Content-Based Image Retrieval

ECE 547

Image Processing

Fall 2004

Charlie Dagli

dagli@uiuc.edu

Today's Topics

- What is CBIR ?
- Image Features
- Feature Weighting and Relevance Feedback
- User Interface and Visualization

What is Content-based Image Retrieval (CBIR)?

- Image Search Systems that search images by image content
  <-> Keyword-based Image Retrieval
  (ex. Google Image Search)

Applications of CBIR

- Consumer Digital Photo Albums
  - Digital Cameras
  - Ex. WWMX by Microsoft Research
- Medical Images
- Digital Museum
- Trademarks Search
- MPEG-7 Content Descriptors

Basic Components of CBIR

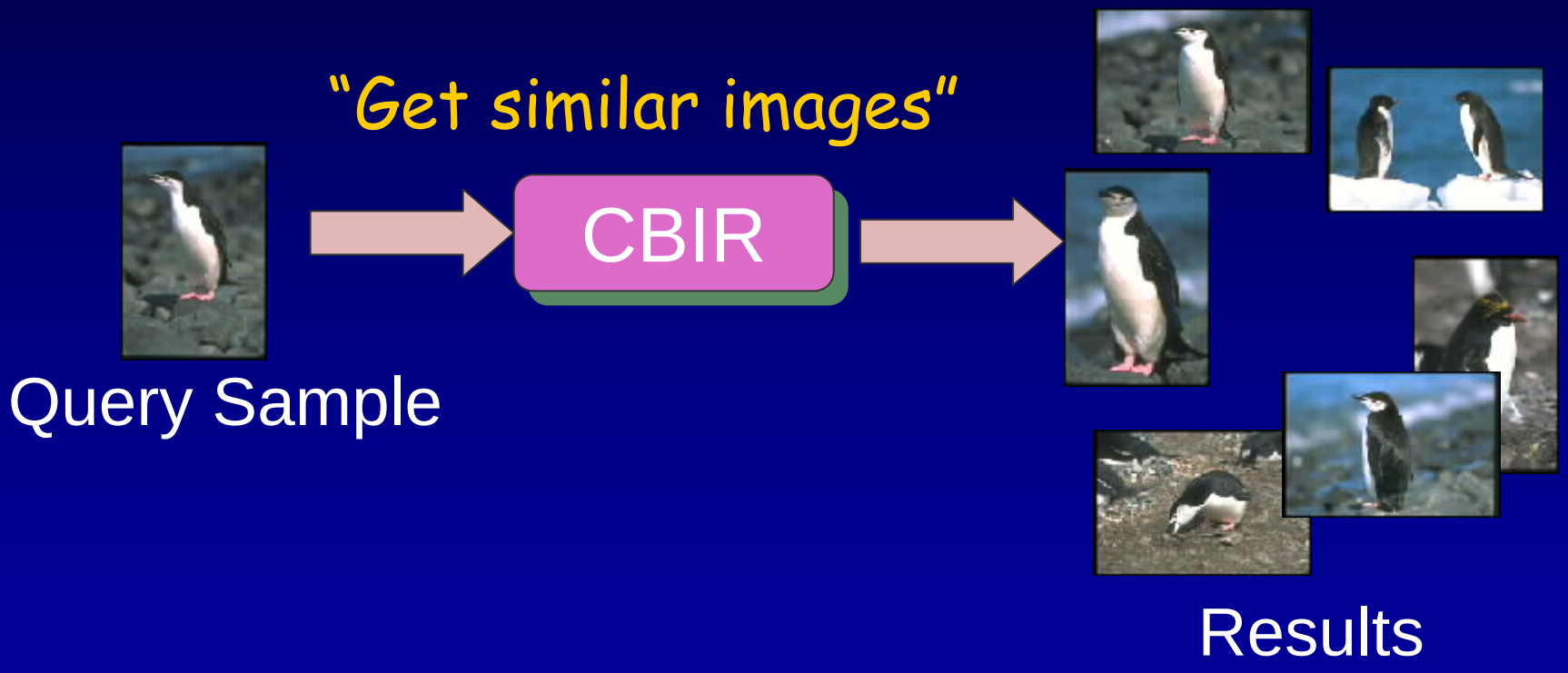
- Feature Extractor
  - Create the metadata
- Query Engine
  - Calculate similarity
- User Interface

How does CBIR work ?

- Extract Features from Images
- Let the user do Query
  - Query by Sketch
  - Query by Keywords
  - Query by Example
- Refine the result by Relevance Feedback
  - Give feedback to the previous result

Query by Example

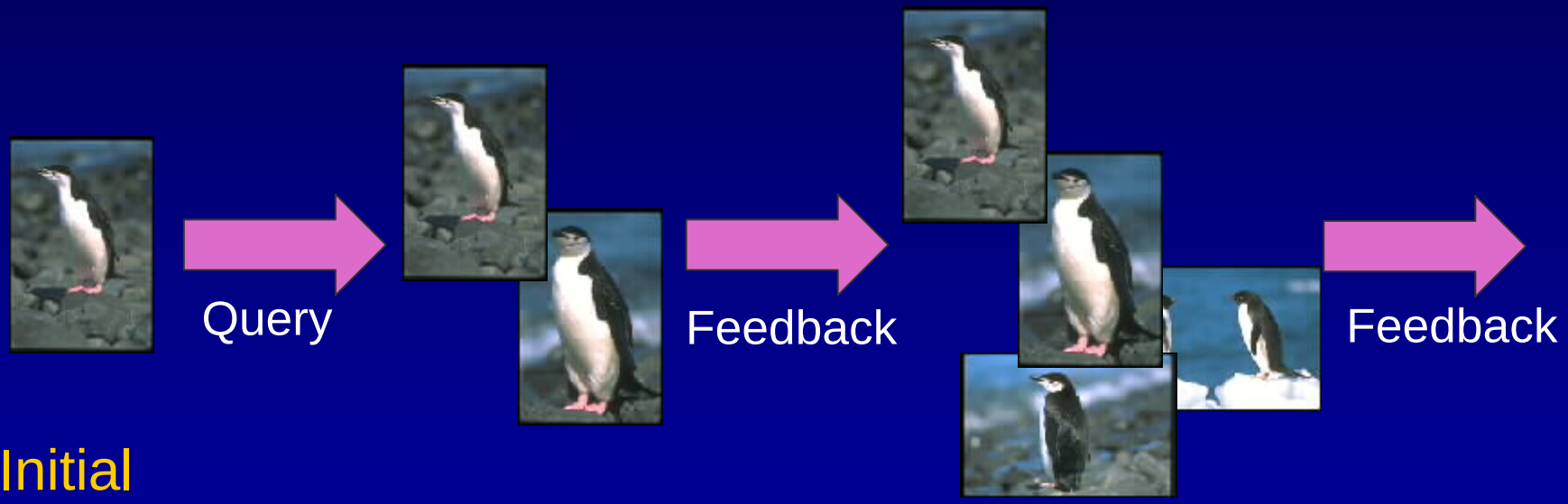
- Pick example images, then ask the system to retrieve "similar" images.



What does "similar" mean?

