Journal of Information Science

http://jis.sagepub.com

Visual image retrieval: seeking the alliance of concept-based and content-based paradigms

Peter Enser

Journal of Information Science 2000; 26; 199
DOI: 10.1177/016555150002600401

The online version of this article can be found at:
http://jis.sagepub.com/cgi/content/abstract/26/4/199

Published by:
SAGE Publications
http://www.sagepublications.com

On behalf of:
cilip

Chartered Institute of Library and Information Professionals

Additional services and information for *Journal of Information Science* can be found at:

Email Alerts: http://jis.sagepub.com/cgi/alerts

Subscriptions: http://jis.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 12 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://jis.sagepub.com/cgi/content/refs/26/4/199

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

Visual image retrieval: seeking the alliance of concept-based and content-based paradigms

Peter Enser

University of Brighton, UK

Received 8 December 1999
Revised 8 April 2000

Abstract.

In the commercial use of picture collections, a heavy dependency continues to be exhibited on a concept-based image retrieval paradigm in which the query is verbalised by the client and resolved as a metadata text-matching operation.

The practical and philosophical challenges posed by the indexing aspect of image metadata construction are significant and frequently expressed. Nevertheless, it has taken image digitisation to bring this particular information retrieval problem to prominence in the research agenda. Metamorphosed into a binary data structure, the digital image offers some enticing processing opportunities which content-based image retrieval techniques are exploiting with developing success.

Drawing on studies of user need, this paper seeks to explain why a heavy dependency will continue to be placed on concept-based rather than content-based image retrieval techniques within archival image collections. In contrast, the promising nature of content-based techniques from the viewpoint of a growing clientele with less traditional visual information needs will also be considered.

The paper concludes by offering the view that, while both concept-based and content-based approaches suffer from operational limitations, the further development of a hybrid image retrieval paradigm which combines the two approaches makes a potentially valuable contribution to the research agenda for visual image retrieval.

Correspondence to: Professor P.G.B Enser, School of Information Management, University of Brighton, Watts Building, Moulescoomb, Brighton BN2 4GJ, UK. Tel: +44 1273 600900. Fax: +44 1273 642405. E-mail: p.g.b.enser@brighton.ac.uk

1. Introduction

We have entered a new era in visual communication. Developing capabilities in information and communications technologies have enabled the visual image to break out of its analogue mould and reap the benefits of digitisation. A whole new engagement with knowledge and its management in the visual medium is unfolding before us.

These rapidly developing capabilities in the processing and management of digital images, and the opportunities which ease of replication and dissemination may provide for widening access and adding value to image collections, have evoked an enthusiastic response from the community of picture archivists, librarians and curators. Although these guardians of the visual archive have long been exercised by the problem of retrieving appropriate images in response to client demand, it has taken image digitisation to bring this variant of the information retrieval problem to prominence in the research agenda.

As a data structure, the digital image offers the computer scientist enticing opportunities for analysis and manipulation. As a result, the digital image has become the focus of intense research activity. Rui *et al*. [1] have suggested that the earliest manifestations of this activity in the context of image retrieval date from the late 1970s, directed towards an image-handling enhancement of database management systems. From the early 1990s, however, research activity intensified following the adoption of a new approach to visual image retrieval characterised as *content-based image retrieval* (CBIR).

This new paradigm of image retrieval marks a significant departure from that traditionally adopted by the practitioner community of picture librarians and archivists. The latter exhibit a continuing dependency on human intervention in the textual characterisation of image content. They also attach significance to the

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

provision of support for intellectual access to image collections, albeit eroded by developing capabilities in the computer-mediated environment of image database and Web-enabled image retrieval systems. CBIR, on the other hand, invokes computer, rather than human, vision. Accordingly, its natural focus has been the computer laboratory, from which it has only recently emerged into some specialised application environments.

Cawkell [2] first noted the lack of effective communication between the practitioner community and the new breed of academic researchers in image retrieval: a problem which remains serious today [3]. There have been recent attempts, however, to present encompassing views of the theory and practice of visual information retrieval [4–7] and the present paper seeks to make a further contribution in this regard.

2. Image versus metadata

A visual image is a data structure characterised by its possession of certain physical attributes (or 'primitive features' [6]), including size, colours, textures, shapes/ regions and their spatial (or, in the case of moving imagery, spatio-temporal) distribution. The whole assembly, which might be the result of a creative act (e.g. a painting or photograph) or automatic visual monitoring process (e.g. inner city closed-circuit television (CCTV) surveillance) lends itself to interpretation in order to derive the attribute of *meaning*.

Meaning is not a well-defined, quantifiable attribute like colour intensity or the spatial distribution of shapes, of course. It is a property ascribed by human analysis of the image, bringing to bear a combination of objective and subjective knowledge in a sociocognitive process, as described by Heidorn [8]. The semantic analysis involved in the indexing of visual imagery is an important component of this process and has been one of the responsibilities of the picture librarian, archivist or curator. The variety of physical carriers of two-dimensional visual imagery, including paintings, prints, maps, photographs, video and film, has led to specialised forms of curatorship. However, there is a common need to construct inventories of such collections. In compiling a catalogue record which allocates an identifier to an image (or compilation of images) and confirms its existence within a collection, an attempt is usually made also to ascribe meaning to the image by means of some level of interpretative annotation. Typically, this might take the form of title, keywords or phrases, caption, synopsis or shot list, or some combination thereof.

Such a semantic record is a (substantially, if not wholly) textual surrogate of the image. In digital image databases, the image and its surrogate are co-located in metadata. The coexistence of these two data structures is represented in Fig. 1. The substantial development of such databases in recent years, as described by Besser [9], has been accompanied by the much-needed promotion of metadata standards, a brief overview of which may be found in [6].

3. Concept-based image retrieval

In the predominant paradigm of visual information retrieval, transactions are conducted with respect to the textual annotations within the metadata of an image collection. The process, usually known as *concept-based image retrieval* and illustrated in Fig. 2, involves a verbal expression of the query, possibly mediated by a thesaurus or classification scheme in order to couch the query in terms of a controlled, or authorised, vocabulary. The (modified) expression is then matched against the textual annotation associated with each image. Any matching expression (or one which matches sufficiently closely to satisfy some similarity threshold) results in the recovery of its associated image, which is then presented to the client for consideration. Such an operation, characterised elsewhere as the 'Linguistic Query, Linguistic Search (LL)' model of visual information retrieval [4], effectively ignores the fact that the information need is in the visual domain and translates the problem into a simple text-matching operation.

