JTAP

JISC Technology Applications

Content-based Image Retrieval

John Eakins
Margaret Graham
University of Northumbria at Newcastle

Report: 39

JISC Technology Applications Programme

Joint Information Systems Committee

October 1999

Content-based Image Retrieval

John Eakins
Margaret Graham
University of Northumbria at Newcastle

The JISC Technology Applications Programme is an initiative of the Joint Information Systems Committee of the Higher Education Funding Councils.

For more information contact:
Tom Franklin
JTAP Programme Manager
Computer Building
University of Manchester
Manchester
M13 9PL

email: t.franklin@manchester.ac.uk
URL: http://www.jtap.ac.uk/

Table of Contents

Executive summary

The aim of this report is to review the current state of the art in content-based image retrieval (CBIR), a technique for retrieving images on the basis of automatically-derived features such as colour, texture and shape. Our findings are based both on a review of the relevant literature and on discussions with researchers and practitioners in the field.

The need to find a desired image from a collection is shared by many professional groups, including journalists, design engineers and art historians. While the requirements of image users can vary considerably, it can be useful to characterize image queries into three levels of abstraction: *primitive* features such as colour or shape, *logical* features such as the identity of objects shown, and *abstract* attributes such as the significance of the scenes depicted. While CBIR systems currently operate effectively only at the lowest of these levels, most users demand higher levels of retrieval.

Users needing to retrieve images from a collection come from a variety of domains, including crime prevention, medicine, architecture, fashion and publishing. Remarkably little has yet been published on the way such users search for and use images, though attempts are being made to categorize users' behaviour in the hope that this will enable their needs to be better met in the future.

Current indexing practice for images relies largely on text descriptors or classification codes, supported in some cases by text retrieval packages designed or adapted specially to handle images. Again, remarkably little evidence on the effectiveness of such systems has been published. User satisfaction with such systems appears to vary considerably.

CBIR operates on a totally different principle from keyword indexing. Primitive features characterizing image content, such as colour, texture, and shape, are computed for both stored and query images, and used to identify (say) the 20 stored images most closely matching the query. Semantic features such as the type of object present in the image are harder to extract, though this remains an active research topic. Video retrieval is a topic of increasing importance – here, CBIR techniques are also used to break up long videos into individual shots, extract still *keyframes* summarizing the content of each shot, and search for video clips containing specified types of movement.

Three commercial CBIR systems are now available – IBM's QBIC, Virage's VIR Image Engine, and Excalibur's Image RetrievalWare. In addition, demonstration versions of numerous experimental systems can be viewed on the Web, including MIT's Photobook, Columbia University's WebSEEk, and Carnegie-Mellon University's Informedia. CBIR systems are beginning to find a foothold in the marketplace; prime application areas include crime prevention (fingerprint and face recognition), intellectual property (trademark registration), journalism and advertising (video asset management) and Web searching. Both the Alta Vista and Yahoo! Search engines now have CBIR facilities, courtesy of Virage and Excalibur respectively.

The effectiveness of all current CBIR systems is inherently limited by the fact that they can operate only at the primitive feature level. None of them can search effectively for, say, a photo of a dog – though some semantic queries can be handled by specifying them in terms of primitives. A beach scene, for example, can be retrieved by specifying large areas of blue at the top of the image, and yellow at the bottom. There is evidence 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.

Standards development relevant to CBIR can be grouped under three headings – image compression, query specification and metadata description. By far the most important emerging standard is MPEG-7, which will define search features of all kinds for both still image and video data.

Our conclusion is that, despite its current limitations, CBIR is a fast-developing technology with considerable potential, and one that should be exploited where appropriate. The report's specific recommendations are as follows:

To users and managers of image collections:

?? Managers of specialist collections such as fingerprints or trademark images should be encouraged to investigate possible adoption of CBIR technology in the near future.

?? Managers of video libraries should investigate the possibility of using a proprietary video asset management package.

?? Managers of general-purpose image collections should be encouraged to keep a watching brief on developments in CBIR.

To software developers or information providers with products designed to handle images, but which currently lack CBIR capabilities:

?? Firms with products or services in specialist areas such as fingerprints or trademark images should investigate the possibility of adding CBIR technology to their products in the near future.

?? Providers of general-purpose multimedia need to keep a watching brief on developments in CBIR, particularly relating to hybrid text/image feature indexing and cross-media retrieval.

To UK government agencies:

?? Funding agencies should consider declaring CBIR research a priority area, as has been done in the USA, and to a lesser extent in the European Community. Topics particularly worth supporting, in the expectation that they will lead to useful results in the long term, include new approaches to semantic image retrieval, cross-media indexing, interface design, studies of image seeking behaviour and use, and evaluation of system effectiveness.

?? Agencies concerned with technology transfer or dissemination of best practice in fields which could potentially benefit from CBIR should consider sponsoring programmes to raise awareness of CBIR technology among leading practitioners in these fields.

To JISC:

?? Further pilot studies of CBIR should be undertaken, with a view to identifying the benefits and costs of CBIR technology, and the specific types of user most likely to benefit.

?? Provided such benefits can in fact be identified, a further programme to raise awareness of the technology and its benefits should be undertaken.

?? Again, provided benefits from the use of CBIR can be demonstrated, large-scale trials of the effectiveness of different ways of delivering CBIR should be undertaken.

?? Since the USA is likely to remain the main source of research activity in the CBIR field, every effort should be made to encourage further co-operation between the UK and USA in this area. In particular, JISC should continue to participate actively in the NSF International Digital Libraries Program.

?? JISC should encourage efforts to create standard collections of images and image queries (similar to those created for the TREC text retrieval experiments) for comparing the effectiveness of different image retrieval techniques.

To all professionals involved in image data management:

?? Image professionals should keep abreast of emerging standards, particularly the new MPEG-7 standard for multimedia content description, and contribute to their future development where appropriate.

1 Introduction

Interest in the potential of digital images has increased enormously over the last few years, fuelled at least in part by the rapid growth of imaging on the World-Wide Web (referred to in this report as 'the Web'). Users in many professional fields are exploiting the opportunities offered by the ability to access and manipulate remotely-stored images in all kinds of new and exciting ways [Gudivada and Raghavan, 1995a]. However, they are also discovering that the process of locating a desired image in a large and varied collection can be a source of considerable frustration [Jain, 1995]. The problems of image retrieval are becoming widely recognized, and the search for solutions an increasingly active area for research and development. Some indication of the rate of increase can be gained from the number of journal articles appearing each year on the subject, growing from 4 in 1991 to 12 in 1994, and 45 in 1998[*].

Problems with traditional methods of image indexing [Enser, 1995] have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as colour, texture and shape – a technology now generally referred to as *Content-Based Image Retrieval* (CBIR). After a decade of intensive research, CBIR technology is now beginning to move out of the laboratory and into the marketplace, in the form of commercial products like QBIC [Flickner et al, 1995] and Virage [Gupta et al, 1996]. However, the technology still lacks maturity, and is not yet being used on a significant scale. In the absence of hard evidence on the effectiveness of CBIR techniques in practice, opinion is still sharply divided about their usefulness in handling real-life queries in large and diverse image collections. Nor is it yet obvious how and where CBIR techniques can most profitably be used [Sutcliffe et al, 1997].

In such a situation, it is difficult for managers and users of image collections to make informed decisions about the value of CBIR techniques to their own work. With this in mind, the JISC Technology Applications Programme has commissioned the Institute for Image Data Research, University of Northumbria, and the Manchester Visualization Centre, University of Manchester, to conduct a detailed investigation into the potential of CBIR technology. The project consists of several parts:

- a report on the current state of the art in CBIR within the UK (this document);
- an evaluation of existing CBIR software packages;
- an investigation of user experience of CBIR technology through a series of pilot applications;
- a programme of publications, workshops and seminars to raise awareness of CBIR in the UK higher education community.

The aim of this report is to clarify some of the issues raised by this new technology, by reviewing its current capabilities and limitations, and its potential usefulness to users in higher education and elsewhere. The report is based both on a review of the research and professional literature, and on discussions with users and managers of large collections of image data, multimedia authors, researchers, software developers, and representatives of standards bodies. A survey was also carried out on a representative sample of UK art libraries and museums listed in the *ARLIS*

[*] *Source: BIDS Science Citation Index database, title search using keywords "image*" and "retriev*".*

Directory 1997/8, to investigate a range of issues surrounding the management of image collections. The views expressed here are purely those of the authors.

2 Background

2.1 The growth of digital imaging

The use of images in human communication is hardly new – our cave-dwelling ancestors painted pictures on the walls of their caves, and the use of maps and building plans to convey information almost certainly dates back to pre-Roman times. But the twentieth century has witnessed unparalleled growth in the number, availability and importance of images in all walks of life. Images now play a crucial role in fields as diverse as medicine, journalism, advertising, design, education and entertainment.

Technology, in the form of inventions such as photography and television, has played a major role in facilitating the capture and communication of image data. But the real engine of the imaging revolution has been the computer, bringing with it a range of techniques for digital image capture, processing, storage and transmission which would surely have startled even pioneers like John Logie Baird. The involvement of computers in imaging can be dated back to 1965, with Ivan Sutherland's *Sketchpad* project, which demonstrated the feasibility of computerised creation, manipulation and storage of images, though the high cost of hardware limited their use until the mid-1980s. Once computerised imaging became affordable (thanks largely to the development of a mass market for computer games), it soon penetrated into areas traditionally depending heavily on images for communication, such as engineering, architecture and medicine. Photograph libraries, art galleries and museums, too, began to see the advantages of making their collections available in electronic form. The creation of the World-Wide Web in the early 1990s, enabling users to access data in a variety of media from anywhere on the planet, has provided a further massive stimulus to the exploitation of digital images. The number of images available on the Web was recently estimated to be between 10 and 30 million [Sclaroff et al, 1997] – a figure which some observers consider to be a significant underestimate.

2.2 The need for image data management

The process of digitisation does not in itself make image collections easier to manage. Some form of cataloguing and indexing is still necessary – the only difference being that much of the required information can now *potentially* be derived automatically from the images themselves. The extent to which this potential is currently being realized is discussed below.

The need for efficient storage and retrieval of images – recognized by managers of large image collections such as picture libraries and design archives for many years – was reinforced by a workshop sponsored by the USA's National Science Foundation in 1992 [Jain, 1993]. After examining the issues involved in managing visual information in some depth, the participants concluded that images were indeed likely to play an increasingly important role in electronically-mediated communication. However, significant research advances, involving collaboration between a number of disciplines, would be needed before image providers could take full advantage of the opportunities offered. They identified a number of critical areas where research was needed, including data representation, feature extractions and indexing, image query matching and user interfacing.

One of the main problems they highlighted was the difficulty of locating a desired image in a large and varied collection. While it is perfectly feasible to identify a desired image from a small collection simply by browsing, more effective techniques are needed with collections containing thousands of items. Journalists requesting photographs of a particular type of event, designers looking for materials with a particular colour or texture, and engineers looking for drawings of a particular type of part, all need some form of access by image content. The existence – and continuing use – of detailed classification schemes such as ICONCLASS [Gordon, 1990] for art images, and the Opitz code [Opitz et al, 1969] for machined parts, reinforces this message.

2.3 Characteristics of image queries

What kinds of query are users likely to put to an image database? To answer this question in depth requires a detailed knowledge of user needs – why users seek images, what use they make of them, and how they judge the utility of the images they retrieve. As we show in section 3.3 below, not enough research has yet been reported to answer these questions with any certainty. Common sense evidence suggests that still images are required for a variety of reasons, including:

?? illustration of text articles, conveying information or emotions difficult to describe in words,

?? display of detailed data (such as radiology images) for analysis,

?? formal recording of design data (such as architectural plans) for later use.

Access to a desired image from a repository might thus involve a search for images depicting specific types of object or scene, evoking a particular mood, or simply containing a specific texture or pattern. Potentially, images have many types of attribute which could be used for retrieval, including:

?? the presence of a particular combination of colour, texture or shape features (e.g. green stars);

?? the presence or arrangement of specific types of object (e.g. chairs around a table);

?? the depiction of a particular type of event (e.g. a football match);

?? the presence of named individuals, locations, or events (e.g. the Queen greeting a crowd);

?? subjective emotions one might associate with the image (e.g. happiness);

?? metadata such as who created the image, where and when.

Each listed query type (with the exception of the last) represents a higher level of abstraction than its predecessor, and each is more difficult to answer without reference to some body of external knowledge. This leads naturally on to a classification of query types into three levels of increasing complexity [Eakins, 1996; Eakins, 1998]:

Level 1 comprises retrieval by *primitive* features such as colour, texture, shape or the spatial location of image elements. Examples of such queries might include "find pictures with long thin dark objects in the top left-hand corner", "find images containing yellow stars arranged in a ring" – or most commonly "find me more pictures that look like this". This level of retrieval uses features (such as a given shade of yellow) which are both objective, and directly derivable from the images themselves, without the need to refer to any external knowledge base. Its use is largely limited to specialist applications such as trademark registration, identification of drawings in a design archive, or colour matching of fashion accessories.

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:

a) retrieval of objects of a given type (e.g. "find pictures of a double-decker bus");

b) 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(b). 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(b), are usually still reasonably objective. This level of query is more generally encountered than level 1 – for example, most queries received by newspaper picture libraries appear to fall into this overall category [Enser, 1995].

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:

a) retrieval of named events or types of activity (e.g. "find pictures of Scottish folk dancing");

b) 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.

As we shall see later, this classification of query types can be useful in illustrating the strengths and limitations of different image retrieval techniques. The most significant gap at present lies between levels 1 and 2. Many authors [e.g. Gudivada and Raghavan, 1995a] refer to levels 2 and 3 together as *semantic* image retrieval, and hence the gap between levels 1 and 2 as the *semantic gap*. Note that this classification ignores a further type of image query – retrieval by associated metadata such as who created the image, where and when. This is not because such retrieval is unimportant. It is because (at least at present) such metadata is exclusively textual, and its management is primarily a text retrieval issue.

2.4 Video queries

Video sequences are an increasingly important form of image data for many users, and pose their own special challenge to those responsible for their storage and retrieval, both because of their additional complexity and their sheer volume. Video images contain a wider range of primitive data types (the most obvious being motion vectors), occupy far more storage, and can take hours to review, while the comparable process for still images takes seconds at most. Hence the process of organizing videos for retrieval is in some ways akin to that of abstracting and indexing long text documents. All but the shortest videos are made up of a number of distinct *scenes*, each of which can be further broken down into individual *shots* depicting a single view, conversation or action. A common way of organizing a video for retrieval is to prepare a *storyboard* of annotated still images (often known as *keyframes*) representing each scene. Another is to prepare a series of short video clips, each capturing the essential details of a single sequence – a process sometimes described as *video skimming*. For a detailed discussion of the issues involved in video data management, and a review of current and emerging techniques, see the reviews by Aigrain et al [1996] and Bolle et al [1998].

Video queries can be categorized by level in just the same way as still images. In this case, level 1 queries obviously include object motion as well as colour, texture or shape (e.g. "find clips of objects moving from top right of the frame to bottom left", which might retrieve videos of downhill skiers). Level 2 queries also include named types of action ("an aeroplane taking off"), though level 3 queries differ little between video and still images. Videos also differ from still images in another respect: few of them consist purely of image data. A typical video will have a soundtrack containing music, speech and other sounds, text appearing in the video sequence (such as a film's opening credits), and possibly *closed-caption text* used to provide subtitles for the hard of hearing. All of these can provide additional cues for retrieval.

2.5 What is CBIR?

The earliest use of the term *content-based image retrieval* in the literature seems to have been by Kato [1992], to describe his 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 (words stored as ASCII character strings) has already been logically structured by the author [Santini and Jain, 1997]. There is no equivalent of level 1 retrieval in a text database.

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

Key research issues in video retrieval include:

?? automatic shot and scene detection

?? ways of combining video, text and sound for retrieval

?? effective presentation of search output for the user.

2.6 Conclusions from previous reviews

Several reviews of the literature on image retrieval have been published, from a variety of different viewpoints. Enser [1995] reviews methods for providing subject access to pictorial data, developing a four-category framework to classify different approaches. He discusses the strengths and limitations both of conventional methods based on linguistic cues for both indexing and search, and experimental systems using visual cues for one or both of these. His conclusions are that, while there are serious limitations in current text-based techniques for subject access to image data, significant research advances will be needed before visually-based methods are adequate for this task. He also notes, as does Cawkell [1993] in an earlier study, that more dialogue between researchers into image analysis and information retrieval is needed.

Aigrain et al [1996] discuss the main principles of automatic image similarity matching for database retrieval, emphasizing the difficulty of expressing this in terms of automatically generated features. They review a selection of current techniques for both still image retrieval and video data management, including video parsing, shot detection, keyframe extraction and video skimming. They conclude that the field is expanding rapidly, but that many major research challenges remain, including the difficulty of expressing semantic information in terms of primitive image features, and the need for significantly improved user interfaces. CBIR techniques are likely to be of most use in restricted subject domains, and where synergies with other types of data (particularly text and speech) can be exploited.

Eakins [1996] proposes a framework for image retrieval (outlined in section 2.3 above), classifying image queries into a series of levels, and discussing the extent to which advances in technology are likely to meet users' needs at each level. His conclusion is that automatic CBIR techniques can already address many of users' requirements at level 1, and will be capable of making a significant contribution at level 2 if current research ideas can be successfully exploited. They are however most unlikely to make any impact at level 3 in the foreseeable future.

Idris and Panchanathan [1997a] provide an in-depth review of CBIR technology, explaining the principles behind techniques for colour, texture, shape and spatial indexing and retrieval in some detail. They also discuss the issues involved in video segmentation, motion detection and retrieval techniques for compressed images. They identify a number of key unanswered research questions, including the development of more robust and compact image content features, more accurate modelling of human perceptions of image similarity, the identification of more efficient physical storage and indexing techniques, and the development of methods of recognizing objects within images. De Marsicoi et al [1997] also review current CBIR technology, providing a useful feature-by-feature comparison of 20 experimental and commercial systems.

In addition to these reviews of the literature, a survey of "non-text information retrieval" was carried out in 1995 on behalf of the European Commission by staff from GMD (Gesellschaft für Mathematik und Datenverarbeitung), Darmstadt and Université Joseph Fourier de Grenoble [Berrut et al, 1995]. This reviewed current indexing practice in a number of European image, video and sound archives, surveyed the current research literature, and assessed the likely future impact of recent research and development on electronic publishing. The survey found that all current operational image archives used text-based indexing methods, which were perceived to have a number of shortcomings. In particular, indexing vocabularies were not felt to be adequate for non-text material. Despite this, users seemed generally satisfied with existing systems. The report concluded that standard information retrieval techniques were appropriate for managing collections of non-text data, though the adoption of intelligent text retrieval techniques such as the inference-based methods developed in the INQUERY project [Turtle and Croft, 1991] could be beneficial.

