DRTC Workshop on
Digital Libraries: Theory and Practice
March 2003
DRTC, Bangalore

Paper: F

Image Indexing and Retrieval

Md. Farooque
Documentation Research and Training Centre
Indian Statistical Institute
Bangalore-560 059
email: *mfarooque@rediffmail.com*

Abstract

The amount of pictorial data has been growing enormously with the expansion of WWW. From the large number of images, it is very important for users to retrieve required images via an efficient and effective mechanism. To solve the image retrieval problem, many techniques have been devised addressing the requirement of different applications. Problem of the traditional methods of image indexing have led to the rise of interest in techniques for retrieving images on the basis of automatically derived features such as color, texture and shape… a technology generally referred as Content-Based Image Retrieval (CBIR). After decade of intensive research, CBIR technology is now beginning to move out of the laboratory into the marketplace. However, the technology still lacks maturity and is not yet being used in a significant scale.

1. INTRODUCTION

Images can be indexed and retrieved by textual descriptors and by image content. In textual queries, words are used to retrieve images or image sets, and in visual queries (content-based retrieval), images or image sets are retrieved by visual characteristics such as color, texture, and shape. Image retrieval strategies based on either text-based or content-based retrieval alone have their limitations. Images are often subject to a wide range of interpretations, and textual descriptions can only begin to capture the richness and complexity of the semantic content of a visual image. Human indexing of images is highly labor-intensive and limiting when large databases are involved. However, retrieval based on visual characteristics is computationally intense and has not yet reached the point where it can be efficiently used to formulate intellectually subtle queries, especially for non-specialist users.

Before going to discuss in detail about image indexing, I would like to give brief **definition** of the terms, which is related to image indexing.

Image

The definition of an image follows its Latin origin, imago, as a pictorial representation of a person, scene, or object. An image is a realistic or semi-realistic representation of a variety of subjects produced by a number of methods and in a number of different styles. The term "picture" is also frequently used in the literature; therefore, the terms "picture" and "image" are both used in appropriate contexts. In other words we can say that an image as described here is any object that could be considered graphical in nature. This includes, but is not limited to, photographs, slides, digital images and any object that is not textual in nature.

Image Attribute

"Attribute" can be defined as: "any kind of feature, component, or property of a stimulus that can be represented by an information processing system." Image attributes are not limited to purely visual characteristics, but include other cognitive, affective, or interpretive responses to an image, such as those describing spatial, semantic, symbolic, or emotional characteristic.

Pixel

The picture elements that make up an image, similar to grains in a photograph or dots in a half-tone. Each pixel can represent a number of different shades or colors, depending upon how much storage space is allocated for it.

8-bit image

A digital image that can include as many as 256 possible colors. In this kind of image, 8 bits are allocated for the storage of each pixel, allowing 2 to the power of 8 (or 256) colors to be represented.

24-bit image

A digital image that can include approximately 16 million possible colors. In this kind of image, 24 bits are allocated for the storage of each pixel, allowing 2 to the power of 24 (or more than 16 million) colors to be represented.

Palette

The set of colors that appear in a particular digital image. Becomes part of a color look-up table.

Adaptive palette

Image-specific set of colors chosen to most closely represent those in the original source. Part of a custom color look-up table.

System palette
A color palette chosen by a computer system and applied to all digital images.

Resolution, Image
Number of pixels (in both height and width) making up an image. The higher the resolution of an image, the greater its clarity and definition.

Bit Mapped Image
An image created from a series of bits and bytes that form pixels. Each pixel can vary in color or gray-scale value. Also known as a raster image.

Vector Graphic
A digital image encoded as formulas that represent lines and curves.

Thumbnail (Browse image)
A small image (usually derived from a larger one). Browse images permit a user to view a dozen or more images on a single screen.

Lossless Compression
Process that reduces the storage space needed for an image file without loss of data. If a digital image that has undergone lossless compression is decompressed, it will be identical to the digital image before it was compressed. Document images (i.e., in black and white, with a great deal of white space) undergoing lossless compression can often be reduced to one-tenth of their original size; continuous-tone images under lossless compression can seldom be reduced to one-half or one-third of their original size.

