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) \( \forall i \in I, (t_i > 0.125) \land (d_i > 30) \land (EV_i \leq 8) \Rightarrow i \in S_n \)

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

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

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

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

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

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

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

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

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

\[
NGD(x, y) = \frac{\max\{\log f(x), \log(y)\} - \log f(x, y)}{\log N - \min\{\log f(x), \log f(y)\}}
\]
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.

\[
w(a \rightarrow b) = |a \rightarrow b| \times \log \left( \sum_{x=1}^{t} \frac{i}{|x \rightarrow b|} \right),
\]

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
<table>
<thead>
<tr>
<th>top</th>
<th>WILM</th>
<th>NGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>downtown (100%)</td>
<td>downtown (100%)</td>
</tr>
<tr>
<td>2</td>
<td>airport (79%)</td>
<td>driving (99%)</td>
</tr>
<tr>
<td>3</td>
<td>street (55%)</td>
<td>harbor (97%)</td>
</tr>
<tr>
<td>4</td>
<td>landmark (54%)</td>
<td>station (96%)</td>
</tr>
<tr>
<td>5</td>
<td>park (52%)</td>
<td>train (93%)</td>
</tr>
<tr>
<td>6</td>
<td>freeway (50%)</td>
<td>street (90%)</td>
</tr>
<tr>
<td>7</td>
<td>urban (48%)</td>
<td>lights (90%)</td>
</tr>
<tr>
<td>8</td>
<td>city (46%)</td>
<td>trails (89%)</td>
</tr>
<tr>
<td>9</td>
<td>transportation (44%)</td>
<td>transportation (87%)</td>
</tr>
<tr>
<td>10</td>
<td>highway (44%)</td>
<td>highway (87%)</td>
</tr>
<tr>
<td>11</td>
<td>tower (43%)</td>
<td>city (84%)</td>
</tr>
<tr>
<td>12</td>
<td>construction (40%)</td>
<td>airport (84%)</td>
</tr>
<tr>
<td>13</td>
<td>hotel (40%)</td>
<td>camp (80%)</td>
</tr>
<tr>
<td>14</td>
<td>road (40%)</td>
<td>bridge (78%)</td>
</tr>
<tr>
<td>15</td>
<td>car (38%)</td>
<td>museum (78%)</td>
</tr>
<tr>
<td>16</td>
<td>railway (35%)</td>
<td>outside (78%)</td>
</tr>
<tr>
<td>17</td>
<td>bridge (34%)</td>
<td>freeway (77%)</td>
</tr>
<tr>
<td>18</td>
<td>architecture (33%)</td>
<td>airplane (77%)</td>
</tr>
<tr>
<td>19</td>
<td>airplane (28%)</td>
<td>urban (77%)</td>
</tr>
<tr>
<td>20</td>
<td>station (27%)</td>
<td>railway (77%)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>413</td>
<td>corinth (0%)</td>
<td>auckland (0%)</td>
</tr>
<tr>
<td>414</td>
<td>destiny (0%)</td>
<td>han (0%)</td>
</tr>
<tr>
<td>415</td>
<td>dual (0%)</td>
<td>berlin (0%)</td>
</tr>
<tr>
<td>416</td>
<td>great (0%)</td>
<td>barcelona (0%)</td>
</tr>
<tr>
<td>417</td>
<td>greek (0%)</td>
<td>new (0%)</td>
</tr>
</tbody>
</table>
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.
  - 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 semantic annotation of images
- a rule-based approach to automatically formulate annotations and search for specific images fully automatically, enabling a new degree of semantic richness to be automatically associated with images which previously can only be performed manually.

Images with metadata embedded

Image processing techniques
- techniques of image processing and visual feature extraction from image contents, such as shape, color and texture.

Metadata extraction
- to extract all image acquisition parameters, including aperture (f), exposure time (t), subject distance (d), focal length (L) and fire activation (V), from metadata of images. Exposure value (EV), location, location could also be annotated based on metadata of images.

Additional Semantics of images

Date and time stamp

GPS Information

Categories and scenes of images

When

What Object and What Action

What Object

Suggested Annotation

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