Content Based Image Retrieval (CBIR)

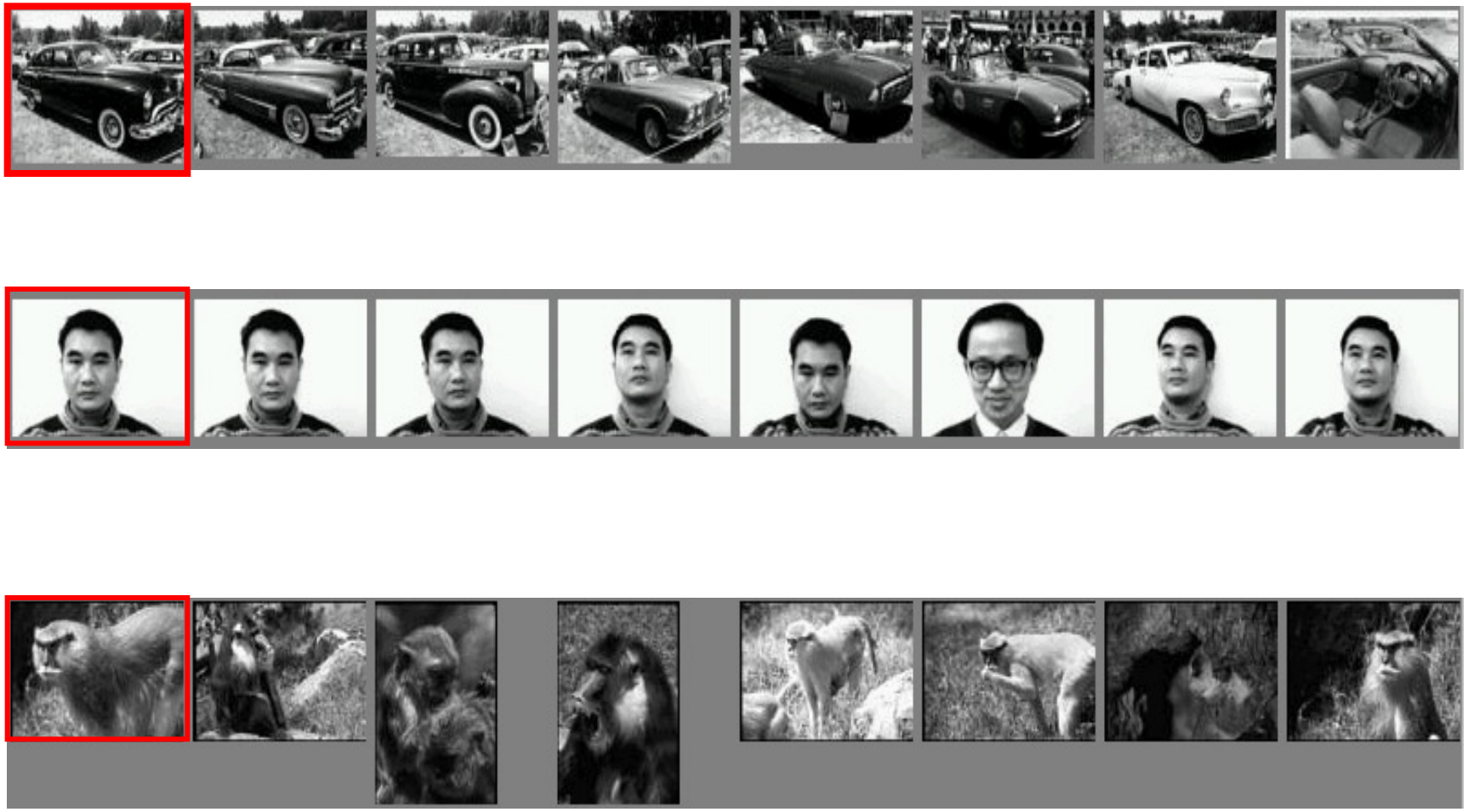
Michele Nappi, Ph.D

mnappi@unisa.it
089-963334

Content Based Image Retrieval

- What is CBIR?
  - Its purpose is to retrieve, from a database, (collections) images that are relevant to a query.
  - Finding images which are “similar” to a query.
    - Query: The whole or parts of an example image.
    - Similarity based on either the whole image or parts of the image.

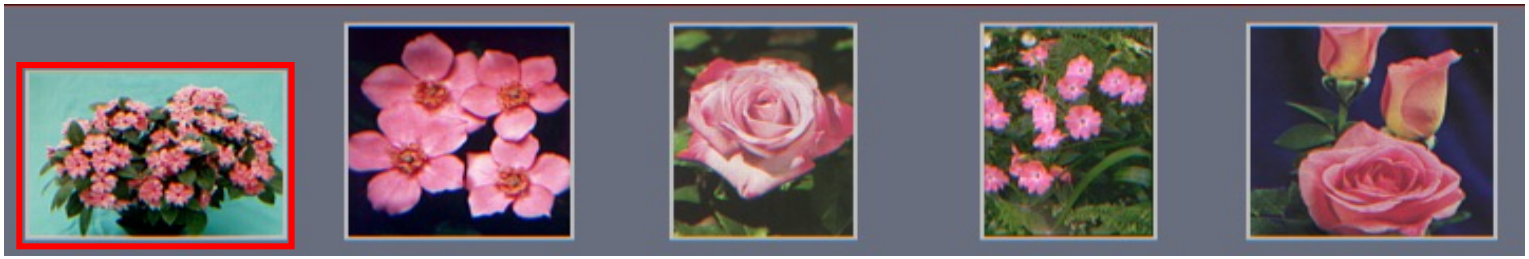
Examples of Image Similarity



Examples of Image Similarity (cont.)

Examples of Image Similarity (cont.)

