MULTIMEDIA RETRIEVAL

Punitha Puttuswamy
Multimedia Information Retrieval Group
Department of Computing Science
punitha@dcs.gla.ac.uk

Multimedia?

- Text, images, drawings(graphics), animation, video, sound(speech)
- PCs, DVDs, games, digital TV, Web surfing etc.

Applications of Multimedia

- Home
  - Video on demand, Interactive TV
  - Electronic album, Personalised electronic journals
- Education and Training
  - Computer Aided Instruction, Multimedia Encylopedias
  - Distant and Interactive training – Teleconference, Distributed Lectures
- Business/Office
  - Co-operative/collaborative Environments
  - Remote consulting systems
  - Document exchange and sharing
  - Advertising/publishing
- Public
  - Digital libraries,
  - Electronic Museum,
  - Network systems – medical, banking, shopping, tourist

Multimedia Retrieval?

- Retrieval of multimedia objects (image,speech,video) from a database (that are relevant to a query)
- Issues/Questions
  - Is it difficult? Why?
    - What is an architecture of a MIR system?
    - How do we index and represent a multimedia object?
    - How do we define/specify a query?

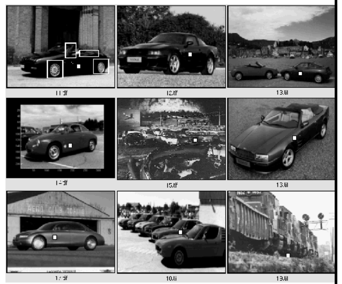
- Content-Based Image Retrieval
  - Architecture of CBIR system
  - Techniques for extraction and representation of image features
  - Research Prototypes/Commercial Systems
- Video Retrieval
  - Video Processing techniques
    - Shot/Scene detection
    - Key Frame selection
    - Video abstracting
    - Video retrieval
    - Interface Issues

Content Based Image Retrieval

- What is CBIR?
  – Its purpose is to retrieve images, from a database (collection), that are relevant to a query.
  – Retrieval of images on the basis of features automatically extracted from the images
  – Finding images which are "similar" to a query.
    - Query: The whole or parts of an example image.

What is "Similarity"

- Ultimately user defines "similarity".
- What is "similar"
  – Cars of a given model or all cars?
  – Red coloured cars?
- Local or Global similarity?
  – Similarity of parts?
  – Similarity of the entire image?



How does one find similarity?
What features?
Metric distance?
Non-metric distance?

Content

- Data which is not directly concerned with image, but in some way related to it is not content but content-independent metadata. (Examples: photographer's name, date, location etc.)

+ Data which is evident from images to human eye is content

+ low intermediate features (colour, texture, shape etc.) are known as content-dependent metadata

The need for content-based Image Retrieval

Large amount of visual data is produced digitally

- Digital cameras at consumer prices
- Publications on the Internet
- Billions of images
- Journalists (Millions of images produced every day)
- Trademarks (>100.000 visual marks in Switzerland alone)
- Hospitals (Geneva radiology: >30,000 images per day)
- Only small part of the images is annotated
- Annotation is expensive, subjective, task dependent
- Not everything can be described by text

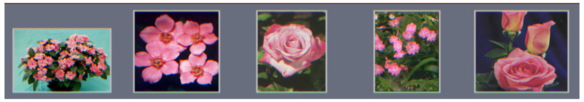
Applications

- **Crime prevention :** (face recognition, fingerprint identification, shoe sole recognition, tyre track identification, iris recognition)
- **Intellectual Property registration :** (trademark registration)
- **Architecture and Design Engineering:** (floor planning)
- **Medicine:** (Teaching/Studying cases, lung CTs, Mammography, tumor detection)
- **GIS, Journalism, Education and Training, Art historians**
- **Fashion, Publishing, Advertisement, Websearching.**

Examples

- Trademark retrieval - Is there a "similar" trademark?

- Flower patents - Is there a "similar" flower or of a given color.



- Face retrieval
  - For identification, security access

  - organizing home collections.

