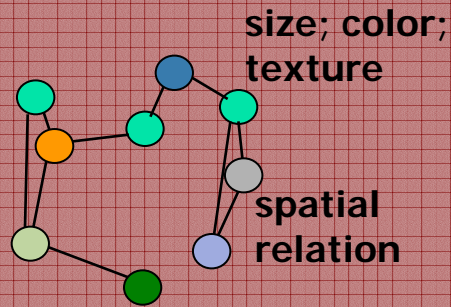
Part-based  
representation



# Representation and Learning



Individual images  
→ Salient points, high entropy regions

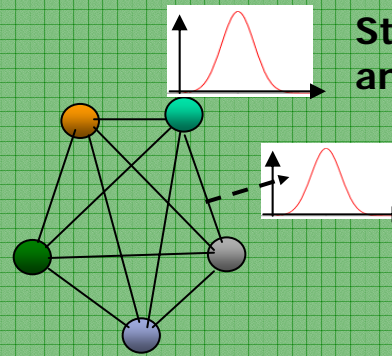


Attributed Relational Graph (ARG)

**Graph Representation of Visual Content**



machine learning

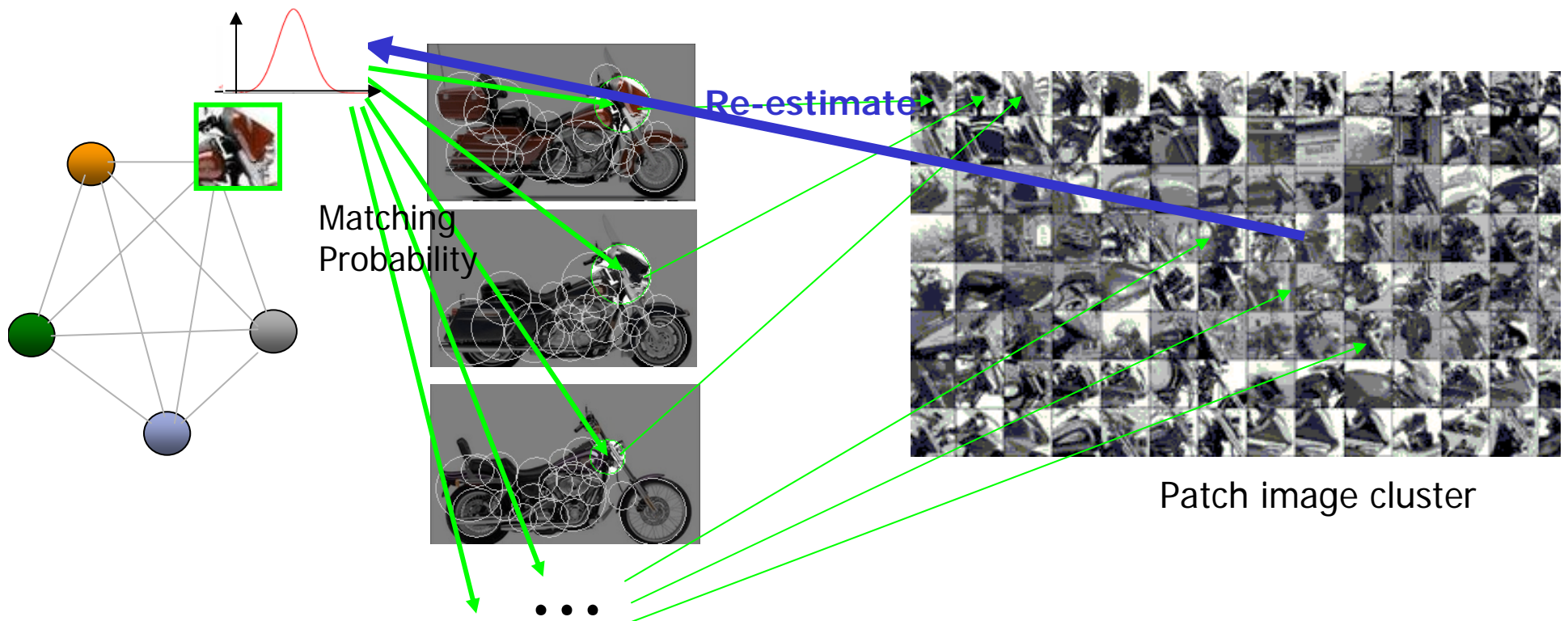


Random Attributed Relational Graph (R-ARG)