3 Image users

3.1 Image use in the community

It is a truism to observe that images are currently used in all walks of life. The influence of television and video games in today's society is clear for all to see. The commonest single reason for storing, transmitting and displaying images is probably for recreational use, though this category includes a wide variety of different attitudes and interaction styles, from passively watching the latest episode of a soap opera to actively analysing a tennis star's shots in the hope of improving one's own game. Images are increasingly used to convey information, in areas as diverse as map-making, weather forecasting and mail-order shopping, and to persuade or convey a mood, as in advertising. They can also be appreciated in their own right, as works of art.

A detailed sociological study of image use would be out of place in this report, particularly as there is currently little evidence for the existence of different user communities with different needs. Most individuals interact with images in different ways at different times, perhaps spending an hour in an art gallery one day, and watching a sports video the next. Trying to categorize such behaviour by user type does not seem very useful.

3.2 Professional groups making use of images

In the realm of professional image use, the situation is rather different. While there are certainly differences in style between individual design engineers, for example, the nature of the design process imposes a number of inescapable constraints within which all engineers must work. Hence it is possible to generalize to some extent about the way images are used by different professions. Since this report is primarily concerned with image storage and retrieval, it makes sense to limit our discussion by concentrating on uses which involve stored collections of images in some way.

Some groups of people use images in their job on a daily basis, such as graphic designers and illustrators, whilst others may never be required to use them, such as bank managers and accountants. There is a wide range of professions lying between these two extremes, including medicine and law. Other groups of workers, such as librarians and museum curators, may be required to find images on behalf of clients rather than for themselves. It is impossible to give a full picture here of the uses being made of visual information. The following examples should be interpreted as being merely a snapshot of the situation:

Crime prevention. The police use visual information to identify people or to record the scenes of crime for evidence; over the course of time, these photographic records become a valuable archive. In the UK, it is common practice to photograph everyone who is arrested and to take their fingerprints. The photograph will be filed with the main record for the person concerned, which in a manual system is a paper-based file. In a computer-based system, the photograph will be digitised and linked to the corresponding textual records. Until convicted, access to photographic information is restricted and, if the accused is acquitted, all photographs and fingerprints are deleted. If convicted, the fingerprints are passed to the National Fingerprint Bureau. Currently, there is a national initiative investigating a networked Automated Fingerprint Recognition system involving BT and over thirty regional police forces. Other uses of images in law enforcement include face recognition [e.g. Wu and Narasimhalu, 1994], DNA matching, shoe sole impressions [e.g. Ashley, 1996], and surveillance systems. The Metropolitan Police Force in London is involved with a project which is setting up an international database of the images of stolen objects (http://www.arttic.com/grasp/).

Medicine. The medical and related health professions use and store visual information in the form of X-rays, ultrasound or other scanned images, for diagnosis and monitoring purposes. There are strict rules on confidentiality of such information. The images are kept with the patients' health records which are, in the main, manual files, stored by unique identifier (NI number). Visual information, provided it is rendered anonymous, may be used for research and teaching purposes. Much of the research effort related to images is undertaken in the medical physics area. Aspects of concern include effective image processing (e.g. boundary/feature detection) systems which aid the practitioner in detecting and diagnosing lesions and tumours and tracking progress/growth.

Fashion and graphic design. Imagery is very important for graphic, fashion and industrial designers. Visualisation seems to be part of the creative process. Whilst there will be individual differences in the way designers approach their task, many use images of previous designs in the form of pictures, photographs and graphics, as well as objects and other visual information from the real world, to provide inspiration and to visualise the end product. 2-D sketches, and, increasingly, 3-D geometric models are used to present ideas to clients and other colleagues. There is also a need to represent the way garments hang and flow.

Publishing and advertising. Photographs and pictures are used extensively in the publishing industry, to illustrate books and articles in newspapers and magazines. Many national and regional newspaper publishers maintain their own libraries of photographs, or will use those available from the Press Association, Reuters and other agencies. The photographic collections will be indexed and filed under, usually, broad subject headings (e.g. local scenes, buildings or personalities as well as pictures covering national and international themes). Increasingly, electronic methods of storage and access are appearing, alongside developments in automated methods of newspaper production, greatly improving the speed and accuracy of the retrieval process. Advertisements and advertising campaigns rely heavily on still and moving imagery to promote the products or services. The growth of commercial stock photograph libraries, such as Getty Images and Corbis, reflects the lucrative nature of the industry.

Architectural and engineering design. Photographs are used in architecture to record finished projects, including interior and exterior shots of buildings as well particular features of the design. Traditionally these photographs will be stored as hardcopy or in slide format, accessible by, say, project number and perhaps name, and used for reference by the architects in making presentations

to clients and for teaching purposes. Larger architects' practices with more ample resources, have introduced digital cameras and the electronic storage of photographs.

The images used in most branches of engineering include drawings, plans, machine parts, and so on. Computer Aided Design (CAD) is used extensively in the design process. A prime need in many applications is the need to make effective use of standard parts, in order to maintain competitive pricing [Bradley et al, 1994]. Hence many engineering firms maintain extensive design archives. CAD and 2-D modelling are also extensively used in architectural design, with 3-D modelling and other visualization techniques increasingly being used for communicating with clients (see, for example, [Ross, 1998] and [Evans, 1996]). A recent survey of IT in architectural firms [Fallon, 1998] emphasized the dominance of CAD (especially 2-D) in the design process, though it concluded that object-based, intelligent 3-D modelling systems will become more important in the future.

Historical research. Historians from a variety of disciplines – art, sociology, medicine, etc. – use visual information sources to support their research activities. Archaeologists also rely heavily on images. In some instances (particularly, but not exclusively, art), the visual record may be the only evidence available. Where access to the original works of art is restricted or impossible, perhaps due to their geographic distance, ownership restrictions or factors to do with their physical condition, researchers have to use surrogates in the form of photographs, slides or other pictures of the objects, which may be collected within a particular library, museum or art gallery. Photographic and slide collections are maintained by a wide range of organisations, including academic and public libraries.

3.3 User needs for image data – research and survey findings

The above discussion should confirm that images are used extensively in many aspects of professional life, as well as emphasizing that they can be used in a variety of different ways. However, little systematic research on image use has yet been published. Most reported research has focused either on specific collections or on specific user types or populations. Examples include: newspaper image archives [e.g. Ornager, 1997; Markkula and Sormunen, 1998]; picture archives [Enser and McGregor, 1992; Enser, 1995]; other historical libraries and archives [e.g. Keister, 1994; Armitage and Enser, 1997]; and museums [McCorry and Morrison, 1995; Sledge, 1995]. The user populations which have received the most attention by researchers in image use are those in the humanities and arts, and in particular, art historians – for example, see Markey [1988].

Ornager [1997], in her study of journalists using newspaper image archives, proposes a user typology, as follows:

- The *specific inquirer* who asks very narrow questions, because he/she has a specific photograph in mind;
- The *general inquirer* who asks very broad questions because they want to make their own choice;
- The *story teller inquirer* who tells about the story and is open to suggestions from the archive staff;
- The *story giver inquirer* who hands the story over to the staff wanting them to choose the photograph(s); and
- The *fill in space inquirer* who only cares about the size of the photograph in order to fill an empty space on the page.

She found that the journalists required the facility to broaden and narrow their queries and/or to search from different aspects (i.e. to look for related concepts), and in some instances to browse.

Markkula and Sormumen [1998] found that journalist requests fell into four categories: concrete objects (i.e. named persons, buildings or places); themes or abstractions interpretable from the photographs; background information on the image (such as documentary information, specific news events and films and television programmes); and known photographs. Requests for the first category dominated the use of the photograph archives and were mostly satisfied using mediated searches. The journalists preferred, on the whole, to undertake their own searches for photographs of themes where there was a need to interpret such aspects as symbolic value, atmosphere and feelings, in the context of a particular illustration task. Searching by themselves allowed the journalists to be more explorative and to rely more on browsing. The authors go on to recommend a 'hybrid' system for the searching of photographs - that is, one which supports traditional concept-based indexing and classification methods and a user interface which permits the browsing of retrieved theme-based sub-sets of the database. Automatic content-based image retrieval methods could then be applied to the browsing stage.

Enser and McGregor [1993] categorised queries put to a large picture archive into those which could be satisfied by a picture of a unique person, object or event (e.g. Kenilworth Castle, Sergei Prokofiev,) and those which could not (e.g. smog, classroom scenes). Uniquely identified subjects dominated the requests. Both categories – 'unique' and 'non-unique' – were subject to refinement in terms of time (e.g. a young D H Lawrence), location (e.g. Eiffel Tower at night, any period), action (e.g. couples dancing the Charleston), event (e.g. Vesuvius erupting) or technical specification (e.g. 1950s fridge). This work was further extended in a study of seven picture libraries which sought to develop a general-purpose categorisation of user requests for still and moving visual images [Armitage and Enser, 1997]. An initial analysis led to the identification of four major types of query:

- image content ('find me some images of …');
- identification/attribution/provenance checking;
- accessibility of image/artist of work (e.g. ownership/viewing availability);
- miscellaneous (e.g. requests for administrative procedures only, or unusable queries).

The authors analysed the image content queries (some 1749 queries across the seven participating libraries) in more depth, categorising them into 'by named artist'; 'known items'; 'unique subject' and 'non-unique subject'.

Hastings [1995] investigated how art historians searched photographic and digital art images, using qualitative techniques which involved observation, interviews and image analysis. In order of frequency, the major classes of queries were found to be: identification, subject, text, style, artist, category, compare and colour. Access points and computer manipulation functions within the database varied with the level of complexity of the queries, ranging from the least complex type such as queries for 'who', 'where' and 'when'; to the most complex queries for 'meaning', 'subject' and 'why'. The highest-level queries could often not be answered from the images or from available textual information and required access to full-text secondary subject resources.

Keister [1994] describes the development of an automated still picture retrieval system at the National Library of Medicine (NLM). She undertook a review of queries over a period which showed that users did not ask for pictures in a consistent, traditional manner. She found that users who were picture professionals (still picture researchers, TV, film, or media personnel) thought visually and used art and/or graphics jargon. Health professionals asked for images in a manner more

in keeping with the Library's orientation (e.g. do you have pictures of cholera?). The museum or academic community often had precise citations to the images it desired. She suggested that words describing concrete image elements comprised a significant proportion of requests and that, since these elements are easily identified and can be agreed upon by different cataloguers, close attention was needed to these elements at the cataloguing stage.

The Catechism Project [McCorry and Morrison, 1995] looked at how information is sought and used in museums. Whilst not specifically focussed on images, the project does have relevance when considering visual information in its broadest definition. The authors analysed 1013 representative user queries from over 100 museums (self selected by the museum personnel). Over two thirds of the questions asked were ones which might be answered by using a typical museum database or other catalogue. The rest were made up of questions which did not refer directly to the objects in the museum's collection and which could be dealt with using a variety of methods, including biographical dictionaries and reference works, and personal knowledge. The largest number of enquiries (about 30%) related to a specific object. Next in order of frequency (around 18%) were questions regarding the names of makers, owners or institutions, followed by place associated with the object, such as place of origin or manufacture (13%). Questions relating to physical description were found not to be a major area of inquiry, although they acknowledge that for some collections it may be vital. The Getty Art History Information Program, which became the Getty Information Institute, used the results of the Catechism Project to inform their *Points of view* project – a series of meetings which aimed to amplify access to cultural heritage information by gaining an understanding of users' needs and expectations [Sledge, 1995].

3.4 How much do we really know about user needs?

Most of the above writers attempt to categorise the uses being made of particular collections by analysing the queries put to the collections, either in the form of written statements by the end users or interpretations put on verbal enquiries by intermediaries (i.e. librarians, archivists, curators). This seeming emphasis on the *expressed need*, as interpreted by the formulation of the query by the end user or the intermediary, still tells us little about what the *actual need* is for the images and, indeed, *what use* will be made of retrieved images. Users' expressed needs are likely to be heavily biased by their expectations of the kinds of query the system can actually handle.

The studies above – backed up by our own admittedly limited investigations – confirm that some image users have very specific needs (e.g. an art historian may want a particular painting by Van Gogh; a journalist may want a recent photograph of Tony Blair, not smiling; a social historian may want a picture of sewers in the 18th century; a theology student may want a picture of a female saint with red hair) whilst others will be more interested in material which conveys particular sensations or moods (e.g. happiness, concentration). Fashion designers and illustrators often do not want a specific image; instead, they want to browse through a number of images seeking inspiration for a current project.

Some attempts have been made (e.g. Keister [1994], Ornager [1997]) to develop a more generic framework for understanding image search and use, though it is not clear how widely applicable these are. Too little is still known about how users can usefully be segmented into different types, the needs of these types, and the implications for retrieval systems design. More research is needed in the area of use and user studies, not just to understand more fully how and why people use images, but also to design more effective retrieval systems for image users. If it can be reliably established that different types of users do in fact require different styles of interaction with retrieval systems, the

task of the systems designer will be made considerably easier. Current research in the Institute for Image Data Research at the University of Northumbria aims to address some of these issues (see http://www.unn.ac.uk/iidr/visor.html for details).

4 Current techniques for image and video retrieval

4.1 Organizing an image collection

Whilst this review is primarily focused on techniques for the storage and retrieval of electronic images, it is useful to reflect on the traditional practices of picture and other manual collections of images and videos. Image collections of various types are maintained by a wide range of organisations, of all sizes and in a variety of sectors.

Traditionally, images will be stored in their original analogue form, in wallets, files or folders, which in turn will be arranged on shelves, in drawers or in cabinets. The level of indexing associated with manual image collections will be closely related to the importance of the collection, the way it is used, and the time and resources allocated to the task. Retrieval of particular images from such collections is inherently labour intensive and often serendipitous. Knowledge of the collection usually rests with the librarians, archivists, curators or others responsible for its upkeep and, less often, the actual users. When manual collections are digitised, decisions have to be made about the associated metadata and often it may not be feasible, due to lack of resources, to upgrade the content of the *catalogue* or textual record associated with each image.

The need for efficient storage and retrieval of images has been recognised by managers of large image collections such as picture libraries and design archives for many years. While it is perfectly feasible to identify a desired image from a small collection simply by browsing, more effective techniques are needed with collections containing thousands of items. The normal technique used is to assign descriptive metadata in the form of keywords, subject headings or classification codes to each image when it is first added to the collection, and to use these descriptors as retrieval keys at search time.

4.2 Classification and indexing schemes

Many picture libraries use keywords as their main form of retrieval – often using indexing schemes developed in-house, which reflect the special nature of their collections. A good example of this is the system developed by Getty Images to index their collection of contemporary stock photographs [Bjarnestam, 1998]. Their thesaurus comprises just over 10 000 keywords, divided into nine semantic groups, including *geography, people, activities* and *concepts.* Index terms are assigned to the whole image, the main objects depicted, and their setting. Retrieval software has been developed to allow users to submit and refine queries at a range of levels, from the broad (e.g. “freedom”) to the specific (e.g. “a child pushing a swing”).

Probably the best-known indexing scheme in the public domain is the Art and Architecture Thesaurus (AAT), originating at Rensselaer Polytechnic Institute in the early 1980s, and now used in art libraries across the world. AAT is maintained by the Getty Information Institute and consists of nearly 120,000 terms for describing objects, textural materials, images, architecture and other cultural heritage material. There are seven facets or categories which are further subdivided into 33 subfacets or hierarchies. The facets, which progress from the abstract to the concrete, are: associated concepts, physical attributes, styles and periods, agents, activities, materials, and objects. AAT is available on the Web from the Getty Information Institute at

http://www.ahip.getty.edu/aat_browser/. Other tools from Getty include the *Union List of Artist Names (ULAN)* and the *Getty Thesaurus of Geographic Names (TGN).* Another popular source for providing subject access to visual material is the *Library of Congress Thesaurus for Graphic Materials (LCTGM).* Derived from the *Library of Congress Subject Headings (LCSH),* LCTGM is designed to assist with the indexing of historical image collections in the automated environment. Greenberg [1993] provides a useful comparison between AAT and LCTGM.

A number of indexing schemes use classification codes rather than keywords or subject descriptors to describe image content, as these can give a greater degree of language independence and show concept hierarchies more clearly. Examples of this genre include ICONCLASS (http://iconclass.let.uu.nl/) from the University of Leiden [Gordon, 1990], and TELCLASS from the BBC [Evans, 1987]. Like AAT, ICONCLASS was designed for the classification of works of art, and to some extent duplicates its function; an example of its use is described by Franklin [1998]. TELCLASS was designed with TV and video programmes in mind, and is hence rather more general in its outlook. The Social History and Industrial Classification, maintained by the Museum Documentation Association, is a subject classification for museum cataloguing (http://www.holm.demon.co.uk/shic/). It is designed to make links between a wide variety of material including objects, photographs, archival material, tape recordings and information files.

A number of less widely-known schemes have been devised to classify images and drawings for specialist purposes. Examples include the Vienna classification for trademark images [World Intellectual Property Organization, 1998], used by registries worldwide to identify potentially conflicting trademark applications, and the Opitz coding system for machined parts [Opitz et al, 1969], used to identify families of similar parts which can be manufactured together.

A survey of art librarians conducted for this report (see section 4.3 for details) suggests that, despite the existence of specialist classification schemes for images, general classification schemes, such as *Dewey Decimal Classification (DDC), Library of Congress (LC), BLISS* and the *Universal Decimal Classification (UDC),* are still widely used in photographic, slide and video libraries. The former scheme is the most popular, which is not surprising when one considers the dominance of DDC in UK public and academic library sectors. ICONCLASS, AAT, LCTGM, SHIC are all in use in at least one or more of the institutions in the survey. However, many libraries and archives use in-house schemes for the description of the subject content. For example, nearly a third of all respondents have their own in-house scheme for indexing slides.

4.3 Current indexing practice

When discussing the indexing of images and videos, one needs to distinguish between systems which are geared to the formal description of the image and those concerned with subject indexing and retrieval. The former is comparable to the bibliographical description of a book. However, there is still no one standard in use for image description, although much effort is being expended in this area by a range of organisations such as the Museum Documentation Association (http://www.mda.co.uk), the Getty Information Institute (http://www.gii.getty.edu/), the Visual Resources Association (http://www.vra.oberlin.edu/) the International Federation of Library Association/Art Libraries (http://ifla.inist.fr/) and the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) (http://www.cidoc.icom.org/).

