Multimedia Semantic Web

Ramesh Jain
and
Several Collaborators

Contact: jain@ics.uci.edu
Today

• Data and Users
• Semantic Gap
• Multimedia Semantics
• Events bridge the Semantic Gap
• EMME: Experiential Media Management Environment
  – Approach
  – Current System
  – Directions
Data
The Challenge

Data → Connecting → Users
Semantic Gap

The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. A linguistic description is almost always contextual, whereas an image may live by itself.

Content-Based Image Retrieval at the End of the Early Years
Found in: IEEE Transactions on Pattern Analysis and Machine Intelligence
Arnold Smeulders, et. al., December 2000
Semantic Gap exists in text also. Search engines do little to bridge this gap.
Are we bridging the Semantic Gap in Text?

- Semantic Web tools started helping that.
- XML was the first step towards that.
- RDF/Ontology are very useful steps in that direction.
- Lots of usable knowledge: dbpedia, foaf, wordnet, …
- But, much remains to be done.
Lets look at Multimedia Semantics, however.
Events

Life = +

Experience
Recording experiences

• Visual
• Aural
• Tactile
• (not worry about smell and taste for the next few years)
• Text
• Log of activities
Events, Experiences, and Multimedia

• Experiences are associated with Events.
• We experience using our sensors and capture them using multimedia data.

• Isolated Events are good, but
• Things really become interesting when we create an EventWeb. Events are somehow connected and intertwined in more complicated Webs.
Multimedia is exponentially increasing on the Web and in our life.
But, too much of a good thing may lead to problems.
Multimedia Semantics: Approaches

• Traditional Content Analysis: Models, features, and descriptions.
• MPEG 7
• COMM
• LSCOM
Figure 3. General object-recognition system structure. $I$, image-formation process; $M$, world-modeling process; $D$, description process; $U$, understanding process; $R$, model-rendering process.
VIMSYS Data Model

In VLDB 1991

Domain Knowledge

Domain Objects

Domain Events

Image Objects

Image Representation
Models

• A model in science is a physical, mathematical, or logical representation of a system of entities, phenomena, or processes. Basically a model is a simplified abstract view of the complex reality.

• For the scientist, a model is also a way in which the human thought processes can be amplified.

• Models in software allow scientists to leverage computational power to simulate, visualize, manipulate and gain intuition about the entity, phenomenon or process being represented.
Descriptions

The purpose of description is to re-create, invent, or visually present a person, place, event, or action so that the reader may picture that which is being described.
Models bridge the Semantic Gap.
MPEG 7

• Generic Multimedia Content Description Standard offers a comprehensive set of multimedia description tools to create descriptions that enable quality access to content.
• Facilitates exchange and reuse of multimedia content across different application domains.
• Supports a range of abstraction levels, from low-level signal characteristics to high-level semantic information.
• Adopted the XML Schema as the basis for the MPEG-7 DDL.
MPEG 7 Description Tools

**Contact organization**
- Collection and classification
- Models

**Content management**
- Media information
- Creation information
- Usage information

**Content description**
- Spatio-temporal structure
- Audio and visual features
- Semantic structure

**Navigation and access**
- Summaries
- Partitions and decompositions
- Variations

**User interaction**
- User preferences
- Usage history

**Basic elements**
- Data types and structures
- Link and media localization
- Roots and top-level elements

**Schema tools**
- Packages
Segment Relationship Graphs in Video
COMM: Core Ontology on MultiMedia

• Based on both the MPEG-7 standard and the DOLCE foundational ontology.
• Came from Semantic Web community interested in multimedia.
COMM: Design Rationale

• **Approach:**
  – NO 1-to-1 translation from MPEG-7 to OWL/RDF
  – Need for patterns: use DOLCE, a well designed foundational ontology as a modeling basis

• **Design patterns:**
  – Ontology of Information Objects (OIO)
    • Formalization of information exchange
    • Multimedia = complex compound information objects
  – Descriptions and Situations (D&S)
    • Formalization of context
    • Multimedia = contextual interpretation (situation)

• Define **multimedia patterns** that translate MPEG-7 in the DOLCE vocabulary

COMM: Designing a Well-Founded Multimedia Ontology for the Web
By Arndt, Staab, Vacura, Troncy, and Hardman.
LSCOM

• Came from VACE/TRECVID community.
• Uses MPEG-7 for descriptions.
• Target of 1000 concepts.
• Produce OWL export of the relevant Cyc subset for the LSCOM concepts that OWL supports.