Related areas of CBIR

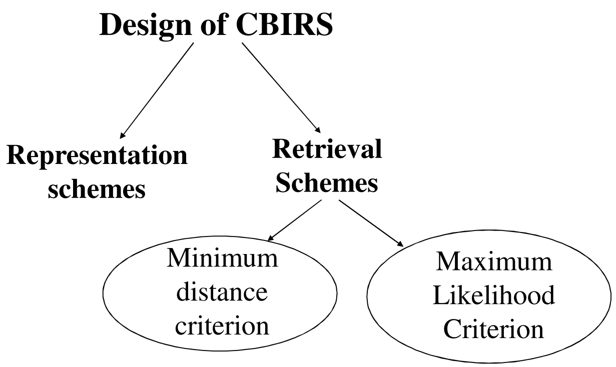
• **Evolution :**
is an active area since 1970, thrust from two major areas Database Management (text based) and Computer Vision (visual based)

• lies at the crossroads of multiple disciplines
> Database
> Artificial intelligence
> Image Processing
> Statistics
> Computer Vision
> High performance Computing
> Cognitive Science
> Human-Computer intelligent interaction … etc

Issues in the design of CBIR

- Understanding users' needs and information seeking behavior
- identification/extraction of suitable features/ways of describing images
- Perception of knowledge embedded in images
- Efficient Storage of images
- Correctness and Effectiveness in image representation
- Family of queries allowed
- Designing Robust search techniques

In other words,



Representation schemes

Representation should minimize

- **Sensory gap**
  - i.e., the gap between the object in the real world and information in recorded scene.
- **Semantic gap**
  - i.e., the lack of coincidence between the information that one can extract from the visual data and the interpretation that the data have in a given situation (for an user)

The Effects of Aging

- Can you guess which person on the right matches the person on the left?
  - The pictures on the left are of high school students.
  - The pictures on the right were taken 20 yrs later.
- This is hard.
- If this is hard for people, how can an image retrieval system do this?

Retrieval Schemes

- Search by association (iterative refinement of search)
- Search for precise copy of an image in mind
- Search for an image, a member of a specific class

Queries can be characterized into three levels of abstraction

L1. Search based on primitive features such as colour, texture, shape or spatial relationship

L2. Logical features such as identity of objects shown

L3. Abstract attributes such as the significance of the scenes depicted

Image Representation and Associated Retrieval Schemes

- **Entirety**
  - Vast memory requirement
  - Encumbers retrieval as it is based on model matching
- **Keywords or Caption Representation**
  - can be traced back to 1970's
  - framework for image representation and retrieval was to annotate images by text and then use text based DBMS to perform retrieval.
  - E.g., Getty information institute - Art and Architecture Thesaurus (use 1,20,000 terms)
  - **Other tools from Getty include**
    1. ULAN : Union List of Artist Names
    2. TGN: Getty Thesaurus of Geographic Names
    3. LCTGM: Library of Congress Thesaurus for Graphic materials
    4. LCSH: Library of Congress Subject Headings

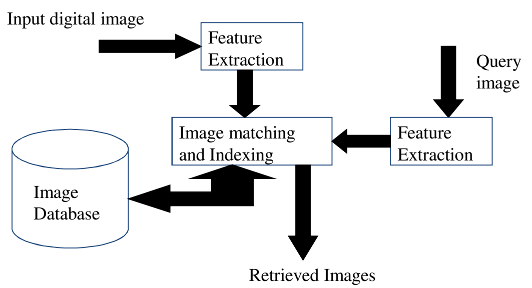
Pros and Cons of keyword/classification code indexing

+ Keywords have high expressive power
+ can be used to describe almost any aspect of image content
+ easily extendible to accommodate new concepts
+ can be used to describe image content at varying degrees of complexity

