TRECVID 2005 Workshop

Columbia University TRECVID 2005
Search Task

Shih-Fu Chang, Winston Hsu, Lyndon Kennedy,
Akira Yanagawa, Eric Zavesky, Dong-Qing Zhang

Digital Video and Multimedia Lab
Columbia University

Nov. 14 2005
http://www.ee.columbia.edu/dvmm
Columbia Video Search 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

Automatic/Manual Search
- mining query topic classes
- cue-X re-ranking
- user search pattern mining

Interactive Search
- text search
- concept search
- Image matching
- story browsing
- Near-duplicate search

Feature extraction (text, video, prosody)
Automatic story segmentation
Concept detection
Near-duplicate detection

Video
Speech
Text
Information Bottleneck principle

Cue-X Information-theoretic Framework

Information Bottleneck principle

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

\( Y = \) search relevance
\( Y = \) “demonstration”
\( Y = \) story boundary

\( \phi(Y|X = x_1) \)
\( \phi(Y|X = x_2) \)
\( \phi(Y|C = c_k) \)

\( y_0 = \) topic “Arafat”

cluster cond. prob.
(relevance to semantic label)

\( C_1 \)  \( C_2 \)  \( \ldots \)  \( C_k \)  \( \ldots \)  \( C_{K-1} \)  \( C_K \)

\( X \)

low-level features

cue-X clusters automatically discovered via
Information Bottleneck principle & Kernel Density Estimation (KDE)
News Story Segmentation in TRECVID 2005

• Cue-X framework effectively applied to discover salient features and achieve accurate story segmentation
  – Focus on visual and audio (prosody) features only
  – Without a priori manual selection of features
  – High accuracy across multi-lingual data sources

• TRECVID 2005
  – Dataset
    • 277 videos, 3 languages (ARB, CHN, and ENG),
    • 7 channels, 10+ different programs
    • Poor or missing ASR/MT transcripts
  – Accuracy on the validation set
    • Cue-X features + prosody features (no text features!)
    • ARB-0.87, CHN-0.84, and ENG-0.52 (F1 measure)
  – Results donated to whole TRECVID 2005 community

• Story boundary results available for download at http://www.ee.columbia.edu/dvmm/downloads/cuex_story.htm
In other news, Pope John Paul II will get his first look at the shroud of Turin today. The piece of linen, many believe was the burial cloth of Jesus, is on public display for the first time in twenty years. It has already drawn up million visitors. The Pope’s visit to northwest Italy has also included beatification services for three people. The Vatican says John Paul II is now the longest serving Pope this century. He has surpassed Pope Pius the Twelfth, who served for nineteen years, seven months and seven days.

- Stories define an intuitive unit with coherent semantics.
- Story boundaries are effectively detected by Cue-X using audio-visual features.
- Improves text search by more than 100% in TRECVID 2005 automatic search.
- Major contributor to good performance of interactive video search.

Relative contributions from different search tools.
Enhancing Semantic Concept Detection Performance Using Local Features and Spatial Context

**Traditional**
- Color Moment
- Global or block-based features:
  - Difficult to achieve robustness against background clutter
  - Difficult to model object appearance variations

**Enhanced**
- Part-based model:
  - Eliminate background clutter
  - Model part appearance more accurately
  - Model part relation more accurately

- Part
- Part relation
Extracting Graphical Representations of Visual Content and Learning Statistical Models of Content Classes

Individual images → Salient points, high entropy regions

Attributed Relational Graph (ARG)

- size; color; texture
- spatial relation

Graph Representation of Visual Content

Collection of training images

Random Attributed Relational Graph (R-ARG)

Statistics of attributes and relations

Statistical Graph Representation of Model

machine learning
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.
Search Components: Detecting Image Near Duplicates (IND)

- 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

Duplicate detection is the single most effective tool in our Interactive Search
Concept Search

Query

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

Part-of-Speech Tags - keywords
“road car”

Map to concepts
WordNet Resnik semantic similarity

Concept Metadata
Names and Definitions

Concept Space
39 dimensions

(1.0) road
(0.1) fire
(0.2) sports
(1.0) car
.....
(0.6) boat
(0.0) person

Documents

Subshots

Confidence for each concept

Concept Space
39 dimensions

(0.9) road
(0.1) fire
(0.3) sports
(0.9) car
.....
(0.2) boat
(0.1) person

• Map text queries to high-level feature 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** - Can help for queries with related concepts

"Find shots of boats."

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Text</td>
<td>.169</td>
</tr>
<tr>
<td>CBIR</td>
<td>.002</td>
</tr>
<tr>
<td>Concept</td>
<td>.115</td>
</tr>
<tr>
<td>Fused</td>
<td>.195</td>
</tr>
</tbody>
</table>

"Find shots of a road with one or more cars."