Statistics of attributes and relations

**Statistical Graph Representation of Model**

# Learning Object Model



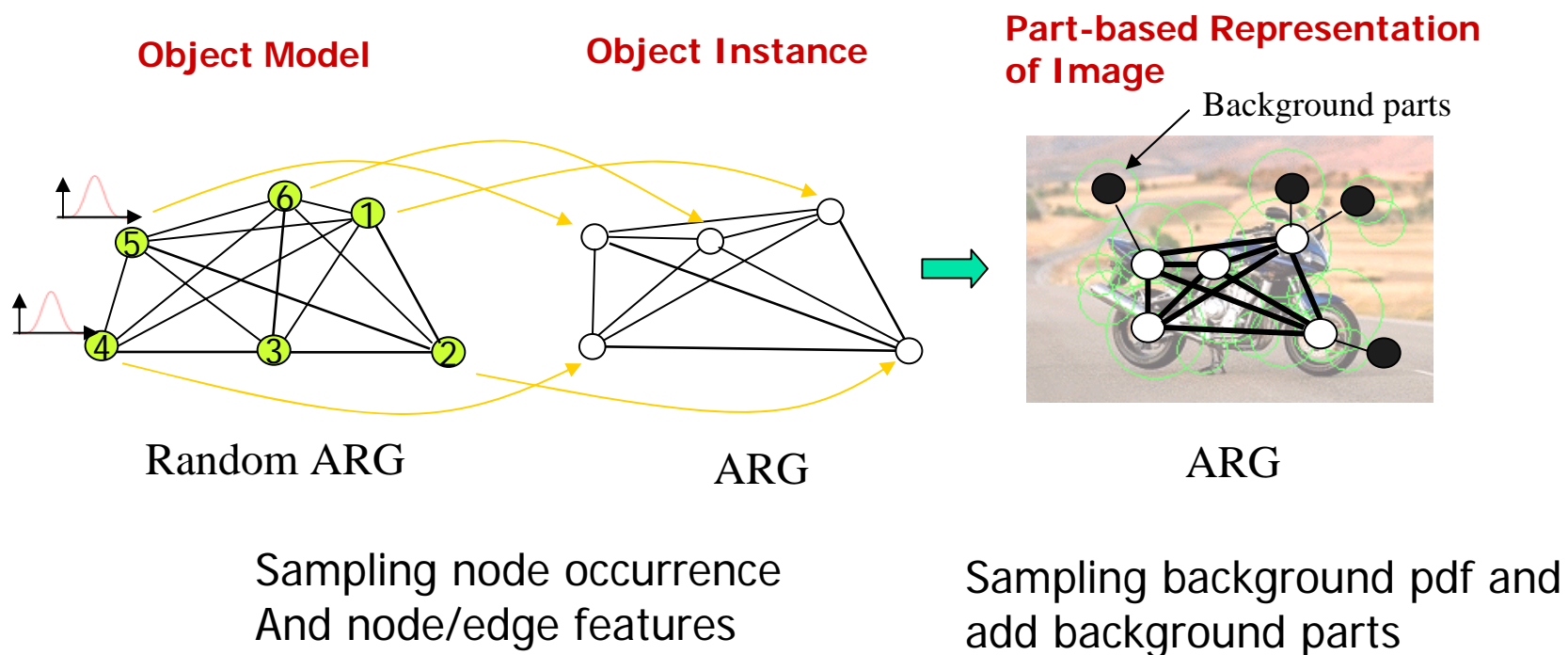
- **Challenge :** Finding the correspondence of parts and computing matching probability are NP-complete
- **Solution :**
  - Apply and develop advanced machine learning techniques – Loopy Belief Propagation (LBP), and Gibbs Sampling plus Belief Optimization (GS+BO)

(demo)

# Role of RARG Model:

## Explain object generation process

- Generative Process : From object model to image



# Object Detection by Random AG

Binary detection problem :  $\frac{\text{contain}}{H=1}$  or  $\frac{\text{not contain}}{H=0}$  an object ?

Likelihood ratio test :

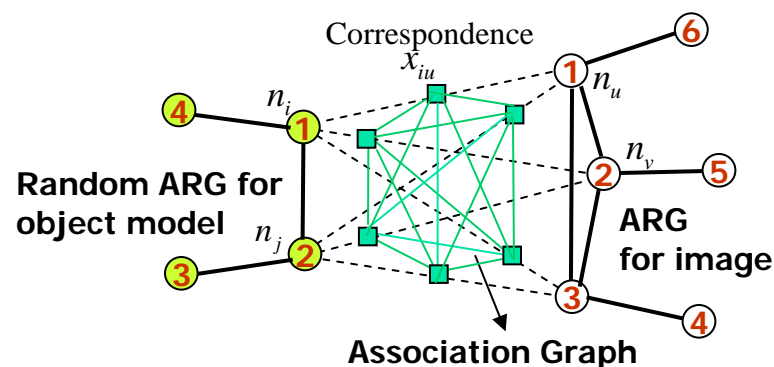
$$\frac{p(O|H=1)}{p(O|H=0)} > \frac{p(H=0)}{p(H=1)} = \lambda \quad , O: \text{input ARG}$$



