Retrieval of Still Images by Content

John P. Eakins

Institute for Image Data Research, University of Northumbria at Newcastle
Newcastle upon Tyne NE1 8ST, U.K.
john.eakins@unn.ac.uk

Abstract This chapter summarises the current state of the art in content based image retrieval (CBIR). It discusses the need for image retrieval by content, and the types of query which might be encountered. It describes the main techniques currently used to retrieve images by content at both primitive and semantic levels, describes the features of some commercial and experimental CBIR systems, assesses the capabilities of current technology, and outlines possible future developments the field.

1 Introduction

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. Users are increasingly discovering that the process of locating a desired image in a large and varied collection can be a source of considerable frustration. Traditional methods of image indexing based on classification schemes or keywords have severe limitations [18]. This has led to the development of automatic techniques for retrieving images on the basis of features extracted from those images themselves – a technology now generally referred to as *Content-Based Image Retrieval* (CBIR). CBIR is an exciting field for research, but as yet has delivered few operational systems capable of meeting real user needs.

1.1 User Needs

Many different groups of users make use of images in a professional capacity. The police use visual information such as photographs to identify individuals and to record scenes of crimes. The use of fingerprints and shoeprints to identify criminals is widespread. In medicine, visual information in the form of X-rays, ultrasound or other types of imaging is routinely used in diagnosis and in monitoring patients' progress. Fashion and graphic designers gain inspiration from images of previous designs, and use sketches and 3-D models to present ideas to clients and colleagues. Photographs and pictures are used extensively in

M. Agosti, F. Crestani, and G. Pasi (Eds.): ESSIR 2000, LNCS 1980, pp. 111–138, 2000.

the publishing world, to illustrate books and magazine articles. Most newspaper publishers maintain their own libraries of photographs, supplementing these where necessary with those from outside sources such as stock photo agencies. Computer-generated images are extensively used in architectural and engineering design, both to specify requirements to those building or manufacturing the design, and to illustrate the end-product to potential customers. Many manufacturing firms maintain design archives of standard components for reuse. And historians from a variety of disciplines use images extensively in their research. Nearly all these professions require access to images from archives at some point, often (though not always) by content rather than just by identifier.

Before discussing techniques for image retrieval, it is first necessary to understand the types of query such users might to put to an image database. Several researchers have addressed this question, though no clear consensus on user needs has yet emerged. For example, Markkula and Sormumen [57] 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. Enser and McGregor [19] 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) and those which could not (e.g. classroom scenes). Both categories were subject to refinement in terms of time, location, or action. Hastings [31], investigating how art historians searched photographic and digital art images, found the major classes of queries to be identification, subject, text, style, artist, category, and colour. Keister [41], reviewing queries put to an automated still picture retrieval system at the National Library of Medicine (NLM), found wide variations in the way users asked for pictures. Users who were picture professionals thought visually and used art and/or graphics jargon. Health professionals asked for images relating to specific diseases or treatments. The museum or academic community often had precise citations to the images it desired. Words describing concrete image elements appeared to make up a significant proportion of requests.

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 search intermediaries. This seeming emphasis on the expressed need tells us little about what the actual need is for the images, or 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. Despite attempts to develop a more general framework for understanding image searching and use (e.g. [41]), we still know too little about the information needs of different types of image user to draw any firm conclusions for retrieval system design.

1.2 Characteristics of Image Queries

As indicated above, insufficient evidence is yet available to categorize in any depth the types of query users might to put to an image database. In the meantime, it has been found useful to classify image queries into three levels of increasing complexity [14]:

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 [18].

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:

(1) retrieval of named events or types of activity (e.g. "find pictures of Scottish folk dancing");
(2) 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. [27]) refer to levels 2 and 3 together as *semantic* image retrieval, and hence the gap between levels 1 and 2 as the *semantic gap*.

2 Traditional Methods of Image Data Management

2.1 Classification and Indexing Techniques

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.

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 [4]. Their thesaurus comprises just over 10,000 keywords, divided into nine semantic groups, including *geography*, *people*, *activities* and *concepts*.

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 consists of nearly 120,000 terms for describing objects, textural materials, images, architecture and other cultural heritage material. The terms are arranged into hierarchies covering concepts such as physical attributes, styles and periods, and materials. Another popular source for providing subject access to visual material is the Library of Congress Thesaurus for Graphic Materials (LCTGM). See Greenberg [26] for a 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 from the University of Leiden [25], and TELCLASS from the BBC [20]. Like AAT, ICONCLASS was designed for the classification of works of art, and to some extent duplicates its function; TELCLASS was designed with TV and video programmes in mind, and is hence rather more general in its outlook.

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 [91], used by registries worldwide to identify potentially conflicting trademark applications, and the Opitz coding system for machined

parts [64], used to identify families of similar parts which can be manufactured together.