Relevance Feedback

- User gives a feedback to the query results
- System recalculates feature weights



Basic Components of CBIR

- Feature Extractor
  - Create the metadata
- Query Engine
  - Calculate similarity
- User Interface

Image Features (Metadata)

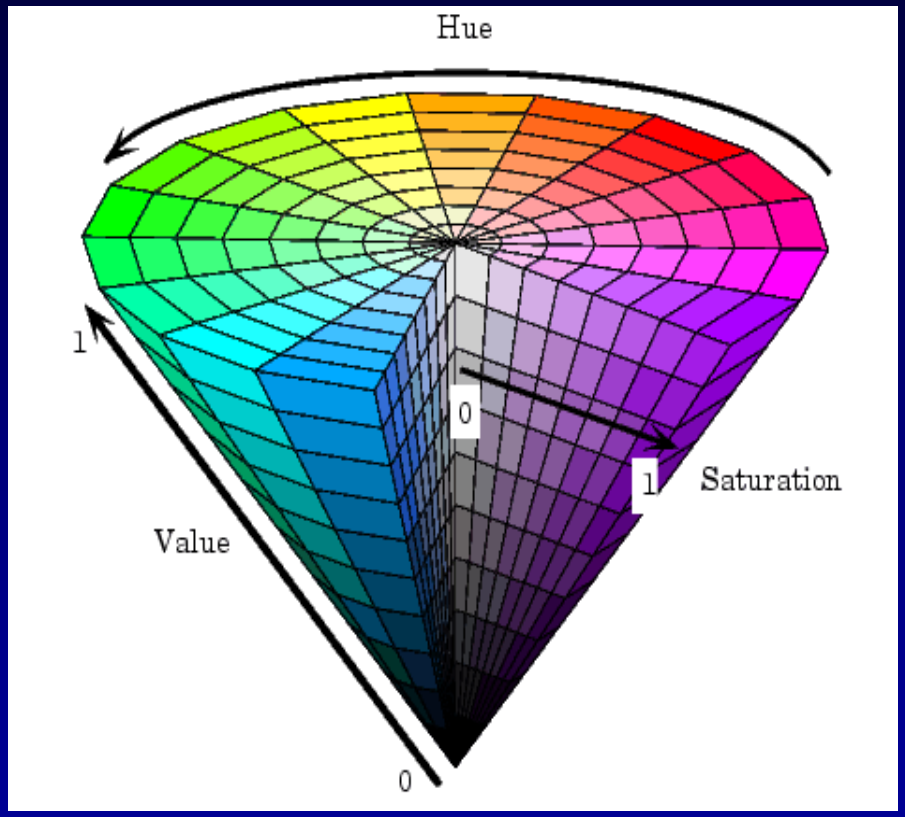
- Color
- Texture
- Structure
- etc

Color Features

- Which Color Space?
  - RGB, CMY, YCrCb, CIE, YIQ, HLS, …
- Our Favorite is HSV
  - Designed to be similar to human perception

HSV Color Space

- H (Hue)
  - Dominant color (spectral)
- S (Saturation)
  - Amount of white
- V (Value)
  - Brightness



How to Use This?

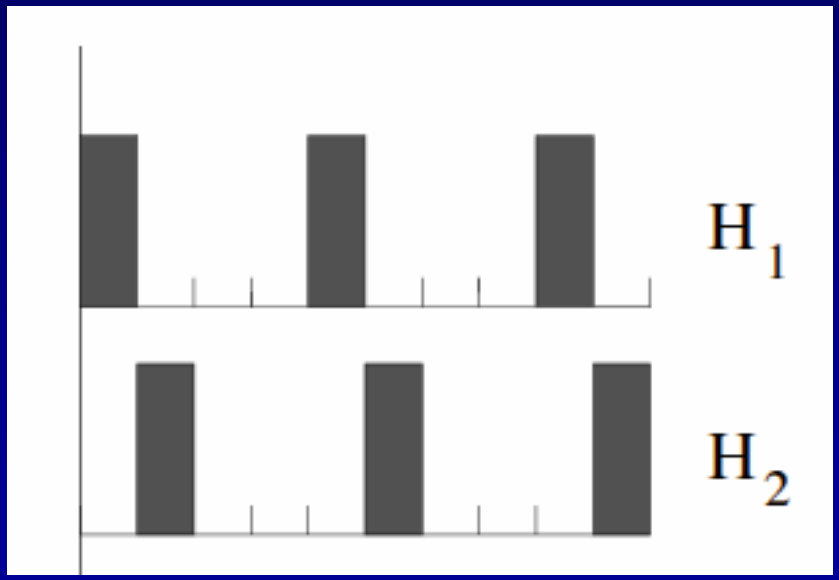
Munehiro Nakazato and Charlie Dagli

Straightforward way to use HSV as color features

- Histogram for each H, S, and V
- Then compare in each bin
- Is this good idea?

Are these two that different?

- Histogram comparison is very sensitive



Color Moments [Stricker '95]

- For each image, the color distribution in each of H, S and V is calculated
  - 1st (mean), 2nd (var) and 3rd moment for HSV

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2\right)^{1/2}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3\right)^{1/3}$$

i : color channel {*i*=h,s,v}
N = # of pixels in image
Total 9 features

Shape Features

- Region-Based Shape
  - Outer Boundary
- Contour-Based Shape
  - Features of Contour
- Edge-Based Shape
  - Ex. Histogram of edge length and orientation

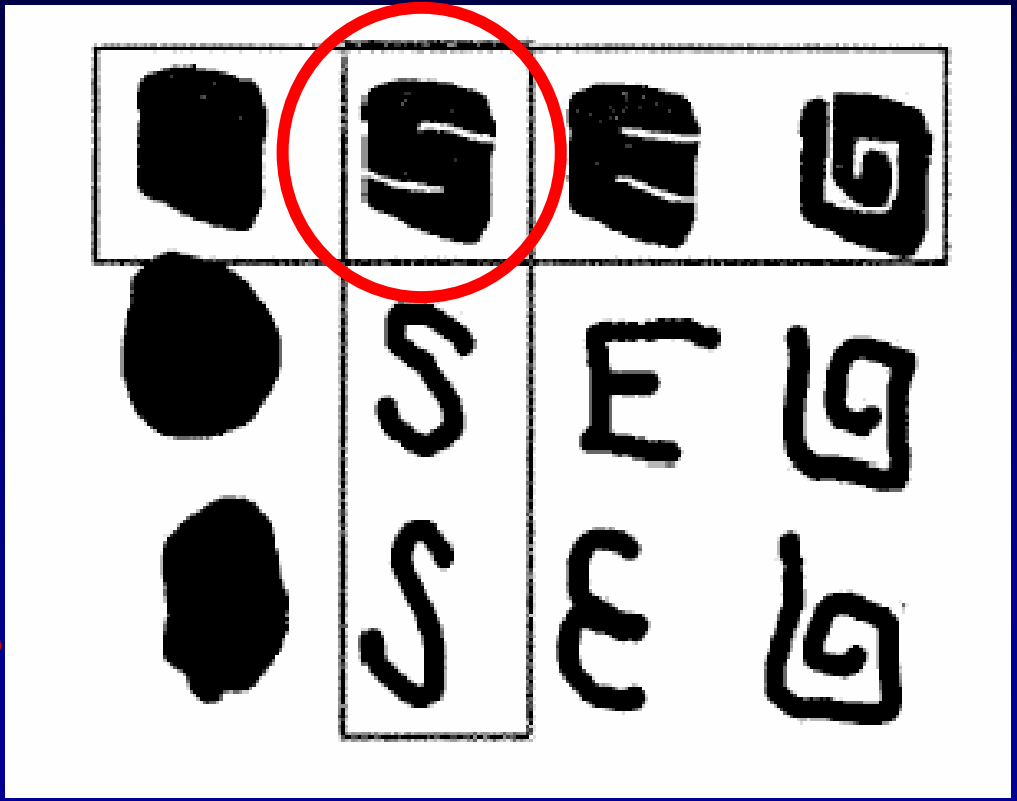
Region-based vs. Contour-based

- Region-based
  - Suitable for Complex objects with disjoint region
- Contour-based
  - preserve semantics