Object likelihood :

$$P(O | H = 1) = P(X | H = 1)P(O | X, H = 1)$$

**X modeled by Association Graph**



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

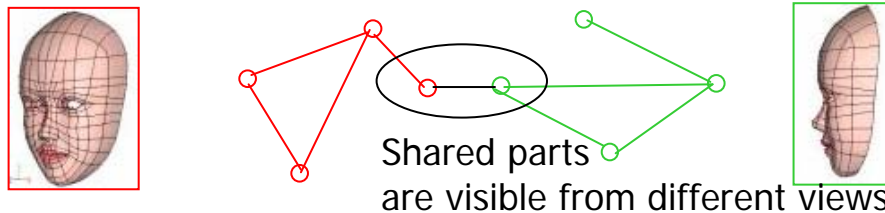


# Extension to Multi-view Object Detection

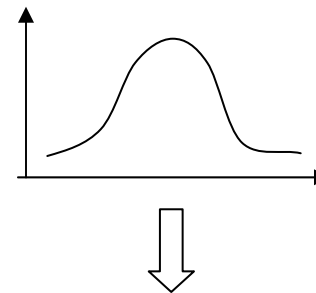
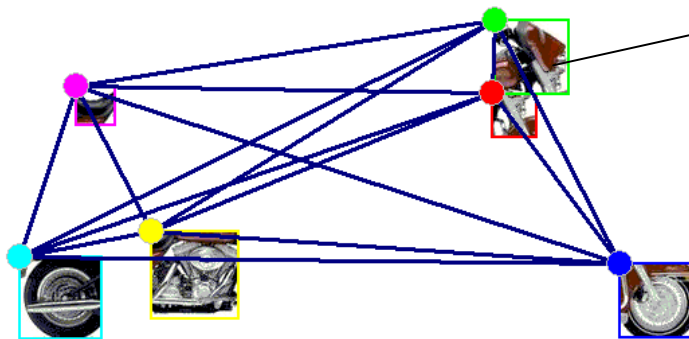
## Challenge of multi-view object/scene detection

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

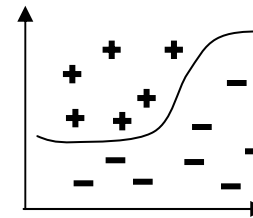
Structure variation could be handled by Random ARG model (each view covered by a sub-graph)



# Adding Discriminative Model for Multi-view Concept Detection

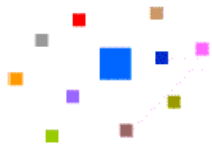


**Previous :**  
Part appearance  
modeling  
by Gaussian  
distribution



**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

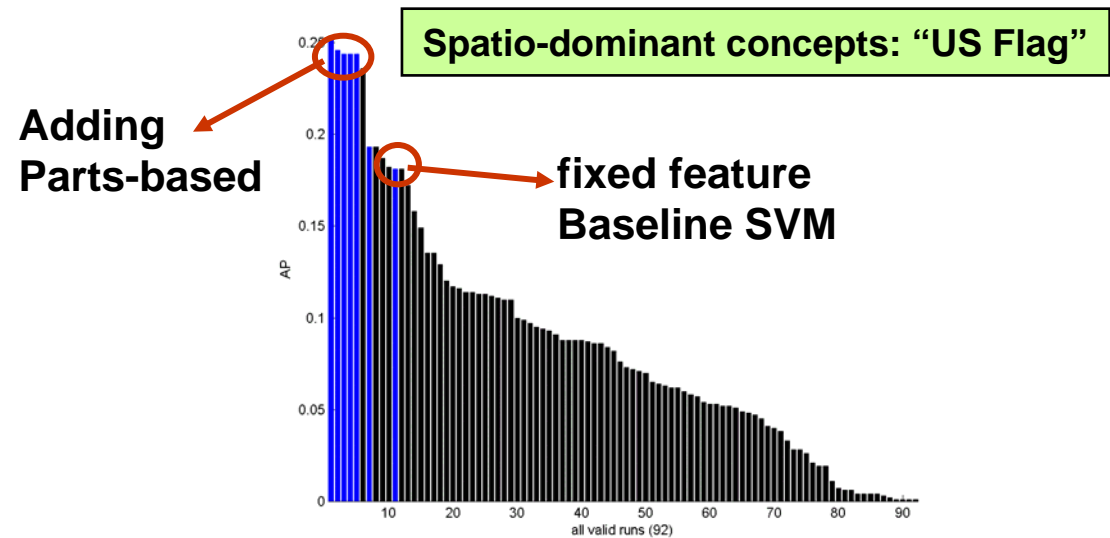
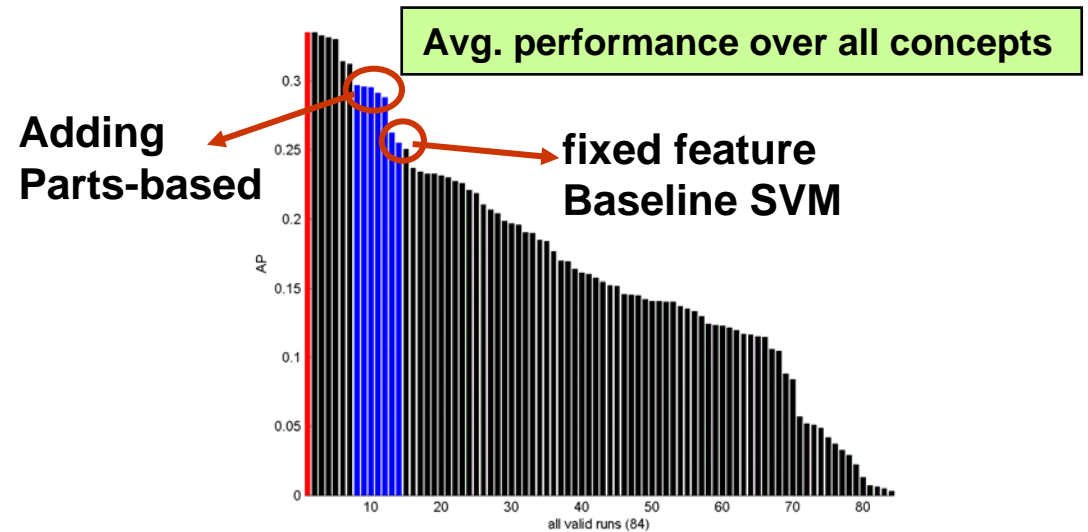