The descriptive cataloguing of photographs presents a number of special challenges. Photographs, for example, are not self-identifying. Unlike textual works that provide such essential cataloguing aids

as title pages, abstracts and table of contents, photographs often contain no indication of author or photographer, names of persons or places depicted, dates, or any textual information whatever. Cataloguing of images is more complex than that for text documents, since records should contain information about the standards used for image capture and how the data is stored as well as descriptive information, such as title, photographer (or painter, artist, etc). In addition, copies of certain types of images may involve many layers of intellectual property rights, pertaining to the original work, its copy (e.g. a photograph), a digital image scanned from the photograph, and any subsequent digital image derived from that image.

Published reviews of traditional indexing practices for images and video include Rasmussen [1997], Lancaster [1998], and Cawkell [1993]. Many writers, including Svenonius [1994], Krause [1988], and Shatford [1986], discuss the difficulties of indexing images. Besser [1990] discusses the problems of managing a large image collection. He notes that, unlike books, images make no attempt to tell us what they are about and that often they may be used for purposes not anticipated by their originators. Images are rich in information and can be used by researchers from a broad range of disciplines. As Besser comments:

> "A set of photographs of a busy street scene a century ago might be useful to historians wanting a 'snapshot' of the times, to architects looking at buildings, to urban planners looking at traffic patterns or building shadows, to cultural historians looking at changes in fashion, to medical researchers looking at female smoking habits, to sociologists looking at class distinctions, or to students looking at the use of certain photographic processes or techniques."

Svenonius [1994] discusses the question of whether it is possible to use words to express the "aboutness of a work in a wordless medium, like art…". To get around the problem of the needs of different users groups, van der Starre [1995] advocates that indexers should "stick to 'plain and simple' indexing, using index terms accepted by the users, and using preferably a thesaurus with many lead-ins," thus placing the burden of further selection on the user. Shatford Layne (1994) suggests that, when indexing images, it may be necessary to determine which attributes provide useful groupings of images; which attributes provide information that is useful once the images are found; and which attributes may, or even should, be left to the searcher or researcher to identify. She also advocates further research into the ways images are sought and the reasons that they are useful in order to improve the indexing process. Constantopulos and Doerr (1995) also support a user centred approach to the designing of effective image retrieval systems. They urge that attention needs to be paid to the intentions and goals of the users, since this will help define the desirable descriptive structures and retrieval mechanisms as well as understanding what is 'out of the scope' of an indexing system.

When it comes to describing the content of images, respondents in our own survey seem to include a wide range of descriptors including title, period, genre, subject headings, keywords, classification and captions (although there was some variation by format). Virtually all maintain some description of the subject content of their images. The majority of our respondents maintain manual collections of images, so it is not surprising that they also maintain manual indexes. Some 11% of respondents included their photographs and slides in the online catalogues, whilst more than half added their videos to their online catalogues. Standard text retrieval or database management systems were in use in a number of libraries (with textual descriptions only for their images). Three respondents used specific image management systems: Index+, iBase and a bespoke in-house system. Unsurprisingly, none currently use CBIR software.

4.4 Software for image data management

The last decade has seen the appearance of a number of commercial image data management systems. These systems normally store representations of pictorial documents (such as photographs, prints, paintings, drawings, illustrations, slides, video clips, and so on) in static archival databases, and incorporate multimedia database management systems in the storage of, and provision of wider access to, these repositories [Furner, 1998]. It should however be noted that none of these systems provide CBIR facilities – all rely on text keywords which have to be added by human indexers to provide retrieval of stored images. Some of these systems are described below.

iBase. iBase Image Systems (http://www.ibase.com/) was set up in 1992. The iBase package is their standard product and claims to be one of the first multi-platform systems that integrates image capture, manipulation, storage, retrieval and display in one application. Unlike conventional text based databases, the iBase Image Database has been designed from its inception to deal with objects and not just text. The interface allows searching by keyword, data match and global text search. The package has been enormously successful and since its formation, iBase has trebled its turnover annually. Their client list includes The British Library, Mercury One2One, The Wellcome Trust, West Yorkshire Police, The Guildhall Library London and Christies. Beasley [1998] describes the development of the pilot image database at the Natural History Museum.

Index+. System Simulation's Index+ is a powerful software toolkit for creating systems to manage structured and unstructured text, data, still and moving images, sound and video. It features fast searching, high storage capacity, a robust, network orientated, server-client architecture and a range of application development tools. Index+ is designed for Unix, MS-Windows, NT, Power Macintosh and mixed environments. For information retrieval, Index+ is designed for both text and numeric searches. Search commands are specified using simple English-like terms or by filling in search screen forms. Proximity, partial matching and use of wildcard characters anywhere in the search specification are catered for. Index+ also supports multilingual text (including the Chinese character set BIG5) and a thesaurus facility for controlling vocabulary at both data input and search stages. Index+ forms the basis of MUSIMS – Museum Information Management System – which offers an integrated approach to collections management and access for museums, galleries, image libraries and archives. The client base includes The Victoria and Albert Museum, The British Museum, The National Gallery, Getty Images, SCRAN (Scottish Cultural Resources Access Network) and the ADAM Information Gateway. The company Web site is at http://www.ssl.co.uk/.

Digital Catalogue. Image Resource Ltd (http://www.imageres.com/) provide a range of products based on their Digital Catalogue software, to cater for the single user/single platform system, through multi-user, cross platform systems, to full Internet publishing services. The system, which uses object oriented technology, is modular and upgradeable, and can be used for storing, retrieving, browsing, viewing and searching an image collection, both online and on CD-ROM. The online version uses Java technology. Images may be searched using numbers, dates, categories, sub-categories, keywords or free text. The system is aimed at picture libraries, medical illustration departments, personnel records, police, product catalogues, museums, galleries, designers, photographers and any major image collection. The database is now being used in a site for The British Tourist Authority called TIMMS (Tourist Information MultiMedia System) [Bell, 1998]. The catalogue contains pictures of hotels, attractions and other tourist related images. The system is also being employed by the National Tramway Museum in Derbyshire (http://www.tramway.co.uk) to make available its picture archive via the Web.

Fastfoto. Picdar (http://www.picdar.co.uk/index2.html) supplies multimedia asset management and image library systems, particularly within the newspaper and pre-press market where two thirds of the UK national newspapers currently use Picdar systems. There are three main products – Fastfoto Image Library, Fastfoto Picture Desk, and Fastfoto 5000. The Fastfoto technology is utilised in A*MMOS, a powerful asset management system which was launched in 1998. A*MMOS can handle from hundreds to millions of digital objects, including images, documents, sound and movie recordings and raw text. Picdar systems are used by large and small businesses in a broad range of industries including regional newspapers, other publishers, broadcasters, reprographic houses, facilities management businesses, picture and news agencies, manufacturers, retailers, charities and police forces. Clients include the Manchester Evening News, the Financial Times Picture Desk and Image Library, Irish Independent Newspapers, Catalyst Creative Imaging, and the Royal Society for the Protection of Birds. A demonstration of the Fastfoto system is available at http://www.g8pix.com/.

FotoWare. FotoWare AS (http://www.fotoware.com/) is based in Norway; its UK agent is Metasys (http://metasys.co.uk/). FotoWare is the name of both the company and a family of software products, including FotoStation Pro (a digital light-box), FotoWeb, for those who want a Web interface, and Index Manager, a search engine providing a full-text search facility for use by FotoStation and FotoWeb. These are aimed at making image archiving, administration and processing easy and powerful, on standard platforms. All applications can be used stand alone, they can also work together forming powerful imaging systems. They have a worldwide client base. Their UK clients include the Daily Express, Daily Star, Sunday Express, Meridian Television, Sky TV, Leicester University, Automobile Association, North Yorkshire Police, and IPC magazines.

Signpost. Signpost (http://www.signpost.com/) has developed a suite of PC and Mac software that allows users to browse for multimedia library files for professional use. Signpost operate the service which is accessible over modems, ISDN and the Internet, for their customers, including British Steel Multimedia Library (http://www.britishsteel.co.uk/) and the Pickhall Picture Library

Cumulus. Canto Software was founded in 1990 as a private software development company with offices in San Francisco and Berlin (http://www.canto-software.com/Pages/framecall.html). Their primary product line is Cumulus Media Management System. This is a powerful system which organises graphic and media files into customisable, hierarchical categories for instant access. The system allows the searching of any field and permits browsing and previewing media files in thumbnail and full resolutions. The company claims to have 85,000 users worldwide; these include advertising agencies and design firms; pre-press and print production; newspaper publishing; book, magazine and catalogue publishers; web and multimedia publishing; technical design and illustration; health/ medical imaging; and architecture and construction.

4.5 Research into indexing effectiveness

Current image indexing techniques have many strengths. Keyword indexing has high expressive power – it can be used to describe almost any aspect of image content. It is in principle easily extensible to accommodate new concepts, and can be used to describe image content at varying degrees of complexity. There is a wide range of available text retrieval software to automate the actual process of searching. But the process of manual indexing, whether by keywords or classification codes, suffers from two significant drawbacks. Firstly, it is inherently very labour-intensive. Indexing times quoted in the literature range from about 7 minutes per image for stock photographs at Getty Images, using their in-house system, to more than 40 minutes per image for a

slide collection at Rensselaer Polytechnic, using AAT. Manual indexing times for video are likely to be even longer.

Secondly, manual indexing does not appear to be particularly reliable as a means of subject retrieval of images. Markey [1984] found that, in a review of inter-indexer consistency, there were wide disparities in the keywords that different individuals assigned to the same picture. Similar results emerge from studies of the usefulness of assigned keywords in answering user queries in picture libraries. Enser and McGregor [1993] found a poor match between user queries and one of the indexing languages in place in the Hulton Deutsch Collection, even though it had been designed for the collection; the scheme served merely as a pointer to regions of the collection, which required further browsing to identify relevant images. Their rather depressing conclusion is that "if the retrieval utility of an image suffers from low predictability, the subject indexing of that image must have low utility" [Enser, 1995]. Ogle and Stonebraker [1995] reinforce this view, pointing out the limitations of keyword indexing in a large collection of digitised images, including: misspellings of keywords (by both indexer and searcher); dictionaries failing to overcome effects of inaccurate descriptions; use of a thesaurus not overcoming incomplete descriptions; and older images digitised with minimal or no descriptive data. These limitations mean that retrieval of images has to rely on the knowledge and experience of staff.

Seloff [1990] describes the inherent problem of image retrieval within many visual archives as being "both the sparsity and inconsistency of textual descriptions evoked by the visual content of the image." The descriptive cataloguing of similar images can vary widely particularly if carried out at separate times. Over a period, vocabulary changes and evolves along with the discipline. The major problem however is that the task of describing image content is highly subjective. The viewpoint of the cataloguer or indexer may be different from the perspective of the user, who him or herself may be an expert in the discipline. A picture can mean different things to different people, and it will also mean different things to the same person at different times. How can an indexer satisfy all the different users and uses of images [van der Starre, 1995]? Keister's [1994] view summarises this major aspect of the indexing of images: "[it] is not so much that a picture is worth a thousand words, for many fewer words can describe a still picture for most retrieval purposes, the issue has more to do with the fact that those words vary from one person to another…"

Our own survey of art librarians tends to confirm this view. Over half our respondents were dissatisfied with the indexing of their image collections. Whilst acknowledging resource constraints on staffing, the majority of comments centred on the indexing process itself – the difficulties in determining the subject, consistency between staff, level of indexing, and the need for more, in-depth indexing. User requirements – matching indexing terms with needs of users – were also mentioned. In nearly all cases, users were encouraged to do their own searching, although librarians, curators and archivists also assisted in the searching for images because they claimed to have a better knowledge of either the collection or the indexes/catalogues available. Users were not always able to express their information need adequately, either because they were not sure what they were looking for, or because they did not understand how to use the indexes. Inadequacies in the indexes themselves or in the physical arrangements of the collections were also issues. Some respondents remarked that users often did not know what they wanted until they saw it, and so needed additional support either in the form of captions or a browsing facility.

It is perhaps surprising that the role of the intermediary in image searching has not been more widely discussed, in view of evidence that it can affect the way in which users interact with the retrieval process in text retrieval systems [Spink et al, 1998]. With the move towards digital libraries and an

ethos of self-service in academic institutions, the role of the intermediary is increasingly being called into question. A systematic study of the value of using intermediaries for image indexing and searching seems long overdue. By contrast, the involvement of the user in the description and indexing of images has been the topic of a number of studies. For example, Dyson [1992] studied the attributes of images and derived a classification scheme from users' perceptions of their features within graphic, semantic and bibliographic dimensions and found that general impressions were particularly important when describing a symbol. A follow up study demonstrated that the participation of users in the development of the symbol classification was beneficial [Dyson and Box, 1997]. Brown et al [1996] advocate user-based *democratic indexing* in the development of an image database.

In contrast with the situation in text retrieval, where index language effectiveness has been the subject of intensive study for more than thirty years (see [Cleverdon et al, 1966] for an account of the classic Cranfield experiments), there is little hard evidence on the effectiveness of visual information retrieval systems of any kind. One reason for this is the lack of any universally-accepted methodology for evaluating multimedia system effectiveness. Doubts have been raised about the applicability of the Cranfield approach to the evaluation of real-life interactive systems [e.g. Ellis, 1996]. Much of the criticism has concerned the validity of deriving quantitative performance measures from subjective relevance judgments. Reliance on predefined requests, with little end-user involvement or interaction, has also been criticised. This has led to the rise of more user-centred approaches, emphasising the use of multiple evaluation methods which take account of the interactive nature of modern systems, the user's work and task context, and the variability of information needs over time [Harter and Hert, 1997]. One example of an evaluation technique based on these principles is that of Borlund and Ingwersen [1997], which aims to provide realistic evaluations of system effectiveness based on users' own queries. Another is the user-centred, task-oriented technique used by Jose et al [1998] to compare the acceptability of two versions of an image retrieval system, one set up to enable spatial queries only, the other allowing textual queries only.

The inescapable conclusion from the work discussed here is that there is very little firm evidence that current text-based techniques for image retrieval are adequate for their task. What evidence there is tends to suggest that the effectiveness of many existing systems is worryingly low. There is certainly little support for the somewhat complacent conclusions of the 1995 ELPUB report to the European Commission [Berrut et al, 1995]. More research into the area of indexing effectiveness would certainly be beneficial, though the problem is compounded by the lack of consensus over what constitutes a valid evaluation method. Further research is needed here too!

5 Content-based image and video retrieval

5.1 Current level 1 CBIR techniques

In contrast to the text-based approach of the systems described in section 4.4 above, CBIR operates on a totally different principle, retrieving stored images from a collection by comparing features automatically extracted from the images themselves. The commonest features used are mathematical measures of colour, texture or shape; hence virtually all current CBIR systems, whether commercial or experimental, operate at level 1. A typical system (see section 5.5 below for details of current CBIR systems) allows users to formulate queries by submitting an example of the type of image being sought, though some offer alternatives such as selection from a palette or sketch input. The system then identifies those stored images whose feature values match those of the query most

closely, and displays thumbnails of these images on the screen**Fehler! Verweisquelle konnte nicht gefunden werden.**. Some of the more commonly used types of feature used for image retrieval are described below.

5.1.1 Colour retrieval

Several methods for retrieving images on the basis of colour similarity have been described in the literature, but most are variations on the same basic idea. Each image added to the collection is analysed to compute a *colour histogram* which shows the proportion of pixels of each colour within the image. The colour histogram for each image is then stored in the database. At search time, the user can either specify the desired proportion of each colour (75% olive green and 25% red, for example), or submit an example image from which a colour histogram is calculated. Either way, the matching process then retrieves those images whose colour histograms match those of the query most closely. The matching technique most commonly used, histogram intersection, was first developed by Swain and Ballard [1991]. Variants of this technique are now used in a high proportion of current CBIR systems. Methods of improving on Swain and Ballard's original technique include the use of cumulative colour histograms [Stricker and Orengo, 1995], combining histogram intersection with some element of spatial matching [Stricker and Dimai, 1996], and the use of region-based colour querying [Carson et al, 1997]. The results from some of these systems can look quite impressive.

5.1.2 Texture retrieval

The ability to retrieve images on the basis of texture similarity may not seem very useful. But the ability to match on texture similarity can often be useful in distinguishing between areas of images with similar colour (such as sky and sea, or leaves and grass). A variety of techniques has been used for measuring texture similarity; the best-established rely on comparing values of what are known as *second-order statistics* calculated from query and stored images. Essentially, these calculate the relative brightness of selected *pairs* of pixels from each image. From these it is possible to calculate measures of image texture such as the degree of *contrast*, *coarseness*, *directionality* and *regularity* [Tamura et al, 1978], or *periodicity*, *directionality* and *randomness* [Liu and Picard, 1996]. Alternative methods of texture analysis for retrieval include the use of Gabor filters [Manjunath and Ma, 1996] and fractals [Kaplan et al, 1998]. Texture queries can be formulated in a similar manner to colour queries, by selecting examples of desired textures from a palette, or by supplying an example query image. The system then retrieves images with texture measures most similar in value to the query. A recent extension of the technique is the texture thesaurus developed by Ma and Manjunath [1998], which retrieves textured regions in images on the basis of similarity to automatically-derived codewords representing important classes of texture within the collection.

5.1.3 Shape retrieval

The ability to retrieve by shape is perhaps the most obvious requirement at the primitive level. Unlike texture, shape is a fairly well-defined concept – and there is considerable evidence that natural objects are primarily recognized by their shape [Biederman, 1987]. A number of features characteristic of object shape (but independent of size or orientation) are computed for every object identified within each stored image. Queries are then answered by computing the same set of features for the query image, and retrieving those stored images whose features most closely match those of the query. Two main types of shape feature are commonly used – *global* features such as aspect ratio, circularity and moment invariants [Niblack et al, 1993] and *local* features such as sets of

consecutive boundary segments [Mehrotra and Gary, 1995]. Alternative methods proposed for shape matching have included elastic deformation of templates (Pentland et al [1996], del Bimbo et al [1996]), comparison of directional histograms of edges extracted from the image (Jain and Vailaya [1996], Androutsas et al [1998]), and *shocks*, skeletal representations of object shape that can be compared using graph matching techniques (Kimia et al [1997], Tirthapura et al [1998]). Queries to shape retrieval systems are formulated either by identifying an example image to act as the query, or as a user-drawn sketch (Hirata and Kato [1992], Chan and Kung [1997]).