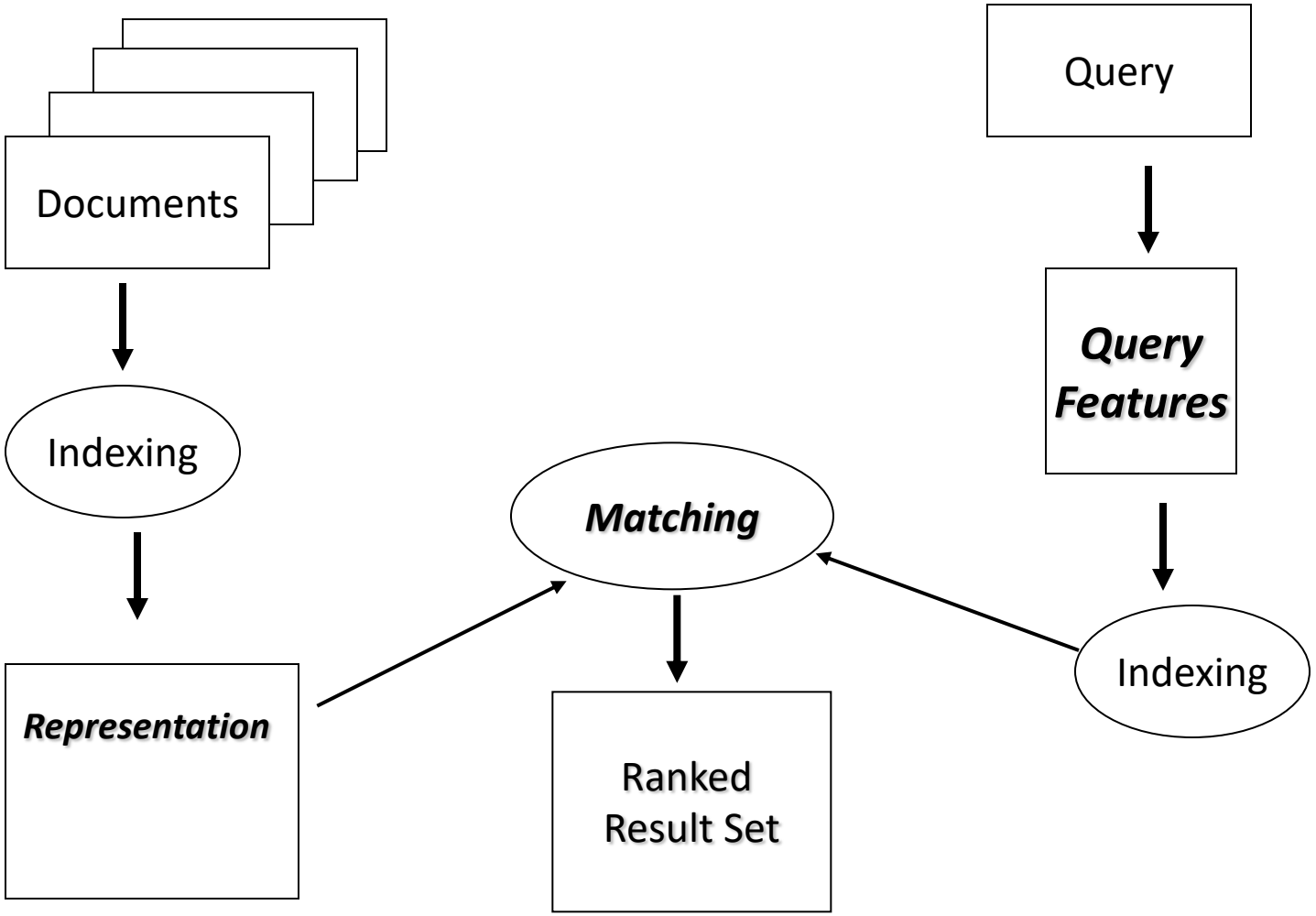
Examples of Image Similarity (cont.)



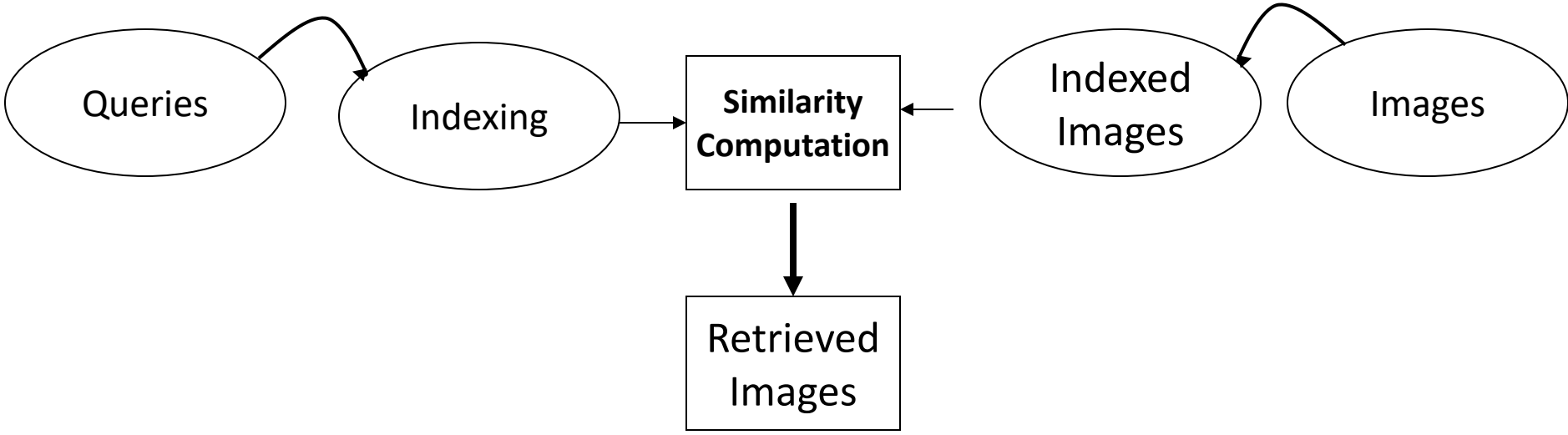
Issues/Questions

- How is it different from text retrieval?
- Is it difficult? Why?
  - How do we define/specify a query?
  - How do we index and represent a multimedia object?
  - What is an architecture of a CBIR system?

Information Retrieval



Conceptual view of CBIR



Why is Image Retrieval Difficult?