2.2 Effectiveness of Manual Techniques

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 for still images range from about 7 to 40 minutes per image [17]. Secondly, it does not appear to be particularly reliable as a means of subject retrieval of images. Markey [56] found that, in a review of inter-indexer consistency, there were wide disparities in the keywords that different individuals assigned to the same picture. Enser and McGregor [19] found a poor match between the wording of user queries and one of the indexing languages in place in the Hulton Deutsch Collection, even though it had been specially designed for the collection. There is little or no firm evidence at present that text-based techniques for image retrieval are adequate for the task in hand.

3 Content-Based Image Retrieval (CBIR)

3.1 Introduction

The limitations of the text-based approach described above have led to an upsurge of interest in CBIR, now an extremely active area for research and development. Most CBIR techniques are based on principles which are markedly different from those used in text retrieval. Features considered to capture essential aspects of image content are extracted automatically from all images in the collection. All subsequent retrieval is based on these features. More formally, feature matching involves calculating and storing a *feature vector* characterising selected aspects of the appearance of each image in the database, and then calculating the similarity between the feature vector computed from the query with that of each image in the database, using some measure such as Euclidean distance $L2 = ||\mathbf{v_i} - \mathbf{v_j}||$, where $\mathbf{v_i}$ and $\mathbf{v_j}$ represent the feature vectors of images i and j.

The commonest features used are mathematical measures of image appearance, such as colour, texture or shape; hence virtually all current CBIR systems, whether commercial or experimental, operate at level 1. A typical CBIR system allows users to formulate queries by submitting an example of the type of image being sought, though some offer alternatives such as selection of a desired colour from a palette, or input of a rough sketch of a desired shape. 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. Some of the more commonly used techniques used for image retrieval are described below.

3.2 Retrieval by Colour

The ability to retrieve images on the basis of colour similarity is intuitively quite appealing, so it is no surprise that considerable effort has been devoted to research in this area. Colour queries can be formulated either by choosing from a palette of possible colour combinations, or by submitting an example image which is then colour matched with those in the database. Most techniques for colour retrieval 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 [83]. Variants of this technique are now used in a high proportion of current CBIR systems (see Section 5 below). Formally, a colour histogram $H(I)$ of an image I is a vector $(h_1, h_2, \ldots, h_j, \ldots, h_n)$, where each element represents the count of pixels falling within partition j of some suitable colour space, such as RGB or HSV. The similarity of two histograms A and B is then given by their intersection, defined as:

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

Swain and Ballard used relatively fine histograms, partitioning the three axes rg, by and wb of opponent colour space into 16, 8 and 8 bins respectively – a total of 2048. Later workers have tended to use somewhat coarser histograms, with apparently satisfactory results. Methods of improving on Swain and Ballard's original technique include the use of cumulative colour histograms $(g_1, g_2, \ldots, g_j, \ldots, g_n)$, where

$$g_j = \sum_{i=1}^{j} h_i$$

and colour moments

$$E_n = \frac{1}{M}\sum_{m=1}^{M} p_{mn},$$

$$\sigma_n = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(p_{mn} - E_n)^2} \quad \text{and}$$

$$s_n = \sqrt[3]{\frac{1}{M}\sum_{m=1}^{M}(p_{mn} - E_n)^3}$$

and representing the distribution of image pixels within each colour channel n [82]. Experiments suggested that colour moments based on HSV colour space could give particularly good results.

Colour matching of images can be applied either at the whole image or region level. A good example of the latter approach is that of Stricker and Dimai [81], who divide each image into five fuzzy regions and then compare colour moments from each of these regions. Other researchers base colour matching on automatically-segmented image regions, including Smith and Chang [77], who use *colour sets* (essentially colour histograms containing binary values) to provide rapid colour indexing of individual image regions. Corridoni et al [10] go further, using theories of human colour perception to formulate a query language which allows users to search on subjective attributes such as colour warmth or contrast as well as objective colour combinations.

3.3 Retrieval by Texture

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). Techniques developed for texture retrieval have often proved useful in matching more general aspects of an image's appearance. Texture queries can be formulated in a similar manner to colour queries, either by selecting examples of desired textures from a palette, or by supplying an example query image. A variety of techniques has been used for measuring texture similarity; the best-established rely on comparing values of 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 which can be used to compare image similarity. Well-established measures include the set defined by Tamura et al [85], which includes:

$$\text{coarseness} \quad \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} S_{\max}(i,j)$$

where m and n define image size, and $S_{\max}$ the neighbourhood size giving greatest separation of average intensity either side of any given pixel,

$$\text{directionality} \quad 1 - rn_p \sum_{p=1}^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi)$$

where n_p is the number of peaks and ϕ_p the direction of the pth peak in the density gradient histogram H_D, and

$$\text{contrast} \quad \sum_n n^2 \Bigg(\sum_i \sum_j p(i,j) : |i-j| = n \Bigg)$$

where $p(i,j)$ is the (i,j)th entry of the $n \times n$ spatial dependence matrix defined by HA number of measures are derived from pixel intensity transformations. One of the most effective methods of texture analysis for retrieval is the use coefficients derived from image transformations using Gabor filters [54].