Shape matching of three-dimensional objects is a more challenging task – particularly where only a single 2-D view of the object in question is available. While no general solution to this problem is possible, some useful inroads have been made into the problem of identifying at least some instances of a given object from different viewpoints. One approach has been to build up a set of plausible 3-D models from the available 2-D image, and match them with other models in the database [Chen and Stockman, 1996]. Another is to generate a series of alternative 2-D views of each database object, each of which is matched with the query image [Dickinson et al, 1998]. Related research issues in this area include defining 3-D shape similarity measures [Shum et al, 1996], and providing a means for users to formulate 3-D shape queries [Horikoshi and Kasahara, 1990].

5.1.4 Retrieval by other types of primitive feature

One of the oldest-established means of accessing pictorial data is retrieval by its position within an image. Accessing data by spatial location is an essential aspect of geographical information systems (section 5.6.8), and efficient methods to achieve this have been around for many years (e.g. Chock et al [1984], Roussopoulos et al [1988]). Similar techniques have been applied to image collections, allowing users to search for images containing objects in defined spatial relationships with each other (Chang et al [1988], Chang and Jungert [1991]). Improved algorithms for spatial retrieval are still being proposed [Gudivada and Raghavan, 1995b]. Spatial indexing is seldom useful on its own, though it has proved effective in combination with other cues such as colour (Stricker and Dimai [1996], Smith and Chang [1997a]) and shape [Hou et al, 1992].

Several other types of image feature have been proposed as a basis for CBIR. Most of these rely on complex transformations of pixel intensities which have no obvious counterpart in any human description of an image. Most such techniques aim to extract features which reflect some aspect of image similarity which a human subject can perceive, even if he or she finds it difficult to describe. The most well-researched technique of this kind uses the *wavelet transform* to model an image at several different resolutions. Promising retrieval results have been reported by matching wavelet features computed from query and stored images (Jacobs et al [1995], Liang and Kuo [1998]). Another method giving interesting results is *retrieval by appearance*. Two versions of this method have been developed, one for whole-image matching and one for matching selected parts of an image. The part-image technique involves filtering the image with Gaussian derivatives at multiple scales [Ravela and Manmatha, 1998a], and then computing differential invariants; the whole-image technique uses distributions of local curvature and phase [Ravela and Manmatha, 1998b].

The advantage of all these techniques is that they can describe an image at varying levels of detail (useful in natural scenes where the objects of interest may appear in a variety of guises), and avoid the need to segment the image into regions of interest before shape descriptors can be computed. Despite recent advances in techniques for image segmentation [Campbell at al, 1997], this remains a troublesome problem.

5.2 Video retrieval

As indicated above (section 2.4), video sequences are an increasingly important form of image data. Several groups of researchers have investigated ways in which CBIR techniques can be adapted for video retrieval. While their approaches differ in detail, most have hit on remarkably similar ways of dealing with the problem [Ardizzone and La Cascia, 1997]. The first step is to divide up the video into individual shots. Normally, the change from one shot to the next involves a sudden change in screen image content and camera angle. Such changes can be detected automatically through analysis of colour histograms, texture and motion vectors from sequences of individual frames. From each shot, a single representative *keyframe* is then selected (again, by analysis of colour and texture). The complete set of keyframes for the video thus forms a storyboard for the video, which can then be manually annotated, or stored in an image database for browsing or content-based retrieval. Most commercial CBIR vendors, including Excalibur and Virage, now have products automating at least part of the video data management process.

The field of video data management continues to advance rapidly [Bolle et al, 1998]. One recent advance in technology is the automatic identification and separation of whole scenes, as opposed to individual shots [Yeo and Yeung, 1998]. Another is the ability to automatically extract short video clips representing the key features of much longer sequences, providing users with far more information than the still keyframes they replace [Smith and Kanade, 1998]. Perhaps even more useful is the query-by-motion-example provided by the experimental VideoQ system [Chang et al, 1997]. This allows users to specify the way an object moves across the screen during a video clip, as well as its colour or shape. Thus one can retrieve clips of high-jumpers by specifying a parabolic motion arc for the principal object in a scene.

Still within the sporting field, CBIR techniques have been used in the automatic analysis of tennis shots [Sudhir et al, 1998] and basketball games [Saur et al, 1997]. Finally, the Informedia project at Carnegie-Mellon University [Wactlar et al, 1996] has demonstrated how much synergy can be achieved in video retrieval when information from video, speech and closed-caption text is combined. The ability to use speech information to resolve ambiguities in video interpretation (and vice versa) has led to some quite impressive results.

5.3 Retrieval by semantic image feature

5.3.1 Level 2

The vast majority of current CBIR techniques are designed for primitive-level retrieval. However, some researchers have attempted to bridge the gap between level 1 and level 2 retrieval. One early system aimed at tackling this problem was GRIM_DBMS [Rabbitti and Stanchev, 1989], designed to interpret and retrieve line drawings of objects within a narrow predefined domain, such as floor plans for domestic buildings. The system analysed object drawings, labelling each with a set of possible interpretations and their probabilities. These were then used to derive likely interpretations of the scene within which they appeared.

More recent research has tended to concentrate on one of two problems. The first is scene recognition. It can often be important to identify the overall type scene depicted by an image, both because this in an important filter which can be used when searching, and because this can help in identifying specific objects present. One system of this type is IRIS [Hermes et al, 1995], which uses colour, texture, region and spatial information to derive the most likely interpretation of the scene, generating text descriptors which can be input to any text retrieval system. Other researchers have

identified simpler techniques for scene analysis, using low-frequency image components to train a neural network [Oliva et al, 1997], or colour neighbourhood information extracted from low-resolution images to construct user-defined templates [Ratan and Grimson, 1997].

The second focus of research activity is object recognition, an area of interest to the computer vision community for many years (e.g. Brooks [1981], Connell and Brady [1987], Strat and Fischler [1991]). Techniques are now being developed for recognizing and classifying objects with database retrieval in mind. The best-known work in this field is probably that of Forsyth et al [1997], who have attracted considerable publicity for themselves by developing a technique for recognizing naked human beings within images, though their approach has been applied to a much wider range of objects, including horses and trees. Haering et al [1997] have also developed a method for identifying deciduous trees via their foliage. The ImageMiner system [Alsuth et al, 1998] aims to extend similar techniques into the video domain. All such techniques are based on the idea of developing a model of each class of object to be recognized, identifying image regions which might contain examples of the object, and building up evidence to confirm or rule out the object's presence. Evidence will typically include both features of the candidate region itself (colour, shape or texture) and contextual information such as its position and the type of background in the image.

In contrast to these fully-automatic methods is a family of techniques which allow systems to learn associations between semantic concepts and primitive features from user feedback. The earliest such system was FourEyes from MIT [Minka, 1996]. This invites the user to annotate selected regions of an image, and then proceeds to apply similar semantic labels to areas with similar characteristics. The system is capable of improving its performance with further user feedback. Another approach is the concept of the *semantic visual template* introduced by S F Chang et al [1998]. Here, the user is asked to identify a possible range of colour, texture, shape or motion parameters to express his or her query, which is then refined using relevance feedback techniques. When the user is satisfied, the query is given a semantic label (such as "sunset") and stored in a query database for later use. Over time, this query database becomes a kind of *visual thesaurus*[*], linking each semantic concept to the range of primitive image features most likely to retrieve relevant items.

5.3.2 Level 3

Reports of automatic image retrieval at level 3 are very rare. The only research that falls even remotely into this category has attempted to use the subjective connotations of colour (such as whether a colour is perceived to be warm or cold, or whether two colours go well with each other) to allow retrieval of images evoking a particular mood (Kato et al [1992], Corridoni et al [1998]). It is not at present clear how successful this approach will prove.

5.4 General issues

5.4.1 Interfacing

The ability for users to express their search needs accurately and easily is crucial in any retrieval system. Image retrieval is no exception to this, though it is by no means obvious how this can be achieved in practice. The use of SQL-like query languages was advocated in some early systems like GRIM_DBMS [Rabbitti and Stanchev, 1989], though keyboard input hardly seems an obvious choice for formulating visual queries. The most appealing paradigm in many ways is query-by-

[*] *a concept first proposed by Hogan et al [1991].*

example: providing a sample of the kind of output desired and asking the system to retrieve further examples of the same kind – an approach pioneered by Chang and Fu [1981] with their QPE (query by pictorial example) interface. Virtually all current CBIR systems now offer query-by-example searching, where users submit a query image and the system retrieves and displays thumbnails of (say) the 20 closest-matching images in the database.

However, users will not always have an example image to hand. Several alternative query formulation methods have been proposed here, most based on ideas originally developed for IBM's QBIC system [Flickner et al, 1995]. The original QBIC interface allowed users to specify colour queries either by sliders varying the relative amounts of red, green and blue in the query, or by selecting a desired colour from a palette. Texture queries could also be specified by choosing from a palette, and shape queries by sketching the desired object on the screen [Lee et al, 1994]. These methods proved adequate but often cumbersome, and later versions of the QBIC system have adopted a set of rather more intuitive *pickers* for query specification [Niblack et al, 1998]. Several other authors have developed sketch-based interfaces for shape retrieval (e.g. Chans et al [1997]); others, recognizing that many users have limited artistic ability, allow query shapes to be built up on the screen from primitives such as rectangles and circles (e.g. Smith and Chang [1997b], Sclaroff et al [1997]).

The ability to refine searches online in response to user indications of relevance, known as *relevance feedback*, is particularly useful in image retrieval, even though it was originally developed for text [Salton, 1971]. This is firstly because users can normally judge the relevance of a set of images displayed on the screen within seconds, and secondly because so few current systems are capable of matching users' needs accurately first time round. The usefulness of relevance feedback for image retrieval was first demonstrated within a keyword-based system [Price et al, 1992]; the facility has now been successfully implemented in several CBIR systems (e.g. Smith and Chang [1997b], Rui et al [1998]). However, there is still considerable scope for more research into improved interfaces for image retrieval systems, in particular the development of better methods for users to convey individual notions of image similarity [Santini and Jain, 1997].

Designing interfaces for video retrieval is a more complex task still. Videos are inherently multimedia objects, combining moving images with other media such as text, music and speech. Interfaces need to allow querying by text, and if possible speech, as well as by image feature. Display of retrieved output is also more complex. Some systems display storyboards of still keyframes in chronological sequence [Niblack et al, 1998]; others provide structured graphs illustrating similarities between shots [Yeo and Yeung, 1998]; yet others provide video skims of each scene [Smith and Kanade, 1998].

5.4.2 Search efficiency

A significant limitation of current CBIR technology is the problem of efficiently retrieving the set of stored images most similar to a given query. One of the many fundamental ways in which CBIR differs from text retrieval is that it is based on a fundamentally different model of data. Most text retrieval systems associate each document with a variable number of descriptors representing its content. A given descriptor is either present or absent in a given document. Searching essentially consists of identifying those documents associated with a given set of descriptors, and is thus governed primarily by the rules of symbolic logic. In such systems, search efficiency can be increased by the use of devices such as *inverted file indexes*, each of which holds a list of document identifiers associated with a given descriptor. Boolean searches can readily be implemented by comparing the

indexes for each search term, yielding result sets which can be used directly to address the documents themselves. Probabilistic retrieval systems [Robertson and Sparck Jones, 1976] operate in a rather more complex way, but still rely on a variable number of descriptors to represent each document.

Most current CBIR systems work on a completely different principle. Stored images are typically characterized by fixed-length real-valued multi-component feature vectors, each image having a value for every feature in the database. In this case, searching consists of calculating the similarity between feature vectors from query and stored images, a process of numerical computation. As Santini and Jain [1997] point out, the prime aim of traditional text retrieval systems is to *partition* a database into two sets – relevant items and non-relevant items – even if members of the first set may later be ranked by relevance. By contrast, the prime aim of CBIR systems is to *sort* the database in order of similarity to the query.*

Finding index structures which allow efficient searching of an image database is still an unsolved problem [Faloutsos et al, 1994]. None of the index structures proposed for text retrieval has proved applicable to the problem. The most promising approach so far has been multidimensional indexing, using structures such as the R*-tree [Beckmann et al, 1990], the TV-tree [Lin et al, 1994] and the SS$^+$-tree [Kurniawati et al, 1997], but the overheads of using these complex index structures are considerable. A more recent approach, which seems to offer better prospects of success, is the use of similarity clustering of images, allowing hierarchical access for retrieval and providing a way of browsing the database as a bonus (Jin et al [1998], Vellaikal and Kuo [1998]).

5.5 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.

5.5.1 Commercial systems

QBIC. IBM's QBIC† system [Flickner et al, 1995] 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 [Faloutsos et al, 1994]. At search time, 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 [Niblack et al, 1998]. 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/.

---

* *The situation is not completely clear-cut. A few text retrieval methods use fixed-length feature vectors to describe semantic content (notably latent semantic indexing [Deerwester et al, 1990]), and hence work on exactly the same principle as CBIR systems. But these are exceptions.*

† *Query By Image Content*

Virage. Another well-known commercial system is the VIR Image Engine from Virage, Inc [Gupta et al, 1996]. 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 [Hampapur et al, 1997]; 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 [Feder, 1996]. 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 management system Screening Room (http://www.excalib.com/products/video/screen.html).

5.5.2 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 [Pentland et al, 1996] 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 features relevant to a particular type of search to be computed at search time, 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 [Picard, 1996]. 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 [Ogle and Stonebraker, 1995], 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 [Smith and Chang, 1997a] 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 [Smith and Chang, 1997b] 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 [Chang et al, 1997], a video search engine allowing users to specify motion queries, and MetaSEEk [Beigi et al, 1998], a meta-search engine for images on the Web.

MARS. The MARS* project at the University of Illinois [Huang et al, 1997] 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 [Rui et al, 1998]. 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 [Wactlar et al, 1996] 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 (e.g. Christel et al [1997]). 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.

---

* *Multimedia Analysis and Retrieval System*

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 [Ma and Manjunath, 1997]. 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* (section 5.1.4) using whole image matching [Ravela and Manmatha, 1998b]. A demonstration of Synapse in action with a variety of different image types can be found at http://cowarie.cs.umass.edu/~demo/.

5.6 Practical applications of CBIR

A wide range of possible applications for CBIR technology has been identified (e.g. Gudivada and Raghavan [1995a]). Potentially fruitful areas include:

- 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.

Closer examination of many of these areas reveals that, while research groups are developing prototype systems, and practitioners are experimenting with the technology, few examples of fully-operational CBIR systems can yet be found. A search of public-domain sources, including the trade and scientific literature and the Web, suggests that the current state of play in each of these areas at the end of 1998 is as follows:

5.6.1 Crime prevention

Law enforcement agencies typically maintain large archives of visual evidence, including past suspects' facial photographs (generally known as mugshots), fingerprints, tyre treads and shoeprints. Whenever a serious crime is committed, they can compare evidence from the scene of the crime for its similarity to records in their archives. Strictly speaking, this is an example of *identity* rather than *similarity* matching, though since all such images vary naturally over time, the distinction is of little practical significance. Of more relevance is the distinction between systems designed for verifying the identity of a known individual (requiring matching against only a single stored record), and those

capable of searching an entire database to find the closest matching records. As discussed in section 2.5, only the latter task can truly be called CBIR.

The basic techniques for automatic fingerprint matching were worked out in the 1980s [IEEE, 1985], and systems based on this technology are now in routine use at the FBI in Washington, as well as a number of police forces around the world. A number of AFIS (automatic fingerprint identification systems) are now commercially available, including AFIX Tracker from the Phoenix Group Inc, Pittsburg, Kansas (http://www.afix.com/) and the Finger Search Engine from East Shore Technologies, Inc. of New York (http://www.east-shore.com/). Both are capable of searching large databases of fingerprints.

Face recognition is also a reasonably mature technology. Most current systems use either a version of the *eigenface* method initially developed for the Photobook system at MIT [Pentland et al, 1996], or local feature matching along lines proposed by Bach et al [1993], The former method is most successful with mugshots, where lighting and pose can be carefully controlled; the latter method is more robust where faces may appear at any angle under a variety of lighting conditions, as in security videos. MIT's eigenface technology has been commercialized by Viisage Technology of Littleton, Massachusetts (http://www.viisage.com) in their FaceID package, used (among others) by the Los Angeles Police Department [Strother-Vien, 1998]. Local feature analysis forms the basis of the FaceIt system from Visionics Corp of Jersey City, New Jersey (http://www.faceit.com), which has been used for immigration control in both the USA and Malaysia [Okon, 1998]. Similar systems are under evaluation by a number of UK police forces.

5.6.2 The military

Military applications of imaging technology are probably the best-developed, though least publicized. Recognition of enemy aircraft from radar screens, identification of targets from satellite photographs, and provision of guidance systems for cruise missiles are known examples – though these almost certainly represent only the tip of the iceberg. Many of the surveillance techniques used in crime prevention could also be relevant to the military field.

5.6.3 Intellectual property

Trademark image registration, where a new candidate mark is compared with existing marks to ensure that there is no risk of confusion, has long been recognized as a prime application area for CBIR. Copyright protection is also a potentially important application area. Enforcing image copyright when electronic versions of the images can easily be transmitted over the Internet in a variety of formats is an increasingly difficult task. There is a growing need for copyright owners to be able to seek out and identify unauthorised copies of images, particularly if they have been altered in some way.

General-purpose CBIR packages such as QBIC and Virage have been tested on trademark image databases. (A demonstration trademark image search system based on QBIC can be found at http://wwwqbic.almaden.ibm.com/tmdemo/). Several experimental systems specifically designed for trademark searching have also been described in the literature. These include TRADEMARK from the Electrotechnical Laboratory, Tsukuba, Japan [Kato, 1992], STAR from the University of Singapore [Wu et al, 1996], and ARTISAN from the University of Northumbria at Newcastle, UK [Eakins et al, 1998]. Both STAR and ARTISAN were developed in close collaboration with the respective national Patent Offices. Encouraging results have been reported, and it is expected that at

least one commercial trademark search system using CBIR technology will be on the market before the end of 1999.

Prototype systems specifically using CBIR techniques for identifying illicit copies of images on the Web are also at an advanced stage of development [e.g. Chang, E Y et al, 1998].

5.6.4 Architectural and engineering design

Architectural and engineering design share a number of common features – the use of stylized 2- and 3-D models to represent design objects, the need to visualize designs for the benefit of non-technical clients, and the need to work within externally-imposed constraints, often financial. Such constraints mean that the designer needs to be aware of previous designs, particularly if these can be adapted to the problem at hand. Hence the ability to search design archives for previous examples which are in some way similar, or meet specified suitability criteria, can be valuable.

Despite some early experimentation, such as SAFARI [Eakins, 1993], AUGURS [Yang et al, 1994], and more recent research into 3D information management by Ramesh Jain's group at the University of California at San Diego (http://vision.ucsd.edu/~dwhite/#3Dcbr), there is no evidence of commercial or near-market development activity in this area at present. This is in many ways surprising, given the potential savings which could accrue from successful design reuse. Possible reasons for this are the difficulty of handling 3-D object and query representations in a systematic way, and the specialized nature of the market. To gain widespread acceptance among designers, any such software tool would need to be closely integrated with existing CAD packages.

