Content-based Image Retrieval (CBIR)

Searching a large database for images that *match* a query:

- What kinds of databases?
- What kinds of queries?
- What constitutes a match?
- How do we make such searches efficient?

Applications

- Art Collections

  e.g. Fine Arts Museum of San Francisco
- Medical Image Databases

  CT, MRI, Ultrasound, The Visible Human
- Scientific Databases

  e.g. Earth Sciences
- General Image Collections for Licensing

  Corbis, Getty Images
- The World Wide Web

What is a query?

- an image you already have
- a rough sketch you draw
- a symbolic description of what you want

  e.g. an image of a man and a woman on a beach, an image of a dad and a daughter on a beach

SYSTEMS

Usage: I: Get Info C: Color Histogram L: Layout T: Texture S: Special Hybrid

Keywords: Previous Next

Query was:
Random

Some Systems You Can Try

Corbis Stock Photography and Pictures

http://www.corbisimages.com

- Corbis sells high-quality images for use in advertising, marketing, illustrating, etc.
- Search is entirely by keywords.
- Human indexers look at each new image and enter keywords.
- A thesaurus constructed from user queries is used.

QBIC

IBM's QBIC (Query by Image Content)

http://www.research.ibm.com/topics/popups/deep/manage/html/qbic.html

- The first commercial system.
- Uses or has-used color percentages, color layout, texture, shape, location, and keywords.

Informedia

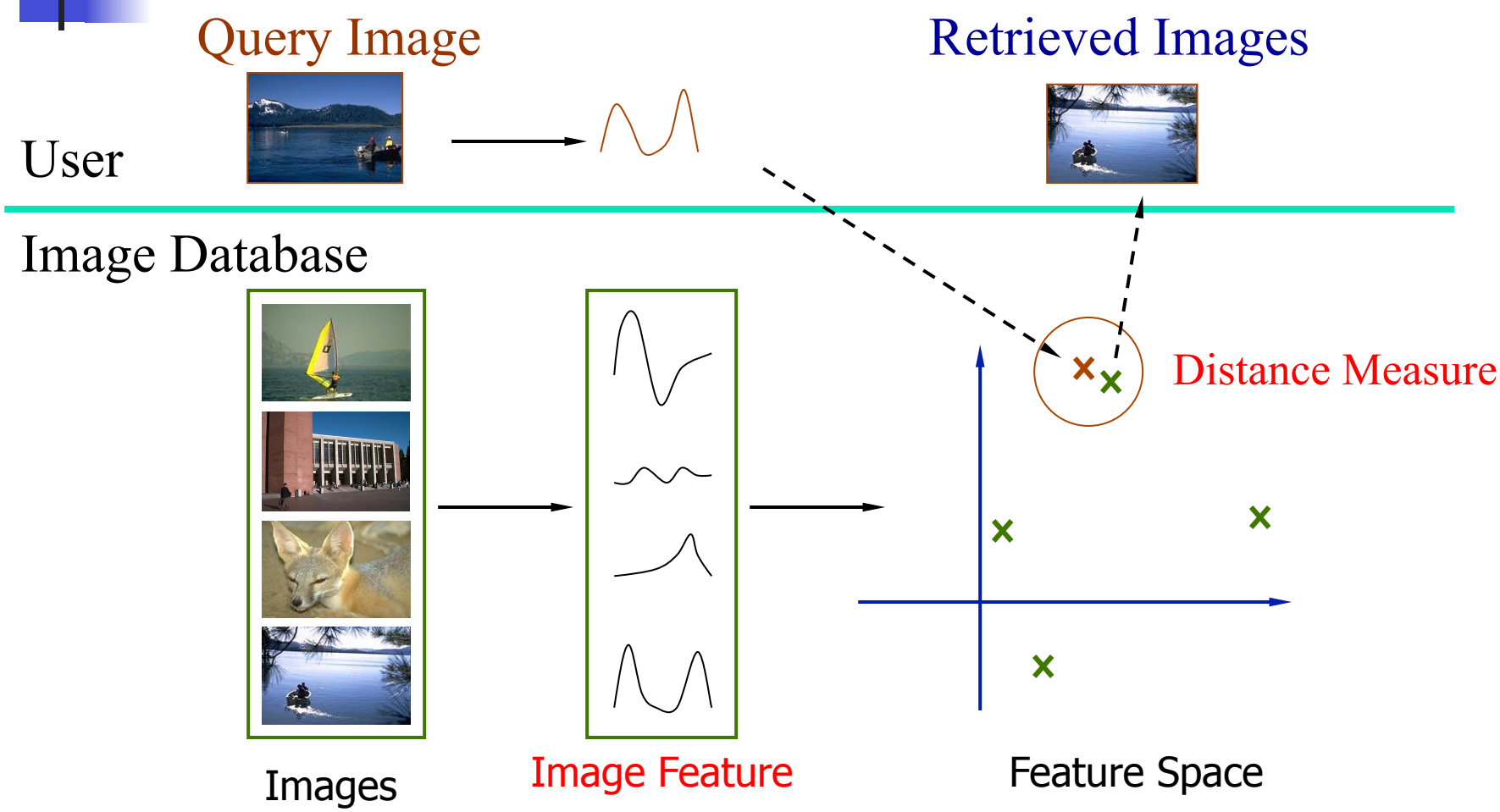
CMU's informedia project

http://www.informedia.cs.cmu.edu/

- Started in 1994 and active till 2007
- The largest academia research project on image retrieval

Google Image search:
https://www.google.com/imghp?gws_rd=ssl

Image Features / Distance Measures

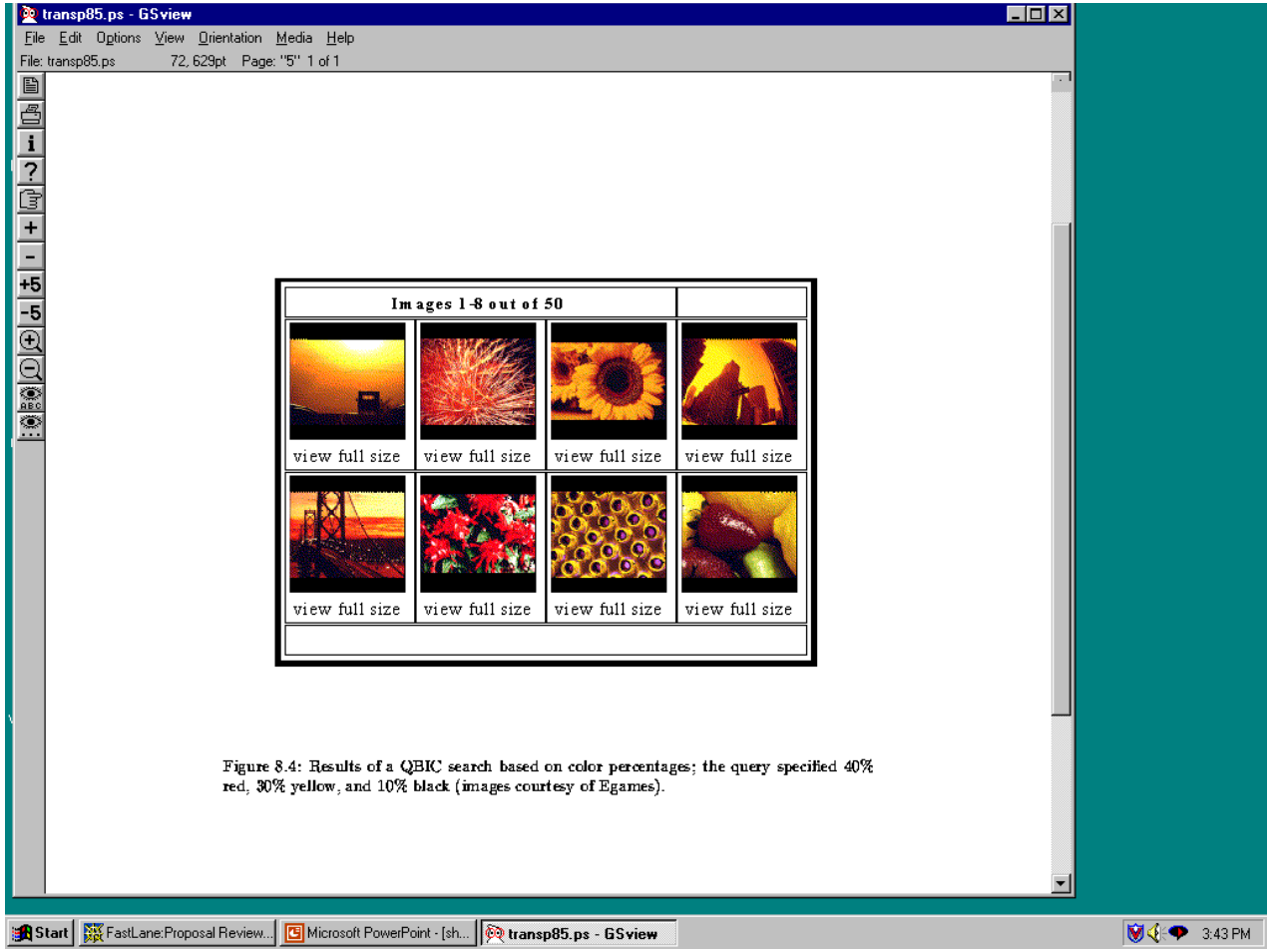


Features

- Color (histograms, gridded layout, wavelets)
- Texture (Laws, Gabor filters, local binary partition)
- Shape (first segment the image, then use statistical or structural shape similarity measures)
- Objects and their relationships

This is the most powerful, but you have to be able to recognize the objects!

Color Histograms



QBIC's Histogram Similarity

The QBIC color histogram distance is:

$$d_{hist}(I,Q) = (h(I) - h(Q))^T \mathbf{A} (h(I) - h(Q))$$

- h(I) is a K-bin histogram of a database image
- h(Q) is a K-bin histogram of the query image
- A is a K x K similarity matrix

Similarity Matrix

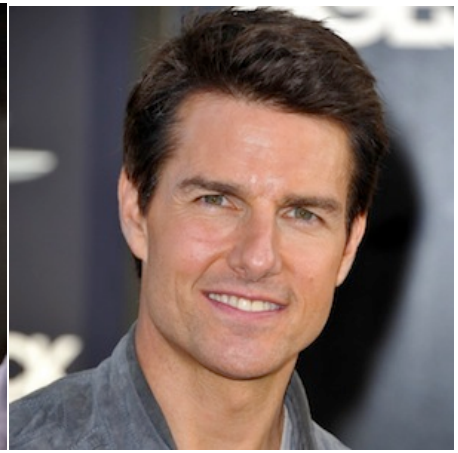
	R	G	B	Y	C	V
R	1	0	0	.5	0	.5
G	0	1	0	.5	.5	0
B	0	0	1		?	
Y				1		
C		?			1	
V						1

How similar is blue to cyan?