$$G(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) e^{\left(2\pi i W x - \frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right)}$$

A bank of Gabor filters can be generated by scaling and rotating this function to different degrees, effectively yielding a set of orientation and scale-dependent edge and line detectors. The mean and standard deviations of filter outputs have been shown to give good discrimination between different kinds of texture in an image. A recent extension of this technique is the texture thesaurus developed by Ma and Manjunath [53], which retrieves textured regions in images on the basis of similarity to automatically-derived codewords representing important classes of texture within the collection.

Another popular approach to texture analysis and classification is the use of the wavelet transformation (see Section 3.5 below), which has been used successfully to characterize image texture (e.g. [76]). Yet another approach is the use of the Wold decomposition [51], which has been applied to identify features characterized as *periodicity*, *directionality* and *randomness*, for use in matching images by texture similarity.

3.4 Retrieval by Shape

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 [3]. As well as the ability to match human similarity judgements, the ideal shape matching technique needs to fulfil several other criteria, such as robustness to noise or small deformations in an image, and (for most applications) invariance to translation, rotation and scaling.

A wide variety of techniques meeting at least some of these criteria has been described in the literature. One important class of methods is based on direct

matching of complete (information-preserving) representations of object shape, such as chain-codes or splines. Such methods can have high discriminating power, at least when matching highly similar shapes, but are often computationally very expensive. A second class of methods is based on the extraction and comparison of features such as edge direction histograms or moment invariants, which may capture important aspects of an object's appearance, but which cannot be used to reconstitute its entire shape. These methods often have lower discriminating power, but tend to scale up better to large image collections. Techniques which involve direct matching of information-preserving representations of shape boundaries include:

- **String-matching of chains of boundary pixels.** Cortelazzo et al [11] suggest a number of ways of measuring the distance between two shape boundaries represented as pixel chains, based on summation of substring differences or string rewriting rules. All can be rendered invariant to translation, rotation, scaling, and choice of starting point for string matching – though this is not a trivial problem.
- **Measurement of turning angle.** For any given shape, it is possible to represent its boundary by the turning function $\Theta(s)$, measuring the angle of the tangent to the boundary as a function of s, the normalized distance along the boundary from a given reference point. The difference in shape between two objects a and b can thus be computed [1] as

$$\int_0^1 |\Theta_a(s) - \Theta_b(s)| \mathrm{d}s$$

  Such measures are inherently invariant to translation or scaling, and can be rendered invariant to rotation given an appropriate choice of starting point.
- **Elastic deformation of templates.** A potentially powerful, though computationally expensive technique for matching unknown and query shapes is to deform the boundary of the query shape until it matches a given stored shape, and then to measure some function Φ which gives an indication of the cost of the deformation process. A good example of this technique is that of Jain et al [38], who apply displacement functions to the query template in order to compute its goodness of fit with a given image region.

Matching using non-information-preserving features involves calculating and matching a shape feature vector as outlined in Section 3.1 above. Commonly used types of feature include:

- **Simple global features.** Several computationally simple measures of a region's overall shape have been proposed over the years [47]. These include aspect ratio (L/W), circularity ($4\pi A/P^2$), and transparency (A/H), used in the ARTISAN trademark image retrieval system [16].
- **Local features.** Features representing shape characteristics of small regions of an image can often act as a useful complement to global measures. Examples include the line-angle line triplet features devised by Eakins [13], and the longer segment sequences used by Mehrotra and Gary [58].

- **Edge direction histograms.** Another indirect measure of shape within an image is to compute a histogram of edge directions. This can give an indication of directionality within the image, though not necessarily the shape of any object it depicts. Jain and Vailaya's [39] technique identifies edge pixels, computes edge directions, and then accumulates these into bins at 5° intervals.
- **Fourier descriptors.** A very popular way of representing a region's overall shape is to represent the cumulative curvature around the boundary as a function of curve length, and expand this function as a Fourier series [92]:

$$\theta(t) = \mu_0 + A_k \cos(kt - a_k)$$

  The coefficients A_k and a_k, the kth harmonic amplitude and phase angle respectively, known as the *Fourier descriptors* of the curve, provide a description of the curve which appears to reflect its overall shape fairly consistently.
- **Moment invariants.** For any digital image $I(x, y)$, it is possible to compute a series of central moments μ_{pq}, defined as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y)$$

  from which a series of *moment invariants* ϕ_n can be derived which characterize shape in a manner which is invariant to scaling, rotation and translation [33]. Moment invariants have been widely used in image analysis for many years.
- **Zernike moments.** The Zernike moment of order n with repetition m for image $I(r, \theta)$ is defined as:

$$A_{nm} = \frac{n+1}{\pi} \sum_\rho \sum_\theta (R_{nm}(r) e^{im\theta})^* I(r, \theta) | r < 1$$

  where $R_{nm}(r)$ are the set of radial polynomials originally defined by Zernike [86]. Zernike moments have the useful property of orthogonality; their use in trademark image retrieval has been investigated by Kim and Kim [42].