5.6.5 Fashion and interior design

Similarities can also be observed in the design process in other fields, including fashion and interior design. Here again, the designer has to work within externally-imposed constraints, such as choice of materials. The ability to search a collection of fabrics to find a particular combination of colour or texture is increasingly being recognized as a useful aid to the design process.

So far, little systematic development activity has been reported in this area. Attempts have been made to use general-purpose CBIR packages for specific tasks such as colour matching of items from electronic versions of mail-order catalogues [Petkovic, 1996], and identifying textile samples bearing a desired pattern [Bird et al, 1996], but no commercial use appears to be made of this at present.

5.6.6 Journalism and advertising

Both newspapers and stock shot agencies maintain archives of still photographs to illustrate articles or advertising copy. These archives can often be extremely large (running into millions of images), and dauntingly expensive to maintain if detailed keyword indexing is provided. Broadcasting corporations are faced with an even bigger problem, having to deal with millions of hours of archive video footage, which are almost impossible to annotate without some degree of automatic assistance.

This application area is probably one of the prime users of CBIR technology at present – though not in the form originally envisaged. In the early years of CBIR development, hopes were high that the technology would provide efficient and effective retrieval of still images from photo libraries, eliminating or at least substantially reducing the need for manual keyword indexing. Disillusionment set in as the realization spread that the CBIR techniques under development were of little use for retrieval by semantic content. Stock shot agencies now seem likely to base their retrieval systems on

manual keywording for many years to come, though a few are experimenting with the use of CBIR software as adjuncts to keyword indexing.

However, the area of *video asset management*, the organization for efficient reuse of archive video footage, has benefited enormously from what might be described as spin-offs from the CBIR development programmes. As indicated above, CBIR techniques can be used to break up a video sequence into individual shots, and generate representative keyframes for each shot. It is therefore possible to generate a storyboard for each video entirely by automatic means. Even if one then uses traditional methods to index and classify the video, there can be large savings in time and cost. As a result, TV companies in the USA are now using this technology in a big way, and their counterparts in Europe are beginning to follow suit. Leading established products in this area include Virage's Videologger (http://www.virage.com/market/cataloger.html), and Excalibur's Screening Room (http://www.excalib.com/products/video/screen.html), both of which automatically create storyboards of thumbnail images which users then annotate manually. Further advances in technology, allowing direct search of video content with a much reduced level of manual annotation, can confidently be expected in the near future. Recently-announced products offering a degree of automatic video indexing and retrieval include the Mediakey Digital Video Library System from Islip Media, Inc (http://www.islip.com/fprod.htm), based on Carnegie-Mellon University's highly-successful Informedia technology [Wactlar, 1996], and Visionary from the Israel-based Media Access Technologies Ltd [Wilf, 1998].

5.6.7 Medical diagnosis

The increasing reliance of modern medicine on diagnostic techniques such as radiology, histopathology, and computerised tomography has resulted in an explosion in the number and importance of medical images now stored by most hospitals. While the prime requirement for medical imaging systems is to be able to display images relating to a named patient, there is increasing interest in the use of CBIR techniques to aid diagnosis by identifying similar past cases.

Most development work in the PACS (picture archiving and communication systems) area is still directed towards providing basic functionality (ensuring that medical images can be successfully digitized, stored and transmitted over local area networks without loss of quality) and usability (providing user-centred interfaces and integrating image storage and retrieval with wider aspects of patient record management). However, experimental content-based retrieval systems are beginning to have some impact. Examples of this include the I^2C system for retrieving 2-D radiological images from the University of Crete [Orphanoudakis et al, 1994], and the 3-D neurological image retrieval system currently being developed at Carnegie-Mellon University [Liu et al, 1998], both developed with the aim of assisting medical staff in diagnosing brain tumours.

5.6.8 Geographical information systems (GIS) and remote sensing

Although not strictly a case of *image* retrieval, managers responsible for planning marketing and distribution in large corporations need to be able to search by spatial attribute (e.g. to find the 10 retail outlets closest to a given warehouse). And the military are not the only group interested in analysing satellite images. Agriculturalists and physical geographers use such images extensively, both in research and for more practical purposes, such as identifying areas where crops are diseased or lacking in nutrients – or alerting governments to farmers growing crops on land they have been paid to leave lying fallow.

Commercial GIS systems such as ArcInfo have provided the capacity to search spatially referenced data by location or attribute for many years. This is an extremely useful function, but whether it can be considered a form of CBIR is a moot point. Experimental systems aimed at identifying objects or regions within satellite images or digitized maps by shape, colour or texture similarity have frequently been reported in the literature (e.g. Kitamoto et al [1993], Soffer and Samet [1996], Ma and Manjunath [1998] and Li et al [1998]), though these have not so far resulted in commercial products.

5.6.9 Cultural heritage

Museums and art galleries deal in inherently visual objects. The ability to identify objects sharing some aspect of visual similarity can be useful both to researchers trying to trace historical influences, and to art lovers looking for further examples of paintings or sculptures appealing to their taste. However, many of the image queries put to art libraries are at levels 2 or 3 as defined in section 2.3 above, well beyond the capabilities of the current generation of CBIR systems.

This has not deterred Kato and his colleagues (Kato and Kurita [1990], Hirata and Kato [1992]) from developing an experimental system for retrieving paintings similar to a given sketch, though few details of its retrieval effectiveness are available. More recently, IBM's QBIC system has received extensive trials in managing art library databases [Holt and Hartwick, 1994], and has proved an extremely useful browsing tool even if its retrieval effectiveness has been limited. Jain et al [1997] have applied CBIR techniques to the management of image and video data relating to a Hindu temple in India. Their article highlights both the opportunities and problems associated with cultural heritage applications of CBIR.

5.6.10 Education and training

It is often difficult to identify good teaching material to illustrate key points in a lecture or self-study module. The availability of searchable collections of video clips providing examples of (say) avalanches for a lecture on mountain safety, or traffic congestion for a course on urban planning, could reduce preparation time and lead to improved teaching quality. In some cases (complex diagnostic and repair procedures) such videos might even replace a human tutor.

Reports of the application of CBIR technology to education and training have so far been sparse – though Carnegie-Mellon University's Informedia system is being trialled at a number of universities, including the Open University in the UK [van der Zwan et al, 1999]. It appears to be too early to form any definite conclusions about the system's effectiveness in practice.

5.6.11 Home entertainment

Much home entertainment is image or video-based, including holiday snapshots, home videos and scenes from favourite TV programmes or films. This is one of the few areas where a mass market for CBIR technology could develop. Possible applications could include management of family photo albums ('find that photo of Aunt Sue on the beach at Brighton') or clips from commercial films ('play me all the car chases from James Bond movies').

Despite a lack of published information about developments in this area, a number of large commercial organizations are known to be devoting substantial development effort into this problem at present, and are believed to be making significant progress. Despite some formidable difficulties – the software will need to offer effective semantic-level retrieval, be far easier to use than any of

today's systems, and come at an affordable price – the rewards for success could be enormous. This application area has the potential to drive virtually all future CBIR development activity if it ever takes off.

5.6.12 Web searching

Cutting across many of the above application areas is the need for effective location of both text and images on the Web, which has developed over the last five years into an indispensable source of both information and entertainment. Text-based search engines have grown rapidly in usage as the Web has expanded; the well-publicized difficulty of locating images on the Web [Jain, 1995] indicates that there is a clear need for image search tools of similar power. Paradoxically, there is also a need for software to *prevent* access to images which are deemed pornographic.

Several experimental systems for content-based image searching on the Web have been demonstrated over the last two to three years. These include WebSEEk [Smith and Chang, 1997b] and ImageRover [Sclaroff et al, 1997], both of which use remote agents to seek out and index images, which can then be searched by keyword or image content similarity. Both systems also provide a relevance feedback facility to refine search results. The MetaSEEk project has also demonstrated the feasibility of implementing a meta-search engine for images on the Web [Beigi et al, 1998].

Two commercial Web search engines now offer a CBIR option. 1998 saw the launch of both the Yahoo! Image Surfer, based on Excalibur technology, and AltaVista's AV Photo Finder, using Virage technology. (See section 5.5 above for relevant URLs). These systems inevitably offer a more limited range of features than experimental systems such as WebSEEk (for example, relevance feedback is not provided for efficiency reasons), but offer users a useful extra method of locating desired images on the Web.

Automatic pornography filters for images are a significant research topic, but the problems involved in defining what constitutes pornography in a way that image processing algorithms can understand guarantee that progress will be slow. Most of the work to date in this area (e.g. [Forsyth et al, 1997], Chan et al, 1999]) has concentrated on automatically identifying images of naked humans by skin tone and limb shape.

5.6.13 Conclusions

The extent to which CBIR technology is currently in routine use is clearly still very limited. In particular, CBIR technology has so far had little impact on the more general applications of image searching, such as journalism or home entertainment. Only in very specialist areas such as crime prevention has CBIR technology been adopted to any significant extent. This is no coincidence – while the problems of image retrieval in a general context have not yet been satisfactorily solved, the well-known artificial intelligence principle of exploiting natural constraints has been successfully adopted by system designers working within restricted domains where shape, colour or texture features play an important part in retrieval.

5.7 Current research trends

The volume of research activity into CBIR techniques and systems continues to grow, though much recent work seems rather derivative, providing research training for PhD students rather than significantly advancing the state of the art. It is notable that MIT, originators of Photobook, have now

decided to move out of CBIR into another research field, on the grounds that too many other groups are getting in on the act. Despite this, significant research problems remain to be addressed, including better methods of segmenting images to distinguish objects of interest from their background (or alternatively, improved techniques for feature extraction which do not rely on segmentation), new paradigms for user interaction with CBIR systems, and better ways of representing human judgements of image similarity. Above all, there is a need to bridge the semantic gap, bringing some measure of automation to the processes of indexing and retrieving images by the type of object or scene depicted.

Research into all these problems is actively under way, particularly in the USA, which can boast over 20 large well-funded research teams in universities and industrial research centres across the country. Many of these have an interdisciplinary research focus, combining experts from image processing, information retrieval, psychology, and other disciplines – a marked contrast to the situation in the early 1990s, when most systems were developed by image processing experts with little or no interest in human factors. Synergy between researchers from different disciplines has already shown beneficial results in areas such as relevance feedback, which has the potential to improve system effectiveness out of all recognition. More can be expected in the future, as the barriers between different research communities noted by Cawkell [1993] gradually come down.

Research topics receiving substantial attention at present include:

?? improved methods for Web searching, allowing users to identify images of interest in remote sites by a variety of image and textual cues

?? improved video retrieval techniques, including automatic segmentation, query-by-motion facilities, and integration of sound and video searching

?? better user interaction, including improved techniques for image browsing and exploiting user feedback

?? automatic or semi-automatic methods of capturing image semantics for retrieval.

Outside the USA, major research activity is largely limited to a few centres of excellence in Japan, Singapore, Australia, France, Italy, Holland, and Germany. The UK's research impact in world terms has so far been negligible, at least in part because of a less than generous funding climate, though research groups with potential are now beginning to emerge here as well. It is noteworthy that the most successful groups outside the USA all maintain close links with USA-based research groups.

6 Implications for systems developers and users

6.1 Effectiveness of current CBIR techniques

Hard information on the effectiveness of automatic CBIR techniques is difficult to come by. Few of the early systems developers made serious attempts to evaluate their retrieval effectiveness, simply providing examples of retrieval output to demonstrate system capabilities. The QBIC team were among the first to take the question of retrieval effectiveness seriously [Faloutsos, 1994]; though even they glossed over some of the problems of determining whether a given image did in fact answer a given query. The position is changing as more researchers with an information retrieval background enter the field, though, as discussed in section 4.5, the problems of evaluating multimedia information retrieval systems are substantial (see also the discussions of the ESPRIT-funded MIRA group at http://www.dcs.gla.ac.uk/mira/workshops/). System developers do now generally report

effectiveness measures such as precision and recall with a test database, though few discuss subjective measures of user satisfaction. In the absence of comparative retrieval effectiveness scores measuring the effectiveness of two different systems on the same set of data and queries, it is difficult to draw many firm conclusions. All that can be said is that retrieval effectiveness scores reported on image retrieval systems (e.g. Manmatha and Ravela [1997], Eakins et al [1997]) are in the same ball park as those commonly reported for text retrieval.

However, the main drawback of current CBIR systems is more fundamental. It is that the only retrieval cues they can exploit are primitive features such as colour, texture and shape. Hence current CBIR systems are likely to be of significant use only for applications at level 1. This restricts their prime usefulness to specialist application areas such as fingerprint matching, trademark retrieval or fabric selection. IBM's QBIC system (see section 5.5 above) has been applied to a variety of tasks, but seems to have been most successful in specialist areas such as colour matching of items in electronic mail-order catalogues, and classification of geological samples on the basis of texture. Similarly, the main commercial application of MIT's Photobook technology has been in the specialist area of face recognition.

Within specialist level 1 applications, CBIR technology does appear to be capable of delivering useful results, though it should be borne in mind that some types of feature have proved much more effective than others. It is generally accepted that colour and texture retrieval yield better results (in that machine judgements of similarity tally well with those of human observers) than shape matching [Faloutsos et al, 1994]. Part of the problem with shape matching lies in the difficulty of automatically distinguishing between foreground shapes and background detail in a natural image [Flickner et al, 1995]. Even when faced with stylized images, or scenes where human intervention has been used to distinguish foreground from background, though, shape retrieval systems often perform poorly. A major contributing factor here is almost certainly the fact that few, f any, of the shape feature measures in current use are accurate predictors of human judgements of shape similarity [Scassellati et al, 1994]. Santini and Jain [in press] go further, and show that none of the commonly-used models of image similarity is capable of capturing the full complexity of the process. More recent techniques based on wavelets [Liang and Kuo, 1988] or Gaussian filtering [Manmatha and Ravela, 1997] appear to perform well in the retrieval experiments reported by their creators – though again, it is difficult to compare their effectiveness with more traditional methods, as no comparative evaluation studies have been performed. An image retrieval version of the TREC text retrieval experiments [Sparck Jones, 1995] might well prove useful here.

Although current CBIR systems use only primitive features for image matching, this does not limit their scope exclusively to level 1 queries. With a little ingenuity on the part of the searcher, they can be used to retrieve images of desired objects or scenes in many cases. A query for beach scenes, for example, can be formulated by specifying images with blue at the top and yellow underneath; a query for images of fish by sketching a typical fish on the screen. Images of specific objects such as the Eiffel Tower can be retrieved by submitting an accurate scale drawing, provided the angle of view is not too different. A skilled search intermediary could thus handle some level 2 queries with current technology, though it is not yet clear how large a range of queries can be successfully handled in this way. Alternatively, if an image database that has been indexed using keywords or descriptive captions is available, it is possible to combine keyword and image similarity querying (sometimes known as *hybrid image retrieval*). This issue is explored in some detail in section 6.2 below.

Overall, current CBIR techniques may well have a part to play in specialist colour or shape-matching applications. It is also possible that they could be of use in enhancing the effectiveness of general-

purpose text-based image retrieval systems. But major advances in technology will be needed before systems capable of automatic semantic feature recognition and indexing become available. Hence the chances of CBIR *superseding* manual indexing in the near future for general applications handling semantic (level 2 or 3) queries look remote. Research into semantic image retrieval techniques is beginning to gather momentum, particularly in restricted domains (such as identifying unclothed human bodies) where it is possible to develop detailed models of the objects involved. But it will take a considerable time before such research finds its way into commercially-available products.

6.2 CBIR *vs* manual indexing

At the present stage of CBIR development, it is meaningless to ask whether CBIR techniques perform better or worse than manual indexing. Potentially, CBIR techniques have a number of advantages over manual indexing. They are inherently quicker, cheaper, and completely objective in their operation. However, these are secondary issues. The prime issue has to be retrieval effectiveness – how well does each type of system work? Unfortunately, the two types of technique cannot be sensibly compared, as they are designed to answer different types of query. Given a specialist application at level 1, such as trademark retrieval, CBIR often performs better than keyword indexing, because many of the images cannot adequately be described by linguistic cues. But for a level 2 application like finding a photograph of a given type of object to illustrate a newspaper article, keyword indexing is more effective, because CBIR simply cannot cope. It should be remembered, though, that manual classification and indexing techniques for images also have their limitations, particularly the difficulty of anticipating the retrieval cues future searchers will actually use [Enser, 1995]. As observed above (see section 4.5), there is remarkably little hard evidence on the effectiveness of text keywords in image retrieval.

Attempts to retrieve images by the exclusive use of keywords or primitive image features have not met with unqualified success. Is the use of keywords and image features *in combination* likely to prove any more effective? There are in fact several reasons for believing this to be the case. Firstly, keyword indexing can be used to capture an image's semantic content, describing objects which are clearly identifiable by linguistic cues, such as trees or cars. Primitive feature matching can usefully complement this by identifying aspects of an image which are hard to name, such as a particular shape of roof on a building. Secondly, evaluation studies of the Chabot system [Ogle and Stonebraker, 1995] showed that higher precision and recall scores could be achieved when text and colour similarity were used in combination than when either was used separately. Finally, theoretical support for this idea comes from Ingwersen's [1996] cognitive model of IR, which predicts that retrieval by a combination of methods using different cognitive structures is likely to be more effective than by any single method.

Several other CBIR systems besides Chabot provide such capabilities – including QBIC, the Yahoo! Image surfer, Alta Vista's AV Photo Finder, and VisualSEEk. Users of these systems typically submit a keyword query to retrieve an initial set of images, each of which can then be used to initiate a similarity search, retrieving additional images which may have been indexed under different keywords. Here, CBIR techniques are being used as an adjunct to traditional search methods, as a recall-enhancing device. An alternative approach, which appears to hold significant promise, is the use of CBIR techniques in combination with latent semantic indexing for images retrieval on the Web [la Cascia et al, 1998].

Further synergies between text and image feature indexing are possible. The specialized Piction system described by Srihari [1995] uses text from image captions to guide face recognition for image

indexing. The SemCap system of Gudivada and Jung [1996] uses techniques from clinical psychology to derive semantic features from images to enhance primitive feature matching. The VisualSEEk system allows users to add descriptive keywords to an entire set of similar images in a single operation [Smith and Chang, 1997a], greatly speeding up the process of manually indexing an image collection. A further possible refinement is the development of search expansion aids such as the *visual thesaurus* [Hogan et al, 1991], designed to link similar-looking objects. Finally, the Informedia project [Wactlar et al, 1996] has shown that automatic indexing of videos through the simultaneous analysis of both images and speech on the sound track can significantly improve indexing effectiveness.