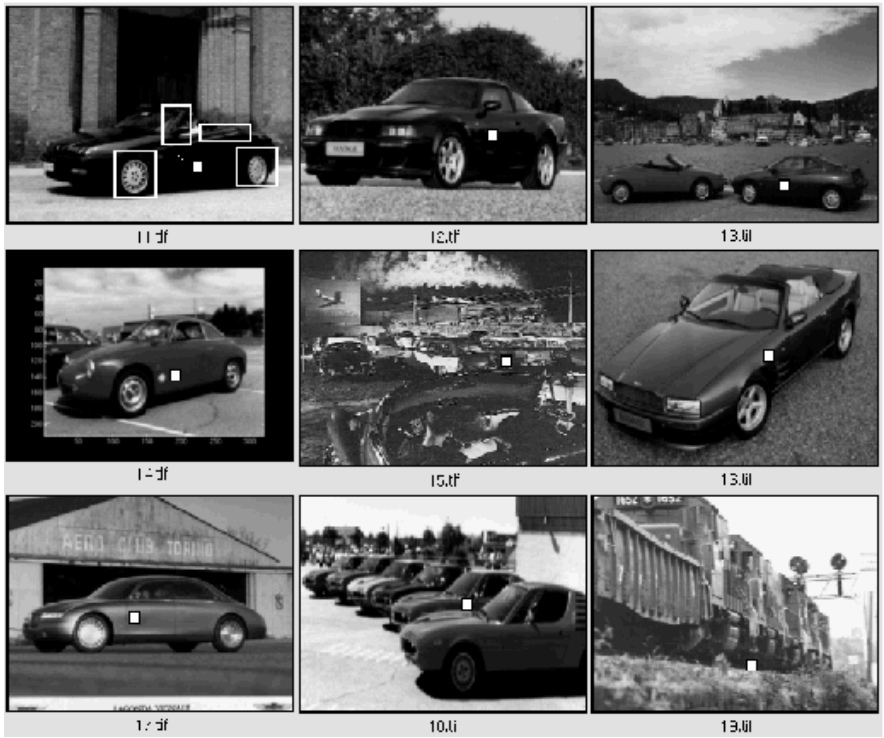
- Text retrieval
  - Word is a natural unit.
  - A word has a semantic meaning.
- Image retrieval
  - No natural unit.
  - A pixel has no semantic meaning.
  - Parts of objects
    - Example: Parts of a human body, parts of an animal
    - Difficult to automatically segment.
  - An object's image depends on many factors
    - Viewpoint, illumination, shadows etc.
  - Other complications like background.

Who want CBIR technology?

- Crime prevention
  - Fingerprint and face matching
- Journalism and Publishing
- Stock photo libraries
- Medical diagnosis
- Geographical Information Systems
- Cultural heritage
- Education and training
- Home entertainment (Video on Demand)
- Web searching

What is “Similarity”

- Ultimately user defines “similarity”.
- What is “similar”
  - Cars of a given model or all cars?
  - Red colored 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?

What is “Similarity” (cont.)

- False Alarms
  - (not qualifyng image) Immagini non significative inserite nell’insieme risposta (answer set)
- False Dismissals
  - (Qualifyng but not retrieved images) Immagini significative non inserite nell’insieme risposta

Examples of Image Similarity (cont.)