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 [9]. Another is to generate a series of alternative 2-D views of each database object, each of which is matched with the query image [12]. Direct matching of 3-D shapes defined as VRML (virtual reality markup language) primitives has also been attempted [65].

3.5 Retrieval by Other Types of Primitive Features

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, and efficient methods to achieve this have been around for many years. Similar techniques have been applied to image collections, allowing users to search for images containing objects in defined spatial relationships with each other. Spatial indexing appears particularly effective in combination with other cues such as colour [77] or shape [32].

One well-established technique for rapid matching of images on the basis of similarity of spatial layout is 2-D iconic indexing, introduced by Chang et al [8] This generates a string representation of the partial ordering of objects within an image along both $x-$ and $y-$ axes, which can readily be used as the basis for similarity matching. Its lack of rotational invariance can be a problem in some contexts. An alternative method described by Gudivada and Raghavan [28] relies on computing edge graphs between the centroids of every significant object in the image. Query and stored images can then be matched by comparing the relative orientation of corresponding edges. Unlike Chang's method, the technique is insensitive to rotation – though it does require all image objects to be labelled before matching can begin.

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, which can be used to express any function as the sum of a set of orthonormal basis functions:

$$f(c) = c_{00}\phi(x) + \sum_m \sum_n d_{mn}\phi_{mn}(x)$$

where ϕ_{mn}, the wavelet function, is defined as:

$$\phi_{mn}(x) = 2^{-m/2}\phi(2^{-m}x - n)$$

and ϕ is a scaling function. Statistics such as the mean and variance of the wavelet coefficients d_{mn} can model a number of aspects of image appearance, including shape and texture, at different resolutions. Promising retrieval results have been reported by matching wavelet features computed from query and stored images (e.g. [36, 80]). 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 [67], and then computing differential invariants; the whole-image technique uses distributions of local curvature and phase [68].

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 (e.g. [75]) this remains a troublesome problem.

4 Retrieval by Semantic Feature

Retrieval of images containing a specified object, scene or event is a much more formidable task than retrieval by similarity of appearance. Despite views expressed in some quarters that by image retrieval by semantic content is simply not feasible (see, for example, [72]), research in this area is beginning to gather momentum. Several different lines of investigation can be distinguished:

4.1 Automatic Whole-Image Scene Classification

Automatic classification of scenes (into general types such as *indoors*, *city street* or *beach*) can be useful, both because this is an important filter which can be used when searching, and because this can help in identifying specific objects present. Techniques of this kind permit automatic assignment of keywords such as *beach*, *mountain* or *city scene* to appropriate images. The most popular approach has been to use some combination of primitive features to train a classifier to distinguish between different kinds of scene – such as city *vs* landscape or mountain *vs* beach. For example, Szummer and Picard [84] used a combination of colour histograms and texture measures to train a nearest-neighbour classifier to distinguish between indoor and outdoor scenes with 90% accuracy. Oliva et al [63] used shape characteristics of whole-image power spectra sampled with Gabor filters to classify scenes by placing them on appropriate points on two semantic axes: artificial *vs* natural, and open *vs* closed. Vailaya et al [87] have developed a Bayesian classifier to group images into a number of semantically meaningful categories, including city *vs* landscape and forest *vs* mountain, using codebook vectors generated by vector quantization from feature vectors based on colour moments and Gabor coefficients. Reported accuracy was better than 90% for most classification tasks.

4.2 Automatic Object Classification Based on Detailed Object Models

The ability to identify a given type of object in a scene is clearly important for semantic image retrieval, both as an end in itself, and as an intermediate step in the interpretation of more complex scenes. One potentially powerful technique for object recognition in an image is to specify a model for each type of object of interest, and then examine the image for regions conforming to that model. An early system embodying these principles was ACRONYM [5], which used generalized shape modelling to identify and locate instances of desired objects in aerial photographs. After an initial edge detection step, a set of production

rules was used to infer the presence of specified types of aircraft from the patterns of lines derived from the image. The more-recently developed PICTION system [79] identifies human faces in natural scenes by matching candidate face shapes generated by multi-resolution edge detection techniques with a simple three-contour model of hairline and left and right face contours.

