

Evaluating the Impact of 374 Visual-based LSCOM Concept Detectors on Automatic Search

Shih-Fu Chang, Winston Hsu, Wei Jiang, **Lyndon Kennedy**, Dong Xu, Akira Yanagawa, and Eric Zavesky

Digital Video and Multimedia Lab
Columbia University

NIST TRECVID Workshop
November 14, 2006

Video / Image Search

- **Objective:** Semantic access to visual content

- **Stop-gap solutions**

- **Text Search**

- Not always useful

- Text not available in all situations

- **Query-by-Example**

- Lacking in semantic meaning

- Example images not readily available

- **Concept Search:** exciting new direction

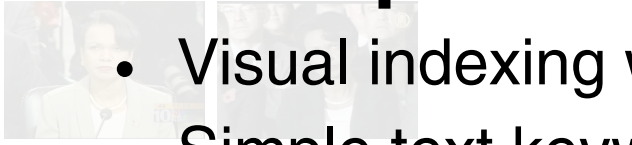
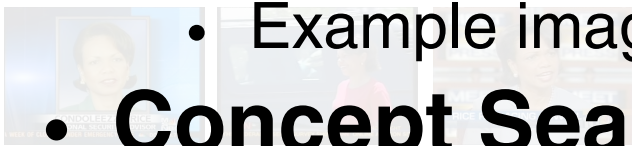
- Visual indexing with concept detection: high semantic meaning
 - Simple text keyword search

Multimodal Queries

Find shots of tennis players on the court from news videos at same time.



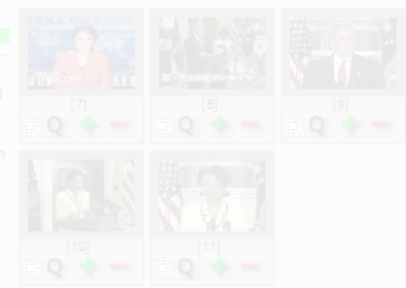
Find shots of Condoleezza Rice.



Video / Image Index

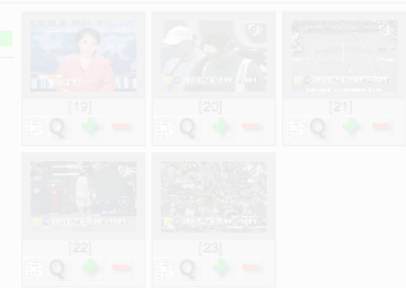
NTDNEWS12, NTDTV (2004-11-17 (2:00:00)) (12 of 13 subshots)

President Bush Tuesday nomination National Security Adviser Condoleezza Rice served as secretary of state from their posts does not mean that from deep and long-lasting firmly believe that freedom of value and charm but President Bush this appointment has to go through approved by the Senate fuji elements nominated by President Tuesday National Security Adviser Condoleezza Rice, president of received in order to life we sculpture itself is responsible for training 4 is Hum is here rise Bush again at the White House announced for anti-corruption war and the Israeli-Palestinian conflict challenges of the foreign policy is ... (more)



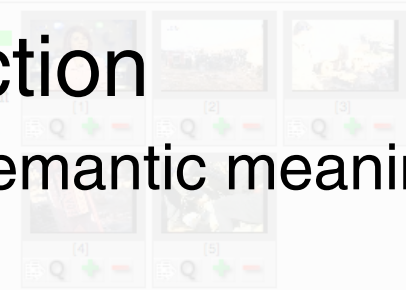
NTDNEWS12, NTDTV (2004-11-17 (2:00:00)) (9 of 21 subshots)

total prize of high 3.7 million of US dollars Tennis Masters Cup ATT finals in the United States Tuesday Houston City formally began its world ranking first in Switzerland players singing in antiphonal style flag to 2 successful than zero light Syrian French tennis champion Argentina were high-profile in the first match fee dreux Swiss players to more than 6 defeated 17.6 Gambia, the Argentine people were injured in 6 months to serve the Dishi greet cup finals ... (more)



LBCNAHAR, LBC (2004-11-28 (4:00:07)) (11 of 13 subshots)

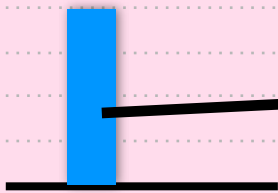
two cars overturned after the Syrian driver insisted, respectively, the Lebanese security sources said among those killed in three Americans, including two children injured Father



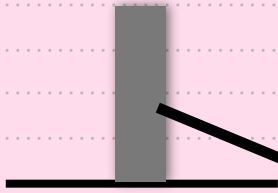
Concept Search Framework

Text Queries

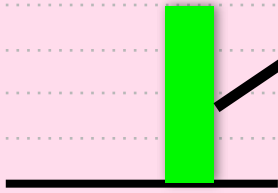
Find shots of snow.



Find shots of soccer matches.



Find shots of buildings.



Concept Detectors

- Anchor
- Snow
- Soccer
- Building
- Outdoor

Image Database



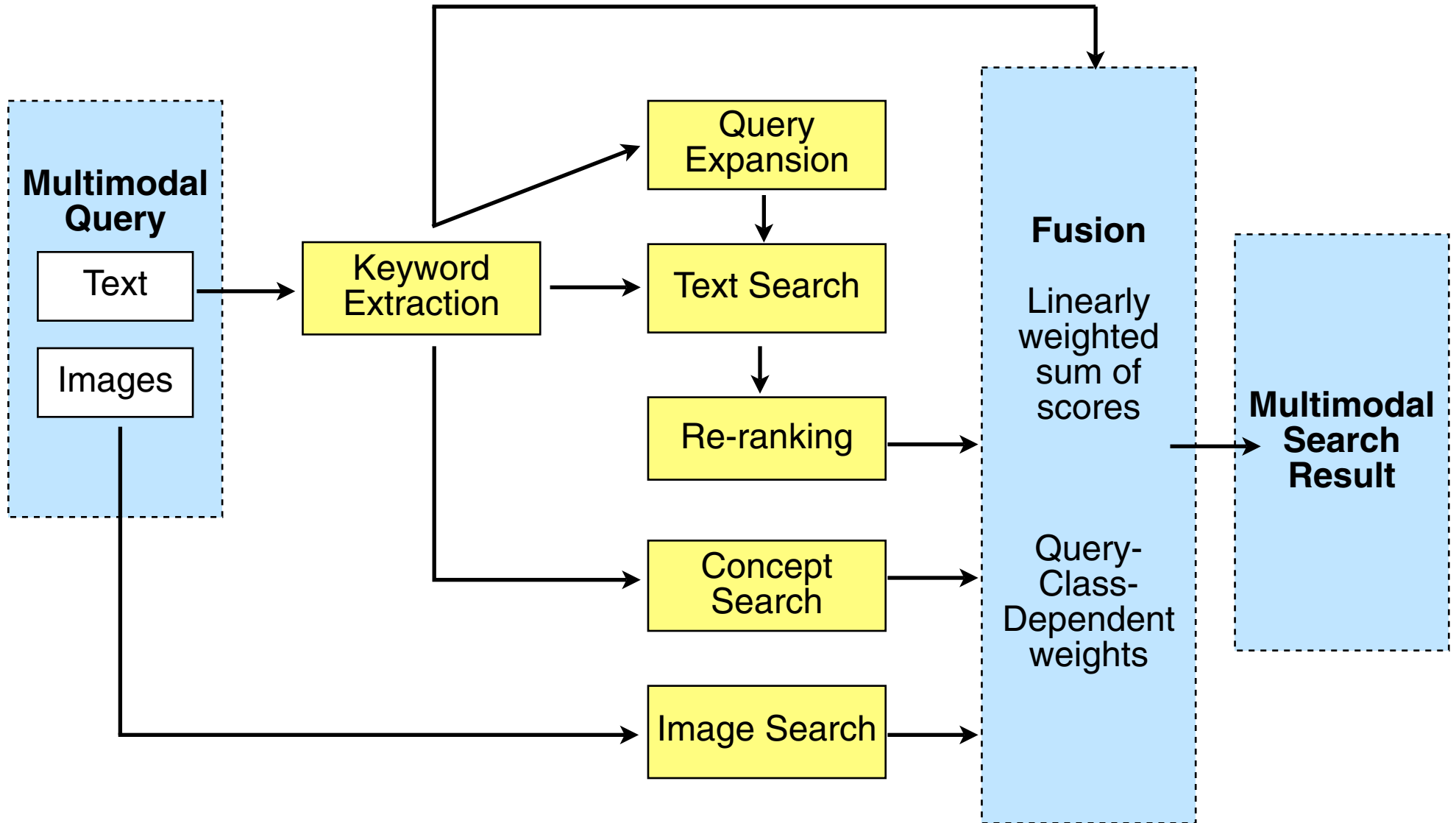
Concept Search

- Text-based queries against visual content
- Index video shots using visual content only
 - run many concept detectors over images
 - treat scores as likelihood of containing concept
- Allow queries using text keywords (no examples)
 - map keywords to concepts
 - use fixed list of synonyms for each concept
- Many concepts available
 - LSCOM 449 / MediaMill 101
 - TRECVID 2006: first opportunity to put large lexicons to the test

Concept Search: TRECVID Perspective

- Concept search is powerful and attractive
 - But unable to handle every type of query
- Text and Query-by-Example still very powerful
 - Want to exploit any/all query/index information available
- Impact of methods varies from query to query
 - Text: named persons
 - Query-by-Example: consistent low-level appearance
 - Concept: existence of matching concept
- Propose: *Query-Class-Dependent* model

