Ontology-enriched Semantic Space for Video Search

Xiao-Yong WEI, Chong-Wah Ngo
CS. Dept., City University of Hong Kong

ACM Multimedia 2007, Augsburg, Germany
Vehicle and weapon... they might see a military vehicle.
Semantic Reasoning

vehicle + weapon = ?
Outline

- Introduction
- Ontology-enriched Semantic Space
- Exploring OSS
  - Concept development
  - Query-to-concept mapping
  - Word sense disambiguation
  - Multi-modality fusion
- Experimental Results
- Conclusions
Introduction - Background

- Video Search vs. Semantic Gap

User Level

Query

Query

Query

Query

Find a car

Tag

woman

man

AD

car

Multimedia Level

YouTube
Introduction - Background

- Video Search vs. Semantic Gap

User Level

Query Query Query Query

Multimedia Level
Introduction - Background

- Video Search vs. Semantic Gap

- Semantic Gap

- Natural language

- Machine computable

- User Level

- Query

- Query

- Query

- Query

- Low-Level Features

- Text

- Image

- Motion

- Audio

- Multimedia Level
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search

Concept-based Video Search user interface representation.
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search

How many and which detectors should be developed?
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search

Which concepts should be selected to describe query?
How many and which detectors should be developed?
Introduction - Background

- Video Search vs. Semantic Gap
- Concept-based Video Search

Linguistic Computing

How many and which detectors should be developed?
Which concepts should be selected to describe query?
How to refine the imprecise and incomplete user queries?
Related Work: Concept Development

- Large scale concept ontology for multimedia (LSCOM)
- MediaMill – 101
- TRECVID
Related Work: *Query Answering*

- Query-to-concept mapping
Related Work: *Query Answering*

- Query-to-concept mapping
- Ontology reasoning: Resnik, JCN, WUP

Queries:
- sports
  - ...
Related Work: Query Answering

- Query-to-concept mapping
- Ontology reasoning: Resnik, JCN, WUP
- Comparing to text descriptions of concepts

Queries:
- Sports
- ...

Descriptions:
- Animal: is a ...
- Bus: is a ...
- Baseball: is a ...sports

Concepts:
- Animal
- Football
- Bus
- Soccer
- Baseball
- Car
Related Work: *Query Answering*

- Query-to-concept mapping
  - Ontology reasoning: *Resnik, JCN, WUP*
  - Comparing to text descriptions of concepts
  - Query items expansion

Queries:
- sports
- ...

... = ...
... = ...
football = soccer

Concepts:
- animal
- football
- bus
- soccer
- baseball
- car
Related Work: *Query Answering*

- Query-to-concept mapping
  - Ontology reasoning: Resnik, JCN, WUP
  - Comparing to text descriptions of concepts
  - Query items expansion
  - Co-occurrence statistics (e.g., by Internet)

Queries: 
- sports

Soccer and sports frequently occur together…

Concepts:
- animal
- football
- bus
- soccer
- baseball
- car
Related Work: Query Answering

- Query-to-concept mapping
  - Ontology reasoning: Resnik, JCN, WUP
  - Comparing to text descriptions of concepts
  - Query items expansion
  - Co-occurrence statistics (e.g., by Internet)
  - Model scores (for image and video query examples)
    - [John R. Smith, ICME’03]
Problem

- Concepts/detectors
- query items

Vocabularies Set (General Knowledge)

High-Level Concepts

Low-Level Features

User Level

General Vocabularies

Semantic Gap

Data Flow

Low-Level Representations

Multimedia Level

Query

Query

Query

Query

Concept

Concept

Concept

Concept

Text

Image

Motion

Audio

Which concepts should be selected to describe query?

How many and which detectors should be developed?

Inconsistency if not considered together !!!

query-concept mapping

concepts development
Motivation

- Ontology-enriched Semantic Space (OSS)

**Reasoning:** A computable platform that allows an uniform and global way of choosing detectors.

**Development:** Generalization power of detectors in spanning the Semantic Space.
Outline

- Introduction
- Ontology-enriched Semantic Space
- Exploring OSS
- Experimental Results
- Conclusions
OSS - Global Consistence

- Conventional ontology reasoning
  - Path-based measures

\[ WUP = \frac{2 \cdot \text{depth}(\text{weapon})}{\text{len}(\text{gun, tank}) + 2 \cdot \text{depth}(\text{weapon})} \]
OSS - Global Consistence