Hair Verification

- Humans use hair for recognition[1]
- Face is not always suitable for recognition
  - Occlusions due to sunglasses, scarves, etc.
  - Low image resolution/quality
  - Extreme pose variation



Wright and Sladden, "An own gender bias and the importance of hair in face recognition"

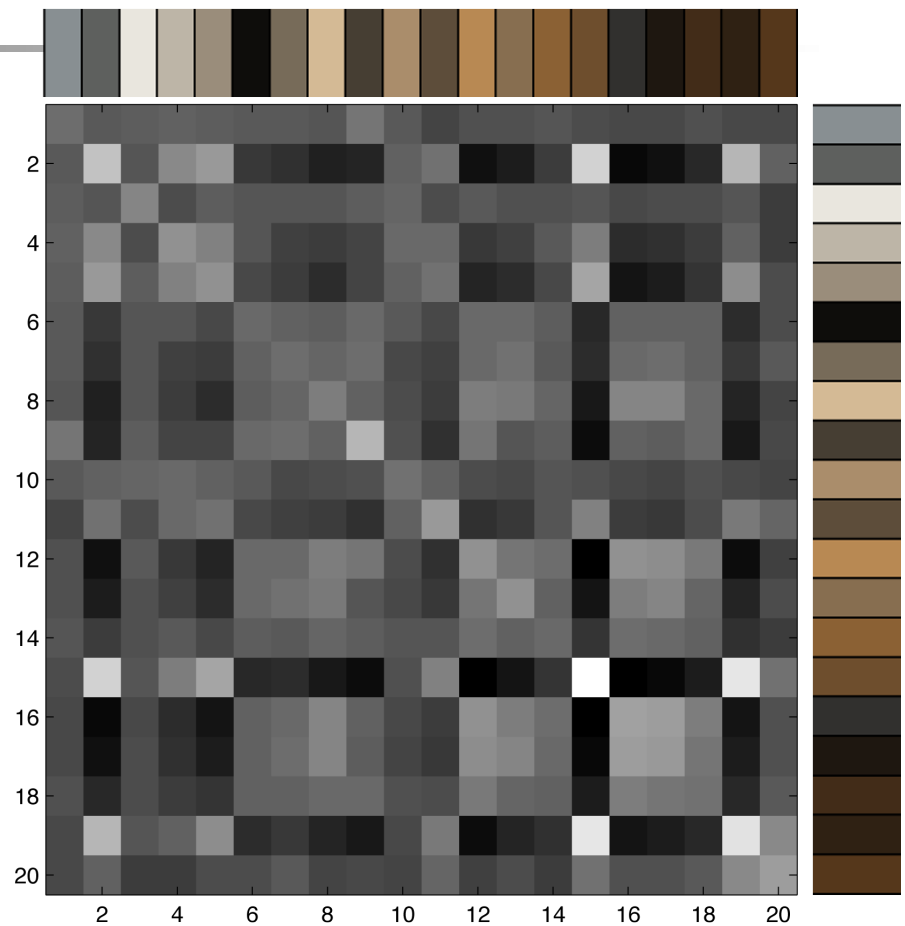
Hair Color

- BoW features
  - Patch sizes: 41, 21
- Cluster in HSV
- Distance Metric Learning

$$\min_{A} \Sigma (f_1 - f_2)\, A\, (f_1 - f_2)^T$$

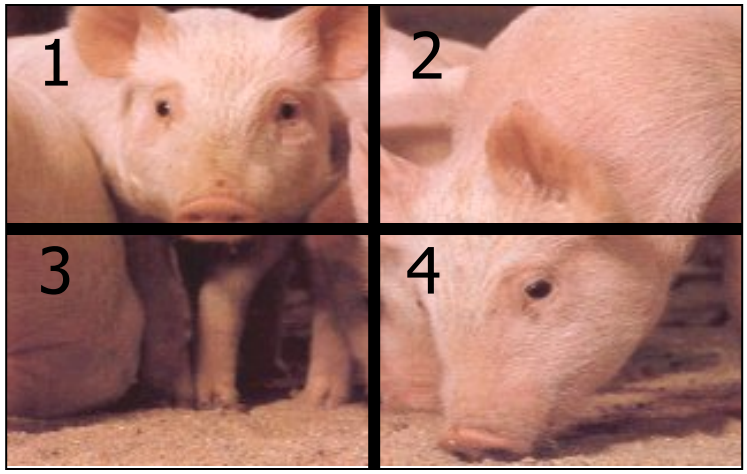
Hist. from image 1

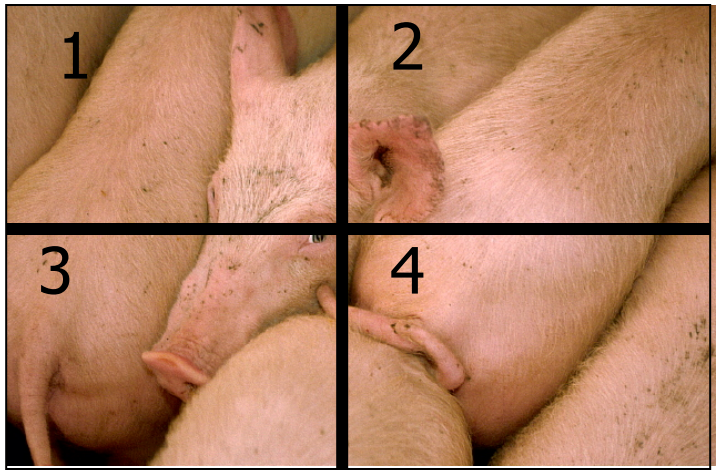
Hist. from image 2



Gridded Color

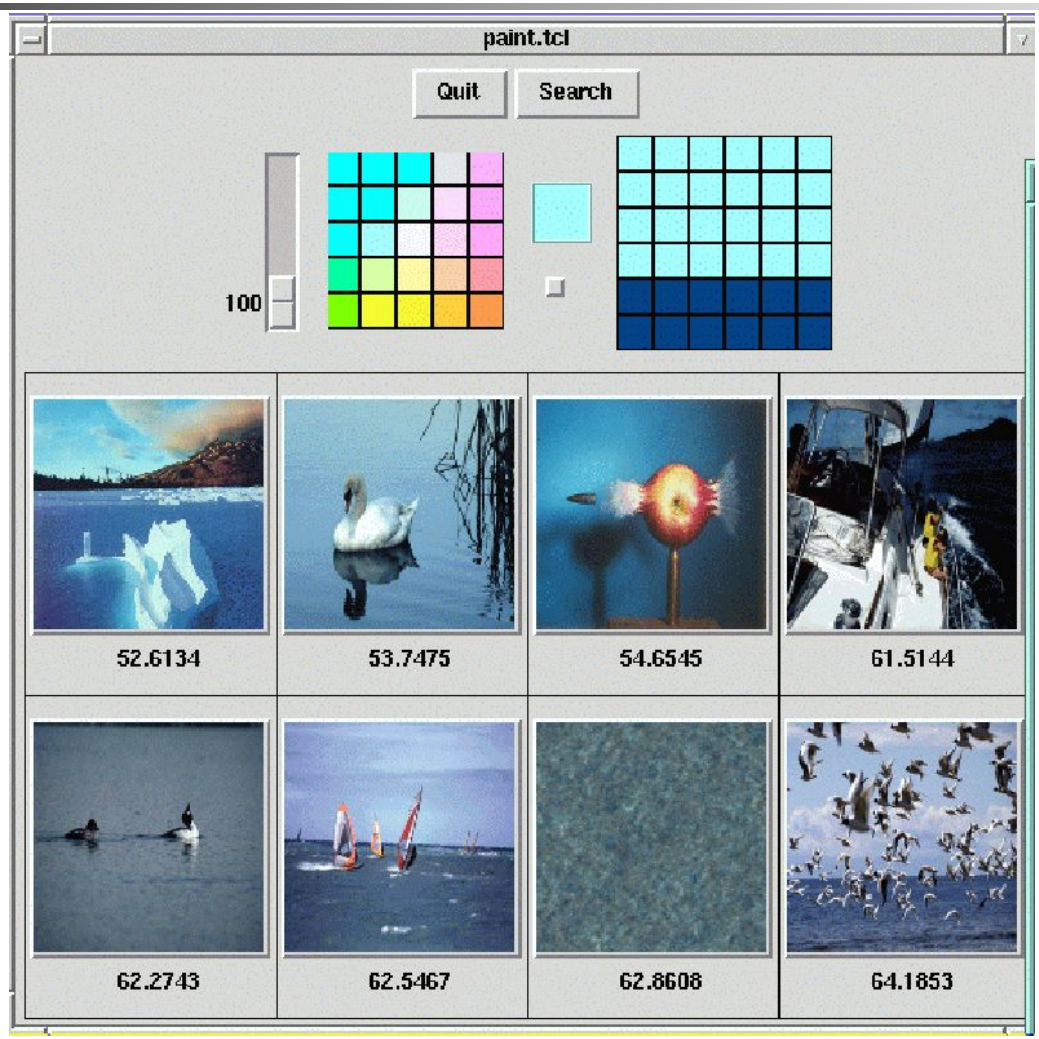
Gridded color distance is the sum of the color distances in each of the corresponding grid squares.



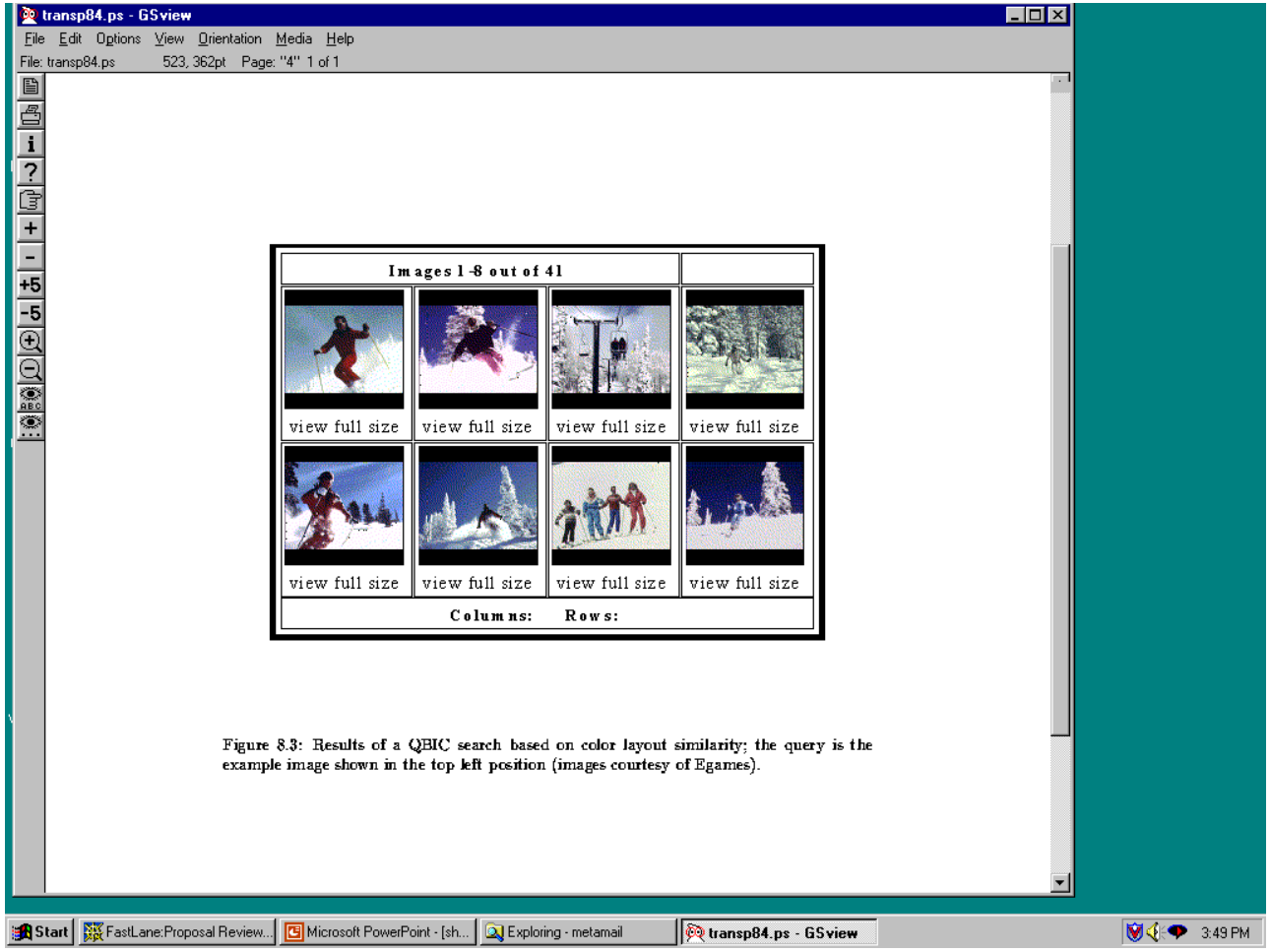


What color distance would you use for a pair of grid squares?