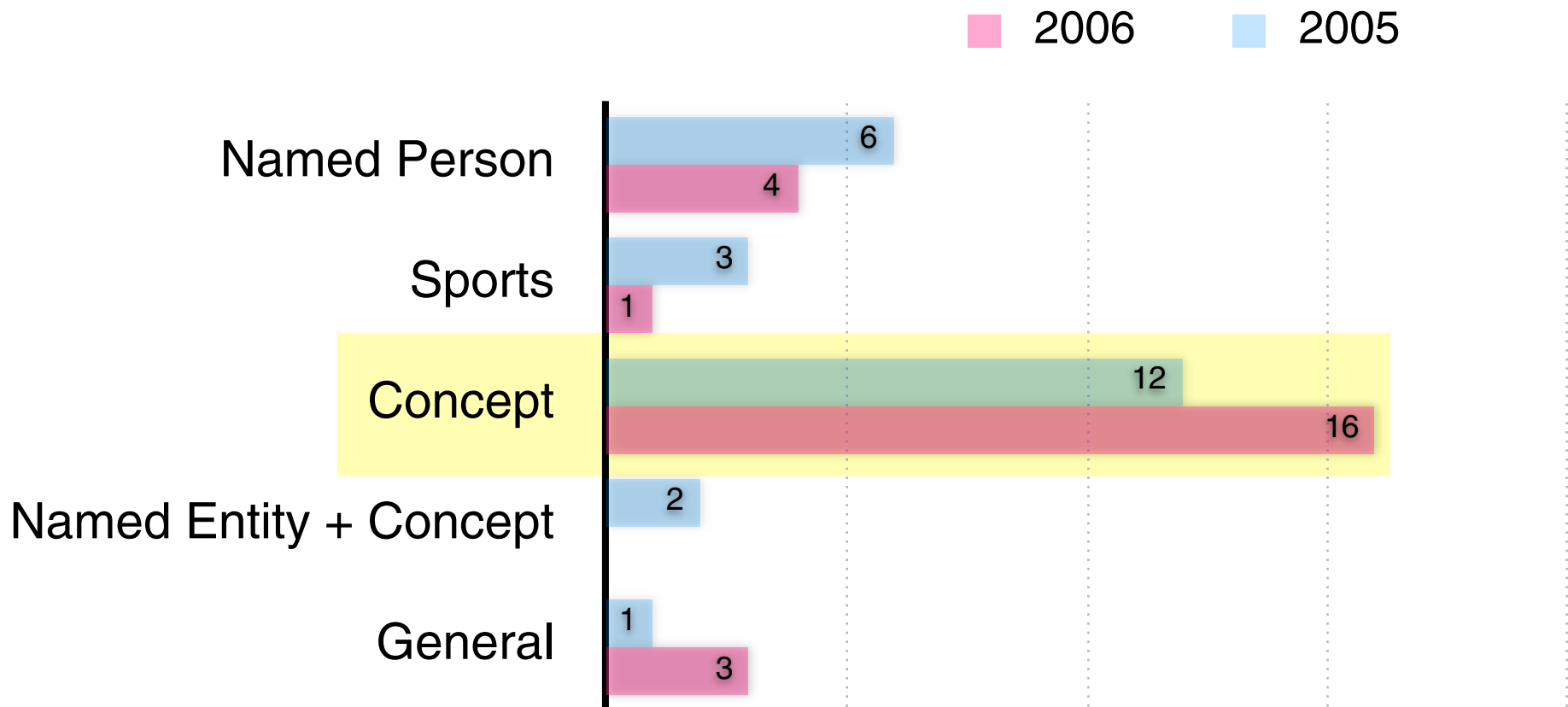
Query-Class-Dependent Search



Query Classes

- **Named Person**
 - if named entity detected in query. Rely on text search
- **Sports**
 - if sports keyword detected in query. Rely on visual examples.
- **Concept**
 - if keyword maps to pre-trained concept detector. Rely on concept search.
- **Named Person + Concept**
 - if both named entity and concept detected. Combine text and concept search equally.
- **General**
 - for all other queries. Combine text and visual examples equally.

Query Class Distribution



Query Processing / Classification

Incoming Query Topic

Natural language
statement

.....
Find shots of
Condoleezza Rice.

.....
Find shots of scenes
with snow.

.....
Find shots of one or more
soccer goalposts.

Keyword Extraction

Part-of-Speech Tagging
Named Entity Detection

.....
condoleezza rice

.....
scenes snow

.....
soccer goalposts

Query Classification

Named Entity?
Matching concept?
Sports word?

.....
Named Entity

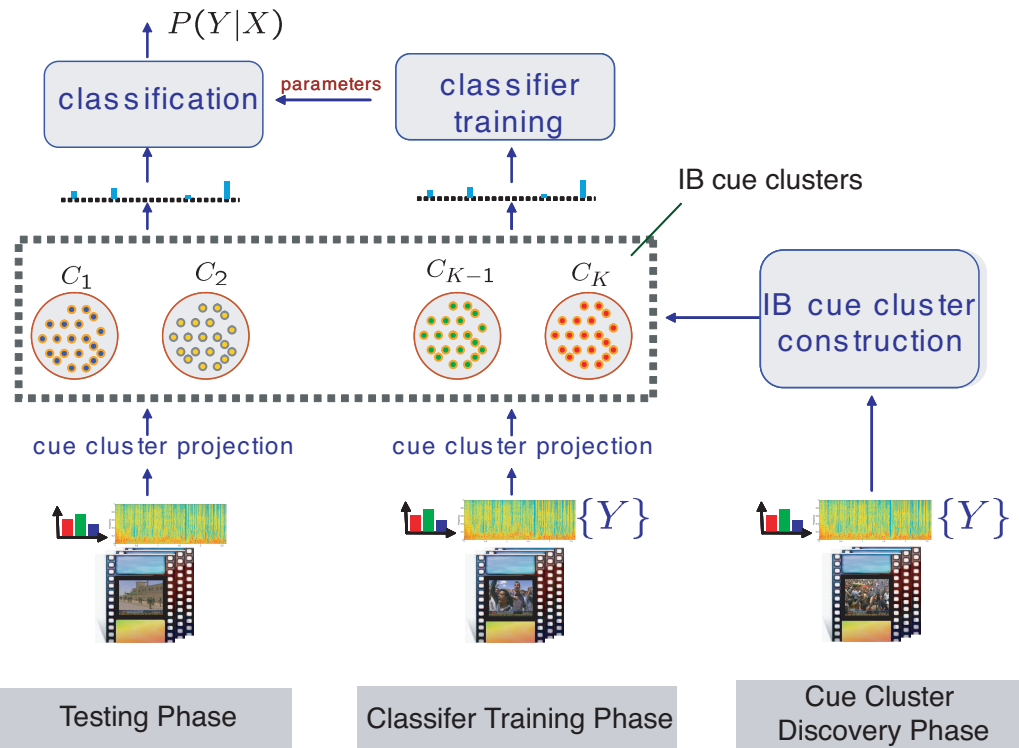
.....
Matching Concept

.....
Sports

Text Search

- Extract named entities or nouns as keywords
- Keyword-based search over ASR/MT transcripts
 - use story segmentation
- Most powerful individual tool
 - Named persons: “Dick Cheney,” “Condoleezza Rice,” “Hussein”
 - Others: “demonstration/protest,” “snow,” “soccer”

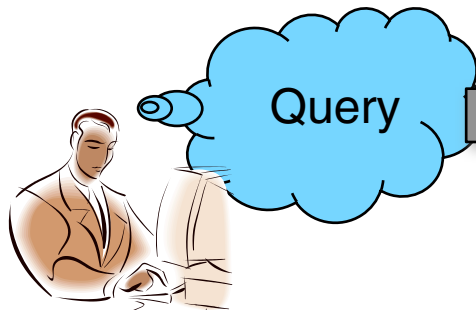
Story Segmentation



[Hsu, CIVR 2005]

- Automatically detect story boundaries.
- low-level features: color moments, gabor texture
- IB framework: discover meaningful mid-level feature clusters
- high-performing in TV2004
- results shared with community in 2005 + 2006
- **Stories for text search**
 - typically 25% improvement
 - TV2006: 10% improvement

Named Entity Query Expansion



Query

Query Processing

Stemmed: shots tennis court players

Search

Internal Text

Single source: target documents

Find shots of a tennis court with both players visible.

Method: Detected named entities from internal and external sources. Secondary search with discovered entities in external text

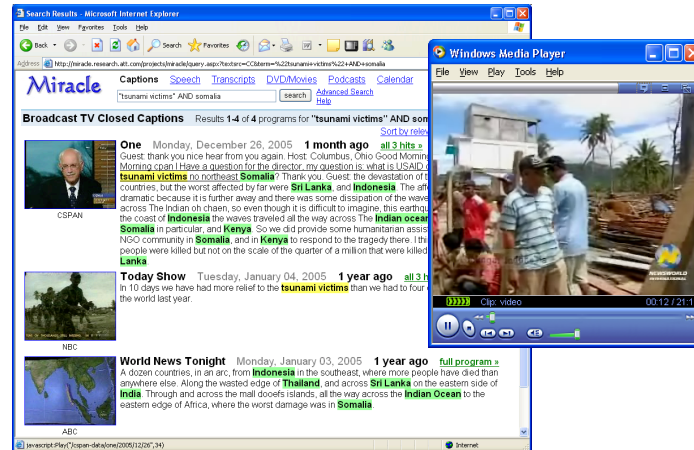
Named Entities

External Text

Story

Stories: ASR / CC

Internal Text



AT&T Labs: Miracle multimedia platform. *

Search with both keywords and entities

Result

Joint work with AT&T Labs: Liu, Gibbon, Zavesky, Shahraray, Haffner, TV2006

Information Bottleneck Reranking



(a)

(a) Search topic -
“Find shots of Tony Blair”
& search examples



(b) Text search results

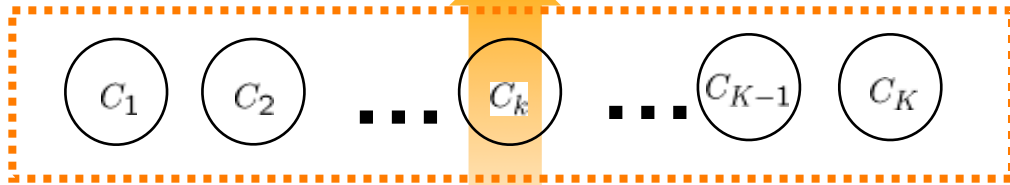
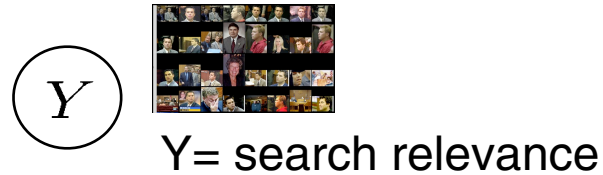


(c) IB reranked results + text search

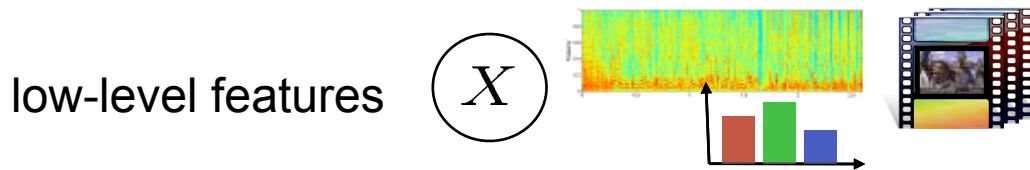
Information Bottleneck Reranking

Information Bottleneck principle

$$C^* = \underset{C|R}{\operatorname{argmin}} \{I(X; Y) - I(C; Y)\}$$



Clusters automatically discovered via Information Bottleneck principle & Kernel Density Estimation



- Re-order text results
 - make visually similar clusters
 - preserve mutual information with estimated search relevance
- Improve 10% over text alone
 - lower than past years
 - text baseline is (too) low

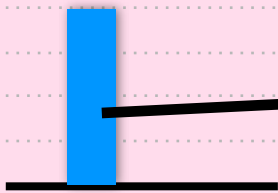
Visual Example Search

- **Fusion of many image matching and SVM-based searches**
[IBM, TV2005]
- **Feature spaces:**
 - 5x5 grid color moments, gabor texture, edge direction histogram
- **Image matching:**
 - euclidean distance between examples and search set in each dimension
- **SVM-based:**
 - Take examples as positives (~5), randomly sample 50 negatives
 - Learn SVM, repeat 10 times, average resulting scores
 - Independent in each feature space
- **Average** scores from 3 image matching and 3 SVM-based models
- **Least-powerful method**
 - best for “soccer”

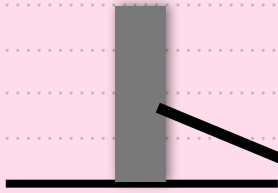
Reminder: Concept Search Framework

Text Queries

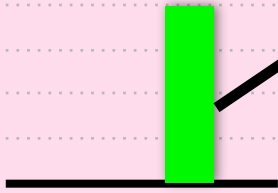
Find shots of snow.



