Multimedia Semantics – From MPEG-7 to Web 3.0

Jane Hunter
jane@itee.uq.edu.au
Agenda

• MPEG-7
• MPEG-7 Ontologies
• Semantic Gap Approaches
• Web 2.0
• Web 3.0
  – New forms of multimedia
  – Hybrid classification
  – Weighted classifications
• Conclusions
Multimedia Semantics

• How to find user-relevant multimedia
• Using search terms meaningful to the user
• Content-based search
  – Events – tries, coral bleaching
  – People – “Barak Obama”
  – Objects – astrocytoma, cancerous cells
  – Scenes – church scenes in “In Bruges”

(Focus is not on metadata – format, creator, date
Focus is not on query-by-example)
The Semantic Gap

Human Interaction

Bibliographic:
Creator, Publisher
Date, Published

Semantics:
Objects, Events
People, Places

Structural:
Regions, Segments

Features:
Colours, Textures
Shapes, Motion,
Pitch, Tempo,
Volume

Automatic Extraction

Semantic search and retrieval e.g.
“Give me high porosity fuel cell images”

Heterogeneous Complex Multimedia Data

video
images
graphs
SMIL files
web pages

semantic gap
MPEG-7 Standard

• Multimedia Content Description Interface
• ISO/IEC standard by MPEG
• Objectives are to:
  – Standardize content-based descriptions for audiovisual information
  – Address a wide range of applications
  – Describe images, audio (speech, music), video, graphics, 3D models, compositions
  – Be independent of storage, coding, display, transmission, medium, or technology
MPEG-7 Components

Describers:
(Syntax & semantics of feature representation)

Description Schemes
(Relationships between Ds and DSs)

Instantiation

Definition

Tags

Description Definition Language

⇒ extension

Structuring

Encoding & Delivery

Video-on-Demand
TV-Anytime
Multimedia DSs

Content organization

Collection & Classification

Analytic Model

Navigation & Access

Summary

Variation

User preferences

User

Creation & production

Content management

Content description

Media

Structural aspects

Conceptual aspects

Content Usage

Datatype & structures

Basic elements

Link & media localization

Basic DSs
### Visual and Audio Descriptors

<table>
<thead>
<tr>
<th>Feature</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>DominantColor</td>
</tr>
<tr>
<td></td>
<td>ScalableColor</td>
</tr>
<tr>
<td></td>
<td>ColorLayout</td>
</tr>
<tr>
<td></td>
<td>ColorStructure</td>
</tr>
<tr>
<td></td>
<td>GoFGoPColor</td>
</tr>
<tr>
<td>Texture</td>
<td>Homogeneous Texture</td>
</tr>
<tr>
<td></td>
<td>TextureBrowsing</td>
</tr>
<tr>
<td>Shape</td>
<td>EdgeHistogram</td>
</tr>
<tr>
<td></td>
<td>RegionShape</td>
</tr>
<tr>
<td></td>
<td>ContourShape</td>
</tr>
<tr>
<td></td>
<td>Shape3D</td>
</tr>
<tr>
<td>Motion</td>
<td>CameraMotion</td>
</tr>
<tr>
<td></td>
<td>MotionTrajectory</td>
</tr>
<tr>
<td></td>
<td>ParametricMotion</td>
</tr>
<tr>
<td></td>
<td>MotionActivity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silence</td>
<td>Silence</td>
</tr>
<tr>
<td>Timbre</td>
<td>InstrumentTimbre</td>
</tr>
<tr>
<td></td>
<td>HarmonicInstrumentTimbre</td>
</tr>
<tr>
<td></td>
<td>PercussiveInstrumentTimbre</td>
</tr>
<tr>
<td>Speech</td>
<td>Phoneme</td>
</tr>
<tr>
<td></td>
<td>Articulation</td>
</tr>
<tr>
<td></td>
<td>Language</td>
</tr>
<tr>
<td>Musical</td>
<td>MelodicContour, Rhythm</td>
</tr>
<tr>
<td>Structure</td>
<td></td>
</tr>
<tr>
<td>SoundEffects</td>
<td>Reverberation, Contour, Noise, Pitch</td>
</tr>
</tbody>
</table>