However, little systematic evaluation of the effectiveness of such techniques has yet been undertaken. Hence key questions such as "can CBIR techniques bring about worthwhile improvements in performance with real-life image retrieval systems?" and "how can any such synergies most effectively be exploited?" thus remain unanswered. The need for research in this area is underlined by Sutcliffe et al [1997], who write: *"Content-based retrieval techniques are still in their infancy. Furthermore, we have little understanding of when such techniques may be profitably used in combination with, or instead of, more traditional query languages".*

6.3 CBIR in context

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. Video editors are unlikely to be satisfied simply with the ability to view retrieved video sequences; they want to extract them for incorporation in future programmes. It is important to stress that CBIR is never more than the means to an end.

One implication of this is that a prime future use of CBIR is likely to be the retrieval of images by content in a multimedia system. We have already discussed possible synergies between text and image searching in section 6.2. Opportunities for synergy in true multimedia systems will be far greater, as already demonstrated by the Informedia project [Wactlar et al, 1996], which combines still and moving image data, sound and text in generating retrieval cues. One example of such synergy revealed by their retrieval experiments was that in the presence of visual cues, almost 100% recall could be achieved even with a 30% error rate in automatic word recognition.

One aspect of multimedia systems that could be much more widely exploited than at present is their use of hyperlinks to point readers to related items of data, whether elsewhere in the same document or at a remote location. The Microcosm project [Hill et al, 1993] pioneered the concept of the *generic link*, which uses a given text string as source rather than a specific document location. This allows users to follow links from *any* occurrence of that word, irrespective of whether the document author has specifically indicated that link. This is achieved by storing all links in a separate *link database*, which can be queried either by highlighting a particular word in a source document, or by direct keyword search of the database. Either way, all links from that word will be retrieved.

This concept has now been extended by the development of MAVIS, a multimedia architecture which allows generic navigation by image content (shape, colour or texture) as well as text [Lewis et al, 1996]. The authors term this process *content-based navigation* (CBN). A further development of this principle is the *multimedia thesaurus* [Lewis et al, 1997], which allows a system administrator to specify semantic relationships* between source items in the link database, whether text, image or sound. An indication of how such a thesaurus can be used to support multimedia search and navigation is given by Dobie et al [1999]. Such a thesaurus is in principle very powerful, as it allows system users to create a database of semantic relationships between text terms and their corresponding images. It represents yet another possible route to closing the semantic gap discussed in section 5.3 above.

6.4 Standards for image data management

6.4.1 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 a different matter – see below). 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. However, these are likely to pale into insignificance in comparison with the efforts now under way by the Moving Picture Experts Group (MPEG) to develop a standard for data description and

* *such as a given item's set of synonyms, broader and narrower terms*

representation specifically geared to the needs of CBIR. The new MPEG-7 standard, due to be approved in September 2001 if all goes according to schedule, will have substantial effects on almost every aspect of CBIR activity. Some of its likely implications are discussed in section 6.4.5 below.

6.4.2 Image compression

High-resolution images can occupy large amounts of storage (around 17.5 Mb for one A5 colour image scanned at 600 dpi). The need to compress image data for machine processing, storage and transmission was therefore recognized early on. Many standard file formats for still images, such as TIFF, BMP and PCX, allow for images to be stored either in compressed or uncompressed mode; some, like GIF and JPEG, always store images in compressed form. (For further details of these and other formats, see standard texts such as Kay and Levine [1992]). For video images the problem of data volume is much more acute, leading to the development of video compression standards such as MPEG-1 (http://drogo.cselt.stet.it/mpeg/standards/mpeg-1/mpeg-1.htm), a generic standard for compression of video and associated audio, and the more recent MPEG-2 (http://drogo.cselt.stet.it/mpeg/standards/mpeg-2/mpeg-2.htm), designed for use in high-bandwidth environments. Both standards work in a similar way, storing full representations only of a small proportion of frames; using these to predict the content of intermediate frames.

Techniques for directly extracting colour and texture features for retrieval from still images compressed in the well-known JPEG format [Wallace et al, 1991] have been reported by several researchers (e.g. Chang [1995], Wan and Kuo [1996]). Wavelet-based techniques which combine compression with feature extraction for retrieval have also been proposed [Idris and Panchanathan, 1997b]. However, most such research has been concentrated on content extraction from MPEG-compressed moving images. The reasons for this are not difficult to understand. Firstly, the process of extracting retrieval features from MPEG-compressed video is potentially far quicker than first decompressing the files, and then processing enormous volumes of uncompressed data. Secondly, compressed-domain feature extraction can take advantage of one of MPEG's most useful features – *motion vectors* giving a direct indication of how blocks of corresponding pixels move between frames. These can be directly interrogated to allow automatic shot boundary detection [Zhang et al, 1995], query-by-motion-example [S F Chang et al, 1997], or object tracking [Schonfield and Lelescu, 1998].

Commercial video search and analysis systems are already following the lead of such research efforts, and basing their algorithms on MPEG-compressed rather than raw video data. The obvious implication of this is that future developments in compression standards such as MPEG-2 could have a very significant impact on video search system technology.

6.4.3 Query specification

To search a networked image database, whether over the Web or a local area network, requires some means of transmitting users' queries from the terminal to the server holding the database. (It also requires mechanisms for transmitting retrieved images in the opposite direction, and displaying them on the terminal screen, but these are general issues, not specific to CBIR). As users become more sophisticated, they will increasingly feel the need to search a whole variety of databases for the images they want. Since they will not want to reformulate their search from scratch every time they access a new database, some form of standardization is required.

In the text retrieval area, one such standard already exists - Z39.50 (http://lcweb.loc.gov/z3950/agency/). The Z39.50 standard is an applications layer protocol designed to allow program-to-program communication between networked bibliographic retrieval systems. It provides a standard framework for transmitting search statements and retrieved documents between hosts, even though these may use quite different internal representations. Hence it allows searchers to put the same query to multiple remote databases, and receive results in a common format.

The original standard was developed long before CBIR came on the scene, and was designed exclusively with text retrieval in mind. It is therefore hardly surprising that CBIR searching cannot be supported by Z39.50-compliant systems at present, other than as a very special case. As discussed in section 5.4.2 above, the vast majority of current CBIR systems are based on a completely different model of data from the bibliographic systems that Z39.50 was designed to handle. None of the common query types supported by the Z39.50 standard can handle the search arguments needed for CBIR (arrays of real numbers representing extracted image features, and specifications of similarity matching algorithms). In theory, the user-defined type-0 query could handle image data, but since this can be used only to communicate with remote sites using essentially identical software, all the benefits of standardization are lost.

Further extensions to the standard are thus needed before it can handle content-based search and retrieval of images. At present, there are no indications that such extensions are planned. Developments in this area will almost certainly have to wait for progress on standardization within the CBIR community itself (see section 6.4.5) if they are to have any chance of success. It is therefore most unlikely that Z39.50-compliant CBIR will be possible within the next few years.

Similar arguments apply to database query languages such as SQL, even though the philosophy behind these is very different from Z39.50. SQL (Structured Query Language) is an ISO standard for database query specification [Date and Darwen, 1992]. It was originally designed as an end-user interface for specifying search and retrieval operations on relational database, but has evolved over the years into a much more complex language for data manipulation, including automatic integrity checking, and support for abstract data types and large binary objects such as images. (See the SQL website at http://www.jcc.com/sql_stnd.html for further details). Few, if any, end-users now use SQL directly, but it remains important as a way of transmitting queries specified by a point-and-click interface to a database.

Like Z39.50, SQL in its native form is incapable of directly expressing CBIR search arguments. In contrast to Z39.50, however, considerable thought has gone into ways in which it could be extended. Pictorial database languages based on SQL have been proposed at regular intervals over the last 20 years (see [Chang and Fu, 1981] for one of the earliest examples). A number of prototype extensions to the language have been described in the literature, including both spatial reasoning [Roussopoulos et al, 1988] and image feature similarity matching [Ogle and Stonebraker, 1995]. However, none of these extensions has yet found its way into the SQL standard. Hence the comments made above about the Z39.50 standard apply equally to SQL.

6.4.4 Metadata description

Metadata ('data about data') form an essential part of any data archive, providing descriptive data about each stored object which form the main basis for their retrieval. Such metadata typically include catalogue information such as the object's creator, form and date of creation (*this item is a 24 ? 36 mm colour slide, taken on 15/1/94 by John Eakins*) – but may also contain information

about the object's content (*this picture shows Trafalgar Square under snow*). Although the first type of metadata has little specific relevance for CBIR, the second type does impinge to some extent. One might thus expect standards activity in the metadata field to have some relevance to CBIR, but for its impact to be limited.

An example of this is the Dublin Core metadata set (http://purl.org/metadata/dublin_core). Although not an official standard, this set of metadata elements is now widely used for describing Web documents. Dublin Core elements include creator, title, subject keywords, resource type and format. The only image-related aspects of the element set are the ability to define the existence of an image in a document ("DC.Type = image"), and to describe an image's subject content by the use of keywords. It is hard to see how such a framework could usefully be extended to handle CBIR-related concepts.

Of potentially greater relevance is the work currently under way by the World-Wide Web Consortium (W3C – see http://www.w3.org/ for a description of their aims and current activities) in developing a standard Resource Description Framework (RDF) for metadata. The aim of this Framework (see http://www.w3.org/TR/PR-rdf-syntax/ for full details) is to provide a foundation for processing metadata in Web documents, allowing interoperability between applications exchanging machine-readable information. It will specify a framework for detailed description of all kinds of object stored on the Web, allowing search engines to identify relevant content with much greater precision than is at present possible. The specification allows users to define attribute types and values relevant to their own needs, hence providing sufficient extensibility to meet a whole range of specialist needs. The preferred language for writing RDF schemas is the Extensible Markup Language XML (http://www.w3.org/XML/). This, like the better-known Web page authoring language HTML, is in fact a derivative of the generalized text markup language SGML (http://www.oasis-open.org/cover/). It provides a flexible language in which to define a whole range of document types (unlike HTML, which can define only a single type) and their associated metadata.

Although RDF and XML are essentially text-based, their extensibility implies that it should be possible to use them to encapsulate most types of data required for CBIR. How elegantly this can be accomplished remains to be seen. For example, the colour histogram of an image might be specified by defining a ColourHistogram data type within XML. Each instance of this type would then be defined by indicating parameters such as the colour space on which the histogram was based, and the number and extent of bins on each axis, followed by a tagged array of numbers representing the actual histogram. CBIR search engines could access such definitions via appropriate APIs such as W3C's Document Object Model (http://www.w3.org/T/PR-DOM-Level-1/introduction.html), translate them into their own internal format, and compare them with corresponding data from other images. Such a definition would be extremely verbose (particularly if all numeric data were stored as text strings), and only marginally human-readable – but it would accomplish its purpose. There are indications that the Description Definition Language for the emerging MPEG-7 standard (see below) may in fact be based on XML. In this respect, therefore, RDF and XML may well influence (and be influenced by) future developments in CBIR.

6.4.5 MPEG-7

Unlike most of the standards discussed above, one emerging standard – MPEG-7 – is set to make a major impact in the CBIR field. The standard is currently at a very early stage: the first preliminary Working Draft is not due to appear until July 1999, with the final version of the ISO standard

scheduled for November 2000. Its importance stems form the fact that it is the only standard specifically aimed at representing multimedia content – the core of CBIR technology.

Previous standards from MPEG have concentrated on image compression (MPEG-1, MPEG-2), and ways of separately representing foreground objects and background (MPEG-4). These standards have had little impact on the information retrieval community, even those members dealing with image data. The new MPEG-7 standard (http://drogo.cselt.stet.it/mpeg/standards/mpeg-7/mpeg-7.html) is the first to address the issue of multimedia content at all seriously. It aims to set up a standard framework for describing all aspects of a multimedia item's content, including:

- low-level descriptions of each individual object in a scene, such as shape, size, colour, position and movement;
- high-level abstract descriptions of the scene, the objects it contains, and the event(s) taking place;
- audio information such as key, mood and tempo.

Thus an MPEG-7 description of a video clip might consist of a set of codes conveying information such as: "This scene contains a barking brown dog on the left and a blue ball dropping on the right, with the sound of passing cars in the background". It would also contain associated metadata such as the format of the video, when and by whom it was recorded, and copyright information. In principle, then, the MPEG-7 specification will cover the entire range of features required for CBIR and content-based audio retrieval (CBAR).

The standard aims to define a Description Definition Language (DDL) in which to write Descriptor Schemas (DSs), specifying the set of features which describe a video's image and audio content. For further details see http://drogo.cselt.stet.it/mpeg/public/w2461.html. It is emphatically *not* intended to cover either the methods by which those features will be extracted, or the way in which search engines make use of the features for retrieval. Any level of feature may be defined in the standard – not all will be suitable for automatic extraction. While the most common use of MPEG-7 will probably be to describe the content of digitized video, no limits are set on the medium of the data being described. MPEG-7 codes can quite legitimately be used to describe the content of still images recorded on paper, described using terms assigned by a human indexer.

The potential benefits of the new standard are considerable. It should make the process of searching for a desired image a great deal easier, since future MPEG-7-based search engines will simply need to process values of defined standard parameters, rather than computing search features from scratch. For the same reasons, the standard will enormously enhance system interoperability, since all search engines will potentially be using compatible features. This is likely to have a major impact on image searching on the Web, which will become a far more efficient process once a significant number of images enhanced with MPEG-7 metadata become available. If the designers of the standard can get it right by choosing a rich and robust set of feature types and representation methods, it should boost the acceptance of CBIR in the marketplace quite considerably.

Inevitably, reservations have been expressed about MPEG-7. Chief among these is the worry that the field of image retrieval is not yet sufficiently mature for a set of preferred retrieval features to be defined. For example, colour histograms are widely used for similarity matching of images, though many variants on Swain and Ballard's original technique have been proposed. Should just one of these variants be adopted for the standard? If so, which one? Although the MPEG development team is aiming to submit all proposals to impartial evaluation, current methods for comparing the

retrieval effectiveness of different retrieval techniques are still relatively undeveloped, and it cannot be guaranteed that the "best" representations will always be selected. Another worry is that the MPEG policy of rigorously distinguishing between image representations and the feature extraction methods used to generate them cannot always be implemented in practice. For example, wavelet coefficients, another form of image representation under consideration for the standard, tend to be defined largely in terms of the algorithm used to calculate them. At this stage, no-one outside the MPEG-7 development team knows enough about the details of the new standard to judge whether these fears are justified.

Whatever its merits and drawbacks, it seems certain that MPEG-7 will dominate the development of the next generation of image retrieval systems. It has the support of an impressive range of academic and industrial backers, including systems designers such as IBM and Virage, and large-scale video users such as the BBC and INA, the French national film and TV archive. This will virtually guarantee its widespread adoption over the next few years, and few developers of image management software will be able to ignore it. The effects of the new standard are likely to be felt most rapidly in the research area. New experimental search engines based on MPEG-7 descriptors will probably start appearing before the end of 1999. Research into better techniques for extracting MPEG-7 preferred features will also be stimulated – though research into feature types *not* supported by MPEG-7 may well suffer, at least in the short term. Commercial vendors are more likely to wait until the standard has been finally agreed before launching new products on the market.

It is important to remember that MPEG-7 will not of itself offer any solutions to the problems of feature extraction and matching discussed in previous sections of this report. Issues such as how to identify the most appropriate similarity measures for image matching, and how best to combine evidence from different sources into a single matching score, will remain within the research domain well into the future. In particular, the semantic gap will remain as wide as ever. If MPEG-7 compliant systems want to use semantic retrieval cues, these will still have to be added manually.

7 Conclusions and recommendations

CBIR at present is still very much a research topic. The technology is exciting but immature, and few operational image archives have yet shown any serious interest in adoption. The crucial question that this report attempts to answer is whether CBIR will turn out to be a flash in the pan, or the wave of the future. Our view is that CBIR is here to stay. It is not as effective as some of its more ardent enthusiasts claim – but it is a lot better than many of its critics allow, and its capabilities are improving all the time. And as we argue in section 4.5 above, most current keyword-based image retrieval systems leave a great deal to be desired. In hard-nosed commercial terms, only one application of CBIR (video asset management) appears to be cost-effective – but few conventional image management systems could pass the test of commercial viability either.

The application areas most likely to benefit from the adoption of CBIR are those where level 1 techniques can be directly applied. Trademark image searching is an obvious example – while the technology of shape retrieval may not be perfect, it is already good enough to be useful in a commercial environment. Other areas where retrieval by primitive image feature is likely to be beneficial are crime prevention (including identification of shoe prints and tyre tracks as well as faces and fingerprints), architectural design (retrieval of similar previous designs and standard components) and medical diagnosis (retrieval of cases with similar features). One caveat here is that the storage, retrieval and use of images in some of these application areas (such as surveillance video monitoring)

can have significant privacy and civil liberties implications. Such implications need to be explored thoroughly before any decision on adoption is taken. It is also unlikely that general-purpose image retrieval software will meet the needs of these user communities without a significant degree of customization. Each of these application areas has their own range of special needs and constraints. Software solutions that fail to address these needs are unlikely to perform well enough to convince users that they are worth adopting.

Video asset management is an area that is already benefiting from CBIR technology, in the form of shot boundary detection and keyframe extraction. Even if manual annotation techniques are used for the remainder of the indexing process, considerable cost savings can be demonstrated. As indicated in section 5.6.6 above, such software is already in widespread use in the USA, and seems likely to be adopted by a number of UK organizations as well. The availability of more sophisticated products such as Islip's Mediakey Digital Video Library System, the commercial version of Carnegie-Mellon University's Informedia (section 5.5.2), is likely to increase the importance of CBIR in video asset management still further.

Whether more general image database users such as stock shot agencies, art galleries and museums can benefit from CBIR is still an open question. Clearly, there is no prospect of CBIR technology *replacing* more traditional methods of indexing and searching at this level in the near future. However, as discussed in section 6.2, there are strong indications that the combined use of text and image features might well yield better performance than either type of retrieval cue on its own. Similarly, the combined use of content-based retrieval and content-based navigation (section 6.3) promises to be a very powerful technique for identifying desired items of any type in multimedia systems. The problem at present with both approaches is that there is as yet no body of knowledge about how these different types of access method can best be combined. Developers and collection managers wishing to use combined techniques are thus inevitably working in the dark.