# Evaluation in TRECVID 2005

---

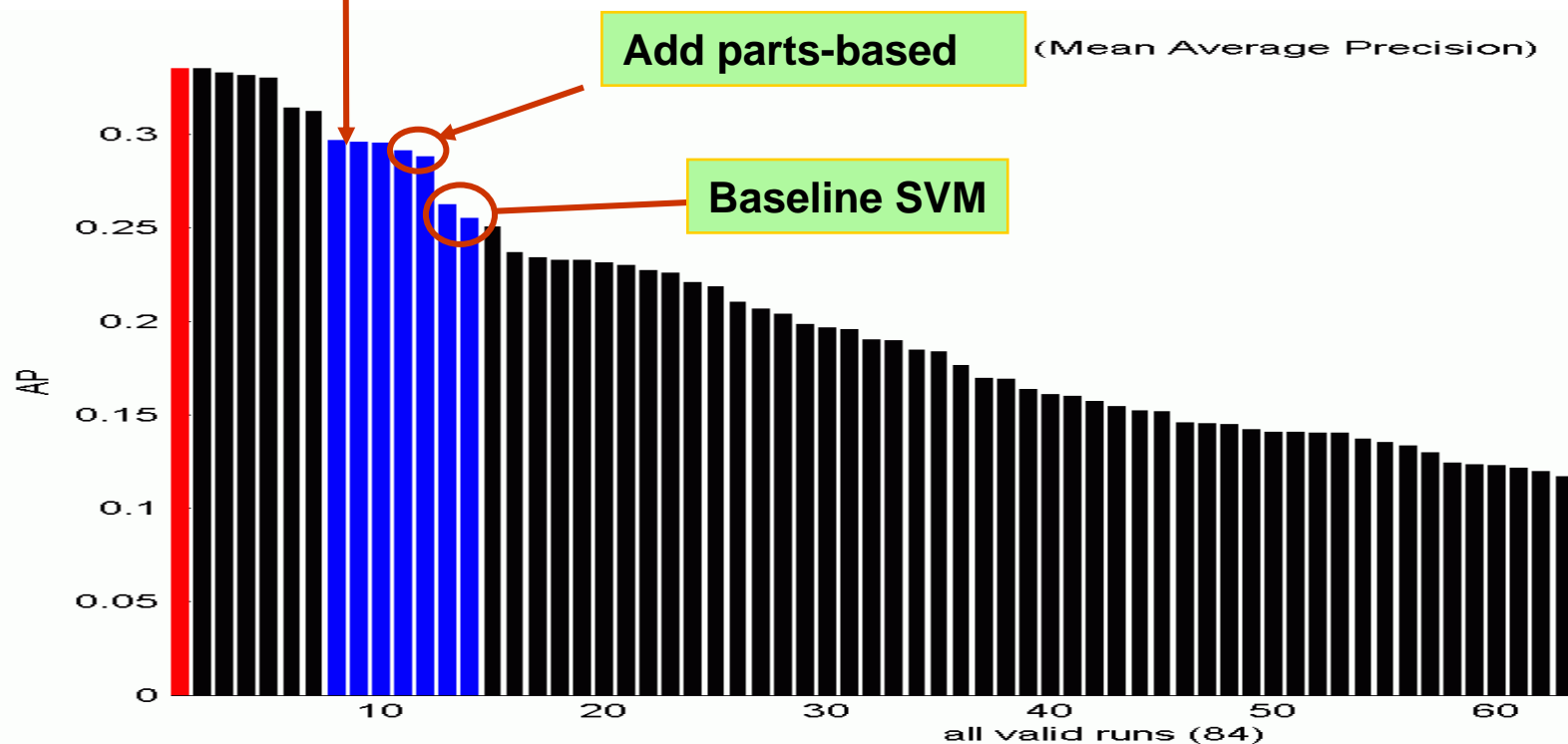
## Parts-based detector performance in TRECVID 2005