Forsyth et al [24] have described a highly sophisticated approach based on developing a model of each class of object to be recognized, and then building up evidence for or against the presence of objects conforming to the model. Evidence includes low-level features of the candidate region itself, and contextual information such as its position and the type of background in the image. Object classification is a three-stage process: (a) segmenting images into coherent regions using a combination of edge, colour and texture information, (b) fusing colour, texture and shape information to identify possible descriptions of each region (for example, as a human arm), and (c) classifying objects from their constituents in terms of component descriptions. The method has been applied with some success to the identification of a range of object types, including unclothed human bodies, horses and trees, though retrieval effectiveness scores are fairly modest at present (15% recall at 66% precision with the horse finder, for example).

4.3 Automatic Object Classification Using Statistical Approaches

A conceptually simpler approach to image interpretation, which does not require the construction of any high-level object model, is the use of statistical techniques (often very similar to those used in scene classification) to assign appropriate semantic labels to individual regions within an image. An example of this is the method of Campbell et al [7], who use a combination of colour and texture features to train an radial basis function network to distinguish between 11 different types of region in a scene, including sky, vegetation, road, and building. They report over 80% classification accuracy for the method. Leung and Malik [46] have developed a method for identifying material within textured regions of an image (such as leather, cork, plaster, etc) using microstructures known as *3-D textons* derived from primitive texture measures.

Schiele and Crowley [74] propose a method using statistically-generated *visual classes* for object recognition. This aims to get round the problem of variability in appearance of objects such as chairs by defining more specific *visual classes*, each of which is sufficiently homogeneous to be identified purely by visual appearance. Buijs and Lew [6] have developed a method for recognising objects (such as oranges) or types of material (such as sand) in an image by inducing 'simple semantics' from primitive image features. They do this by identifying both positive and negative example images, identifying a subset of primitive features with high discriminating power, and using these to train a minimum distance classifier.

4.4 Methods for Learning and Propagating Labels Assigned by Human Users

The problems of achieving effective semantic image retrieval by purely automatic means have led many researchers to investigate methods which are capable of continuous learning through run-time interaction with end-users. Most of these are based on extensions of the principle of *relevance feedback* (see Section 6.1 below). One of the earliest systems to provide this kind of interaction was FourEyes [59], permitting a user to assign semantic labels such as *grass* or *sky* to selected image regions. Once a sufficient number of regions has been labelled, the system attempts to induce grouping rules from the positive and negative examples at its disposal. These rules can then be used to assign labels to new examples sharing the same range of feature values. Effectively, then the system can learn what areas of grass and sky look like, and can then search for images containing such areas.

Lee et al [45] also use relevance feedback to capture semantic information about an image collection. This incorporates a second feedback loop so that users' input is remembered permanently, and used to store semantic links between images as well as similarity of appearance. Initially, images are clustered purely on the basis of primitive feature similarity. Users who search the system are asked to indicate which retrieved images are relevant and which irrelevant. This information is then used to split and merge clusters of similar images, gradually introducing an element of semantic similarity in the process.

Jaimes and Chang's [37] Visual Apprentice aims to provide users with a general framework for building up visual classes which can represent specified types of object or scene. Users can define a visual class by specifying labels for objects and their key constituent parts, together with a set of training examples in which image regions are labelled according to the class definition. The system then uses a combination of lazy learning, decision trees and genetic algorithms to build up a hierarchical object definition in which image regions generated by primitive-level segmentation routines are grouped progressively into *perceptual areas* (groups of regions likely to be perceived as a whole), object parts, whole objects and scenes.

5 Current CBIR Applications and Systems

5.1 Commercial Systems

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. The most well-known commercial systems are:

QBIC. IBM's Query By Image Content system [23] 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 [21]. 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, and the ability to search grey-level images [61]. An online demonstration, together with information on how to download an evaluation copy of the software, is available on the World-Wide Web at `http://wwwqbic.almaden.ibm.com/`.

Virage. Another well-known commercial system is the VIR Image Engine from Virage, Inc [29]. 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.

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 [22]. 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/`.

5.2 Experimental Systems

Prominent experimental CBIR systems, most of which are available as demonstration versions on the Web, include:

Photobook. The Photobook system [66] 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 different application areas, involving retrieval of image textures, shapes, and human faces, each using features based on a different model of the image. Further information on the Photobook system, together with an online demonstration, can be found at `http://www-white.media.mit.edu/vismod/demos/photobook/`.