- Requires vast amount of labor in manual image annotation
- causes wide disparities in the keywords assigned to the same picture by different individuals
- Keywords do not allow unanticipated searching
- Subjectivity of human perception cause mismatch in retrieval
- The descriptive cataloguing of similar images can vary widely particularly if carried out at different times
- Entails describing every color, texture, shape and object in the image for complete annotation
- **RELY ON KNOWLEDGE AND EXPERIENCE OF THE STAFF**

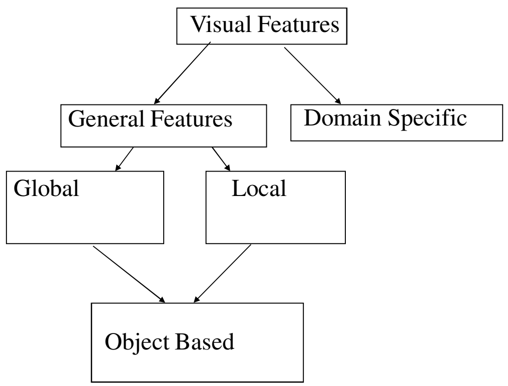
Birth of CBIR

Instead of being manually annotated by text based keywords , images were indexed by their own visual content such as color, texture, shape and spatial relations



General schema of CBIR

Feature Extraction



Global Features

- Color
  - **Histograms**, Dominant Color
- Texture
  - Repetitiveness, directionality, granularity, coarseness
  - Probabilistic/statistical, spectral, structural
  - Morphology, Fractals
- Transforms
  - Fourier, Wavelet, DCT, Gabor, Haar

Local Feature

- Color
  - Color Layout, Scalable Color, Color coherence
- Shape
  - Contour based, Region based

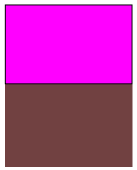
Object Based

- Spatial Reasoning

Content Based Image Retrieval – Based on Colour Features

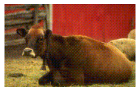
Retrieval based on colour

- Finding images containing a specified colour in an assigned proportion



Retrieval based on colour

- Finding images containing a specified colour in an assigned proportion
- Finding images whose colours are similar to those of an example image

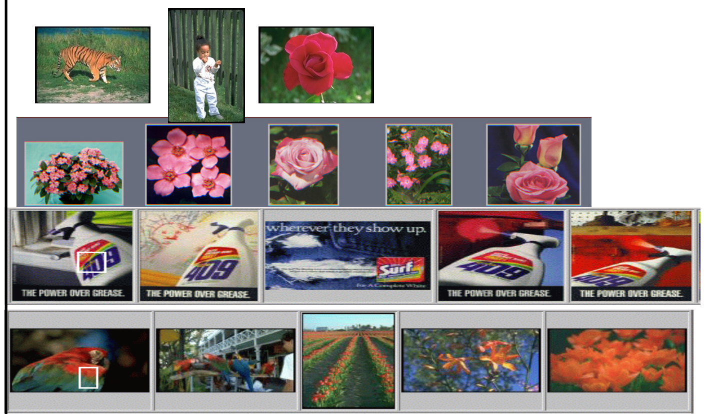


Retrieval based on colour

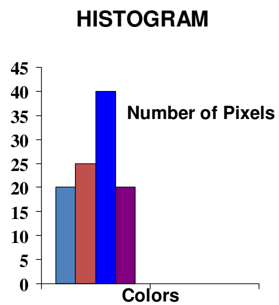
- Finding images containing a specified colour in an assigned proportion
- Finding images whose colours are similar to those of an example image
- Finding images containing colour regions as specified in a query

Retrieval based on colour

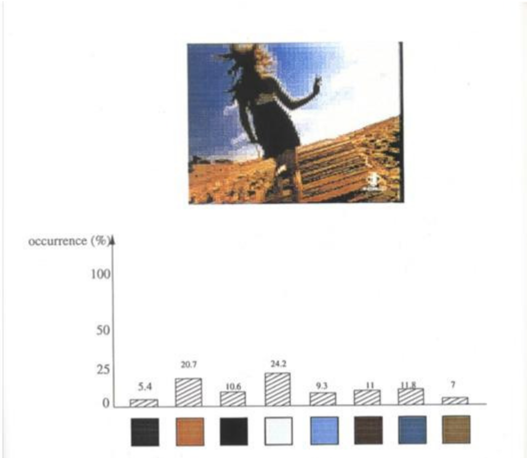
- Finding images containing a specified colour in an assigned proportion
- Finding images whose colours are similar to those of an example image
- Finding images containing colour regions as specified in a query
- Finding images containing a known object based on its colour properties