- 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.

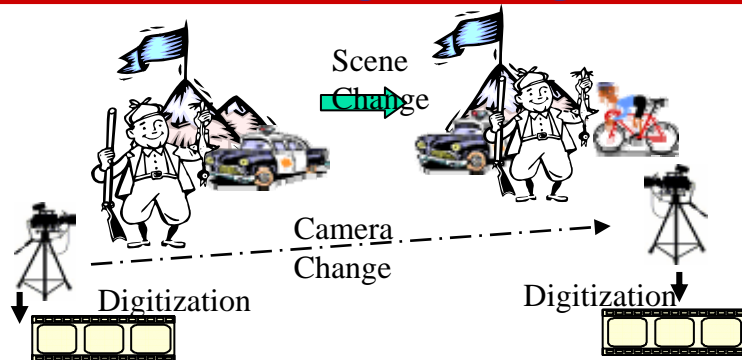


# Relative contributions

- Add text or change fusion models does not help



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



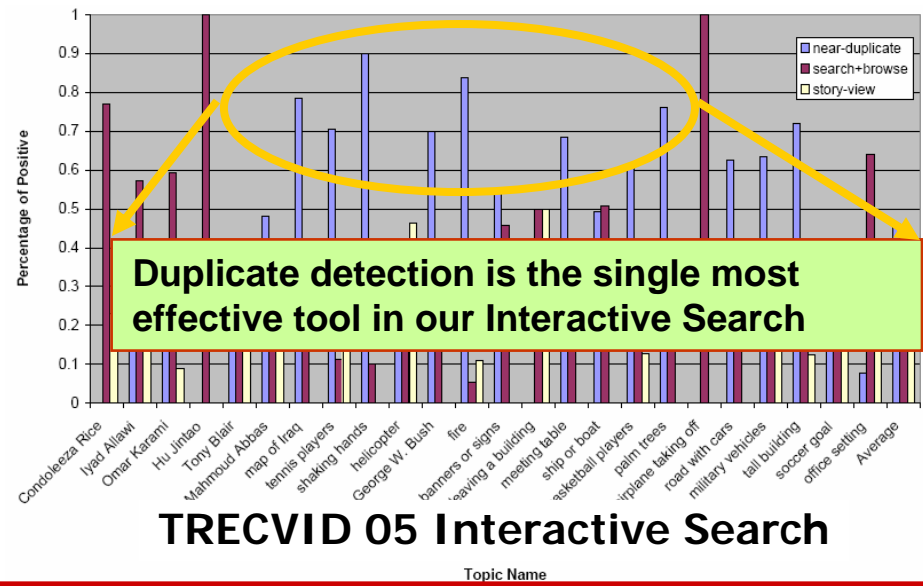
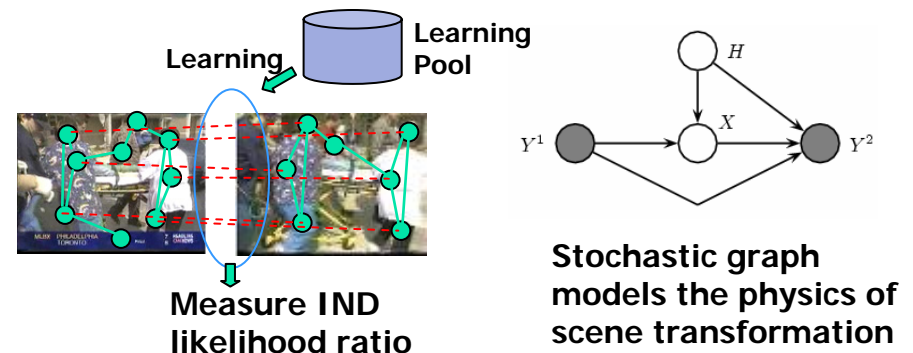
Many Near-Duplicates in TRECVD 05



- 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





# Near Duplicate Benchmark Set

(available for download at Columbia Web Site)