Lossy Compression
A process that reduces the storage space needed for an image file. If a digital image that has undergone lossy compression is decompressed, it will differ from the image before it was compressed (though this difference may be difficult for the human eye to detect). The most effective lossy-compression algorithms work by discarding information that is not easily perceptible to the human eye.

Gray Scale
The range of shades of gray in an image. The number of grays, or steps between black and white determines the gray scales of scanners and terminals that they can recognize and reproduce

2. TYPES OF IMAGES FOUND IN LIBRARIES ARE

- Slides
- Photographs (+ve)
- Photographs (-ve)
- Video
- Posters
- Prints
- Paintings
- Drawings
- Transparencies
- Fabrics Films, etc

3. IMAGE ATTRIBUTES

In a study conducted by School of Information and Library Studies, University at Buffalo, image attributes as described by participants fell into three main types: "Perceptual," " Interpretive," and "Reactive." Perceptual attributes are those named in direct response to a visual stimulus, such as Color ("red") or Object ("confetti"). Interpretive attributes are those, which require both interpretation of perceptual cues and application of a general level of knowledge or inference from that knowledge to name the attribute. This category includes such attributes as Style ("modern art") and Atmosphere ("dreamy"). Reactive attributes describe personal reactions to the image, such as uncertainty, confusion, and "liking" the image. Attributes were then grouped into higher-level Classes based upon conceptual and functional relationships. The Classes were the most useful level for describing attribute distributions.

Forty-eight image attributes and twelve higher-level Classes of attributes were derived from the data, using the techniques described above. The twelve higher-level Classes, with brief definitions, are listed below.

- **Literal Object (perceptual)**.
  This class contains items, which are classified as being literal (visually perceived) objects.
- **People (perceptual)**
  The presence of a human form.
- **People Qualities (Interpretive)**
  Interpretive qualities such as the nature of the relationship among people depicted in an image, their mental or emotional state, or their occupation.
- **Art Historical Information (Interpretive)**
  Information, which is related to the production context of the representation, such as Artist, Medium, Style, and Type.
- **Color (Perceptual)**
  Includes both specific named colors and terms relating to various aspects of color value, hue, and tint.
- **Location (Perceptual)**
  Includes attributes relating to both general and specific locations of picture components.
- **Visual Elements (Perceptual)**
  Includes those percepts such as Orientation, Shape, Visual Component (line, details, lighting) or Texture.
- **Description (Perceptual)**
  Includes descriptive adjectives and words referring to size or quantity.
- **Abstract Concepts (Interpretive)**
  Abstract, thematic, and symbolic image descriptors.
- **Content/Story (Interpretive)**
  Attributes relating to a specific instance being depicted, such as Activity, Event, and Setting.
- **Personal Reaction**
  Personal reactions to the image.
- **External Relationship**
  Comparison of attributes within a picture or among pictures or reference to an external entity.

It is found that these twelve attributes constitute around 96.2% user approach for retrieving the images.

The distribution is given below,

Class	Percentage (%)
Literal Object	17.7
Content/Story	14.9
Abstract	14.4
Color	12.3
People Qualities	8.1
People	7.2
Description	5.2
Art Historical Information	5.8
Personal Reaction	5.0
External Relation	4.1
Visual Elements	3.8
Location	1.6

4. IMAGE INDEXING

The objective of image indexing is to retrieve similar images from an image database for a given query image (i.e., a pattern image). Each image has its unique feature. Hence image indexing can be implemented by comparing their features, which are extracted from the images. The criterion of similarity among images may be based on the features such as color, intensity, shape, location and texture, and above mentioned other image attributes. Current Image indexing techniques are of two types,

1. Textual (manual)
2. Content- based (automated)

4.1. Textual

It is very simple techniques; keeping in mind the user approach keywords are given for a particular image. These includes

- Caption indexing
- Keyword additions
- Standard subject headings, Classification, etc.

The problem with this indexing is that it is

- Labor intensive
- More prone to inter indexer consistency problems than indexing of text
- Of-ness, thing-ness, about-ness ambiguities

4.2. Content-based

It is also known as automated indexing, in this technique images are indexed based on their content like color, shape, direction, texture, spatial relation etc. This kind of indexing is taken care by software itself, algorithms are developed which can differentiate the color, shape, textures etc. The image retrieved through this technique is known as Content Based Image Retrieval (CBIR).