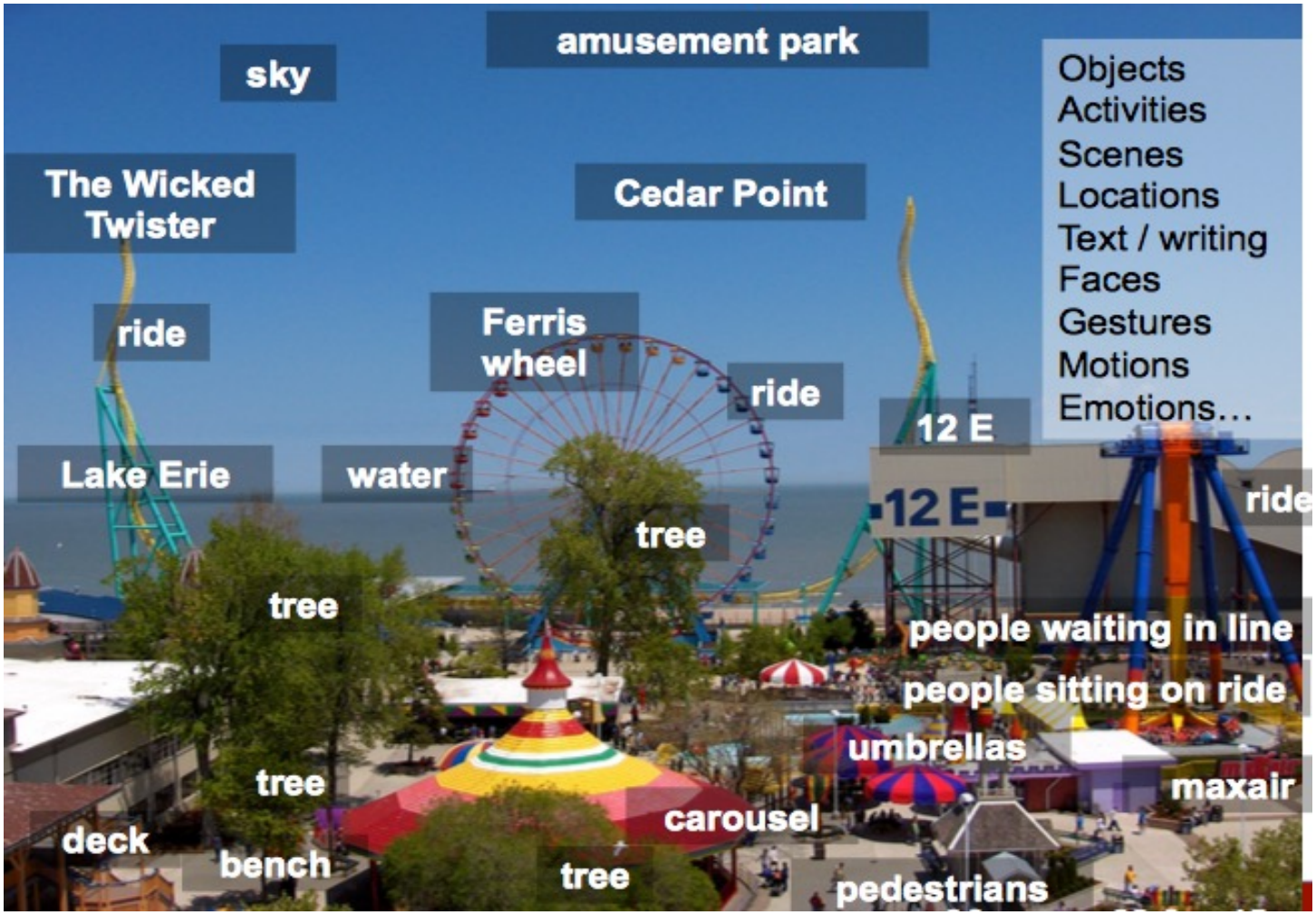


Image Features

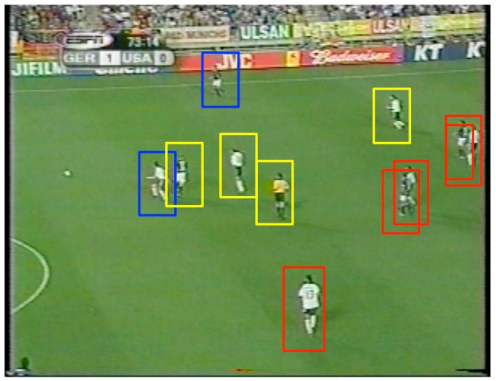
- An image is likely to be interpretable at more than one level
- An image might satisfy a visual information need by dint of its generic, specific or abstract content
- The delight and frustration of pictorial resources is that a picture can mean different things to different people

Image Features (cont.)

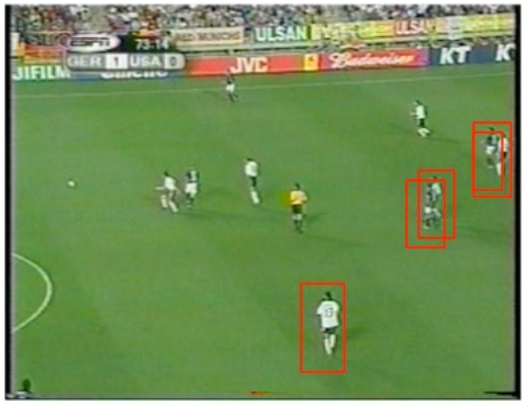
- Using Primitive features
  - Texture, Colors, Shapes, Spatial Relationships…
- Using Derived (logical) features
  - Object of a given type, person…
- Using Abstract features
  - Events, emotional significance

Image Features (cont.)

- Given an image or video choice between detection and recognition generally depends on the desired output



- This image contains *5 men walking, 4 jogging, 2 running* (object/action recognition)



- This image contains *5 men walking, 4 jogging, 2 running* (object/action recognition)
- The 5 walking men *are here* (object/action detection)

Image Features: Primitive (cont.)

1. *Find all images that have 20% of purple and 40% of green*
2. *Find all images that have 60% of green*

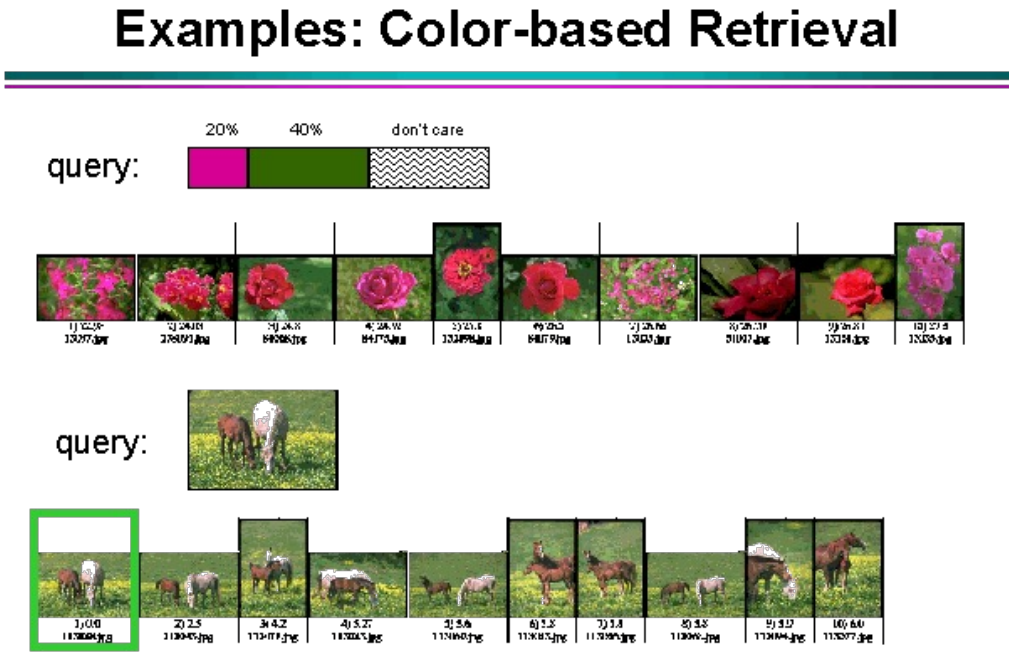


Image Features: Primitive (cont.)