Chabot. Another early system which has received wide publicity is Chabot [62], 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 [77] 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 [78] 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 select categories of interest; the system then displays images from 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/`

MARS. The Multimedia Analysis and Retrieval Ssystem project at the University of Illinois [34] 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 [71]. A demonstration of the MARS system can be viewed at `http://jadzia.ifp.uiuc.edu:8001/`

Surfimage. An example of European CBIR technology is the Surfimage system from INRIA, France [60]. 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 [52]. 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* (see above) using whole image matching [68]. A demonstration of Synapse in action with a variety of different image types can be found at `http://cowarie.cs.umass.edu/~demo/`.

6 General Issues

6.1 Interfacing and Search Efficiency

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, though keyboard input hardly seems an obvious choice for formulating visual queries. The most appealing paradigm in many ways is query-by-example: providing a sample of the kind of output desired and asking the system to retrieve further examples of the same kind. 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 [23]. 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 [44]. 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 [61]. Some systems provide users with the ability to build up query shapes on the screen from primitives such as rectangles and circles (e.g. [78]).

The ability to refine searches online in response to user indications of relevance, known as *relevance feedback*, is particularly useful in image retrieval. 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 has been demonstrated in several CBIR systems (e.g. Smith and Chang [78], Rui et al [70]). 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 [72].

A continuing challenge facing current CBIR technology is that of efficiently retrieving the set of stored images most similar to a given query. Finding index structures which allow efficient searching of an image database is still an unsolved problem [21]. None of the index structures proposed for text retrieval has proved applicable to the problem, since CBIR techniques are based on a fundamentally different model of data. The most promising approach so far has been multidimensional indexing, using structures such as the R*-tree [2], the TV-tree [50] and the SS+-tree [43], though the overheads of using these complex index structures are considerable. Alternative approaches, which appear to avoid these problems include similarity clustering of images [40], and the use of *vantage objects* [89].

6.2 CBIR Effectiveness

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 [21], though even they glossed over some of the problems of determining whether a given image did in fact answer a given query. 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 [55], Eakins et al [15]) 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 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 [21]. Part of the problem with shape matching lies in the difficulty

of automatically distinguishing between foreground shapes and background detail in a natural image [23]. 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, if any, of the shape feature measures in current use are accurate predictors of human judgements of shape similarity [73].

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.

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. As discussed above, research into semantic image retrieval techniques gathering momentum. But it will take a considerable time before such research finds its way into commercially-available products.

6.3 CBIR and 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 [18]. As

observed above, 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 [62] 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 [35] 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. 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.

6.4 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. 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. Opportunities for synergy in true multimedia systems will be far greater, as demonstrated by the Informedia project [90], 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.

Another 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. This concept has been exploited in the development of MAVIS, a multimedia

architecture which allows generic navigation by image content (shape, colour or texture) as well as text [48]. The authors term this process *content-based navigation* (CBN). A further development of this principle is the *multimedia thesaurus* [49], which allows a system administrator to specify semantic relationships between source items in the link database (such as a given item's set of synonyms, broader and narrower terms), whether text, image or sound.

7 Current Status of CBIR Technology

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 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). It is unlikely, however, that general-purpose image retrieval software will meet the needs of these user communities without a significant degree of customization.

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, 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 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.

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.

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.

The volume of research into improved techniques for CBIR is increasing every year. How much of it is likely to make a real difference to the capabilities of CBIR technology? This is a difficult question to answer. Much current research into improved methods of primitive-level retrieval appears to be concerned with minor modifications to existing techniques. While it would be nice to have better methods for colour, texture and (particularly) shape matching, further research in this area is unlikely to lead to significantly more useful operational systems. One possible exception is research into modelling human perception of image features such as colour [10] or shape [69]. This could lead to systems capable of matching images the way people actually perceive them – what one might call *retrieval by subjective appearance*. Another is research into interface design: despite over ten years' development of CBIR systems, no really satisfactory way has yet been found to formulate a visual query. Overshadowing all these in potential importance is the fast-growing area of semantic image retrieval. While the problems involved are formidable, the potential reward – the development of CBIR systems which meet genuine user needs – is great. There are grounds for cautious optimism that advances in this area will be significant enough to feed

into commercially-available CBIR technology within the next ten years. If this does happen, CBIR will indeed have come of age.

Acknowledgements

Much of the material appearing in this chapter was originally published as JISC Technology Applications Programme Report 39 (see [17] in list of references). It is reproduced here by kind permission of JISC.

8 Further Reading

- del Bimbo A. (1999) Visual Information Retrieval. Morgan Kaufmann, New York.
- Eakins J. P. (2000) Towards intelligent image retrieval. *Pattern Recognition*, in press.
- Eakins, J. P. and Graham, M. E. (1999) Content-Based Image Retrieval. *JISC Technology Applications Programme Report 39*. Available at `http://www.unn.ac.uk/iidr/CBIR/report.html`.
- Lew M. S. ed (2000) Principles of Visual Information Retrieval. Springer-Verlag, Berlin.
- Rui Y. et al (1999) Image retrieval: current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62.

