Incorporating Concept Ontology into Multi-level Image Indexing

Roger Wong

Principal Supervisor: Prof. Clement Leung

31st Aug 2009

10th Postgraduate Research Symposium

Department of Computer Science

Hong Kong Baptist University

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

Research objective

- To develop a novel framework to enable more effective interpretation of semantic concepts and multi-level annotation
- To resolve issues of scenes detections in Automatic Semantic Annotation
- Propose extensions of image annotation models which are using ontology-based expansion and contextual feature-based expansion.
- To manage 100,000+ web images and 1,000,000 tags

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

Automatic Semantic Annotation (ASA) Approach

- a rule-based approach to formulate annotations for images fully automatically
- based on the use of image parametric dimensions and metadata

Rules for scenes of images

Rule 1) $$\forall i \in I, (t_i > 0.125) \wedge (d_i > 30) \wedge (EV_i \leq 8) \Rightarrow i \in S_n$$

Rule 2) $$\forall i \in I, (d_i > 30) \wedge (EV_i > 8) \wedge (t_i \leq 0.125) \Rightarrow i \in S_d$$

Rule 3) $$\forall i \in I, (f_i > 20) \wedge (d_i > 50) \wedge (EV_i > 11) \Rightarrow i \in S_{ss}$$

Rule 4) $$\forall i \in I, [(f_i \leq 5.6) \wedge (5 < d_i \leq 8)] \wedge \{[(t_i \leq 0.00625) \wedge (L_i \leq 30)] \vee [(30 < L_i \leq 182) \wedge (ISO_i \leq 250)] \vee (L_i > 182) \vee (t_i \leq 0.003125)\} \Rightarrow i \in S_{op}$$

Rule 5) $$\forall i \in I, (f_i > 5.6) \wedge (L_i \leq 25) \wedge (5 < d_i \leq 8) \wedge (t_i > 0.003125) \Rightarrow i \in S_{oe}$$

Rule 6) $$\forall i \in I, (f_i > 5.6) \wedge (0.003125 < t_i \leq 0.011111) \wedge (5 < d_i \leq 8) \wedge (L_i > 25) \Rightarrow i \in S_{ip}$$

Rule 7) $$\forall i \in I, (5 < d_i \leq 8) \wedge \{(f_i \leq 5.6) \wedge \{[(L_i \leq 30) \wedge (t_i > 0.00625)] \vee [(ISO_i > 250) \wedge (30 < L_i \leq 182)]\}\} \vee [(h_i = 1) \wedge (f_i > 5.6) \wedge (L_i > 25) \wedge (t_i < 0.011111)] \Rightarrow i \in S_{ie}$$

Rule 8) $$\forall i \in I, (d_i > 10) \wedge (150 < L_i \leq 400) \wedge (t_i \leq 0.005) \Rightarrow i \in S_s$$

Rule 9) $$\forall i \in I, (d_i \leq 5) \wedge (EV_i > 9) \Rightarrow i \in S_m$$

Rule 10) $$\forall i \in I, (L_i > 450) \wedge (d_i > 20) \Rightarrow i \in S_w$$

Results of Semantic Queries

"Night scene in Hong Kong in Summer"

"Sunset by the sea in New York in Autumn"

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

Ontology-based tree expansion

- **Ontology-based tree expansion**:
  - An ontology is a representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain.
  - The presence of particular objects in an image often implies the presence of other objects.
  - For example: an ”orchestra” expanded to violins and trumpets

Semantic Knowledge from WordNet

- **WordNet**
  - is one of these applications of semantic lexicon for the English language and general knowledge base and commonsense reasoning engine.

Normalized Google Distance

- **Normalized Google Distance**
  - Normalized Google distance(NGD) is proposed to calculate the relationship between two concepts by their correlation in the search result from Google search engine when querying both concepts.

$$NGD(x,y) = \frac{max\{logf(x), \log(y)\} - logf(x,y)}{logN - min\{lgof(x), logf(y)\}}$$

Wikipedia Distance

- **Wikipedia Distance**
  - Wikipedia is the world largest collaboratively edited source of encyclopedic knowledge.
  - The Wikipedia Link Vector Model (WLVM) uses Wikipedia to provide structured world knowledge about the terms of interest. Their approaches are using the hyperlink structure of Wikipedia rather than its category hierarchy or textual content.