In order for the text-matching operation to retrieve salient images in response to a statement of visual information need, the linguistic representation of the

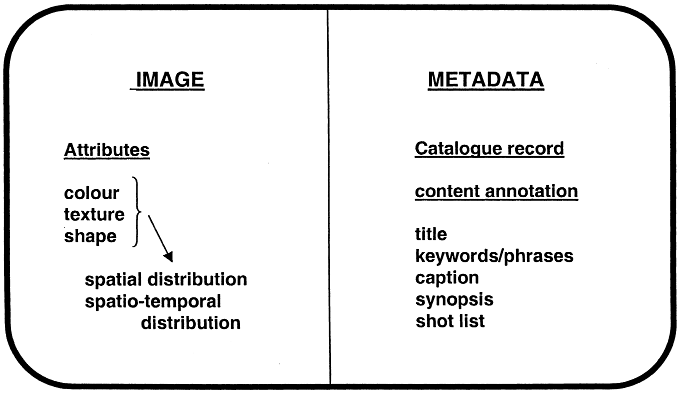

Fig. 1. The image and its metadata.

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

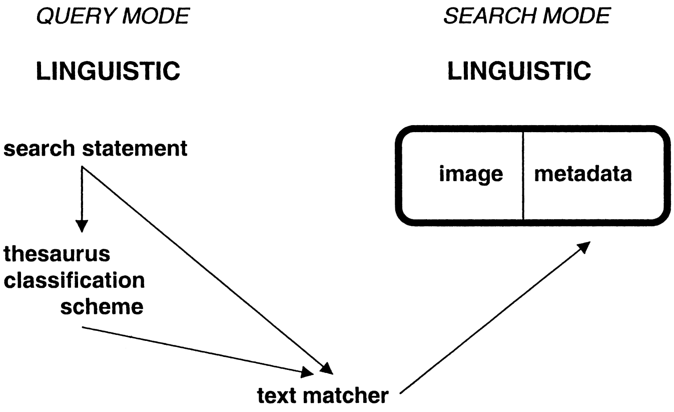

Fig. 2. The concept-based image retrieval model.

content or meaning of an image must be effective. However, capturing in words the content or meaning of an image is a significant intellectual challenge.

The nature of this challenge has been widely discussed (see Enser [4] and Rasmussen [5] for a comprehensive coverage). In essence, semantic analysis of an image typically identifies more than one layer of meaning. The pre-iconographic, iconographic and iconological levels of analysis proposed by Panofsky [10] and generalised by Shatford [11] offer a convenient formalism for the notion that an image is not a single semantic unit but an amalgam of generic, specific and abstract semantic content. Furthermore, the attribution of some of that content (notably, Panofsky's expressional pre-iconographical, and iconological levels) will reflect a purely subjective response to the image by the beholder.

In Shatford's terms: 'The delight and frustration of pictorial resources is that a picture can mean different things to different people' [11]. Indeed, a picture can mean different things to the same person at different times or under different circumstances of need. We capture the essence of these troublesome conclusions, of course, in the old adage 'a picture is worth a thousand words'.

The important, if dispiriting, conclusion to which one is drawn is that *the retrieval utility of an image is inherently unpredictable*. To use an example from Besser [12]: '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' – or, indeed, to a host of other people whose needs for the photographs we cannot foresee. Although the utility of the images is here expressed in terms of types of user, the retrieval utility actually derives from the change in such a street scene occasioned by the passage of time. When the images were first created a century ago, their future interest could not have been predicted in other than a general sense; moreover, the annotation applied at that time would have reflected no more than a current perception of the scene. Forsyth made a similar point when noting that images may acquire retrieval utility through their depiction of previously unknown individuals who have suddenly been made famous by a news event [13].

This unpredictability of retrieval utility is a characteristic of image material which, arguably, is not manifest to anything like the same degree in textual material. This, in turn, must influence our view of the utility of subject indexing of image material, since, in the abstract, there can be no means of determining the appropriate level of indexing exhaustivity. In practice, the situation is ameliorated by (i) the specialisation of subject focus adopted by many picture libraries and archives and/or (ii) their adoption of subject classification or thesaural devices which condition the interpretation placed upon the meaning or significance of images within their collections. Overviews of such devices, together with relevant website addresses, may be found in the comprehensive report to the Joint Information Systems Committee (JISC) Technology Applications Programme by Eakins and Graham [6].

To these reflections must be added concerns surrounding indexer subjectivity, which, following Markey [14], one might expect to be especially pronounced in the visual medium. Attention has also been drawn frequently to the potentially serious time/cost implications of manually indexing these visual resources.

The intellectual and practical challenges posed by the semantic indexing of images are accentuated when those images are encapsulated within film or video footage. The rate at which video material is being generated and submitted for archiving presents some organisations (national television broadcasting agencies being an obvious example) with a huge practical problem. As to the indexing methodology, while it may be appropriate to conceptualise the moving imagery as a set of static shots or keyframes, there is the added need to represent the semantic continuity, i.e. the storyboard, of those discrete units. The spatio-temporal properties of indexable objects – their motion or animation and

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

relationships – may also be highly significant as a visual feature, but problematical to capture verbally. Two other kinds of motion information in videos, i.e. camera movement and post-processing effects like warping, add further complexity [15]. Film/Video librarians respond to these challenges by means of synopses or shot lists of greatly varying exhaustivity; the ratio of cataloguing time to transmission time can be as high as 30:1 for a major national television broadcaster.

These considerations resonate with Svenonius' perception of the philosophical challenge posed by any attempt to express in words the 'aboutness' of a work which is cast in a wordless medium [16]. Although her argument is couched in terms of art or music, it applies with equal force to creative and digitally enhanced imagery, as Fig. 3 testifies.

From the foregoing, it is clear that there are significant problems associated with concept-based image retrieval. These problems have been well rehearsed in the literature and interest has been expressed in their alleviation by eschewing the verbal expression of information need and image content.

