Multimedia semantic analysis technologies and their potential uses

Industry Day SAMT 2006

Yiannis Kompatsiaris
Multimedia Knowledge Group
CERTH - Informatics and Telematics Institute
Multimedia Content

Ontologies
Vocabularies

Networks
Web 2.0 photo - video applications

Storage & Devices

Segmentation
KA Analysis

Labeling

Cross-media
analysis

Context

Reasoning

Metadata
Generation & Representation

Application

Semantic Middleware

Application Server
Database
Operating System
tomorrow

Content adaptation and distribution - Multiple Terminal & Networks

Hybrid / Content-based
recommendations
and personalization

Semantic technology
in Markets

Multimedia Knowledge Group
Informatics and Telematics Institute
Evolution of Content

- 1-2 exabytes (millions of terabytes) of new information produced world-wide annually
- 80 billion of digital images are captured each year
- Over 1 billion images related to commercial transactions are available through the Internet
- This number is estimated to increase by ten times in the next two years.
- 4,000 new films are produced each year
- 300,000 world-wide available films
- 33,000 television stations and 43,000 radio stations
- 100 billions of hours of audiovisual content
Need for annotation + metadata

“The value of information depends on how easily it can be found, retrieved, accessed, filtered or managed in an active, personalized way”
Text-based indexing

- **Manual annotation**
  - + Straightforward
  - + High/Semantic level
  - + Efficient during content creation
- Most commonly used
- Necessary in a number of applications
- - Time consuming
- - Operator-application dependent
- - Text related problems (synonyms etc)

- **Annotation using captions and related text**
  - Web, Video, Documents etc
  - + Straightforward
  - + High/Semantic level
  - + Multimodal approach
  - - Text processing restrictions and limitations
  - - Captions must exist
Text-based indexing
Collaborative tagging

e.g. Flickr
Web 2.0 applications
Addressing the **Semantic Gap**

- **Semantic Gap** for multimedia: To map automatically generated numerical low level-features to higher level human-understandable semantic concepts

This image contains a sky region and is a holiday image

```xml
<?xml version='1.0' encoding='ISO-8859-1' ?>
<Mpeg7 xmlns...>
  <DescriptionUnit xsi:type = "DescriptorCollectionType">
    <Descriptor xsi:type = "DominantColorType">
      <SpatialCoherency>31</SpatialCoherency>
      <Value>
        <Percentage>31</Percentage>
        <Index>19 23 29</Index>
        <ColorVariance>0 0 0</ColorVariance>
      </Value>
    </Descriptor>
  </DescriptionUnit>
</Mpeg7>
```

Dominant Color Descriptor of a sky region
Problem definition

• **Semantic image analysis**: how to translate the automatically extracted visual descriptions into human-like conceptual ones

• **Low-level features** provide **cues** for strengthen/weaken evidence based on visual similarity

• **Prior knowledge** is needed to support semantics disambiguation
Use of ontologies

- Metadata representation
  - interoperability
- Ontology-driven analysis
- Reasoning
  - Extracting higher-level annotations
- Retrieval
- Personalization
- Semantic Web
Indexing using Low-Level Visual Features

- **Low-level features** (color, texture, shape, edges, motion, etc)
  - + automatic extraction
  - + computation efficiency
  - Suitable for many applications
  - - not semantic
  - - algorithm complexity

- **representation**
  - features
  - color, texture space
  - invariance
  - compactness

- **indexing (MPEG-7)**
- **database**
- **matching – distance**
- **global – local features** (segmentation)
Region-Based Query-by-Example

Results:
25 items total (1 - 5) Search completed in 4688 secs...

Original Image

PFZL (TUM)

113026.jpg 86.88%
192x128px, 8KB Find Similar

113026.jpg 82.37%
192x128px, 8KB Find Similar

113026.jpg 81.80%
192x128px, 8KB Find Similar

113001.jpg 81.37%
192x128px, 7KB Find Similar

113050.jpg 80.88%
192x128px, 8KB Find Similar

113032.jpg 80.25%
192x128px, 7KB Find Similar

SCHEMA MPEG-7 XM based Reference System

http://media.iti.gr/site/SchemaXM
Query-by-Example Application (Like.com)
Scene Classification