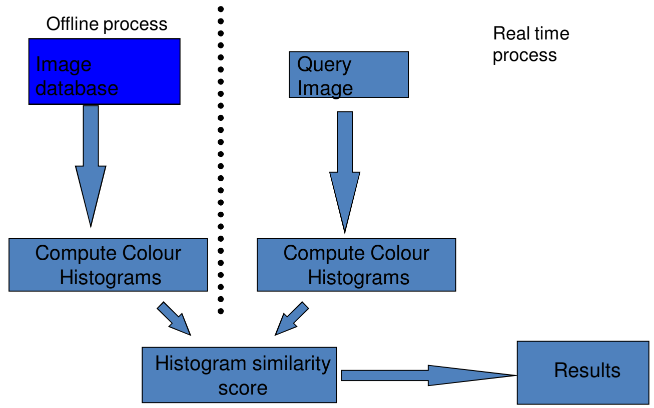
Histogram



- How many pixels of a given color or a given intensity?
- Two images are “similar” if they have the same distribution i.e. histograms.
- Simplest kind of non-parametric density estimate.



Retrieval using Histograms

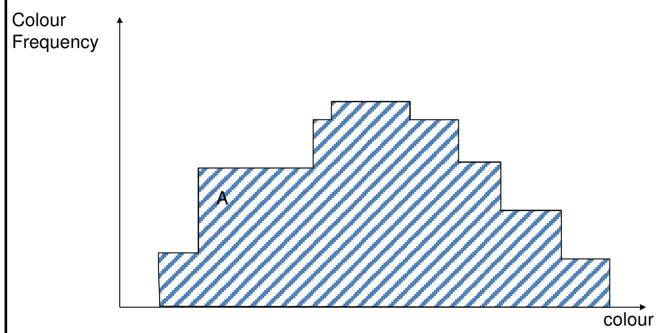


Distance Measures

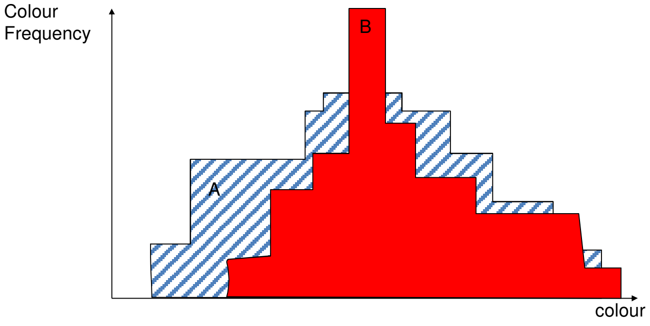
- Bin by Bin dissimilarity measures:
  - L1 norm (absolute deviation) $\sum_i | A_i - B_i |$
  - L2 norm (Euclidean distance) $\left(\sum_i | A_i - B_i |^2 \right)^{\frac{1}{2}}$
  - Minkowski Distance