Region-based vs. Contour-based

Region

Contour

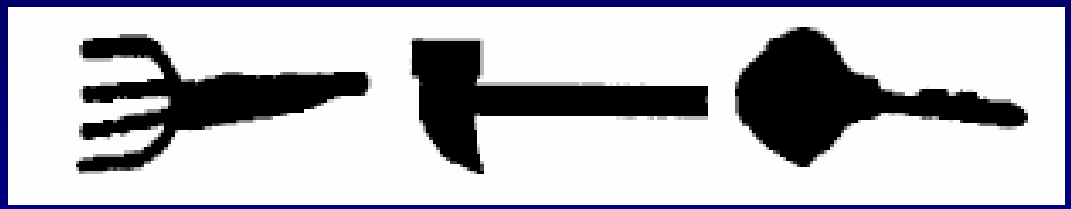


Region-based vs. Contour-based

Good Examples for Region-based shape

Good Examples for Contour-based shape

Similar Region Shape, Different Contour



Angular Radial Transformation (ART)
[Kim'99]

- A Region-based shape
- Calculate the coefficients based on image intensities in polar coordinates (n<3, m<12)

$$F_{nm} = \int_0^{2\pi} \int_o^1 V_{nm}(\rho,\theta) f(\rho,\theta) \rho d\rho d\theta$$

$f(\rho,\theta) \cdots$ image intensity in polar coordinates

$V_{nm}(\rho,\theta) \cdots$ ART basis function

$$V_{nm}(\rho,\theta) = 1/2\pi \exp(jm\theta) R_n(\rho)$$

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(\pi n \rho) & n \neq 0 \end{cases}$$

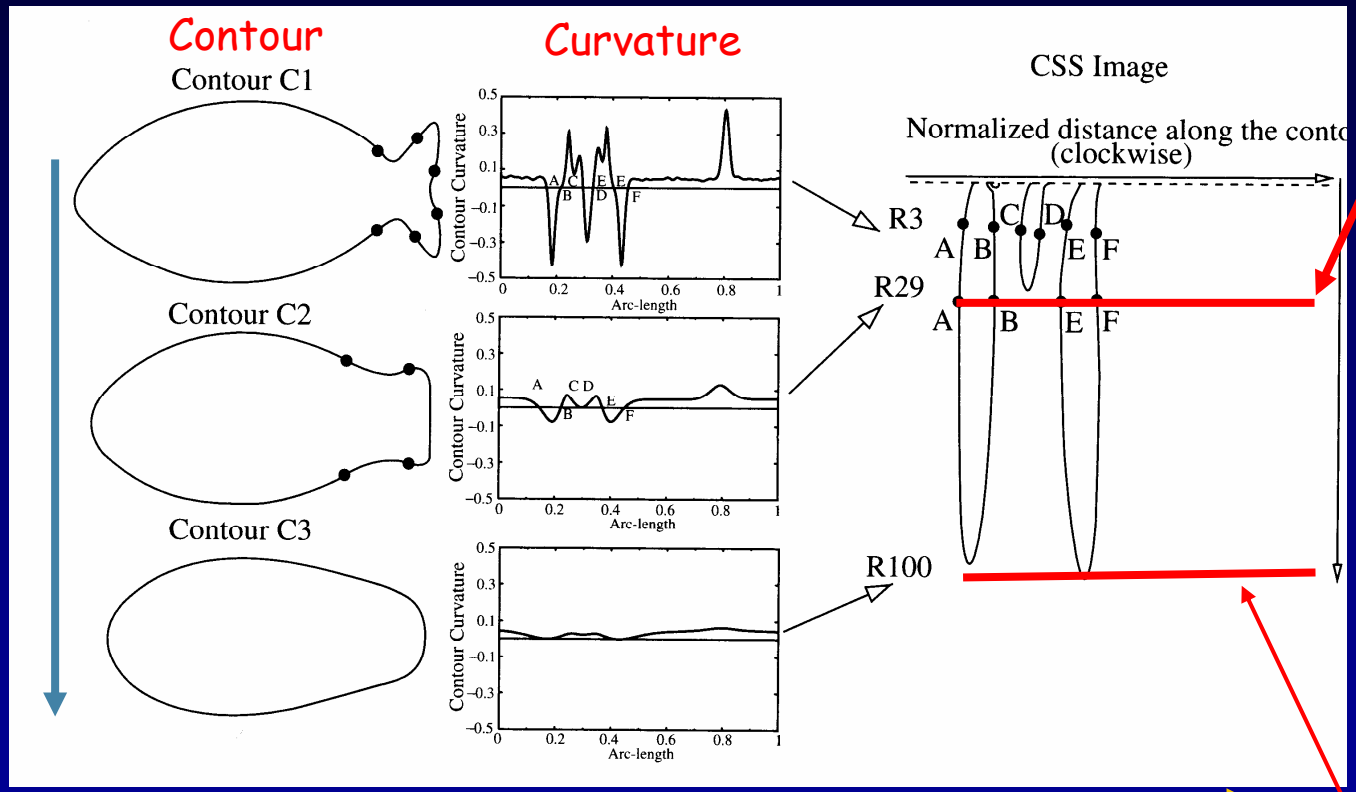
Total 35 coefficients in 140 bits (4 bits/coeff)

Curvature Scale-Space (CSS) [Mokhtaarian '92]

- A contour-based shape
  1) Select N points in the contour
  2) Apply lowpath filter repeatedly until concave contours smoothed out
  3) "How contours are filtered" becomes the features
     - Zero crossing in the curvature functions after each application of the lowpass filter
     - CSS Image

CSS Image

Zero crossing in curvature is plotted



After 29 iteration

Amount of Smoothing

Distance along Contour

After 100 iteration

Munehiro Nakazato and Charlie Dagli

CSS Features

- # of peaks in CSS images
- Highest peak
- Circularity (perimeter^2/ area)
- Eccentricity
- Etc.

Texture Features

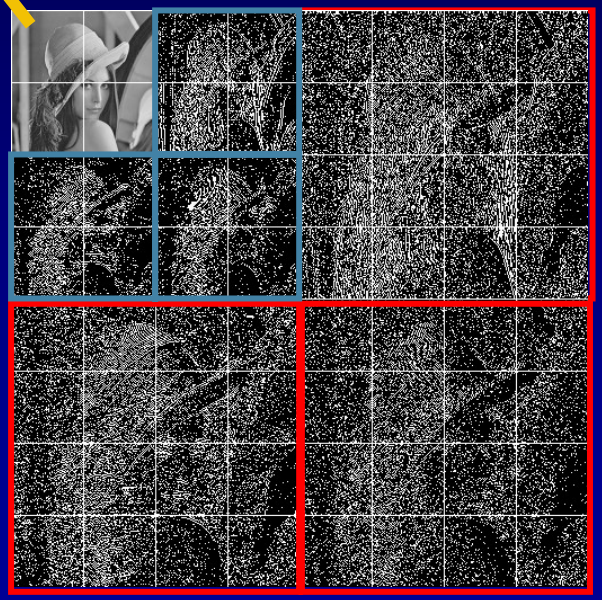
- Wavelet-based Texture Features [Smith'94]

Wavelet Filter Bank

Coarse Info (low freq)

