CBIVR: Content-Based Image and Video Retrieval

Prepared by Stan Sclaroff

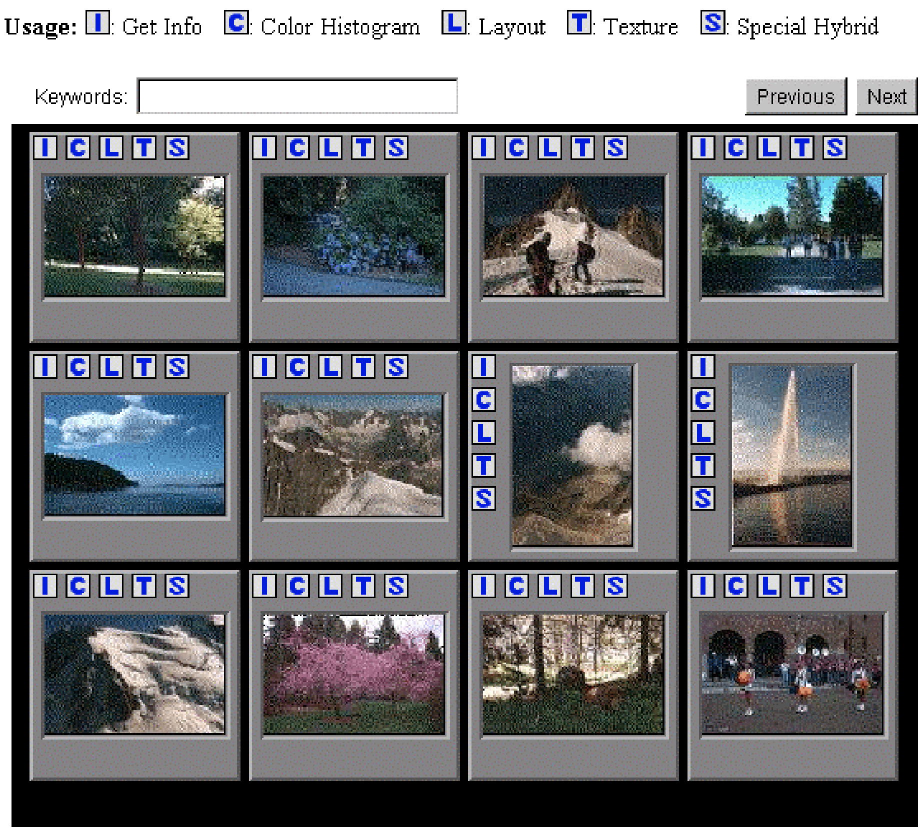
(with a few slides from Linda Shapiro)

for 6.801/6.866

December 3, 2002

QBIC: Query by Image Content

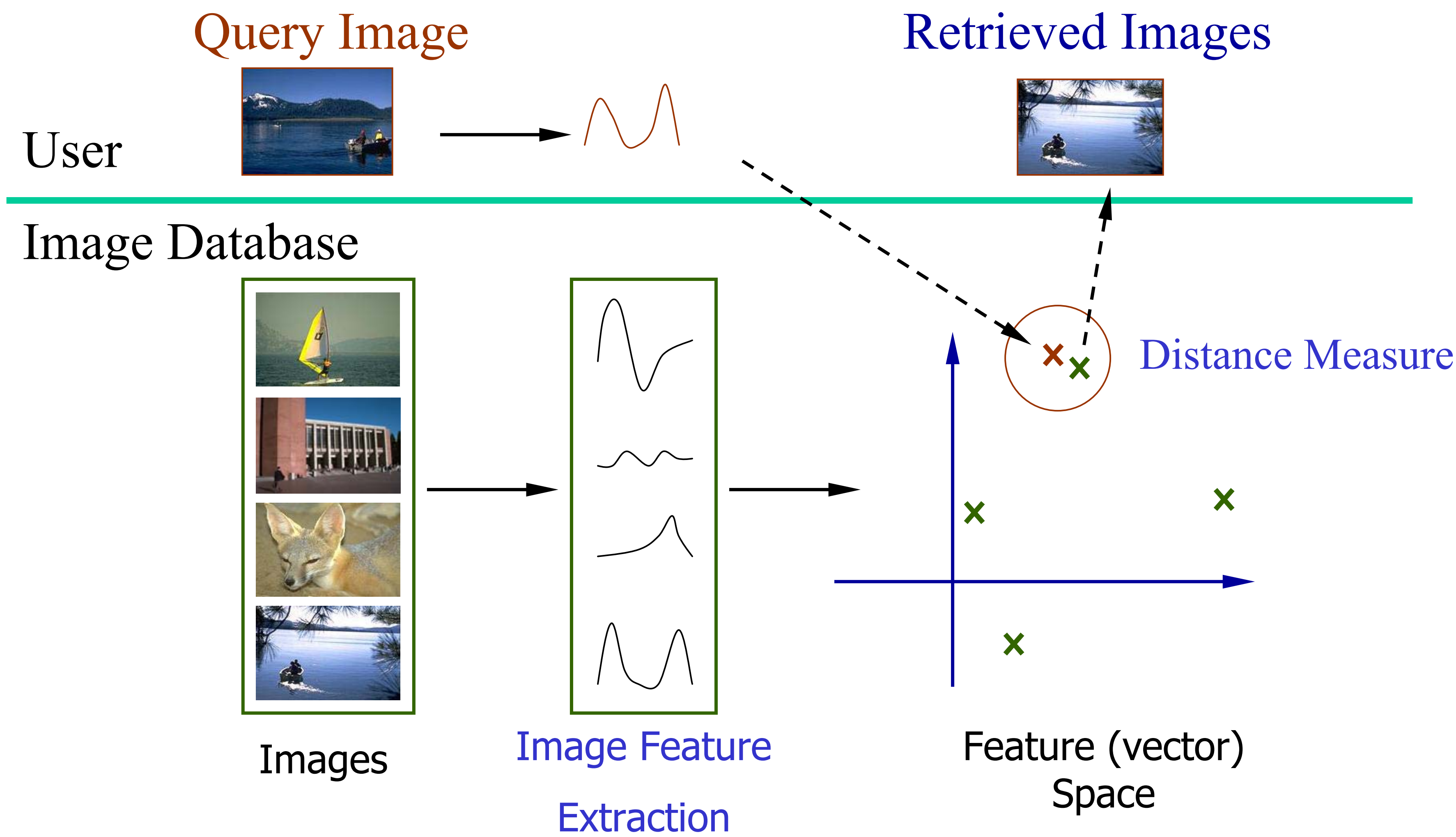
IBM QBIC™



- First commercial system
- Search by:
    - color percentages
    - color layout
    - texture
    - shape/location
    - keywords

Try their demo: http://wwwqbic.almaden.ibm.com

Image Features / Distance Measures



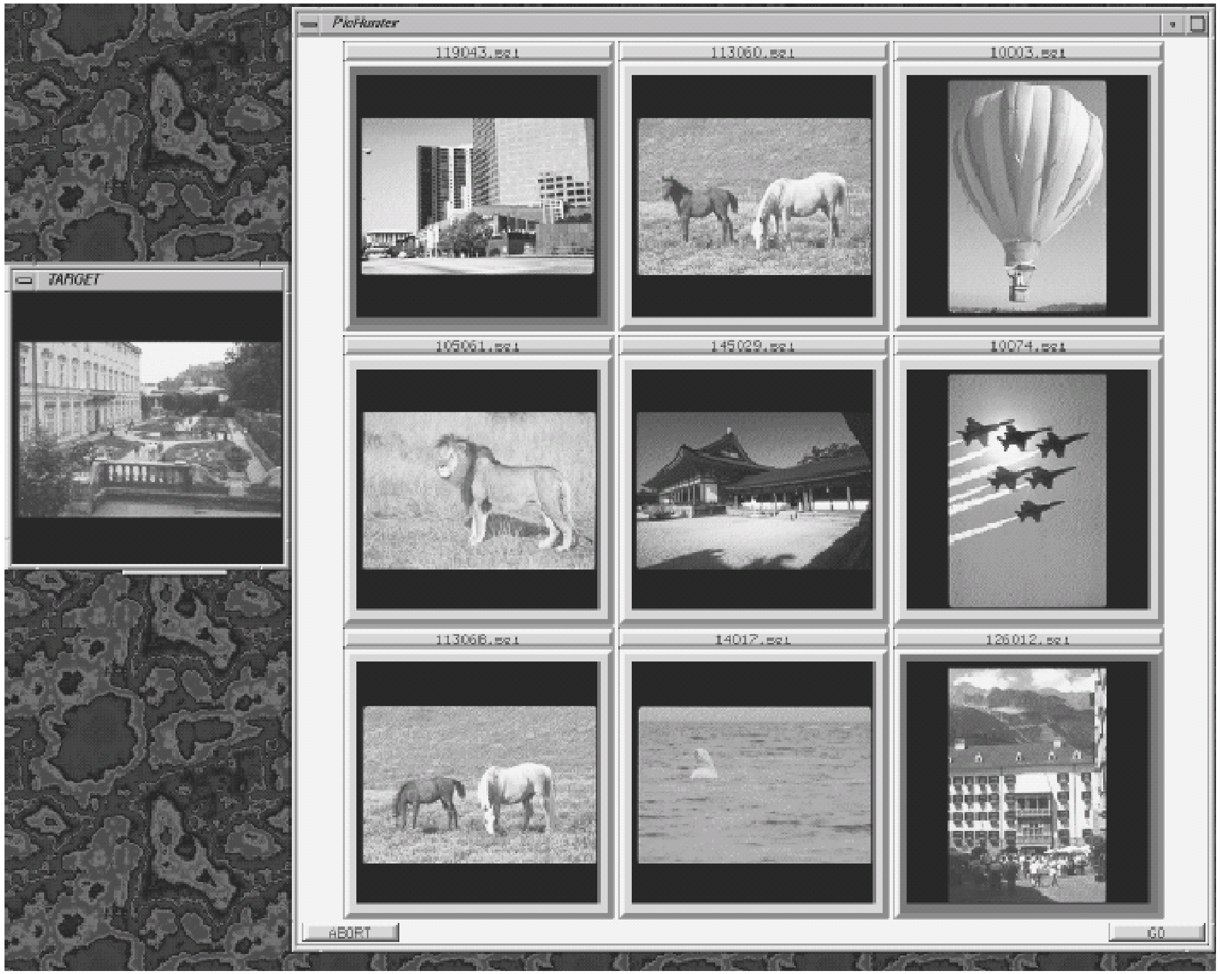
Query Formulation Methods

- QBE: Query by Example
  - Positive and negative examples
- Text description
- Query by sketch
- Cluster-based retrieval
- Relevance feedback

Query by Sketch

Example taken from Jacobs, Finkelstein, & Salesin
Fast Multi-Resolution Image Querying, SIGGRAPH 1995

Relevance Feedback



Example taken from Cox, Miller, Minka, Papathomas, and Yianilos, "The Bayesian Image Retrieval System, *PicHunter*: Theory, Implementation, and Psychophysical Experiments*,"* *IEEE T-IP, 2000.*

Application Areas

- Images and video on the web
- *Igrep*: Images and video in email and local files
- Individual collections of video or family photos
- Military intelligence, homeland security
- Archives: stock photos, stock film/video footage
- Access to museum collections
- Trademark and copyright infringement
- Medical information systems

CBIVR: Some Key Issues

Searching a large database for images or video clips that match a query:

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

CBIVR: Some Key Issues

Searching a large database for images or video clips that match a query:

- What kinds of databases?
- What kinds of queries?
- What constitutes a match?
- How to make such searches efficient?
- How to quantitatively evaluate performance?

Quantitative Evaluation of CBIVR Performance

A Standard Information Retrieval Evaluation Measure

For a given query q:

R_a = set of relevant documents in answer set A

R = set of relevant documents for q

$$Recall = \frac{|R_a|}{|R|} \qquad Precision = \frac{|R_a|}{|A|}$$

Ideally these values are close to one.

A Standard Information Retrieval Evaluation Measure

For a given query q:

R_a = set of relevant documents in answer set A

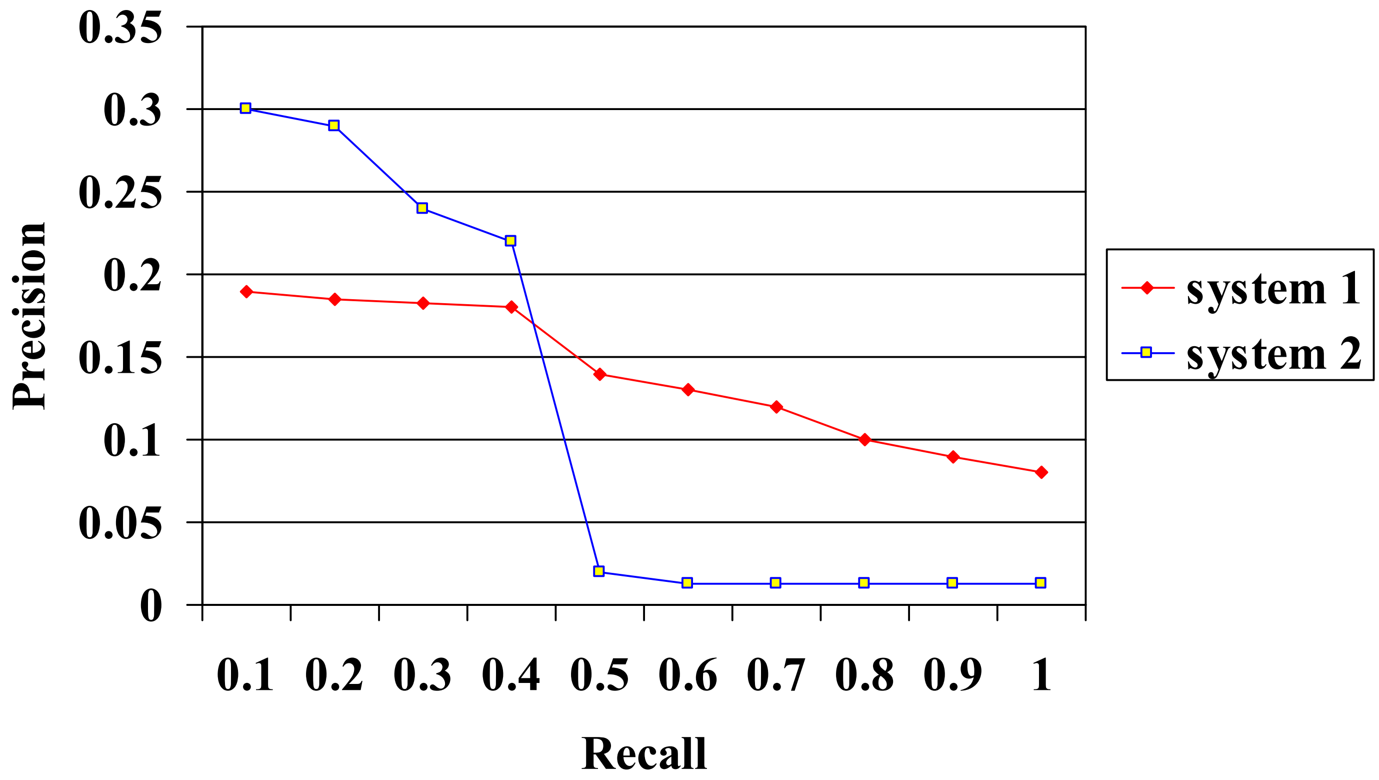
R = set of relevant documents for q

$$Recall = \frac{|R_a|}{|R|} \qquad Precision = \frac{|R_a|}{|A|}$$

Problem: What is relevant?

The relevance judgments of competent human informants can differ.

Average Precision vs. Recall



Generally, as the recall level rises, the level of precision falls.

Which system is better?

Desirable Precision vs. Recall is Application-Dependent!

- Images and video on the web
- *Igrep*: Images and video in email and local files
- Individual collections of video or family photos
- Military intelligence, homeland security
- Archives: stock photos, stock film/video footage
- Access to museum collections
- Trademark and copyright infringement
- Medical information systems

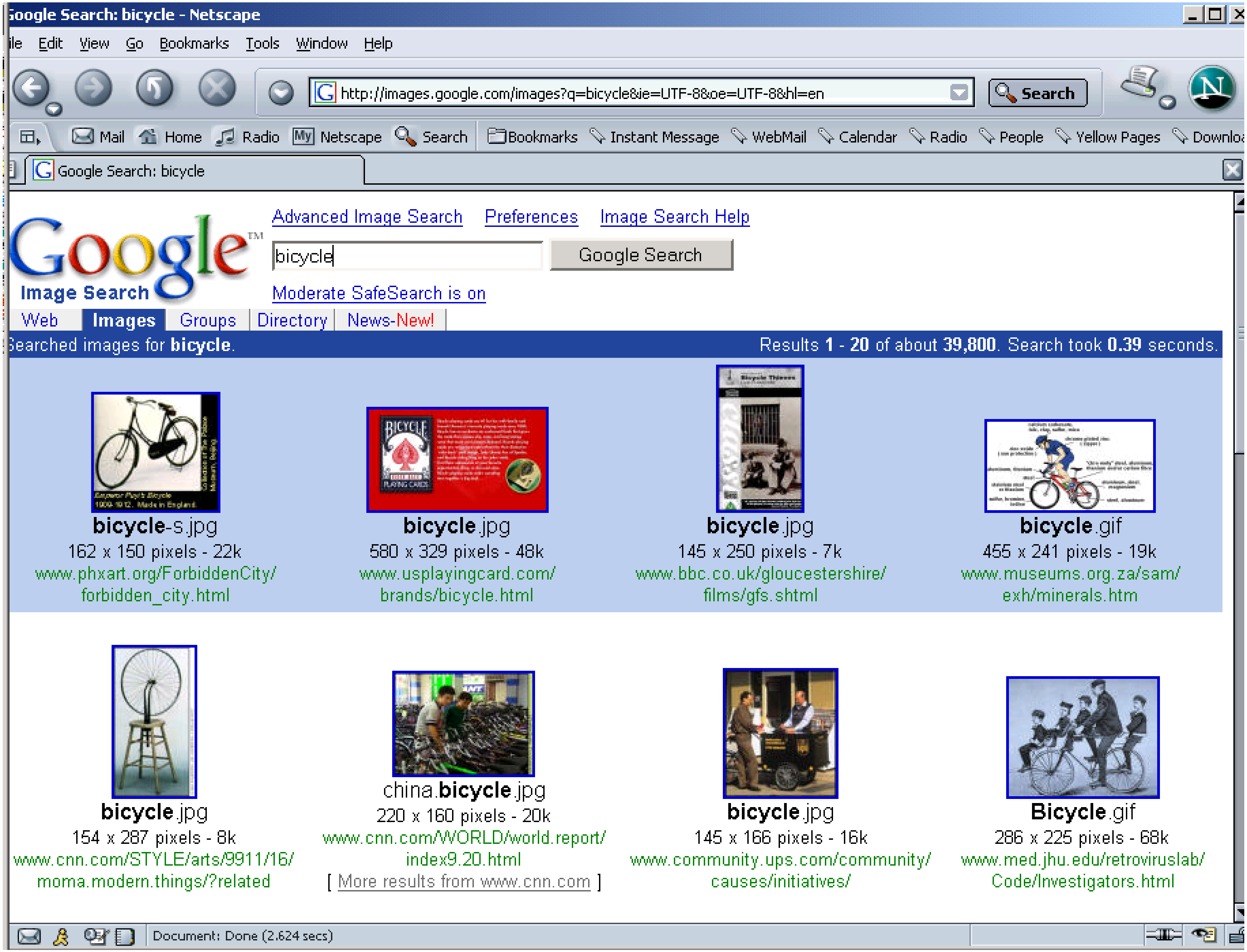
Example Application:

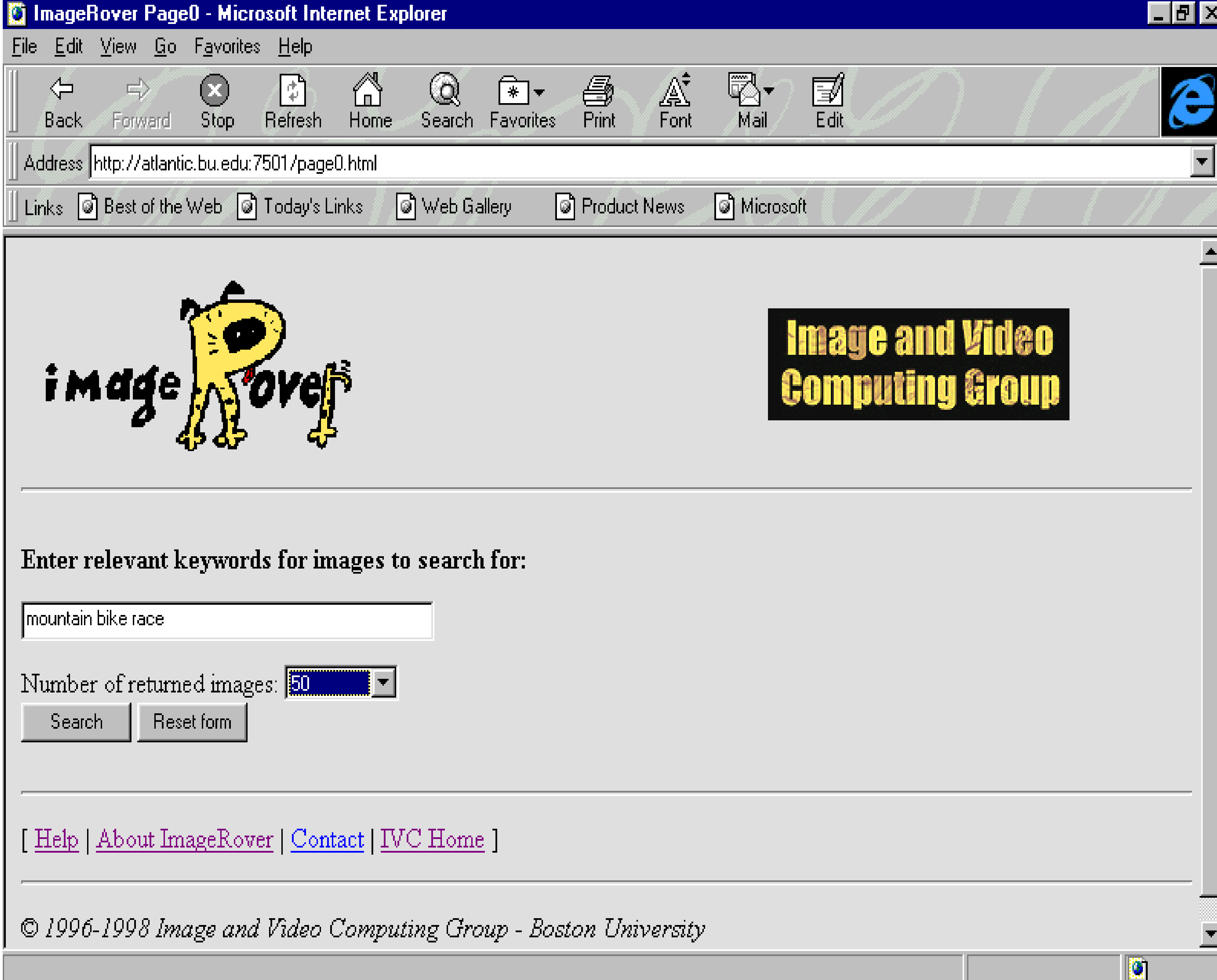
CBIVR for the WWW

CBIVR for the WWW?

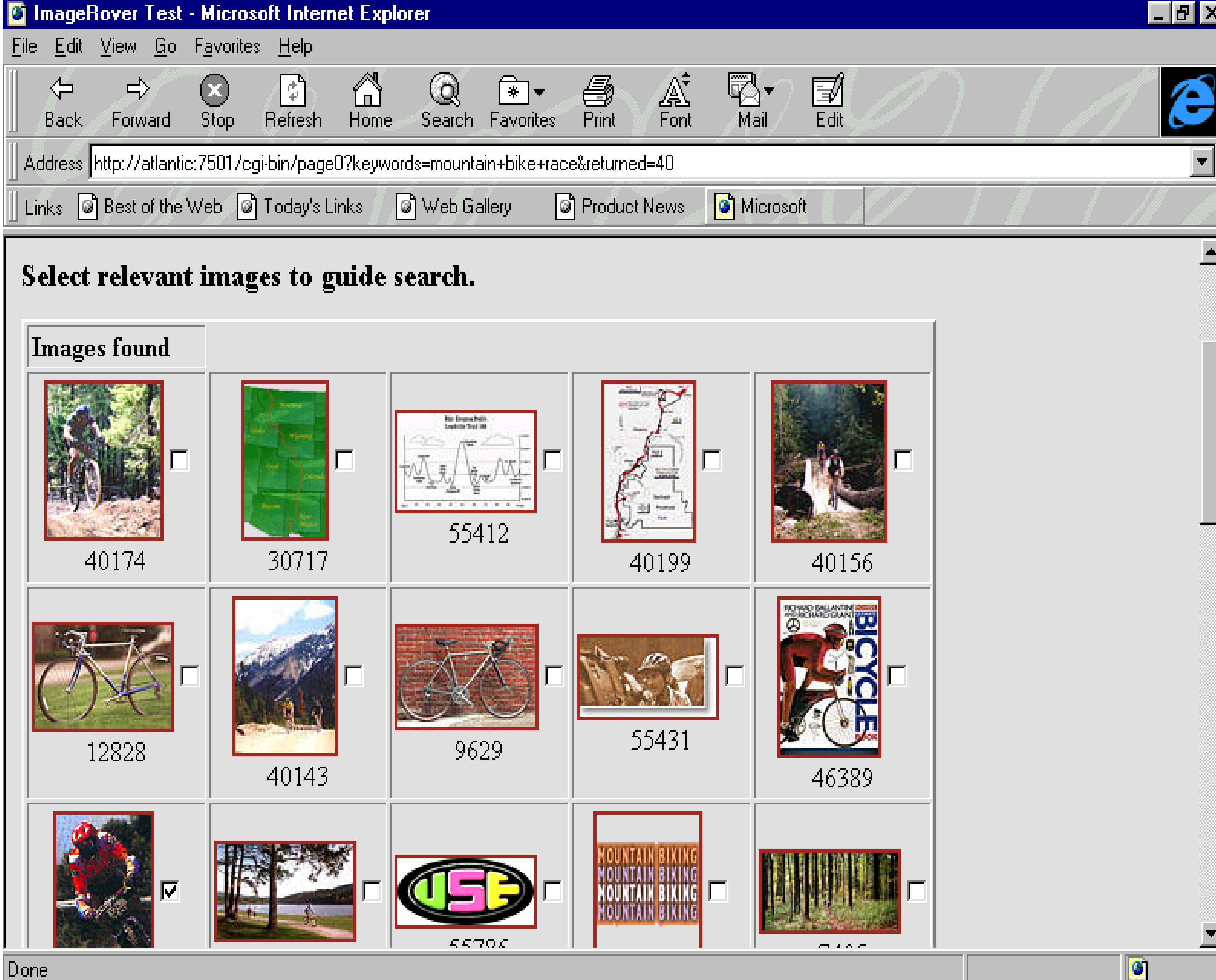
- Very large, unstructured database
- Diverse content
- No single standard image nor video format
- No standard illumination
- Images/video can be altered in Photoshop, etc.
- Lossy compression, color quantization, scanned

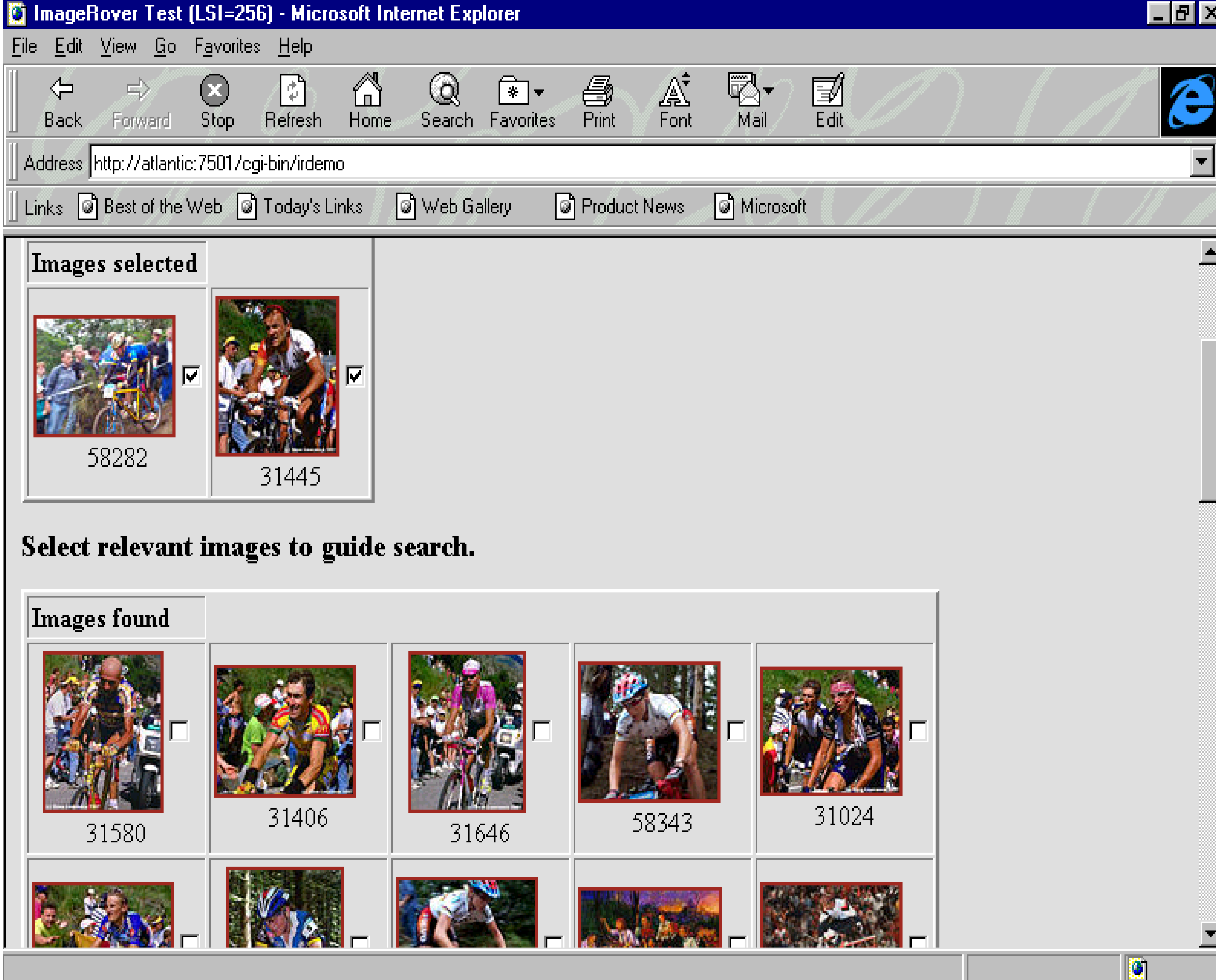
What are the precision and recall requirements?

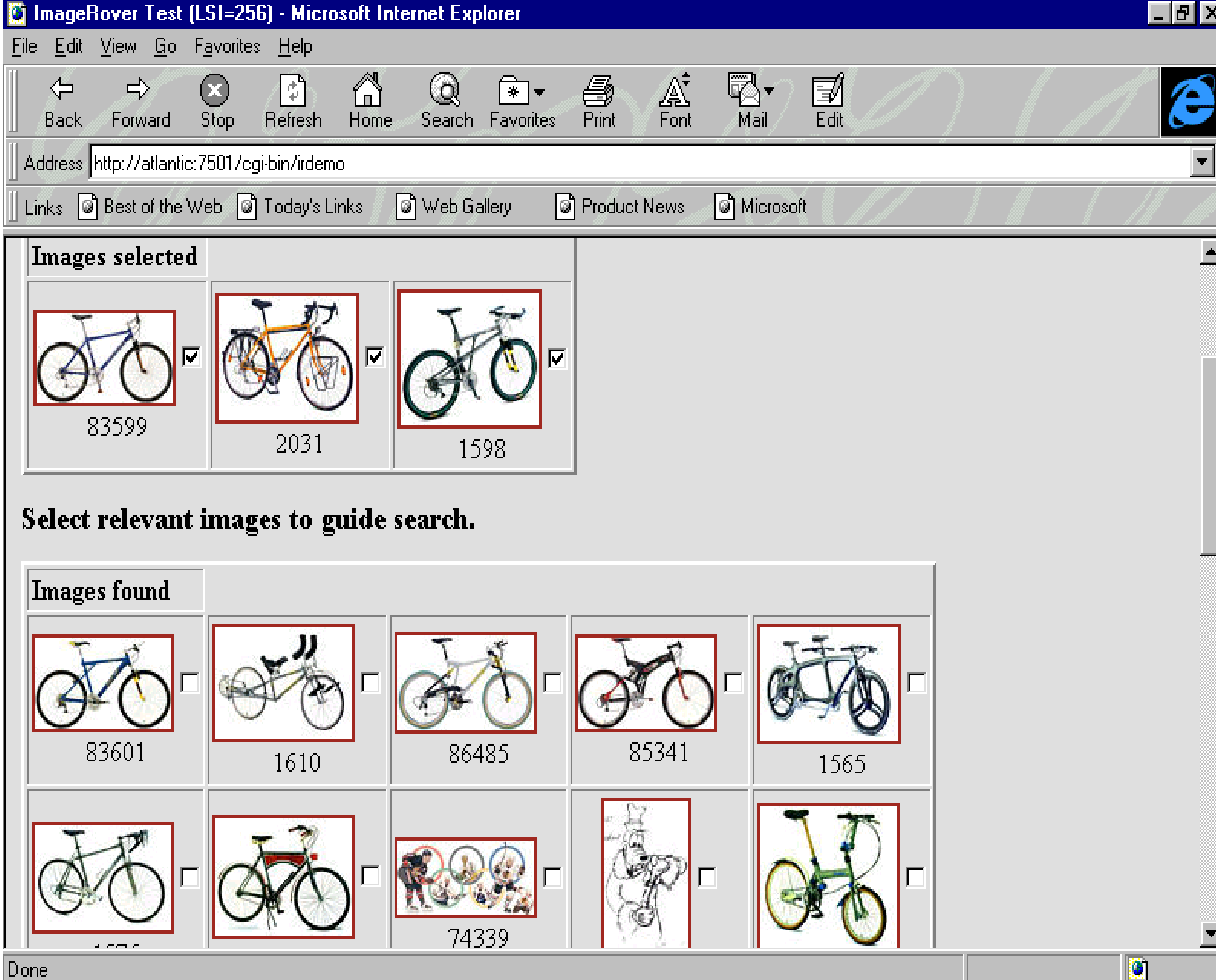

Google Search: bicycle - Netscape
File Edit View Go Bookmarks Tools Window Help
http://images.google.com/images?q=bicycle&ie=UTF-8&oe=UTF-8&hl=en
Search
Mail Home Radio Netscape Search Bookmarks Instant Message WebMail Calendar Radio People Yellow Pages Downloa
Google Search: bicycle
Google Image Search
Advanced Image Search Preferences Image Search Help
bicycle
Google Search
Moderate SafeSearch is on
Web Images Groups Directory News-New!
Searched images for bicycle.
Results 1 - 20 of about 39,800. Search took 0.39 seconds.
bicycle-s.jpg
162 x 150 pixels - 22k
www.phxart.org/ForbiddenCity/forbidden_city.html
bicycle.jpg
580 x 329 pixels - 48k
www.usplayingcard.com/brands/bicycle.html
bicycle.jpg
145 x 250 pixels - 7k
www.bbc.co.uk/gloucestershire/films/gfs.shtml
bicycle.gif
455 x 241 pixels - 19k
www.museums.org.za/sam/exh/minerals.htm
bicycle.jpg
154 x 287 pixels - 8k
www.cnn.com/STYLE/arts/9911/16/moma.modern.things/?related
china.bicycle.jpg
220 x 160 pixels - 20k
www.cnn.com/WORLD/world.report/index9.20.html
[More results from www.cnn.com]
bicycle.jpg
145 x 166 pixels - 16k
www.community.ups.com/community/causes/initiatives/
Bicycle.gif
286 x 225 pixels - 68k
www.med.jhu.edu/retroviruslab/Code/Investigators.html
Document: Done (2.624 secs)



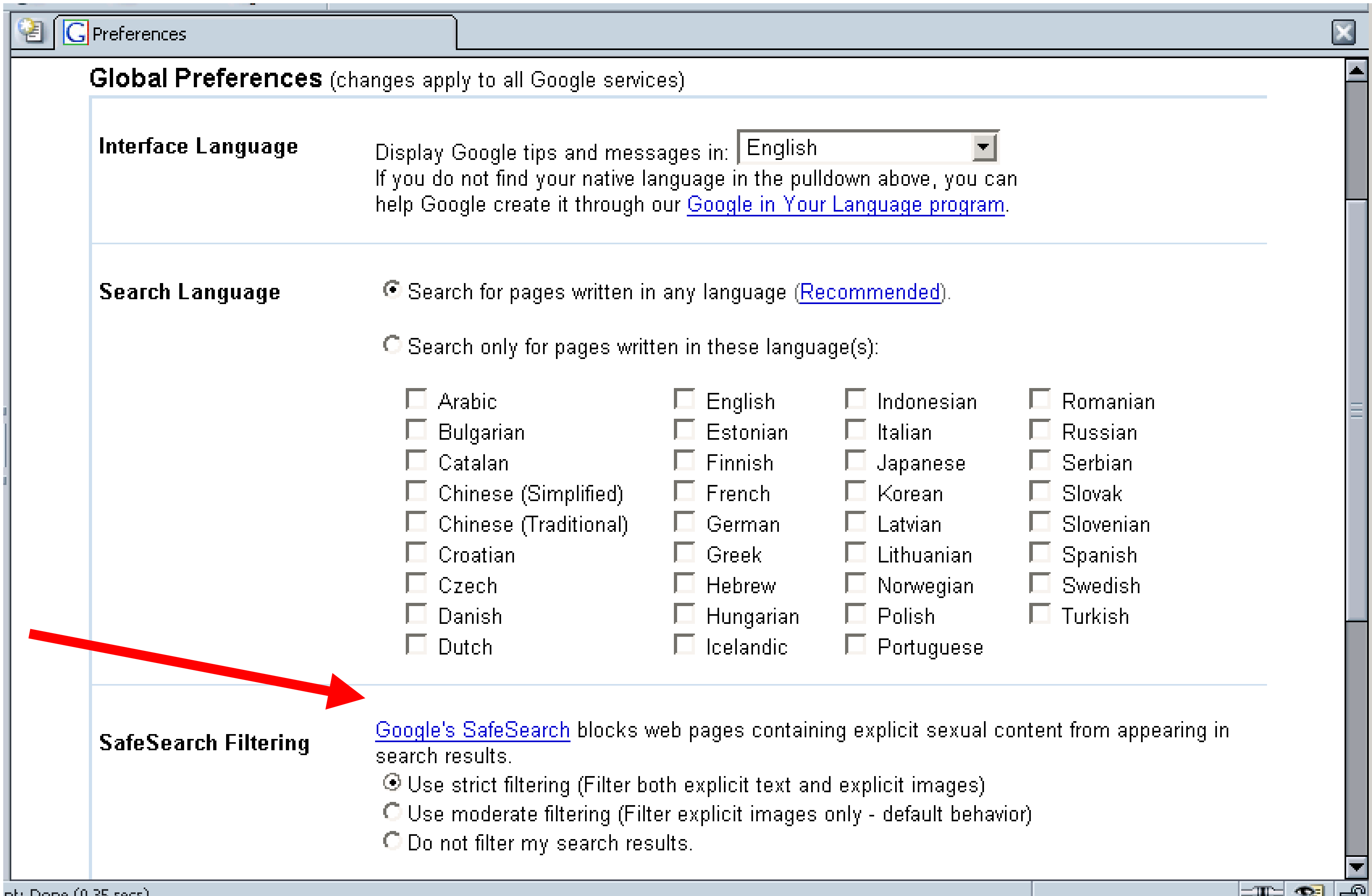
ImageRover Page0 - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites Print Font Mail Edit
Address http://atlantic.bu.edu:7501/page0.html
Links Best of the Web Today's Links Web Gallery Product News Microsoft
imageRover
Image and Video Computing Group
Enter relevant keywords for images to search for:
mountain bike race
Number of returned images: 50
Search Reset form
[Help | About ImageRover | Contact | IVC Home]
© 1996-1998 Image and Video Computing Group - Boston University


ImageRover Test - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites Print Font Mail Edit
Address http://atlantic:7501/cgi-bin/page0?keywords=mountain+bike+race&returned=40
Links Best of the Web Today's Links Web Gallery Product News Microsoft
Select relevant images to guide search.
Images found
40174
30717
55412
40199
40156
12828
40143
9629
55431
46389
BICYCLE
MOUNTAIN BIKING
Done


ImageRover Test (LSI=256) - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites Print Font Mail Edit
Address http://atlantic:7501/cgi-bin/irdemo
Links Best of the Web Today's Links Web Gallery Product News Microsoft
Images selected
58282
31445
Select relevant images to guide search.
Images found
31580
31406
31646
58343
31024


ImageRover Test (LSI=256) - Microsoft Internet Explorer
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites Print Font Mail Edit
Address http://atlantic:7501/cgi-bin/irdemo
Links Best of the Web Today's Links Web Gallery Product News Microsoft
Images selected
83599
2031
1598
Select relevant images to guide search.
Images found
83601
1610
86485
85341
1565
74339
Done

Google *SafeSearch* Filtering



Finding Naked People

[Fleck,Forsyth, and Bregler 1996]

- Convert RGB color to HIS color
- 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.
Use simple grouping rules for limbs/trunk/legs (see their paper).