Similar considerations apply to the use of intermediaries. It has been traditional in image libraries for the custodian to perform much of the searching on behalf of users. This made excellent sense when such collections were small, and the librarian could recall the contents of most, if not all images in the collection from memory. The trend away from isolated collections and towards networked resources which can be accessed directly from users' own terminals inevitably throws the responsibility for devising an effective search strategy back on to the user. But it is questionable whether this is in fact the most effective approach. CBIR systems are not particularly easy for inexperienced end-users to understand. It is certainly not obvious to the casual user how to formulate and refine queries couched in terms of colour, texture or shape features. The use of relevance feedback can obviously help, but it is no panacea. Unless the set of retrieved images converges fairly quickly on what the user wants, disillusionment will set in quite quickly. There is thus an argument for the involvement of an experienced search intermediary who can translate a user's query into appropriate image primitives, and refine the search in consultation with the user in the light of output received. This kind of role is less fanciful than it might seem – it is simply trying to exploit the same approach as some researchers into semantic image retrieval (section 5.3). The only difference is that it would use humans instead of machines, and therefore probably has a higher chance of success. Such intermediaries would be difficult to train because the requisite body of knowledge exists only in fragmented form, if at all. But they could be enormously helpful in making CBIR systems usable.

For image database users such as graphic designers, the ability to retrieve specific images is of marginal usefulness. The role of images in stimulating creativity is little understood – images located by chance may be just as useful in providing the designer with inspiration as those retrieved in

response to specific queries. In these circumstances search intermediaries are likely to be of little use, and the often capricious performance of CBIR becomes an advantage. The ability of systems like QBIC to display sets of images with underlying features in common, even if superficially dissimilar, may be just what the designer needs, particularly if any retrieved image may be used to start a further search. Such *content-assisted browsing* might turn out to be a valuable, if unforeseen, application of CBIR. There is of course a risk that future improvements in CBIR technology, enabling more accurate searching, will erode its usefulness here!

Searching the Web for images is such a chaotic process that almost any advance on current technology is likely to be beneficial. Improved search engines, capable of using both text and image features for retrieval, will become commonplace within the next few years. Users may still need considerable stamina to find the images they want, particularly if relevance feedback techniques remain too computationally expensive to operate over the Web. A variety of specialized search engines are likely to appear on the Web, such as duplicate image detectors to seek out and report on unauthorized copies of copyright material, and possibly filters to detect and block pornographic images. Pornography filters based on current CBIR technology are not likely to be very effective, as this verges on a level 3 application.

Specific uses of CBIR technology in UK higher education are difficult to pinpoint. Lecturers in general certainly look for both still and moving images to illustrate their teaching material, though the variety of subjects and types of material involved is such that queries at all three levels are likely to be involved. As argued above, the best prospects for the effective use of CBIR techniques for still images in this context is as an adjunct to text descriptors. Video users could well benefit from the video asset management software discussed above. In certain specialized areas, however, the situation is rather different. The ability to retrieve and compare past diagnoses of similar-looking radiographs could provide valuable training for medical students. Retrieval of buildings with similar appearance to compare function, or to trace the development of architectural styles, could be useful for architecture students. Comparison of designs on Etruscan pottery could be of use to trainee archaeologists. The list of such applications may not be endless, but is quite lengthy. Few studies of the effectiveness of CBIR in these areas have been reported; more would be welcome.

The volume of research into improved techniques for CBIR is increasing every year. Even allowing for the fact that much of it is concerned with minor modifications to existing methods, there is enough genuine innovation in areas such as semantic image retrieval, cross-media content description, and user interaction for us to be confident that there will be significant advances in commercially-available CBIR technology over the next ten years. We cannot of course predict what these will be. It is therefore important that even those who are sceptical about the technology maintain a watching brief on developments in this area. CBIR may have nothing to offer some types of image user at present – though it is important to stress that the situation could change very rapidly if even a single large research project makes a major breakthrough.

Finally, our discussions with image managers and users in the UK did not reveal any great depth of understanding of the capabilities of CBIR technology among any of these communities. This suggests that few of them will be in any position to respond adequately to any future breakthrough in technology. This could lead to UK image collection managers, multimedia developers and other professional image users losing out on opportunities to exploit new technology to better-informed members of the same professions in the USA. To prevent this, a strong case can be made for a continued programme to raise awareness of CBIR techniques across the whole spectrum of image generators and users.

Our specific recommendations are as follows:

Users and managers of image collections need to be aware of the capabilities of CBIR technology, and to be capable of making informed decisions about adoption. Specifically:

?? Managers of image collections in specialist areas such as fingerprints or trademark images, involving image matching by appearance, should be encouraged to investigate possible adoption of CBIR technology in the near future.

?? Managers of video libraries should certainly investigate the possibility of using one of the proprietary video asset management packages described in section 5.6.6.

?? Managers of general-purpose image collections such as art galleries or photographic libraries should be encouraged to keep a watching brief on developments in CBIR, through articles in the specialist press and conferences relating to image retrieval.

Software developers or information providers with products designed to handle images, but which currently lack CBIR capabilities, also need to make informed decisions about whether CBIR would add value to their products. Specifically:

?? Organizations offering products or services in specialist areas such as fingerprints or trademark images, involving image matching by appearance, need to investigate the possibility of adding CBIR technology to their products in the near future. Their competitors may well be doing this already.

?? Providers of general-purpose multimedia software for managing collections such as art galleries or photographic libraries need to keep a watching brief on developments in CBIR, particularly relating to hybrid text/image feature indexing and cross-media retrieval.

UK government agencies can influence both the rate of future development of CBIR technology in the UK, and the level of awareness and adoption among influential users. Research funding bodies are in a position to stimulate technology development through the level of funding they are prepared to allocate to research in the field, and the type of projects selected for funding. The USA's current pre-eminence in the CBIR field owes a great deal to the generous level of funding for the subject from NSF and DARPA, two of their largest grant-giving agencies. NSF's Digital Libraries Program has had a particularly beneficial effect in stimulating the development of new image retrieval techniques and their application to the management of large image collections. Bodies such as the Library and Information Commission are in a position to assist the growth of awareness of CBIR among opinion-formers in the library profession. We therefore recommend that:

?? Funding agencies should consider declaring CBIR research a priority area, as has been done in the USA, and to a lesser extent in the European Community. The field appears to be generating interesting and valid results, even though it has so far led to few commercial applications. Topics particularly worth supporting, in the expectation that they will lead to useful results in the long term, include:

    ?? studies of image seeking behaviour and use,

    ?? evaluation of system effectiveness,

    ?? new approaches to semantic image retrieval,

    ?? techniques for cross-media indexing,

?? improved methods for interface design.

?? Agencies concerned with technology transfer or dissemination of best practice in fields which could potentially benefit from CBIR (including management of image collections and drawing archives, electronic publishing and multimedia content creation) should consider sponsoring programmes to raise awareness of CBIR technology among leading practitioners in these fields.

JISC can influence the take-up of CBIR within higher education and elsewhere, through pilot studies and dissemination activities. It is already committed to funding at least one series of pilot studies on the feasibility of using CBIR techniques within higher education. Our recommendations in this area are as follows:

?? Further pilot studies of CBIR should be undertaken, with a view to identifying the benefits and costs of CBIR technology in the context of higher education, and the specific types of user most likely to benefit.

?? Provided significant benefits can in fact be identified, a further programme to raise awareness of CBIR and its benefits both within and outside the higher education community should be undertaken.

?? Again, provided benefits from the use of CBIR can be demonstrated, large-scale trials of the effectiveness of different ways of delivering CBIR should be undertaken. These should include studies of the effectiveness of combined text/primitive feature indexing of still images, cross-media searching of video and other multimedia documents, and the role of search intermediaries in image retrieval.

?? Since the USA is likely to remain the main source of research activity in the CBIR field, every effort should be made to encourage further co-operation between the UK and USA in this area. In particular, JISC should continue to participate actively in the NSF International Digital Libraries Program.

?? One particularly useful area for international co-operation would be the provision of standard collections of images and image queries for comparing the effectiveness of different image retrieval techniques (similar to those created for the TREC text retrieval experiments). JISC should encourage efforts to create such collections.

Finally, ***all professionals involved in image data management*** need to be aware of standards development in the area, and be prepared to influence their future development if appropriate.

?? Image professionals of all kinds (including managers of image collections) should keep abreast of emerging standards, particularly the new MPEG-7 standard for multimedia content description, and contribute to their future development where necessary.

In conclusion, CBIR is clearly a technology with potential. The next five to ten years will reveal whether this potential can be turned into solid achievement. Our view is that at present the omens are favourable.

8 References

Aigrain, P et al (1996) "Content-based representation and retrieval of visual media – a state-of-the-art review" *Multimedia Tools and Applications* **3**(3), 179-202

Alsuth, P et al (1998) "On video retrieval: content analysis by ImageMiner" in *Storage and Retrieval for Image and Video Databases VI*, Proc SPIE 3312, 236-247

Androutsas, D et al (1998) "Image retrieval using directional detail histograms" in *Storage and Retrieval for Image and Video Databases VI*, Proc SPIE 3312, 129-137

Ardizzone, E and La Cascia, M (1997) "Automatic video database indexing and retrieval" *Multimedia Tools and Applications* **4**, 29-56

Armitage, L and Enser, P G B (1997) "Analysis of user need in image archives." *Journal of Information Science*, **23**(4), 287-299

Ashley, W (1996) "What shoe was that? The use of computerised image database to assist in identification." *Forensic Science International*, **82**, 7-20

Bach J R et al (1993) "A visual information management system for the interactive retrieval of faces" *IEEE Transactions on Knowledge and Data Engineering* **5**(4), 619-628

Beasley, M (1998) "What's cooking at the NHM?" *Vine*, No.**107**, 16-20

Beckmann, N (1990) "The R*-tree: an efficient and robust access method for points and rectangles" *ACM SIGMOD Record* **19**(2), 322-331

Beigi, M et al (1998) "MetaSEEk: a content-based meta-search engine for images" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 118-128

Bell, I (1998) "Images for the tourist: the experiences of the British Tourist Authority and English Tourist Board in setting up a digital image collection." *Vine*, No.**107**, 21-24

Berrut, C et al (1995) "Status review on non-text information retrieval" Report ELPUB106, European Commission, Brussels

Besser, H (1990) "Visual access to visual images: the UC Berkeley Image Database Project." *Library Trends*, **38**(4), 787-798

Biederman, I (1987) "Recognition-by-components: a theory of human image understanding" *Psychological Review* **94**(2), 115-147

del Bimbo, A et al (1996) "Image retrieval by elastic matching of shapes and image patterns" in *Proceedings of Multimedia '96*, 215-218

Bird, C et al (1996) "User interfaces for content-based image retrieval" in *Proceedings of IEE Colloquium on Intelligent Image Databases*, IEE, London, 8/1-8/4

Bjarnestam, A (1998) "Description of an image retrieval system", presented at *The Challenge of Image Retrieval* research workshop, Newcastle upon Tyne, 5 February 1998

Bolle, R M et al (1998) "Video query: research directions" *IBM Journal of Research and Development* **42**(2), 233-252

Borlund, P and Ingwersen, P (1997) "The development of a method for the evaluation of interactive information retrieval systems" *Journal of Documentation* **53**(3), 225-250

Bradley, S et al (1994) "Intelligent engineering component catalogs" in *Artificial intelligence in design '94*. (Gero, J S and Sudweeks, eds.) Kluwer, 641-658

Brooks, R A (1981) "Symbolic reasoning among 3-D models and 2-D images" *Artificial Intelligence* **17,** 285-348

Brown, P et al (1996) "The democratic indexing of images." *New Review of Hypermedia and Multimedia: Applications and Research*, **2**, 107-120

Campbell, N W et al (1997) "Interpreting Image Databases by Region Classification" *Pattern Recognition* **30**(4), 555-563

Carson C S et al (1997) "Region-based image querying" in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, 42-49

La Cascia, M et al (1998) "Combining textual and visual cues for content-based image retrieval on the world-wide web" *Proc IEEE Workshop on Content-based Access of Image and Video Libraries,* 24-28

Cawkell, A E (1993) "Indexing collections of electronic images: a review" British Library Research Review 15, British Library, London.

Chan, Y and Kung, S Y (1997) "A hierarchical algorithm for image retrieval by sketch" in *First IEEE Workshop on Multimedia Signal Processing*, 564-569

Chan, Y et al (1999) "Building systems to block pornography" to be presented at *CIR-99: the Challenge of Image Retrieval,* Newcastle upon Tyne, February 25-26, 1999

Chang, E Y et al (1998) "RIME, a replicated image detector for the World-Wide Web" in *Multimedia Storage and Archiving Systems III* (Kuo, C C J et al, eds), Proc SPIE 3527, 58-67

Chang N S and Fu K S (1981) "Picture query languages for pictorial data-base systems" *IEEE Computer* **14(**11) 23-33

Chang, S F (1995) ""Compressed domain techniques for image/video indexing and manipulation" *Proceedings of IEEE International Conference on Image Processing*, Washington, DC, 314-317

Chang, S F et al (1997) "VideoQ: an automated content based video search system using visual cues" *Proceedings of ACM Multimedia 1997*, Seattle, WA, 313-324

Chang, S F et al (1998) "Semantic visual templates: linking visual features to semantics" in *IEEE International Conference on Image Processing (ICIP'98)*, Chicago, Illinois 531-535

Chang, S K et al (1988) "An intelligent image database system" *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **14**(5), 681-688

Chang, S K and Jungert, E (1991) "Pictorial data management based upon the theory of symbolic projections" J*ournal of Visual Languages and Computing* **2**, 195-215

Chans, Y et al (1997) "A feature-based approach for image retrieval by sketch" in *Multimedia Storage and Archiving Systems II* (Kuo, C C J et al, eds), Proc SPIE 3229, 220-231

Chen, J L and Stockman, C C (1996) "Indexing to 3D model aspects using 2D contour features" in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco,* 913-920

Chock, M et al (1984) "Database structure and manipulation capabilities of the picture database management system PICDMS" *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6**(4), 484-492

Christel, M G et al (1997) "Multimedia abstractions for a digital video library" in *ACM Digital Libraries '97* (Allen, R B and Rasmussen, E , eds) 21-29. ACM, New York

Cleverdon, C W et al (1966) "Factors determining the performance of indexing systems" Cranfield College of Aeronautics, Cranfield, UK.

Connell, J H and Brady, J M (1987) "Generating and generalizing models of visual objects" *Artificial Intelligence* **31**(2) 159-183

Constantopulos, P and Doerr, M (1995) "An approach to indexing annotated images" in *Multimedia Computing and Museums: selected papers from the 3rd International conference on Hypermedia and Interactivity in Museums (ICHIM'95 / MCN'95),* San Diego, California, October 9-13, 1995. (Bearman, D, ed) Archives and Museum Informatics, 278-298

Corridoni, J M et al (1998) "Image retrieval by color semantics with incomplete knowledge" *Journal of the American Society for Information Science* **49**(3), 267-282

Date, C J and Darwen H (1992) "A guide to the SQL standard" 3rd edition, Addison-Wesley, Reading, MA

Deerwater, S et al (1990) "Indexing by latent semantic analysis" *Journal of the American Society for Information Science* **41**(6), 391-407

Dickinson S et al (1998) "Viewpoint-invariant indexing for content-based image retrieval" in *IEEE International Workshop on Content-based Access of Image and Video Databases (CAIVD'98)*, Bombay, India, 20-30

Dobie, M et al (1999) "A flexible architecture for content and concept-based multimedia information exploration" to be presented at *CIR-99: the Challenge of Image Retrieval*, Newcastle upon Tyne, February 25-26, 1999

Dyson, M C (1992) "How do you describe a symbol? The problems involved in retrieving symbols from a database." *Information Services and Use*, **12**, 65-76

Dyson, M C and Box, H (1997) "Retrieving symbols from a database by their graphic characteristics: are users consistent?" *Journal of Visual Languages and Computing*, **8**(1), 85-107

Eakins J P (1993) "Design criteria for a shape retrieval system" *Computers in Industry* **21**, 167-184

Eakins, J P (1996) "Automatic image content retrieval – are we getting anywhere?" *Proceedings of Third International Conference on Electronic Library and Visual Information Research (ELVIRA3)*, De Montfort University, Milton Keynes, pp 123-135

Eakins J P (1998) "Techniques for image retrieval" *Library and Information Briefings*, in press

Eakins J P, Graham M E and Boardman J M (1997) "Evaluation of a trademark retrieval system", in 19th BCS IRSG Research Colloquium on Information Retrieval, Robert Gordon University, Aberdeen, *electronic* Workshops in Computing, http://www.springer.co.uk/ewic/workshops/IRR97/. Springer-Verlag, Berlin

Eakins J P, Boardman J M and Graham M E (1998) "Similarity retrieval of trade mark images" *IEEE Multimedia*, **5**(2), 53-63

Ellis, D (1996) "The dilemma of measurement in information retrieval research" *Journal of the American Society for Information Science* **47**, 23-36

Enser P G B (1995) "Pictorial information retrieval" *Journal of Documentation*, **51**(2), 126-170

Enser, P G B and McGregor, C G (1992) "*Analysis of visual information retrieval queries*." British Library. (British Library Research and Development Report, 6104)

Evans, A (1987) "TELCLASS: a structural approach to TV classification." *Audiovisual Librarian*, **13**(4), 215-216

Evans, B (1996) "Imaging the future." *The Architect's Journal*, **204** , 46-47

Fallon, K K (1998) "Look for higher technology benefits." *Architectural Record*, December, 139-142

Faloutsos, C et al (1994) "Efficient and effective querying by image content" *Journal of Intelligent Information Systems* **3**, 231-262

Feder, J (1996) "Towards image content-based retrieval for the World-Wide Web" *Advanced Imaging* **11**(1), 26-29

Flickner, M et al (1995) "Query by image and video content: the QBIC system" *IEEE Computer* **28**(9), 23-32

Forsyth, D A et al (1997) "Finding pictures of objects in large collections of images" in *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing* (Heidorn, P B and Sandore, B, eds), 118-139. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Franklin, A (1998) "Image indexing in the Bodleian ballads project." *Vine*, No.**107**, 51-57

Furner, J (1998) "Digital images in libraries: an overview." *Vine*, No.**107**, 3-7

Gordon, C (1990) "An introduction to ICONCLASS" in *Terminology for Museums, Proceedings of an International Conference, Cambridge, 1988* (Roberts, D A, ed), 233-244. Museum Documentation Association

Greenberg, J (1993) "Intellectual control of visual archives: a comparison between the Art and Architecture Thesaurus and the Library of Congress Thesaurus for Graphic Materials." *Cataloging & Classification Quarterly*, **16**(1), 85-101

Gudivada, V N and Jung, G S (1996) "An architecture for and query processing in distributed content-based image retrieval" *Real-Time Imaging* **2**, 139-152

