ECE160
Multimedia

Lecture 15: Spring 2011
Image Recognition and Retrieval
National Research Priorities

- Energy Technologies
  - Fuel efficient engines
  - Replacement energy to fossil fuels
  - Lighter, longer-duration batteries
- Bioengineering/Bioinformatics
  - Genes $\rightarrow$ disease
  - Disease $\rightarrow$ medicine
- Search with Multimedia Content
  - Video surveillance
  - Photo interpretation
Audio Search

- Much video already has subtitles
- Speech recognition
- Then use text search
Multimedia Recognition

- Video surveillance
- Photo interpretation
- Search of photo and video archives
Wide-area Surveillance
Surveillance Scenarios

(1) Intrusion Detection
- Monitor and alert on tailgating, loitering, exit/closed entry, other unauthorized access

(2) Passenger Screening
- Use biometric facial recognition to identify individuals of interest through existing closed circuit TV surveillance

(3) Perimeter Monitoring
- Object tracking and biometric facial recognition to determine vehicles and humans exhibiting suspicious behavior

(4) Unattended Baggage
- Identify unattended baggage (or other objects) left for long periods of time
Surveillance in London

- 45,000+ television cameras in the street
- Images recorded for subsequent analysis
- Sophisticated software to track a suspect from one camera to the next
- Matching of track of suspect to mobile phone records to identify suspect
Multimedia Recognition

- Video surveillance
- Photo interpretation
- Search of photo and video archives
How to Organize these Photos?
Image Organization & Retrieval

- **Keyword-based**
  - Manual labeling is subjective, cumbersome
  - The aliasing problem
- **Content-based**
  - Promising for general semantics: outdoor, landscape, flowers, people, etc.
  - Not enough for *wh*-queries (where, who, when, or what)
EXTENT™ = contEXT + contENT

- **Context**
  - Spatial (location)
  - Temporal
  - Social
  - Others

- **Content**
  - Perceptual features, such as color, texture, and shape
  - Holistic features and local features
Augmented Images

Cameras with high-quality lenses can record location, time, camera parameters, and voice.
Context from Space/Time

- GPS or CellID data
  - Into place names
- Time-based grouping
  - Into meaningful “events”
- From place names and time
  - Time of day
  - Weather
Example of Using Three Pieces of Information

Spatial  Temporal

Content
Maui Sunsets can be obtained from Space/Time
Use content for verification
Use content to transfer metadata
Summarize of the example

- Derived from Context
  - Derive time of the day
  - Obtain weather
  - Verify content
- Use of Content
  - Verify context
  - Transfer context
- Much more...
Are They Similar?
Are They Similar?
Are They Similar?

- In terms of what?
- What is the user’s perception?
Conveying Perception

- Image Databases
  - Conveyed via Examples
- Use a sunset picture (or pictures) to find more sunset images
- Where does the perfect example come from?
Conveying Perception

- Internet Searches
  - Conveyed via Keywords
Keyword Retrieval

- **Pros**
  - A user-friendly paradigm

- **Cons**
  - Annotation is a laborious process
  - Annotation quality can be subpar
  - Annotation can be subjective
  - Synonyms
Conveying Perception

- Image Databases
  - Conveyed via Examples
- Use a sunset picture (or pictures) to find more sunset images
- Where does the perfect example come from?
Are They Similar?
Are They Similar?
Are They Similar?

- In terms of what?
- What is the user’s perception?
Recognition of Content
Recognition
Recognition
clouds vs. waves
Web 1.0 vs. Web 2.0

Diagram showing the difference between Web 1.0 and Web 2.0:
- Web 1.0: User interacts with content primarily in text format.
- Web 2.0: Users interact with content that includes images and videos, creating a more interactive and participatory web experience.
Web 2.0

- Content + Users + Interactions
- Collect rich, organized content
- Attract users & interactions
  - To provide metadata
  - To provide new content
- Improve search quality
  - With new metadata and data
  - Via social-network structure