<table>
<thead>
<tr>
<th>Method</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Text</td>
<td>.053</td>
</tr>
<tr>
<td>CBIR</td>
<td>.009</td>
</tr>
<tr>
<td>Concept</td>
<td>.090</td>
</tr>
<tr>
<td>Fused</td>
<td>.095</td>
</tr>
</tbody>
</table>

**Manual / Interactive**

Manual keyword selection allows more relationships to be found.

Query Text

“Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people”

Concepts

Office

Query Text

“Find shots of one or more people entering or leaving a building”

Concepts

Person, Building, Urban

Query Text

“Find shots of a graphic map of Iraq, location of Baghdad marked - not a weather map”

Concepts

Map

Query Text

Find shots of people with banners or signs

Concepts

March or protest
Cue-X Reranking by Pseudo-Labeling

- Learn the recurrent relevant and irrelevant low-level patterns from the estimated pseudo-labels
- Reorder shots by the smoothed cluster relevance

Query: "AL clinic bombing"

rough shot list

soft pseudo-labeling
pseudo-label, random variable: $Y$
use $\bar{p}(Y | X)$ only

feature & pseudo-label smoothing w/ kernels

$p(x, y) = \frac{1}{Z(x, y)} \sum_{x_i \in S} K_\sigma(x - x_i) \cdot \bar{p}(y | x_i)$

low-level feature: $X$

(1) Text Search
- OKAPI text query
- Yahoo
- Google

(2) estimated from rough search results (e.g., text search scores), user feedbacks, etc.

(3) cue-x clustering & ranking clusters by $P(Y = ' + | c_i)$

(4) rank within-cluster features by density prob.
Effect of Cue-X Reranking in Video Search

- Improvement over story-based text search (in automatic search TRECVID 2005)
  - 17% in MAP, 46% in soccer (171), 36% in helicopter (158), 32% in Blair (153), 28% in Abbas (154), etc.
  - No search examples provided but discovered automatically
Automatic Discovery of Multimodal Query Classes

- Distinct query classes use customized fusion strategies
- How to automatically discover query classes?
- When and how does each modality help for each query?
- Existing methods: define query classes using human knowledge.
- New method: discover queries according to performance and semantics of searches.
Auto. Discovered Query Clusters

- Learned over a large query topic pool
- Text search and person-X
  - named persons
- Image search
  - named objects,
  - sports, and
  - generic scene classes
- Automated term expansion
  - Google class for cats, birds and airport terminals.
Interactive Activity Logging

Example Log Detail

Post-Mortem Analysis

- Analyze inter-labeler disparity
- Find difficult search topics by high common error rate
- Discover where certain tools failed
- In the future, use actions as passive relevance feedback rounds

Ground truth included in label actions

Detailed search and topic criterion

Aggregate tool actions by search time

Monitor labeling to understand interface usage
Automatic Search (Performance Breakdown)

- Largest improvement from story segmentation
- Noticeable improvements from other components
  - especially cue-x rerank and concept search

<table>
<thead>
<tr>
<th>MAP</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.039</td>
<td>Text</td>
</tr>
<tr>
<td>0.087</td>
<td>Text+Story</td>
</tr>
<tr>
<td>0.095</td>
<td>Text+Story+Anchor Removal</td>
</tr>
<tr>
<td>0.107</td>
<td>Text+Story+Anchor Removal +CueX Re-rank</td>
</tr>
<tr>
<td>0.111</td>
<td>Text+Story+Anchor Removal +CueX Re-rank +CBIR</td>
</tr>
<tr>
<td>0.114</td>
<td>Text+Story+Anchor Removal +CueX Re-rank +CBIR+Concept Search</td>
</tr>
</tbody>
</table>

- + concept search
- + Cue-X re-rank (visual features)
- + story boundary

Run
Automatic Search

Topic

- Rice
- Ailawi
- Karami
- Jintao
- Blair
- Abbes
- Baghdad map
- tennis
- shaking hands
- helicopter
- Bush
- fire
- banners
- enter building
- meeting
- boat
- basketball
- palm trees
- airplane
- road cars
- military vehicles
- building
- soccer
- office
Interactive Tool Contribution

Varied search strategies
- User 1: prefers story browsing, duplicate and traditional search
- User 2: no story discovery, use lots of duplicate browsing

Strategy dynamic for each topic
- Common visual concepts good candidates for duplicates
- Temporal events best suited for discovery by story browsing
- Named entities or specific actions usually best in traditional search methods
Formula for Success:
1. Find positives through any search method
2. Iteratively browse through the near-duplicates or story browsing

Best Overall Performance
160 (fire), 164 (boat), and 162 (entering building)

Close to Best
149 (Rice), 151 (Karami), 153 (Blair), 154 (Abbas), 157 (shaking hands), 161 (banners), 166 (palm trees), 168 (roads/cars), 169 (military vehicles), and 171 (soccer)