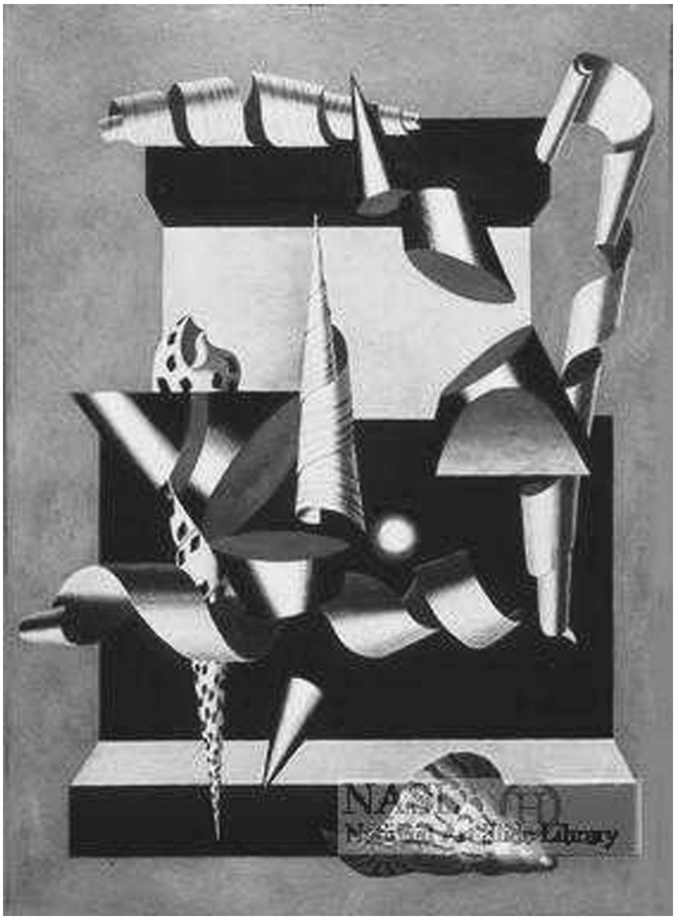

Fig. 3. 'Abstract', a watercolour painting by Edward Wadsworth (National Arts Slide Library).

In the absence of words, image indexing and retrieval must operate directly on the image rather than on its metadata. Central to this approach is the notion of the query assuming a visual form, i.e. itself an image that the client can use as an example of the sort of image which he or she wishes to be shown or which contains a feature of interest. The query is then digitised and it participates as an input data structure to a similarity matching process conducted on the target collection of digitised images. We have entered the CBIR arena.

4. Content-based image retrieval

The CBIR matching process, which is represented in Fig. 4, is conducted on those image attributes of colour, texture and shape; the latter elaborated by spatial (or spatio-temporal) distribution, which are amenable to quantification and, thereby, automatic indexing. Since this process is conducted on unstructured arrays of pixel intensities, in contrast to the logically structured data (ASCII character strings) which populate text databases, CBIR at this level is said to have no parallel in text-based information retrieval [6, 17].

Colour has been a widely used attribute in CBIR studies and, arguably, that with which the most encouraging results have been achieved thus far. This approach envisages a colour space, which is modelled in terms of colour channels, e.g. red, green, blue (RGB) or hue saturation value (HSV) [8]. The colour space is partitioned into n gradations of hue (bins), while the query image and target collection images are usually represented as colour histograms, each assuming the

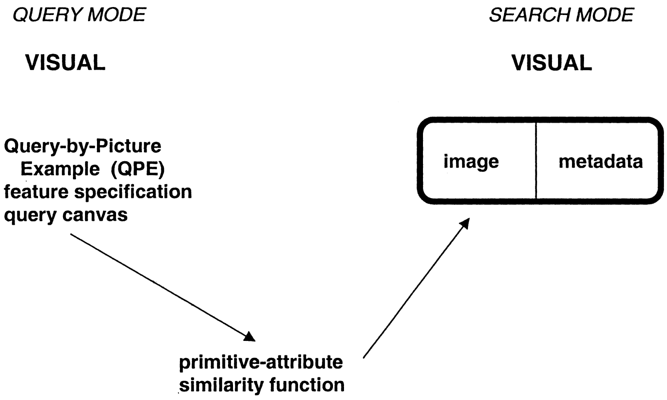

Fig. 4. The content-based image retrieval model.

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

form of a vector (h_1, h_2, ..., h_n), where h_j represents the number of pixels within the image assigned to bin j, $1 \leq j \leq n$. These vectors act as image (and query) surrogates, between which similarity analysis is conducted, typically using histogram intersection techniques. There are echoes here of the document vector processing employed in text-based information retrieval [18]; furthermore, the process of allocating colours to bins – colour quantising – may be likened to a text conflation procedure.

Storing the histograms within the image database as indexing surrogates also facilitates image retrieval in response to requests for images which have specified colour percentages. However, colour histograms lack colour layout cues and cannot differentiate objects which share the same proportion of specific colours but in different spatial distributions, such as the British and French flags [13, 19]. Improvements to the original histogram intersection technique of Swain and Ballard [20], and especially the use of colour correlograms, address this problem and that of distortions due to quantisation [1, 6, 13, 15, 21]. However, the distortion of pixel values induced by variations in illumination are not so amenable to solution [15].

In step with the growing opportunities to gain access to stores of digitised images, there is a developing body of users for whom the ability to retrieve images in response to colour specification is advantageous: applications in fields as diverse as medical diagnosis, fashion and interior design, art history, journalism and advertising have been identified [22].

Progress has also been made in the automatic indexing of images on the basis of the texture attribute. In Forsyth's terms: 'Most people know texture when they see it, though the concept is either difficult or impossible to define' [13]. Nevertheless, a formal definition is provided by Rui *et al.* [1], who describe texture as an innate property of virtually all surfaces, identified as visual patterns having 'properties of homogeneity that do not result from the presence of only a single colour or intensity'. Forsyth's characterisation of texture as the difference between a dalmatian and a zebra helps to make the point [13].

In the context of visual image retrieval, emphasis has been placed on computational approximations to a number of visually meaningful texture properties, among which coarseness, contrast and directionality have been shown in psychophysical studies to be of particular significance to the human visual system. Typically, these three texture features are computed from local neighbourhood analysis of each of an image's pixels. The set of feature vectors generated from all of the image's pixels forms a three-dimensional (3D) texture histogram, which may be used in image similarity analysis [21]. References to specific techniques for texture analysis may be found in [1, 6, 13].

One of the most potentially valuable approaches to automatic image retrieval by primitive feature involves shape analysis. Shape is generally defined in terms either of boundaries (the outer perimeter of a shape feature) or regions (the entire 'footprint' of the shape feature). Unfortunately, at present, serious difficulties in disambiguating foreground from background content, and in registering the perceptual similarity of shape features perturbed by rotation, scaling, occlusion or translation (repositioning), all severely constrain effective, automatic feature extraction from real-scene imagery.