5. WHAT IS CBIR?

The earliest use of the term *content-based image retrieval* in the literature seems to have been to describe the experiments into automatic retrieval of images from a database by colour and shape feature. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as colour, texture and shape) that can be automatically extracted from the images themselves. The features used for retrieval can be either primitive or semantic, but the extraction process must be predominantly automatic. Retrieval of images by manually-assigned keywords is definitely not CBIR as the term is generally understood – even if the keywords describe image content.

CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. One of the key issues with any kind of image processing is the need to extract useful information from the raw data (such as recognizing the presence of particular shapes or textures) before any kind of reasoning about the image's contents is possible. Image databases thus differ fundamentally from text databases, where the raw material are in pixel form (words stored as ASCII character strings)]. CBIR draws many of its methods from the field of image processing and computer vision, and is regarded by some as a subset of that field. It differs from these fields principally through its emphasis on the retrieval of images with desired characteristics from a collection of significant size. Image processing covers a much wider field, including image enhancement, compression, transmission, and interpretation. While there are grey areas (such as object recognition by feature analysis), the distinction between mainstream image analysis and CBIR is usually fairly clear-cut. An example may make this clear. Many police forces now use automatic face recognition systems. Such systems may be used in one of two ways. Firstly, the image in front of the camera may be compared with a single individual's database record to verify his or her identity. In this case, only two images are matched, a process few observers would call CBIR. Secondly, the entire database may be searched to find the most closely matching images. This is a genuine example of CBIR.

Research and development issues in CBIR cover a range of topics, many shared with mainstream image processing and information retrieval. Some of the most important are:

- understanding image users' needs and information-seeking behaviour
- identification of suitable ways of describing image content
- extracting such features from raw images
- providing compact storage for large image databases
- matching query and stored images in a way that reflects human similarity judgements
- efficiently accessing stored images by content
- providing usable human interfaces to CBIR systems

5.1. CBIR Techniques

Most of the CBIR indexing techniques based on color, shape and texture.

5.1.1. Color

Retrieving images based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors). Current research is attempting to segment color proportion by region and by spatial relationship among several color regions

5.1.2. Texture

Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated (Tamura, Mori & Yamawaki, 1978). However, the problem is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as ``silky'', or ``rough''.

5.1.3. Shape

Queries for shapes are generally achieved by selecting an example image provided by the system or by having the user sketch a shape. The primary mechanisms used for shape retrieval include identification of features such as lines, boundaries, aspect ratio, and circularity, and by identifying areas of change or stability via region growing and edge detection. Of particular concern is the problem of dealing with images having overlapping or touching shapes.

5.2. What Standards are Relevant to CBIR?

Potentially, a number of different types of standard could affect, and be affected by, developments in CBIR technology. These include:

- network protocols such as TCP/IP, governing the transmission of data between hosts holding stored data and clients running applications making use of such data;
- image storage formats such as TIFF or JPEG, specifying how images should be encoded for long-term storage or transmission;
- image data compression standards such as JPEG and MPEG-2, specifying standard methods for compressing image (and video) data for efficient transmission;
- database command languages such as SQL, providing a standard syntax for specifying queries to a relational database;
- metadata standards such as RDF, providing a framework for describing the content of multimedia objects, and languages such as XML in which to write content descriptions.

Some of these standards are unlikely to pose any implications for the development of CBIR. For example, low-level network transmission protocols such as TCP/IP handle all types of data in the same way, regarding them simply as packets of binary data whose meaning, if any, is left to the sending and receiving applications to sort out. CBIR applications are no different from any others in this respect. (Higher-level protocols such as Z39.50 are different issue). Similarly, storage formats for image data are not really a CBIR issue either. All commercial and many experimental CBIR systems can accept images in a wide variety of formats, converting them to their own native format for feature extraction if required. Image matching and retrieval is always performed on a database of extracted features, with the original images used purely for display purposes. Hence the format in which these images are stored has no effect on the operations of query formulation, matching or retrieval.