Original Image

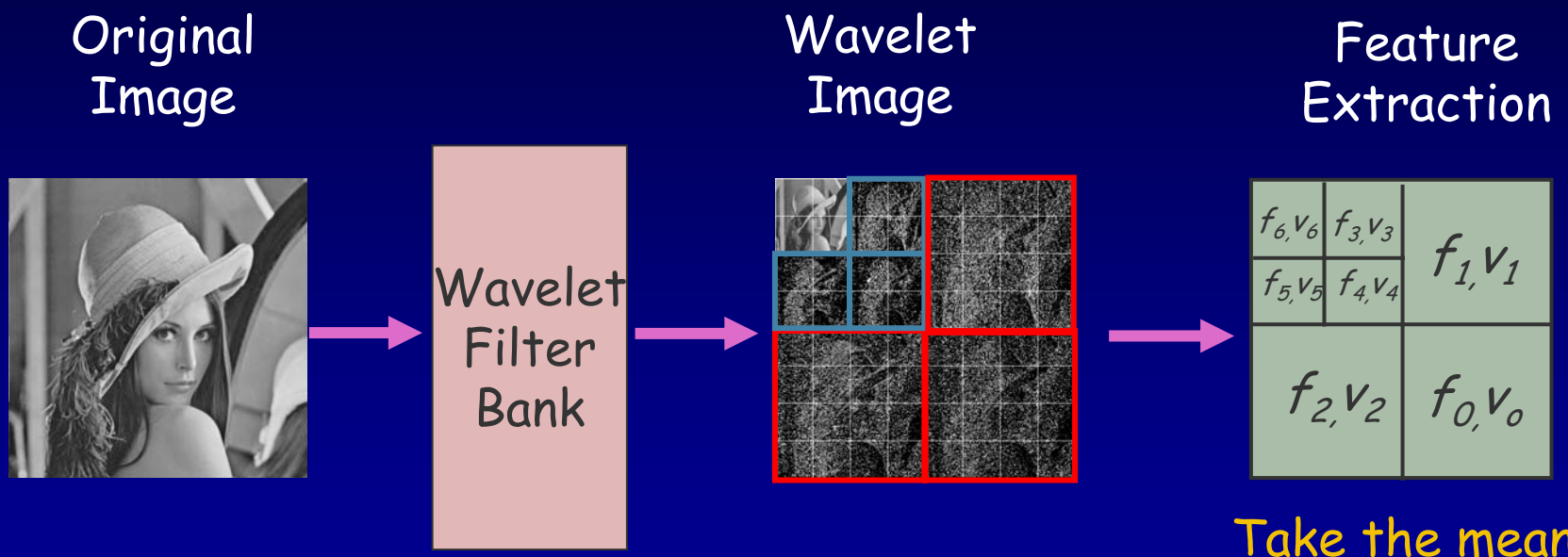
Wavelet Filter



Detail (high freq)

Munehiro Nakazato and Charlie Dagli

Texture Features from Wavelet



Munehiro Nakazato and Charlie Dagli

Other approaches: Region-Based

- Global features often times fail to capture local content in an image

GLOBAL DESCRIPTION

{Green, Grassy, Hillside}

color, texture, shape

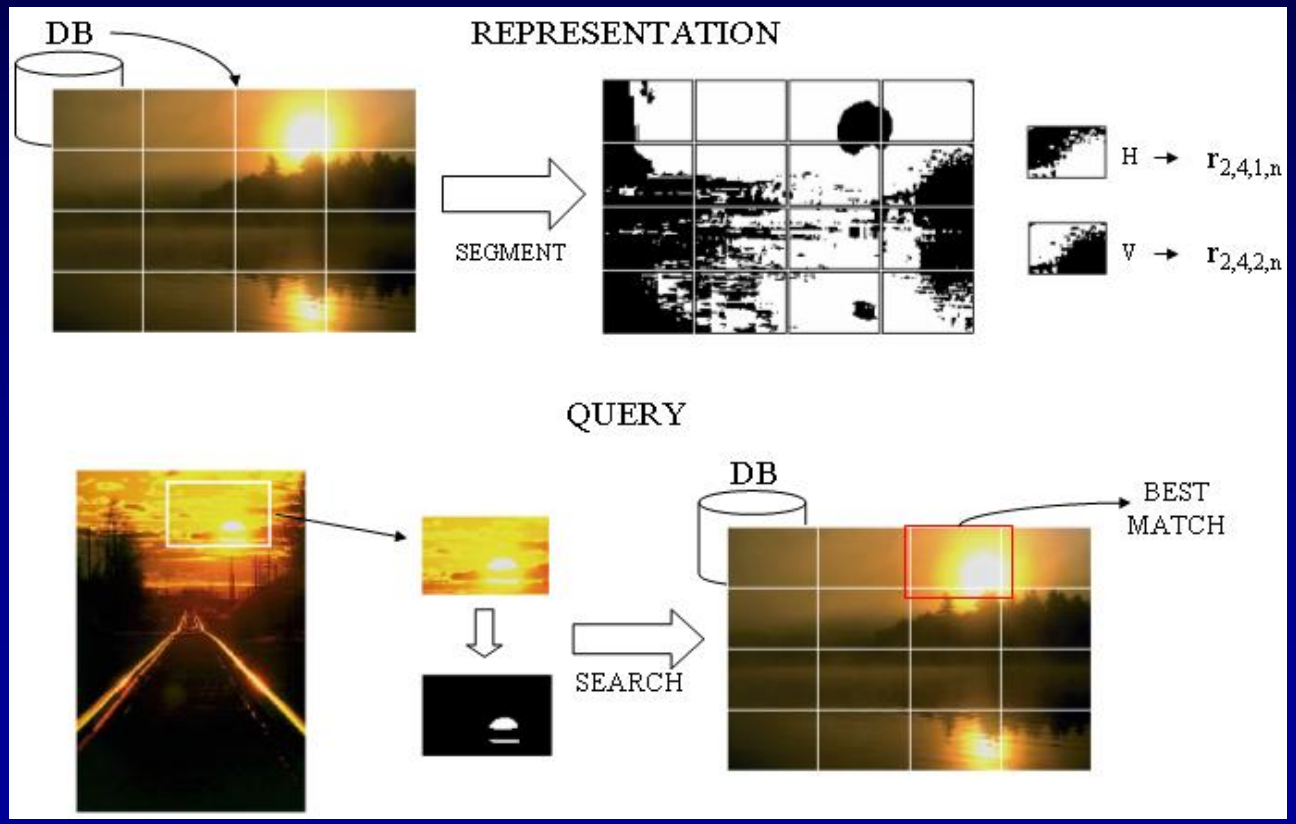
No sheep? No fence? No houses?

Other approaches: Region-Based

- Segmentation-Based
  - Images are segmented by color/texture similarities: Blobworld [Carson '99], Netra [Ma and Manjunath '99]
- Grid-Based
  - Images are partitioned, features are calculated from blocks: [Tian '00],[Moghaddam '99]

Other approaches: Region-Based

- Combine Grid and Segmentation: [Dagli and Huang, '03]



Basic Components of CBIR

- Feature Extractor
  - Create the metadata
- Query Engine
  - Calculate similarity
- User Interface

Now, We have many features (too many?)

- How to express visual "similarity" with these features?

Visual Similarity ?

- "Similarity" is Subjective and Context-dependent.
- "Similarity" is High-level Concept.
  - Cars, Flowers, …
- But, our features are Low-level features.
  - Semantic Gap!
- Still, we have to struggle with them.

Which features are most important?

- Not all features are always important.
- "Similarity" measure is always changing
- The system has to weight features on the fly.