In general, use of shape (and colour layout) features in CBIR has to confront the challenge of segmenting the image in order to identify possible objects of interest as spatial regions within colour-texture space. Much ongoing research of a deeply mathematical nature is being directed at this problem by the computer vision research community, recent overviews or more detailed descriptions of which may be found in [1, 5, 6, 13, 15, 21, 23, 24]. Eakins and Graham [6] also reported on promising techniques which avoid the 'troublesome problem' of needing to segment the image before shape descriptors can be computed.

This literature is directed at the computer scientist and, in particular, the computer vision community. Without the necessary mathematical armoury, the reader is likely to be greatly challenged by the techniques and terminology used – an explanatory factor in the lack of effective communication between the practitioner and research communities in visual image retrieval, to which reference was made earlier.

The application of CBIR techniques to moving imagery is also being pursued with increasing vigour. Sometimes characterised as *video asset management* (a term which, in the author's view, shares with *content-based* image retrieval the misfortune of implying an operation other than that actually involved!), automatic procedures have been developed for shot boundary detection, keyframe selection and scene clustering. Such automatic processing of the raw or Motion Picture Experts Group (MPEG) compressed footage can represent a considerable saving in time and cost of total indexing effort, enabling users faced with short deadlines to browse through a large video archive in order rapidly to retrieve particular stories or types of scene [25]. A number of CBIR-based products have gravitated from the laboratory to the marketplace in response to the needs of television production companies [6].

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

Reference has been made in Section 3 above to the difficulty of achieving a semantic indexing of the spatio-temporal properties of objects featured on video. As Bolle *et al.* [26] have observed, the semantic content of video, while static at the frame level, is dynamic at shot and scene levels. The CBIR research community has harnessed motion vectors – an MPEG compression feature which models the inter-frame dynamics of pixel structures ('optic flow' [13]) – as a means of representing the appearance, disappearance, actions and interactions of objects. These motion vectors can be interrogated to provide responses to certain types of video retrieval queries, such as query-by-motion-example and object tracking [6].

CBIR techniques have been applied in a number of well-known experimental systems, including IBM's Query By Image Content (QBIC) [27, 28], VisualSEEK [25], Virage [29], Photobook [30, 31], Excalibur [32] and MARS [21]. Helpful overviews of these systems are provided by Rui *et al.* [1] and Eakins and Graham [6] and some applications are reported in [5].

Scaling up from such experimental systems to operational image retrieval remains a major challenge because of what Huang *et al.* [21] call the 'dimensionality curse' of the image feature space. They observe that the most robust indexing methods work well only for multidimensional feature spaces with dimensionality around 20. However, Rui *et al.* [1] have reported that the dimensionality of the feature vectors is normally of order 10^2. Techniques for transforming high-dimensional feature spaces to lower dimensions do exist, e.g. QBIC employs such a technique, but there is uncertainty about the robustness of such techniques [1, 21].

5. Visual image retrieval from the user's perspective

At present, there remains the worry that CBIR techniques are being tested against artificial queries which bear little resemblance to those encountered among the clients served by the image practitioner community. A number of projects have been reported which have offered some insights into the characteristics of real client need in the visual medium [33–44].

These studies have often emphasised the high incidence of requests for images which feature named people ('Sergei Prokofiev'), places ('Yardley brickworks'), events ('July 3, 1951: Judy Garland . . . at Birmingham Hippodrome') and objects ('HMS Volunteer'). The incidence of query refinement, whereby the depiction of the target object, event or person is constrained in terms of time period ('Blackpool holiday enjoyment, 1949–1951'), location ('Napoleon at Jaffa, 1799'), action ('Edward Heath gesticulating') or other desired aspect, has also been highlighted [4, 33–35].

Clearly, these findings reflect the nature of the collections sampled and the procedures employed for the recording of the requests. With regard to the former, the collections were archival in nature, as opposed to stock shot providers or other minimally indexed image collections. For such archival collections, a high incidence of requests for images which feature appellation (the naming of uniquely defined people, places, events and objects) is to be expected. As regards the procedures employed for request recording, it is important to note that the majority of the sampled requests were mediated by a picture researcher or receptionist, the recorded request inevitably having gained in both specificity and refinement as a result.

Notwithstanding the methodological constraints, these user studies do serve to emphasise the fact that the retrieval of the iconographic content of image material, where specific identification of some feature is at issue, is dependent on a defining textual expression. In Eakins' terms, 'recognizing and tracing the path of a river or road can be performed automatically; naming the river as the Mississippi or the road as the M4 requires some human intervention' [45].

Studies of journalists' requests to picture libraries add further emphasis to the significance of uniquely defined subject matter and the reliance placed on a text caption for image interpretation and signification [41–44]. Fig. 5 illustrates the point: the image has little significance (except to an aviation expert and to the photographer who created it) without a supporting caption which uncovers the fact that the aircraft has Pinochet on board, for whom this is a flight to freedom from the threat of extradition or a war crimes trial. Given this 'invisible' fact, the image acquired sufficient metaphorical significance to be reproduced on the front page of various national newspapers.

For any visual feature which can be named as a specific instantiation of an entity, there is a parallel interpretation at the generic ('pre-iconographic' [10, 11]) level. Examples of real client queries at this level may be found in [33–35], where the argument is made that such queries are qualitatively different from those which seek specific instantiations. This argument is reinforced in Table 1, in which some examples of requests addressed to a stock shot picture library are reproduced. Eakins and Graham [6] have offered an alternative classification of queries that treats generic