References

1. Arkin, E. M. et al (1991) An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):209–216.
2. Beckmann, N., Kriegel, H.-P., Schneider, R., and Seeger, B. (1990). R*-tree: An efficient and robust access method for points and rectangles. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 19(2):322–331.
3. Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147.
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.
5. Brooks, R. A. (1983) Model-based three-dimensional interpretations of two-dimensional images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):140–150.
6. Buijs J. M. and Lew M. S. (1999) Visual learning of simple semantics in ImageScape. in *VISUAL99: 3rd International Conference on Visual Information and Information Systems*. Lecture Notes in Computer Science, 1614:131–138.
7. Campbell, N. W. et al (1997) Interpreting Image Databases by Region Classification. *Pattern Recognition*, 30(4):555–563.

8. Chang, S. K. et al (1987) Iconic indexing by 2-D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3):413–427.
9. 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.
10. Corridoni, J. M. et al (1998) Image retrieval by color semantics with incomplete knowledge. *Journal of the American Society for Information Science*, 49(3):267–2.
11. Cortelazzo, G. et al (1994) Trademark shape description by string-matching techniques. *Pattern Recognition*, 27(8):1005–1018.
12. 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.
13. Eakins, J. P. (1993) Design criteria for a shape retrieval system. *Computers in Industry*, 21:167–184.
14. Eakins J. P. (1998) Techniques for image retrieval. *Library and Information Briefings*, in press.
15. 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.
16. Eakins, J. P., Boardman, J. M., and Graham, M. E. (1998). Similarity retrieval of trademark images. *IEEE Multimedia*, 5(2):53–63.
17. Eakins, J. P., and Graham, M. E. (1999) Content-Based Image Retrieval. *JISC Technology Applications Programme Report*, 39. Available at `http://www.unn.ac.uk/iidr/CBIR/report.html`.
18. Enser P. G. B. (1995) Pictorial information retrieval. *Journal of Documentation*, 51(2):126–170.
19. Enser, P. G. B. and McGregor, C. G. (1992) Analysis of visual information retrieval queries. *British Library Research and Development Report*, 6104.
20. Evans, A. (1987) TELCLASS: a structural approach to TV classification. *Audiovisual Librarian*, 13(4):215–216.
21. Faloutsos, C. et al (1994) Efficient and effective querying by image content. J*ournal of Intelligent Information Systems*, 3, 231–262.
22. Feder, J. (1996) Towards image content-based retrieval for the World-Wide Web. *Advanced Imaging*, 11(1), 26–29.
23. Flickner, M. et al (1995). Query by image and video content: The QBIC system. *Computer*, 28(9):23–32.
24. 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.
25. 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.
26. 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.
27. Gudivada, V. N. and Raghavan, V. V. (1995). Guest editors' introduction: Content-based image retrieval systems. *Computer*, 28(9):18–22.
28. Gudivada, V. N. and Raghavan, V. V. (1995). Design and evaluation of algorithms for image retrieval by spatial similarity. *ACM Trans. on Information Systems*, 13(2):115–144.

29. 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 2670:76–87.
30. Haralick, R. M. et al (1973). Textrual features for image classification. *IEEE Transactions on Systems Man and Cybernetics*, 3(6):610–621.
31. 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*, 32:3–8.
32. 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.
33. Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, IT-8: 179–187.
34. 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.
35. Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52(1):3–50.
36. 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.
37. Jaimes, A. and Chang S. F. (1999). Model-based classification of visual information for content-based retrieval. in *Storage and Retrieval for Image and Video Databases VII*, Proc SPIE 3656:402–414.
38. Jain, A. K. et al (1996). Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):267–277.
39. Jain, A. K. and Vailaya (1996). A Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244.
40. 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.
41. 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, 7–22.
42. Kim, Y. S. and Kim, W. Y. (1998). Content-based trademark retrieval system using a visually salient feature. *Image and Vision Computing*, 16:931–939.
43. 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.
44. 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.
45. Lee, C. S. et al (1999). Information embedding based on users' relevance feedback for image retrieval. in *Multimedia Storage and Archiving Systems IV* (S Panchanathan et al, eds), Proc SPIE 3846:294–304.
46. Leung, T. and Malik J. (1999). Recognizing surfaces using three-dimensional textons. presented at *Seventh IEEE International Conference on Computer Vision (ICCV-99)*, Corfu, Greece, 2:1010–1017.
47. Levine, M. D. (1985). *Vision in man and machine*, ch 10. McGraw-Hill, NY
48. 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.