Color Retrieval - Bad Example

query:

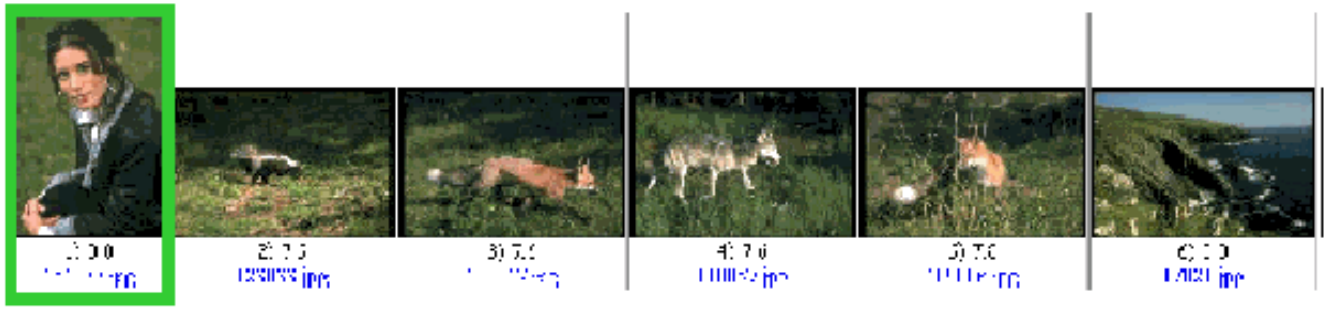
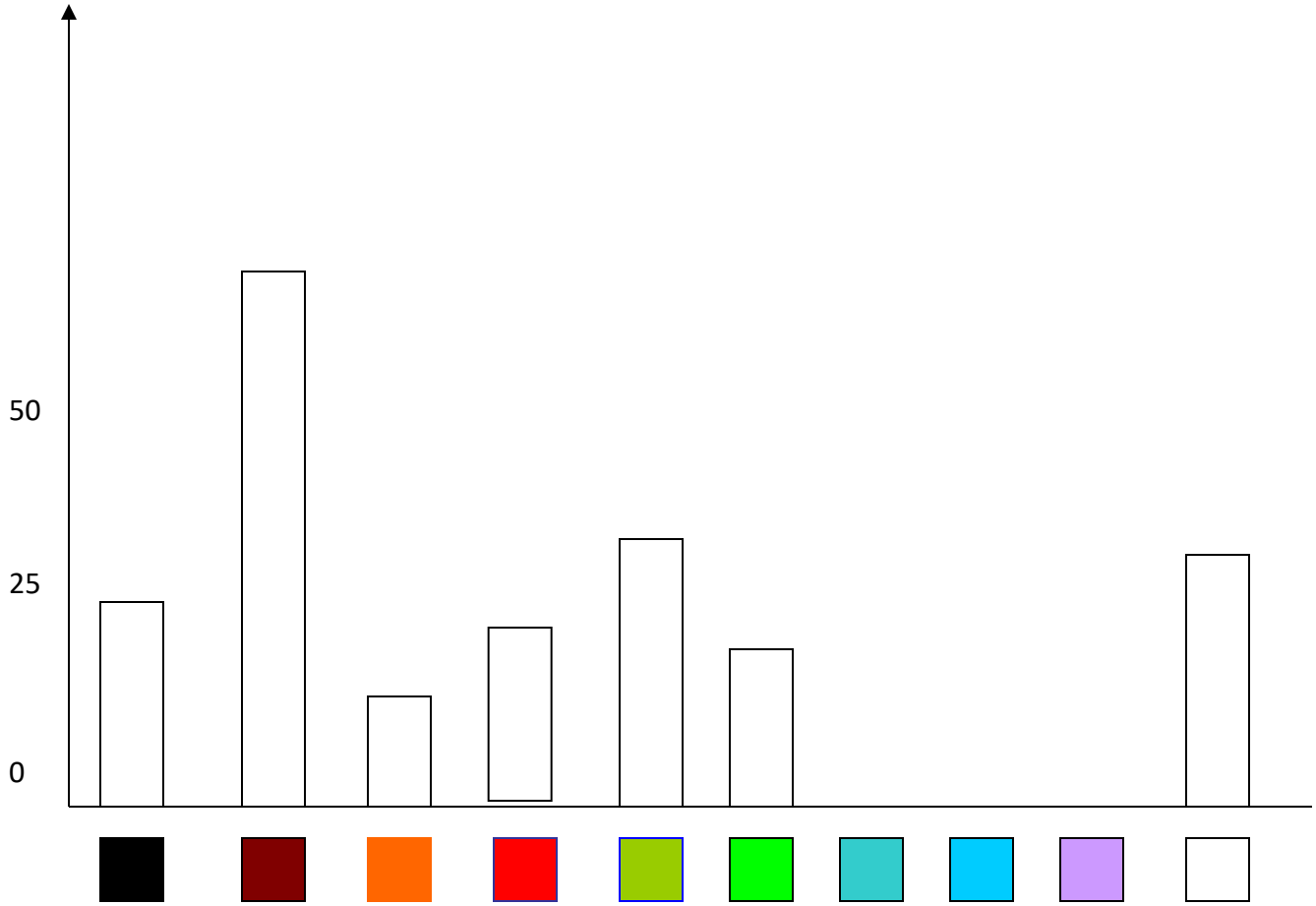


Image Features: Primitive (cont.)

- How can be implemented Retrieval using Color Histograms?

Image Features: Primitive (cont.)



Match Histogram Algorithm

- Euclidean Metric??
  - *n-dimensional space*
    - Image 1: (C_{11}, C_{12}, C_{13},……C_{1n})
    - Image 2: (C_{21}, C_{22}, C_{23},……C_{2n})
  - Huge Computing Time
- DFT + Euclideam Metrics
  - Cut off frequencies (keeping first k coefficients)
    - k<<n

Match Histogram Algorithm
(cont.)