Find shots of soccer matches.



Find shots of buildings.



Concept Detectors

- Anchor
- Snow
- Soccer
- Building
- Outdoor

Image Database



Concept Ontologies

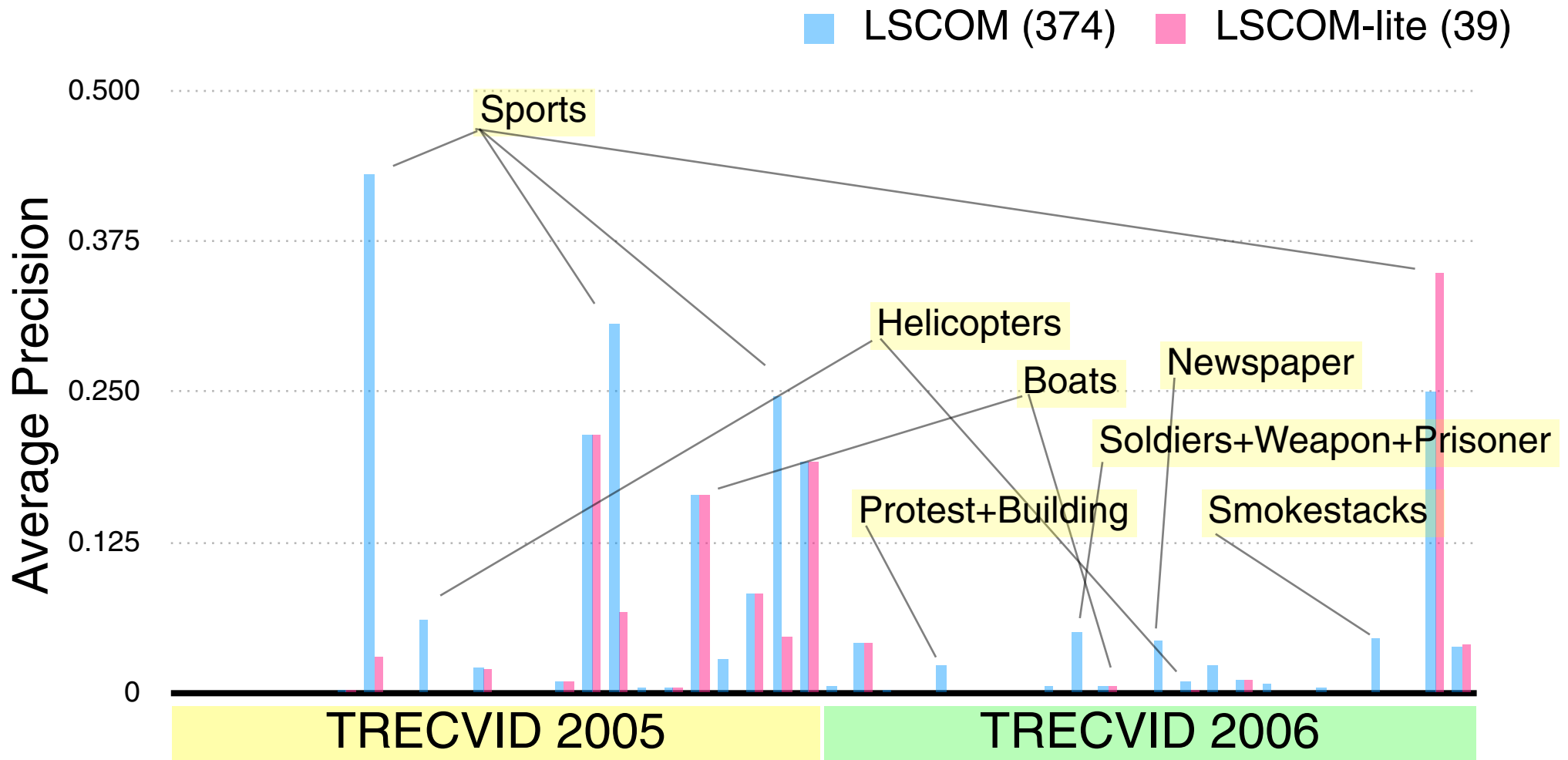
- LSCOM-Lite
 - 39 Concepts (used for TRECVID 2006 High-level Features)
- LSCOM
 - 449 Concepts
 - Labeled over TRECVID 2005 development set
 - 30+ Annotators at CMU and Columbia
 - 33 million judgments collected
 - Free to download (110+ downloads so far)
 - <http://www.ee.columbia.edu/dvmm/lscom/>
 - revisions for “event/activity” (motion) concepts coming soon!

Lexicon Size Impact

- 10-fold increase in number of concepts
Possible effects on search?
- Depends on:
 - How many queries have matching concepts?
 - How frequent are the concepts?
 - How good are the detection results?

Concept Search Performance

Increasing Size of Lexicon



Increasing Lexicon Size

	39 Concepts	374 Concepts
TV 2005	MAP: 0.0353	MAP: 0.0743
TV 2006	MAP: 0.0191	MAP: 0.0244

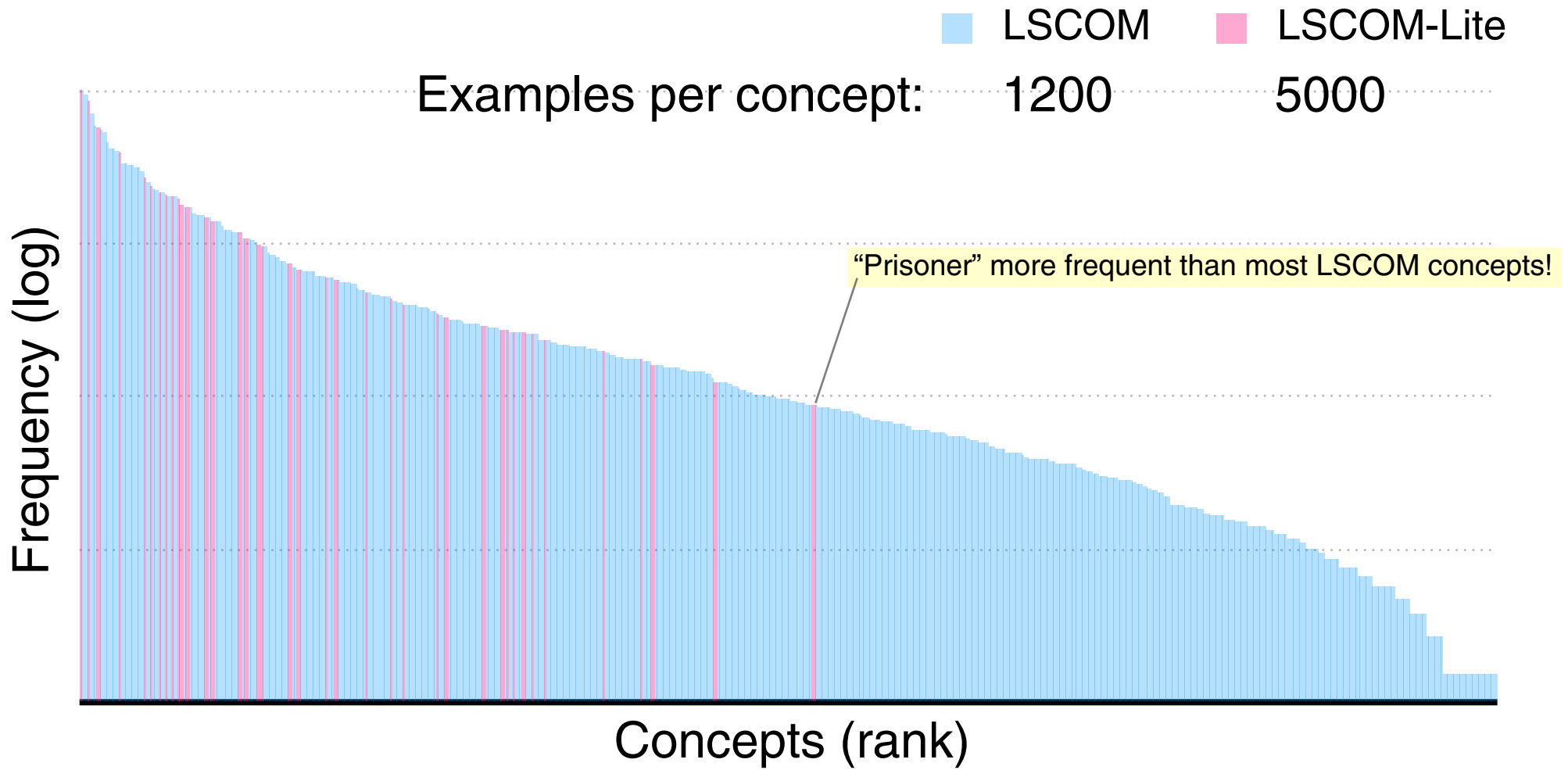
- Large increase in number of concepts,
Moderate increase in search performance
 - **10x** as many concepts in lexicon
 - Search MAP increases by **30% - 100%**