Color Layout



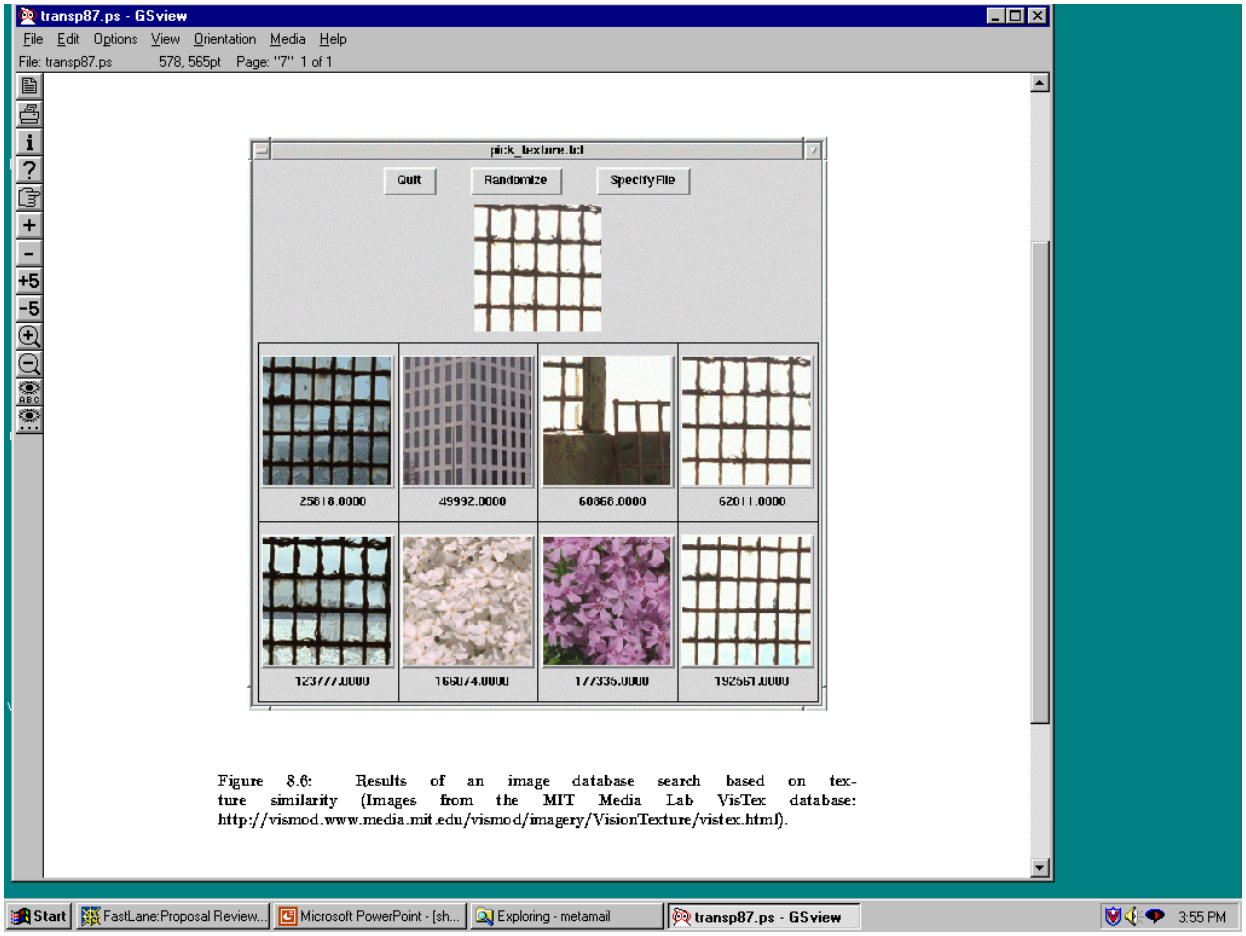
Color Layout (IBM's Gridded Color)



Texture Distances

- Pick and Click (user clicks on a pixel and system retrieves images that have in them a region with similar texture to the region surrounding it.
- Gridded (just like gridded color, but use texture).
- Histogram-based (e.g. compare the LBP histograms).

Laws Texture

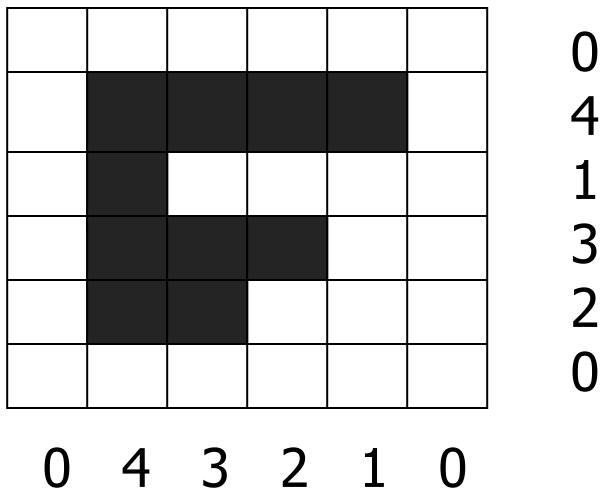


L5 = [+1 +4 6 +4 +1] (Level)
E5 = [-1 -2 0 +2 +1] (Edge)
S5 = [-1 0 2 0 -1] (Spot)
W5 = [-1 +2 0 -2 +1] (Wave)
R5 = [+1 -4 6 -4 +1] (Ripple)

Shape Distances

- Shape goes one step further than color and texture.
- It requires identification of regions to compare.
- There have been many shape similarity measures suggested for pattern recognition that can be used to construct shape distance measures.

Global Shape Properties: Projection Matching

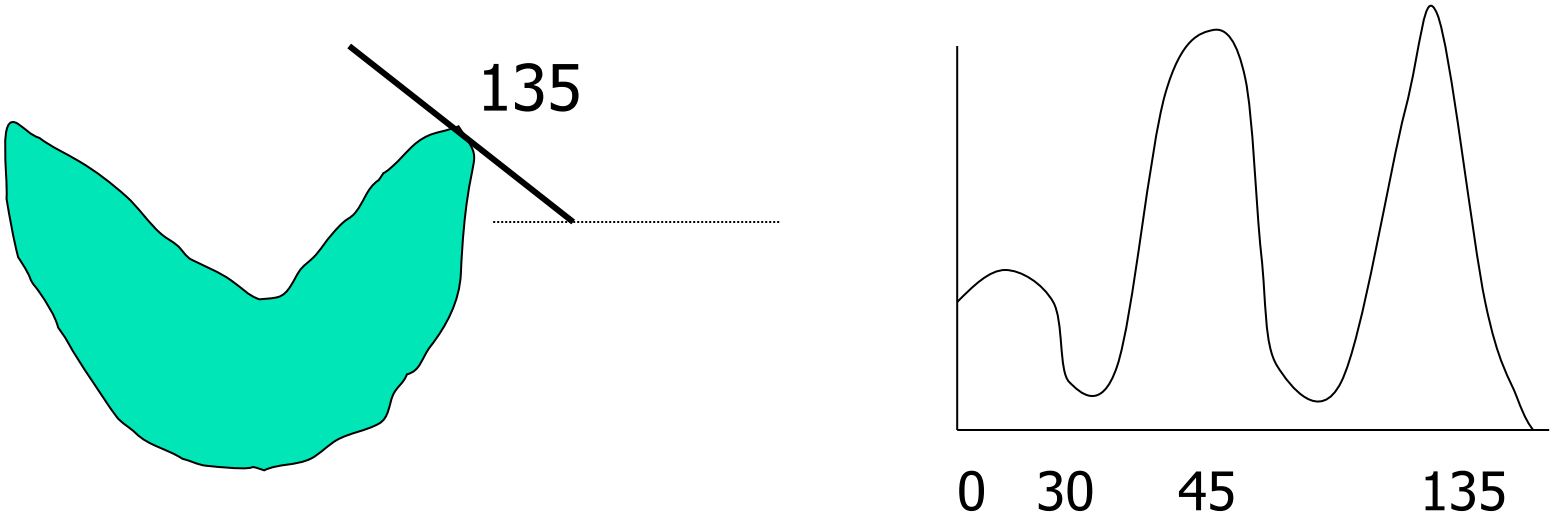


Feature Vector
(0,4,1,3,2,0,0,4,3,2,1,0)

In projection matching, the horizontal and vertical projections form a histogram.

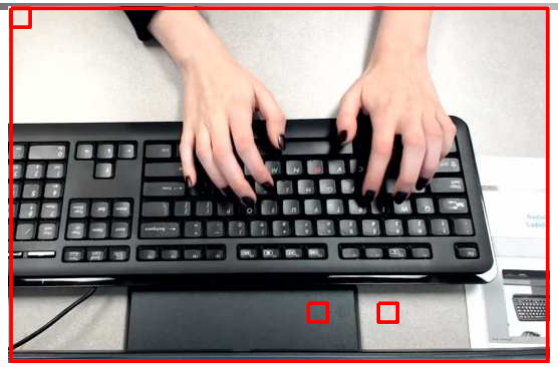
What are the weaknesses of this method? strengths?

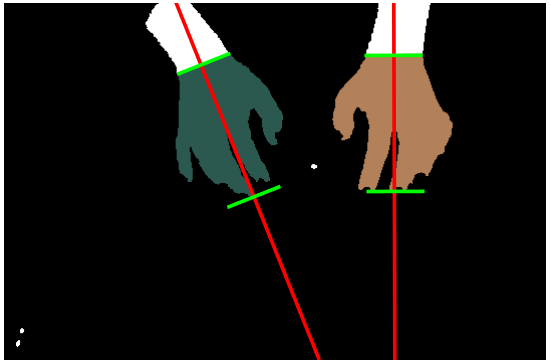
Global Shape Properties: Tangent-Angle Histograms

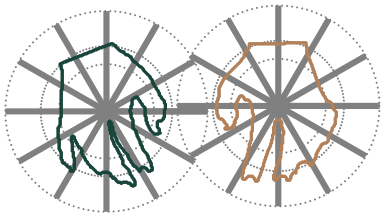


Is this feature invariant to starting point?