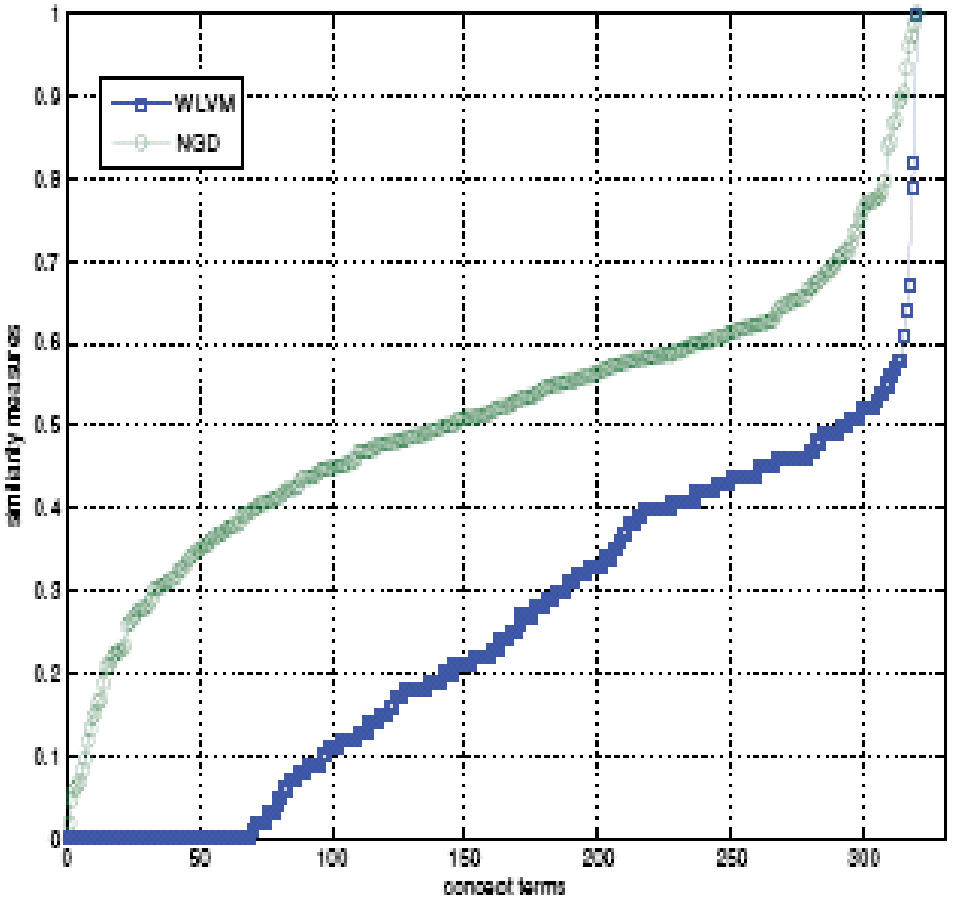
$$w(a \rightarrow b) = |a \rightarrow b| \times log(\sum_{x=1}^{\tau} \frac{t}{|x \rightarrow b|}),$$

where a and b denotes the search terms.

Experimental protocol

- We annotate night scenes based on the prior rule-based approach to extract 422 out of 103,527 images. We also gather 1108 tags associated with those images and totally 417 unique terms are formed.