Concept / Query Coverage

	39 Concepts	374 Concepts
TV 2005	11 Query Matches 1.1 Concepts/Query	17 Query Matches 1.3 Concepts/Query
TV 2006	12 Query Matches 1.8 Concepts/Query	17 Query Matches 2.5 Concepts/Query

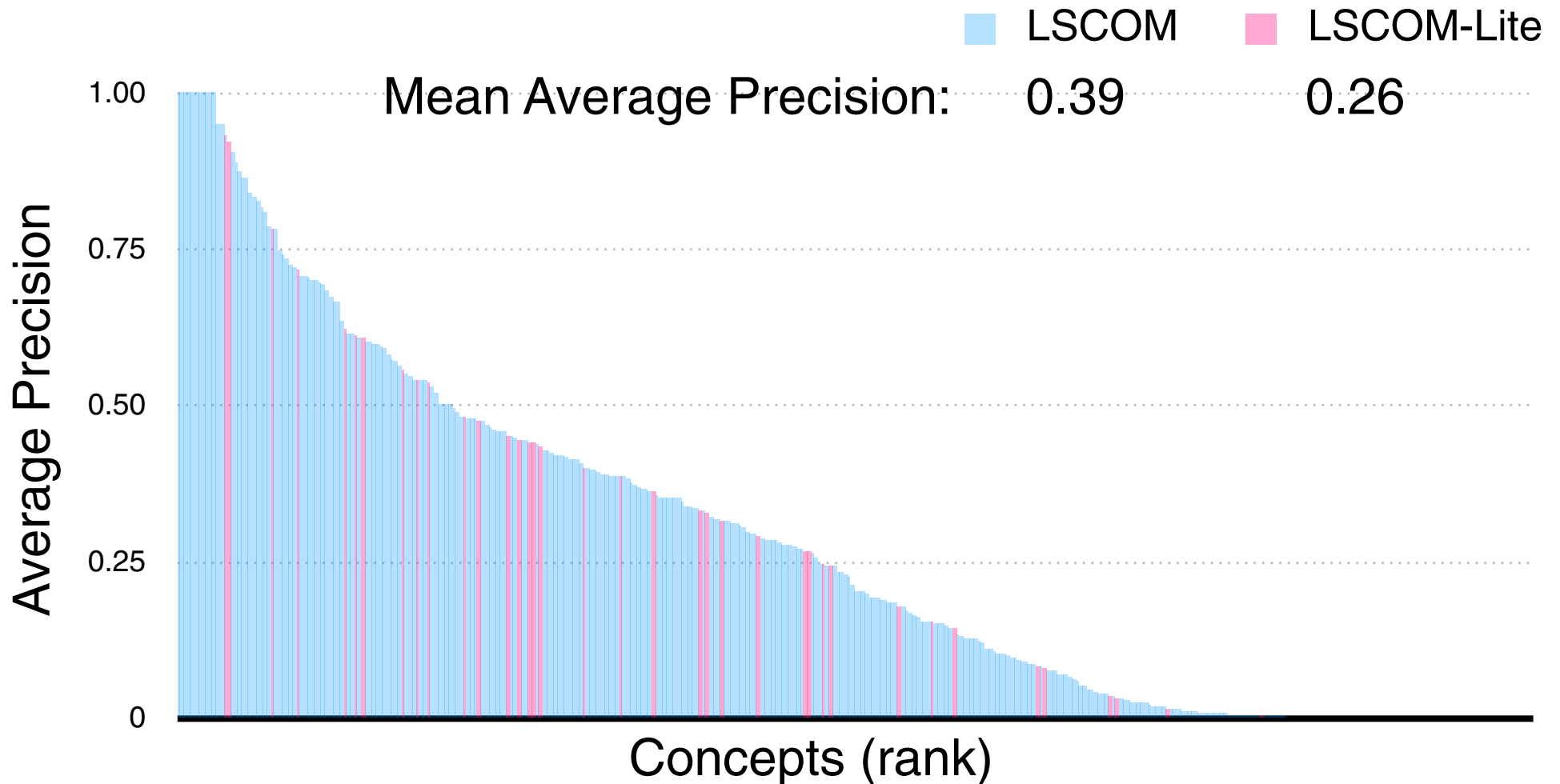
- Large increase in number of concepts,
Small increase in coverage
 - **10x** as many concepts in lexicon
 - **1.5x** as many queries covered
 - **1.2x - 1.4x** as many concepts per covered query

Concept Frequencies



Concept Detection Performance

Internal evaluation: 2005 validation data



Effect of Larger Lexicon

- 10x increase in lexicon size
 - 30% (2006) - 100% (2005) increase in retrieval performance
- Contributing factor:
 - 50% relative increase in query coverage
- Negative factors:
 - concepts in larger lexicon on less frequent (75% decrease)
 - less detectable (33% decrease in MAP)
- Positive effects of matching concepts outweigh problems of detectability

Query Processing

Finds shots with one or more **emergency vehicles** in **motion** (e.g., ambulance, police car, fire truck, etc.)

39: None

374: Emergency_Room, Vehicle

Concept

Find shots with a **view** of one or more tall **buildings** (more than 4 stories) and the top **story** visible.

39: Building

374: Building

Concept

Find shots with one or more people leaving or entering a **vehicle**.

39: None

374: Vehicle

Concept

Find shots with one or more **soldiers**, **police**, or **guards** escorting a **prisoner**.

39: Military_Personnel, Prisoner

374: Guard, Police_Security, Prisoner, Soldier

Concept

Query Processing

Find shots of a daytime **demonstration** or **protest** with at least part of one **building** visible.

39: People_Marching, Building

374: Demonstration_Protest, Building, Protesters

Concept

Find shots of US Vice President **Dick Cheney**.



Named Person

Find shots of **Saddam Hussein** with at least one other person's face at least partially visible.

Named Person

Find shots of multiple people in **uniform** and in **formation**.



General

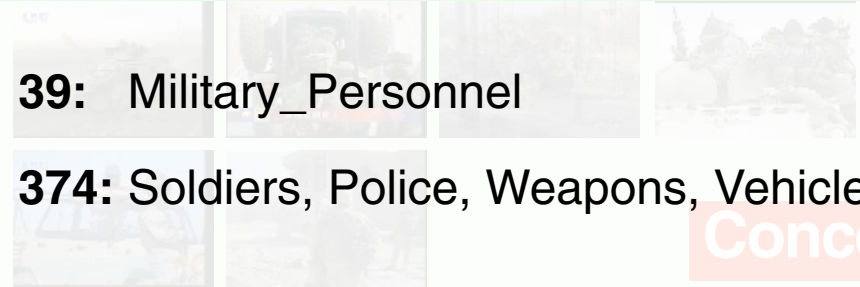
Query Processing

Find shots of US President **George W. Bush, Jr.** walking.



Named Person

Find shots of one or more **soldiers** or **police** with one or more **weapons** and military **vehicles**.

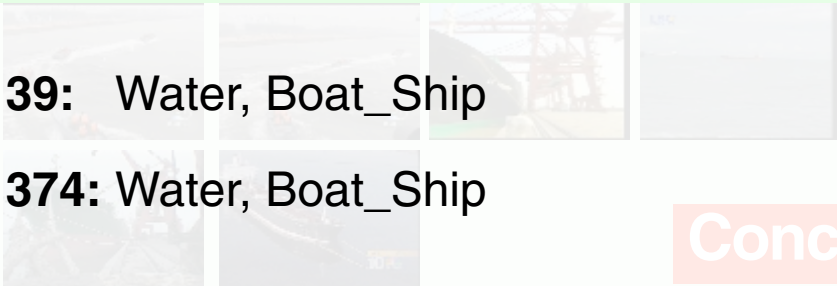


39: Military_Personnel

374: Soldiers, Police, Weapons, Vehicles

Concept

Find shots of **water** with one or more **boats** or **ships**.

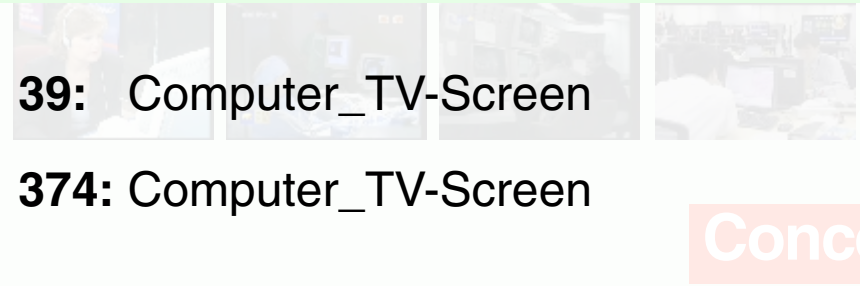


39: Water, Boat_Ship

374: Water, Boat_Ship

Concept

Find shots of one or more people seated at a **computer** with **display** visible.



39: Computer_TV-Screen

374: Computer_TV-Screen

Concept

Query Processing

Find shots of one or more people reading a **newspaper**.

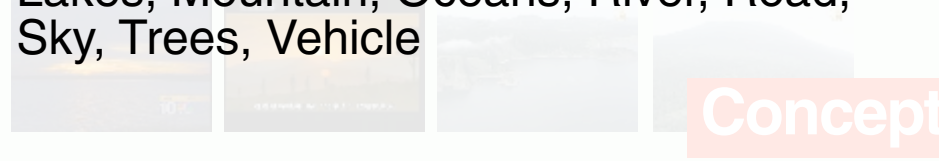


General

Find shots of a natural scene - with, for example, **fields, trees, sky, lake, mountain, water, sun, sunset, buildings, no roads, no vehicles**.

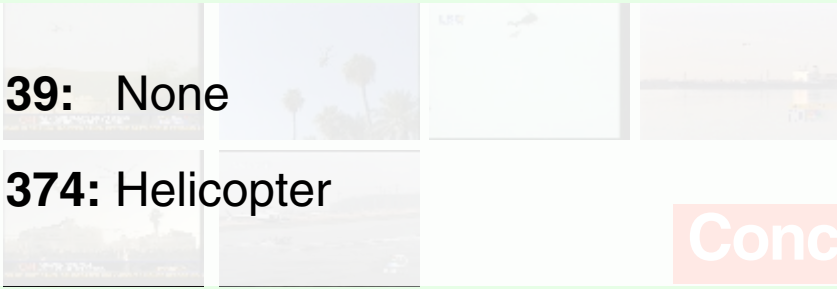
39: Animal, Building, Mountain, Waterscape_Waterfront, Road, Sky

374: Animal, Beach, Building, Fields, Lakes, Mountain, Oceans, River, Road, Sky, Trees, Vehicle



Concept

Find shots of one or more **helicopters** in flight.