Global Shape Properties: Shape Context







Boundary Matching

- Fourier Descriptors

- Elastic Matching

The distance between query shape and image shape has two components:

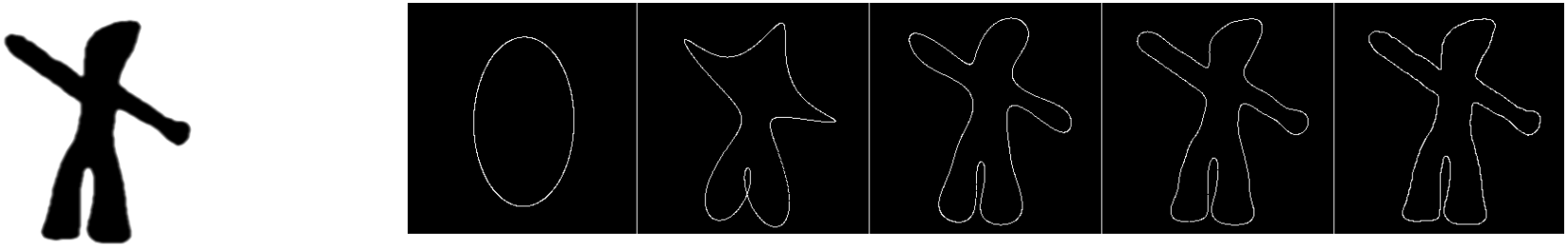
1. energy required to deform the query shape into one that best matches the image shape

2. a measure of how well the deformed query matches the image

Fourier Descriptors

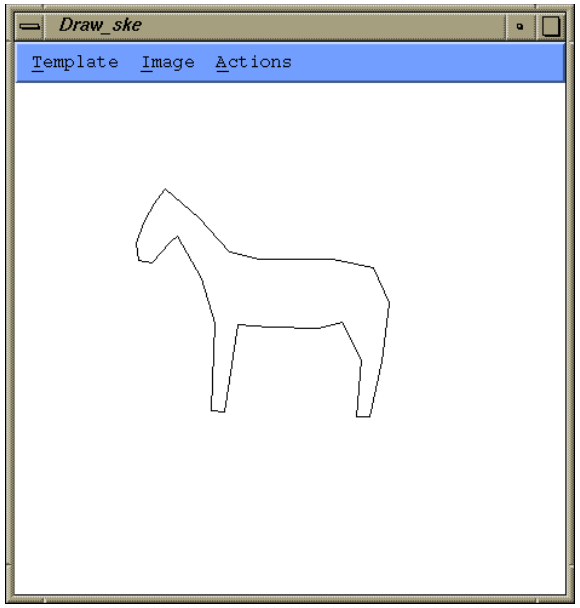
$$z[m] = x[m] + jy[m]$$

$$Z[k] = DFT[z[m]] = \frac{1}{N}\sum_{m=0}^{N-1} z[m]e^{-j2\pi mk/N} \quad (k = 0, \cdots, N-1)$$

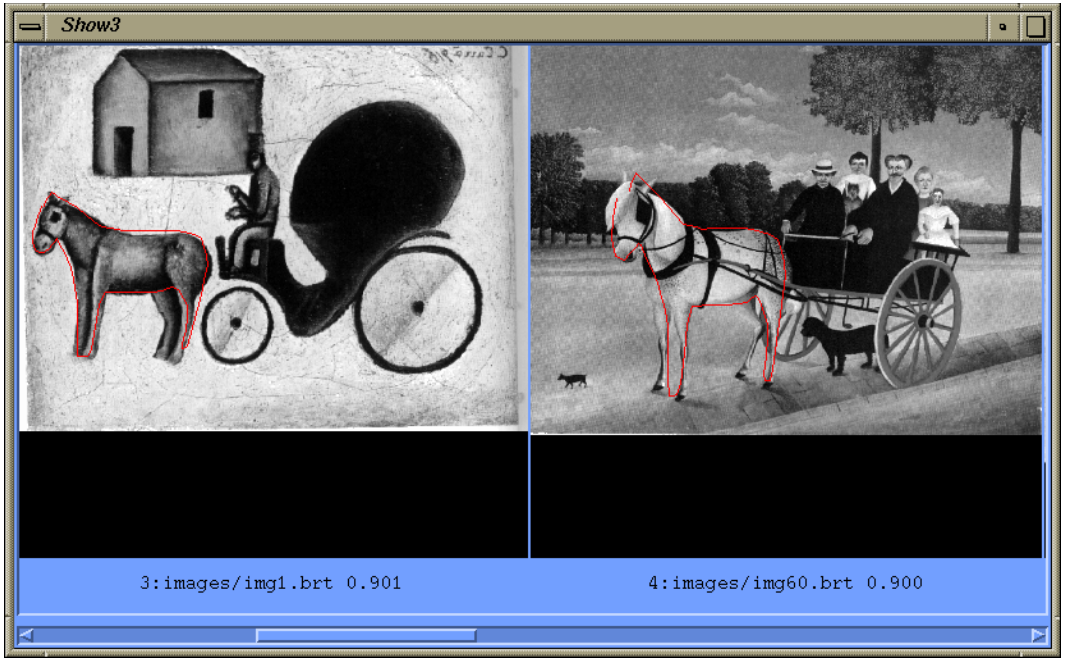


http://fourier.eng.hmc.edu/e161/lectures/fd/node1.html

Del Bimbo Elastic Shape Matching



query



retrieved images

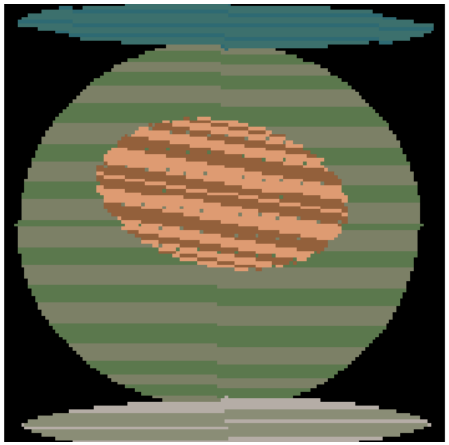
Regions and Relationships

- Segment the image into regions
- Find their properties and interrelationships
- Construct a graph representation with nodes for regions and edges for spatial relationships
- Use graph matching to compare images

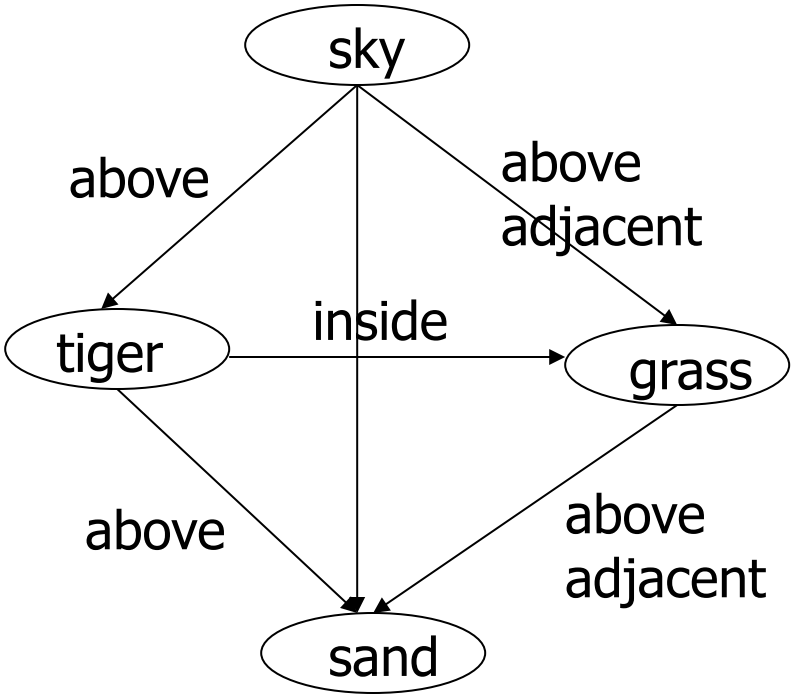
Like what?

Tiger Image as a Graph

image



abstract regions



Object Detection: Rowley's Face Finder

1. convert to gray scale
2. normalize for lighting*
3. histogram equalization
4. apply neural net(s) trained on 16K images

What data is fed to the classifier?

32 x 32 windows in a pyramid structure

* Like first step in Laws algorithm, p. 220

Fleck and Forsyth's Flesh Detector

The "Finding Naked People" Paper

- Convert RGB to HSI
- Use the intensity component to compute a texture map

  texture = med2 (| I - med1(I) |)

  median filters of radii 4 and 6
- If a pixel falls into either of the following ranges, it's a potential skin pixel

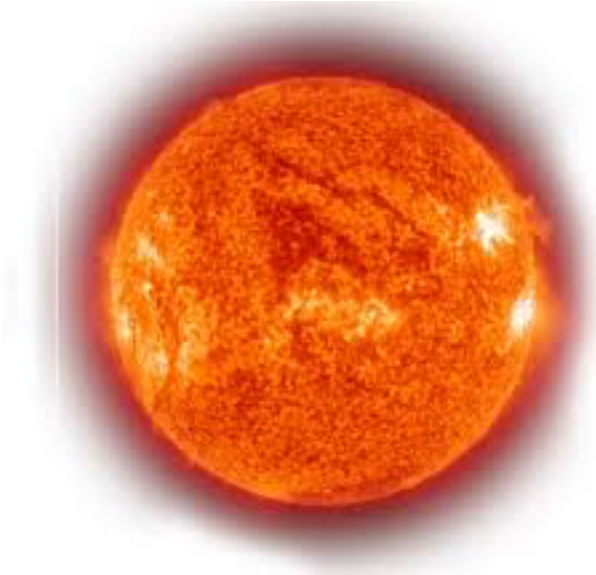
  texture < 5, 110 < hue < 150, 20 < saturation < 60

  texture < 5, 130 < hue < 170, 30 < saturation < 130

Look for LARGE areas that satisfy this to identify pornography.

Weakness of Low-level Features

- Can't capture the high-level concepts



Relevance Feedback

In real interactive CBIR systems, the user should be allowed to interact with the system to "refine" the results of a query until he/she is satisfied.

Relevance feedback work has been done by a number of research groups, e.g.

- The Photobook Project (Media Lab, MIT)
- The Leiden Portrait Retrieval Project
- The MARS Project (Tom Huang's group at Illinois)

Information Retrieval Model*

- An IR model consists of:
  - a document model
  - a query model
  - a model for computing similarity between documents and the queries
- Term (keyword) weighting
- Relevance Feedback

*from Rui, Huang, and Mehrotra's work

Term weighting

- Term weight
  - assigning different weights for different keyword(terms) according their relative importance to the document
- define w_{ik} to be the weight for term t_k ,k=1,2,…,N, in the document *i*
- document *i* can be represented as a weight vector in the term space

$$D_i = \left[w_{i1}; w_{i2}; \ldots; w_{iN} \right]$$

Term weighting

- The query Q also is a weight vector in the term space

$$Q = \lfloor w_{q1}; w_{q2}; \ldots; w_{qN} \rfloor$$

- The similarity between D and Q

$$Sim(D,Q) = \frac{D \cdot Q}{\|D\| \; \|Q\|}$$

Using Relevance Feedback

- The CBIR system should automatically adjust the weights that were given by the user for the relevance of previously retrieved documents

- Most systems use a statistical method for adjusting the weights.