Large Scale Concept Ontology for Multimedia: Naphade, Smith, Chang, Hauptman, Curtis; IEEE Multimedia July 2006.
LSCOM (lite) Taxonomy for TrecVid 2006

Program
- Weather
  - Office
  - Court
  - Meeting
  - Studio
  - Outdoor
  - Road
  - Sky
  - Snow
  - Urban
  - Waterscape
  - Mountain
  - Desert
  - Building
  - Vegetation
- Entertainment
- Sports

Location
- Location
  - Office
  - Court
  - Meeting
  - Studio
  - Outdoor
  - Road
  - Sky
  - Snow
  - Urban
  - Waterscape
  - Mountain
  - Desert
  - Building
  - Vegetation

People
- People
  - Crowd
  - Face
  - Person
  - Roles
  - Government leader
  - Corporate leader
  - Police
  - Military
  - Prisoner

Objects
- Objects
  - Flag
  - Animal
  - Computer
  - Vehicle
  - Airplane
  - Car
  - Boat/ship
  - Bus
  - Truck

Activities and events
- Activities and events
  - People related
    - Walk/Run
    - March
    - Events
    - Explosion
    - Natural disaster

Graphics
- Graphics
  - Maps
  - Charts
Segmentation, Tagging, Annotations
These are all good steps, but

- Are we bridging the semantic gap?

Or

- Are we just refining our techniques on both sides of the semantic gap?
Current Popular approaches:

• Semantic Web tools (Ontologies, RDF, XML) help in creating relationships among ‘symbolic data’.

• Concept detection commonly use machine learning and other media processing to deal with signal data.
Current Approaches

• Semantic Web emphasizes explicit representation of semantics so humans can understand it and machines can use it.

  *Effective at high level.*

• Machine learning analyzes data and builds models for interpreting it. The models built are implicit and very data dependent.

  *Effective at low level.*
These approaches are good and have been effective.

But, can they bridge the semantic gap?
What is a Dog?

Can we create a model of a dog that can help in recognizing it in photos?
Let's revisit the problem of Multimedia Semantics.
• Contenxt = Content + Context

• Context is as powerful, possibly more, as content in understanding audio-visual information
We need *models* that

1. Capture the Web of Symbolic Information.
2. Represent content/signal characteristics using features and other signal characteristics.
3. Represent context and knowledge and their use in selection of appropriate processing to match symbolic and signal characteristics to cross the Gap.
Multimedia Information

Images
Sequences

Audio

Mail

Text

Photos
Indexing audio, images, and video has been extremely difficult.
Technology tamed knowledge in text because in text the experiential data (i.e. speech) is converted to symbols by humans.
Semantics

Multimedia Data
Bridge: Unified Indexing

Images Sequences
Audio
Mail
Text
Photos

Index
Index
Index
Index
Index

Events
Transformations

Users

Semantic Gap

Data

Text, Images, Audio, Video, Tactile...

Alphanumeric, Pixel, Characters

Bits and Bytes

ISWC 2008
Objects and Event

- Objects and Events are strongly related and support each other.
- DOLCE and other approaches recognize this too:
  - Endurants
  - Perdurants
- Object oriented approaches are good for dealing with STATIC situations.
- Events are good for dealing with dynamic situations and relationships.
- Events offer a strong model to develop insights in many applications.
Event Representation

- Documenting Media/ Sensor Data (Different Modalities, Locations And Granularities)
- Media /Sensor Data Metadata (Format, Size, Encoding Feature)
- Physical Time (Time Stamp Interval Frames)
- Logical Time (Temporal Domain Concepts)
- Relative Time (Temporal Relationships to Other Events)
- Relative Location (Spatial Relationships to Other Events)
- Logical Location (Spatial Domain Concept)
- Physical Location (GPS position, Geographic region, Frame Region)
- Event Causality (Chain of Causing Events)
- Involved Actors (Domain Concepts, Attribute Value Tags)
- Involved Entities (Domain Concepts, Attribute Value Tags)
- Event Type
- Composite Event Structure (Substructure)
Events
1-dimensional Space

Time
EventWeb
1-dimensional Space
Multimedia Storytelling

• Collect information about events
  – Select relevant events
  – For each event, select appropriate information
  – In right media

• Stories are sequence of coherent events.
  – Stories/Novels
  – Drama
  – Movies

Present right event information using right media in right order.
The universe is made of stories, not atoms.

- Muriel Rukeyser
Multimedia Storytelling
1-dimensional Space

Time

Photo

Text

Video
Experiential Media Management Environment

• Event-based
• Should be able to deal with ‘multimedia’
  – Photos
  – Audio
  – Video
  – Text
  – Information and data
  – …
• Searching based on events and media.
• Storytelling

First Photos then other media.
Events: Ontological Modeling

- We are creating an Upper Ontology for events
- We are creating an XML sublanguage to specify event instances
  - Constrained by the Ontology
  - Allows events to contain media properties
- We are developing a framework for using ideas from ontological approaches for modeling as well as description.
Composite Events

- Structuring the real world events.
- Include as many complex relationships as required for recognition of events.
- Incorporate ontologies.
- Formalizing event predicates with Event-Owl.
Temporal Relations (Allen)
Spatial Relations (RCC8)
Spatiotemporal Relations
Wedding Event Ontology - 2
Modern Cameras

- Are more than ‘Camera Obscura’: They capture an event.
- Many sensors capture scene context and store it along with intensity values.
- EXIF data is all metadata related to the Event.