			
19980202_ABC_013352 19980213_CNN_037692	19980202_ABC_013965 19980416_CNN_037364	19980205_CNN_013805 19980214_CNN_010550	19980206_ABC_043621 19980326_ABC_019566
			
19980207_ABC_021562 19980227_CNN_005290	19980207_ABC_022138 19980209_ABC_018538	19980209_CNN_008858 19980214_CNN_010325	19980209_CNN_016138 19980211_CNN_017866
			
19980209_CNN_016234 19980211_CNN_017962	19980211_ABC_009082 19980311_ABC_012250	19980211_ABC_009178 19980311_ABC_012346	19980211_ABC_009202 19980311_ABC_012370
			
19980211_ABC_009610 19980311_ABC_011818	19980212_CNN_001243 19980212_CNN_004003	19980212_CNN_008952 19980213_CNN_013188	19980213_CNN_013553 19980216_ABC_018308

# Examples of Near Duplicate Search in TRECVID 05

Displaying results 1 - 40 of 1000 from 1000 documents. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25  
[\[Return to Grid/Story Browse\]](#) [\[Only Input\]](#) [\[Only Positive\]](#) [Next >>](#)

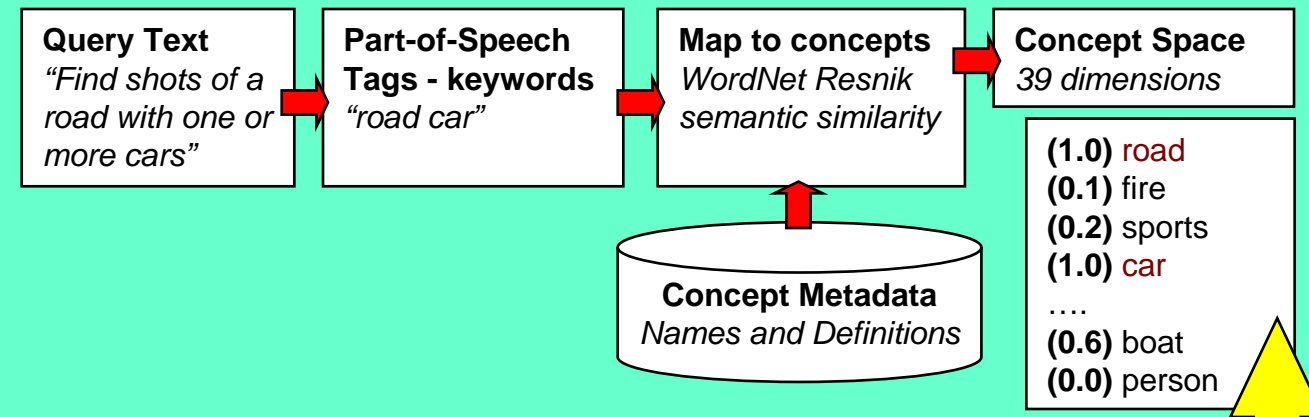
The main grid displays 15 video shot thumbnails arranged in 3 rows and 5 columns. Each thumbnail shows a still from a video with a timestamp in brackets and a shot ID below it. The thumbnails are as follows:

- Row 1:
  - [15.0000] shot14\_16\_RKF
  - [14.0000] shot6\_81\_RKF
  - [13.0000] shot6\_82\_RKF
  - [12.0000] shot6\_85\_RKF
  - [12.0000] shot19\_101\_RKF
- Row 2:
  - [12.0000] shot19\_150\_RKF
  - [12.0000] shot2\_1\_RKF
  - [10.0000] shot6\_84\_RKF
  - [10.0000] shot5\_20\_RKF
  - [10.0000] shot2\_7\_RKF
- Row 3:
  - [9.0000] shot8\_90\_RKF
  - [8.0000] shot2\_210\_RKF
  - [8.0000] shot5\_15\_RKF
  - [8.0000] shot5\_10\_RKF
  - [7.0000] shot9\_125\_RKF