The other types of standard listed above could have implications for CBIR (and *vice versa),* even though these implications may well turn out to be minimal in practice. Three main areas of potential impact can be identified: image compression, query specification, and metadata description.

5.3. Practical Applications of CBIR

A wide range of possible applications for CBIR technology has been identified. Some of these are listed below.

- Crime prevention
- The military
- Intellectual property
- Architectural and engineering design
- Fashion and interior design
- Journalism and advertising
- Medical diagnosis
- Geographical information and remote sensing systems
- Cultural heritage

- Education and training
- Home entertainment
- Web searching.

5.4. Available CBIR Software

Despite the shortcomings of current CBIR technology, several image retrieval systems are now available as commercial packages, with demonstration versions of many others available on the Web. Some of the most prominent of these are described below.

5.4.1. Commercial systems

QBIC

IBM's QBIC system is probably the best-known of all image content retrieval systems. It is available commercially either in standalone form, or as part of other IBM products such as the DB2 Digital Library. It offers retrieval by any combination of colour, texture or shape – as well as by text keyword. Image queries can be formulated by selection from a palette, specifying an example query image, or sketching a desired shape on the screen. The system extracts and stores colour, shape and texture features from each image added to the database, and uses R*-tree indexes to improve search efficiency . At the time of searching, the system matches appropriate features from query and stored images, calculates a similarity score between the query and each stored image examined, and displays the most similar images on the screen as thumbnails. The latest version of the system incorporates more efficient indexing techniques, an improved user interface, the ability to search grey-level images, and a video storyboarding facility . An online demonstration, together with information on how to download an evaluation copy of the software, is available on the World-Wide Web at http://wwwqbic.almaden.ibm.com/ .

Virage

Another well-known commercial system is the VIR Image Engine from Virage, Inc. This is available as a series of independent modules, which systems developers can build in to their own programs. This makes it easy to extend the system by building in new types of query interface, or additional customized modules to process specialized collections of images such as trademarks. Alternatively, the system is available as an add-on to existing database management systems such as Oracle or Informix. An on-line demonstration of the VIR Image Engine can be found at http://www.virage.com/online/. A high-profile application of Virage technology is AltaVista's AV Photo Finder (http://image.altavista.com/cgi-bin/avncgi), allowing Web surfers to search for images by content similarity. Virage technology has also been extended to the management of video data ; details of their commercial Videologger product can be found on the Web at http://www.virage.com/market/cataloger.html.

Excalibur

A similar philosophy has been adopted by Excalibur Technologies, a company with a long history of successful database applications, for their Visual RetrievalWare product . This product offers a variety of image indexing and matching techniques based on the company's own proprietary pattern recognition technology. It is marketed principally as an applications development tool rather then as a standalone retrieval package. Its best-known application is probably the Yahoo! Image Surfer, allowing content-based retrieval of images from the World-wide Web. Further information on Visual RetrievalWare can be found at http://www.excalib.com/, and a demonstration of the Yahoo! Image Surfer at http://isurf.yahoo.com/. Excalibur's product range also includes the video data managementsystem Screening Room (http://www.excalib.com/products/video/screen.html).

5.4.2. Experimental Systems
A large number of experimental systems have been developed, mainly by academic institutions, in order to demonstrate the feasibility of new techniques. Many of these are available as demonstration versions on the Web. Some of the best-known are described below.

Photobook
The Photobook system from Massachusetts Institute of Technology (MIT) has proved to be one of the most influential of the early CBIR systems. Like the commercial systems above, aims to characterize images for retrieval by computing shape, texture and other appropriate features. Unlike these systems, however, it aims to calculate *information-preserving* features, from which all essential aspects of the original image can in theory be reconstructed. This allows computation of features relevant to a particular type of search at the time of search, giving greater flexibility at the expense of speed. The system has been successfully used in a number of applications, involving retrieval of image textures, shapes, and human faces, each using features based on a different model of the image. More recent versions of the system allow users to select the most appropriate feature type for the retrieval problem at hand from a wide range of alternatives . Further information on Photobook, together with an online demonstration, can be found at http://www-white.media.mit.edu/vismod/demos/photobook/. Although Photobook itself never became a commercial product, its face recognition technology has been incorporated into the FaceID package from Viisage Technology (http://www.viisage.com/), now in use by several US police departments.