$$d_{L_r}(A,B) = \left(\sum_i | A_i - B_i |^r \right)^{1/r}$$

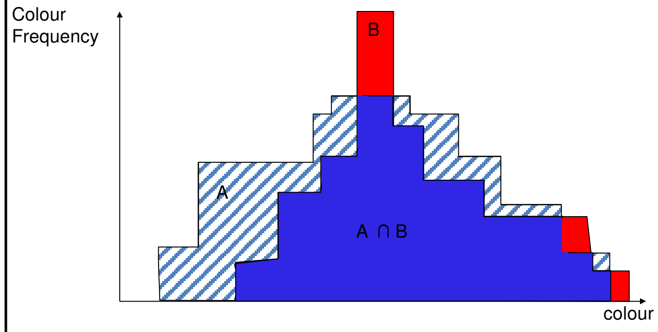
Histogram intersection



Histogram intersection



Histogram intersection



- Swain & Ballard

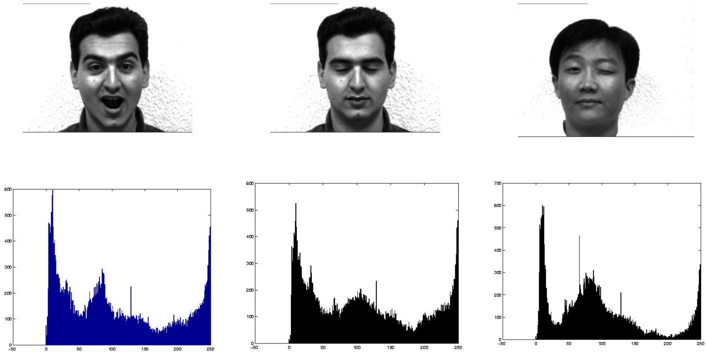
$$d\ (A,B) = 1 - \frac{\sum_i \min(A_i, B_i)}{\sum_i B_i}$$

- ❖ Can handle partial matches when the areas of the histograms are different.
- Colours that are not present in the query image do not contribute to the intersection distance

Image Representation using colour histograms

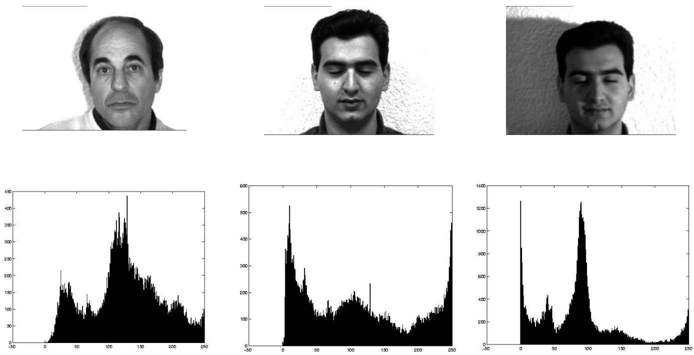
- Most traditional way of representing low-level properties of an image
- Computation is trivial
- Fairly insensitive to variation originated by camera rotation (up to 45 degree)
- Image resolution
- Zooming (1,3)
- Partial occlusions

Why are Histograms Useful?



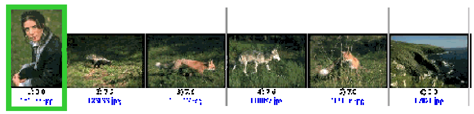
Why is Intensity a Bad Feature?

Why is Intensity a Bad Feature?

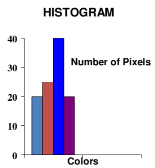


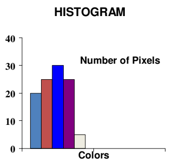
Color Retrieval - Bad Example

query: 



Problems with histograms.





- Qunatization must be fine enough that distinct colours are not in the same bin
- Corresponding bins in the two graphs are not identical.
  - The histogram may look different if the bin origins are changed.
  - The histogram may look different if the bin sizes are changed.
- Choice of representative colours affects the perception of similar images
- How does one compute similarity?
  - Use a good measure of distance?

```
Int Histogram[256][256][256];

For rows = 1 to Row_max
For cols = 1 to Col_max
{
        R = image[row][col].red;
        G = image[row][col].green;
        B = image[row][col].blue;

        Histogram[R][G][B] ++;

}
```

In order to reduce computation time, a 256x256x256 = 16777216 color image is quantized into a 8 x 8 x 8 = 512 color image in RGB color space.

