One picture is worth a 1,000 words, goes the popular saying. However, a picture is worth far fewer than a thousand words if you cannot find it!!!!

Can you always describe the image content using words?

Multimedia Information Retrieval & Filtering

Increasing availability of MM information demands efficient tools for finding images/video:

- 15% and more Web content is multimedia
- Content search tools make a billion $ market

Information Overload or Information Scarcity?

Image Collections are diverse

What are the tasks?

- Large, diverse image collections are common
- Text is sometimes unreliable or unavailable
- Users want "things," not "stuff"

- More than overall image properties
- Traditional object recognition won't work

Two choices:
- Rely on text to identify objects
- Look at regions (objects or parts of objects)
**Image Access at Multiple Levels**

- **Visual Object**
  - “Find more objects like this…”

- **Visual Feature**
  - (texture) [image]
  - (color) [image]
  - (shape) [image]
  - (spatial layout, motion, camera operations)

- **Textual Feature**: keywords, transcripts
  - “Clinton”, “Budget”, “the Pentagon”

- **Semantic**
  - “Picture of Golden Gate Bridge”

**Different Search Methods**

- **Conceptual-level Search**
  - Keyword Search, Keyword Spotting from Speech, Subject Navigation
  - “Find videos of President Clinton discussing budget at a press conference.”

- **Syntactic-level Summary**
  - Video shot segmentation, key frame extraction, storyboard construction
  - “Show me the visual summary of the video clips.”

- **Similarity-based Search**
  - Query images/video by example, sketch, animation

- **Interactive Browsing and Navigation**
  - Random display, relevance feedback, iterative browsing

**Image Meanings**

- **Multiple levels & different dimensions**
  - Image meanings are very rich (multi-level, multi-dimension)
  - User needs are dependent on task, mood, and other factors
  - Levels: Pre-iconographic, Iconographic, Iconology
  - Who, What, Where, When (Generic or Specific)

**Is video retrieval different from image retrieval?**

- **Static scenes**
  - ability to obtain a continuously varying set of views on the scene

- **Dynamic objects**
  - ability to capture the temporal (or dynamic) evolution of phenomena

- **Extra of video over images: context**
  - allows to put dynamic objects in their static context

**What is CBIR?**

- **Allows queries of large image databases based on visual image content** -- color percentages, color layout, textures, shape, etc.

- **In contrast to the text-based approach CBIR retrieves images from a collection by comparing features automatically extracted from the images themselves.**

- **Content-based queries are often combined with text and keyword predicates to get powerful retrieval methods for image and multimedia databases.**

No one-fits-all solutions!!!
Each image added to the collection is analyzed to compute a color histogram which shows the proportion of pixels of each color within the image. At search time, the user can either specify the desired proportion of each color or submit an example image from which a color histogram is calculated. Either way, the matching process then retrieves those images whose color histograms match those of the query most closely.

**Color Histogram**
- Represent each pixel as a quantized color – e.g., 256 colors ranging from red through violet
- Count the number of pixels in each color bin – Produces vector representations
- Compute vector similarity – e.g., normalized inner product

**Color histogram: Example**
- Histogram is a summary of the data summarizing in this case color characteristics
Color histogram: Example

GODZILLA

Color histogram retrieval

How color histogram works?

Different color histogram $\rightarrow$ Different images

Difficulty of color histogram

similar color histogram $\Rightarrow$ similar images

The Problem

Color alone is not always sufficient to characterize an image. How about use color and spatial information in retrieval? How about use texture and shape?

Visual similarity search
Texture matching

- Texture characterizes small-scale regularity
  - Color describes pixels, texture describes regions
- Described by several types of features
  - e.g., smoothness, periodicity, directionality
- Match region size with image characteristics
  - Computed using filter banks, Gabor wavelets, ...
- Perform weighted vector space matching
  - Usually in combination with a color histogram

Shape Retrieval

- A number of features characteristic of object shape are computed for every object identified within each stored image.
- Queries are answered by computing the same set of features for the query image, and retrieving those images whose features most closely match those of the query.
Matching Shapes

- Find correspondences between points on shape
- Estimate transformation
- Measure similarity

Outlier Test Example

Shape retrieval: Example

Iconic-based retrieval

Points-based retrieval

Corner extractors (Harris)
- Drawbacks for natural images
  - visual features are not real corners
  - everything may be a corner (textured regions)
- In general, corners are not spread in the image
Salient points

Demands:
- Salient points should be located in any visually interesting part of the image (at any resolution)
- Salient points should be spread in the image
Wavelet transform
- Frequency localization
- Multi-resolution
- High coefficients correspond to high variations in the image: salient points