Chabot
Another early system which has received wide publicity is Chabot , which provided a combination of text-based and colour-based access to a collection of digitized photographs held by California's Department of Water Resources. The system has now been renamed Cypress, and incorporated within the Berkeley Digital Library project at the University of California at Berkeley (UCB). A demonstration of the current version of Cypress (which no longer appears to have CBIR capabilities) can be found at http://elib.cs.berkeley.edu/cypress.html. Rather more impressive is UCB's recently-developed Blobworld software, incorporating sophisticated colour region searching facilities (http://elib.cs.berkeley.edu/photos/blobworld/).

VisualSEEk
The VisualSEEk system is the first of a whole family of experimental systems developed at Columbia University, New York. It offers searching by image region colour, shape and spatial location, as well as by keyword. Users can build up image queries by specifying areas of defined shape and colour at absolute or relative locations within the image. The WebSEEk system aims to facilitate image searching on the Web. Web images are identified and indexed by an autonomous agent, which assigns them to an appropriate subject category according to associated text. Colour histograms are also computed from each image. At search time, users are invited to select categories of interest; the system then displays a selection of images within this category, which users can then search by colour similarity. Relevance feedback facilities are also provided for search refinement. For a demonstration of WebSEEk in action, see http://disney.ctr.columbia.edu/WebSEEk/ Further prototypes from this group include VideoQ , a video search engine allowing users to specify motion queries, and MetaSEEk, a meta-search engine for images on the Web.

MARS

The MARS project at the University of Illinois is aimed at developing image retrieval systems which put the user firmly in the driving seat. Relevance feedback is thus an integral part of the system, as this is felt to be the only way at present of capturing individual human similarity judgements. The system characterizes each object within an image by a variety of features, and uses a range of different similarity measures to compare query and stored objects. User feedback is then used to adjust feature weights, and if necessary to invoke different similarity measures . A demonstration of the MARS system can be viewed at http://jadzia.ifp.uiuc.edu:8001/

Informedia

In contrast to the systems described above, the Informedia project was conceived as a multimedia video-based project from the outset. Its overall aims are to allow full content search and retrieval of video by integrating speech and image processing. The system performs a number of functions. It identifies video scenes (not just shots) from analysis of colour histograms, motion vectors, speech and audio soundtracks, and then automatically indexes these 'video paragraphs' according to significant words detected from the soundtrack, text from images and captions, and objects detected within the video clips. A query is typically submitted as speech input. Thumbnails of keyframes are then displayed with the option to show a sentence describing the content of each shot, extracted from spoken dialogue or captions, or to play back the shot itself. Many of the system's strengths stem from its extensive evaluation with a range of different user populations . Its potential applications include TV news archiving, sports, entertainment and other consumer videos, and education and training. The Informedia website is at http://informedia.cs.cmu.edu/; the Mediakey Digital Video Library System from Islip Media, Inc, a commercially-available system based on Informedia technology, is at http://www.islip.com/fprod.htm.

Surfimage

An example of European CBIR technology is the Surfimage system from INRIA, France [Nastar et al, 1998]. This has a similar philosophy to the MARS system, using multiple types of image feature which can be combined in different ways, and offering sophisticated relevance feedback facilities. See http://www-syntim.inria.fr/htbin/syntim/surfimage/surfimage.cgi for a demonstration of Surfimage in action.

Netra

The Netra system uses colour texture, shape and spatial location information to provide region-based searching based on local image properties. An interesting feature is its use of sophisticated image segmentation techniques. A Web demonstration of Netra is available at http://vivaldi.ece.ucsb.edu/Netra.

Synapse

This system is an implementation of *retrieval by appearance* using whole image matching. A demonstration of Synapse in action with a variety of different image types can be found at http://cowarie.cs.umass.edu/~demo/.

6. SEARCH ENGINES

Some of the search engines for image searching are listed below:

Picsearch

http://www.picsearch.com/

Image search engine that is family friendly, simple to use and informative, for finding images, pictures, photographs or animations on the web. The search interface is available in English, German and Swedish.