The Idea of Gaussian Normalization

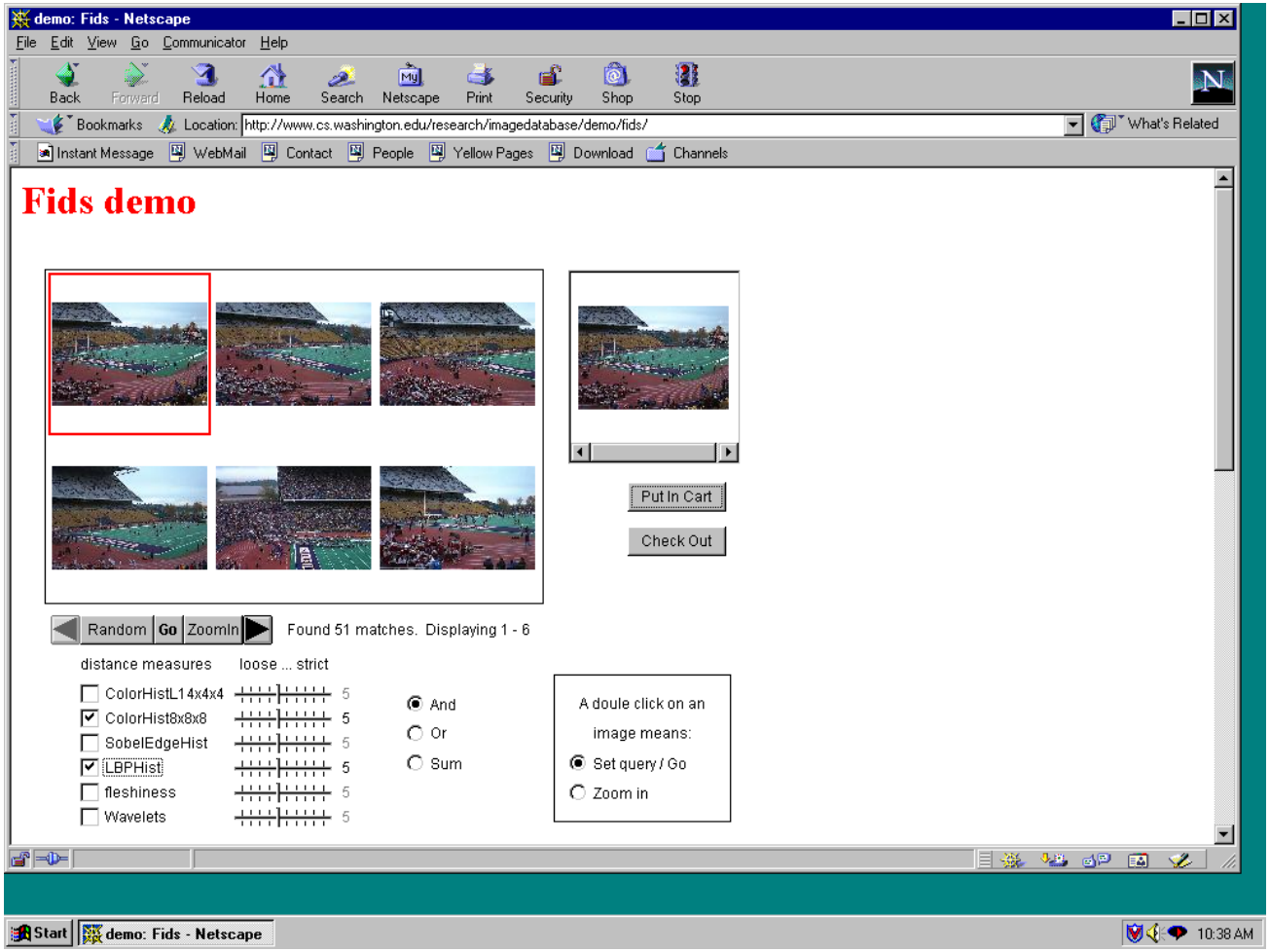
- If all the relevant images have similar values for component *j*
  - the component *j* is relevant to the query
- If all the relevant images have very different values for component *j*
  - the component *j* is not relevant to the query
- the inverse of the standard deviation of the related image sequence is a good measure of the weight for component *j*
- the smaller the variance, the larger the weight

Relevance Feedback Example

http://www.ifp.illinois.edu/~xzhou2/demo/cbir.html

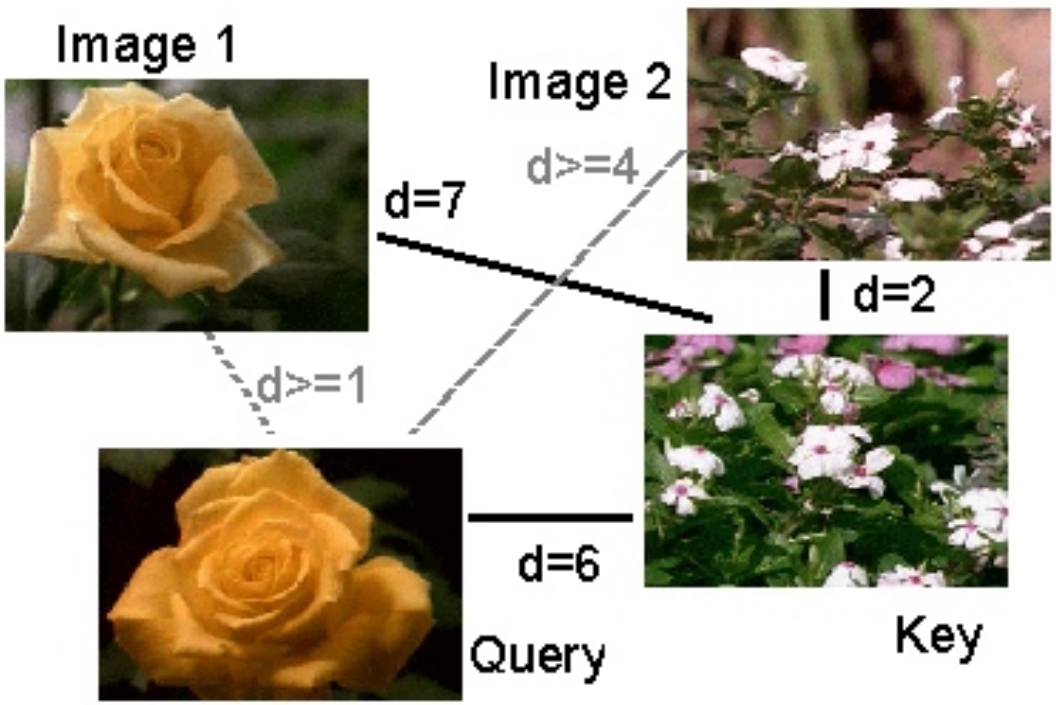
Andy Berman's FIDS System

multiple distance measures
Boolean and linear combinations
efficient indexing using images as keys



Andy Berman's FIDS System:

Use of key images and the triangle inequality for efficient retrieval.



Andy Berman's FIDS System:

Bare-Bones Triangle Inequality Algorithm

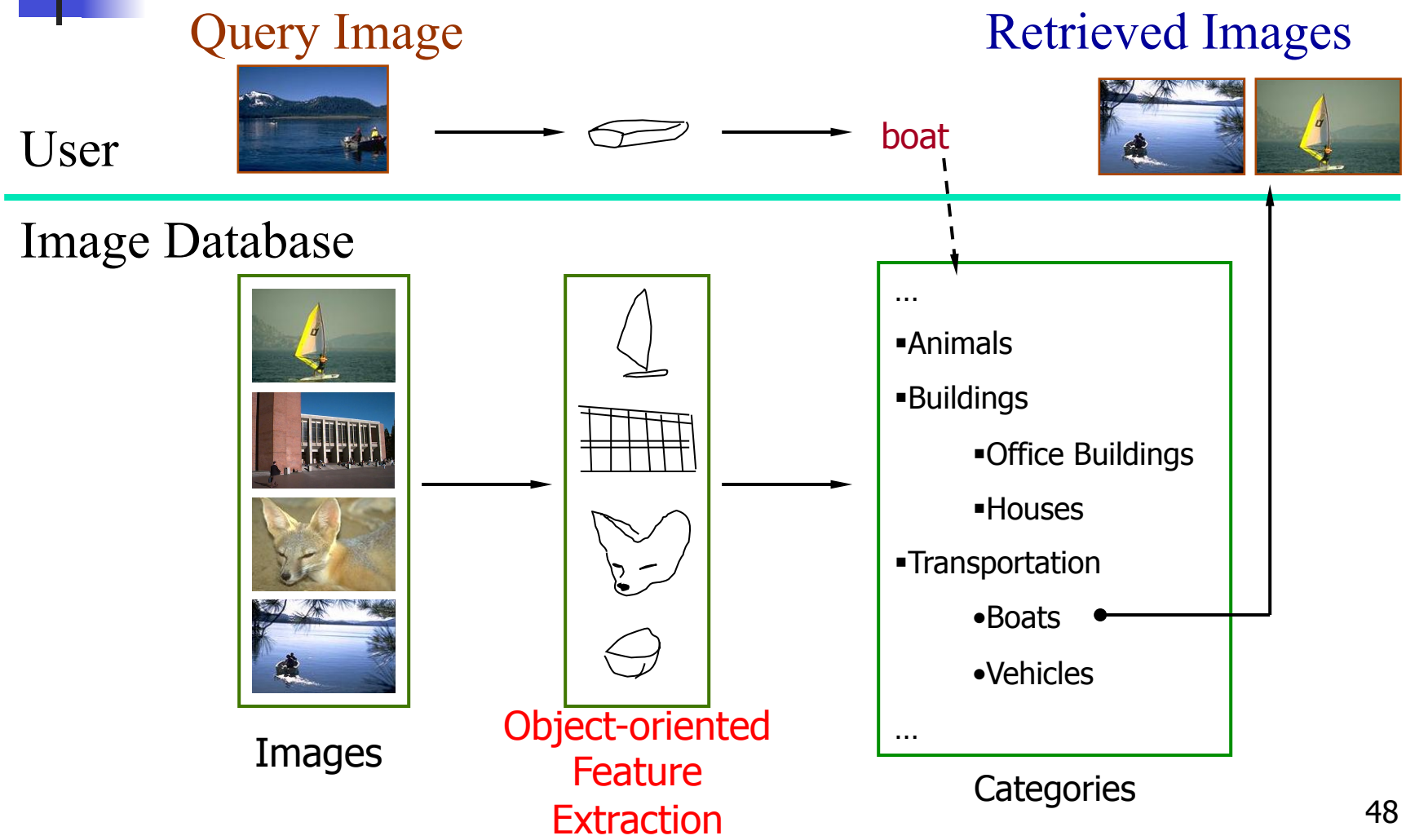
Offline

1. Choose a small set of key images
2. Store distances from database images to keys

Online (given query Q)

1. Compute the distance from Q to each key
2. Obtain lower bounds on distances to database images
3. Threshold or return all images in order of lower bounds