- Th.del Parseval:
  - Let *X* be the Discrete Fourier Transform of the sequence *x*. Then we have:

$$\sum_{i=0}^{n-1} x_i^{\,2} = \sum_{u=0}^{n-1} X_u^2$$

$$D\left(\vec{x}, \vec{y}\right) = D\left(\vec{X}, \vec{Y}\right)$$

where $\vec{X}, \vec{Y}$ are Fourier transform of $\vec{x}$ and $\vec{y}$ respectively

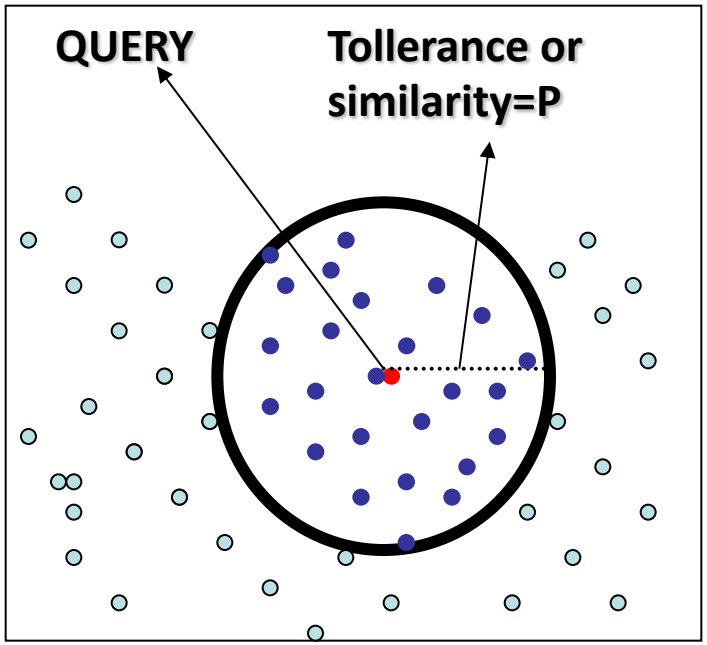
- No False Dismissals

$$D_k\left(\vec{X}, \vec{Y}\right) \le D_n\left(\vec{X}, \vec{Y}\right)$$

where $\vec{X}, \vec{Y}$ are Fourier transform of $\vec{x}$ and $\vec{y}$ respectively, and $k \le n$

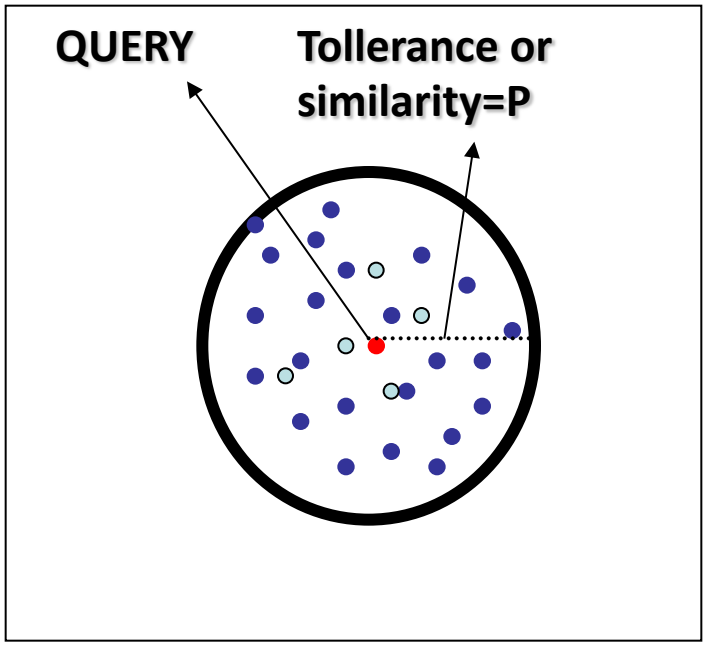
Parseval Theorem: Application

- Query
- Included in answer set [Distance from query ≤ P]
- Not included in answer set (n-dimensional) or false alarms (k-dimensional)



Space/Time domain:

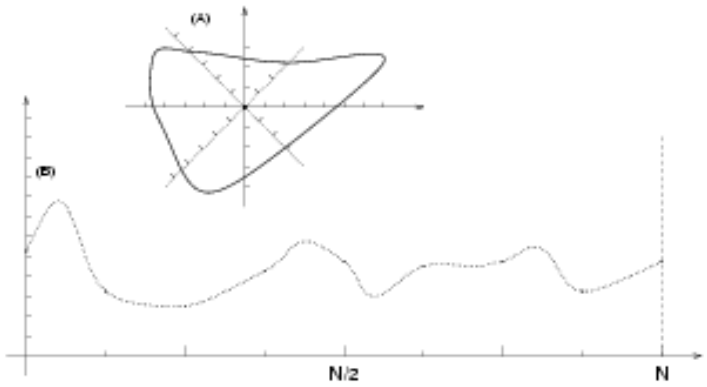
n-dimensional space



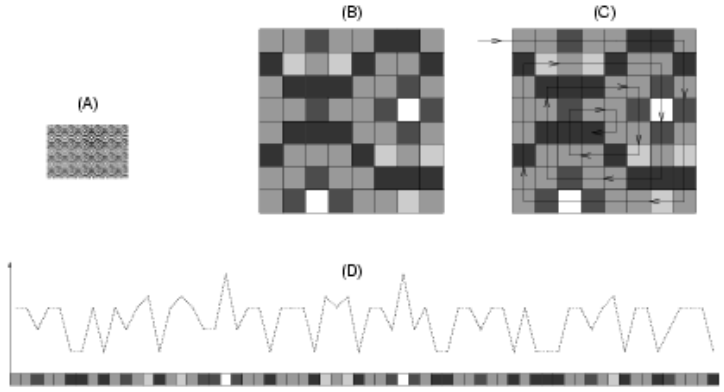
Frequencies Domain:

k-dimensional space (k<<n)

Perseval Theorem: Application (cont.)