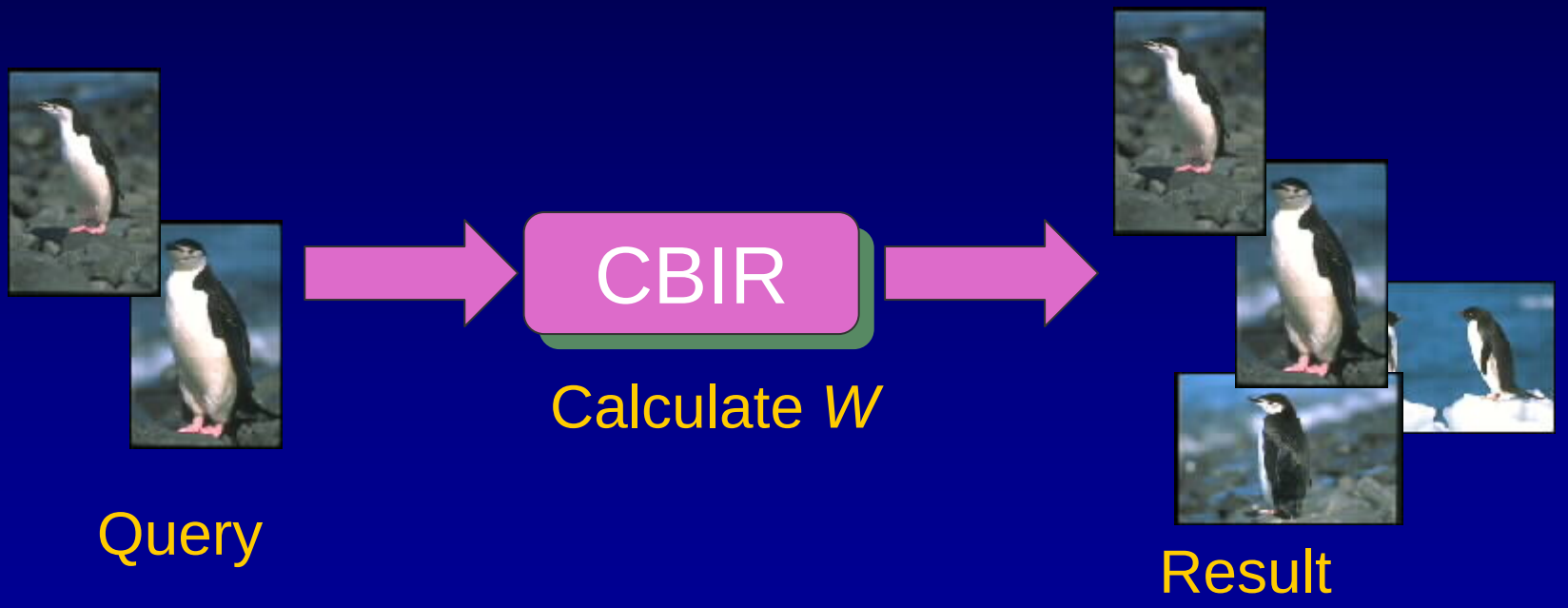
How ?

Online Feature Weighting

- Approach #1 - Manual
  - Ask the user to specify number
    "35% of color and 50% of texture…"
  - Very difficult to determine the numbers
- Approach #2 - Automatic
  - Learn feature weights from examples
  - Relevance Feedback

Online Feature Weighting

- From Query Examples, the system determines feature weighting matrix *W*



$$distance(\vec{x},\vec{y}) = (\vec{x}-\vec{y})^T W (\vec{x}-\vec{y})$$

How to Calculate *W*?

- No Negative Examples (*1-class*)
- Positive and Negative Examples (*2-class*)
- One Positive and Many Negative classes *(1+x)-class*
- Many Positive and Many Negative classes *(x+y)-class*

When there are only relevant images available…

- We want to give more weights to common features among example images.
- Use the variance.
  - Features with low variance
    -> Common features
    -> Give higher weight

One Class Relevance Feedback in MARS [Rui '98]

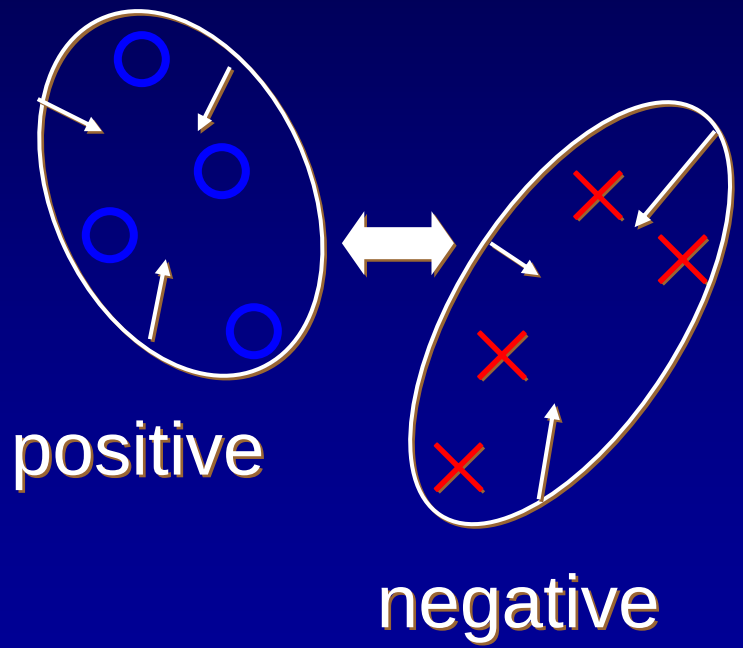
- Calculates the Variance among relevant examples.
- The inverse of variance becomes the weight of each feature.
- This means "common features" between positive examples have larger weights.

$$W = \begin{bmatrix} 1/\sigma_1^2 & & & & 0 \\ & 1/\sigma_2^2 & & & \\ & & 1/\sigma_3^2 & & \\ & & & \ddots & \\ 0 & & & & 1/\sigma_k^2 \end{bmatrix}$$

W is a k x k diagonal matrix

Relevance Feedback as Two-Class Problem (positive and negative)

Fisher's Discriminant Analysis (FDA)



- Find a *W* that …
- minimizes the scatter of each class cluster (within scatter)
- maximizes the scatter between the clusters (between scatter)

Two-Class problem

- Target function
- W is full matrix

$$W = \underset{W}{\operatorname{argmax}} \left| \frac{W^T S_B W}{W^T S_W W} \right|$$

$S_B \cdots$ Between Scatter Matrix

$S_W \cdots$ Within Scatter Matrix

$$S_W = \sum_{i=1}^{2} \sum_{j \in \text{group } \# i} (x_j - m_i)(x_j - m_i)^T$$