Image Representations for CBIR

Image Representations

There are roughly three levels of image representation used for CBIR:

1. Iconic – exact pixel values
2. Compositional – overall image appearance
3. Objects – things depicted in the image, their properties, and their relationships

Iconic Matching

Example applications:

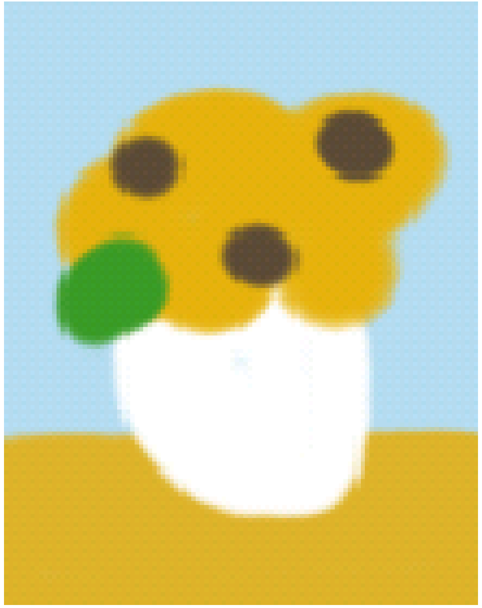
- Copyright and trademark protection
- Duplicate removal
- Linking images used in evidence, for example child pornography

Problems in finding "exact" matches:

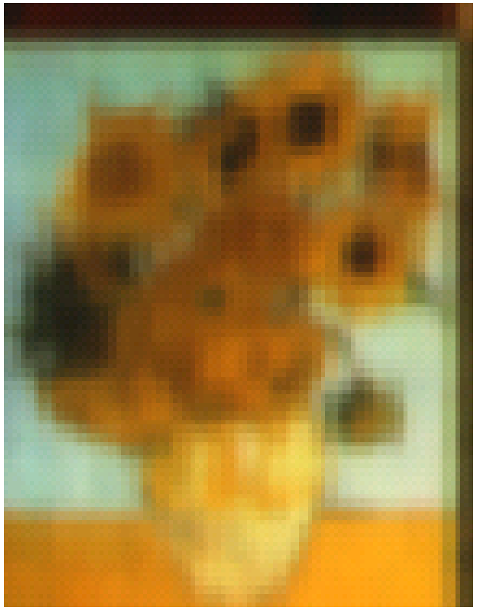
- Lossy compression, image scanning
- Color space conversion
- Photoshop-style transforms: blur, scale, rotate, warp, crop, cut, etc.

Iconic Matching

painted



scanned

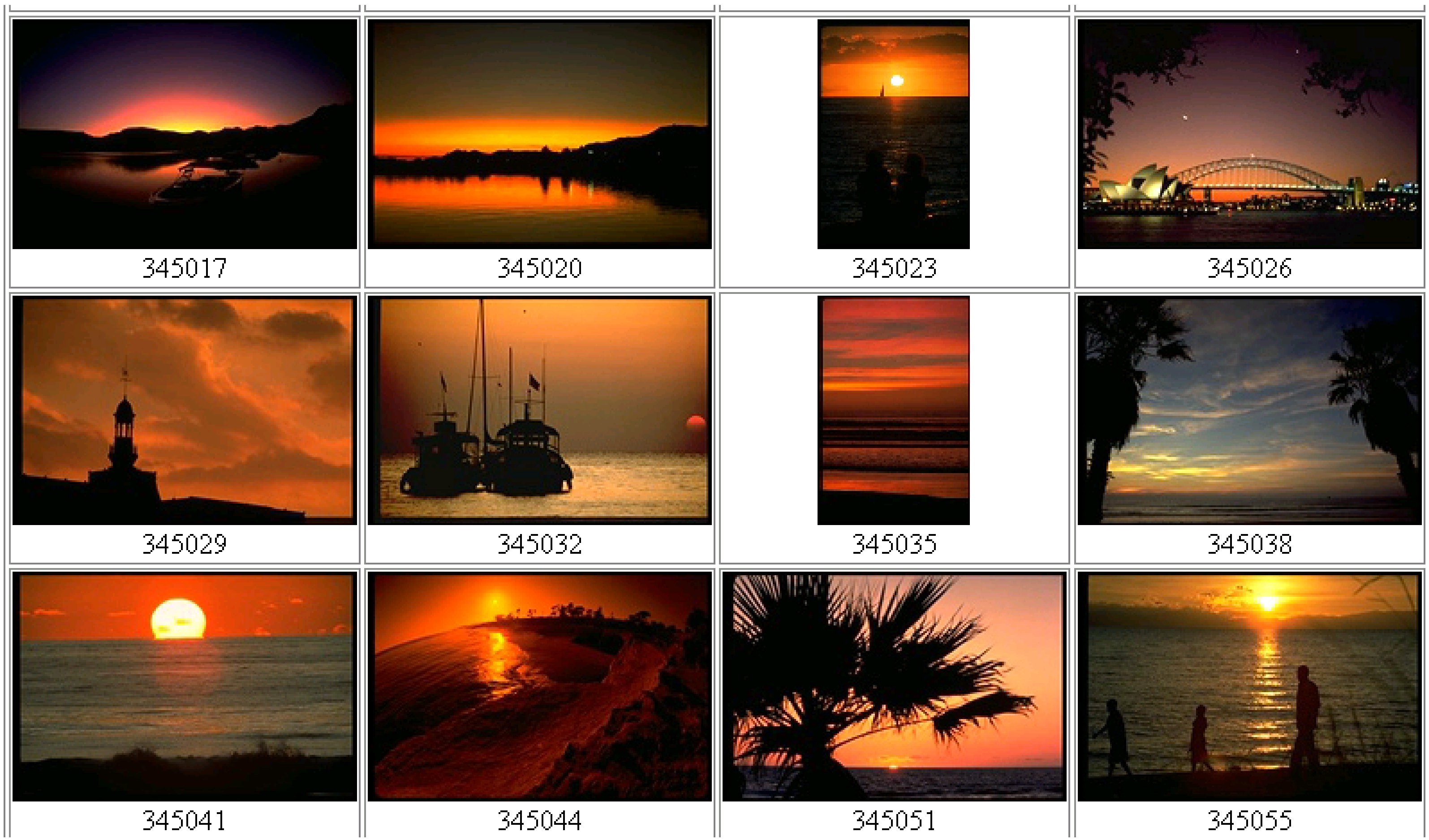


target

From Jacobs, Finkelstein, & Salesin
Fast Multi-Resolution Image Querying, SIGGRAPH 1995

Summary Representations of a Whole Picture: Color Histograms

[Swain and Ballard, IJCV 1991]



Color Histograms

```
Off-line, for each image
    create histogram with a bin for each color
    initialize each bin counter = 0
        for each pixel in image:
            increment bin counter corresponding to pixel
            color
        end
```

On-line, use histograms in image similarity measure:
Euclidean, dot product, histogram intersection, etc.

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
- $\mathbf{A}$ is a K x K similarity matrix

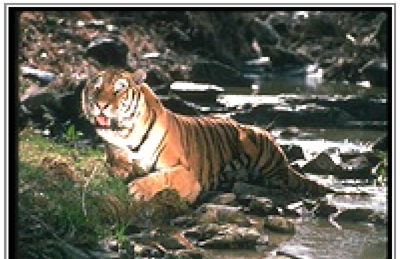
Similarity Matrix: A

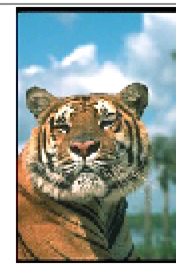
	R	G	B	Y	C	V
R	1	0	0	.5	0	.5
G	0	1	0	.5	.5	0
B	0	0	1		?	
Y				1		
C		?			1	
V						1

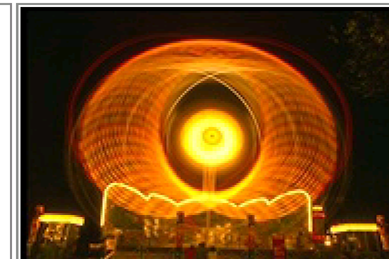
How similar is blue to cyan?

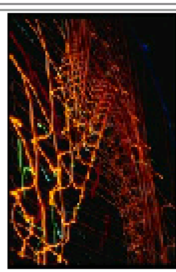
Images Classified as Sunsets using Overall Color Histograms

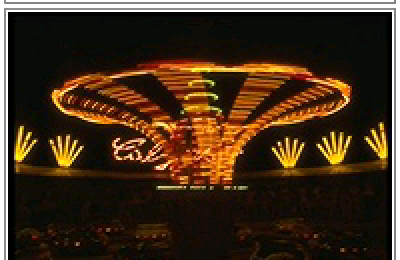

10085

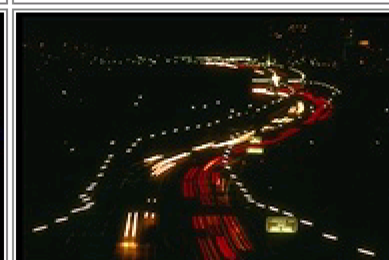

108047

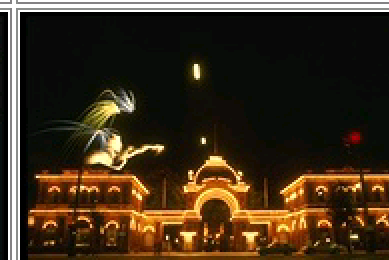

108099

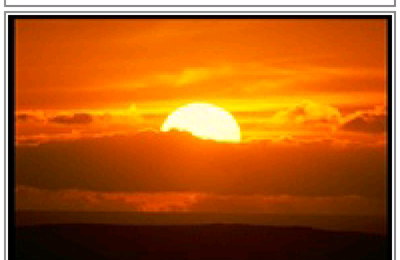

287029

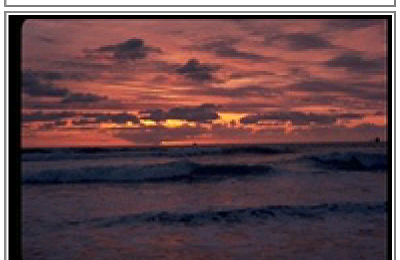

287040

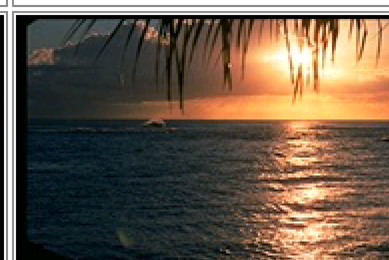

287048

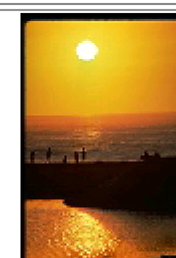

287057


287092


345002

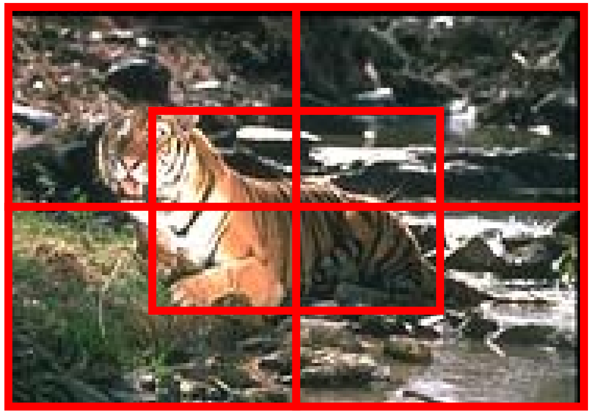

345005

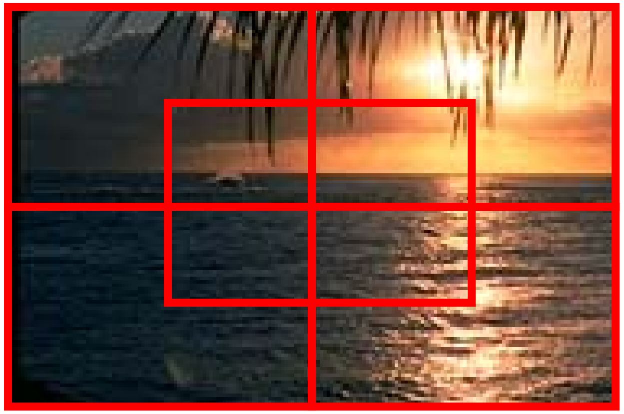

345011


345014

Histograms of Partitioned Image

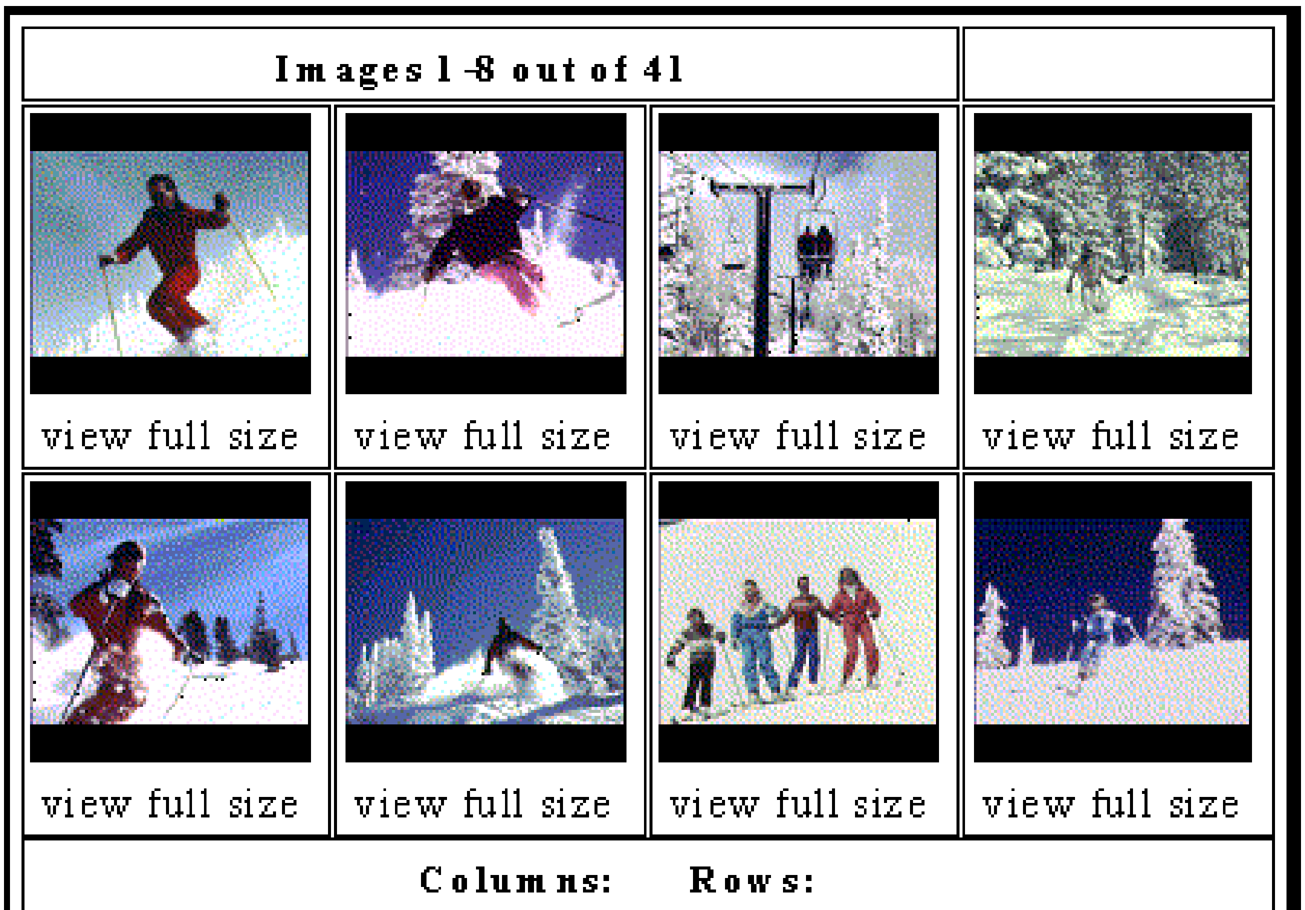
Divide image up into rectangles.
Compute separate histogram for each partition.





Rectangles can overlap.

Retrieval by “color layout” in IBM’s QBIC



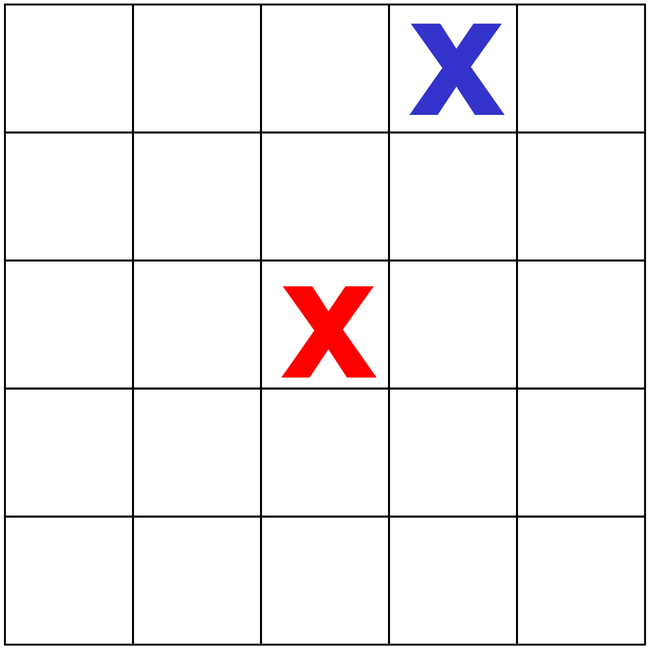
Indexing with Color Correlograms
[Zabih, et al.]

