



# Retrieval of Still Images by Content

*John P Eakins*

Institute for Image Data Research  
University of Northumbria at Newcastle  
United Kingdom



## Focus of talk

- What techniques are currently used for image retrieval?
- How do they work?
- How successful are they?

Emphasis is on **Content-based image retrieval (CBIR)**: *retrieval on the basis of features automatically extracted from the images themselves*

# Retrieval needs of image users

Images and image archives in widespread use.

Applications include:

- Crime prevention (fingerprint and face matching)
- Intellectual property (trademark image registration)
- Engineering and architectural design
- Journalism and publishing (stock shot photographs)
- Medical diagnosis (radiography)
- Web searching

*Remarkably little currently known about user needs*

## Levels of image query

Three levels proposed, in order of increasing abstraction:

- **Level 1** - retrieval by low-level or *primitive* attributes
  - shape
  - texture
  - colour
  - spatial location
- **Level 2** - retrieval by derived or *logical* attributes
  - objects of given type
  - named individuals or objects
- **Level 3** - retrieval by subjective or *abstract* attributes
  - specified events or types of activity
  - emotional or symbolic significance

## Traditional methods of image retrieval

### *Available techniques include:*

- Human memory
- Keyword indexing
- Browsing aids
- Classification schemes

### *Advantages:*

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

### *Disadvantages:*

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

## Content-based Image Retrieval

- Provides objective and automatic image matching
- Queries can be formulated visually
- CBIR now reasonably well established at level 1 (retrieval by *primitive* feature):
  - at least three commercial systems available
  - over 50 experimental systems reported
- Large volume of research into improved techniques

## How does CBIR work?

- Most techniques operate by comparing *feature vectors* measuring key aspects of image appearance such as colour, texture or shape, e.g.

Feature:	Col1	Col2	Tex1	Tex2	Shp1	Shp2
Image 1	0.36	0.72	0.18	0.22	0.94	0.31
Image 2	0.36	0.71	0.93	0.75	0.31	0.88

- Overall similarity between images 1 & 2 then computed using suitable metric such as Euclidean distance:

$$L_2 = \sqrt{\sum_{i=1}^n (f_i - g_i)^2}$$

## Retrieval by colour

- Commonest form of matching compares histograms of colour distribution in query and stored images
- Other techniques include colour moments and colour sets
- Works well, especially if spatial distribution of colour taken into account
- Does not depend on accurate image segmentation
- Histogram intersection technique calculates colour similarity between two images **A** and **B** as

$$\sum_{j=1}^n \min(A_j, B_j)$$

## Retrieval by texture

- Matching typically compares statistics computed from intensity distribution in query and stored images
- Works well on own or in combination with colour
- Accurate image segmentation not usually required

## Texture measures

- Earliest texture measures (e.g. Tamura's *coarseness*, *contrast* and *directionality*) based on second-order statistics of intensity distribution
- More recent measures based on coefficients derived from image transformations e.g. using Gabor filters (Manjunath and Ma)

## Retrieval by shape

- Matching typically compares shape features computed from selected regions within query and stored images
- Works adequately for isolated objects, less well for natural scenes
- Critically dependent on accurate image segmentation

## Shape matching

- Most popular techniques rely on calculating and comparing shape feature vectors
- Features may be simple (aspect ratio, circularity) or complex (Fourier descriptors, Zernike moments)
- Features may be global (describing entire shape) or local (describing part of object boundary)

# Image segmentation

## ***Problem:***

- Object boundaries can't always be extracted reliably from natural images by automatic means

## ***Ways of overcoming:***

- Identify more intelligent segmentation techniques
  - significant research activity, but no breakthroughs yet
- Use techniques to reveal local image structure which don't depend on segmentation, e.g:
  - wavelet transformation
  - filtering with Gaussian derivatives

## Retrieval by appearance (Manmatha & Ravela)

- Technique based on filtering by Gaussian derivatives
- Requires no prior segmentation of image
- Looks promising!

## Query formulation

- Most current CBIR systems based on query-by-example paradigm
- Alternative techniques include:
  - Colour and texture palettes
  - Drawing tools
  - Freehand sketch
  - Structured browsing
- Relevance feedback facilities provided in many experimental systems

## Efficiency considerations

- Sequential searching of large databases computationally infeasible
- Traditional text-based approaches such as inverted files not generally usable
- Multidimensional indexes such as R-trees, TV-trees or SS-trees usable, but not very efficient with large numbers of features
- Newer techniques such as vantage objects seem promising



## Commercial CBIR software

Three main commercial systems for still images:

- QBIC (IBM)
  - <http://www.qbic.almaden.ibm.com/>
- VIR Image Engine (Virage)
  - <http://www.virage.com/online>
- VisualRetrievalWare (Excalibur)
  - <http://www.excalib.com/>