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

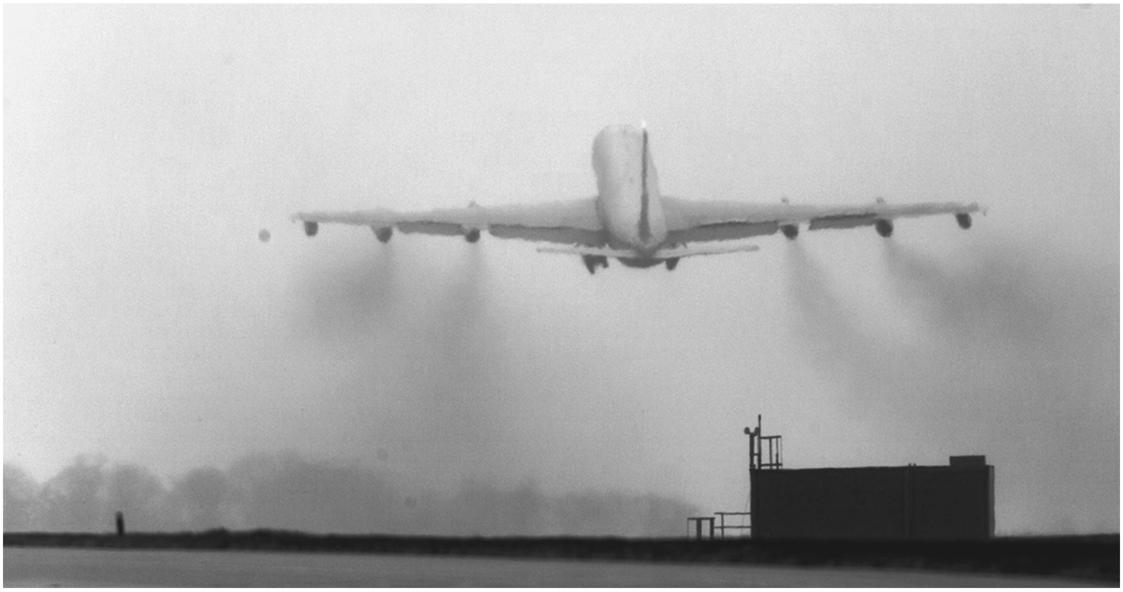

Fig. 5. 'General Pinochet leaving on the Chilean Airforce jet at RAF Waddington, Lincolnshire, today' (Alban Donohoe Picture Service).

objects (including people) and their specific instantiations as a single category of 'derived (logical) feature'. The fact that the satisfaction of requests for images of objects, whether generic or specific, is a matter of identification in both cases does provide a warranty for such an approach. However, there is some value in following Panofsky [10], Shatford [11] and others in distinguishing between the two. The classificatory or thesaural tools used in metadata construction are typically cast in terms of generic features; only in constrained domains of knowledge is it an economic proposition for such structures to represent a variety of specific instantiations. It follows that the retrieval of images may be more readily assisted by the use of finding aids such as classification schemes and/or thesauri when the query relates to a generic feature. The search for named, specific objects and persons is more likely to involve browsing and/or text matching.

Table 1
Sample of requests addressed to a stock shot picture library

Gutsy fat blokes – doing anything – full length or just parts of body – quirky or obscene
Person scratching an itch
Bird asleep on a perch
Search for cookery feature on stirfries – young, trendy feel – 25–33, girls night in, couples, dinner parties with friends – each group in kitchen cooking stirfry – holding, shaking, stirring pans – laughing, having fun – directed towards 'friends' sitcom
Sausage and egg
Pretty girl doing something active, sporty in a summery setting, beach – not wearing lycra, exercise clothes – more relaxed in tee-shirt. Feature is about deodorant so girl should look active – not sweaty, but happy, healthy, carefree – nothing too posed or set up – nice and natural looking
Stressed women
Campaign cars, with megaphones
Person having blood pressure taken

At the highest level of abstraction in the interpretation of image meaning or content, i.e. that which corresponds with Panofsky's iconological level, the human reasoning based on tacit or world knowledge which underpins image indexing and retrieval operations poses a seemingly insurmountable obstacle to the application of CBIR techniques. At this level, we humans are able to 'see' within the primitive attributes of two-dimensional imagery the portrayal of love, power, benevolence, hardship, discrimination, triumph, persecution and a host of other aspects of the human condition. We are enabled, through the visual medium, to exercise skills in semiological analysis – the shared connotation of the icon, metonym and metaphor, the understanding and appreciation of two conceptually related but antithetical images such as those reproduced in Figs. 6 and 7.

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

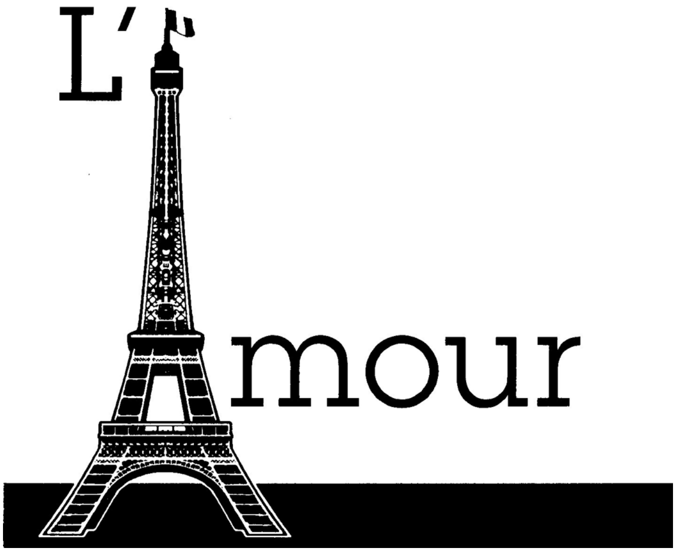

Fig. 6. Ideagram (First Choice Holidays and Flights Ltd).

Requests which address the iconological property of images conform with Fidel's observation that images may be sought on the basis of their holistic content or message, as opposed to the information embedded within them by dint of their depiction of certain features [39].

The incidence of queries which relate to the image as artefact have also been noted [35, 37, 38], although here a quite different information need addresses the image as a physical object rather than the conveyor of a message in the visual medium. Such queries rise to particular significance in the case of image collections in the field of art and art history, of course, where aspects of the provenance of an image (biographical check of artist/movement, attribution check via image/style/ subject, subject check via content identification, attribution check via inscription / signature, bibliographic check of artist/movement/subject, (past) ownership check, object dimensions check, (past) collection check) or its accessibility (location/viewing query, (current) ownership query, (current) collection query, copyright query) are frequently encountered queries.

In reality, then, user needs for recorded visual knowledge are diverse and complex. We cannot be surprised that their satisfaction has traditionally been entrusted to that highly capable, intelligent device known as the picture researcher! The picture researcher brings three fundamentally important attributes to the task of mediated visual information retrieval: collection knowledge, expert (domain) knowledge and world knowledge. In combination, they arm the human intermediary with the in-depth familiarity with the conceptual and physical attributes of a collection of image material, the underpinning expertise in a given subject domain and the tacit knowledge which informs our interpretation and signification of any given scene.