Colour Spaces

- Points in three dimensional space
- Calorimetric models
  - CIE Chromaticity diagram
- Physiologically inspired models
  - CIE XYZ, RGB
- Psychological models
  - HSV
- Hardware-oriented models
  - RGB, CMY, YIQ
- User-oriented models
  - HLS, HSV, HSB

Non-parametric Measures

- Kullback-Leibler divergence
  - How much does two distribution (histogram) agree
  - Sensitive to histogram binning, Asymmetric

$$d_{KL}(A,B)=\sum_i A_i \log \frac{A_i}{B_i}$$

- Jeffrey divergence
  - Symmetric version of KL. Empirically derived. More robust.

$$d_{KL}(A,B)=\sum_i \left(A_i \log \frac{A_i}{M_i}+B_i \log \frac{B_i}{M_i}\right) \qquad M_i=(A_i+B_i)/2$$

- χ^2 statistics
  - How likely is it that one distribution is drawn from the population represented by the other?

$$d_{\chi^2}(A,B)=\sum_i \frac{(A_i-M_i)^2}{M_i}$$

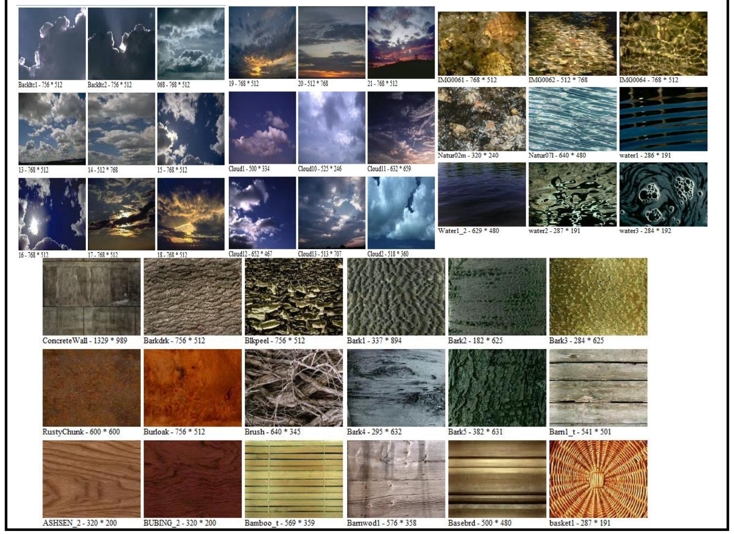
More Distance Measures

- Cross-bin distance measure: Quadratic-form distances $d_Q(A,B) = \sqrt{(A-B)^T Q(A-B)}$
  - Q = [q_{ij}] denotes similarity between bins i and j.
  - Use $q_{ij} = 1 - d_{ij}/d_{max}$ where $d_{max} = \max(d_{ij})$,d_{ij} distance between bins i and j.
  - A measure of how bins i and j are related.

In Summary

- Colour is a visual feature which is immediately perceived
- Distances in colour space should correspond to human perceptual distance
- Salient chromatic properties are captured
- Presence and distributions of colours induce sensations and conveys meanings in the observer according to specific rules
- Retrieval according to the meaning they convey or sensations they produce

Content Based Image Retrieval – Based on Texture Features



- Structured Textures usually have dominant periodic patterns
- A periodic or repetitive patterns can be captured by the filtered images
- Dominant scale and orientation can also be captured

2D texture Histogram

- Directionality d
- Edge separation e (repetitiveness)

1. Extract Edge map using any edge operator (Sobel, Canny...)
2. Find the number of edge having same direction d
3. Find how many pairs of edges with same orientation are separated by same distance

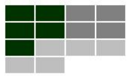
More Methods

Second order statistics based... **Gray level Co-occurrence matrix**

GLCM texture considers the relation between two pixels at a time, called the **reference** and the **neighbour** pixel.

For instance, the neighbour pixel is chosen to be the one to the east (right) of each reference pixel. This can also be expressed as a (1,0) relation: 1 pixel in the x direction, 0 pixels in the y direction.

Each pixel within the window becomes the reference pixel in turn, starting in the upper left corner and proceeding to the lower right. Pixels along the right edge have no right hand neighbour, so they are not used for this count.