Gudivada V N and Raghavan V V (1995a) "Content-based image retrieval systems" *IEEE Computer* **28**(9), 18-22

Gudivada V N and Raghavan V V (1995b) "Design and evaluation of algorithms for image retrieval by spatial similarity" *ACM Transactions on Information Systems* **13**(2), 115-144

Gupta, A et al (1996) "The Virage image search engine: an open framework for image management" in *Storage and Retrieval for Image and Video Databases IV*, Proc SPIE 2670, pp 76-87

Haering, N et al (1997) "Locating deciduous trees" in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, June 1997, 18-25

Harter, S P and Hert, C A (1997) "Evaluation of information retrieval systems: approaches, issues and methods" *Annual Review of Information Science and Technology* **32**, 3 - 94

Hastings, S K (1995) "Query categories in a study of intellectual access to digitized art images." *ASIS '95: proceedings of the 58th ASIS Annual Meeting*, vol **32**, 1995, 3-8

Hermes, T et al (1995) "Image retrieval for information systems" in *Storage and Retrieval for Image and Video Databases III* (Niblack, W R and Jain, R C, eds), Proc SPIE 2420, 394-405

Hill, G J et al (1993) "Open and reconfigurable hypermedia systems: a filter-based model" *Hypermedia* **5**(2), 103-118

Hirata, K and Kato, T (1992) "Query by visual example – content-based image retrieval" in *EDBT'92, Third International Conference on Extending Database Technology*, 56-71

Hogan M, et al (1991) "The visual thesaurus in a hypermedia environment" *Proc. International Conference on Hypermedia and Interactivity in Museums, Pittsburgh, USA*, 202-221

Holt, B and Hartwick, L (1994) "Retrieving art images by image content: the UC Davis QBIC project" Proc First International Conference on Electronic Library and Visual Information System Research, de Montfort University, Milton Keynes, 93-100

Horikoshi, T and Kasahara, H (1990) "3-D shape indexing language" in *Ninth Annual Phoenix Conference on Computers and Communications*, Los Alamitos, CA, 493-499

Hou, Y T et al (1992) "A content-based indexing technique using relative geometry features" in *Image Storage and Retrieval Systems*, Proc SPIE 1662, 59-68

Huang, T et al (1997) "Multimedia Analysis and Retrieval System (MARS) project" in *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing* (Heidorn, P B and Sandore, B, eds), 101-117. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Idris F and Panchanathan, S (1997a) "Review of image and video indexing techniques" *Journal of Visual Communication and Image Representation* **8**(2) 146-166

Idris F and Panchanathan, S (1997b) "Storage and retrieval of compressed images using wavelet vector quantization" *Journal of Visual Languages and Computing* **8**, 289-310

IEEE (1985) "Computer graphics in the detective business" *IEEE Computer Graphics and Applications* **5**(4), 14-17

Ingwersen, P (1996) "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory" *Journal of Documentation* **52**(1), 3-50

Jacobs, C. E. Et al (1995) "Fast Multiresolution Image Querying" *Proceedings of SIGGRAPH 95, Los Angeles, CA* (ACM SIGGRAPH Annual Conference Series, 1995), 277-286.

Jain, A K and Vailaya, A (1996) "Image retrieval using color and shape" *Pattern Recognition* **29**(8), 1233-1244

Jain, A K et al (1997) "Multimedia systems for art and culture: a case study of Brihadisvara Temple" in *Storage and Retrieval for Image and Video Databases V* (Sethi, I K and Jain, R C, eds), Proc SPIE 3022, 249-261

Jain, R (1993) "Workshop report: NSF Workshop on Visual Information Management Systems" in *Storage and Retrieval for Image and Video Databases* (Niblack, W R and Jain, R C, eds), Proc SPIE 1908, 198-218

Jain, R (1995) "World-wide maze" *IEEE Multimedia* **2**(3), 3

Jin, J S et al (1998) "Using browsing to improve content-based image retrieval" in *Multimedia Storage and Archiving Systems III*, Proc SPIE 3527, 101-109

Jose, J et al (1998) "Spatial querying for image retrieval: a user-oriented evaluation" presented at *21st ACM-SIGIR International Conference on Research & Development in Information Retrieval*, Melbourne, Australia

Kaplan, L M et al (1998) "Fast texture database retrieval using extended fractal features" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 162-173

Kato T (1992) "Database architecture for content-based image retrieval" in *Image Storage and Retrieval Systems* (Jambardino, A A and Niblack, W R , eds), Proc SPIE 1662, 112-123

Kato T and Kurita T (1990) "Visual interaction with the Electronic Art Gallery" in *Database and Expert Systems Applications: Proceedings on an International Conference*, 234-240

Kay, D C and Levine, J R (1992) *Graphics File Formats.* McGraw-Hill, NY.

Keister, L H (1994) "User types and queries: impact on image access systems." In *Challenges in indexing electronic text and images* (Fidel, R et al., eds). ASIS, 1994, 7-22

Kimia, B B et al (1997) "A shock-based approach for indexing of image databases using shape" in *Multimedia Storage and Archiving Systems II* (Kuo, C C J et al, eds), Proc SPIE 3229, 288-302

Kitamoto, A et al (1993) "Similarity retrieval of NOAA satellite imagery by graph matching" in *Storage and Retrieval for Image and Video Databases* (Niblack, W R and Jain, R C, eds), Proc SPIE 1908, 60-73

Krause M G (1988) "Intellectual problems of indexing picture collections" *Audiovisual Librarian*, **14**(2), 73-81

Kurniawati, R et al (1997) "The SS+ tree: an improved index structure for similarity searches in high-dimensional feature space" in *Storage and Retrieval for Image and Video Databases V* (Sethi, I K and Jain, R C, eds), Proc SPIE 3022, 110-120

Lancaster, F W (1998) *Indexing and abstracting in theory and practice*. 2nd ed. Library Association, London

Lee, D et al (1994) "Query by image content using multiple objects and multiple features: user interface issues" in *Proceedings of ICIP-94, International Conference on Image Processing, Austin, Texas*, 76-80

Lewis, P H et al (1996) "Media-based navigation with generic links" in *Proceedings of the Seventh ACM Conference on Hypertext*, New York, 215-223

Lewis, P H et al (1997) "Towards multimedia thesaurus support for media-based navigation" in *Image Databases and Multimedia Search*, (Smeulders, A W M and Jain, R C, eds), 111-118. World Scientific, Amsterdam

Li C S et al (1998) "S-STIR: Similarity Search Through Iterative Refinement" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 250-258

Liang, K C and Kuo, C C J (1998) "Implementation and performance evaluation of a progressive image retrieval system" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 37-48

Lin, K I et al (1994) "The TV-tree: an index structure for high dimensional data" *Journal of Very Large Databases* **3**(4), 517-549

Liu, F and Picard, R W (1996) "Periodicity, directionality and randomness: Wold features for image modelling and retrieval" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(7), 722-733

Liu Y et al (1998) "Content-based 3-D neuroradiologic image retrieval: preliminary results" in *IEEE International Workshop on Content-based Access of Image and Video Databases (CAIVD'98)*, Bombay, India, 91-100

Ma W Y and Manjunath, B S (1997) "Netra: a toolbox for navigating large image databases" *Proc IEEE International Conference on Image Processing (ICIP97),* **1,** 568-571

Ma W Y and Manjunath, B S (1998) "A texture thesaurus for browsing large aerial photographs" *Journal of the American Society for Information Science* **49**(7), 633-648

Manjunath, B S and Ma, W Y (1996) "Texture features for browsing and retrieval of large image data" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 837-842

Manmatha, R and Ravela, S (1997) "A syntactic characterization of appearance and its application to image retrieval" in *Human Vision and Electronic Imaging II* (Rogowitz BE and Pappas TN, eds), SPIE 3016, 484-495

Markey, K (1988) "Access to iconographical research collections." *Library Trends*, **37**(2), 154-174

Markkula, M and Sormunen, E (1998) "Searching for photos – journalists' practices in pictorial IR", presented at *The Challenge of Image Retrieval* research workshop, Newcastle upon Tyne, 5 February 1998

de Marsicoi, M et al (1997) "Indexing pictorial documents by their content: a survey of current techniques" *Image and Vision Computing* **15**, 119-141

McCorry, H and Morrison, I O (1995) "*Report on the Catechism project.*" National Museums of Scotland

Mehrotra, R and Gary, J E (1995) "Similar-shape retrieval in shape data management" *IEEE Computer* **28**(9), 57-62

Minka, T (1996) "An image database browser that learns from user interaction" MIT Media Laboratory Technical Report #365

Nastar, C et al (1998) "Surfimage: a flexible content-based image retrieval system" presented at *ACM Multimedia '98*, Bristol, UK

Niblack, W et al (1993) "The QBIC project: querying images by color, texture and shape" IBM Research Report RJ-9203

Niblack, W et al (1998) "Updates to the QBIC system" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 150-161

Ogle, V E and Stonebraker, M (1995) "Chabot: retrieval from a relational database of images" *IEEE Computer* **28**(9), 40-48

Okon, C (1998) "Mainstreaming face recognition apps" *Advanced Imaging* **13**(4), 16-18

Oliva, A et al (1997) "Real-world scene categorization by a self-organizing neural network" *Perception*, supp **26**, 19

Opitz, H et al (1969) "Workpiece classification and its industrial application" *International Journal of Machine Tool Design Research* **9**, 39-50

Ornager, S (1997) Image retrieval: theoretical and empirical user studies on accessing information in images." In *ASIS '97: proceedings of the 60th ASIS Annual Meeting*, vol. **34**, 202-211

Pentland A et al (1996) "Photobook: tools for content-based manipulation of image databases" *International Journal of Computer Vision* **18**(3), 233-254

Petkovic, D (1996) "Query by Image Content", presented at *Storage and Retrieval for Image and Video Databases IV*, (Sethi, I K and Jain, R C, eds), Proc SPIE 2670

Picard, R W (1996) "A society of models for video and image libraries" *IBM Systems Journal* **35**, 292-312

Price, R et al (1992) "Applying relevance feedback to a photo archival system" *Journal of Information Science* **18**(3), 203-215

Rabbitti, F and Stanchev, P (1989) "GRIM_DBMS: a graphical image database management system" in *Visual Database Systems* (Kunii, T, ed), Elsevier, Amsterdam, 415-430.

Rasmussen, E M (1997) "Indexing images" *Annual Review of Information Science and Technology*, **32**, 169-196

Ratan, A L and Grimson, W E L (1997) "Training templates for scene classification using a few examples" in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, June 1997, 90-97

Ravela, S and Manmatha, R (1998a) "Retrieving images by appearance" in *Proceedings of IEEE International Conference on Computer Vision (IICV98), Bombay, India*, 608-613

Ravela, S and Manmatha, R (1998b) "On computing global similarity in images" in *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV98), Princeton, NJ*, 82-87

Roberston, S E and Sparck Jones, K (1976) "Relevance weighting of search terms" *Journal of the American Society for Information Science* **27**, 129-46

Ross, S S (1998) "Entering the third dimension." *Architectural Record*, June, 175-180

Roussopoulos, N et al (1988) "An efficient pictorial database system for PSQL" *IEEE Transactions on Software Engineering*, **14**(5), 639-650

Rui, Y et al (1997) "Relevance feedback techniques in interactive content-based image retrieval" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 25-36

Salton, G (1971) *The SMART retrieval system – experiments in automatic document processing*. Prentice-Hall, New Jersey

Santini, S and Jain, R C (1997) "The graphical specification of similarity queries" *Journal of Visual Languages and Computing* **7**, 403-421

Santini, S and Jain, R C (in press) "Measures of image similarity" *IEEE Transactions on pattern Analysis and Machine Intelligence*, in press

Saur, D D et al (1997) "Automated analysis and annotation of basketball video" in *Storage and Retrieval for Image and Video Databases V* (Sethi, I K and Jain, R C, eds), Proc SPIE 3022, 176-187

Scassellati, B et al (1994) "Retrieving images by 2-D shape: a comparison of computation methods with human perceptual judgements" in *Storage and Retrieval for Image and Video Databases II* (Niblack, W R and Jain, R C, eds), Proc SPIE 2185, 2-14

Schonfield, D and Lelescu, D (1998) "VORTEX: video retrieval and tracking from compressed multimedia databases – template matching from MPEG-2 video compression standard" in *Multimedia Storage and Archiving Systems III* (Kuo, C C J et al, eds), Proc SPIE 3527, 233-244

Sclaroff S et al (1997) "ImageRover: a content-based image browser for the World-Wide Web" in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, June 1997, 2-9

Seloff, G A (1990) "Automated access to NASA-JSC image archives." *Library Trends*, **38**(4), 682-696

Shatford, S (1986) "Analyzing the subject of a picture: a theoretical approach." *Cataloging and Classification Quarterly*, **6**(3), 39-62

Shatford Layne, S (1994) "Some issues in the indexing of images." *Journal of the Amercan Society of Information Science*, **45**(8), 583-588

Shum, H Y et al (1996) "On 3D shape similarity" in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco*, 526-531

Sledge, J (1995) "Points of view." in *Multimedia Computing and Museums: selected papers from the 3rd International conference on Hypermedia and Interactivity in Museums (ICHIM'95 / MCN'95)*, San Diego, California, October 9-13, 1995. (Bearman, D, ed) Archives and Museum Informatics, 335-346

Smith J R and Chang S F (1997a) "Querying by color regions using the VisualSEEk content-based visual query system" *Intelligent Multimedia Information Retrieval* (Maybury, M T, ed). AAAI Press, Menlo Park, CA, 23-41.

Smith J R and Chang S F (1997b) "An image and video search engine for the World-Wide Web" in *Storage and Retrieval for Image and Video Databases V* (Sethi, I K and Jain, R C, eds), Proc SPIE 3022, 84-95.

Smith, M A and Kanade, T (1998) "Video skimming and characterization though combination of image and language understanding" " in *IEEE International Workshop on Content-based Access of Image and Video Databases (CAIVD'98)*, Bombay, India, 61-71

Soffer, A and Samet, H (1996) "Retrieval by content in symbolic image databases" in *Storage and Retrieval for Image and Video Databases IV*, (Sethi, I K and Jain, R C, eds), Proc SPIE 2670, 144-155)

Sparck Jones, K (1995) "Reflections on TREC" *Information Processing and Management* **31**(3) 291-314

Spink, A et al (1998) "Elicitation behaviour during mediated information retrieval." *Information Processing and Management*, **34**(2/3), 257-273

Srihari, R K (1995) "Automatic indexing and content-based retrieval of captioned images" *IEEE Computer* **28**(9), 49-56

van der Starre, J H E (1995) "Ceci n'est pas une pipe: indexing of images" in *Multimedia Computing and Museums: selected papers from the 3rd International conference on Hypermedia and Interactivity in Museums (ICHIM'95 / MCN'95)*, San Diego, California, October 9-13, 1995. (Bearman, D, ed) Archives and Museum Informatics, 267-277

Strat, T M and Fischler, M A (1991) "Context-based vision: recognizing objects using information from both 2-D and 3-D imagery" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(10), 1050-1065

Stricker, M and Dimai, A (1996) "Color indexing with weak spatial constraints" in *Storage and Retrieval for Image and Video Databases IV*, (Sethi, I K and Jain, R C, eds), Proc SPIE 2670, 29-40

Stricker, M and Orengo, M (1995) "Similarity of color images" in *Storage and Retrieval for Image and Video Databases III* (Niblack, W R and Jain, R C, eds), Proc SPIE 2420, pp 381-392

Strother-Vien, F (1998) "Mugshot recognition meets witness composite sketches in LA" *Advanced Imaging* **13**(1), 22

Sudhir, G et al (1998) "Automatic classification of tennis video for high-level CBR" in *IEEE International Workshop on Content-based Access of Image and Video Databases (CAIVD'98)*, Bombay, India, 81-90

Sutcliffe, A et al (1997) "Empirical studies in multimedia information retrieval" in *Intelligent Multimedia Information Retrieval* (Maybury, M T, ed). AAAI Press, Menlo Park, CA.

Svenonius, E (1994) "Access to nonbook materials: the limits of subject indexing for visual and aural languages" *Journal of the American Society of Information Science*, **45**(8), 600-606

Swain, M J and Ballard, D H (1991) "Color indexing" *International Journal of Computer Vision* **7**(1), 11-32

Tamura, H et al (1978) "Textural features corresponding to visual perception" *IEEE Transactions on Systems, Man and Cybernetics* **8**(6), 460-472

Tirthapura, S et al (1998) "Indexing based on edit-distance matching of shape graphs" in *Multimedia Storage and Archiving Systems III* (Kuo, C C J et al, eds), Proc SPIE 3527, 25-36

Turtle, H R and Croft, W B (1991) "Efficient probabilistic inference for text retrieval" in *Proceedings of RIAO 3,* Barcelona, Spain, 644-661

Vellaikal, A and Kuo, C C J (1998) "Hierarchical clustering techniques for image database organization and summarization" in *Multimedia Storage and Archiving Systems III*, Proc SPIE 3527, 68-79

Wactlar, H D et al (1996) "Intelligent access to digital video: the Informedia project" *IEEE Computer* **29**(5), 46-52

Wallace, G K (1991) "The JPEG still picture compression standard" *Communications of the ACM* **34**(4), 31-45

Wan X and Kuo, C C J (1996) "Image retrieval based on JPEG compressed data" in *Multimedia Storage and Archiving Systems* (Kuo, C C J et al, eds*)* Proc SPIE 2916, 104-115

Wilf, I (1998) "Computer, retrieve for me the video clip of the winning goal" *Advanced Imaging* **15**(8), 53-55

World Intellectual Property Organization (1998) *InternationalClassification of the Figurative Elements of Marks (Vienna Classification), Fourth Edition.* ISBN 92-805-0728-1. World Intellectual Property Organization, Geneva

Wu, J K and Narasimhalu, A D (1994) "Identifying faces using multiple retrievals." *IEEE Multimedia*, **1**(2), 27-38

Wu J K et al (1996) "Content-based retrieval for trademark registration" *Multimedia Tools and Applications* **3**, 245-267

Yang, D S et al (1994) "An intelligent symbol usage assistant for CAD systems" *IEEE Expert* **10**(3), 32-41

Yeo, B L and Yeung, MM (1998) "Classification, simplification, and dynamic visualization of scene transition graphs for video browsing" in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), Proc SPIE 3312, 60-70

Zhang, H J et al (1995) "Video parsing and browsing using compressed data" *Multimedia Tools and Applications* **1,** 89-111

van der Zwan, R et al (1999) "The Informedia digital video library system at the Open University" to be presented at *CIR-99: the Challenge of Image Retrieval*, Newcastle upon Tyne, February 25-26, 1999