$$S_B = (m_1 - m_2)(m_1 - m_2)^T$$

$m_1, m_2 \cdots$ mean of each class

Solution

- The problem is reduced to generalized eigenvalue problem

$$S_B w_i = \lambda_i S_W w_i$$

$$W = \Phi \Lambda^{1/2}$$

$\Lambda \cdots$ diagonal matrix of eigenvalues

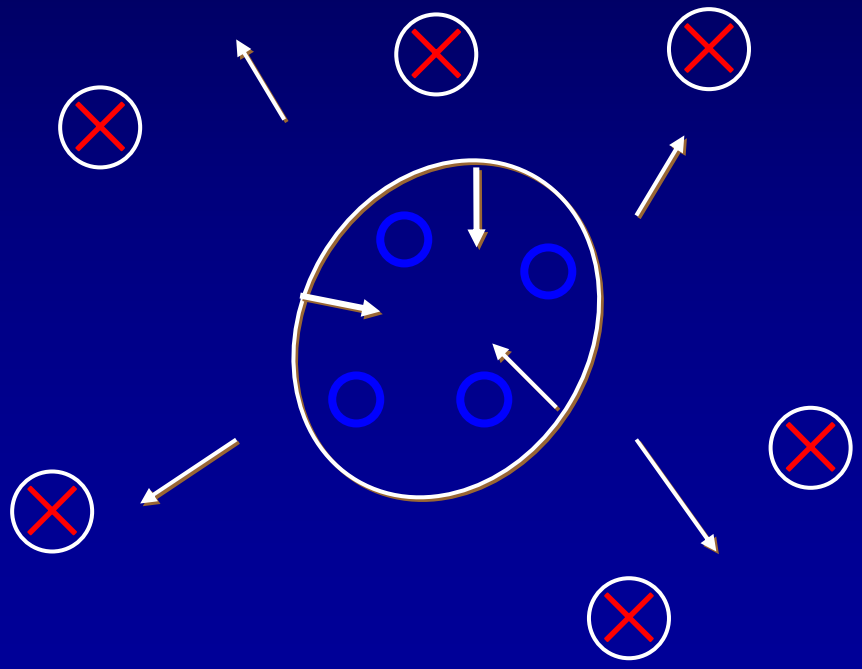
$\Phi \cdots$ eigenvectors

From Two-class to (1+x)-class

- Positive examples are usually from one class such as flower
- Negative examples can be from any classes such as "car", "elephant", "orange"…
- It is not desirable to assume negative images as one class.

RF as (1+x)-Class Problem

- *Biased Discriminant Analysis* [Zhou et al. '01]
- Negative examples can be any images
- Each negative image has its own group



positive

negative

$$S_W = \sum_{x \in positive} (x - m)(x - m)^T$$

$$S_B = \sum_{x \in negative} (x - m)(x - m)^T$$

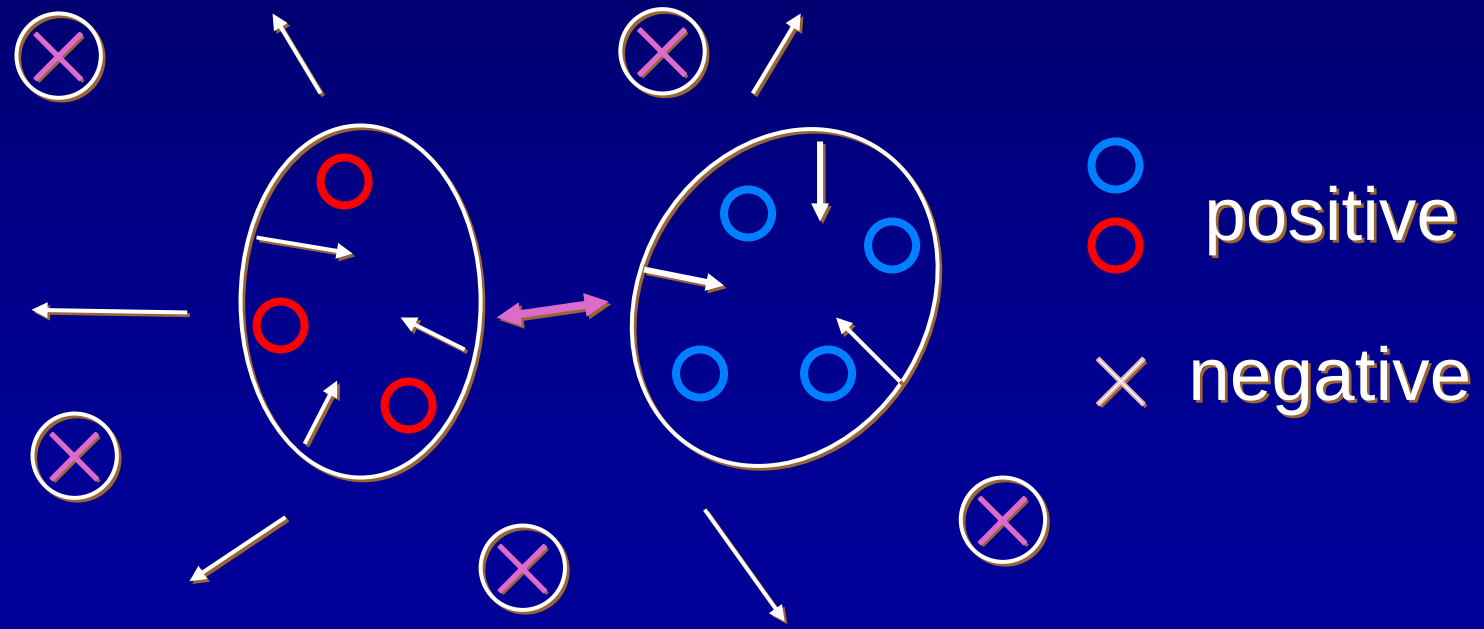
$m \cdots$ mean of positive class

The solution is similar to FDA

Munehiro Nakazato and Charlie Dagli

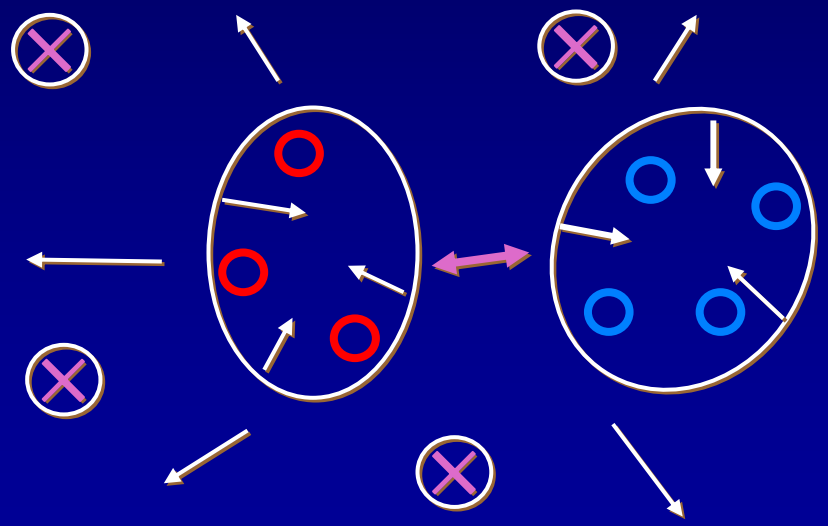
RF as (x+y)-Class Problem

- *Group BDA* [Nakazato, Dagli '03]
- Multiple Positive classes
- Scattered Negative classes



RF as (x+y)-Class Problem

- *Group BDA* [Nakazato, Dagli '03]
- Multiple Positive classes
- Scattered Negative classes