---

*Image: University of Queensland, Australia*
MPEG-7 Ontology

Necessary in order to:
• Bridge the semantic gap
• Enhance discovery
• Add multimedia to the semantic web
• Enable machine reasoning both within and across multimedia content
• Enhance semantic interoperability
• Facilitate integration of multimedia content through common understandings
Colour Descriptors

- Video Segment
- Still Region
- Moving Region
- Visual descriptor
- rdfs:Resource
- Color
  - Dominant Color
  - Scalable Color
  - Color Layout
  - Color Structure
  - GoFGoP Color

subPropertyOf

subClassOf
Semantic Inferencing Architecture

Define RuleML, SWRL rules

Ontologies – MPEG-7 + Domain-specific
Rules-By-Example
Richer User-Centred Queries

How does the mean catalyst size in electrodes of width < 20 microns effect electrode conductivity?
Upper Ontologies

- ABC Ontology
- MPEG-7 Ontology
- CIDOC CRM – Museum content
- FUSION – Fuel Cell images
- OBOE – Environmental images/video

PhysicalEntity
AbstractEntity
InformationObject
Event
Place
Time
Agent

SAMT 2008
Related Approaches

• Low level feature extraction – still room for improvement
• Statistical classification methods
  – Cluster multimedia objects based on similarity
• Machine Learning/Black box approaches
  – Bayesian, Probabilistic models
  – Neural networks
  – Genetic algorithms
  – Decision tree learning
Web 2.0, eScience -> New Content

- 3D – cultural, biomedical, nano-structural
- 4D – Scientific Visualizations
- Compound Objects (Provenance, Learning Objects)
- Access Grid, EVO Sessions
- Multi-sensor networks – webcams, sensors, remote sensing satellite imagery
- YouTube, PodCasts
- Video/Audio Blogs
- FaceBook, MySpace
Ocean_Observation#1

Weather Data supplied by the Heron Island Research Station, QLD.
Date: 02-01-2000
Location: 151 9'00; -23 44'22.

<table>
<thead>
<tr>
<th>Tide (m)</th>
<th>Temp. Max (deg C)</th>
<th>Temp. Min (deg C)</th>
<th>Temp. Observed (deg C)</th>
<th>Wind Direction (Degrees)</th>
<th>Wind Speed (Knots)</th>
<th>State of Sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>31.5</td>
<td>26</td>
<td>27</td>
<td>15</td>
<td>5</td>
<td>calm water and small waves</td>
</tr>
</tbody>
</table>

Directions: To here - From here
Indexing and Search for 3D Cultural Artefacts
AccessGrid Session Recording & Playback

The image shows a video recording interface with thumbnails of different locations, including the University of Queensland (Centre and Rear). Below the thumbnails, there is an XML file snippet that details a session titled "DART Revival Session" with participants Jane Hunter and Russell Crow. The session took place on 2006-08-13T07:02:08 and lasted for 00:00:37.

XML Snippet:
```xml
<session>
  <title>DART Revival Session</title>
  <date>2006-08-13T07:02:08</date>
  <duration>00:00:37</duration>
  <identifier>http://metadatall.net/archive/0813.avi</identifier>
  <language>en</language>
  <agenda>http://metadatall.net/archivie/0813_agenda.html</agenda>
  <minutes>http://metadatall.net/archivie/0813_minutes.html</minutes>
  <participants>
    <participant>
      <name>Jane Hunter</name>
      <email>jane@iteo.uq.edu.au</email>
      <location>University of Queensland</location>
    </participant>
    <participant>
      <name>Russell Crow</name>
      <email>woof@usq.edu.au</email>
      <location>University of Southern Queensland</location>
    </participant>
  </participants>
</session>
```
Image Upload:  W cold fracture tip.tif

Please Complete the Required Metadata:

SAMPLE:  Tip Fracture
EXPERIMENT REFERENCE:  i2c
PROJECT AFFILIATION:  NIDB
INSTRUMENT:  UQJSM6460LA

✓ Metadata Value Changed to "UQJSM6460LA"

OK
Web 2.0

Social networks/community participation
- define information environments
- drive the technologies
- rank information resources
- define own tags (folksonomies)

Distributed Multimedia Collections
Metadata vs Tags

Tags
- Topical, relevant, adaptive, light-weight
- Cheap – generated by masses
- Inconsistent, inaccurate, won’t scale, flat structure
- Systems don’t interoperate (TagCommons)

Authoritative metadata
- Complex, fixed, hierarchical structures
- Don’t evolve, irrelevant, anachronistic terms
- Very expensive
- High quality
Annotations vs Tags

- **Tags**
  - Light-weight, unstructured, organic, folksonomies
- **Annotations**
  - Structured, schemas, controlled vocabs, ontologies
  - creator, date, type, content/description, context
- **Ontology-based tags in RDF enable:**
  - machine-processing of annotations
  - resources to become part of semantic web
  - enhanced discovery and reasoning
Collaborative 3D Object Annotation
System Architecture

- **Web Search Interface**
- **Harvested Metadata Store**
  - Periodically Harvested Metadata
  - OAI-PMH
- **Augmented Metadata Store**
- **Community generating annotations**
- **Authenticated Annotation Service**
  - Shibboleth
  - OAI-PMH Annotation Server

Institutional Repositories/Metadata
Tag Clouds

A access Agriculture AIDS all anabolic and Audio Best biomass blog blogs Blue Body book brain cancer car care climate data design Development diving DIY dmr energy ethanol ethics evolution fitness for Free from fuel gardening globalwarming Green Greenhouse growth guide health health care history HIV Home hosting house imaging in insurance internet keyboard law learning Logitech memory method mobile motor mouse new news of online open access panasonic PET pharmaceutical Power products programming quality rental research review RV science scuba side Skin smoking Snowboarding Social sports states Steroids stock supplements technology the tips to treatment tutorials vacations Warming web Wiki wireless
Government Printing House, N/A


Image data provided by Picture Australia
Issues

• Quality Control on community-generated data
• Ontology-directed folksonomies
• How to identify tags to add to ontology
  – Tag convergence over time
• When, where to add them?
Annotating Relationships

• Tag/Annotate a set of objects (or segments)
• Annotate one multimedia object with another multimedia object
  – Audio description of a photo
• Tag/Annotate relationships between objects or segments – label the relationship
• Kickstart greater semantic inferencing
Annotating Relationships

Co-Annotea User Interface
- attach, share, view, edit associations

Students of Film and Media 101

See this association between scenes in the Seven Samurai and Magnificent Seven

Lecturer

Web Video Server

HTTP

Web Browser + Plug-ins

Secure Annotation Server

Annotating Relationships
Annotating Relationships
Toukelau

Interlinear
Translation

s4 1{ Malen amain na ka tur roat.
s5 1{ go apo pan sGot roat, ka fan slut roat,
s6 1{ Ler mo, ka fo pei m{eel roat, a m{eel roat inom
s7 1{ Amer p{iel ki, ap{iel kin itu.
s8 1{ Me apo mer pan tai lop, A pan tai lop, mai, a preg p{iel, tai p{ielki
s9 1{ Jeg nen kin ka fo tur roat, apreg, aton kin rup{ielkei. Lop ru p{ielkei
s10 1{ Innik wou kin na ka fo preg nantampe neu ru p{ielkei.
inom
s11 1{ Amer pei taakokoi lop ru to m{at.
s12 1{ Ka fo mer pei m{e, m{ pin, tenen ru soso ki 'pin'
s13 1{ A mas pin na inom
We’re going to report three crystal structure complexes of human phenethylamine N-methyltransferase (PNMT). One bound with substrate that incorporates a flexible ethanalamine side chain (p-octopamine), a second bound with a semirigid analogue substrate [cis-(1R,2S)-2-amino-1-tetralol] and a third with trans- (1S,2S)-2-amino-1-tetralol, that can be used as a potential substrate for PNMT.
OAI-ORE - Complex Compound Objects