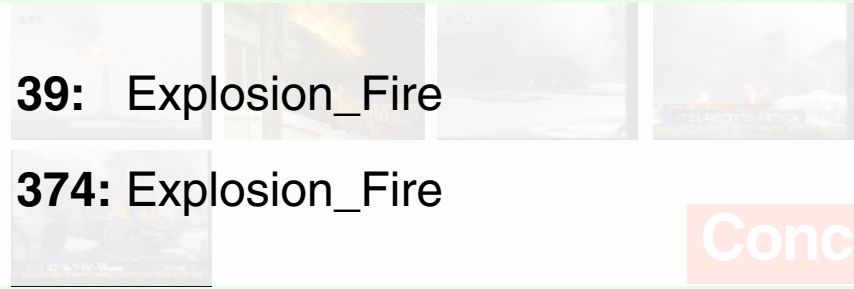


39: None

374: Helicopter

Concept

Find shots of something **burning** with **flames** visible.



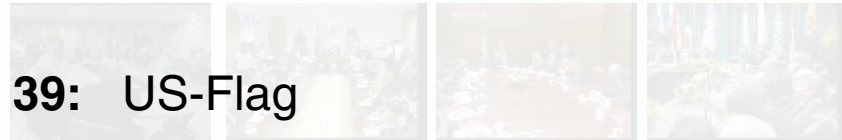
39: Explosion_Fire

374: Explosion_Fire

Concept

Query Processing

Find shots of a group including least four people dressed in **suits**, seated, and with at least one **flag**.



39: US-Flag

374: Flags, Groups, Suits

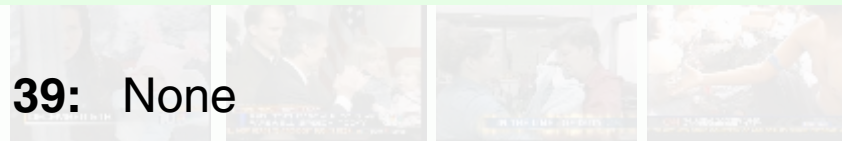
Concept

Find shots of at least one person and at least 10 **books**.



General

Find shots containing at least one **adult** person and at least one **child**.



39: None

374: Adult, Child

Concept

Find shots of a **greeting** by at least one **kiss** on the **cheek**.



39: None

374: Election_Greeting

Concept

Query Processing

Find shots of **smokestacks**, **chimneys**, or **cooling towers** with **smoke** or **vapor** coming out.

39: None

374: Power_Plants, Smoke

Concept

Find shots of **Condoleezza Rice**.



Named Person

Find shots of one or more **soccer** **goalposts**.

39: Sports

374: Soccer

Sports

Find shots of **scenes** of **snow**.

39: Snow

374: Snow

Concept

Example Query: Building

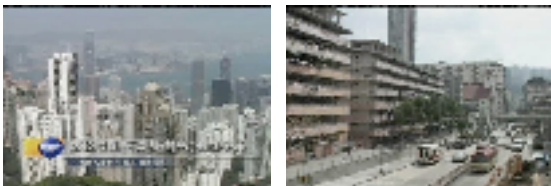
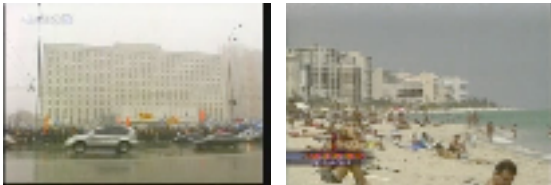
Text:

“Find shots with a **view** of one or more tall **buildings** (at least 5 stories) and the top **story** visible.”

Keywords: view building story

Visual Concepts: building

Visual Examples:



- **Query class:**
 - “concept”
- **Text search: (.05)**
 - “view building story”
- **Concept search: (.85)**
 - “building” concept
- **Visual search: (.10)**
 - 4 examples

Building: Text Search



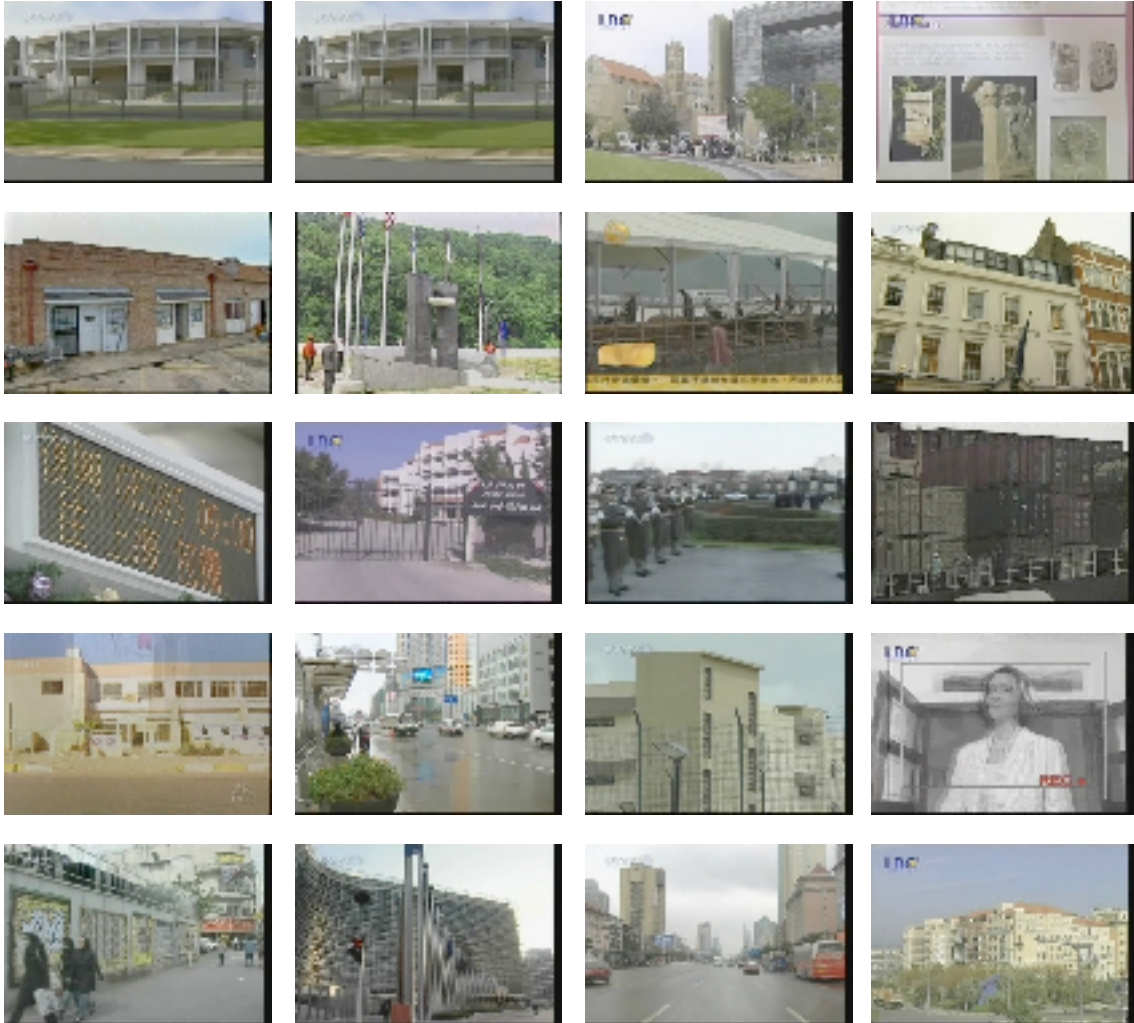
- No meaningful results from text
- No link between transcript and visuals

Building: Information Bottleneck Reranking



- IB reranking ineffective
- Text search ineffective to begin with

Building: Concept Search



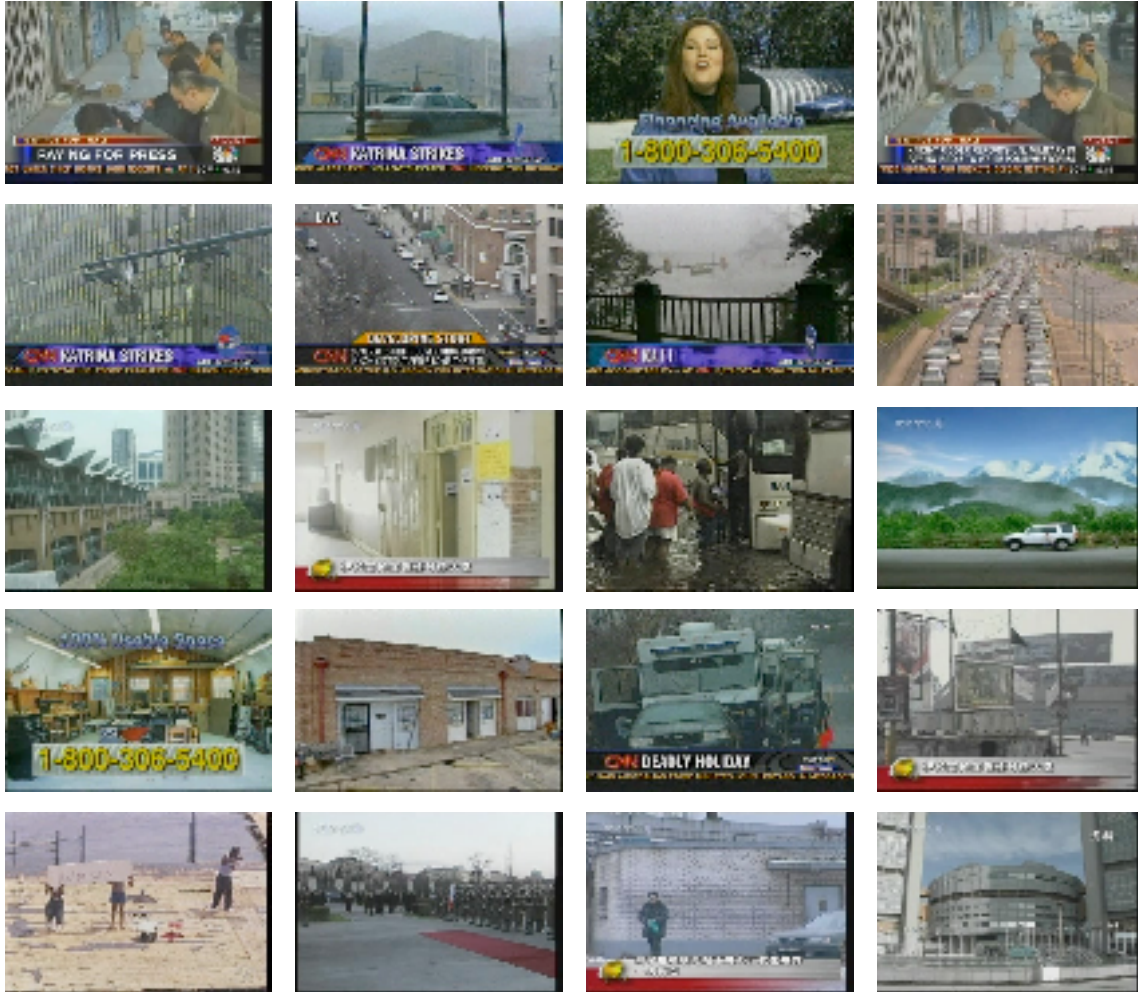
- Map to “building” concept detector
- Strong performance

Building: Visual Example Search



- Strong color and edge cues
- Many false alarms

Building: Fused



- **Concept** - 0.85
 - **Example** - 0.10
 - **Text** - 0.05
-
- little improvement over concept alone
 - no loss, though

Example Query: Snowy Scenes

Text:

“Find shots of **scenes** with **snow**.”