*All operate exclusively at level 1*

## Commercial software - 2

- Offers retrieval by combination of features, e.g. colour, shape, texture, grey level or text keyword
- Available in standalone form, as application development tool, or as add-on to standard DBMS packages
- QBIC used at a number of art libraries
- Virage provides image search capability for Alta Vista Web search engine
- Excalibur provides technology for Yahoo! Image Surfer

## Experimental CBIR software

Several experimental systems available on Web, including

- Photobook (MIT)

- <http://www-white.media.mit.edu/vismod/demos/photobook/>

- WebSEEk (Columbia University)

- <http://disney.ctr.columbia.edu/WebSEEk/>

- MARS (University of Illinois)

- <http://jadzia.ifp.uiuc.edu:8001/>

- Surfimage (INRIA)

- <http://www-syntim.inria.fr/htbin/syntim/surfimage/surfimage.cgi>

## How effective is CBIR at primitive level?

- Current techniques for colour and texture matching reasonably successful
- Shape matching techniques much less effective in mimicking human similarity judgements
- Some newer techniques seem promising - but little independent evidence of effectiveness
- Current techniques good enough for use in specialized applications (e.g. colour matching of fabrics)
- Current techniques not good enough to support end-user searching for general applications

## Retrieval by semantic feature

- Still a research field - commercial application most unlikely in near future
- Research at level 2 just beginning to take off.  
Main current approaches:
  - Whole-image scene classification (normally statistical)
  - Model-based object recognition
  - Statistics-based region classification
  - Adaptive learning of image semantics
- No worthwhile progress yet at level 3

## Techniques for whole-image scene classification

- Scene classification useful both as initial filter and aid to object interpretation
- Examples from literature include:
  - IRIS (Hermes et al) - classifier trained with combined colour, texture, and shape information
  - Power spectrum analysis (Oliva et al) for natural/artificial and open/closed classification
  - Colour and texture-based city/landscape and forest/mountain classification (Vailaya and Jain)

## Automatic scene labelling (IRIS)

- Colour, texture and contour information automatically extracted from image
- Graph grammar used to generate object descriptors from low-level features
- Human experts still needed to create and modify object definitions

## Model-based object recognition

- Technique based on building detailed model of class of object to be recognized
- First reported use of technique in Brooks' aircraft recognition system ACRONYM
- Best-known current approach probably that of Forsyth et al, used to identify horses and unclothed human bodies. Technique involves three stages:
  - segmenting image into regions using edge, colour and texture information
  - generating possible descriptions of each region identified
  - classifying objects from constituents

## Statistical methods for object recognition

- Rely on statistical associations between primitive feature values and known semantic labels
- Allow automatic labelling of image regions either by type of object (car) or material (sand)
- Variety of techniques used, including Bayesian classifiers and neural networks. Examples include:
  - Region labelling with neural networks (Campbell et al)
  - *Simple Semantics* (Lew et al)
  - *Visual classes* (Schiele and Crowley)

## Adaptive semantic retrieval techniques

- Most techniques based on extension of relevance feedback principle. Examples include:
  - FourEyes (Minka and Picard)
  - Semantic Visual Templates (S F Chang et al)
  - warping feature space (Lee et al)
  - *Visual apprentice* (Jaimes and Chang)
- Systems capable of learning from experience - hence potentially quite powerful

## FourEyes (Minka and Picard)

- Tool for power annotation and retrieval of images:
  - User annotates selected regions of image
  - System learns association between semantic labels and primitive image features
  - User can then query database for all images containing specified semantic labels
- Results interesting rather than impressive

## Current state of semantic image retrieval techniques

- Growing body of research into topic - some promising results
- Main current approaches have advantages and drawbacks:
  - model-based techniques can draw on high-level reasoning, *but* are domain-specific and incapable of learning
  - statistical techniques less domain-specific *but* lack deep knowledge or ability to learn
  - adaptive techniques can continue to learn, *but* depend crucially on quality of user input

## CBIR and multimedia

- CBIR may be of most use in combination with other types of retrieval cue in multimedia systems:
  - Some evidence of synergy between text-based and content-based image indexing (e.g. Sclaroff, 1999)
  - Informedia video retrieval system (Wactlar, 1996) successfully combines text, CBIR and sound retrieval
  - Content-based navigation (Lewis, 1997), applying CBIR techniques to hyperlinks, appears promising
- Little yet known about how best to combine different types of retrieval cue in multimedia systems

## Issues for further research

- Are current techniques for shape, colour and texture retrieval going to get much better?
- Are we wasting our time looking for better ways of segmenting images?
- Is automated semantic image retrieval an achievable goal?
- Is content-based navigation a better idea than content-based retrieval?
- Should we be putting all our effort into video rather than still image retrieval techniques?
- Can we train up search intermediaries sufficiently skilled to overcome CBIR's current limitations?
- Do we really know anything about user requirements?