**Exposure Time**
**Aperture Diameter**
**Flash**
**Metering Mode**
**ISO Ratings**
**Focal Length**

**Time**
**Location (soon)**
**Face**
Photos can be tagged using only EXIF

We will also use content features and LSCOM Concepts – will soon start using them.

<table>
<thead>
<tr>
<th>Images</th>
<th>Classes Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Image 1" /></td>
<td>Outdoor Night, People in Restaurants, Theater, Stage Show, Talk By Speaker, Portrait at Night, Public Indoor Places</td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Image 2" /></td>
<td>Daily Life Indoors, Furniture, Party at Day, View of Rooms, Group Photo, People in front of Building</td>
</tr>
<tr>
<td><img src="image3.jpg" alt="Image 3" /></td>
<td>Landscape/ Nature, Single person Outdoors, Daily Life Outdoors, Sky, Poster/Whiteboard, Bird’s Eye View, Sunset, Ocean/Lake, Silhouette</td>
</tr>
<tr>
<td><img src="image4.jpg" alt="Image 4" /></td>
<td>Fireworks, Moonlit Scene, Night Illumination, Group of People Outdoor Night, Sports, Scenery</td>
</tr>
<tr>
<td><img src="image5.jpg" alt="Image 5" /></td>
<td>Daily Life Outdoors, Poster/Whiteboard, Sunset, Street in City, Public Places Outdoors, Outdoor Parties, Outdoor Decorations/Sculpture, Architecture, People in front of Buildings</td>
</tr>
<tr>
<td><img src="image6.jpg" alt="Image 6" /></td>
<td>Street, Public Places Outdoors, Sky, outdoor decorations, Architecture, Theater, People in Public Places</td>
</tr>
<tr>
<td><img src="image7.jpg" alt="Image 7" /></td>
<td>Daily Life Indoors, Signboard, View of Rooms, Slides/displays, Group of People, People in Public Outdoor Places</td>
</tr>
</tbody>
</table>
EMME Event Cycle

Event Presentation/Navigation

Atomic Event Entry

Event Base

Multimedia Events

EXIF Features Tags/Context

Composite Events

Atomic Events

Event Grouping, Linking, Assimilation
Using Context/Models to Build the EventWeb

- *Folder structure*
- Calendar
- Social Network
- EXIF Data
- *Event Ontology*
- Personal annotations

- Photostream Segmentation
- Event Detection from photos
Photo Stream Segmentation

Definition: given a photo stream \( P = \{p_i\} \)
EMME Event Cycle

Event Presentation/Navigation

Atomic Event Entry

Event Grouping, Linking, Assimilation

Minimize Manual Work
**EMME Event Cycle**

- **Event Presentation/Navigation**
  - Explore
  - Search
  - Composite Events
  - User Annotations
  - Event Ontology
  - Event Grouping, Linking, Assimilation

- **Atomic Event Entry**
  - Multimedia Events
  - EXIF
  - Features
  - Tags/Context
  - Atomic Events
  - Photo Stream Segment
Using EMME

• Searching for photo
  – ISWC2008
• Creating Albums:
  – Professional
  – Family
  – Tourism
• Telling stories
  – What did I do in Karlsruhe

• Scenario: In December 2008, I have 20,000 pictures taken in 2008. How do I (semi-automatically) select 25 to send to:
  – My mother
  – The uncle that I hate
  – My personal friend
  – My professional friend
  – ...

Version 0.1 is ready
Welcome to Personal EventWeb

Advanced Search

When?                  Where?

Select Date Range      Country/City /Venue

singapore

Choose Tags

○ Outdoor Day         ○ Indoor
○ No Face             ○ Portraits
○ Outdoor Night       ○ People/Group

Keywords Search

Enter Keyword

Search  Clear all

Search Result        Map View

Events Tree

2008

2008 August

2008 Jan

2008 July

Singapore

2008 Sept

Singapore

2008 April

Beijing WWW 2008
Audio can also be used as experiential data.
Current Status

• Implemented
  – Event model
  – Complete Event Cycle
    • Atomic event ingestion
    • Composite events
    • Navigation/Search environment

• In Progress
  – Event Ontology and its use for composition and recognition
  – Use of ‘concepts’ and other media processing framework.
  – Use of context and various sources of knowledge
Conclusion

• Semantic Multimedia Web requires bridging the Semantic Gap.
• Context and knowledge are important. Many important approaches being developed to address this.
• Equally important are concept detection and other media processing approaches. Many approaches being developed here also.
• We need to bring modeling and description frameworks together to bridge the semantic gap.
Thanks for your time and attention.

For questions: jain@ics.uci.edu
Arguments/Differences

In his third and final debate with Obama this week, McCain cited Joe Wurzelbacher as an example of someone who would be hurt under Obama's tax plan. Wurzelbacher said he intended to buy a plumbing business but believed he would receive a tax increase under Obama's economic plan.

Based on what Wurzelbacher has said publicly about his income, the Obama campaign said he would be eligible for a tax cut, not a tax increase, under the Democrat's proposal.