These three levels of knowledge, when combined with a well-tuned visual memory, facilitate the efficient retrieval of appropriate images with recourse only to the researcher's ability to visualise a depiction which is the subject of a query. This capacity to visualise – to create visual mental models – is a most important cognitive skill in visual information retrieval [8]. It is complemented by our seemingly innate ability to inspect a sequence of images and make extremely rapid judgements as to their relevance in response to a specific visual information need. As a result, and partly in consequence of the difficulties which attend the semantic indexing of image material, browsing usually plays a highly significant role in visual information retrieval. This fact is reflected in the design of the user interface to digital image databases, which typically enables the user to 'speed view' a scrolling sequence of thumbnail images, any one of which can be magnified for more detailed consideration. Such a facility is built

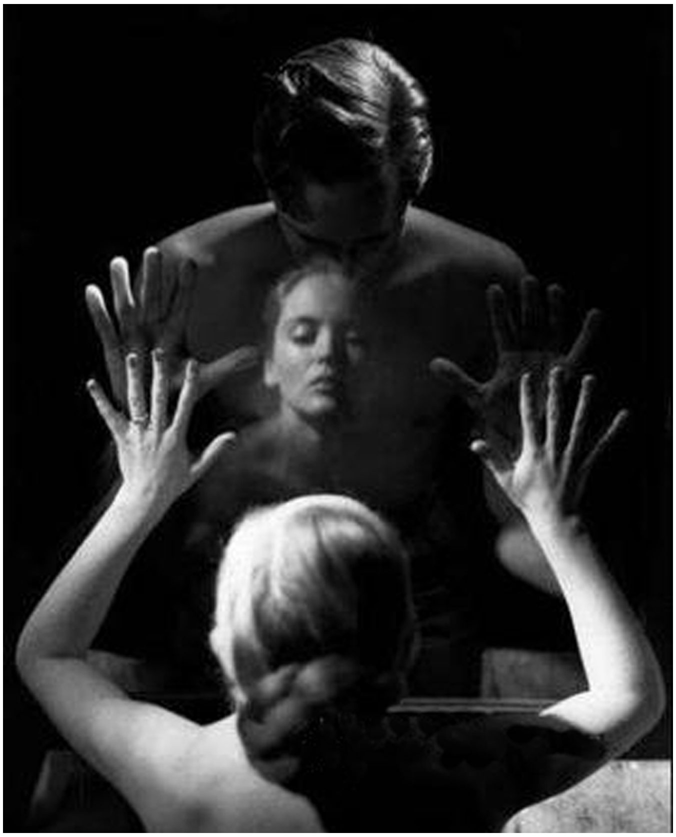

Fig. 7. 'Break-Up', a photograph by Thurston Hopkins (Hulton Getty Picture Post Collection).

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

into almost all current CBIR systems too, the retrieved images being presented in rank order of similarity with the query image. The more sophisticated of such systems also offer relevance feedback, requantifying the comparator attribute of the query image (e.g. colour) to take account of the values for the retrieved images.

6. Hybrid image retrieval systems

One of the most encouraging developments at the interface between theory and practice in image retrieval has been the emergence of the *hybrid image retrieval system*. At their simplest, such systems enable (i) the query to be posed verbally, (ii) a text-matching operation to recover images on the basis of content description in their metadata and (iii) a CBIR technique to accept these images as input to a similarity matching process which might enhance recall by retrieving further images without reference to their indexing. Both the Yahoo! and AltaVista Web search engines incorporate such capabilities, but the reader might like to verify for himself or herself how dramatically poor can be the precision performance of a retrieval tool which responds to a user's request for images similar to an exemplar in which an *object* of interest is depicted by offering images which have similar quantifications of *colour* to those of the exemplar! Forsyth *et al.* [19] have characterised such responses rather kindly as 'eclectic in content'.

A more sophisticated approach to this 'Linguistic Query, Visual Search (LV)' model of image retrieval [4] can be conceived, wherein a visual thesaurus or dictionary embedded within the user interface effects the translation from verbal to visual query. In theory, such an approach offers a means of mitigating the indexing exhaustivity problem to which reference was made earlier. In practice, considerably enhanced functionality of CBIR techniques beyond those which operate on the primitive features of an image is necessary. This enhanced functionality is currently the subject of considerable research effort, often characterised by the CBIR research community as 'semantic image retrieval'.

Typically, semantic-level CBIR is achieved by developing a 'reference' model of the object to be sought, then seeking to recognise the presence of a similar object within a collection of stored images by employing procedures for grouping the images into semantically meaningful categories on the basis of primitive features [6, 19, 46].

Such a feature recognition process avoids the full complexity of automatic identification, therefore, by using a classification procedure. The technique has been applied successfully in a number of highly constrained domains, including recognition of the presence within an image of horses [13], trees [21], people (albeit naked, since, unlike clothed people, 'naked people display a very limited range of colours and are untextured') [19], environmental features (cityscapes, forests, mountains) [46] and crops [47]. Alternative formulations have used neural nets and genetic algorithms as learning tools rather than classification [1]; in other cases, scene recognition rather than object recognition has been the focus of enquiry [6].

Complementary with these approaches to automatic semantic-level CBIR are semi-automatic procedures which seek to bridge the gap between high-level human semantic processing and low-level machine processing of images. Central to these procedures is the initial, human specification of the visual cues which he or she finds significant. This might take the form of annotating regions of interest within an exemplar image or completing a semantic visual template by means of which preferred value ranges on primitive attributes are specified. Human high-level semantic processing is further engaged by means of relevance feedback procedures as a means of augmenting system performance [6, 30, 48].

By means of relevance feedback, such systems 'learn' to relate high-level semantic interpretation to low-level primitive attributes of an image. Having associated a semantic label with a region within an exemplar image, other images which are retrieved on the basis of their having similar primitive attribute values may be indexed semantically by having the matching regions similarly labelled. Once a semantic visual template has been refined to the user's satisfaction as a specification of interest, a semantic label or simple caption can be assigned to the query, which is then stored within a query database that associates each semantic concept with a range of primitive attribute values [6, 30, 48].

Encouraging though such developments may be in the laboratory, their operationalisation for commercial purposes in other than highly constrained domains of knowledge seems very problematic: to the high computational expense of implementing relevance feedback over the Web must be added the user subjectivity which is an essential characteristic of relevance feedback and the difficulty of formulating a robust relationship between a semantic concept and a (fusion of) primitive attribute(s) in real-scene imagery.

Nevertheless, the challenge posed by the indexing and retrieval of moving image material is one particularly susceptible to mitigation by the hybrid model. Documents cast in film and video format are multi-

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

media in nature. In addition to streamed objects in the visual medium, this documentary form usually embodies a soundtrack which conveys speech, music or other aural stimuli and might also offer closed-caption text. Even silent, archival film often has some textual content. In the Informedia project [49], for example, video material is partitioned automatically into scenes by composite analysis of motion vectors, colour histograms and soundtrack signals. The scenes are then indexed by the detected video objects, significant words are extracted from the soundtrack and text is derived from captions. The Informedia-II project incorporates a number of developments, including dynamic story segmentation; speaker, voice and face recognition; video event characterisation and similarity matching [50].