Keywords: scenes snow

Visual Concepts: snow

Visual Examples:



- **Query class:**

- “concept”

- **Text search: (.05)**

- “scenes snow”

- **Concept search: (.85)**

- “snow” concept

- **Visual search: (.10)**

- 7 examples

Snowy Scenes: Text Search

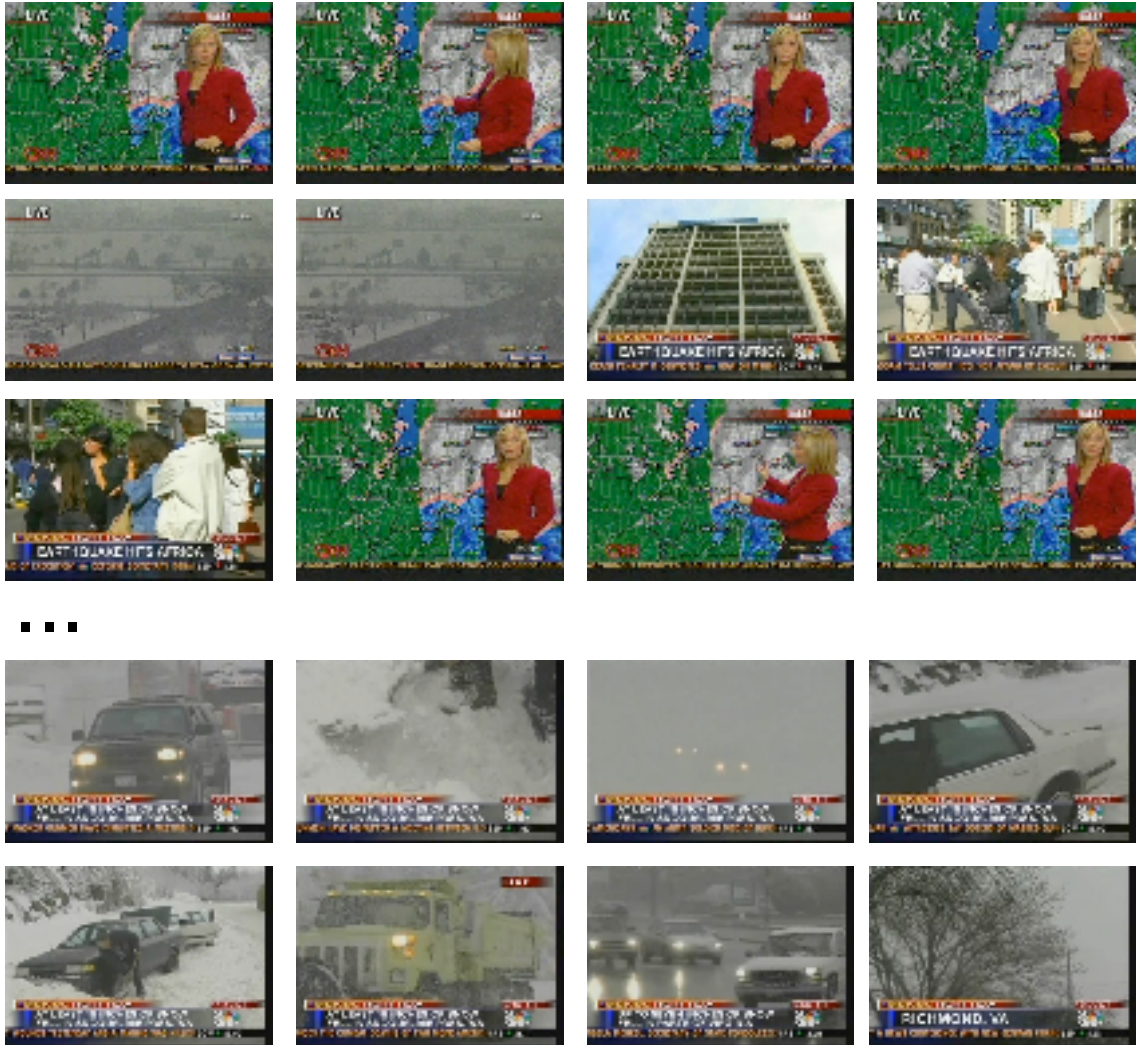


...



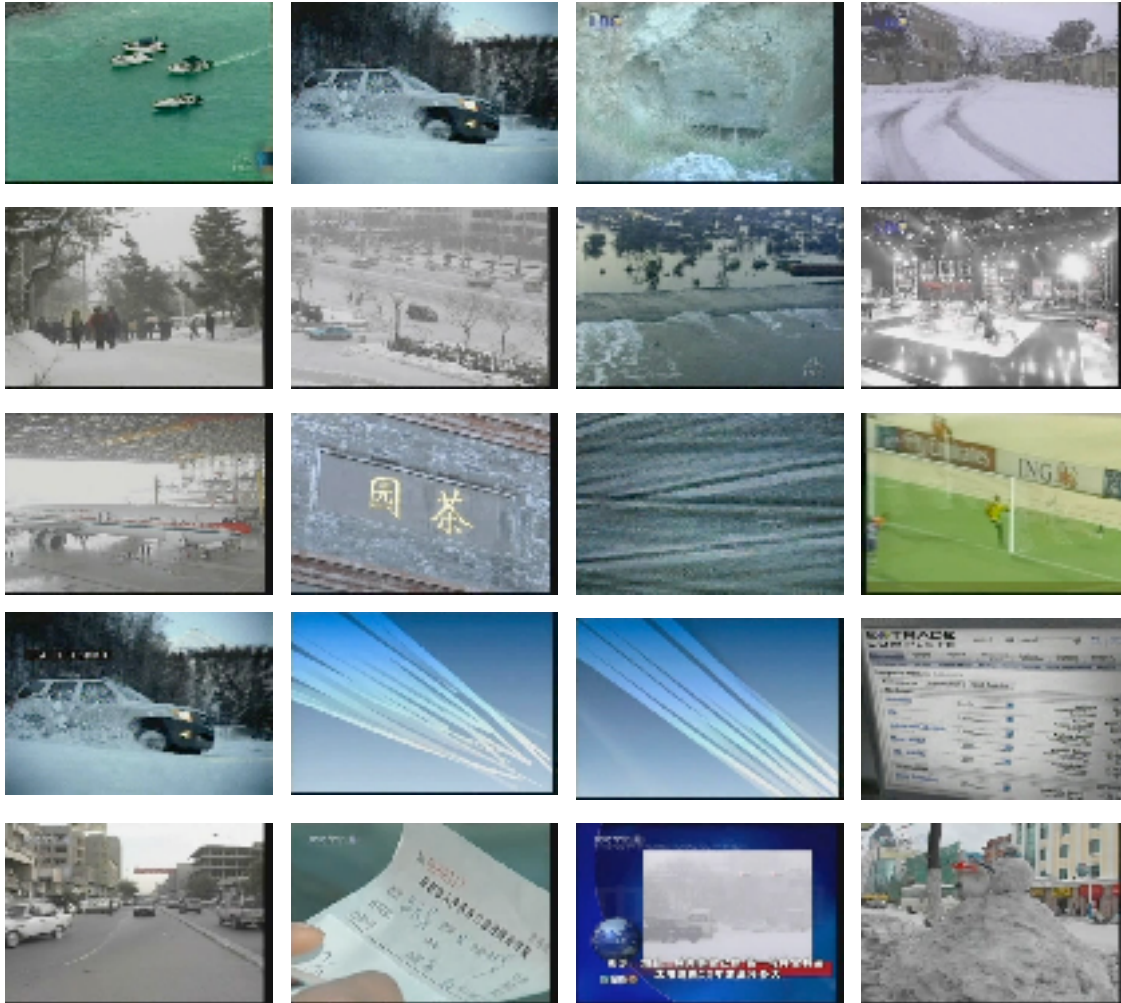
- Stories about snow, some with weather map.
- Text search strong down the list.