Shape Contour



Texture

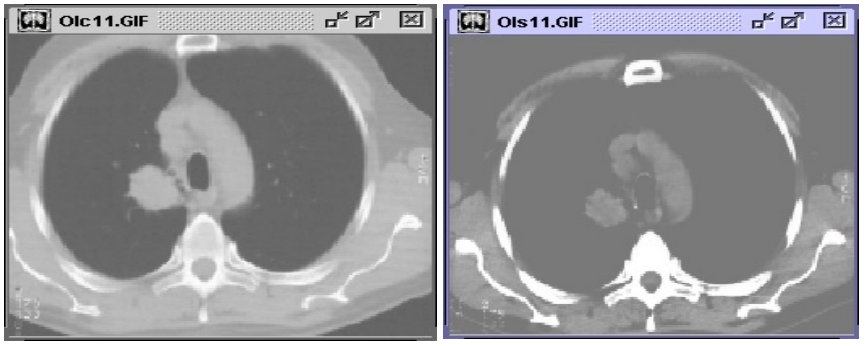
Image Features: Logical

- ***Level 2*** comprises retrieval by *derived* (sometimes known as *logical*) features, involving some degree of logical inference about the identity of the objects depicted in the image. It can usefully be divided further into:
  - retrieval of objects of a given type (e.g. “find pictures of a double-decker bus”);
  - retrieval of individual objects or persons (“find a picture of the Eiffel tower”).
- To answer queries at this level, reference to some outside store of knowledge is normally required – particularly for the more specific queries at level 2. In the first example above, some prior understanding is necessary to identify an object as a bus rather than a lorry; in the second example, one needs the knowledge that a given individual structure has been given the name “the Eiffel tower”. Search criteria at this level, particularly at level 2, are usually still reasonably objective.

Image Features: Logical (cont.)

- *Find all CT images that have a symmetric circular nodule in left lung, vertically included between spine and aorta*.

Query



Answer Set

Image Features: Abstract

- ***Level 3*** comprises retrieval by *abstract* attributes, involving a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. Again, this level of retrieval can usefully be subdivided into:
  - retrieval of named events or types of activity (e.g. “find pictures of Scottish folk dancing”);
  - retrieval of pictures with emotional or religious significance (“find a picture depicting suffering”).
- Success in answering queries at this level can require some sophistication on the part of the searcher. Complex reasoning, and often subjective judgement, can be required to make the link between image content and the abstract concepts it is required to illustrate. Queries at this level, though perhaps less common than level 2, are often encountered in both newspaper and art libraries.

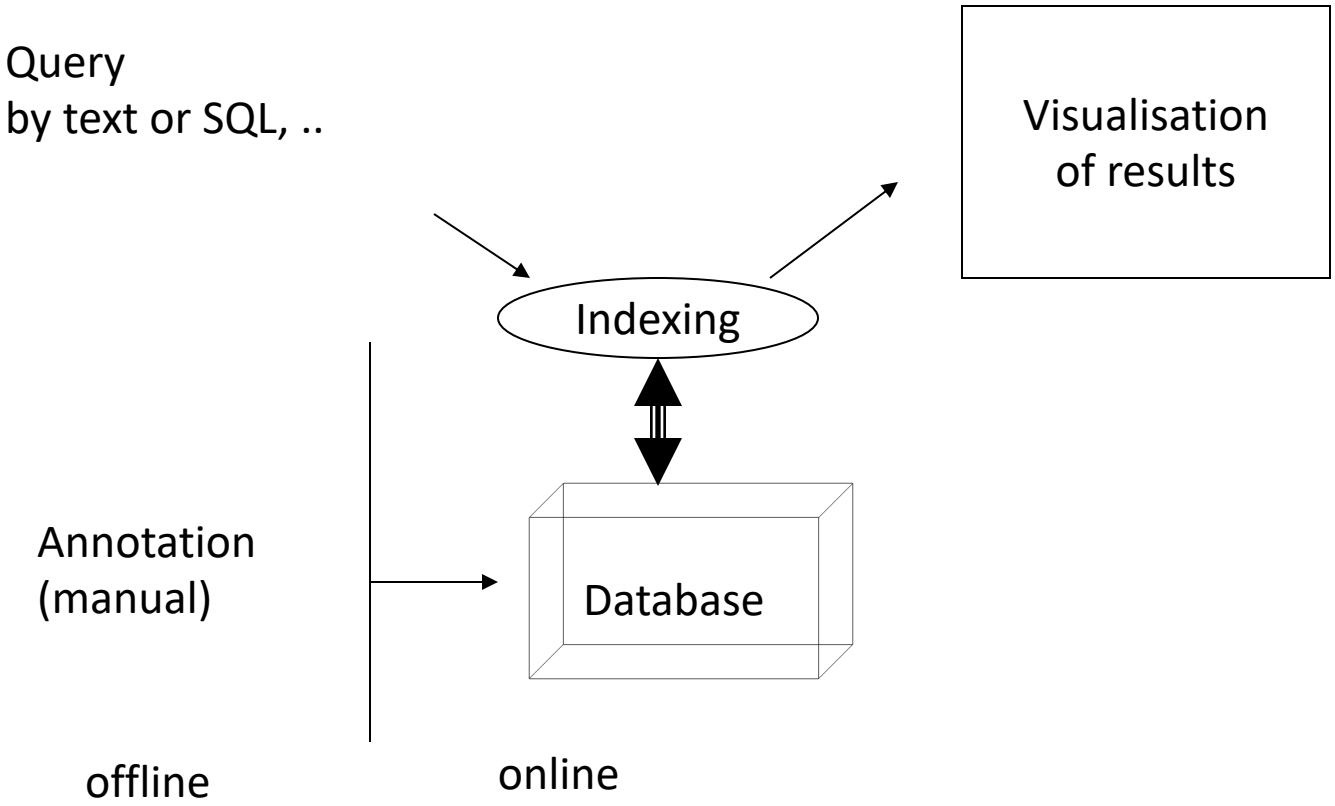
Levels of Image Description

- Content-independent metadata
  - Data which is not directly concerned with image content, but in some way related to it. (Examples: author's name, date, location etc.)
- Data which refers to the visual content of images
  - low intermediate features (colour, texture, shape etc.) known as content-dependent metadata
  - data refers to content semantics; known as content-descriptive metadata

Image Attributes

- Image attributes like color, texture, appearance and shape are often correlated with semantics.
- Examples
  - Retrieve pictures of Bill Clinton in a crowd using similarity by appearance.
  - Pictures of red and blue parrots using color.
  - Pictures of gorillas using texture.

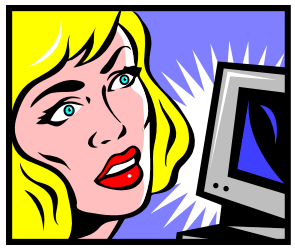
Architecture of an image retrieval system (first-generation)



User in the Loop

- Much easier to build a semi-automatic system than a fully automatic system.
- Assume that a person will use the system and will be able to provide some feedback
  - Exploit this.
  - Facilitate feedback (e.g.. relevance feedback).
  - Good user interfaces.
  - Person can easily modify query to get better results.