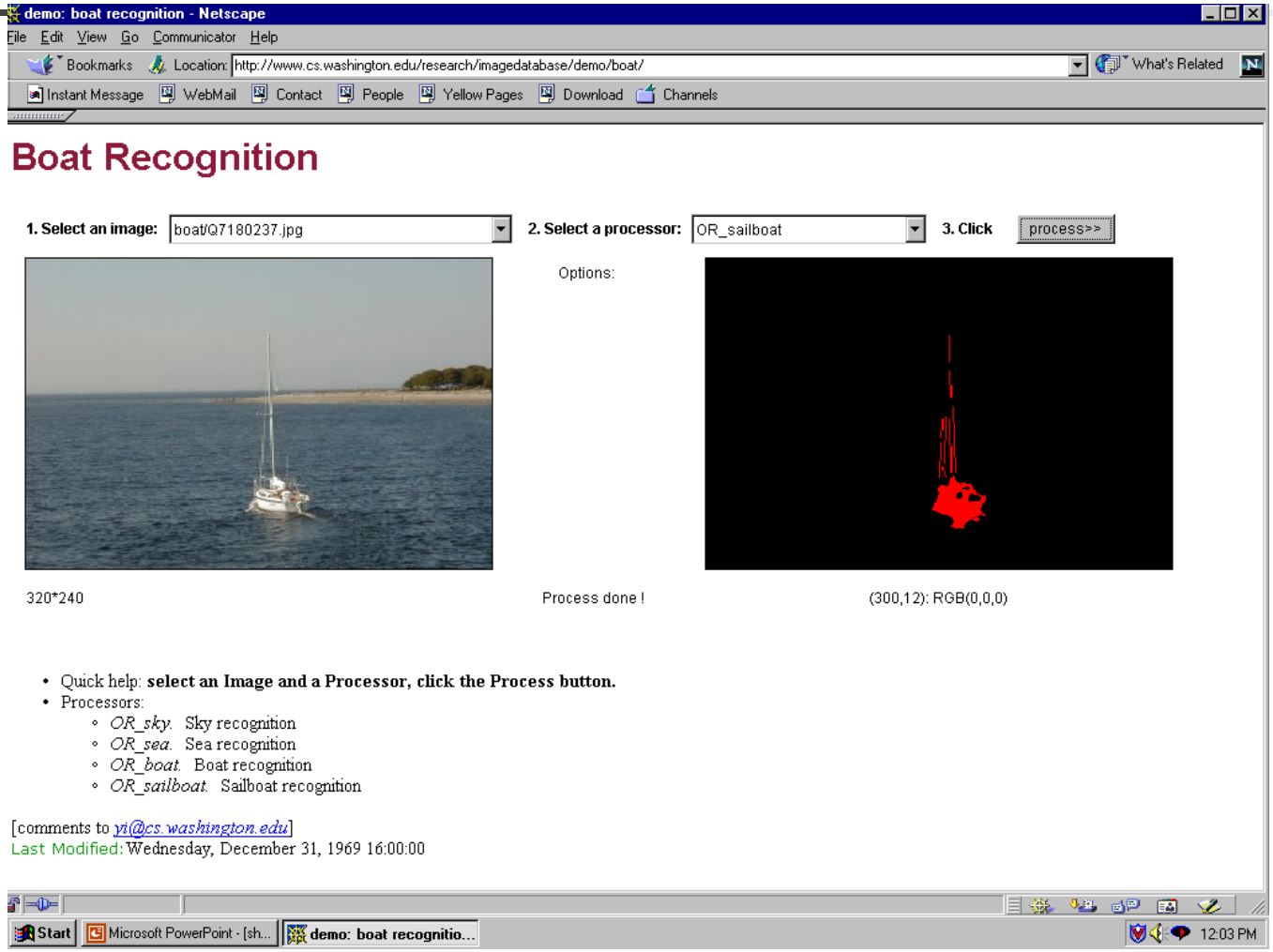
Current Research Objective



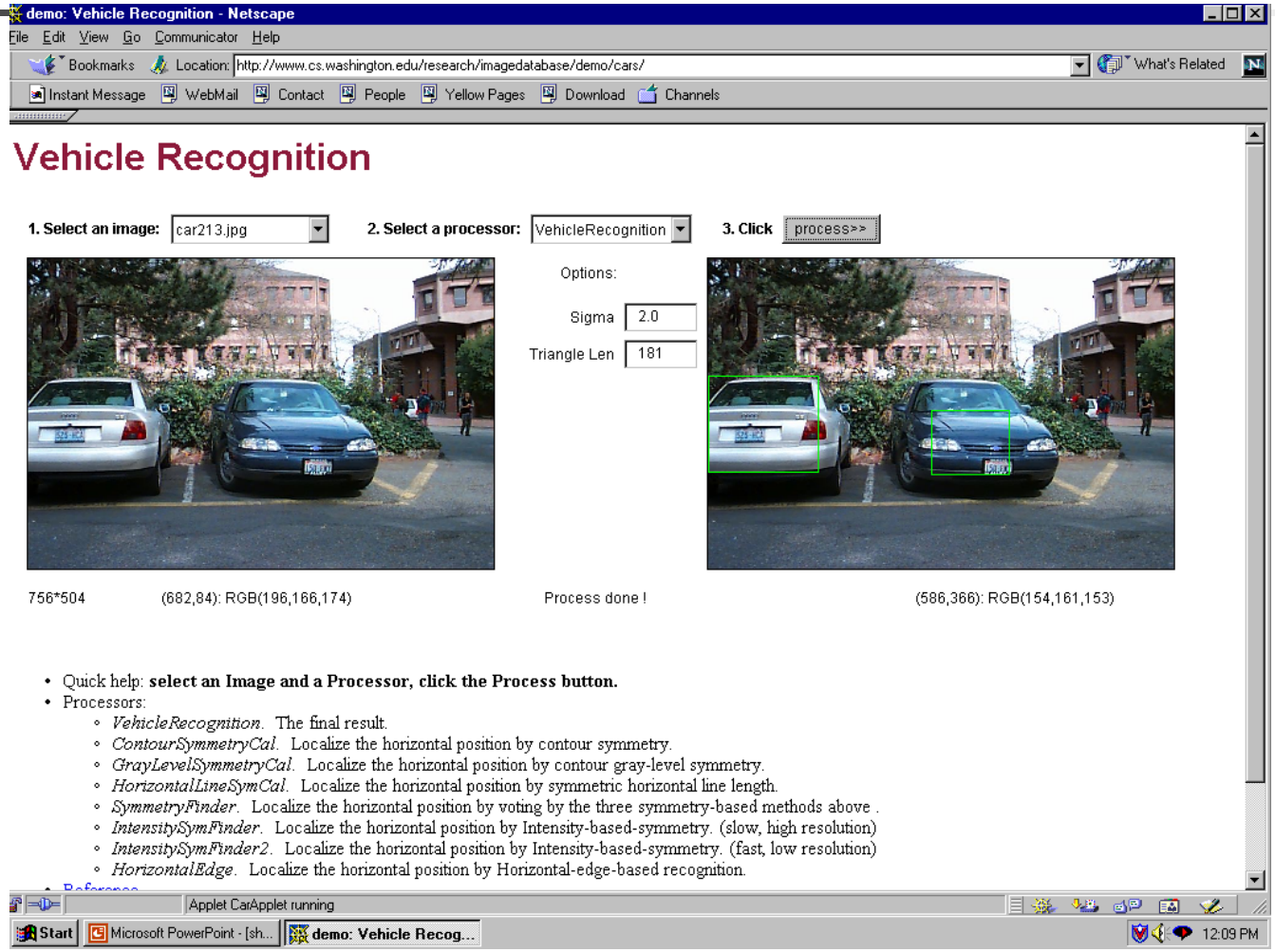
Overall Approach

- Develop object recognizers for common objects
- Use these recognizers to design a new set of both low- and high-level features
- Design a learning system that can use these features to recognize classes of objects

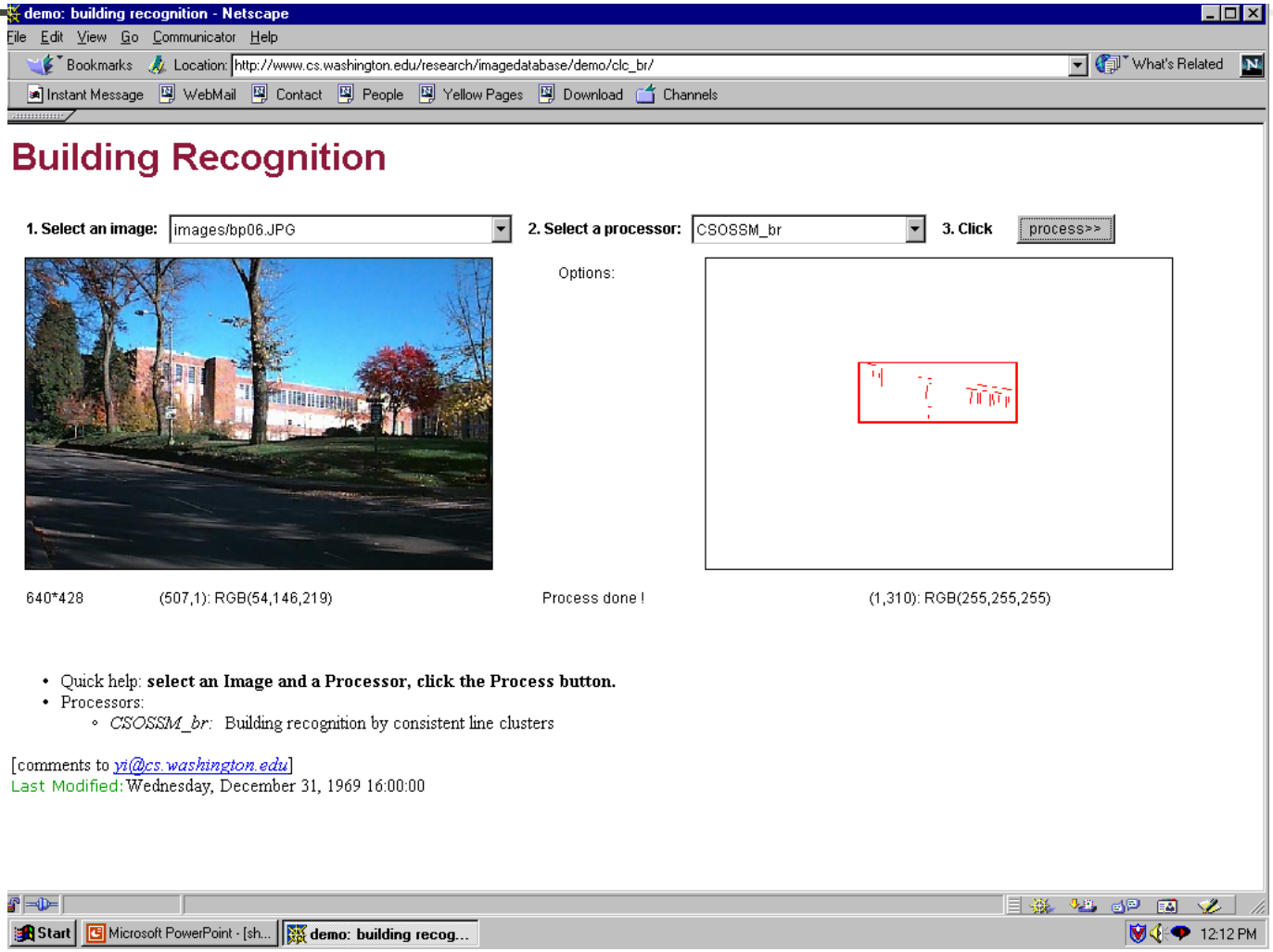
Boat Recognition



Vehicle Recognition



Building Recognition



Building Features: Consistent Line Clusters (CLC)

A **Consistent Line Cluster** is a set of lines that are homogeneous in terms of some line features.

- **Color-CLC**: The lines have the same color feature.
- **Orientation-CLC**: The lines are parallel to each other or converge to a common vanishing point.
- **Spatially-CLC**: The lines are in close proximity to each other.

Color-CLC

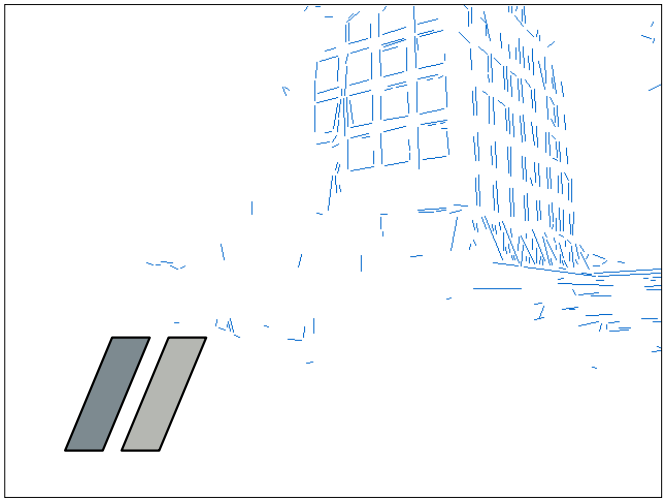
- Color feature of lines: color pair (c_1,c_2)
- Color pair space:

  RGB (256^3*256^3) Too big!