- **General concept-based classification** (people-no people / indoors – outdoors)
  - + automatic
  - + computational efficient
  - + semantic classes
- Appropriate for a number of applications
- - training and classification limitations
- - predefined restricted classes
Semantic Analysis

• Aims to develop automated methods for semantic annotation of multimedia content
  • LL features are analyzed to recognize objects and events
• Object/Events/Relationships knowledge is needed
• Techniques for knowledge extraction and representation → knowledge base
  • Learning techniques, classification, pattern recognition (implicit knowledge)
  • Model-based techniques (explicit knowledge)
• Specific domains (e.g. sports, news)
• Multimodal, context-assisted approaches are usually followed (e.g. audio-assisted video analysis: goal detection)
Context and Reasoning for Analysis

Creation of contextual information

• Use of contextual information
  • From metadata layer
  • spatio-temporal relations
  • Domain knowledge
• Reduction of label sets
• Merging of segments

multimedia reasoning
Automatic annotation of holidays images
Level of automation keeps increasing (IBM-TRECVID)
IBM Marvel

- MPEG-7 Video Search Engine
- Automatic indexing:
  - Shot detection/key-frame extraction
  - Feature Extraction
  - Semantic Concept Detection
- Search methods:
  - Model-based retrieval (MBR) – statistical modeling and detection of semantic concepts - faces, people, outdoors, etc.
  - Content-based retrieval (CBR) - color, texture, edges, etc.
  - Text-based retrieval (TBR) – textual metadata, annotations, speech transcript
  - Model-vector based retrieval (MVBR) = MBR + CBR
- Interaction:
  - Multi-example relevance feedback searching
  - Iterative searching (combination methods and aggregation functions)
- On-line demo:
  - http://mp7.watson.ibm.com
ITI REACH
http://reach.iti.gr

<table>
<thead>
<tr>
<th>ΚΑΚΤΟΣ</th>
<th>ΒΑΖΟ ΜΕ ΓΑΡΥΦΑΛΛΑ</th>
<th>ΑΝΘΗ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Εύρεση Ομοιών</td>
<td>Πληροφορίες</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Αγορά Συναφών Υλικών</td>
<td>Εύρεση Ομοιών</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Εύρεση Ομοιών</td>
<td>Πληροφορίες</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ΤΟΠΙΟΓΡΑΦΙΑ</th>
<th>ΑΓΡΟΤΟΣΠΙΤΟ</th>
<th>ΤΟΠΙΟ ΜΕ ΔΕΝΔΡΑ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Εύρεση Ομοιών</td>
<td>Πληροφορίες</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Αγορά Συναφών Υλικών</td>
<td>Εύρεση Ομοιών</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Εύρεση Ομοιών</td>
<td>Πληροφορίες</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Αγορά Συναφών Υλικών</td>
<td>Εύρεση Ομοιών</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
<tr>
<td>Εύρεση Ομοιών</td>
<td>Πληροφορίες</td>
<td>Προσθήκη σε προσωπικό κατάλογο</td>
</tr>
</tbody>
</table>
ITI REACH
Automatic query generation and recommendation
use case / domain

requirements

different modalities

different datasets

different knowledge

security

medical images

news video

personal photos

scientific data

surveillance

Multimedia Knowledge Group
Informatics and Telematics Institute
Users - Applications

Audiovisual Sector (TV stations, digital - subscriber TV, advertisement, production, designers, photographers, etc)
- Adaptation, summarization
- «I want images of the Prime Minister»
- «I want all goals of the national team in this game»

Mobile Sector
- Personalized content transmission
- Recommendations
Users - Applications

- Search Services (Portals, news sites, libraries, museums, companies, etc)

Medical Applications

Cultural Applications
Users - Services

- Multimedia and Semantic Web
- Personal users and content:
  - «I want photos from my vacations in the Greek islands»
  - «I want video-clips from the Web with sports cars»
Conclusions

• Semantic analysis of multimedia is already providing results
• There is a gap between generic technologies and specific applications
• In many cases automatic analysis can enhance existing applications and not generate new ones
• Have to be integrated as part of a complete system or application
• A lot of factors have to be considered: users, interfaces, infrastructure, scenario, business model
Thank you!

CERTH-ITI / Multimedia Knowledge Group
http://mkg.iti.gr