Architecture of Visual Information retrieval system – New Generation



Query by text/
Visual example

Browsing

Result Viewer / Visualisation

Search Engine
Indexing

Feature
Extraction
(automatic –
IP/IR)

Annotation
(manual)

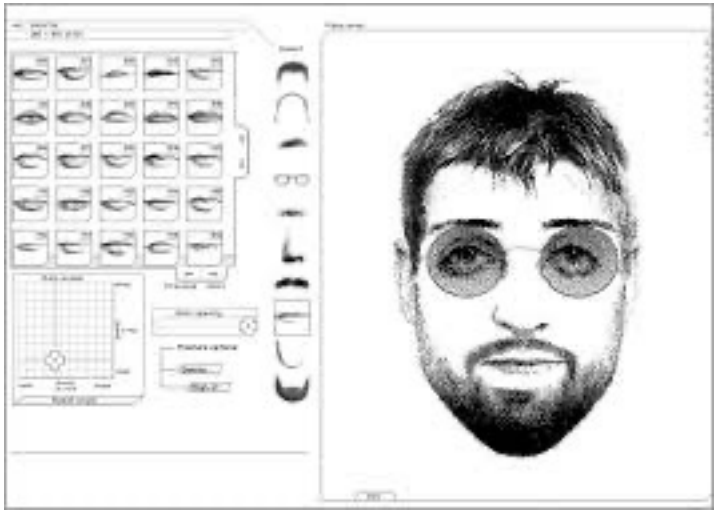
offline

online

Query Paradigms

- QbE
  - Query by Example
    - User provides as example an image

- QbS
  - Query by Sketch
    - User draws a sketch of query image using visual tools



Alternatives to CBIR technology?

- Human memory
- Image browsing
- Keyword indexing
- Classification schemes

Manual Indexing vs CBIR

Manual Indexing

- Indexing expertise widely available in image libraries
- Can use wide range of text retrieval software

BUT:

- Labour intensive
- Can be subjective and unreliable
- Difficult to capture concept of image similarity

CBIR

- Feature matching both objective and automatic
- Query formulation by visual process

BUT:

- Available features don't capture image semantics
- Features don't necessarily match human similarity judgements

Measures for CBIR

- Efficiency
  - Medium response time for retrieval
- Effectiveness
  - Quantity of false alarms and false dismissals
- Minimize False Alarms
- Minimize False Dismissals

Measures for CBIR

- Recall
  - The capability of system to retrieve all relevant images
- Precision
  - The capability of system to retrieve only relevant images

Measures for CBIR: Normalized Recall

- *TOT* is the number of images in the collection (IDB size).
- Relevant images, for each query, are ranked 1, 2,…, *REL*, where *REL* is the number of relevant images.

- Ideal Rank (*IR*)

$$IR = \sum_{r=1}^{REL} \frac{r}{REL}$$

- Average Rank (*AR*)

$$AR = \sum_{r=1}^{REL} \frac{Rank_r}{REL}$$

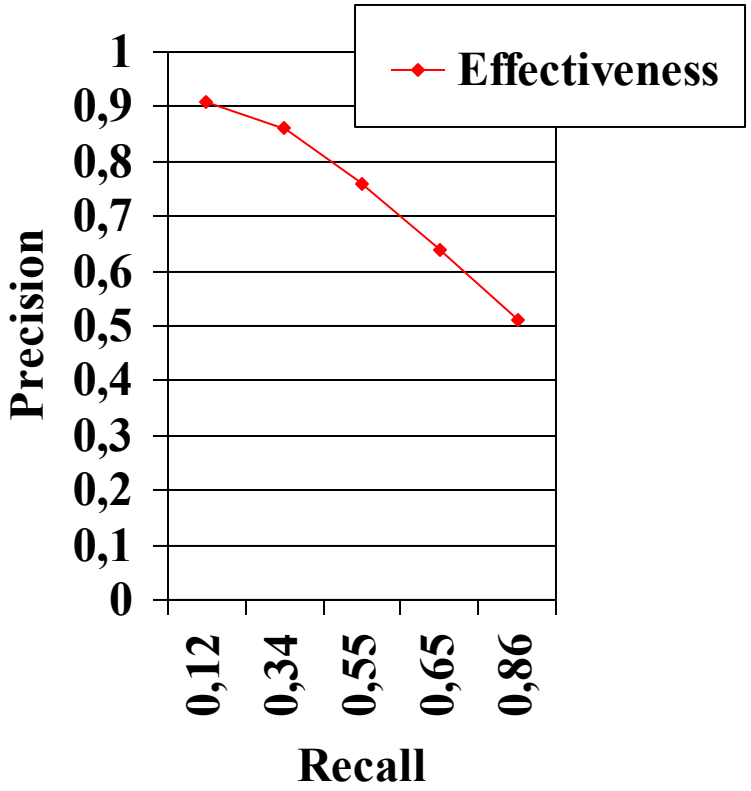
Measures for CBIR: Normalized Recall

- Effectiveness = (*AR* – *IR*)
  - [0; (*TOT-REL*)]
    - 0 (*AR* = *IR*) Perfect Retrieval
    - (*TOT* – *REL*) Worst Case
- Normalized Recall (*NR)*
  - [0;1]

$$NR = 1 - \frac{(AR - IR)}{(TOT - REL)}$$

Measures for CBIR

- Recall vs Precision



CBIR Software

- Three main commercial systems for still images
  - QBIC (IBM)
    - http://wwwqbic.almaden.ibm.com/
  - VIR Image Engine (Virage)
    - http://www.virage.com/online
  - VisualRetrievalWare (Excalibur)
    - http://www.excalib.com/

Research Directions in CBIR.

- New visual interfaces for image access
  - switch back and forth between navigation and browsing
  - query tool locate a set of candidate images and a good visualisation tool to explore this set
- New models of standards for representation of visual content
- Web search and tools
- New methodology for evaluation