  Dominant colors (20*20)
- Finding the color pairs:

  One line $\rightarrow$ Several color pairs
- Constructing Color-CLC: use clustering

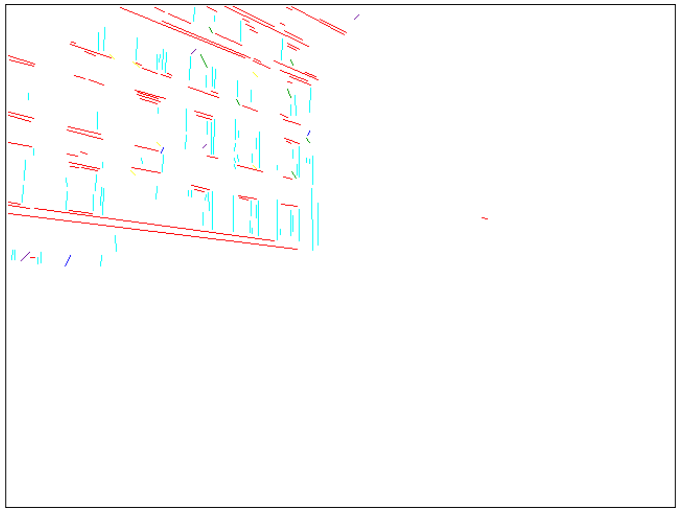
Color-CLC

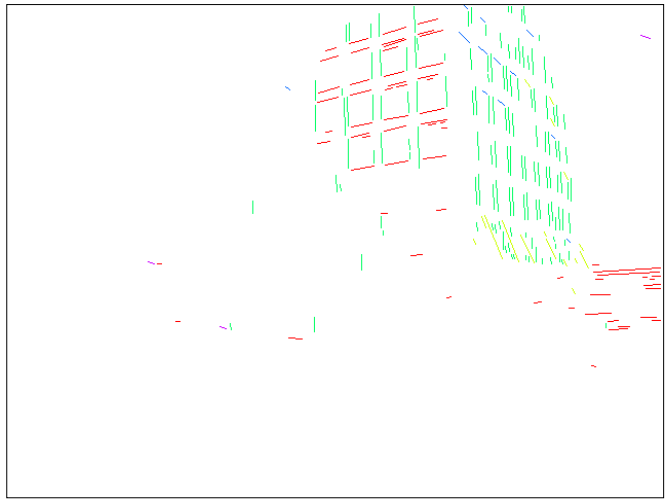


Orientation-CLC

- The lines in an Orientation-CLC are parallel to each other in the 3D world
- The parallel lines of an object in a 2D image can be:
  - Parallel in 2D
  - Converging to a vanishing point (perspective)

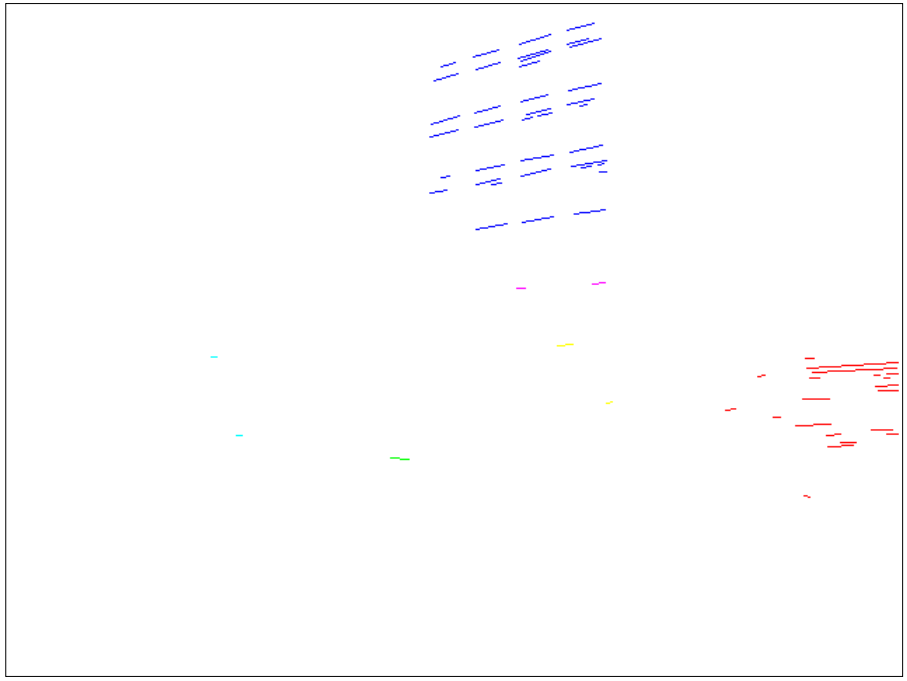
Orientation-CLC





Spatially-CLC

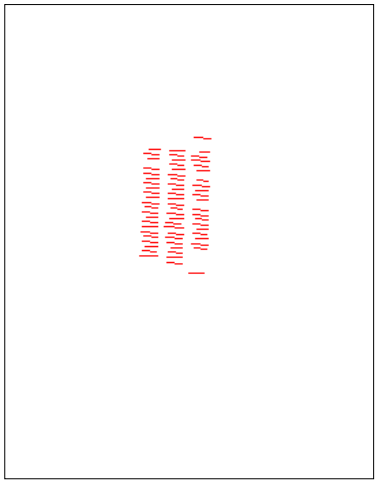
- Vertical position clustering
- Horizontal position clustering

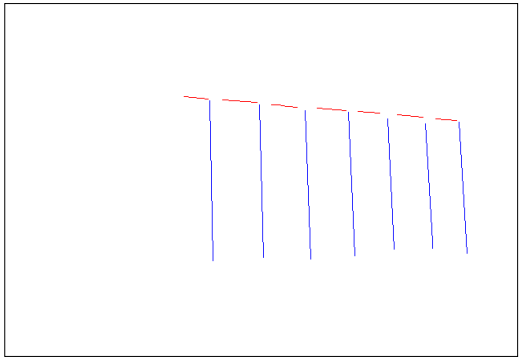


Building Recognition by CLC

Two types of buildings → Two criteria

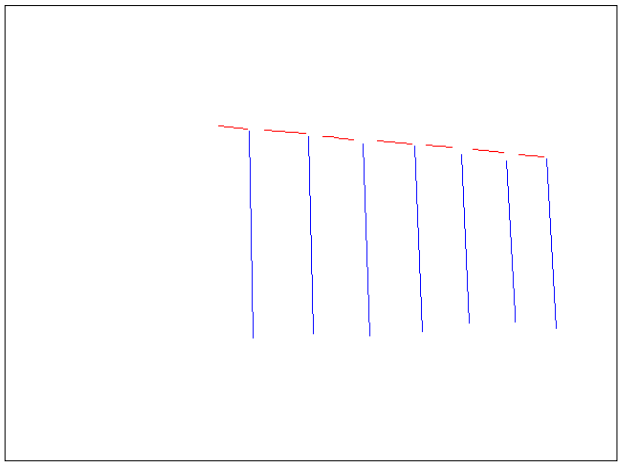
- Inter-relationship criterion
- Intra-relationship criterion





Inter-relationship criterion

$(N_{c1} > T_{i1}$ or $N_{c2} > T_{i1})$ and $(N_{c1} + N_{c2}) > T_{i2}$

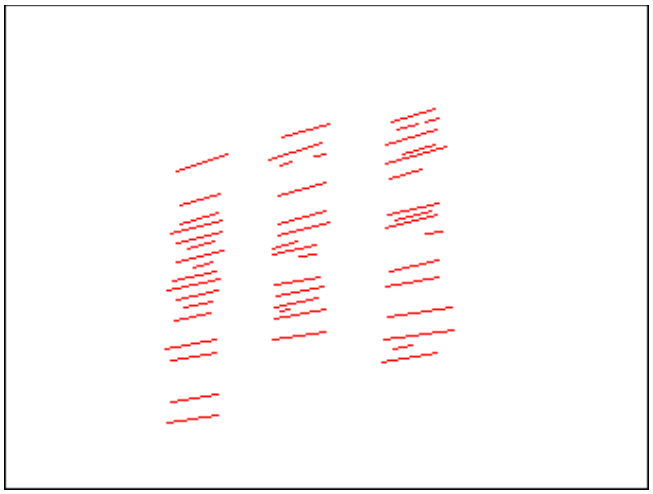


N_{c1} = number of intersecting lines in cluster 1

N_{c2} = number of intersecting lines in cluster 2

Intra-relationship criterion

$$|S_o| > T_{j1} \text{ or } w(S_o) > T_{j2}$$



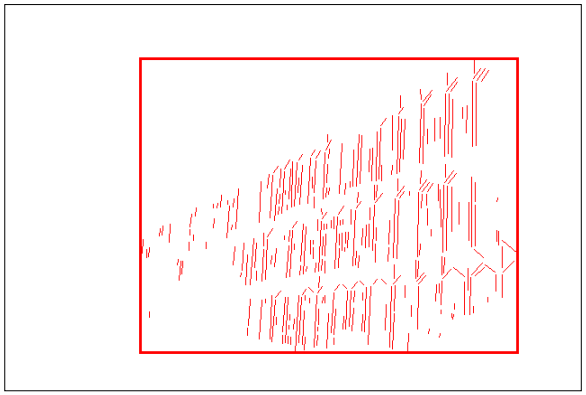
S_0 = set of heavily overlapping lines in a cluster

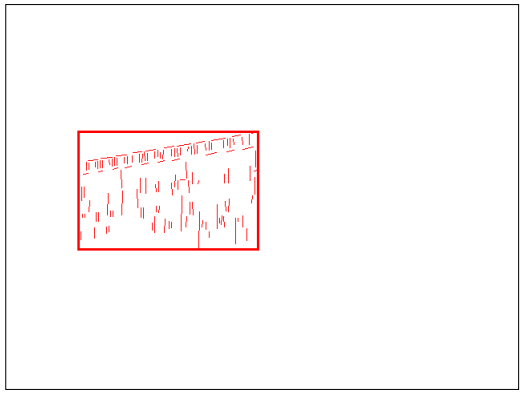
Experimental Evaluation

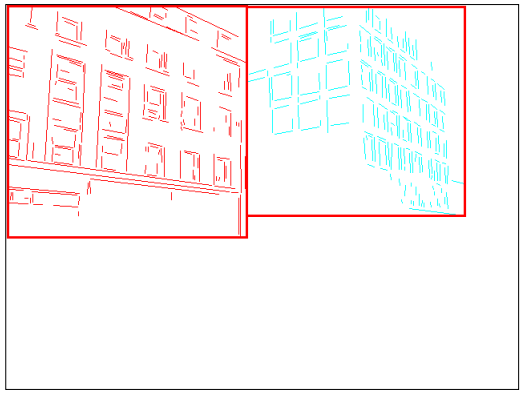
- Object Recognition
  - 97 well-patterned buildings (bp): 97/97
  - 44 not well-patterned buildings (bnp): 42/44
  - 16 not patterned non-buildings (nbnp): 15/16 (one false positive)
  - 25 patterned non-buildings (nbp): 0/25
- CBIR