positive

negative

$$S_W = \sum_i \sum_{x \in i} (x - m_i)(x - m_i)^T$$

$$S_B = \sum_i \sum_{y \in negative} (y - m_i)(y - m_i)^T$$

$m_i \cdots$ mean of positive class i

Basic Components of CBIR

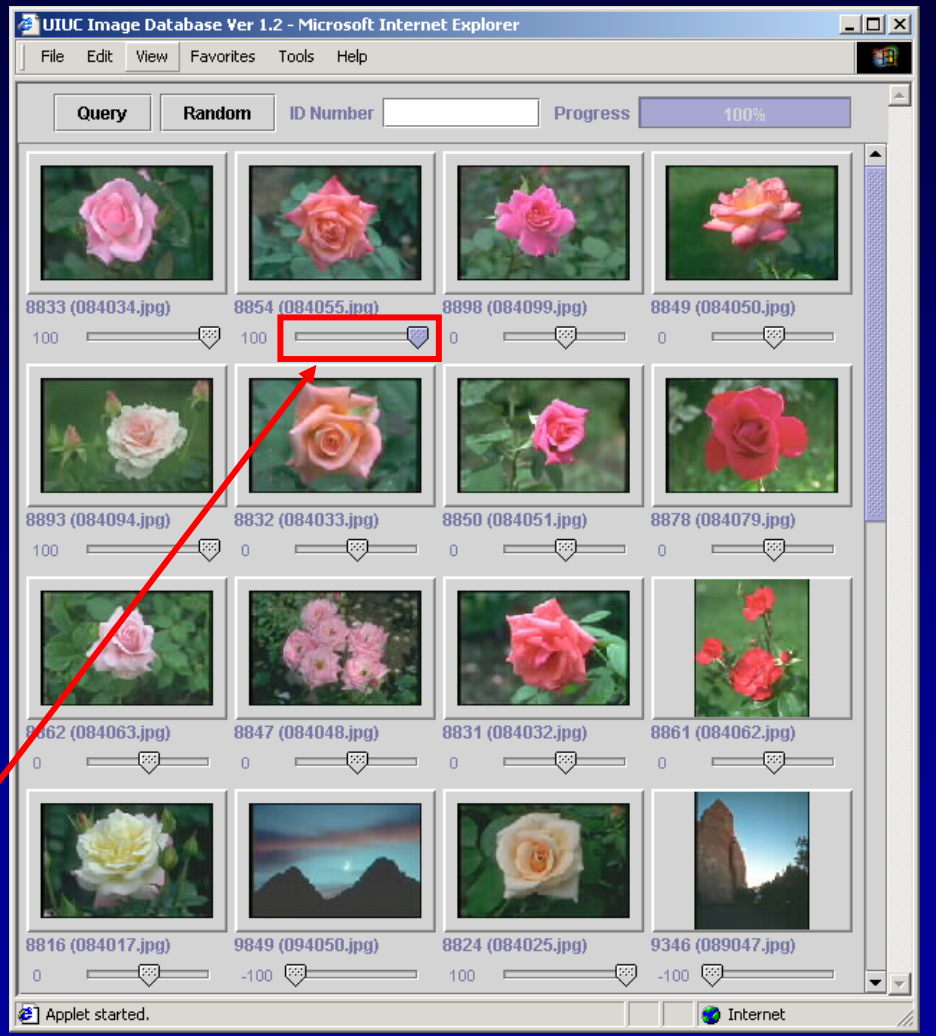
- Feature Extractor
  - Create the metadata
- Query Engine
  - Calculate similarity
- User Interface

User Interface and Visualization

- Basic GUI
- Direct Manipulation GUI
  - El Nino [UC San Diego]
  - Image Grouper [Nakazato and Huang]
- 3D Virtual Reality Display

Traditional GUI for Relevance Feedback

- User selects relevant images
- If good images are found, add them
- When no more images to add, the search converges



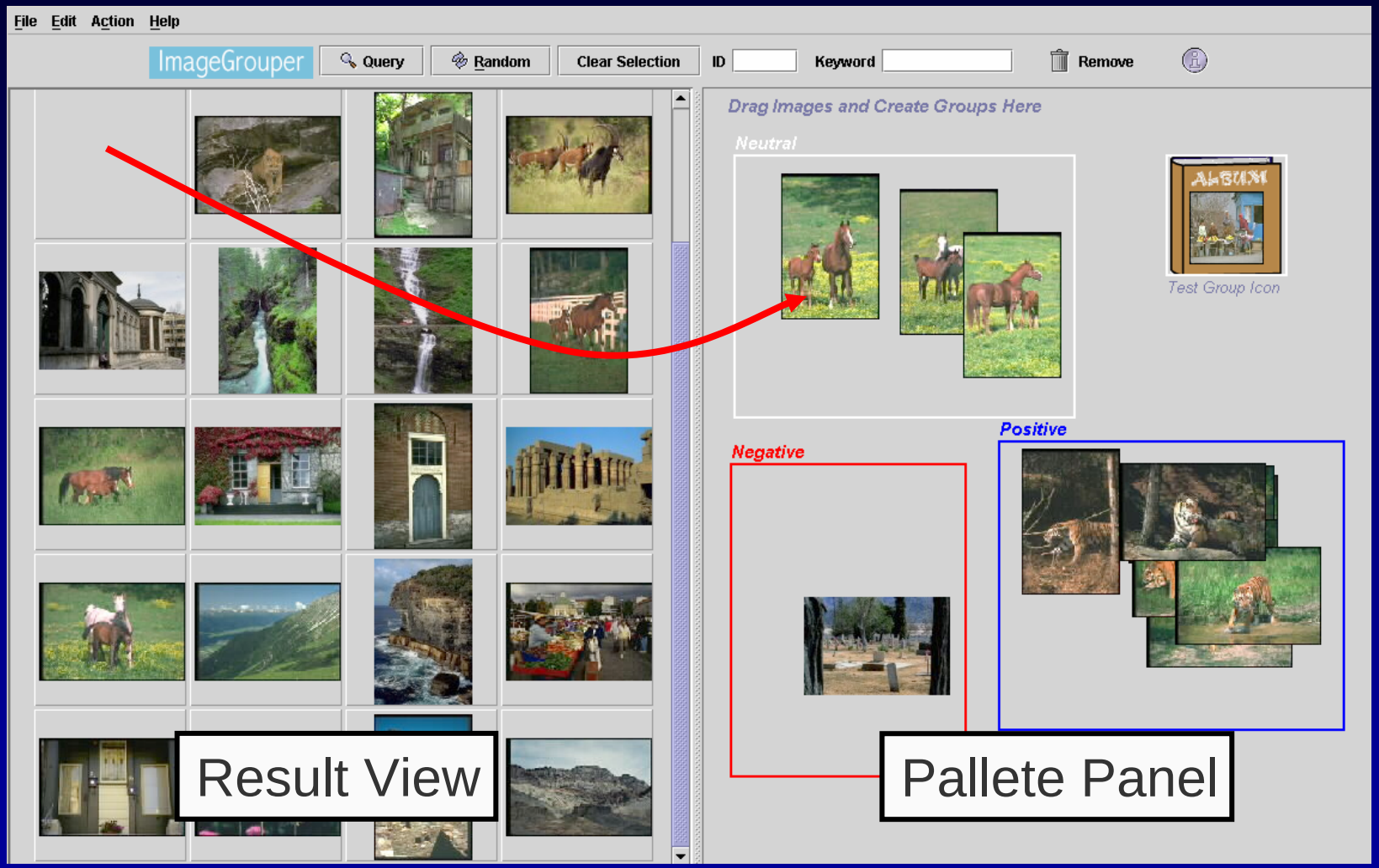
Slider or Checkbox

Munehiro Nakazato and Charlie Dagli

ImageGrouper [Nakazato and Huang]

- Query by Groups
  - Make a query by creating groups of images
  - Easier to try different combinations of query sets (trial-and-Error Query)