Snowy Scenes: Information Bottleneck Reranking



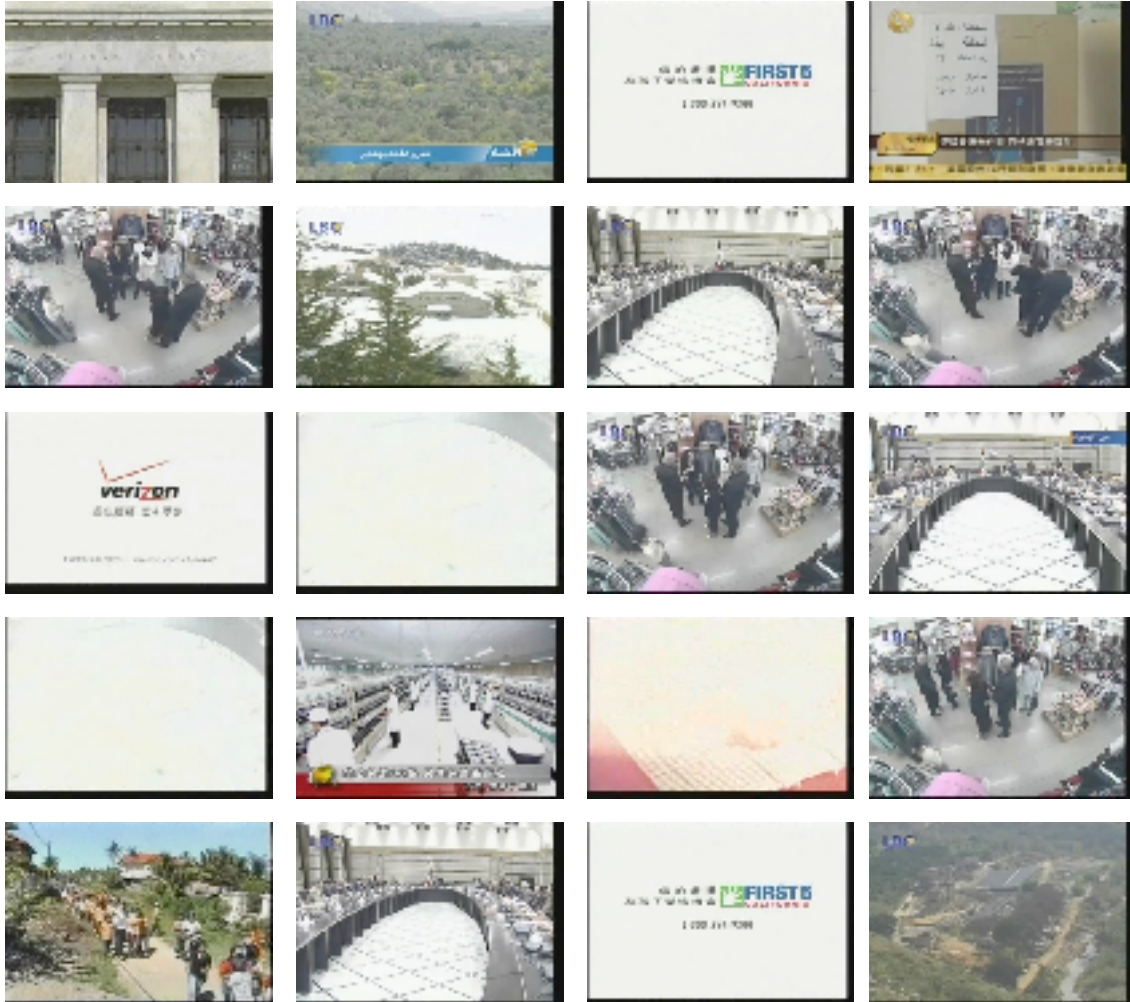
- IB reranking focuses in on:
 - weather map (error)
 - white scene (correct)
- Filters out:
 - anchors
 - other noise

Snowy Scenes: Concept Search



- Map to “snow” concept detector
- Fair performance

Snowy Scenes: Visual Example Search



- Strong white scene cues
- Many false alarms

Snowy Scene: Fused



- **Concept** - 0.85
Example - 0.10
Text - 0.05
- Complementary,
high-precision

Example Query: Soccer

Text:

“Find shots of one or more
soccer goalposts.”

Keywords: soccer goalposts

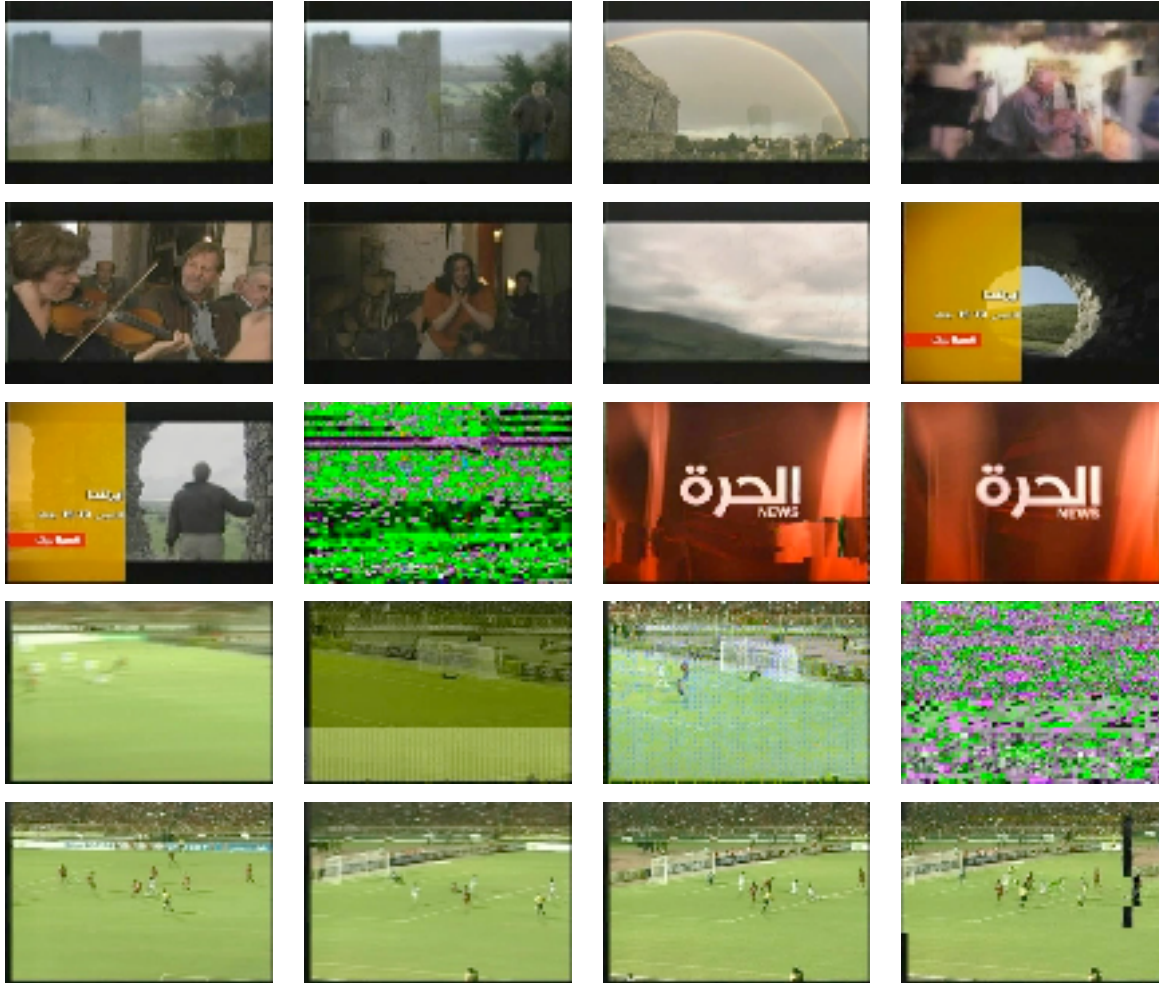
Visual Concepts: soccer

Visual Examples:



- **Query class:**
 - “sports”
- **Text search: (.05)**
 - “soccer goalposts”
- **Concept search: (.15)**
 - “soccer” concept
- **Visual search: (.80)**
 - 4 examples

Soccer: Text Search



- Many stories mention “soccer”
- Most are relevant
- Considerable noise

Soccer: Information Bottleneck Reranking



- Strong green cues

Soccer: Concept Search



- Map to pre-trained “soccer” concept
- Strong performance

Soccer: Visual Examples



- Strong green cues
- Still many errors

Soccer: Fused



- **Concept** - 0.15
Example - 0.80
Text - 0.05
- Complementary,
highly precise

Example Query: Condoleezza Rice

Text:

“Find shots of **Condoleezza Rice**.”

Keywords: condoleezza rice

Visual Concepts: N/A

Visual Examples:



- **Query class:**
 - “Named Person”
- **Text search: (.95)**
 - “condoleezza rice”
- **Concept search: (.00)**
 - N/A
- **Visual search: (.05)**
 - 7 examples

Condoleezza Rice: Text Search



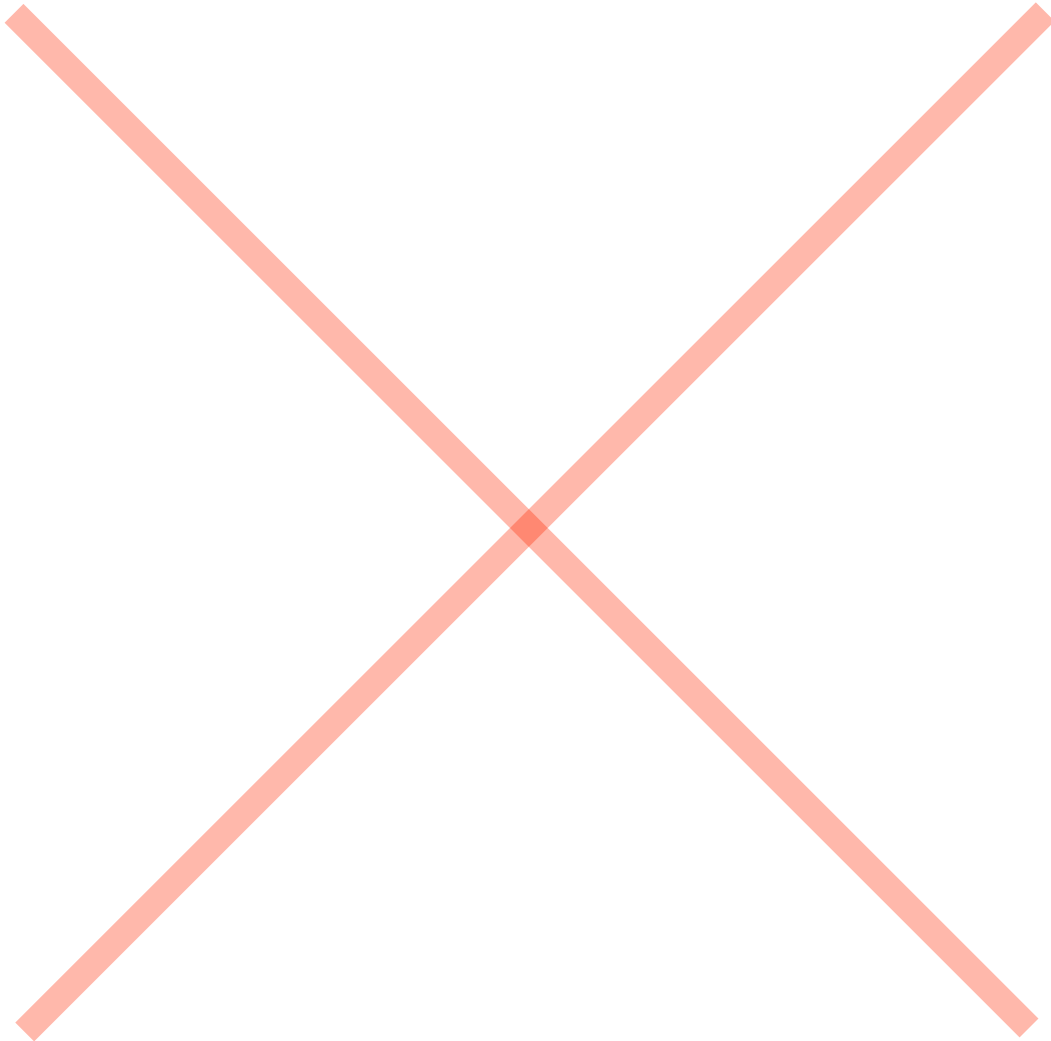
- Text search strong.
- Still many false alarms
 - anchors
 - graphics
 - etc