Problem: Pictures taken from slightly different view positions can look substantially different with a color histogram similarity measure.

Proposed solution: Compute color co-occurance statistics [Haralick 1979].

Color Correlogram
[Zabih, et al.]

For each image, estimate the probability that a pixel of some color lies within a particular distance of pixel of another color.



Estimating Color Correlogram

Consider set of distances of interest $[d]=\{1,2,\ldots,d\}$
Measure pixel distance with L_∞ norm.
Consider m possible colors $c_i \in \{c_1, c_2, \ldots, c_m\}$.

```
Offline, for each image:
    Construct a correlogram that has m×m×d bins, initialize=0.
    For each pixel p_i in the image, find it's color c_i
        for each distance k∈{1,2,...,d}
            for each pixel at distance k from p_i
                increment bin (i,j,k)
            end
        end
    end
    Normalize correlogram by a scale factor (see Zabih, et al.)
```

Similarity Measure

Given correlograms of query and target images, $\gamma(Q)$ and $\gamma(T)$, define similarity:

$$|Q-T|_{\gamma,L_1} \triangleq \sum_{i,j\in[m],k\in[d]} |\gamma_{c_i c_j}^{(k)}(Q) - \gamma_{c_i c_j}^{(k)}(T)|$$

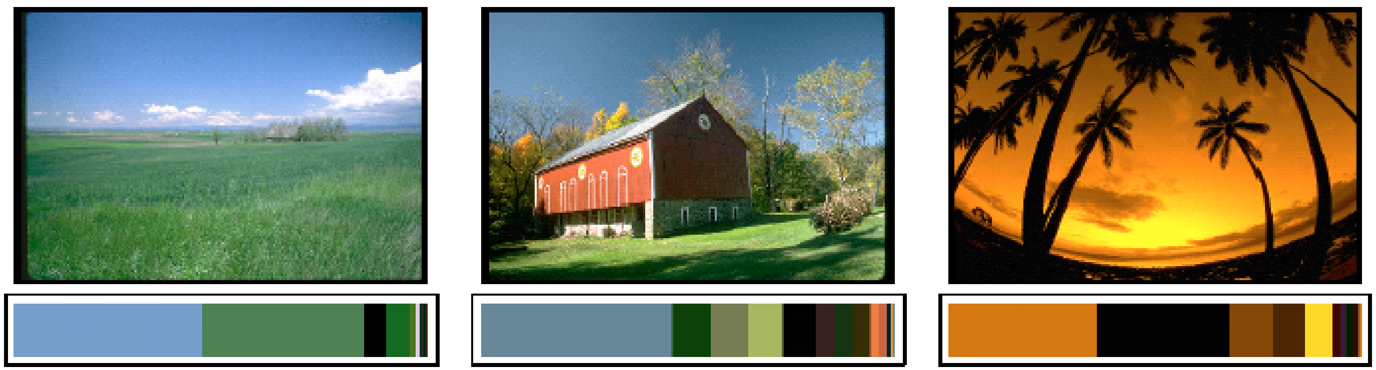
Improved measure includes normalization:

$$|Q-T|_{\gamma,d_1} \triangleq \sum_{i,j\in[m],k\in[d]} \left(\frac{|\gamma_{c_i c_j}^{(k)}(Q) - \gamma_{c_i c_j}^{(k)}(T)|}{1+\gamma_{c_i c_j}^{(k)}(Q) + \gamma_{c_i c_j}^{(k)}(T)} \right)$$

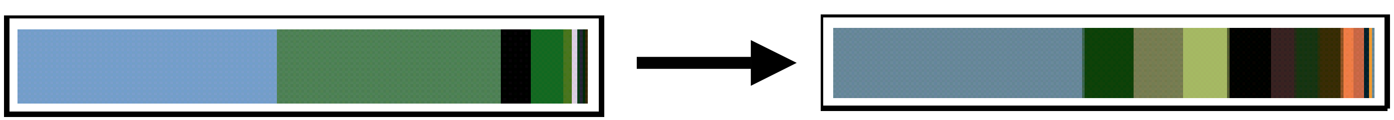
Earth Mover's Distance (EMD)

[Rubner, Guibas, & Tomasi 1998]

For each image, compute color signature:



Define distance between two color signatures to be the minimum amount of "work" needed to transform one signature into another.



Computing the Image Color Signature for EMD

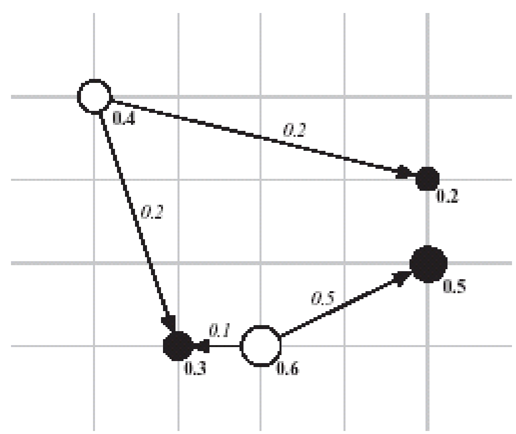
- Transform pixel colors into CIE-LAB color space.
- Each pixel of the image constitutes a point in this color space.
- Cluster the pixels in color space, *k-d* tree based algorithm. Clusters constrained to not exceed 30 units in L,a,b axes.
- Find centroids of each cluster.
- Each cluster contributes a pair (p,w_p) to the signature
  p is the average color.
  w_p is the fraction of pixels in that cluster.

Typically there are 8 to 12 clusters.

EMD of Color Signatures

The **work** needed to move a point, or a fraction of a point, to a new location is the portion of weight being moved, multiplied by the Euclidean distance between the old and new location.

Allow the weight of a single source to be partitioned among several destination points, and vice versa.



Can be solved with linear programming (see Rubner, et al.).

Multiscale EMD Formulation

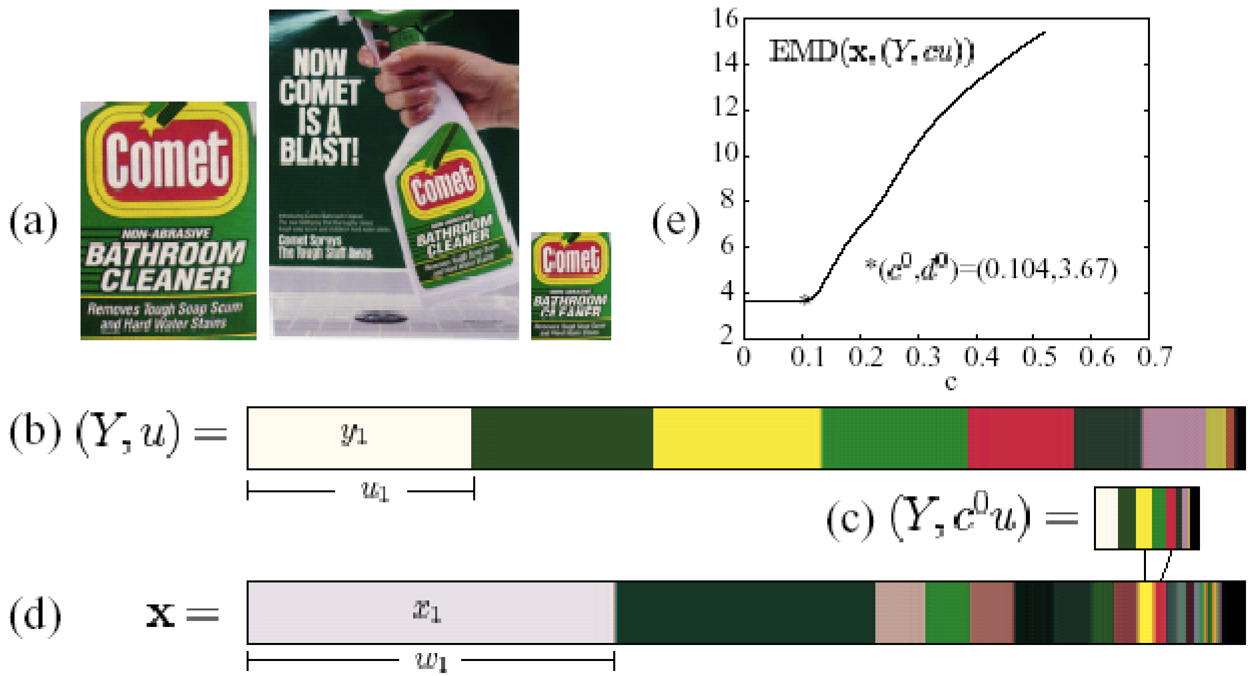
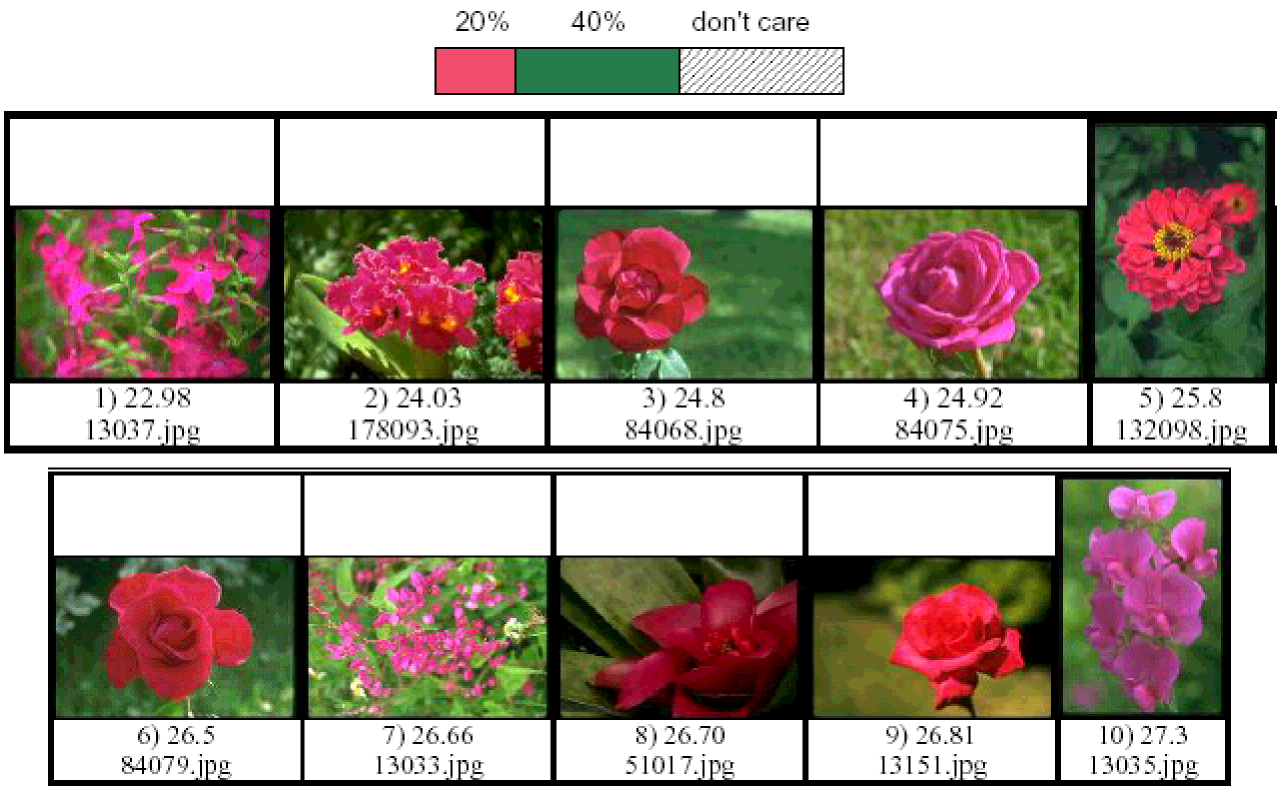
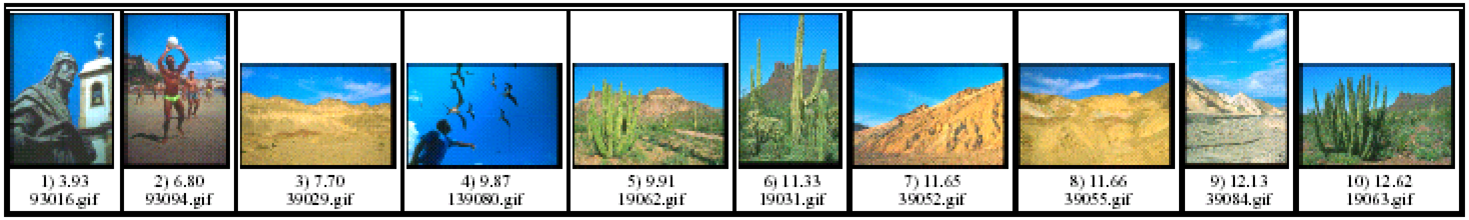


Figure 1. Scale Estimation. (a) pattern, image, and pattern scaled by the scale estimate c^0. (b),(d) pattern, image signatures. (c) pattern signature with weights scaled by c^0. (e) $\mathrm{EMD}(\mathbf{x}, (Y, cu))$ v. c.

Retrieval Example



Example EMD Retrieval Results



(a)

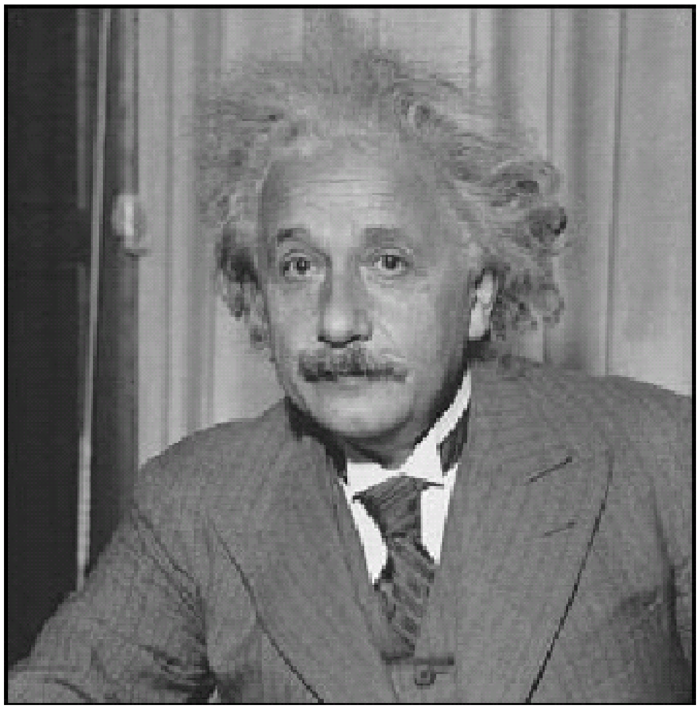
(b)

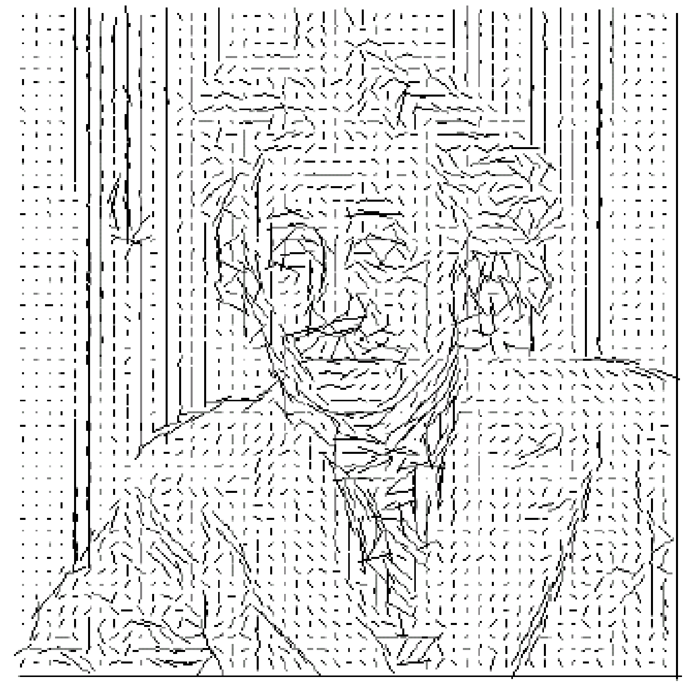
Figure 3: The top ten images for a query that asked for 20% blue and 80% don't care. (a) Traditional display. (b) MDS map.

Visualizing Dataset with EMD and Multidimensional Scaling

Orientation Histograms

Determine local orientation and magnitude at each pixel.





Example images taken from Freeman & Adelson, “The Design and Use of Steerable Filters,” IEEE T-PAMI, 1991.

Multiscale Orientation Histograms

Off-line, for each image:

```
Compute steerable pyramid
For each level in the pyramid
      For each pixel
            Estimate local orientation and magnitude
            If magnitude > threshold
                  increment appropriate histogram bin
            End
      End
      Circular blur histogram
End
```

Fast Multi-Resolution Image Querying

[Jacobs, Finkelstein, & Salesin SIGGRAPH 1995]

Off-line:

For each image

- Compute Haar wavelet decomposition
- Store truncated coefficients (top 60 of largest magnitude in each color channel)
- Quantize remaining coefficients to –1 or 1 (for negative or positive values)
- Store the coefficients as the image *signature*

end

Similarity Measure

Let Q and T represent just a single channel of the wavelet decomposition of the query and target images.

Define the similarity measure:

$$\| Q, T \| = w_{0,0} \, | Q[0,0] - T[0,0] | + \sum_{i,j} w_{i,j} \, | \tilde{Q}[i,j] - \tilde{T}[i,j] |$$

overall average intensity

truncated coefficients

$$\| Q, T \| \approx w_{0,0} \, | Q[0,0] - T[0,0] | + \sum_{i,j} w_{i,j} (\tilde{Q}[i,j] \neq \tilde{T}[i,j])$$

Only consider terms where query has non-zero wavelet coefficients.

Minka and Picard here?