- Conventional ontology reasoning
  - Path-based measures
  - Information content (IC)

Resnik = IC(weapon)

JCN = \( \frac{1}{IC(gun) + IC(tank) - 2*IC(weapon)} \)
OSS - Global Consistence

- Conventional Ontology Reasoning

Query: tank

\[ \text{Sim (tank, gun)} = \text{Sim (tank, armored car)} \]

- gun
- tank
- armored car

- gun?
- armored car?
OSS - Global Consistence

- Conventional Ontology Reasoning
  - Local measure
OSS - Global Consistence

- Conventional Ontology Reasoning
- OSS: multiple “key-concepts”

![Diagram showing relationships between concepts]

|     | weapon | vehicle | ...
|-----|--------|---------|-----
| gun |        |         |     
| tank|        |         |     
| armored car |   |         |     
| ...  |        |         |     

WUP(gun, weapon)  WUP(gun, container)
OSS - Global Consistence

- Conventional Ontology Reasoning
- OSS: multiple “key-concepts”
OSS - Construction

- Conventional Ontology Reasoning
- Multiple “key-concepts”
- Linear Semantic Space

\[
\text{Sim(\text{tank, armored car})} > \text{Sim(\text{tank, gun})}
\]
OSS - Computable

- Metric – Cosine Similarity
OSS - Computable

- Metric – Cosine Similarity
- Semantic Reasoning
  - vehicle + weapon = ?

![Diagram](vehicle + weapon = military vehicle)
Outline

- Introduction
- Ontology-enriched Semantic Space
- Exploring OSS
- Experimental Results
- Conclusions
Exploring OSS
– Concept Development

- Constructing OSS on MediaMill-101
- Hierarchical Clustering
Exploring OSS
– *Concept Development*

- Constructing OSS on MediaMill-101
- Hierarchical Clustering
- Selection of basis concepts
  - Completeness
  - Compactness

<table>
<thead>
<tr>
<th>Basis</th>
<th>Cluster Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>water</td>
<td>river, water, waterfall</td>
</tr>
<tr>
<td>vehicle</td>
<td>tank, bicycle, vehicle</td>
</tr>
<tr>
<td>golf</td>
<td>soccer, football, golf</td>
</tr>
<tr>
<td>entertainment</td>
<td>racing, cycling, sports, entertainment</td>
</tr>
</tbody>
</table>
Exploring OSS
— Guidelines to Concept Development

- **OSS**
  - **Generalization**: given priority to develop general concepts (more training examples)
  - **Feasibility**: concepts not feasible to be developed → use the axes as reference

- [W. H. Lin, Alexander Hauptmann, ICME’06]
  - **Utility**: given priority to frequent and scene-based concepts
Exploring OSS

Query-to-concept Mapping

- Project query items into OSS
- Top-$k$ concepts
  maximize similarities between query items and concepts
Exploring OSS

Word Sense Disambiguation (WSD)

- Maximize similarities among senses
- = Minimize distances among senses
- “Find a graphic map of Iraq”
Exploring OSS

Word Sense Disambiguation (WSD)

- Maximize similarities among senses
- Minimize distances among senses
- “Find a graphic map of Iraq”
Exploring OSS

Word Sense Disambiguation (WSD)

- Maximize similarities among senses
- Minimize distances among senses
- “Find a graphic map of Iraq”
Exploring OSS

Multi-modality Fusion

- Query dependent fusion
  - Clustering
  - Weighting

Text | Image | Motion | Audio
Exploring OSS

*Multi-modality Fusion*

- Query dependent fusion
- Distribution of MediaMill-101 with Multi-Dimensional Scaling (MDS)
Exploring OSS

Multi-modality Fusion

- Distribution of MediaMill-101 with Multi-Dimensional Scaling (MDS)
- Correlation between concepts clusters and multi-modality features – matrix $R$

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>visual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

detector reliability
Exploring OSS

**Multi-modality Fusion**

- Distribution of MediaMill-101 with Multi-Dimensional Scaling (MDS)
- Correlation between concepts clusters and multi-modality features - $R$
- Fuzzy synthetic evaluation

![Diagram showing fuzzy composition operation between query item \( q_i \), belonging scores, and weights for modalities.]