Condoleezza Rice: Information Bottleneck Reranking



- IB reranking focuses in on:
 - recurrent scenes
- Filters out:
 - anchors
 - other noise

Condoleezza Rice: Concept Search



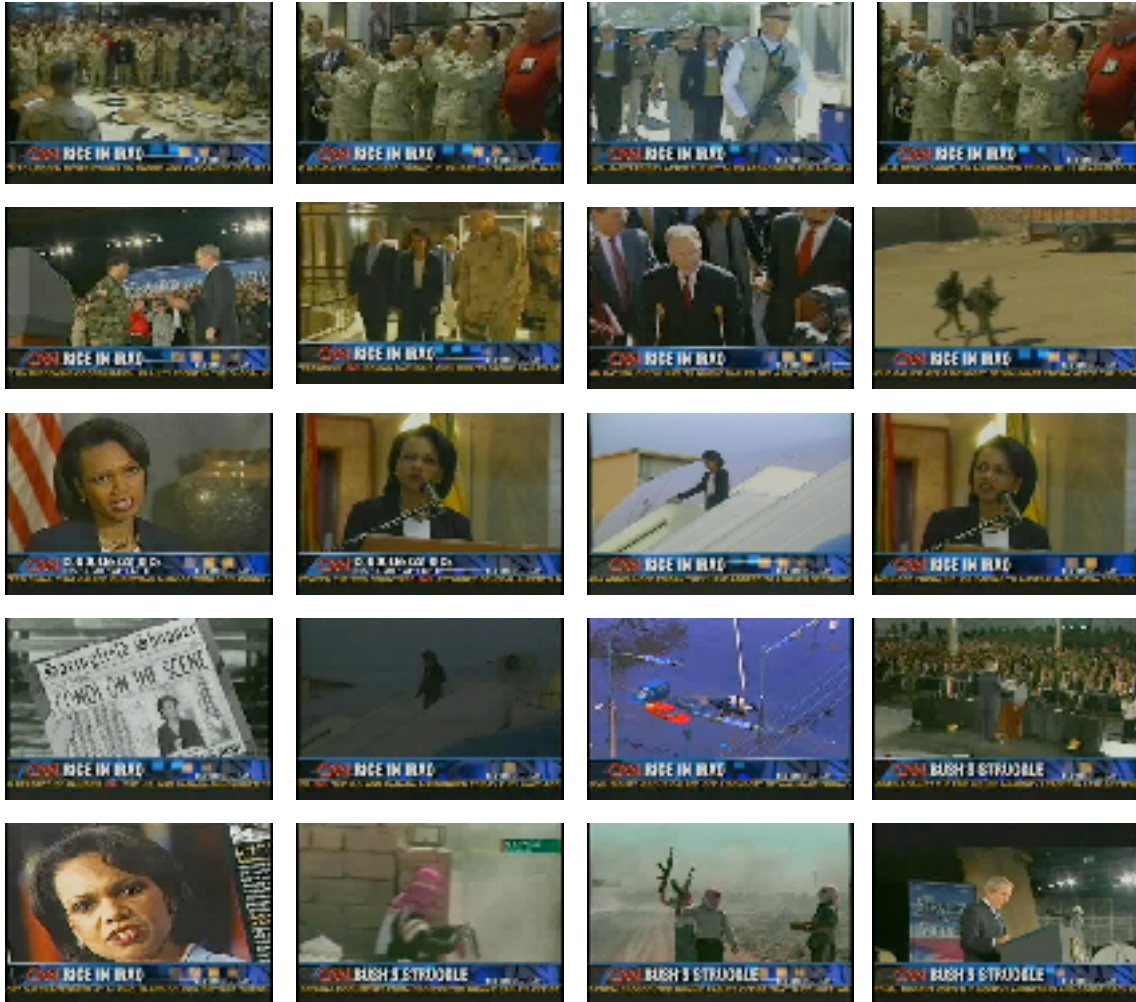
- No matching concepts
- Returns random results

Condoleezza Rice: Visual Example Search



- Non-meaningful results
- Cues
 - anchor background
 - white house

Condoleezza Rice: Fused



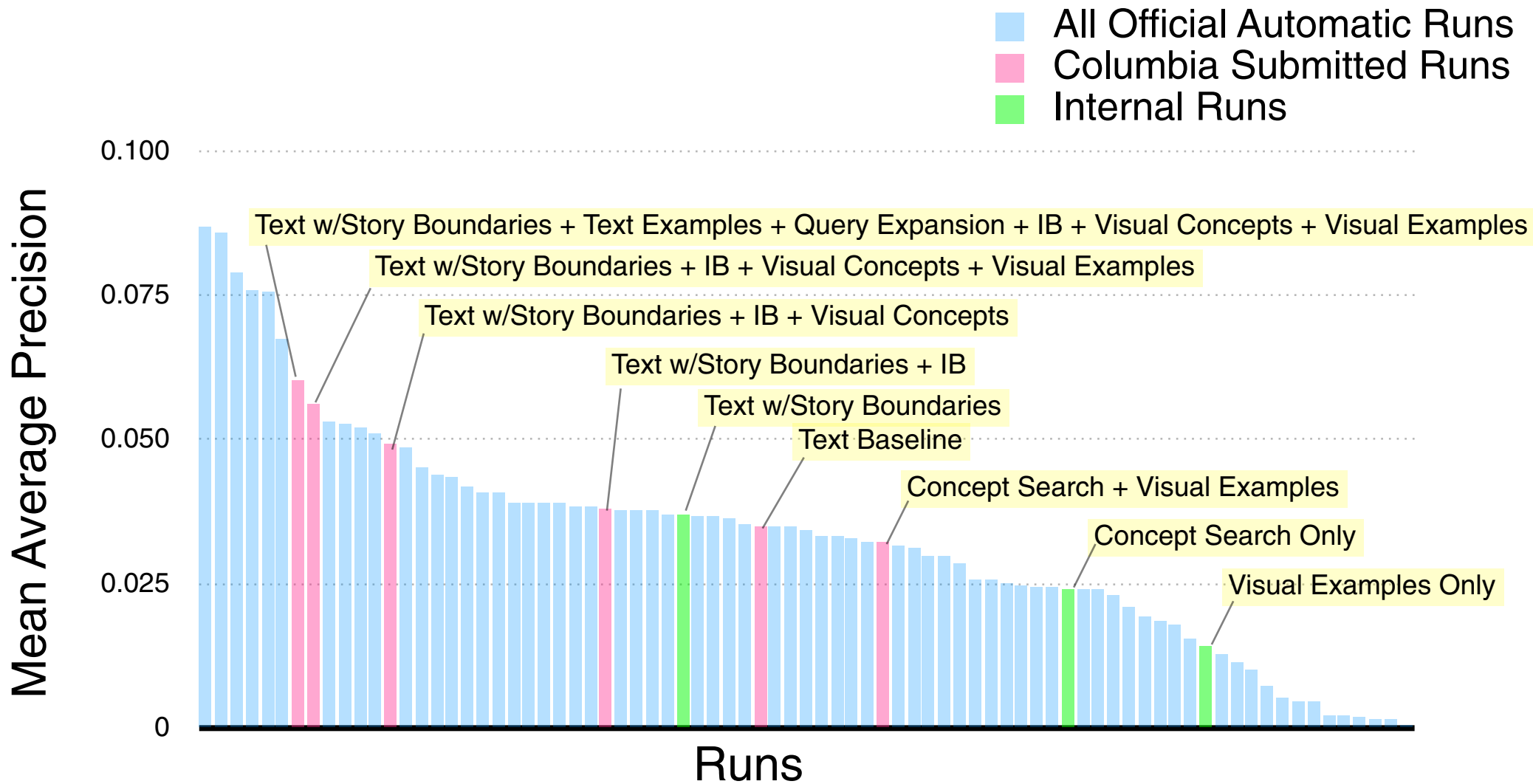
- **Concept** - 0.00
Example - 0.05
Text - 0.95
- little improvement over text alone
- still much noise

Multimodal Search Performance

	Text	IB	Concept	Visual	Fused	Change
Building	0.04	0.03	0.30	0.10	0.33	+10%
Condi Rice	0.19	0.20	0.00	0.01	0.20	+0%
Soccer	0.34	0.42	0.58	0.42	0.83	+43%
Snow	0.19	0.29	0.33	0.03	0.80	+143%
All (2006)	0.08	0.09	0.10	0.07	0.19	+90%

- Values show P@100 (precision of top 100)
- Selective exploitation of stronger methods per query
- Consistent large improvement through class-dependent fusion over best individual method.

Automatic Search Performance



Conclusions: Concept Search

- Concept-based Search is a powerful tool for video search: 30% increase when fused with text.
- Applicable for as many as 2/3 of all queries
- Exploitation of large concept lexicons gives modest improvement: 10x increase in lexicon, 2x increase in search performance
 - **Reason 1:** TRECVID topics biased by LSCOM-lite concepts?
 - **Reason 2:** LSCOM-lite covers most general concepts?
 - Evaluate coverage distances over large *real* query logs?
- Concept search complementary with other search methods (text and visual)

Conclusions:

Query-Class-Dependency

- Multimodal fusion of text, concept, and visual searches improves 70% over text baseline.
 - Concept search accounts for largest portion of the increase.
- Query-Class-Dependency selects right tools for the type of query.
 - Class selection influenced by strengths of individual tools
 - Fusion consistently improves over best individual tool
- Open issues:
 - Mapping queries to concepts.
 - “Find shots of people in uniform in formation” -> soldiers

More...

- LSCOM Annotations and Lexicon
 - <http://www.ee.columbia.edu/dvmm/lscom/>
Update coming soon!
- CuVid Video Search Engine
 - <http://www.ee.columbia.edu/cuvidsearch>
- Digital Video and Multimedia Lab
 - <http://www.ee.columbia.edu/dvmm>