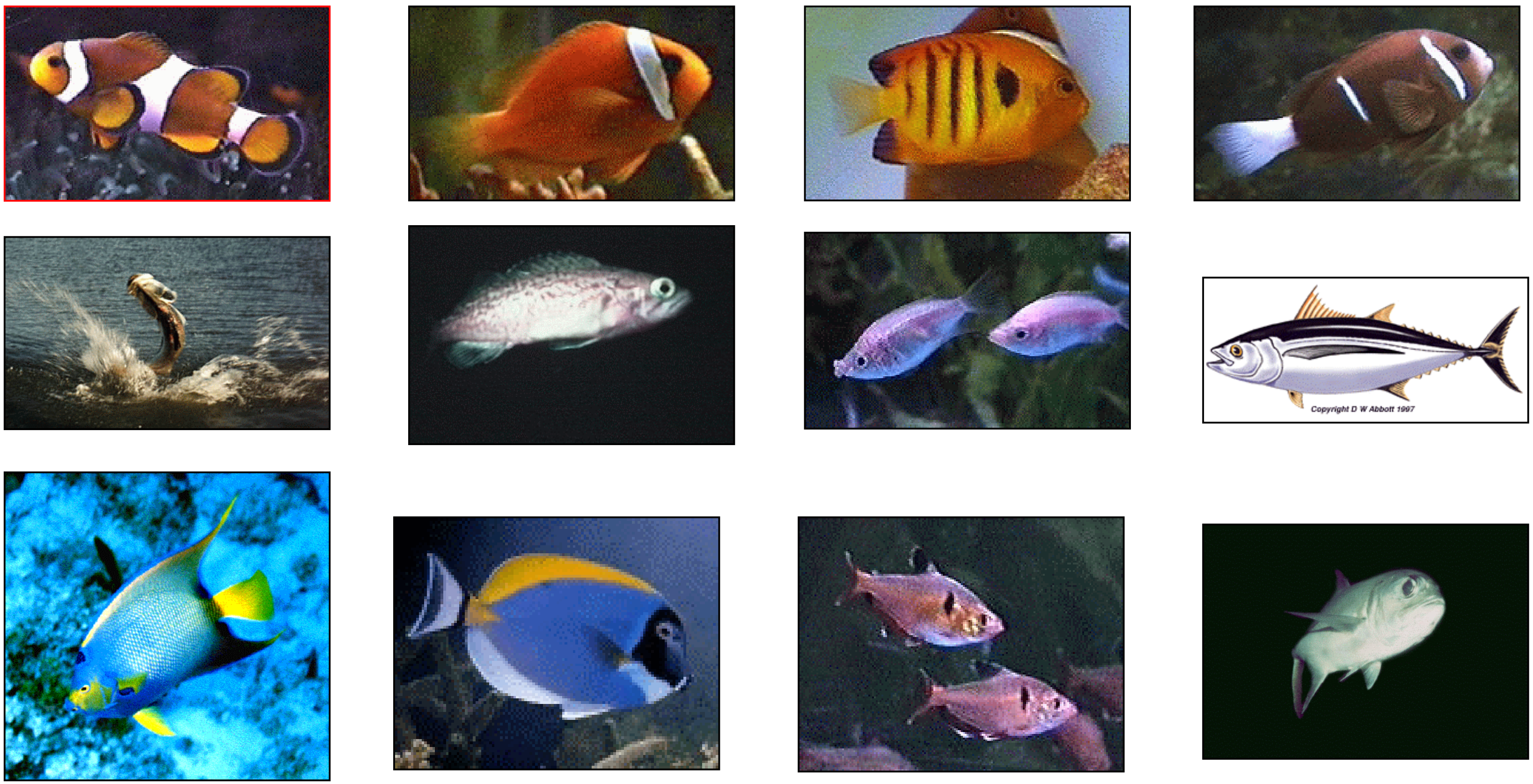
Image Representations

There are roughly three levels of image representation used for CBIR:

1. Iconic – exact pixel values
2. Compositional – overall image appearance
3. Objects – things depicted in the image, their properties, and their relationships

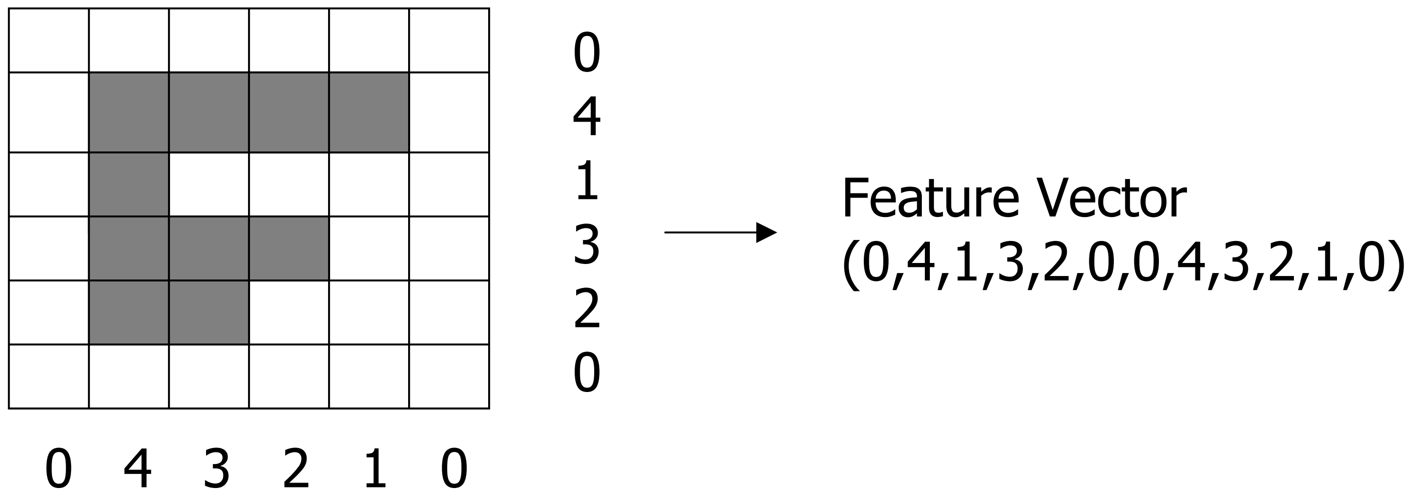
Shape-based retrieval of images

Find more shapes like this



Shape Properties: Projection Matching

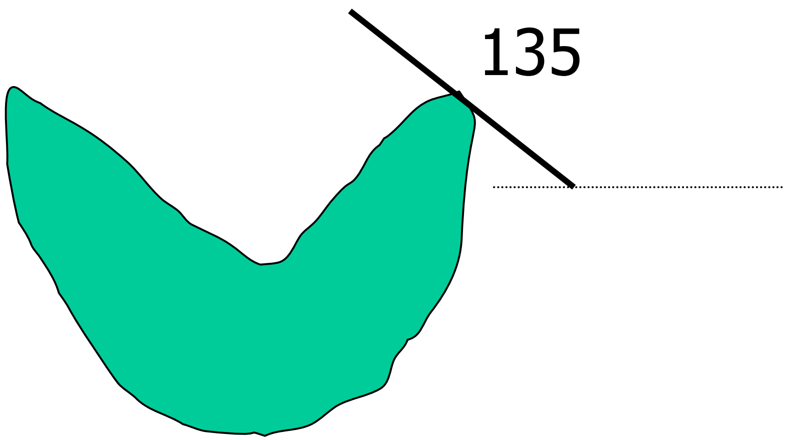
[VisualSeek, Smith&Chang 1996]

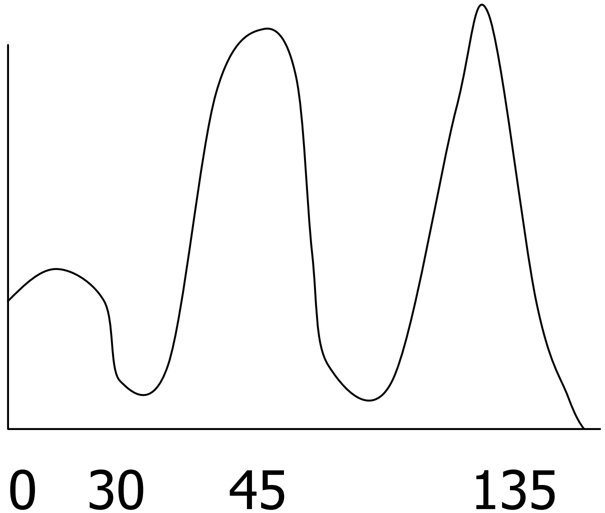


In projection matching, the horizontal and vertical projections form a histogram.

What are the weaknesses of this method? strengths?

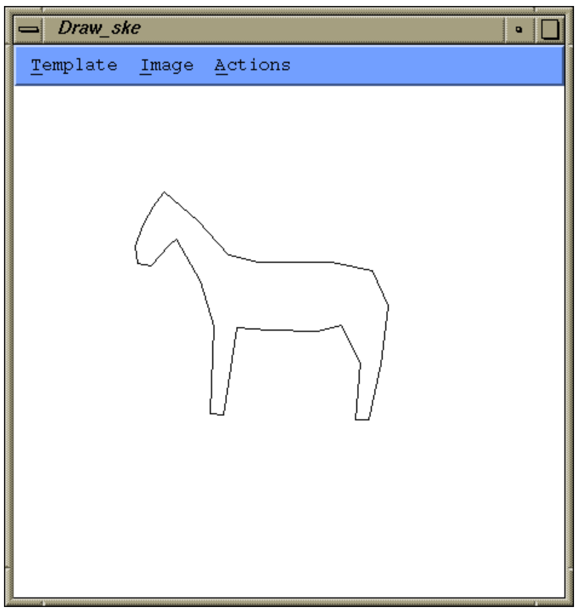
Global Shape Properties: Tangent-Angle Histograms



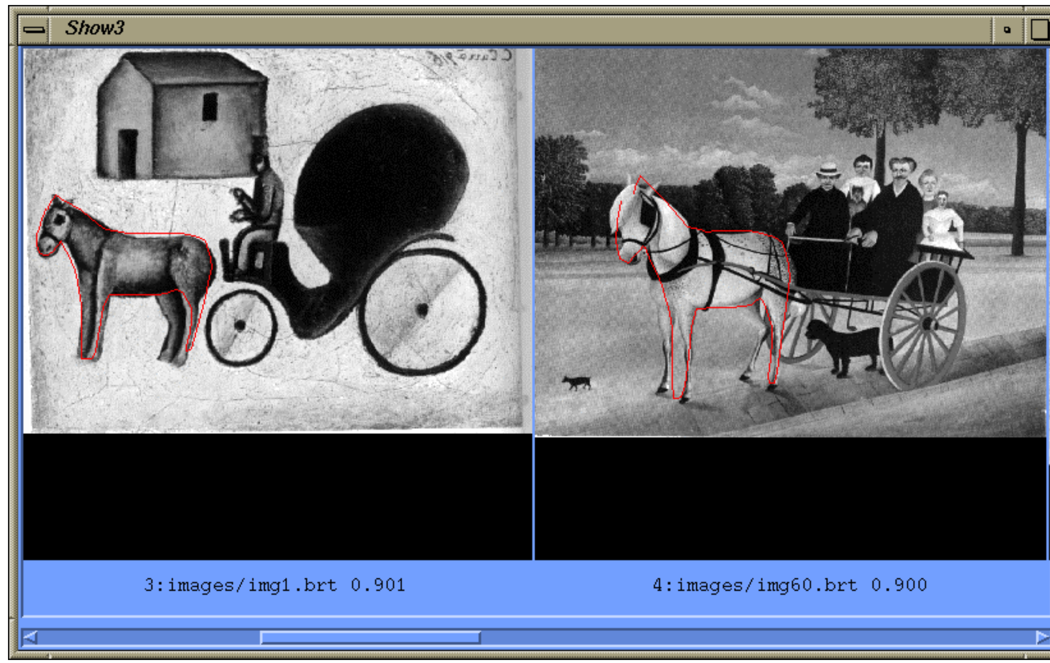


Is this feature invariant to starting point?

Del Bimbo Elastic Shape Matching



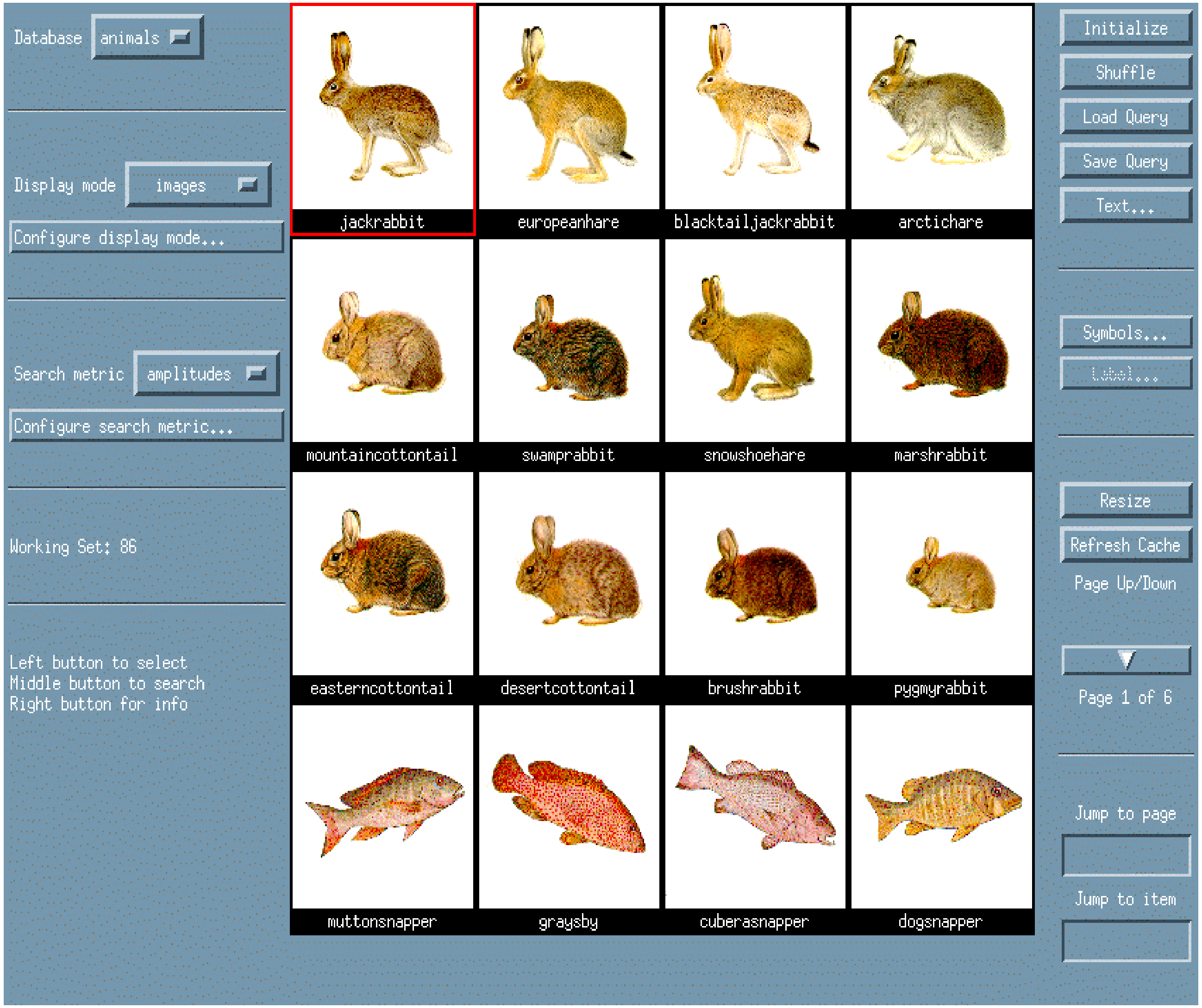
Sketch-based query



retrieved images

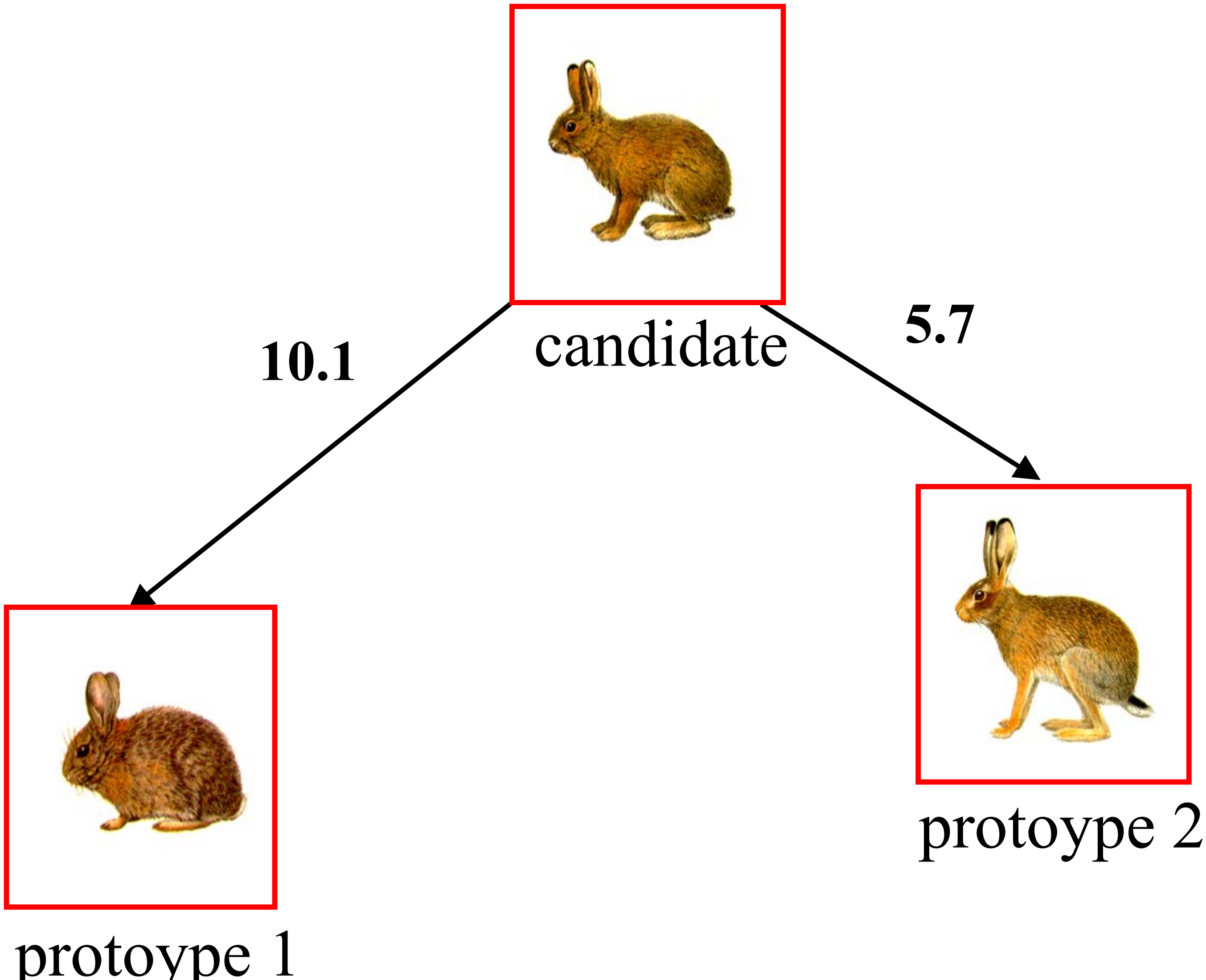
Shape-based Search in Photobook

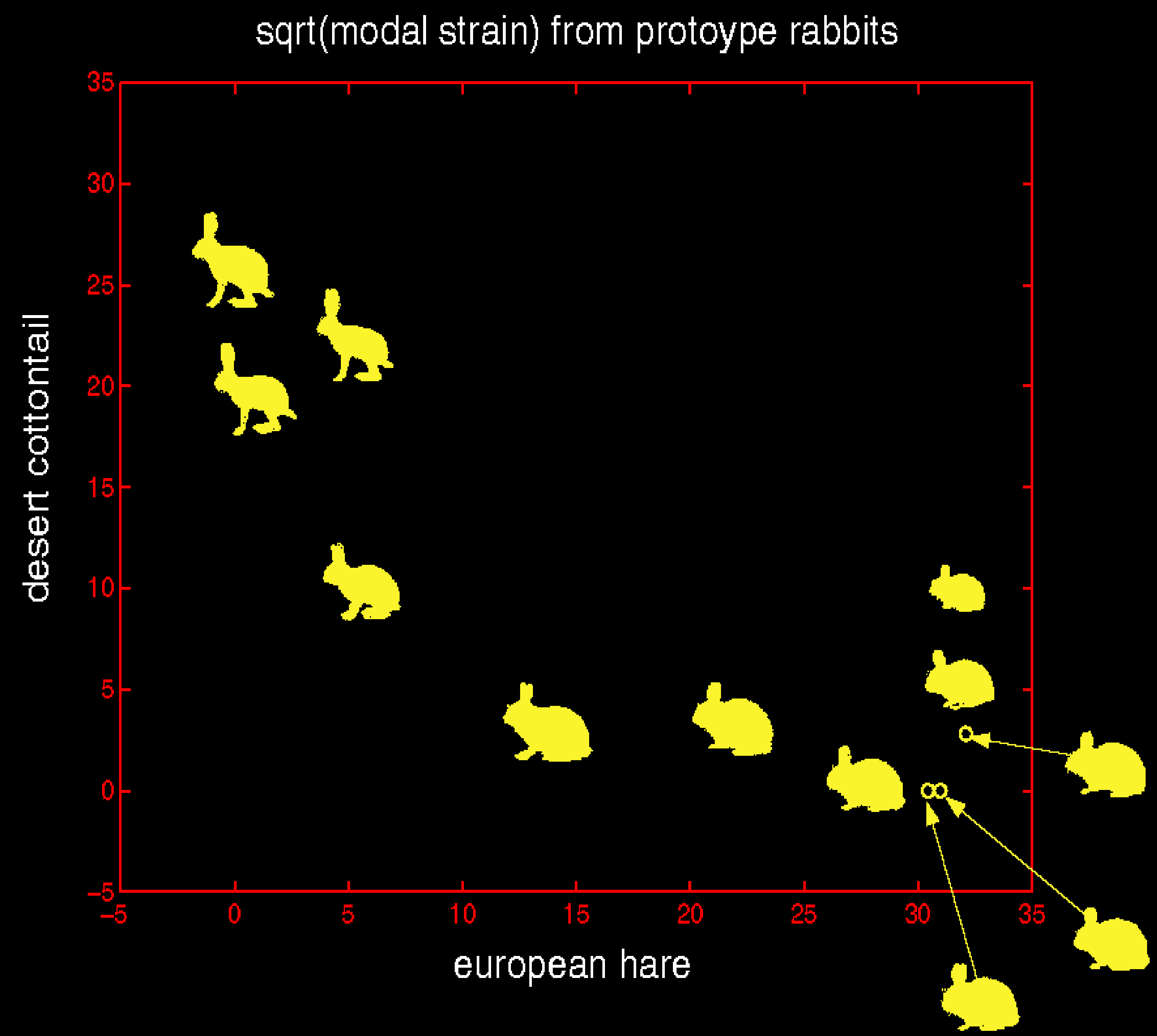
[Pentland, Picard, & Sclaroff 1994]