OAI/ORE Named Graphs/Resource Maps:
- Define set of components
- Typed Relationships between components
- Different views of the compound object
- Metadata attached to compound object
- Publish as RDF Named Graph

Identifier
- URI
- DOI
- PURL


PDF
- cites
- is_derived_from

PS

HTML

MP3

hasRepresentation
- View1.html

hasRepresentation

View2.smil

http://openarchives.org/ore
OAI-ORE Compound Objects
Future -> Hybrid Approaches

• Traditional authoritative metadata – high quality, expensive, hierarchical, rigid

• Combine with:
  – automatic low-level feature extraction – colour, texture, shape, pitch
  – social tagging – topical, cheap, flat, non-conformist
  – machine learning/statistical – need corpus

• Assign weightings based on:
  – reliability of source/method
  – social network information - FOAF

SAMT 2008
Web 3.0 – Social Web meets Semantic Web

Photo in LOC

Library of Congress Curator
  VRA Metadata

FlickrCommons
  Community Tags/Tag Cloud

Image processing + rules
  MPEG-7 Ds
    - Colour histogram
    - Region shape
    -> label

Machine-learning, statistical methods
  Automatic Tags

Aggregated description – that includes sources and weightings
Hybrid Architectures

Collection of images

assign tags

Collective Intelligence

e.g., Flickr

SciVee

New photo

Workflow

Preprocessing

New photo

Place (Geotags), Date, Creator

Format, Size, Regions, MPEG-7

Social tags

Semantic tags

Corpus

Structured, weighted semantically rich machine-processable metadata

Machine Learning Method

Tagged Collection

Tagged Collection

SAMT 2008
Weighted Multimedia Metadata

Source

- Manual
- Automatic
- Semi-Automatic

- Machine Learning
- Bayesian
- Neural Network
- Statistical

- Image Processing
- Audio Processing

Weightings

- Bibliographic: Title, Creator, Subject, Genre, Date_created
  - Weight: High

- Formatting: Format, encoding, storage, dimensions, hardware/software
  - Weight: High

- Life History/Provenance: Events, Rights mgt.
  - Weight: Low-Medium – depends on Trust rating

- Semantics: People, place, objects, events, concepts
  - Weight: Medium

- Structural: Regions, Segments
  - Weight: Medium

- Content: Description, Transcript, Colours, Textures, Shapes, Motion, Pitch, Tempo, Volume
  - Weight: Medium
Hybrid Search and Browse Interfaces

- Tag Clouds + Ontology-based querying
- Combined keyword plus descriptor (e.g., shape) searches
- SPARQL + spatio-temporal search interfaces
  - Give me sub-tropical rainforests in S-E Qld above 3000m with reduced rainfall and containing endangered species
Future Research Challenges I

Increasing volumes of community-generated content + community generated metadata/annotations in multiple formats

Challenges –
1. Identify relationships between resources on the Web (is_about)
2. Extract structured information/metadata/tags from unstructured sources (RDFa?)
3. Assign trust or reliability weighting based on the source

- how to crawl blogs and identify discussions around particular videos
- how to extract structured information about the video/image from a blog
Future Research Challenges II

Multimedia search engines that support:

- Ontology-based searches
  - Ontology is dynamic – incorporates folksonomies over time
  - Combined with content-based, QBE, spatio-temporal searches
- Dynamic inferencing – rules change over time
- Identification of your social network and trust metrics
- Ranking of results based on:
  - tags assigned by users from same social networks
  - weightings applied to metadata that reflect reliability, precision
Acknowledgements

• Australia Research Council
• Dept of Innovation Science and Research
• NCRIS 5.16
• Suzanne Little
• Laura Hollink
• Ronald Schroeter
• Imran Khan
• Ron Chernich
• Anna Gerber
Questions?

http://www.itee.uq.edu.au/~eresearch

j.hunter@uq.edu.au