Salient points extraction

Salient points

Corners Salient

Salient points

Corners Salient

Salient points

Corners Salient
Salient points

- Global techniques alone yield low precision
  - Color & texture characterize objects, not images
- Segment at color and texture discontinuities
  - Like “flood fill” in Photoshop
- Represent size, shape & orientation of objects
  - e.g., Berkeley’s “Blobworld” uses ellipses
- Represent relative positions of objects
  - e.g., angles between lines joining the centers
- Perform rotation- and scale-invariant matching

Objects in image

- We want to find objects by looking for coherent regions

Flood Fill in Photoshop

More sophisticated techniques are needed

Blobworld: represent image regions
1. Expectation-Maximization (EM)
   - Assign each pixel to a Gaussian in color-texture space
     ⇒ segmentation

2. Model selection: Minimum Description Length
   - Prefer fewer Gaussians if performance is comparable

---

### Querying: Let user see the representation

- Current systems act like black boxes
  - User can’t see what the computer sees
  - It’s not clear how parameters relate to the image

- User should interact with the representation
  - Helps in query formulation
  - Makes results understandable
  - Minimizes disappointment

---

### Blobworld vs. global histograms

- Distinctive objects (tigers, cheetahs, zebras):
  - Blobworld does better

- Distinctive scenes (airplanes):
  - Global histograms do better
    - Adding blob size would help (⇒ “large blue blob”)

- Others:
  - Both do somewhat poorly, but Blobworld has room to grow (shape, etc.)
Other approaches

- Iconic matching [Jacobs et al. '95]
  - Match user sketch to coarse representation
  - Look at overall image layout
- Spatial relationships [Lipson et al. '97]
  - Relationships between superpixels for querying and classification
- Region-based approach [Ma & Manjunath '98]
  - Semi-automatic segmentation

Text-based search engines: google

- Piction (http://www.cedar.buffalo.edu/Piction/)
  - News photos of people
  - Relies heavily on text
  - Captions guide image understanding
- WebSeek (http://www.ctr.columbia.edu/webseek/)
  - Index images and videos on the Web
  - Relies heavily on text
  - Some content-based searching

Text-based search engines: yahoo

Sample image retrieval systems

- QBIC (http://wwwqbic.almaden.ibm.com/)
  - General image/video collections
  - Use both content and textual information
  - User labeling is important
- Blobworld (http://elib.cs.berkeley.edu/photos/blobworld/)
  - Relies exclusively on image content
  - Find images regions corresponding to objects
  - Querying based on regions' color, texture, simple shape
Welcome to WebSeek: search for images on the Web.

Collection: 500,000 Web Images and videos
- Filename and surrounding text determine subject
- Simple content-based techniques
  - Global color histograms
  - Color in localized regions

Collect and process images and videos
- Retrieve image or video
- Store visual features and textual information
- Generate icon (motion icon) to represent image (video) compactly

Classify images based on associated text
- Extract terms
  - URL: http://www.host.com/animals/dogs/poodle.gif
  - Alt text: img src=URL alt="picture of poodle"
  - Hyperlink text: <a href=URL>Sally the poodle</a>
- Map terms to subjects using a semi-automatically constructed key-term dictionary
- Store terms, location in subject taxonomy

Content-based information
- Extract information from image
  - Global color histogram
  - Size, location, relationship of color regions
- Automatically determine image type
  - Color photo, color graphic, gray image, B&W image
- Use content-based information when browsing within category hierarchy

Discussion
- How good is the text-based classification?
- Can we always count on having text?
- How much does the content-based information add in this case?

QBIC: image content plus annotation
- Collections:
  - General collections
  - Video clips
  - Fabric samples
  - etc.
- Retrieval based on color, texture, shape, sketch
- Relies (sometimes) on human annotation of image or objects
Populating the database

- Use any available text (title, subject, caption, etc.)
- User selects and labels important regions
- Store information about image content
- Extract key frames and describe video content

Query based on text and content

- Textual information: title, subject, object labels
- Global image properties: color and texture
- Object properties: color and shape
- Sketch: use the sketch as a template to find matches

Discussion

- QBIC's shape description doesn't capture the essence of objects, just the appearance of one view
- How much does QBIC rely on textual information?

Assignment

- Find on the web images of Westminster Abbey
- Find a close-up image of ET
- Find a video with Clinton
- Find this video !!!!
Assignment

1. How did you find the requested image/video?
2. What are the problems with the existing search engines?
3. What improvements would you suggest?

The answers must be placed on the class bulletin board.