7. Conclusion

New capabilities in the digitisation, storage, manipulation and transmission of still and moving images are providing enhanced opportunities to gain *physical* (albeit virtual) access to our visual heritage, to augment our visual culture and to service our knowledge-based economy. By harnessing the forces of information and communication technologies, the research community has made of visual information retrieval (and multimedia asset management in general) a vibrant research topic. We can look with confidence to there being yet more developments in the future. Nor can there be any doubt of our willingness to absorb the fruits of that endeavour: a visually-oriented and Web-based culture is penetrating our society with accelerating force.

There is much to excite us here. Yet, we must not be blinded by the technology into believing that the new capabilities currently offer significant advances in meeting our needs for *logical* (subject) access to our visual resources.

This paper has sought to make the point that, in order to satisfy the majority of currently articulated visual information needs, a rich combination of collection, domain and world knowledge has to be brought to bear and that the knowledge transaction has to be conducted, in part at least, in verbal form. There is a heavy dependency on the quality of the metadata compiled from the manual cataloguing and indexing of image material.

These observations place a significant constraint on the efficacy of the CBIR paradigm. Nevertheless, as this paper has also sought to show, there is a growing population of users who stand to benefit from CBIR processes and for whom a requirement to cast their query/request in a verbal form might involve them in an artificial and problematical translation of their visual information need. There is already a number of specialised application areas, e.g. fingerprint, fabric and trademark matching; face recognition; colour matching of items in electronic mail-order catalogues; texture-based classification of geological samples, etc, where CBIR has been applied with some success [6]. More generally, CBIR might be said to offer the potential for enhanced consistency and productivity of image indexing.

However, if logical access to visually encoded knowledge is to reap the benefits of the great advances made in providing physical (virtual) access to such material, it seems clear that the huge importance of concept-based image retrieval must not be relegated in our enthusiasm to promote research into CBIR. Similarly, the high level of functionality of the picture researcher, librarian, archivist or curator must be acknowledged, not subjugated in the headlong rush towards technologically attractive solutions to the image retrieval problem.

There is, above all, a need to make advances in *both* concept- and content-based image retrieval. This calls for the design of image retrieval interfaces of considerably greater capability than either the operational or experimental ones of today. In this regard, the attention now being paid to hybrid image retrieval systems is to be welcomed. For such retrieval systems to be truly integrative of both concept-based and content-based image retrieval paradigms, however, there has to be effective communication between those who are concerned with the operation of picture archives and text-based retrieval of image resources, on the one hand, and those who are involved in machine vision and automatic image analysis, on the other. Their current paucity of shared perceptions and vocabulary bodes ill for the full exploitation of visual knowledge management which the digital age invites.

For the next generation of visual image retrieval systems, the design challenge is great indeed, but so too is the potential contribution to our visually absorbed society.

Acknowledgements

I am grateful to Christine Sandom and Samantha Gregory for practical assistance in the preparation of

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

this paper. My thanks also to Mark Rorvig for helpful observations and for bringing to my attention the work of the late David Marr, whose computational modelling of human visual information processing provided the foundation for CBIR.

I wish to thank the following organisations for their kind permission to reproduce material included in this paper: De Montfort University and Associates (Fig. 3); Alban Donohoe Picture Service (Fig. 5); Tony Stone Images (Table 1); First Choice Holidays and Flights Ltd (Fig. 6); Getty Communications (Fig. 7).

References

[1] Y. Rui, T.S. Huang and S-F Chang, Image retrieval: current techniques, promising directions, and open issues, *Journal of Visual Communication and Image Representation* 10 (1999) 39–62.

[2] A.E. Cawkell, Selected aspects of image processing and management: review and future prospects, *Journal of Information Science* 18(3) (1992) 179–192.

[3] D.J. Harper and J.P. Eakins (eds), *The Challenge of Image Retrieval: Papers presented at CIR99 – Second UK Conference on Image Retrieval, 25–26 February 1999, Newcastle upon Tyne, UK* (University of Northumbria at Newcastle, 1999) [Preface].

[4] P.G.B. Enser, Pictorial information retrieval [Progress in documentation], *Journal of Documentation* 51(2) (1995) 126–170.

[5] E.M. Rasmussen, Indexing images. In: M.E. Williams (ed.), *Annual Review of Information Science: Volume 32* (Information Today, Medford, NJ, 1997), pp. 169–196.

[6] J.P. Eakins and M.E. Graham, *Content-based Image Retrieval: A Report to the JISC Technology Applications Programme* (Institute for Image Data Research, University of Northumbria at Newcastle, January 1999).

[7] B. Sandore (ed.), Progress in visual information access and retrieval, *Library Trends* 48(2) (1999) 283–524.

[8] P.B. Heidorn, Image retrieval as linguistic and nonlinguistic visual model matching. In: [7], pp. 303–325.

[9] H. Besser, Image databases: the first decade, the present, and the future. In: P.B. Heidorn and B. Sandore (eds), *Digital Image Access and Retrieval: Papers presented at the 1996 Clinic on Library Applications of Data Processing, March 24–26, 1996, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign* (GSLIS, University of Illinois at Urbana-Champaign, 1997), pp. 11–28.

[10] E. Panofsky, *Meaning in the Visual Arts* (Doubleday Anchor Books, Garden City, NY, 1955).

[11] S. Shatford, Analyzing the subject of a picture: a theoretical approach, *Cataloguing and Classification Quarterly* 5(3) (1986) 39–61.

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

[13] D.A. Forsyth, Computer vision tools for finding images and video sequences. In: [7], pp. 326–355.

[14] K. Markey, Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials, *Library and Information Science Research* 6(2) (1984) 155–177.

[15} A. Gupta and R.C. Jain, Visual information retrieval, *Communications of the ACM* 40(5) (1997) 71–79.

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

[17] S. Santini, and R.C. Jain, The graphical specification of similarity queries, *Journal of Visual Languages and Computing* 7 (1997) 403–421.

[18] G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval* (McGraw-Hill, New York, 1983).