49. 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
50. Lin, K.I., Jagadish, H. V., and Faloutsos, C. (1994). The TV-tree—an index structure for high-dimensional data. *VLDB Journal: Special Issue on Spatial Database Systems*, 3(4):517–542.
51. Liu, F. and Picard, R. W. (1996). Periodicity, directionality and randomness: Wold features for image modeling and retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(7):722–733.
52. Ma, W.Y. and Manjunath, B. S. (1997). NeTra: A toolbox for navigating large image databases. In *Proc. of the IEEE Int. Conf. on Image Processing*, 562–571.
53. 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.
54. Manjunath, B. S. and Ma, W.-Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):837–842.
55. 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.
56. Markey, K. (1984). Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research*, 6:155–177.
57. 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, February 1998.
58. Mehrotra, R. and Gary J. E. (1995). Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9):57–62.
59. Minka, T. (1996). An image database browser that learns from user interaction. *MIT Media Laboratory Technical Report*, #365.
60. Nastar, C. et al (1998). Surfimage: a flexible content-based image retrieval system. presented at *ACM Multimedia '98*, Bristol, UK.
61. 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.
62. Ogle, V. E. and Stonebraker, M. (1995). CHABOT: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48.
63. Oliva, A. et al (1999). Global semantic classification of scenes using power spectrum templates. presented at CIR-99: *The Challenge of Image Retrieval*, Newcastle upon Tyne, UK, February 1999.
64. Opitz, H. et al (1969). Workpiece classification and its industrial application. *International Journal of Machine Tool Design Research*, 9:39–50.
65. Paquet, E. and Rioux, M. (1998). *Content-based access of VRML libraries* Lecture Notes in Computer Science 1464:20–32.
66. Pentland, A. et al (1996). Photobook: tools for content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3)233–254.
67. Ravela, S. and Manmatha, R. (1998a). Retrieving images by appearance. in *Proceedings of IEEE International Conference on Computer Vision (IICV98)*, Bombay, India, 608–613.

68. Ravela, S. and Manmatha, R. (1998). On computing global similarity in images. in *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV98)*, Princeton, NJ, 82–87.
69. Ren, M. et al (2000). Human perception of trademark images: implications for retrieval system design. *Journal of Electronic Imaging*, in press.
70. 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.
71. Rui, Y., Huang, T. S., Ortega, M., and Mehrotra, S. (1998). Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Tran on Circuits and Systems for Video Technology*, 8(5):644–655.
72. Santini, S. and Jain, R. (1997). Do images mean anything? In *Proc. of the Int. Conf. on Image Analysis and Processing, ICIP-97*, 564–567.
73. 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.
74. Schiele, B. and Crowley J. L. (1997). The concept of visual classes for object classification. in *Proceedings of SCIA'97, Tenth Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, 43–50.
75. Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.
76. Smith, J. R. and Chang S. F. (1994). Transform features for texture classification and discrimination in large image databases. in *Proceedings ICIP-94*, Austin, Texas, 407–411.
77. 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.
78. 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.
79. Srihari, R. K. (1995). Automatic indexing and content-based retrieval of captioned images. *IEEE Computer*, 28(9):49–56.
80. Stark, H-G (1996). On image retrieval with wavelets. *International Journal of Imaging Systems and Technology*, 7:200–210.
81. 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 2670:29–4.
82. 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 2420:381–392.
83. Swain, M. J. and Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(1):11–32.
84. Szummer, M. and Picard, R. (1998). Indoor-outdoor image classification. in *IEEE International Workshop on Content-based Access of Image and Video Databases (CAIVD98)*, Bombay, India, 42–51.
85. Tamura, H., Mori, S., and Yamawaki, T. (1978). Texture features corresponding to visual perception. *IEEE Trans. on Systems, Man, and Cybernetics*, 8(6):460–473.
86. Teh, C. H. and Chin, R. T. (1988). Image analysis by methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):496–513.

87. Vailaya, A. et al (1998). On image classification: city images vs landscapes. *Pattern Recognition*, 31(12):921–1936.
88. Vailaya, A. and Jain, A. K. (1999). Incremental learning for Bayesian classification of images. presented at *IEEE International Conference on Image Processing (ICIP'99)*, Kobe, Japan, October 1999.
89. Vleugels, J. and Veltkamp, R. (1999). Efficient image retrieval through vantage objects. presented at *VISUAL99: 3rd International Conference on Visual Information and Information Systems*. Lecture Notes in Computer Science 1614:769–776.
90. Wactlar, H. D. et al (1996). Intelligent access to digital video: the Informedia project. *IEEE Computer*, 29(5):46–52.
91. World Intellectual Property Organization (1998). *International Classification of the Figurative Elements of Marks (Vienna Classification)*, Fourth Edition. ISBN 92–805–0728–1. World Intellectual Property Organization, Geneva.
92. Zahn, C. T. and Roskies, C. Z. (1972). Fourier descriptor for plane closed curves. *IEEE Transactions on Computers*, C-21:269–281.