Coordinate in Deformable Prototype Space

[Sclaroff 1995]





sqrt(modal strain) from protoype rabbits
desert cottontail
european hare
35
30
25
20
15
10
5
0
−5
−5
0
5
10
15
20
25
30
35

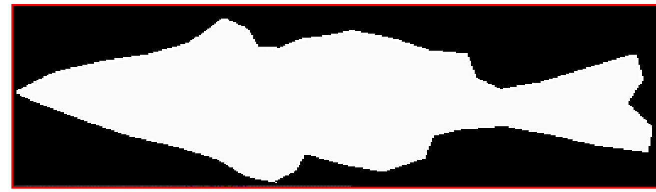
Problems in shape-based indexing

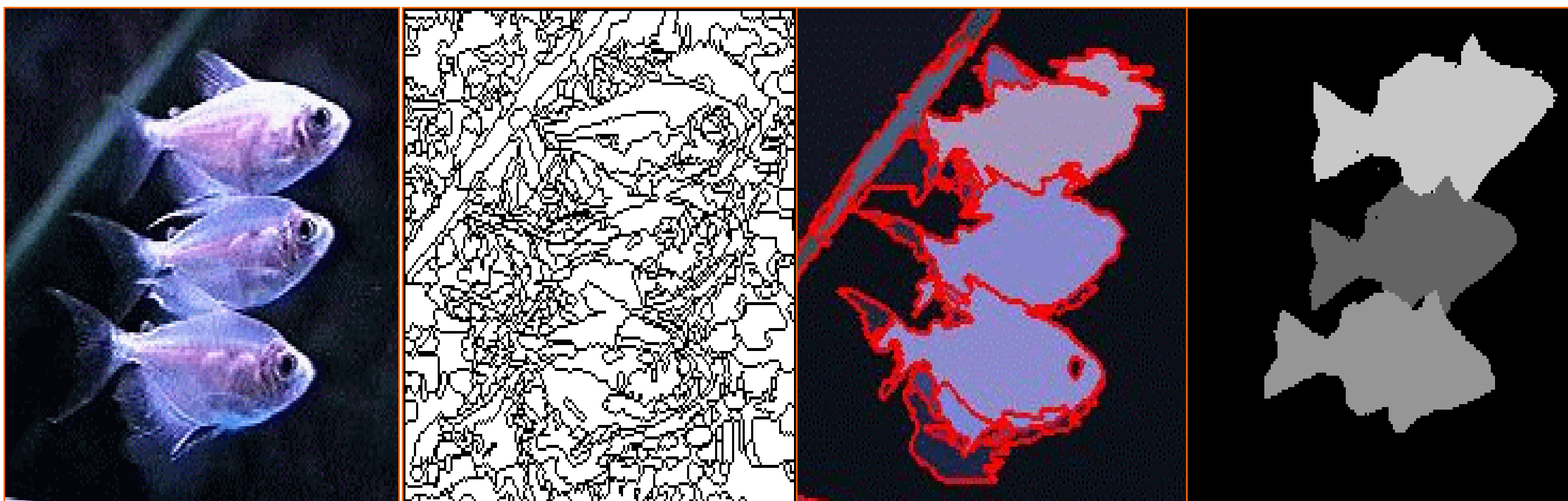
Many existing approaches assume

- segmentation is given, or…
- human operator circles object of interest, or…
- lack of clutter and shadows, or…
- objects are rigid, or…
- planar (2-D) shape models, or…
- models are known in advance

Deformable template-based region grouping
[Sclaroff&Liu, 2001]

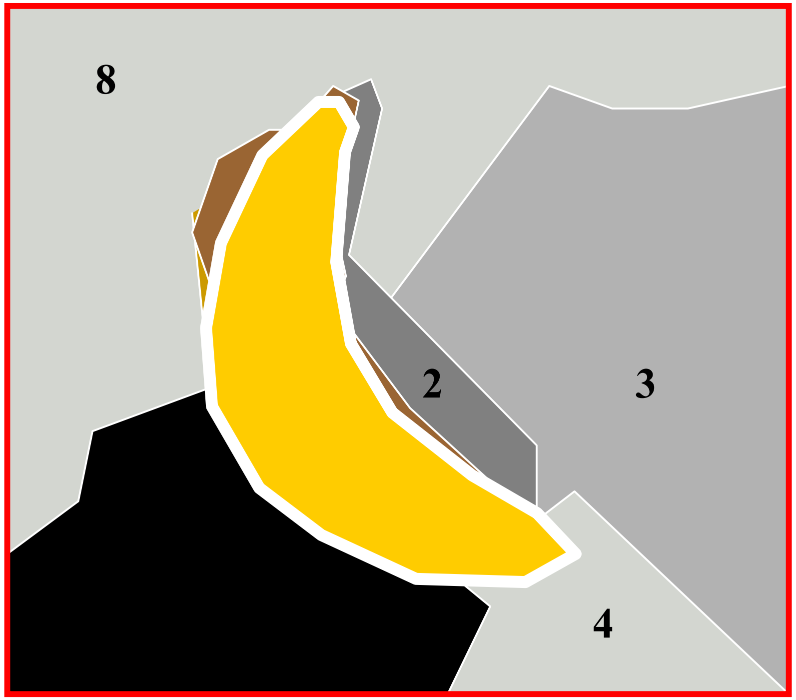
2D deformable template is trained on training data for object class

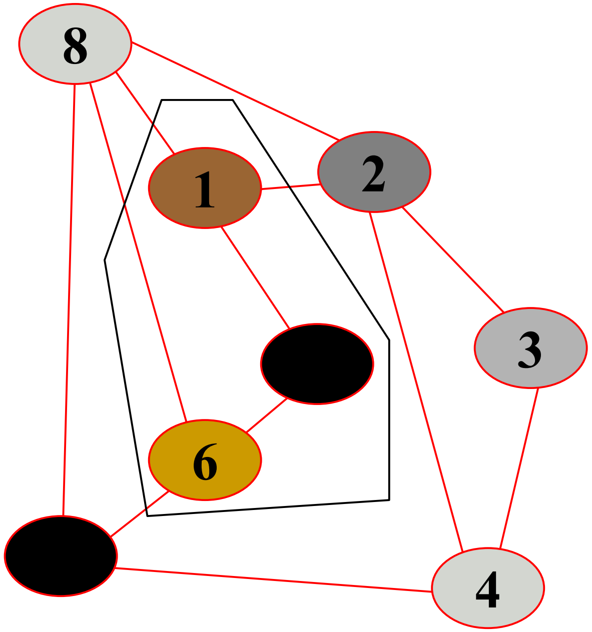




Input image | Over-segmentation | Model-guided merging | Model descriptions

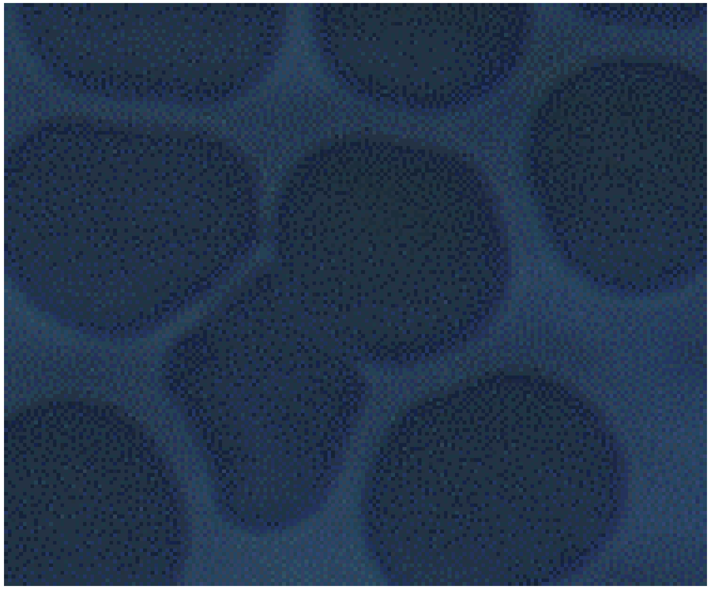
Image Partitioning via Optimization

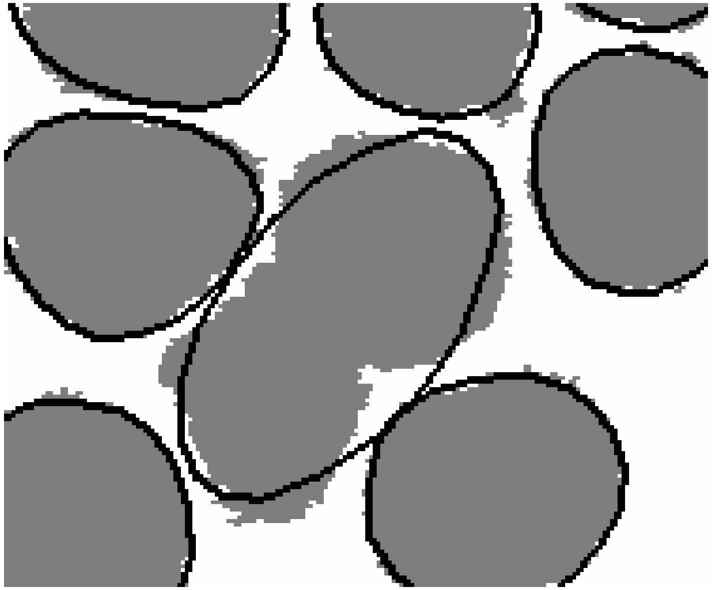




Model-based Region Splitting

- Detect candidates for splitting based on model fitting cost value and a specified threshold.
- Determine candidate cuts based on model and curvature extrema of the region group boundary.



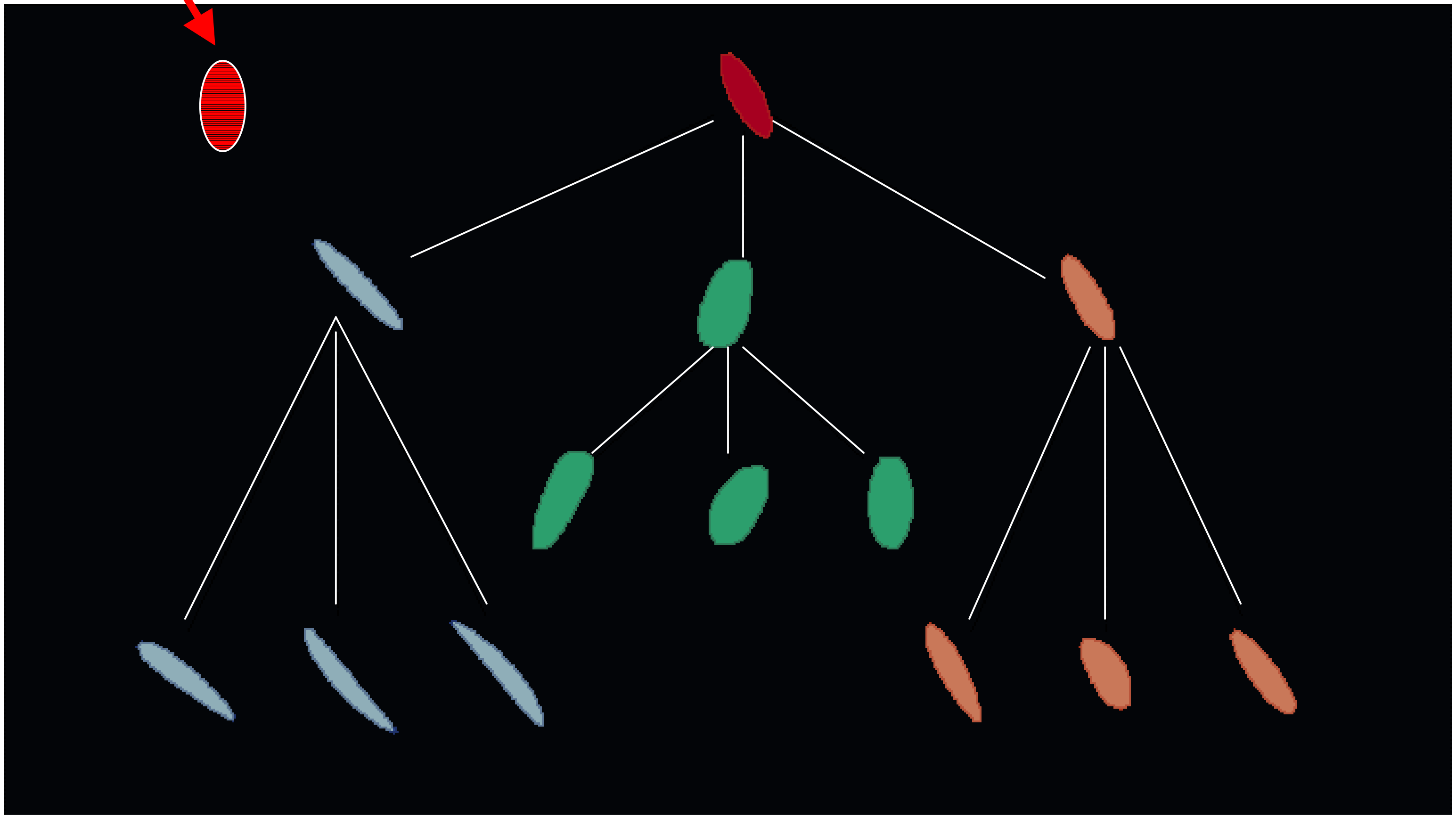


Index trees: basic idea

- Off-line:
  - Generate deformed instances of the object class
  - Compute their shape feature vectors
  - Create hierarchical indexing structure
- On-line:
  - Compute the shape feature vector for a potential region group
  - Fetch the most similar model instance via comparing the shape feature vectors

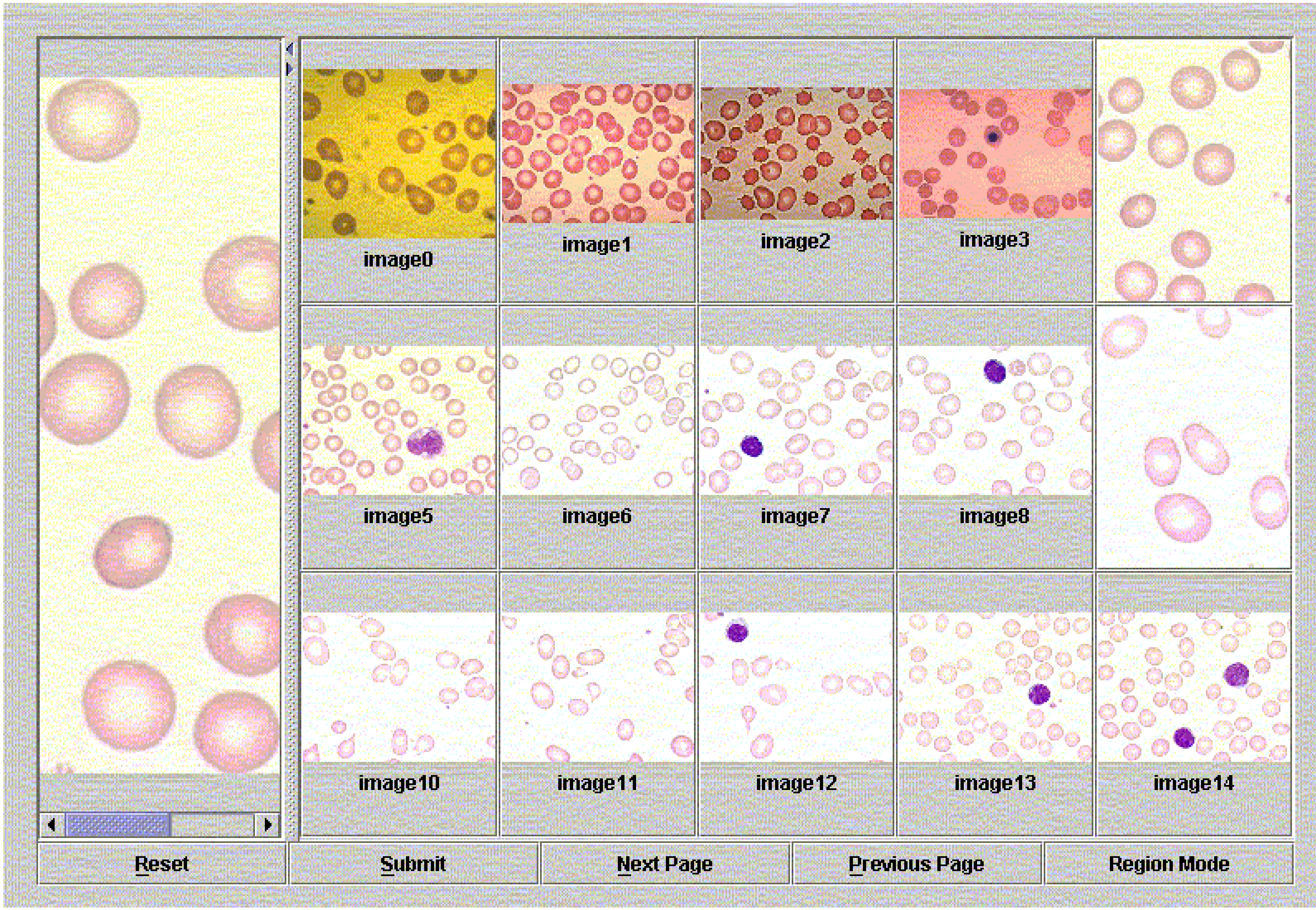
Shape Index Trees

Candidate region group



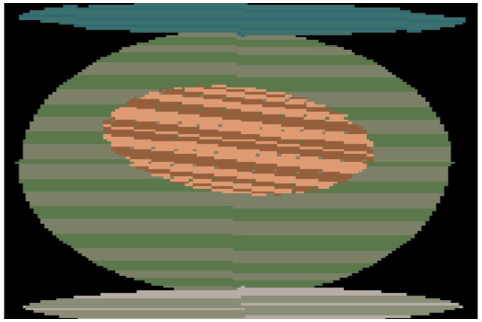
Shape-Population Retrieval

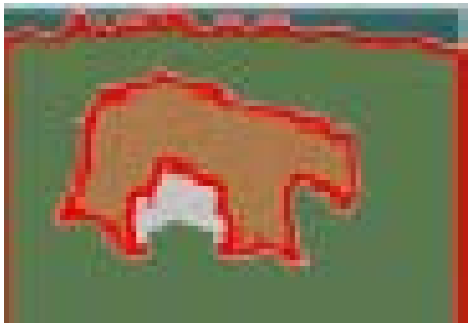
[Liu and Sclaroff 2000]