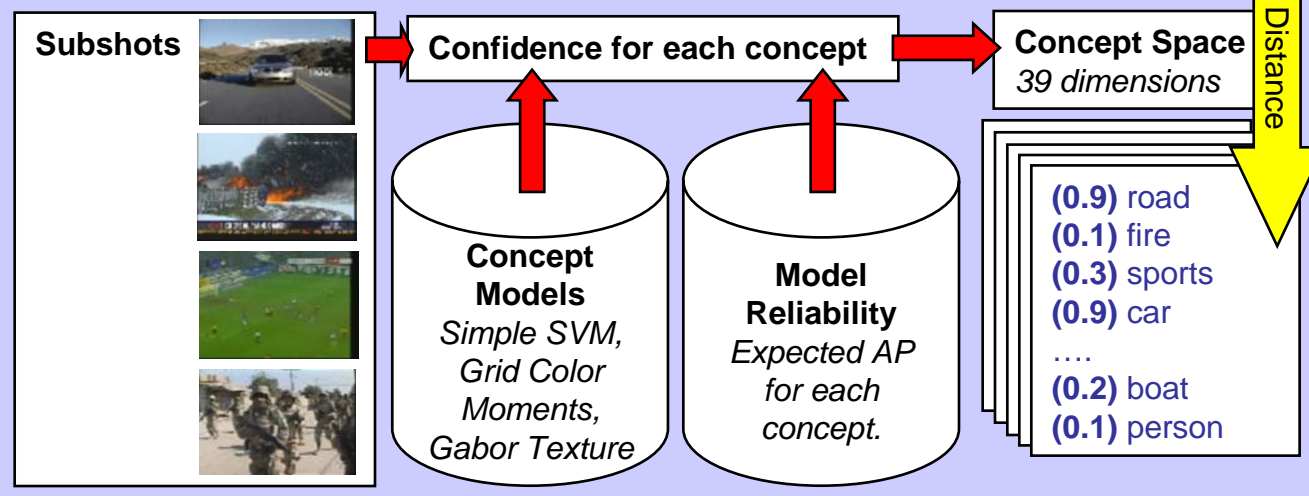
On the right side, the 'Saved Shots' sidebar contains a grid of 8 thumbnails, each with a magnifying glass icon and a red minus sign. Navigation controls for the sidebar include '<< Previous', '1', '2', '3', '4', '5', 'Next >>'.

# Application: Concept Search

## Query



## Documents



Euclidean Distance

- 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.”

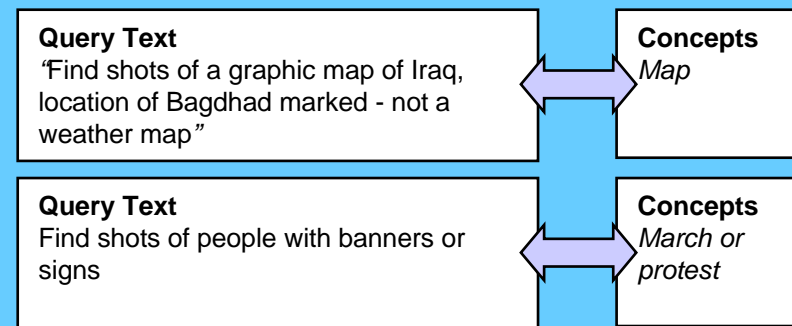
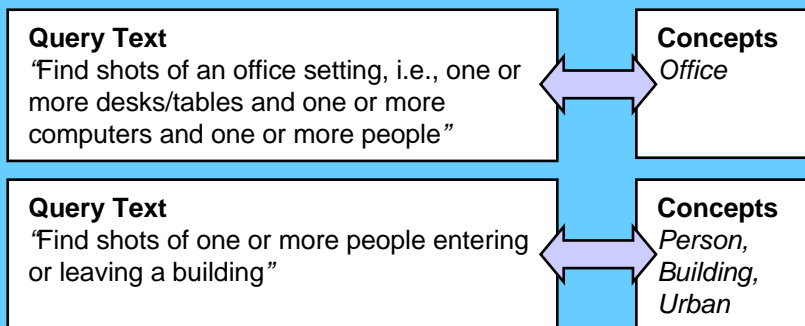
Method	AP
Story Text	.169
CBIR	.002
Concept	.115
Fused	<b>.195</b>