0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

neighbour pixel value -> ref pixel value:	0	1	2	3
0	0,0	0,1	0,2	0,3
1	1,0	1,1	1,2	1,3
2	2,0	2,1	2,2	2,3
3	3,0	3,1	3,2	3,3

Co-occurrence

2	2	1	0
0	2	0	0
0	0	3	1
0	0	0	1

Add to its transpose and obtain symmetric matrix

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

Normalise by dividing each entry by the sum of the elements to obtain normalised GLCM which is the probability matrix

Energy (also called Angular moment and uniformity) measure the uniformity of a pattern.

$$\sum_i \sum_j p_d^2(i, j)$$

• Energy reaches its highest value when gray level distribution has either a constant or a periodic form.

• A homogenous image contains very few dominant gray tone transitions, and therefore the matrix for this image will have fewer entries of larger magnitude resulting in large value for energy feature.

• In contrast, if the P matrix contains a large number of small entries, the energy feature will have smaller value.

Entropy measures the disorder of an image and it achieves its largest value when all elements in P matrix are equal.

$$\sum_i \sum_j p_d(i, j) \log p_d(i, j)$$

• When the image is not texturally uniform many GLCM elements have very small values, which implies that entropy is very large.

• Therefore, entropy is inversely proportional to GLCM energy.

Contrast is a difference moment of the matrix and it measures the amount of local variations in an image $\sum_i \sum_j (i-j)^2 p_d(i, j)$

Inverse difference moment measures image homogeneity.

$$\sum_i \sum_j \frac{p_d(i, j)}{(i-j)^2} \quad i \neq j$$

• This parameter achieves its largest value when most of the occurrences in GLCM are concentrated near the main diagonal.

• IDM is inversely proportional to GLCM contrast

Matching

Weighted differences between the moments of two distributions

$$\sum_{i=1}^{r} w_i |V_d - V_t|$$

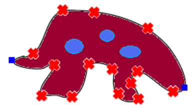
• Texture features serve better when applied for regions.

• This requires Image segmentation

Content Based Image Retrieval – Based on Shape Features

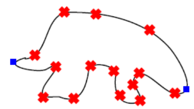
Region Based

- Rectangularity
- Elongatedness
- Graph Representation by thinning
- Medial axis
- Extract features from regions (Color, texture)

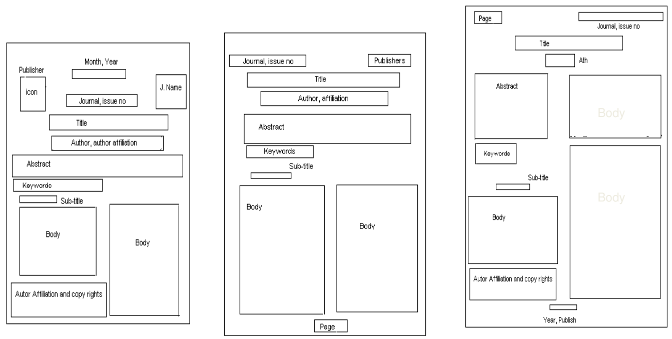


Contour based

- Chain Code
- Signature
- Polygonal Approximation



Content Based Image Retrieval – Based on Spatial Relationship



PRL PR PAMI

Role of Representation

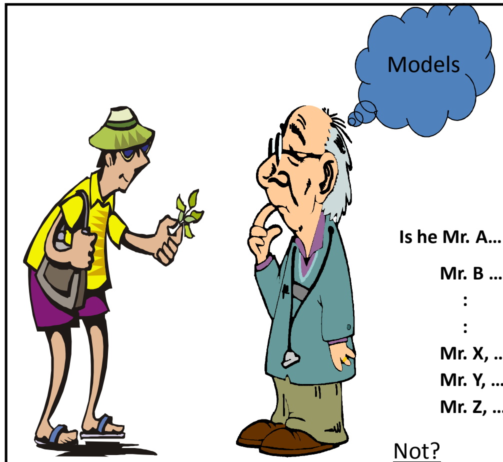
- Library of Models (off line)
- A scene image
- Matching (Hypothesis generation) (on-line)
  - Generally one by one (Sequential)
  - Computationally Expensive

$$O\left(\sum_{i=1}^{n} d_i \cdot s\right) \quad \text{linear w.r.t } n$$

Is there any replacement?