Experimental Evaluation

Well-Patterned Buildings

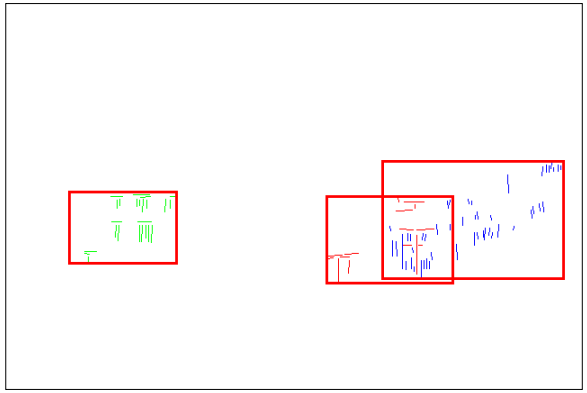


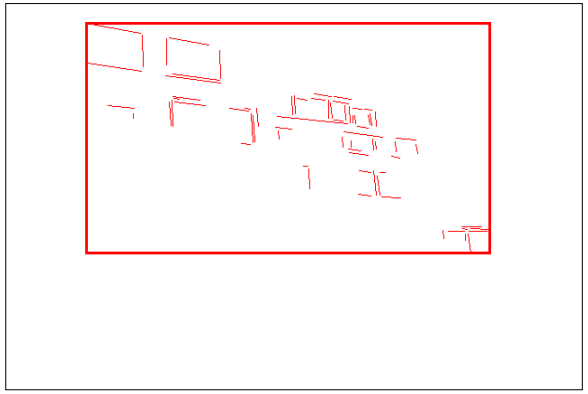


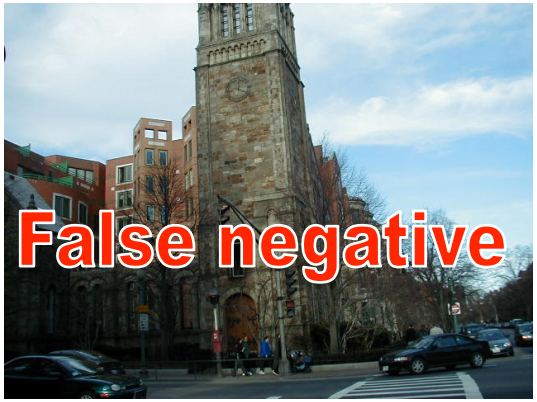


Experimental Evaluation

Non-Well-Patterned Buildings

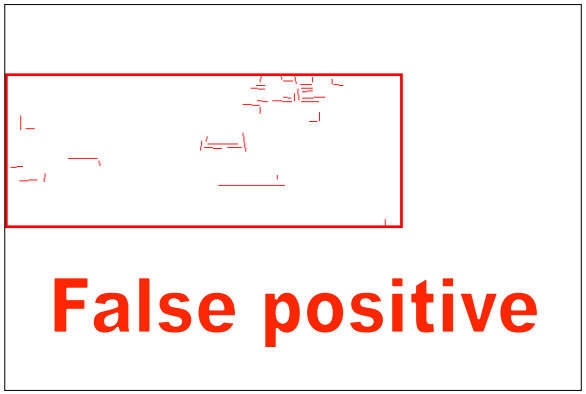






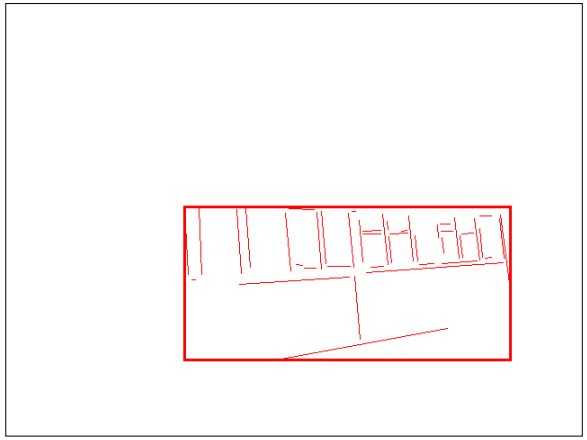
Experimental Evaluation

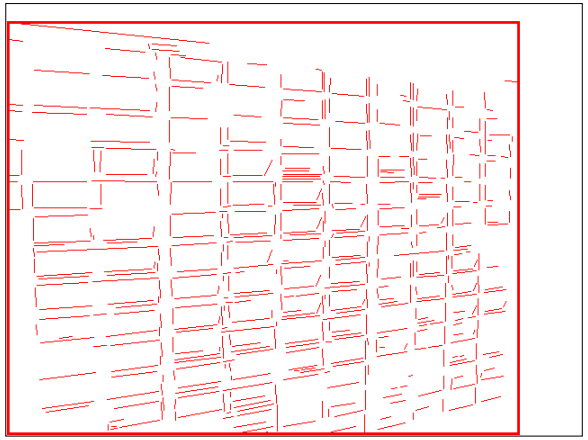
Non-Well-Patterned Non-Buildings

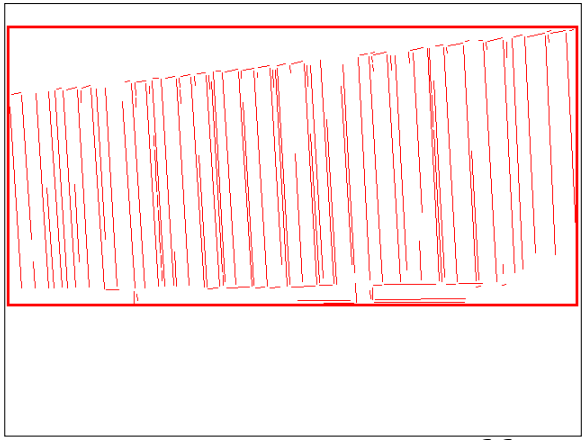


Experimental Evaluation

Well-Patterned Non-Buildings (false positives)







Experimental Evaluation (CBIR)

False positives from Yellowstone

Future Work

- Future Work
  - Constructing hierarchically structured clusters
  - Using CLC on other objects
  - Combining CLC with other features
  - Developing a learning approach using hierarchical, multiple classifiers (Chou 2000)

Text Search

- Documents (Web pages) represented by words
- Inverted index links keywords to the documents that contain them
- Keyword query retrieves documents containing that word

Inverted Index Example

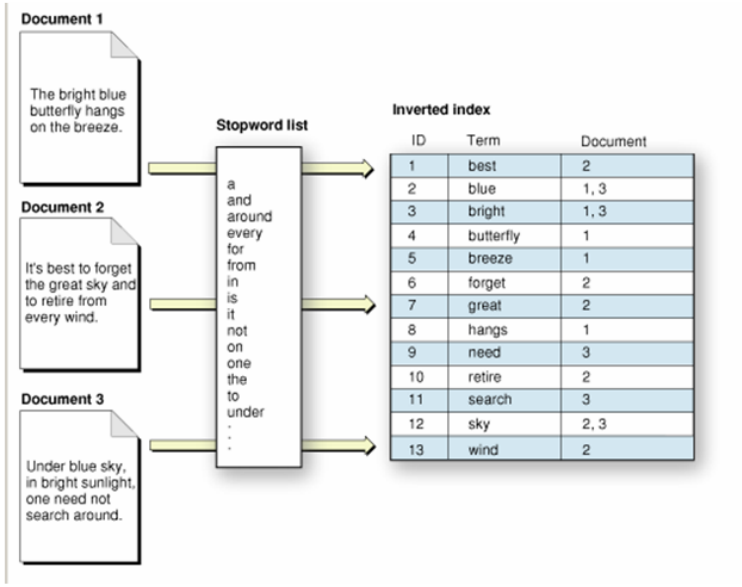


Image from http://developer.apple.com /documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/chapter_2_section_2.htm

Slide from Lynn Wilcox

Image Search - Tags

- Search over tags associated with images
  - Users manually add Tags to images
    - FlickR
    - FaceBook
  - Find images with tags that match the query keyword
- Limitations
  - Tags require human effort to create
  - Tags may be wrong

Slide from Lynn Wilcox

Image Search - Text

- Use text associated with images for search
  - Search web for images
  - Use surrounding text
    - Text in URL for image filename
    - Text in HTML on page
  - Same as text search

- Example: Google Image Search for “Sunset” gives
  - Sunset at Rocky Point in Australia
  - Sunset Beach, Oahu
  - Frank Smiles at Sunset

- Because the keyword “Sunset” was in the title of all these images

Sunset at Rocky Point

Frank Smiles at Sunset

Sunset Beach

Slide from Lynn Wilcox

Image Search – Image Descriptors

- SIFT Features
  - (Scale Invariant Feature Transform) Features
  - (2004: David Lowe, UBC)
- Select keypoints regions in image from extrema in scale space
  - Different images have different numbers of keypoints
- Compute feature vectors X for each keypoint region
  - Feature vectors from histogram of gradient directions near the keypoint
  - SIFT features X are 128-dimensional vectors
- Image described by N SIFT features
  - Features are $X_1, \ldots X_N$
  - N is different for different images

Slide from Lynn Wilcox

Image Search – Visual Words

- Quantize SIFT features to create "visual words" to represent images
  - (2006: Lienhart, University of Augsburg & Slaney, Yahoo!)
- Cluster SIFT features of representative images
  - Features X are in 128-dimensional space
  - Generate W clusters
  - Clusters define "visual words"
  - All features in same cluster are the same "visual word"
- To compute visual words describing an image
  - Compute N SIFT $X_1,\ldots.X_N$ features for the image
  - Find nearest cluster center (codeword) to each features X_j
  - These clusters define the visual words for the image

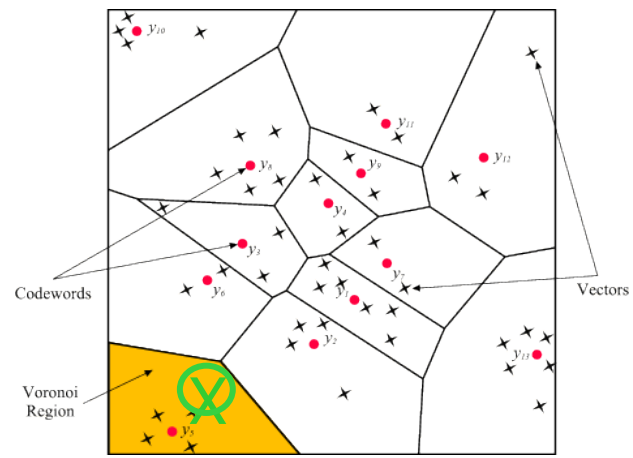


Image Feature ⊗

Visual Word 1

Slide from Lynn Wilcox

Image Retrieval – Visual Words

- Image is described by it's visual words
  - Just like a document is described by the text words
- Create image index
  - Compute visual words for all images
  - Create a visual word index into the images
- Compute visual words for query image
  - Use query words for retrieval
- Just like text!
  - Except the visual words aren't quite as meaningful

Slide from Lynn Wilcox

Image Search –Faces

- Face Detection
  - Find faces in images
  - Search for all images with faces
  - Ex: Google advance search for images with faces
  - Good results!
- Example:
  - FXPAL Photo Application (2004: Girgensohn et al.)

Photo Collection

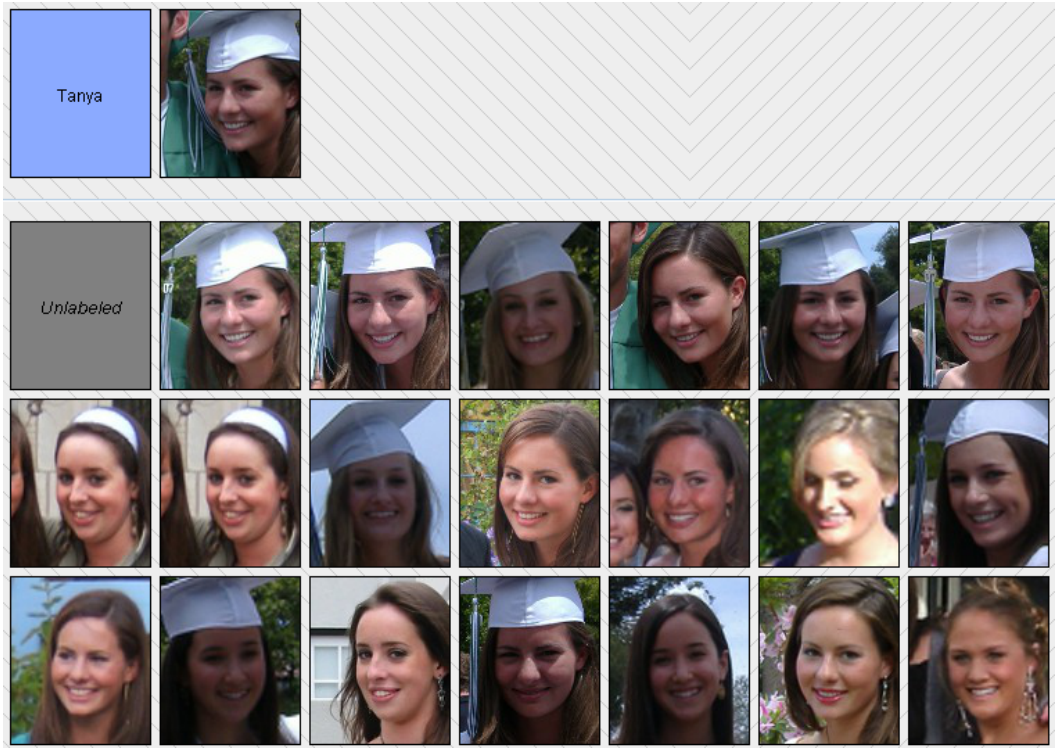
Faces in Photo Collection

Face Detection

Slide from Lynn Wilcox

Image Search –Faces

- Face Recognition
  - Search for all images of a particular person
  - Bad results!
- Face Similarity
  - Similarity search based on face features
  - Use face similarity to help manually label faces
  - Good results!
- User Interface for Labeling Faces
  - Drag face to label



Slide from Lynn Wilcox