Blobworld

[Belongie, et al. 1998]





- Images are segmented on color plus texture
- User selects a region of the query image
- System returns images with similar regions
- Works really well for tigers and zebras

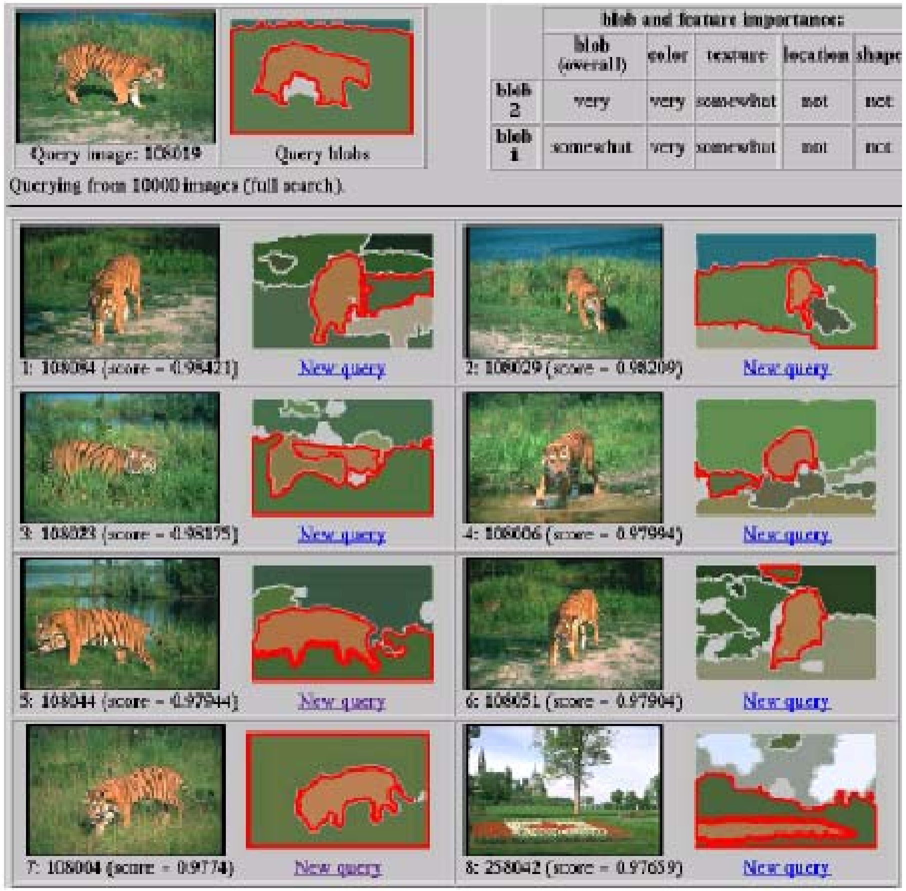
Demo: http://elib.cs.berkeley.edu/photos/blobworld

Blobworld Region Segmentation

- 8D descriptor computed for each pixel:
  - color in L*a*b* space
  - 3 texture features at selected scale: anisotropy, polarity, and contrast
  - pixel position (x,y)
- Represent each image as mixture of Gaussians, estimated via EM algorithm.
- Resulting pixel memberships form a segmentation of the image (after connected components analysis, etc.)

Example Blobworld Segmentation

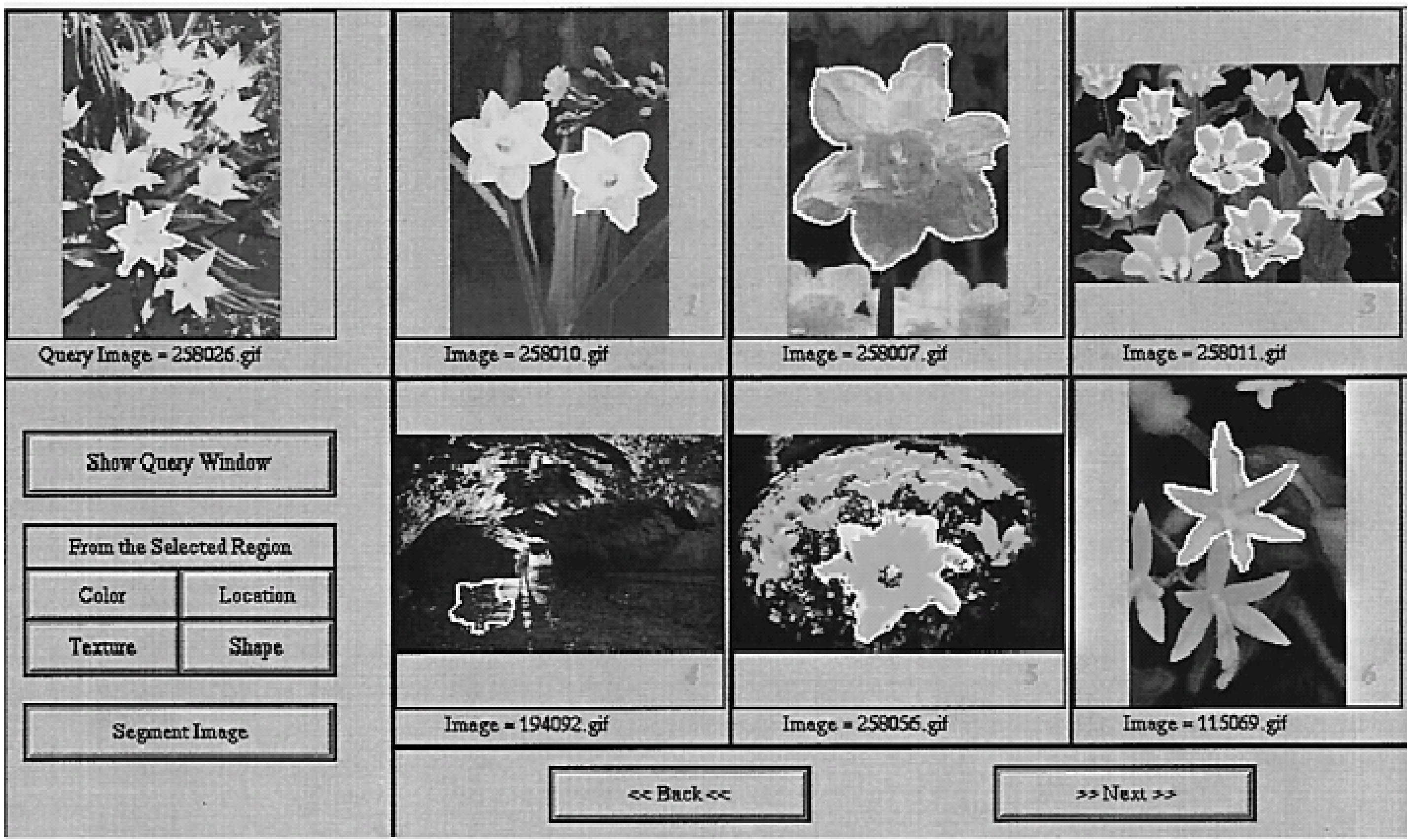
Example Blobworld Queries



The NeTra System

[Ma and Manjunath, 1999]

Retrieve by region color, texture, shape and position.



Demo: http://vision.ece.ucsb.edu/netra/Netra.html

The NeTra System Overview

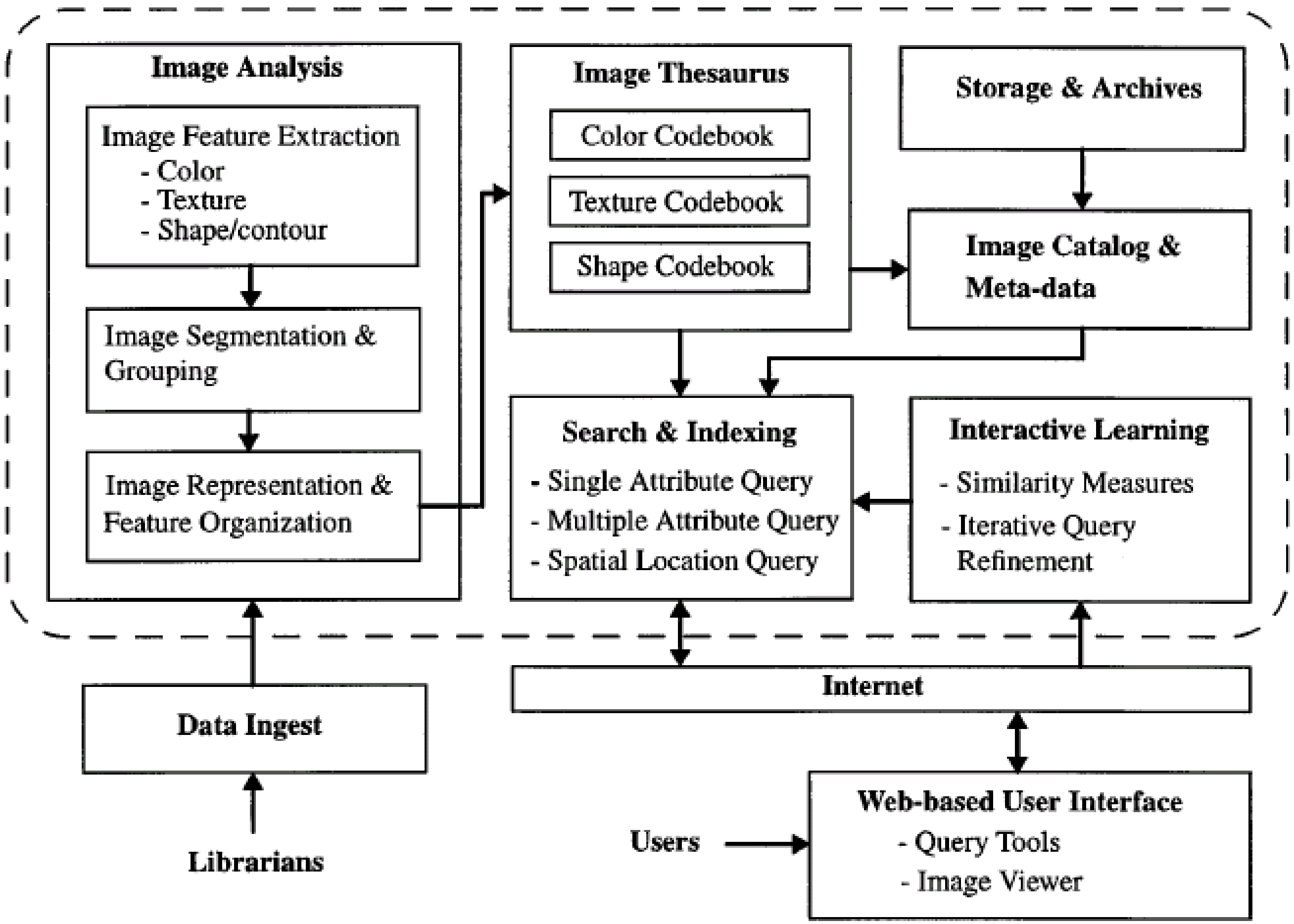
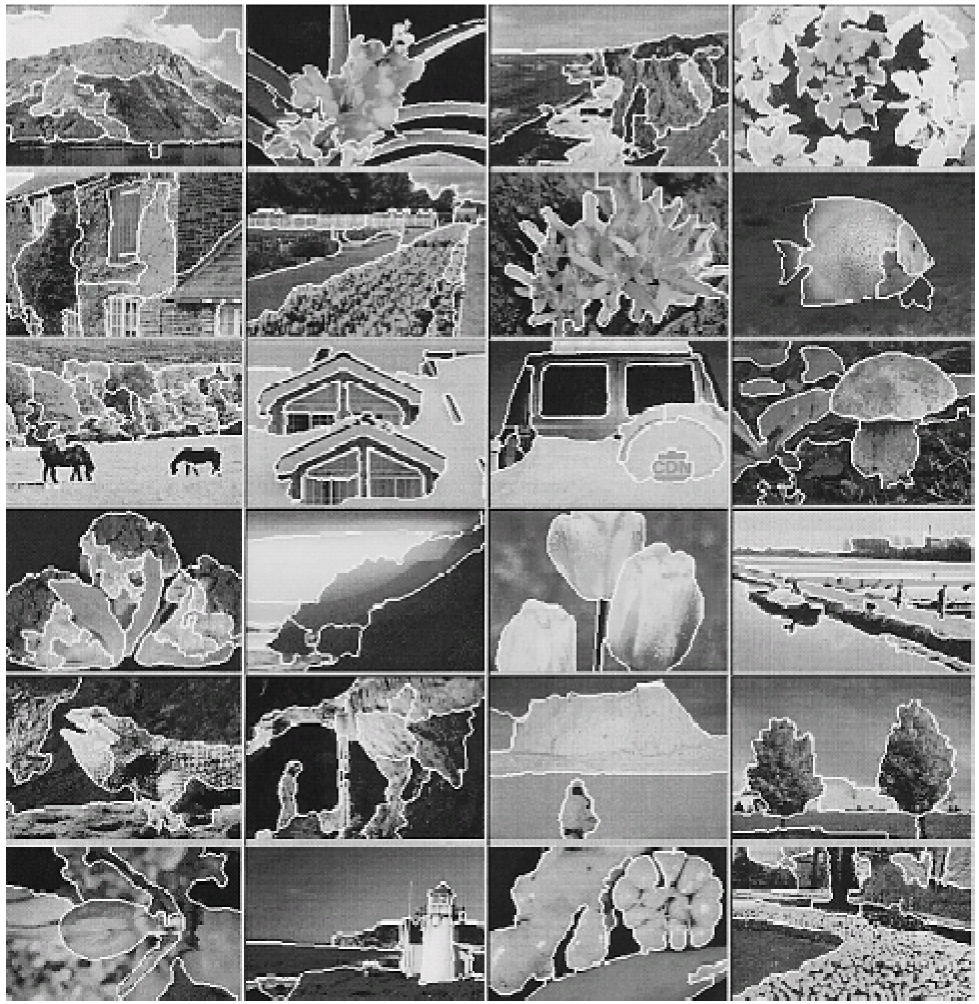


Image Features

- Use vector quantization (VQ) to build code book for RGB color, given training set of images chosen from database.
- Fourier descriptor of region contour is used to represent shape. Similarity measure: Euclidean.
- Gabor decomposition for texture at 4 scales, and 6 orientations. Store means and standard deviations, in 48-D feature vector. Similarity measure: Euclidean.

Example Segmentation Results



Taken from Ma & Manjunath, "NeTra: A toolbox for navigating large image databases," *Multimedia Systems*, 1999.

NeTra Indexing/Retrieval

- Color, texture, and shape are stored in separate index structures.
- Spatial location/size of regions represented by centroid and minimum bounding rectangle.
- Use quad-tree and/or R-trees to organize index for efficient queries.

Spatial Relationships

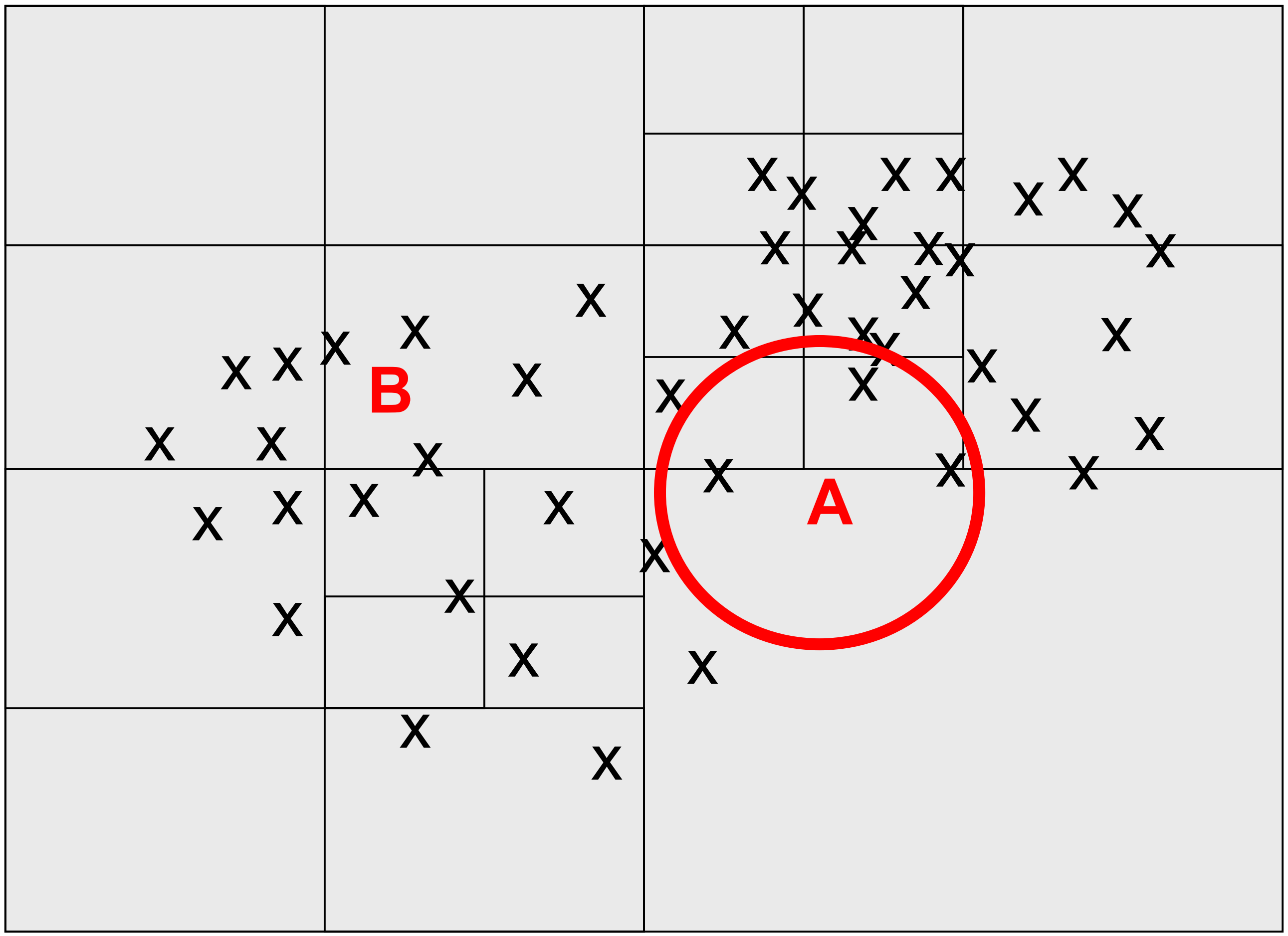
Example queries:

- Find all images where A is within 50 pixels of B.
- Find all images where A appears to the right of B.
- Find all images in where apples are on tables.