- \( U_1(q_i) \) to \( R \) with \( C_1 \)
- \( U_2(q_i) \) to \( R \) with \( C_3 \)
- \( U_3(q_i) \) to \( R \)
- Weights for modalities: \( w_1(q_i) \), \( w_2(q_i) \), \( w_3(q_i) \)

**Modality 1**

**Modality 2**

**Modality 3**
Exploring OSS

**Multi-modality Fusion**

- Distribution of MediaMill-101 with Multi-Dimensional Scaling (MDS)
- Correlation between concepts clusters and multi-modality features – matrix $R$
- Fuzzy synthetic evaluation

Query-concept relation

![Diagram of Query-concept relation]

Query-modality relation

![Diagram of Query-modality relation]
Outline

- Introduction
- Ontology-enriched Semantic Space
- Exploring OSS
- Experimental Results
- Conclusions
Experimental Results

Datasets and Evaluation

- Automatic search task of TRECVID 2005
  - 45,765 shots (about 85 hours)
- Twenty-four search topics (text only)
- Ground truth provide by TRECVID’05
- MediaMill-101 concept detectors
  - 80 after removing concepts not defined in WordNet
- Evaluation: Mean Average Precision (MAP)
Experimental Results

*Basis Selection vs. Search*

- Agglomerative hierarchical clustering
- Inconsistency coefficient - tightness
  - [A. K. Jain, Algorithm for Clustering Data, 1988]
Experimental Results

*Basis Selection vs. Search*

- Agglomerative hierarchical clustering
- Inconsistency coefficient - tightness

[A. K. Jain, Algorithm for Clustering Data, 1988]
Experimental Results

*Basis Selection vs. Search*

- Agglomerative hierarchical clustering
- Inconsistency coefficient
- Search by selecting top-1 concept
Experimental Results

*Word Sense Disambiguation (WSD)*

- Comparing with Lesk (gloss overlap)
  - [S. Banerjee, Computational Linguistics and Intelligent Text Processing, 2002]

<table>
<thead>
<tr>
<th></th>
<th>Query Terms</th>
<th>Correct Sense</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesk</td>
<td>70</td>
<td>56</td>
<td>80.00%</td>
</tr>
<tr>
<td>OSS</td>
<td>70</td>
<td>58</td>
<td>82.85%</td>
</tr>
</tbody>
</table>
Experimental Results

Concept-based Video Search

Comparing with popular ontology measures

- LCH [C. Leacock, 1998] - Path length-based
- WUP [Z. Wu, 1994]
- RES [P. Resnik, 1995]
- LIN [D. Lin, 1997]
- JCN [J. J. Jiang, 1997] - Information Content-based
Experimental Results

Concept-based Video Search

- Comparing with popular ontology measures
- Single concept selection (SCS)
- Multiple concept selection (MCS)

<table>
<thead>
<tr>
<th></th>
<th>LCH</th>
<th>WUP</th>
<th>RES</th>
<th>LIN</th>
<th>JCN</th>
<th>OSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCS</td>
<td>0.0213</td>
<td>0.0213</td>
<td>0.0418</td>
<td>0.0104</td>
<td>0.0104</td>
<td>0.0486</td>
</tr>
<tr>
<td>MCS</td>
<td>0.0460</td>
<td>0.0533</td>
<td>0.0423</td>
<td>0.0344</td>
<td>0.0478</td>
<td>0.0543</td>
</tr>
</tbody>
</table>
## Experimental Results

### Multi-Modality Fusion

- **Retrieval-by-ASR (text)**
- **Retrieval-by-Concept (concept)**

<table>
<thead>
<tr>
<th></th>
<th>Text</th>
<th>WAF</th>
<th>PQF</th>
<th>OSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>0.0601</td>
<td>0.0798</td>
<td>0.0874</td>
<td>0.0922</td>
</tr>
<tr>
<td>Improve</td>
<td>-</td>
<td>32.78%</td>
<td>45.42%</td>
<td>53.46%</td>
</tr>
</tbody>
</table>

**WAF**: weighted average fusion: 0.6 for text, 0.4 for concept

**PQF**: pseudo query-class dependent fusion: queries containing name entities (0.7 for text, 0.3 for concept); otherwise: 0.5 for each
Conclusion

- OSS is a *computable* Platform
- Uniform and *consistent* measurement
- *Guideline* for concept development
- *A feasible solution* for
  - Large-scale video search
  - Concept fusion
  - Query dependent fusion
Conclusion – Possible Extension

• Construction based on
  • Spectral decomposition – orthogonal space
  • Nonlinear assumption
• Co-occurrence statistics of concepts
• Negative concepts
Thanks!