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
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

Multimodal Query
- Text
- Images

Keyword Extraction

Query Expansion

Text Search

Re-ranking

Concept Search

Image Search

Fusion
- Linearly weighted sum of scores
- Query-Class-Dependent weights

Multimodal Search Result
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

- **Named Person**
  - 2006: 4
  - 2005: 6
- **Sports**
  - 2006: 1
  - 2005: 3
- **Concept**
  - 2006: 12
  - 2005: 16
- **Named Entity + Concept**
  - 2006: 2
- **General**
  - 2006: 1
  - 2005: 3
## Query Processing / Classification

<table>
<thead>
<tr>
<th>Incoming Query Topic</th>
<th>Keyword Extraction</th>
<th>Query Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find shots of Condoleezza Rice.</td>
<td>condoleezza rice</td>
<td>Named Entity</td>
</tr>
<tr>
<td>Find shots of scenes with snow.</td>
<td>scenes snow</td>
<td>Matching Concept</td>
</tr>
<tr>
<td>Find shots of one or more soccer goalposts.</td>
<td>soccer goalposts</td>
<td>Sports</td>
</tr>
</tbody>
</table>
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

- 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

[Hsu, CIVR 2005]
Named Entity Query Expansion

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

Find shots of a tennis court with both players visible.

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

AT&T Labs: Miracle multimedia platform. *
Information Bottleneck Reranking

(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

\[ C^* = \arg\min_{C|R} \{I(X; Y) - I(C; Y)\} \]

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

low-level features \( X \)

Y= search relevance
Visual Example Search

- Fusion of many image matching and SVM-based searches
  - 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

- Find shots of snow.
- Find shots of soccer matches.
- Find shots of buildings.
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)
    - 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

Average Precision

TRECVID 2005  TRECVID 2006

LSCOM (374)  LSCOM-lite (39)

Sports
Helicopters
Boats
Newspaper
Soldiers+Weapon+Prisoner
Protest+Building
Smokestacks
# Increasing Lexicon Size

<table>
<thead>
<tr>
<th></th>
<th>39 Concepts</th>
<th>374 Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TV 2005</strong></td>
<td>MAP: 0.0353</td>
<td>MAP: 0.0743</td>
</tr>
<tr>
<td><strong>TV 2006</strong></td>
<td>MAP: 0.0191</td>
<td>MAP: 0.0244</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th></th>
<th>39 Concepts</th>
<th>374 Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV 2005</td>
<td>11 Query Matches</td>
<td>17 Query Matches</td>
</tr>
<tr>
<td></td>
<td>1.1 Concepts/Query</td>
<td>1.3 Concepts/Query</td>
</tr>
<tr>
<td>TV 2006</td>
<td>12 Query Matches</td>
<td>17 Query Matches</td>
</tr>
<tr>
<td></td>
<td>1.8 Concepts/Query</td>
<td>2.5 Concepts/Query</td>
</tr>
</tbody>
</table>

- 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

Examples per concept: 1200

“Prisoner” more frequent than most LSCOM concepts!

Frequency (log)

Concepts (rank)
Concept Detection Performance

Internal evaluation: 2005 validation data

Mean Average Precision: 0.39 0.26

Average Precision

Concepts (rank)