semantic relatedness measures of LVMD and GND



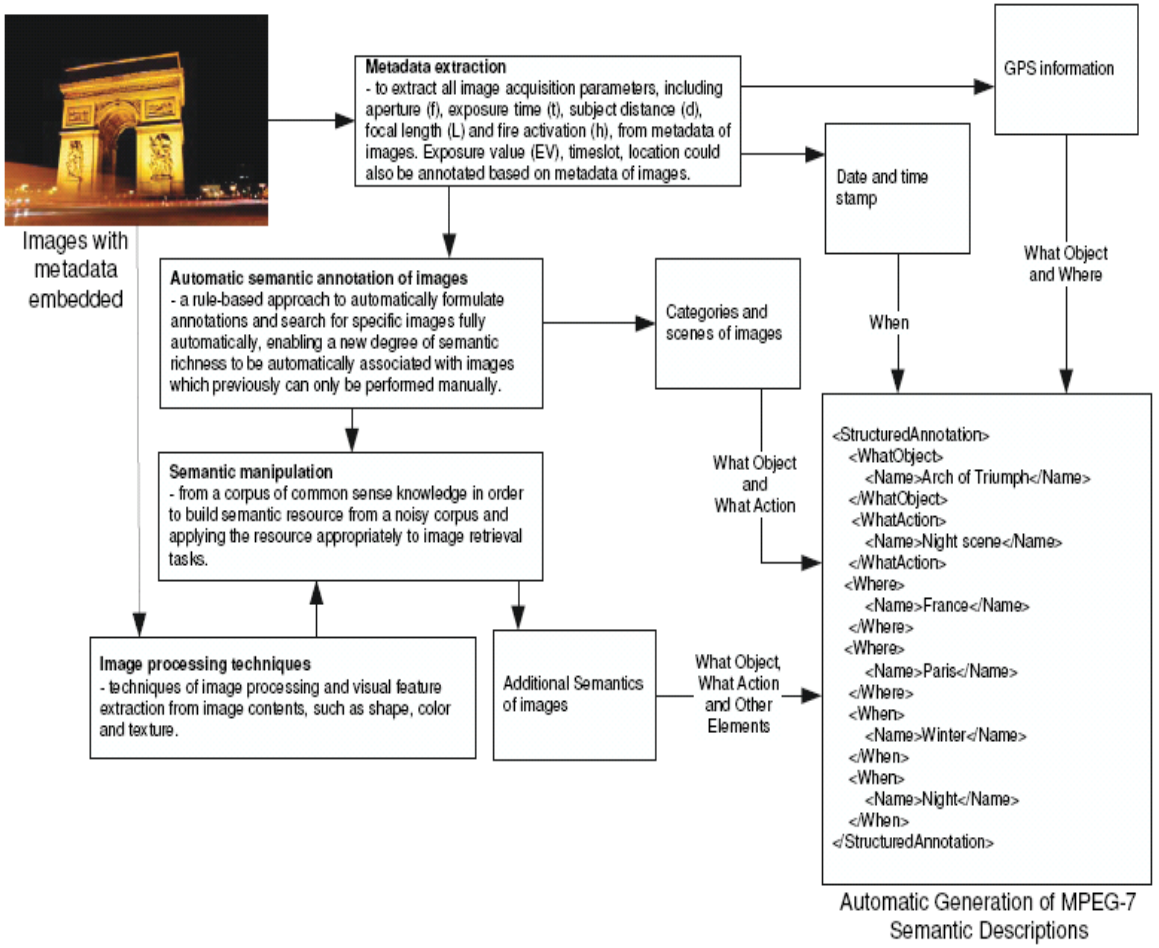
top	WLVM	NGD
1	downtown (100%)	downtown (100%)
2	airport (79%)	driving (99%)
3	street (55%)	harbor (97%)
4	landmark (54%)	station (96%)
5	park (52%)	train (93%)
6	freeway (50%)	street (90%)
7	urban (48%)	lights (90%)
8	city (46%)	trails (89%)
9	transportation (44%)	transportation (87%)
10	highway (44%)	highway (87%)
11	tower (43%)	city (84%)
12	construction (40%)	airport (84%)
13	hotel (40%)	camp (80%)
14	road (40%)	bridge (78%)
15	car (38%)	museum (78%)
16	railway (35%)	outside (78%)
17	bridge (34%)	freeway (77%)
18	architecture (33%)	airplane (77%)
19	airplane (28%)	urban (77%)
20	station (27%)	railway (77%)
...	...	...
413	corinth (0%)	auckland (0%)
414	destiny (0%)	han (0%)
415	dual (0%)	berlin (0%)
416	great (0%)	barcelona (0%)
417	greek (0%)	new (0%)

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

MPEG Standard

- MPEG-1 (1993):
  - The first compression standard for audio and video.
- MPEG-2 (1995):
  - Transport, video and audio standards for broadcast-quality television.
- MPEG-4 (1998):
  - MPEG-4 uses further coding tools with additional complexity to achieve higher compression factors than MPEG-2.
- **MPEG-7:**
  - **A multimedia content description standard.**
- MPEG-21:
  - MPEG describes this standard as a multimedia framework and provides for intellectual property management and protection.

Automatic Generation of MPEG-7 Semantic Descriptions



Automatic Generation of MPEG-7 Semantic Descriptions

```
<StructuredAnnotation>
    <WhatObject>
        <Name>Landscape</Name>
    </WhatObject>
    <WhatAction>
        <Name> Night scene</Name>
    </WhatAction>
    <WhatObject>
        <Name> Eiffel Tower</Name>
    </WhatObject>
    <WhatObject>
        <Name> downtown</Name>
    </WhatObject>
    <Where>
        <Name> Paris</Name>
    </Where>
    <Where>
        <Name>France</Name>
    </Where>
    <When>
        <Name>Summer</Name>
    </When>
    <When>
        <Name>Night</Name>
    </When>
    <When>
        <Name>2008</Name>
    </When>
</StructuredAnnotation>
```

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

Conclusion

- A novel approach for automatic semantic annotation of web images with no human involvement
- An extension of ASA image annotation models with ontology
- Able to automatically fill up the Structured Annotation fields in the MPEG-7 Description Standard which previously could only be performed manually.

- Introduction
- Automatic Semantic Annotation (ASA) Approach
- Ontology-based expansion
- Automatic Generation of MPEG-7 Semantic Descriptions
- Conclusion
- Q & A

Q & A