ImageGrouper



Munehiro Nakazato and Charlie Dagli

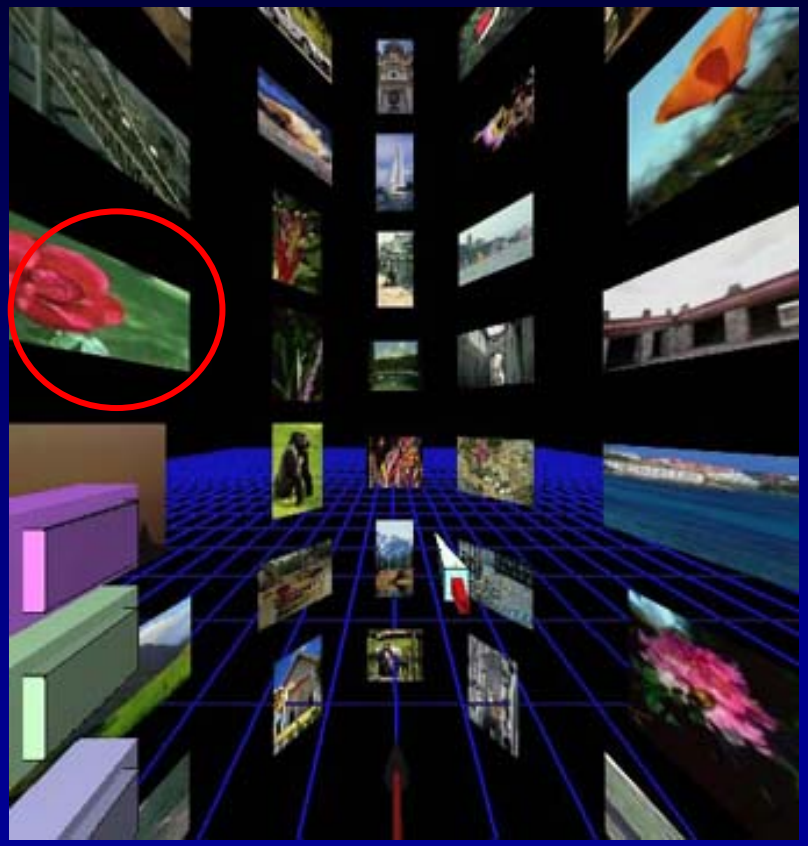
Note

- Trial-and-Error Query is very important because
  - Image similarity is subjective and context-dependent.
  - In addition, we are using low-level image features. (semantic gap)
  - Thus, it is VERY difficult to express the user's concept by these features.

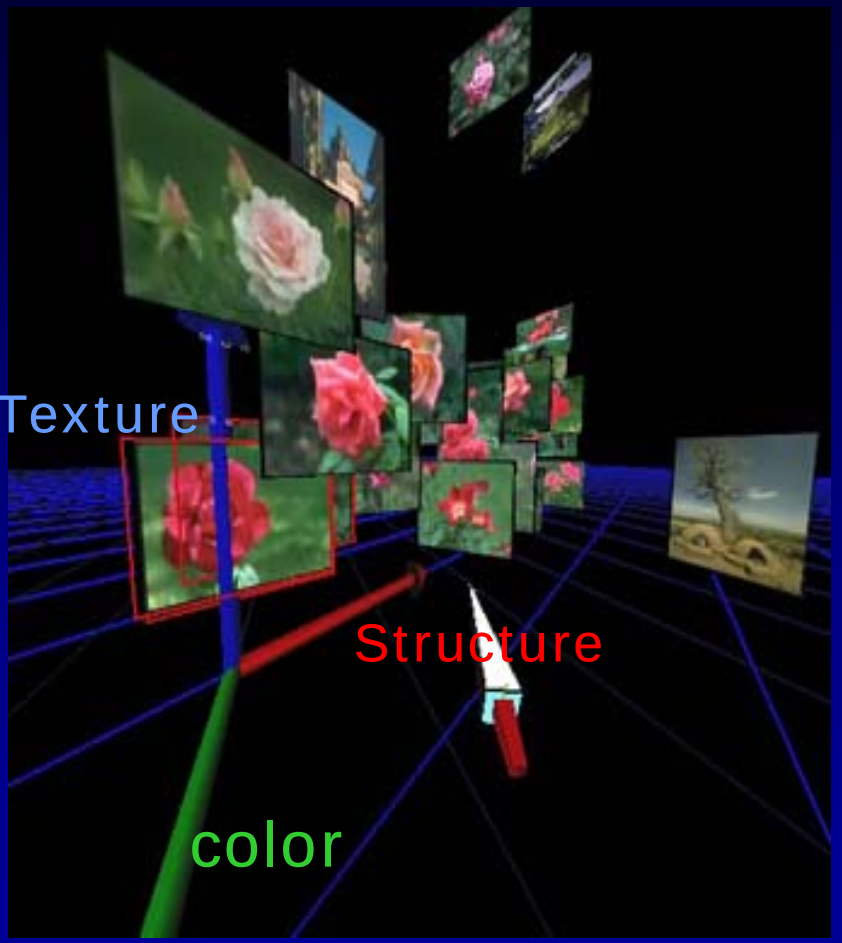
Image Retrieval in 3D

- Image retrieval and browsing in 3D Virtual Reality
- The user can see more images without occlusion
- Query results can be displayed in various criteria
  - Results by Color features, by texture, by combination of color and texture

3D MARS



Initial Display



Result

Munehiro Nakazato and Charlie Dagli

3D MARS in CAVE™

- Shuttered glasses for immersive 3D experience
- Click and Drag images by WAND
- Fly-through by Joystick

Munehiro Nakazato and Charlie Dagli

Demos

- Traditional GUI
  - IBM QBIC
    - http://wwwqbic.almaden.ibm.com/
  - UIUC MARS
    - http://chopin.ifp.uiuc.edu:8080
- ImageGrouper
  - http://www.ifp.uiuc.edu/~nakazato/gouper

References (Image Features)

- Bober, M., "MPEG-7 Visual Descriptors," In IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 6, June 2001.
- Stricker, M. and Orengo, M., "Similarity of Color Images," In Proceedings of SPIE, Vol. 2420 (Storage and Retrieval of Image and Video Databases III), SPIE Press, Feb. 1995.
- Zhou, X. S. and Huang, T. S., "Edge-based structural feature for content-base image retrieval," Pattern Recognition Letters, Special issue on Image and Video Indexing, 2000.
- Smith, J. R. and Chang S-F. Transform features for texture classification and discrimination in large image databases. In Proceedings of IEEE Intl. Conf. on Image Processing, 1994.
- Smith J. R. and Chang S-F. "Quad-Tree Segmentation for Texture-based Image Query." In Proceedings of ACM 2nd International Conference on Multimedia, 1994.
- Dagli, C. K. and Huang, T.S., "A Framework for Color Grid-Based Image Retrieval," (to be submitted ICPR'03).
- Tian, Q. et. al. "Combine user defined region-of-interest and spatial layout in image retrieval," in IEEE Intl. Conf. on Image Processing, 2000.
- Moghaddam B. et. al. "Defining image content with multiple regions-of-interest," in IEEE Wrkshp on Content-Based Access of Image and Video Libraries, 1999.

References (Relevance Feedback)

- Rui, Y., et al., "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," In *IEEE Trans. on Circuits and Video Technology*, Vol.8, No.5, Sept. 1998
- Zhou, X. S., Petrovic, N. and Huang, T. S. "Comparing Discriminating Transformations and SVM for Learning during Multimedia Retrieval." In *Proceedings of ACM Multimedia '01,* 2001.
- Ishikawa, Y., Subrammanya, R. and Faloutsos, C., "MindReader: Query database through multiple examples," In Proceedings of the 24th VLDB Conference, 1998.

References (User Interfaces and Visualizations)

- Nakazato, M. and Huang, T. S. "3D MARS: Immersive Virtual Reality for Content-Based Image Retrieval." In *Proceedings of 2001 IEEE International Conference on Multimedia and Expo (ICME2001)*, Tokyo, August 22-25, 2001
- Nakazato, M., Manola, L. and Huang, T.S., "ImageGrouper: Search, Annotate and Organize Images by Groups," In *Proc. of 5th Intl. Conf. On Visual Information Systems (VIS'02)*, 2002.
- Santini, S. and Jain, R., "Integrated Browsing and Querying for Image Database," *IEEE Multimedia*, Vol. 7, No. 3, 2000, page 26-39.