Certainly !!!

because,



Is he Mr. A....No

Mr. B ... No

: :

: :

Mr. X, ...OK $\Rightarrow$ Verification

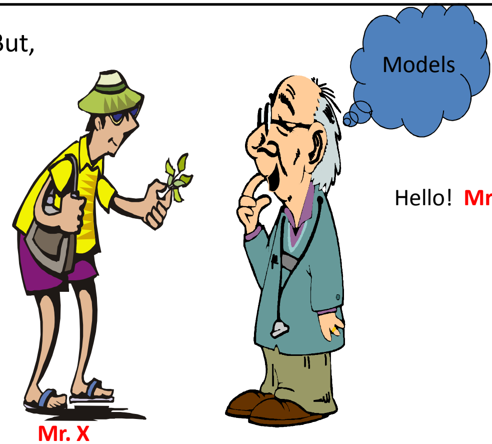
Mr. Y, ... No

Mr. Z, ...No

Not?

Mr. X

But,



Hello! **Mr. X**

Mr. X

How fast is this approach?

Indexing: A quick reference

Cluster based

Hashing based

Neural Network based

Tree based (Actual indexing scheme)

Commercial Systems and Demonstration Versions

- IBM's QBIC (Flickner et al., 1995)
- Virage's VIR Image Engine (Gupta et al, 1996)
- Excalibur's Image Retrieval Ware (Pentland et al., 1996)
- MIT's Photobook (Pentland et al., 1996)
- Chabot, now been renamed as Cypress and incorporated within Berkeley
- Digital Library project at University of California at Berkeley (UCB) (Ogle and Stonebrakers, 1995);
- Columbia University's WebSEEk (Smith and Chang, 1997b) and VisualSeek (Smith and Chang, 1997(a)
- MetaSEEk (Beigi et al., 1998)
- Carneigh-Mellon University's Informedia (Wactlar et al., 1996)
- MARS (Huang et al., 1997), University of Illinois
- Surfimage (Nastar et al, 1998), INRIA, France
- Netra (Ma and Manjunath, 1997)
- Synapse (Ravela and Manmatha, 1998b)
- PCQUERY (Cardenas et al., 1993)

Both **Altavista** and **Yahoo! search engines** use CBIR facilities, courtesy of Virage and Excalibur respectively.

Queries supported…

1. The presence of particular combination of colour, texture or shape features
2. Presence or arrangement of specific types of objects
3. Presence of named individuals to some extent…

But….,

4. locations, or events (act)
5. Depiction of particular event
6. Subjective emotions one may associate with images
7. Metadata such as who created the image, where and when

….. ???

CBIR vs Manual indexing

CBIR often performs better than keyword indexing but is limited to **level-1** searching. Keywords can provide semantics but CBIR features do not.

Avenues

- While the technology behind current CBIR systems is undoubtedly impressive, user take-up of such systems has so far been minimal. This is not because the need for such system is lacking, but because there is a mismatch between the capabilities of the technology and the needs of the users.
- CBIR systems are limited by the fact that they can operate only at primitive feature level.
- There are evidences that combining primitive image features with text keywords or hyperlinks can overcome some of these problems, though little is known about how such features can best be combined for retrieval.
- Shape matching of three dimensional objects is a more challenging task particularly when only a single 2D view is available.
- CBIR + ??? to achieve level 2/3 indexing.
- A change from static to dynamic indexing is required
- Schemes for system evaluation
- Ranking of images based on categorizing pictures
- General method for strong segmentation, where clutter and occlusion are expected
- Will learning methods help ??

Questions ?