[19] D.A. Forsyth *et al.*, Finding pictures of objects in large collections of images. In: P.B. Heidorn and B. Sandore (eds), *Digital Image Access and Retrieval: Papers presented at the 1996 Clinic on Library Applications of Data Processing, March 24–26, 1996, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign* (GSLIS, University of Illinois at Urbana-Champaign, 1997), pp. 118–139.

[20] M.J. Swain and D.H. Ballard, Color indexing, *International Journal of Computer Vision* 7(1) (1991) 11–32.

[21] T. Huang, S. Mehrotra and K. Ramchandran, Multimedia Analysis and Retrieval System (MARS) Project. In: P.B. Heidorn and B. Sandore (eds), *Digital Image Access and Retrieval: Papers presented at the 1996 Clinic on Library Applications of Data Processing, March 24–26, 1996, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign* (GSLIS, University of Illinois at Urbana-Champaign, 1997), pp. 100–117.

[22] V.N. Gudivada and V.V. Raghavan, Content-based image retrieval systems, *Computer* 28(9) (1995) 18–22.

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

[24] J.P. Eakins, Automatic image content retrieval – are we getting anywhere? In: M. Collier and K. Arnold (eds), *Proceedings of the Third International Conference on Electronic Library and Visual Information Research (ELVIRA3)* (De Montfort University, Milton Keynes, 1996), pp. 123–135.

[25] S-F. Chang, J.R. Smith and J. Meng, Efficient techniques for feature-based image/video access and manipulation. In: P.B. Heidhorn and B. Sandore (eds), *Digital Image Access and Retrieval: Papers presented at the 1996 Clinic on Library Applications of Data Processing, March 24–26, 1996, Graduate School of Library and Information Science, University of Illinois at Urbana-*

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.

Champaign (GSLIS, University of Illinois at Urbana-Champaign, 1997) pp. 86–99.

[26] R.M. Bolle, B-L. Yeo and M.M. Yeung, *Video Query: Research Directions* (1998). Available at: http://www.research.ibm.com/journal/rd/422/bolle.txt

[27] Available at: http://www.qbic.almaden.ibm.com/

[28] M. Flickner *et al.*, Query by image and video content: the QBIC system, *Computer* 28(9) (1995) 23–32.

[29] Available at: http://www.virage.com/

[30] Available at: http://www-white.media.mit.edu/vismod/demos/photobook/

[31] A. Pentland, R. Picard and S. Sclaroff, Photobook: tools for content-based manipulation of image databases, *International Journal of Computer Vision* 18(3) (1996) 233–254.

[32] Available at: http://www.excalib.com/products/vrw/vrw.html

[33] P.G.B. Enser and C.G. McGregor, *Analysis of Visual Information Retrieval Queries* (British Library Research and Development Report 6104) (British Library, London, 1992).

[34] L.H. Armitage and P.G.B. Enser, Analysis of user need in image archives, *Journal of Information Science* 23(4) (1997) 287–299.

[35] L.H. Armitage and P.G.B. Enser, *Information Need in the Visual Document Domain* (British Library Research and Innovation Report 27) (British Library, London, 1996).

[36] L.H. Keister, User types and queries: impact on image access systems. In: R. Fidel *et al.* (eds), *Challenges in Indexing Electronic Text and Images* (ASIS Monograph Series) (Learned Information, Medford, NJ, 1994), pp. 7–22.

[37] S.K. Hastings, Query categories in a study of intellectual access to digitized art images, *ASIS '95: Proceedings of the 58th ASIS Annual Meeting* 32 (1995) 3–8.

[38] R. Hidderley *et al.*, Capturing iconology: a study in retrieval modelling and image indexing. In: M. Collier and K. Arnold (eds), *Proceedings of the Third International Conference on Electronic Library and Visual Information Research (ELVIRA3)* (De Montfort University, Milton Keynes, 1996), pp. 123–135.

[39] R. Fidel, The image retrieval task: implications for the design and evaluation of image databases, *The New Review of Hypermedia and Multimedia* (1997) 181–199.

[40] S.K. Hastings, Evaluation of image retrieval systems: role of user feedback. In: [7], pp. 438–452.

[41] M. Markkula and E. Sormunen, Searching for photos – journalists' practices in pictorial IR. In: J.P. Eakins, D.J. Harper and J. Jose (eds), *The Challenge of Image Retrieval: Papers presented at a Workshop on Image Retrieval, 5 February 1998, University of Northumbria at Newcastle, Newcastle upon Tyne, UK* (University of Northumbria at Newcastle, 1998).

[42] M. Markkula and E. Sormunen, End-user searching challenges indexing practices in the digital newspaper photo archive, *Information Retrieval* 1(4) (1999) 259–286.

[43] S. Ornager, The newspaper image database: empirical supported analysis of users' typology and word association clusters. In: *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, 9–13 July 1995* (1995), pp. 212–218.

[44] S. Ornager, Image retrieval: theoretical and empirical user studies on accessing information in images. In: *ASIS' 97: Proceedings of the 60th ASIS Annual Meeting* 34 (1997) 202–211.

[45] J.P. Eakins, Pictorial information systems – prospects and problems. In: A. McEnery (ed.), *Proceedings of the 14th British Computer Society Information Retrieval Specialist Group Research Colloquium, University of Lancaster, 13–14 April 1992* (British Computer Society, London, 1992), pp. 102–123.

[46] A. Vailaya, A. Jain and H.J. Zhang, On image classification: city images vs. landscapes, *Pattern Recognition* 31(12) (1998) 1921–1935.

[47] G.W. Horgan, M. Talbot and J.C. Davey, Towards automatic recognition of plant varieties. In: J.P. Eakins, D.J. Harper and J. Jose (eds), *The Challenge of Image Retrieval: Papers presented at a Workshop on Image Retrieval, 5 February 1998, University of Northumbria at Newcastle, Newcastle upon Tyne, UK* (University of Northumbria at Newcastle, 1998).

[48] S-F. Chang *et al.*, Semantic visual templates: linking visual features to semantics. In: *ICIP'98: Proceedings of the 1998 International Conference on Image Processing Chicago, Illinois, October 4–7, 1998* (IEEE Computer Society, Los Alamitos, CA, 1998), pp. 531–535.

[49] H.D. Wactlar *et al.*, Intelligent access to digital video: the Informedia project, *Computer* 29(5) (1996) 46–52.

[50] Available at: http://www.informedia.cs.cmu.edu/

Downloaded from http://jis.sagepub.com at PENNSYLVANIA STATE UNIV on April 17, 2008
© 2000 Chartered Institute of Library and Information Professionals. All rights reserved. Not for commercial use or unauthorized distribution.