Appropriate indexing structures:

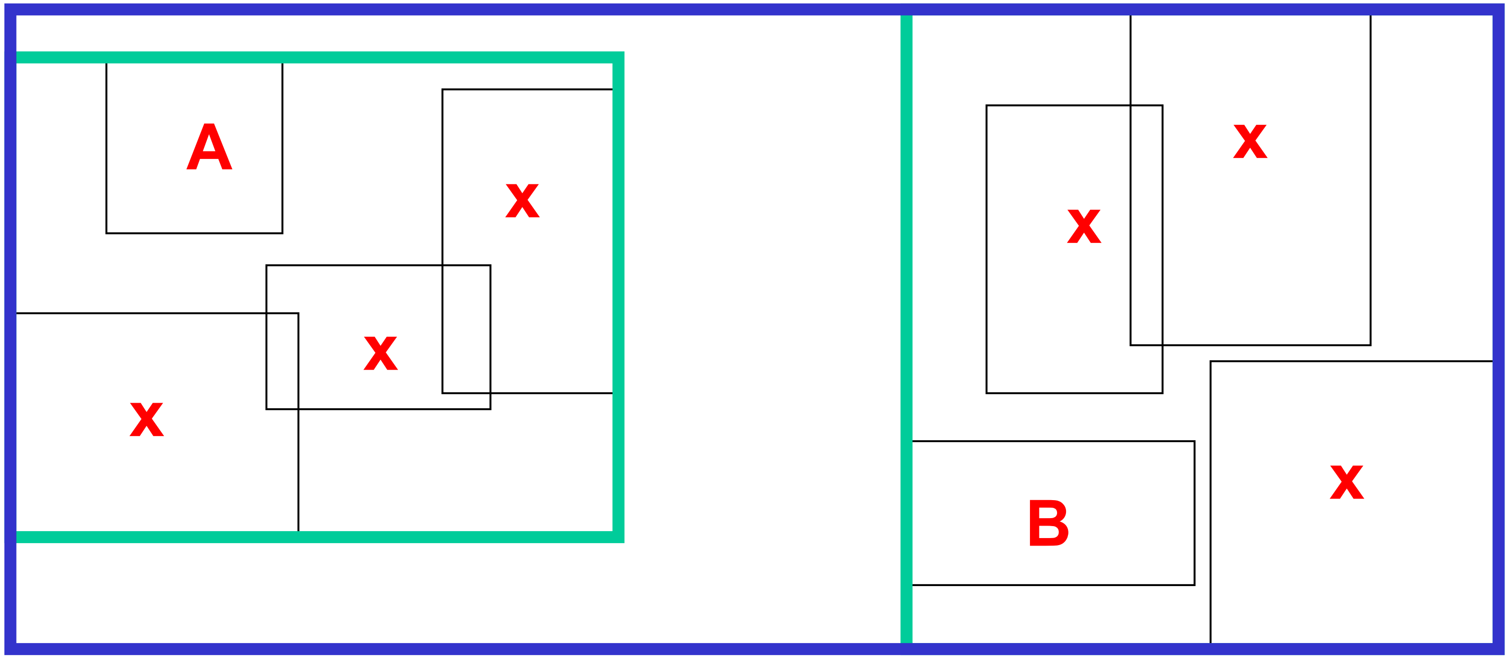
- Quad-trees
- R-trees, R*-trees
- K-d trees
- etc.

Quad-tree



Find all images where the centroid of A is within 50 pixels of B.

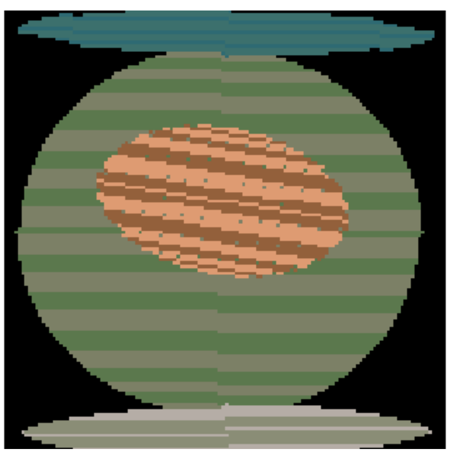
Hierarchical Minimum Bounding Rectangles



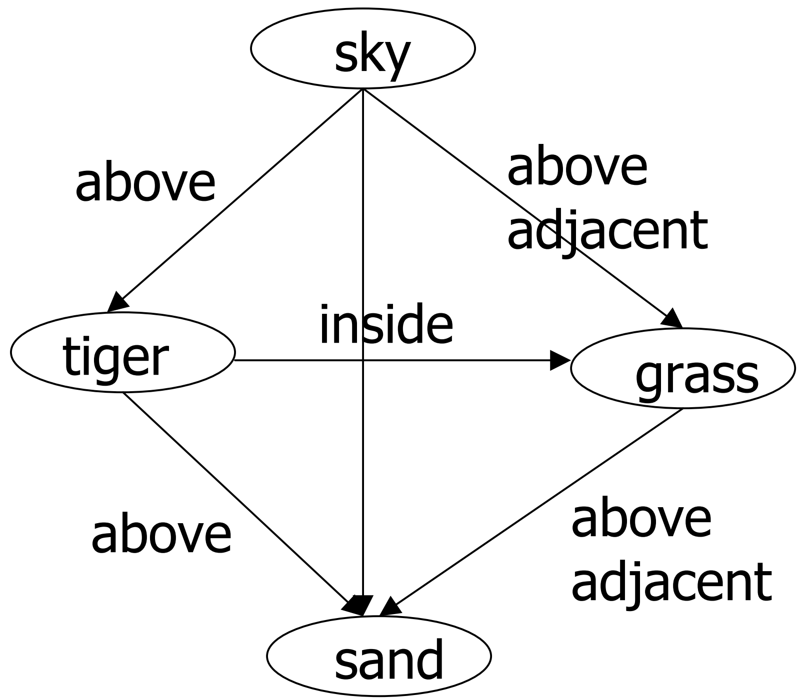
Find all images where A is within 50 pixels of B.

Region Relation Graph

image



abstract regions

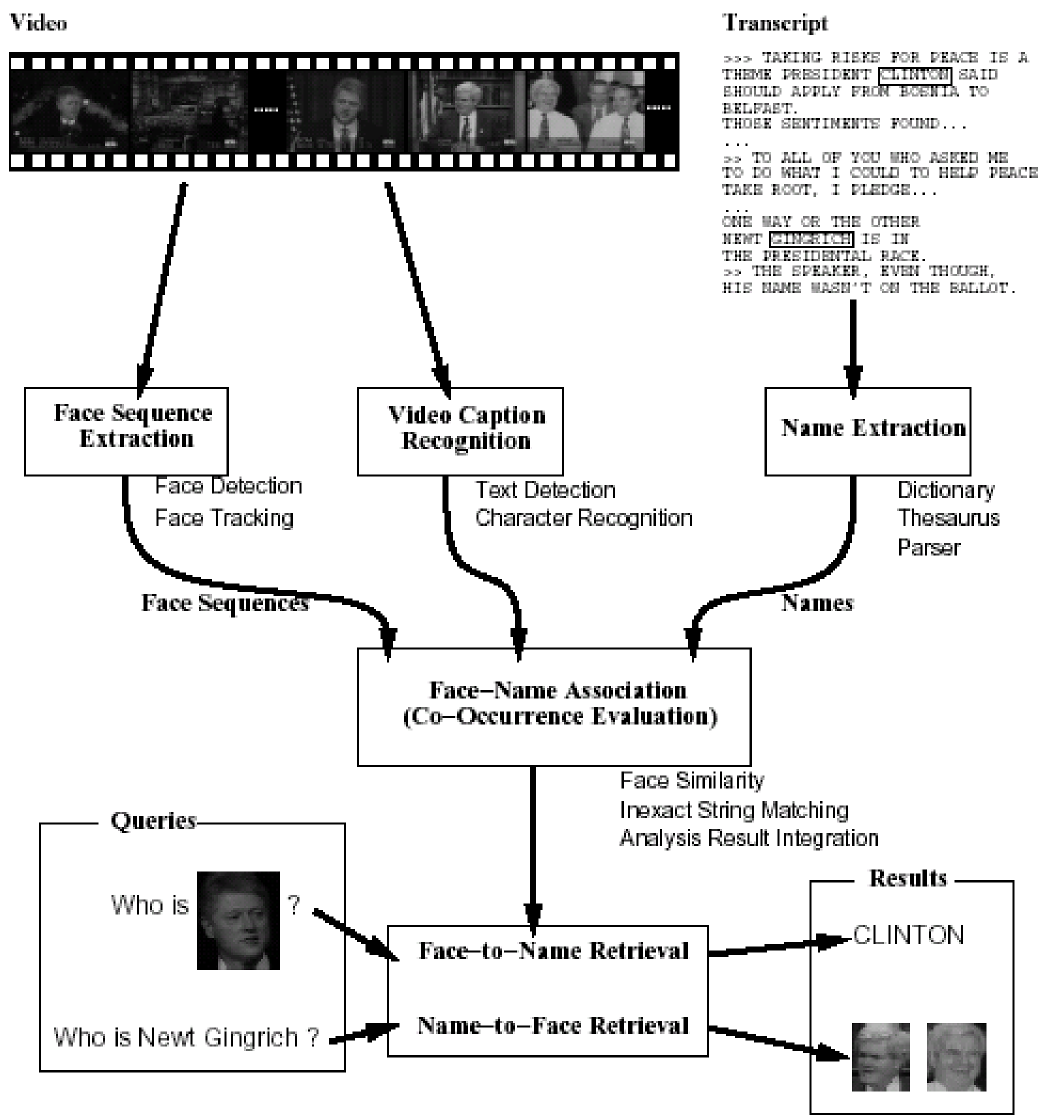


Eigenfaces in Photobook
[Pentland, Picard, Sclaroff 1995]

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

Name-It [Satoh, Nakamura, Kanade 1999]



Face Tracking and Frontal View Extraction

	start	end	frontal
(a)			
(b)			
(c)			
(d)			

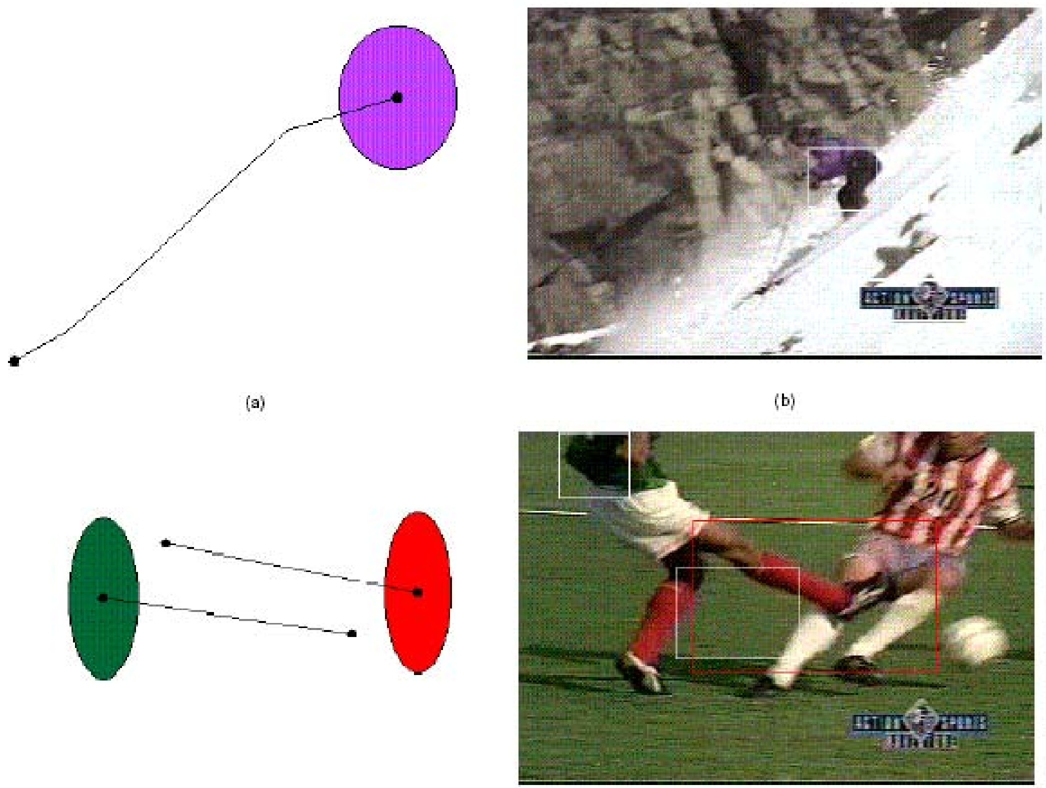
OCR for Video Captions

Figure 9: Typical Video Caption

Content-based Video Retrieval

VideoQ: Query by Sketch

[S.F. Chang, et al.]



Demo: http://www.ctr.columbia.edu/videoq/

Shot Boundary Detection Methods

Assumption: shot boundaries are discontinuities in space-time.

- Compare color and orientation histograms in adjacent frames
- Motion (flow) analysis
- Multimodal approach: video + audio track

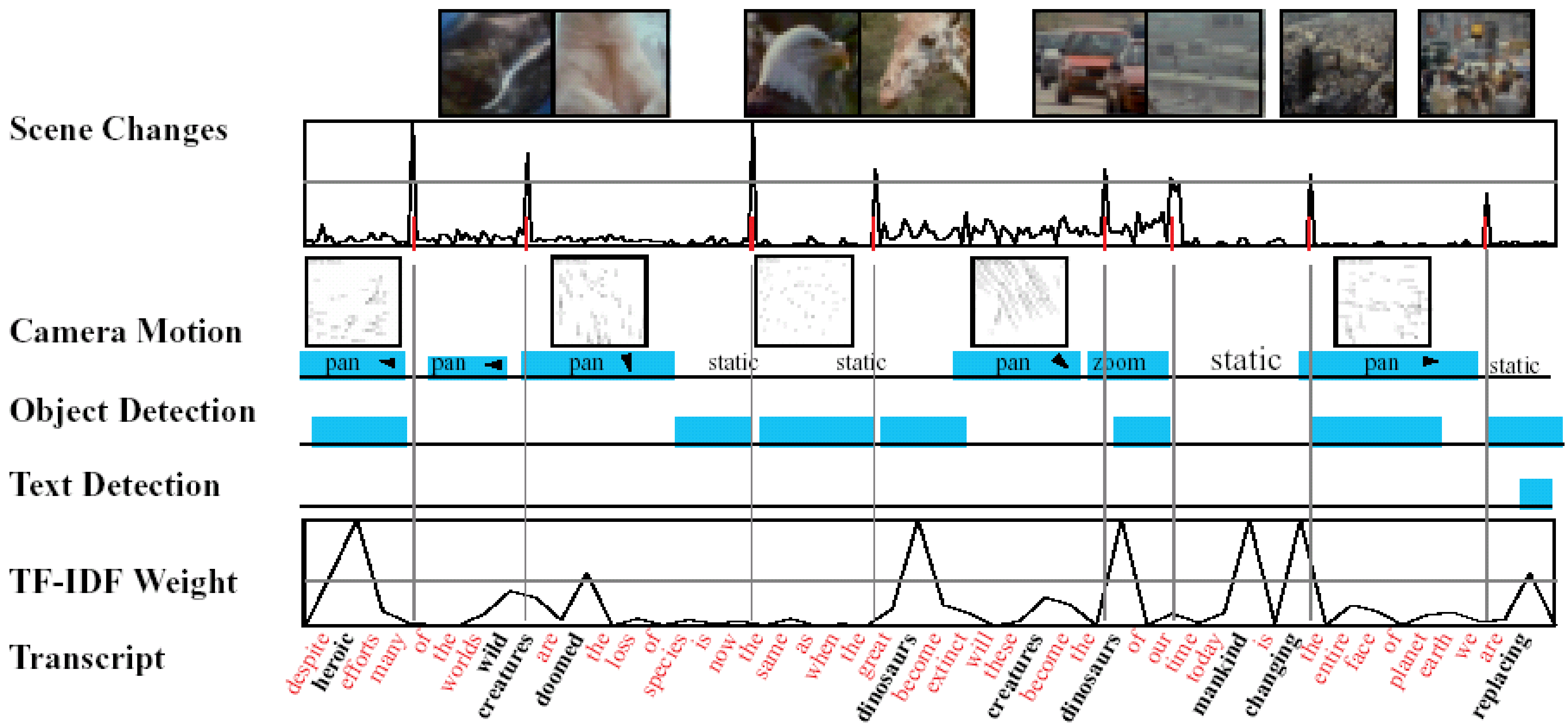
Other applications:

Keyframe extraction. Intelligent fast forward.

Problems:

- Transitions like wipe, fade, cross-dissolve
- Camera motions: pan, zoom, etc.
- Moving objects occupy large percentage of image

Video Skimming (CMU Informedia)



Event Detection, Indexing and Retrieval

Assign semantic labels to significant events in video:

- Explosion, car crash, door slam (audio/video track)
- Marilyn Monroe enters scene
- Pele scores goal
- Jay Leno tells joke and then delivers punch-line
- Two people exchange a briefcase in park
- etc.

There are events that are "latent" in the video database that are not of interest now, but may become interesting later.

Relevance Feedback

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 ImageRover Project at Boston U.
- The MARS Project (Tom Huang’s group at Illinois)
- PicHunter (Cox, et al. at NEC)

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

Term weighting in Info Retrieval

- 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*
- Target document *i* can be represented as a weight vector in the term space

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

Term weighting

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

$$Q = [w_{q1}; w_{q2}; \ldots; w_{qN}]$$

- The similarity between T and Q

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

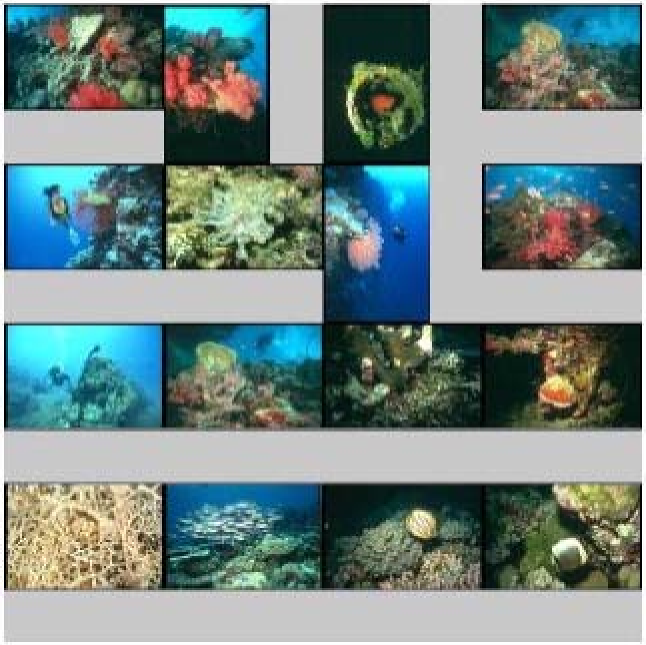
Using Relevance Feedback

- 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.

What are the problems in applying the IR relevance feedback paradigm in image and video retrieval?

Clustering Images [Barnard&Forsyth 2001]



Cluster on text only.

Cluster on image features only.

Clustering Images

Two clusters obtained using both text and image segment features.

Image/Video Databases

- Since databases can be large, computational complexity is very important
- Spatial data structures can help
- Hierarchical data structures, clustering
- Multiple metric strategies
- Embeddings