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

Method	AP
Story Text	.053
CBIR	.009
Concept	.090
Fused	<b>.095</b>

## Manual / Interactive

**Manual keyword selection allows more relationships to be found.**





# Columbia Video Search Engine System Overview

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

## User Level Search Objects

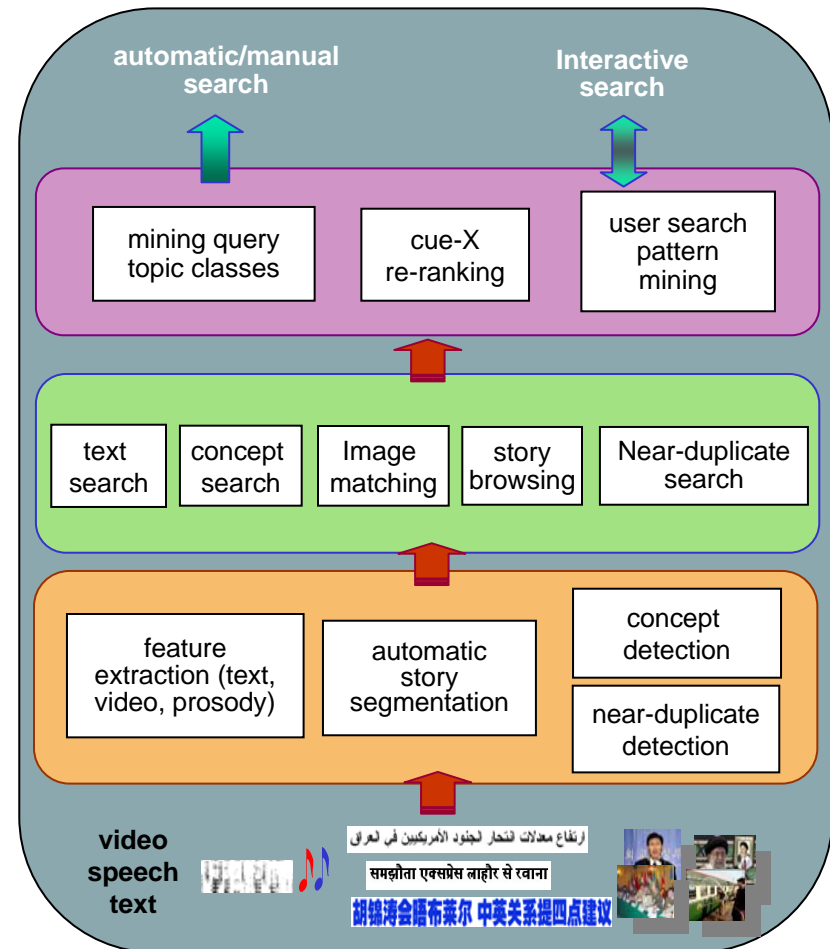
- Query topic class mining
- Cue-X reranking
- Interactive activity log

## 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

**Query Input** [Reset] Search

abbas arafat palestinian january mahm

☐ Pseudo Rel.  
☐ Google  
☐ Exclude Neagive  
☐ Exclude Positive

☐ WordNet  
☒ Exclude Anchor  
☐ Sound Similarity

**CueX Reraking** Start

**Fusion Weighting** Off 25 50 75 Full

Concept ☐ Image ☐ Full-text ☐

**Suggestions**

**Google:** edit election gaza fatah yasser united states see abbas' september said people palestinians holocaust groups west term security president power

**WordNet:** yasser instance arab arabian jan gregorian state yisrael authorization authorisation quantity number palestine liberation curate religion

**NLP Keywords:** abbas arafat january mahmoud israel authority plo minister

**Original Query:** abbas arafat palestinian january mahmoud israel authority prime plo minister

Execution Time: 0.591857s  
Started: 09/19 12:21:15 pm

[XML] [Browse]  
[Logout eric]

**Query Images**

1



Displaying results 1 - 10 of 68 from 68 documents.  
[All By Time] | [All Duplicate Shots] | [Grid Browse]

1 2 3 4 5 6 7 Next >>

**LBCNAHAR, LBC (2004-11-28 14:00:01) (23 of 25 subshots)**

Good tribute of neutrality in the One that both the Israeli **Prime Minister** Ariel Sharon and the President of the Palestine Liberation Organization **Mahmoud Abbas**, a candidate to succeed their willingness to hold a meeting between Arafat, to try to coordinate an Israeli withdrawal from the Gaza Strip in the meantime, held the ABWMAZN [بومازن] spiritual head of the **Palestinian Authority** advances and **prime minister** Ahmed Korei talks with Egyptian President Hosni Mubarak in Cairo He said ABWMAZN [بومازن] it addressed the issues of security and national unity and the presidential elections next ABWMAZN [بومازن] stressed that the presidential election ... (more)

 [1]  
 [2]  
 [3]  
 [4]  
 [5]

**DAILY\_NEWS, CCTV4 (2004-11-28 15:00:00) (8 of 10 subshots)**

According to the US Newsweek magazine Reports Nos. 28 **PLO** Executive Committee Chairman **Abbas** and Israeli **Prime Minister** Ariel Sharon separately in an interview with the magazine said they are prepared to meet with each other Sharon to the reports in an interview that he is ready and **Abbas** and is willing to meet with **Palestinian** new government in Israel's withdrawal from the Gaza Strip Strategic Plan pass unimpeded harm that **Israel** will take necessary measures to the **Palestinian** without interference to the general elections **Abbas** in an interview with the Palestinians her at the junction of the Provisional yen ... (more)

 [14]  
 [222]  
 [239]  
 [664]  
 [760]

**LBCNEWS, LBC (2004-11-28 20:00:00) (25 of 26 subshots)**

 [760]  
 [760]  
 [760]

**Saved Shots**

1 2 Next >>

 [1]  
 [2]  
 [3]  
 [4]  
 [5]  
 [6]  
 [7]  
 [8]  
 [9]  
 [10]  
 [11]  
 [12]





# 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