LSCOM  LSCOM-Lite
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

<table>
<thead>
<tr>
<th>Description</th>
<th>Example Images</th>
<th>Concept Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finds shots with one or more emergency vehicles in motion (e.g., ambulance, police car, fire truck, etc.)</td>
<td><img src="image1.png" alt="Example Image" /> <img src="image2.png" alt="Example Image" /></td>
<td><img src="image3.png" alt="Concept Image" /> <img src="image4.png" alt="Concept Image" /></td>
</tr>
<tr>
<td>39: None</td>
<td><strong>374: Emergency_Room, Vehicle</strong></td>
<td><strong>Concept</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible.</td>
<td><img src="image5.png" alt="Example Image" /> <img src="image6.png" alt="Example Image" /></td>
<td><img src="image7.png" alt="Concept Image" /> <img src="image8.png" alt="Concept Image" /></td>
</tr>
<tr>
<td>39: Building</td>
<td><strong>374: Building</strong></td>
<td><strong>Concept</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find shots with one or more people leaving or entering a vehicle.</td>
<td><img src="image9.png" alt="Example Image" /> <img src="image10.png" alt="Example Image" /></td>
<td><img src="image11.png" alt="Concept Image" /> <img src="image12.png" alt="Concept Image" /></td>
</tr>
<tr>
<td>39: None</td>
<td><strong>374: Vehicle</strong></td>
<td><strong>Concept</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find shots with one or more soldiers, police, or guards escorting a prisoner.</td>
<td><img src="image13.png" alt="Example Image" /> <img src="image14.png" alt="Example Image" /></td>
<td><img src="image15.png" alt="Concept Image" /> <img src="image16.png" alt="Concept Image" /></td>
</tr>
<tr>
<td>39: Military_Personnel, Prisoner</td>
<td><strong>374: Guard, Police_Security, Prisoner, Soldier</strong></td>
<td><strong>Concept</strong></td>
</tr>
</tbody>
</table>
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.

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

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

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

39:  Water, Boat.Ship

374: Water, Boat.Ship

39:  Computer_TV-Screen

374: Computer_TV-Screen
## Query Processing

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

**General**

| Find shots of one or more helicopters in flight. | Find shots of a natural scene - with, for example, **fields**, **trees**, **sky**, **lake**, **mountain**, **Waterscape**,**Waterfront**, **Road**, **Sky**.

**Concept**

| Find shots of something burning with flames visible. | Find shots of an explosion or fire. |
## Query Processing

<table>
<thead>
<tr>
<th>Concept</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find shots of a group including least four people dressed in suits, seated, and with at least one flag.</td>
<td>Find shots of at least one person and at least 10 books.</td>
</tr>
<tr>
<td>Find shots containing at least one adult person and at least one child.</td>
<td>Find shots of a greeting by at least one kiss on the cheek.</td>
</tr>
</tbody>
</table>

### Conceptual Results

<table>
<thead>
<tr>
<th>Query</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find shots of a group including least four people dressed in suits, seated, and with at least one flag.</td>
<td>US-Flag</td>
</tr>
<tr>
<td>Find shots containing at least one adult person and at least one child.</td>
<td>None</td>
</tr>
<tr>
<td>Find shots of at least one person and at least 10 books.</td>
<td>None</td>
</tr>
<tr>
<td>Find shots of a greeting by at least one kiss on the cheek.</td>
<td>Election_Greeting</td>
</tr>
</tbody>
</table>
Query Processing

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

- **39:** None

- **374:** Power_Plants, Smoke

Find shots of Condoleezza Rice.

Find shots of scenes of snow.

- **39:** Snow

- **374:** Snow
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

- Query class:
  - “concept”
- Text search: (0.05)
  - “view building story”
- Concept search: (0.85)
  - “building” concept
- Visual search: (0.10)
  - 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

DVMM Lab, Columbia University
Lyndon Kennedy
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: (0.05)
  - “soccer goalposts”
- Concept search: (0.15)
  - “soccer” concept
- Visual search: (0.80)
  - 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:**

- Various images of Condoleezza Rice in different settings.

- **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

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

- 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

- All Official Automatic Runs
- Columbia Submitted Runs
- Internal Runs

Mean Average Precision vs. Runs

Text w/Story Boundaries + Text Examples + Query Expansion + IB + Visual Concepts + Visual Examples
Text w/Story Boundaries + IB + Visual Concepts + Visual Examples
Text w/Story Boundaries + IB
Text w/Story Boundaries
Text Baseline
Concept Search + Visual Examples
Concept Search Only
Visual Examples Only
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