Ditto

http://ditto.com/
The visual search engine that provides a unique way to visually search the web for pictures, images and photos.

Animation Search
http://www.animationlibrary.com/
Over 13,000 animations and a very powerful search engine to navigate through them.

Kamat House of Pictures
http://www.kamat.com/picturehouse/
Photographic and picture search engine on India.

Fagan Finder – Images
http://www.faganfinder.com/img/index.shtml
Search for graphics, photos, and clip-art, with the Internet's largest picture search tools, all on one page.

Ithaki Meta Search
http://www.ithaki.net/images/
Metacrawler that finds pictures in 8 image search engines simultaneously.

3Dup
http://3dup.com/
Search engine for 3D, multimedia and audiovisual industry. 11 Languages.

Landscape Architecture Image Resource
http://www.lair.umd.edu/default.htm
The LAIR website is a searchable database of over 3,200 catalogued landscape images which are available for use for educational purposes.

PicturesNOW!
http://www.picturesnow.com
A powerful visual search engine and content solution for browsing and searching the webs largest free picture network. Still and streaming media includes sports, entertainment, autos, fine art, animals.

Lockheed Martin Digital Photo Collection
http://www.photos.external.lmco.com/
Image base of aviation images. Approximately 1000 images in this collection, images are free for educational or editorial use.

Mukpuddy Animation
http://www.mukpuddy.com
Insanity is free, come get some! Totally original web cartoons starring Dirk Banzai and Skid & Hokey

Internet Image Search Engine
http://www.photoseek.net/
Search engine for finding free photos, images and pictures on many different subjects.

Google Image Search

http://images.google.com/
Beta version works on text in file names and captions only not content-based.

PixelFind Web Graphics Search Engine
http://www.pixelfind.com/index.shtml
Find free graphics for web sites: backgrounds, lines, bullets, icons, clip-art, more. Completely searchable and categorized.

AltavistaImageSearch
http://www.altavista.com/cgi-bin/query?mmdo=1&stype=simage
Find images and photos on the web, or through one of Alta Vista's media partners.

7. CONCLUSION

Although university researchers may experiment with standalone image retrieval systems to test the effectiveness of search algorithms, this is not at all typical of the way they are likely to be used in practice. The experience of all commercial vendors of CBIR software is that system acceptability is heavily influenced by the extent to which image retrieval capabilities can be embedded within users' overall work tasks. Trademark examiners need to be able to integrate image searching with other keys such as trade class or status, and embed retrieved images in official documentation. Engineers will need to modify retrieved components to meet new design requirements. It is important to stress that CBIR is never more than the means to an end.

8. REFERENCES

1. Leu, J -G. (2001). On indexing the periodicity of image textores. *Image and vision computing, 19*, 987-1000.
2. Cheu, H -L. (2001). An analysis of image retrieval tasks in the field of art history. *Information processing and management, 37*, 701-720.
3. Goodrum, A., & Spink, A. (2001). Image searching on the Excite Web search engine. *Information processing and management, 37*, 295-311.
4. Roth, V. (1999). Content based retrieval from digital video. *Image and vision computing, 17*, 531-540.
5. Nes, N., & Kersten, M. *Region-based Indexing in an Image Database*, from http://www.cwi.nl/~niels/Pub/cisst/node1.html
6. Tsai, C -Y., Chen, A. L. P., & Essing, K.*Efficient image retrieval approaches for different similarity requirement*, from http://db.nthu.edu.tw/alp/course/cs4570-89/image.doc
7. Marsicoi, M. d., Cinque, L., & Levialdi, S. (1997). Indexing pictorial document by their content: a survey of current techniques journal. *Image and vision computing, 15*, 119-141.
8. Shih, T. k., Wang, C.-S., & Kao, C.-H.*A flexible content-based image retrieval system integrating with color, shape and spatial relation*, from http://www.mine.tku.edu.tw/mns2001/MNS06/doc
9. *Photo study collection database*. from http://getty.edu/research/tools/psc
10. *QBIC(TM) -- IBM's Query By Image Content*. from http://wwwqbic.almaden.ibm.com/