The evolution of visual information retrieval

Peter Enser

Journal of Information Science 2008 34: 531 originally published online 13 June 2008
DOI: 10.1177/0165551508091013

JIS celebrating 5 years

The evolution of visual information retrieval

Peter Enser

University of Brighton, United Kingdom

Abstract.

This paper seeks to provide a brief overview of those developments which have taken the theory and practice of image and video retrieval into the digital age. Drawing on a voluminous literature, the context in which visual information retrieval takes place is followed by a consideration of the conceptual and practical challenges posed by the representation and recovery of visual material on the basis of its semantic content. An historical account of research endeavours in content-based retrieval, directed towards the automation of these operations in digital image scenarios, provides the main thrust of the paper. Finally, a look forwards locates visual information retrieval research within the wider context of content-based multimedia retrieval.

Keywords: visual information retrieval; image retrieval; video retrieval; semantic image retrieval; content-based retrieval

1. The context of visual information retrieval

The retrieval of images or image sequences that are relevant to a query is a long-established activity which has evolved quite remarkably during the last 50 years, from the special preserve of a relatively few professional practitioners to the forefront of research in computer vision and a leading edge domestic application of information technology. This extension of traditional information retrieval activity includes both still and moving images, the former usually characterized in the literature as 'image retrieval', the latter as 'video retrieval', and the two in combination, sometimes, as 'visual information retrieval' [1,2].

The literature of visual information retrieval has grown at a stupendous rate. To quote Jörgensen, in her landmark text within the field:

> Adjectives such as 'vast' are often applied to the various literatures … related to image processing, but even this designation is an understatement [3: p. 199].

More remarkable still is the fact that almost all of that growth has taken place since the early 1990s, and reflects those technological advances which brought the digital image to the attention of the computer scientist. Greatly increased availability of images via the Internet, then via mobile platforms, and most recently as an aspect of the social networking phenomenon, has been said to place us

Correspondence to: Professor P.G.B. Enser, School of Computing, Mathematical and Information Sciences, University of Brighton, Watts Building, Moulsecoomb, Brighton BN2 4GJ, UK.
Email: peter.enser@btinternet.com

> on the hinge of an important historical swing back towards to what may be called the primacy of the image [3: p. ix].

Jörgensen's observation reflects the huge upsurge in image and video retrieval activity by the general public, which finds expression in such diverse activities as searching for visual materials using search engines such as Google Images (http://images.google.co.uk/) or a social networking facility such as YouTube (http://uk.youtube.com/), the browsing of online television broadcast archives, and the recovery of images from increasingly voluminous personal stores of digital photographs.

Visual images exist in a wide variety of forms, but it is those whose features can be captured and/or viewed by unenhanced human vision, encountered typically as photographs or artwork, which have predominated in the literature of image retrieval. The curatorial or commercial imperative to collect other types of still image, including those the features of which must be captured and/or viewed by means of equipment which expands the range of human vision, such as microscopes, telescopes and electronic imaging devices, has been less pronounced. In part this is because some classes of image, notably in the medical, architectural and engineering domains, tend to occur as adjuncts to parent records, and it is these parent records which are usually the object of retrieval, rather than the images themselves. However, researchers in medicine – and in defence and criminology – came to an early realization that images within those domains must be treated as important information objects in their own right, rather than mere appendices to other database information, leading to the formation of specialized collections for research and training purposes [3: p. 139].

The technology to support the display of sequences of images in rapid succession in order to create the illusion of moving imagery has only given rise to collections of film and video material in more recent times. Because of their scale and growth rate, however, such collections have also figured significantly in the literature of visual information retrieval, and the locating of that activity within the wider context of multimedia retrieval.

In the pre-digital era, requests of varying degrees of urgency would be addressed to image repositories in the form of telephone calls, written specifications and sometimes by the presence of the client in person. Where necessary, the repository's picture researchers would act as mediators, seeking to introduce greater precision into the natural language requests, perhaps translating them into the terminology of a controlled vocabulary associated with the repository's classification scheme, and helping the client towards an explicit articulation of the mental image for which the client was seeking some physical realization [4]. In other words, this mediation process exactly paralleled that of the reference librarian in a traditional library, except that it was conducted among hanging files or archival boxes stuffed with monochrome prints and colour transparencies, backed up by a store of negatives, in a scenario engagingly captured, albeit with some dramatic license, in Stephen Poliakoff's television play *Shooting the Past* [5].

The success with which material appropriate to a client's request could be extracted from such stores reflected the picture researcher's knowledge of the collection, and of the classification and indexing practices adopted by the repository; it also reflected the researcher's judgement, based on visual inspection of any candidate images.

Film and video libraries presented a different appearance, their shelves laden with tins containing reels of film, the chemical properties of which called for special knowledge and a controlled environment [6]. Prior to viewing, determination of the potential relevance of complete films was assisted by the short synopses which sometimes augmented their catalogue records. The retrieval of image sequences, as opposed to whole films, was more challenging. Protracted viewing of material in order to make selections might be assisted by time-coded listings of each shot within a film, but the compilation of such tools was itself a highly labour-intensive operation, the undertaking of which reflected a clear commercial imperative.

In general, although a number of cataloguing standards have been developed for image and film material, and are comprehensively described by the Technical Advisory Service for Images (TASI) [7], visual asset management has lacked the adherence to universal standards of cataloguing and classification which characterized traditional library practice with text-based material. In large

measure this reflected the problems posed in attempting to capture in indexing language the semantic content of images. These problems have been a recurring theme in the literature of image and video retrieval, and an understanding of their nature is central to an appreciation of the evolution of visual information retrieval.

2. Image indexing

Greisdorf and O'Connor [8] and Jörgensen [3: pp. 7–68), in particular, have drawn on the literature of cognitive psychology to assist our understanding of how humans interact with images. They describe an initial physiological response to the visual primitives of colour, texture and the spatial distribution of blobs and regions within an image. This perception of the syntactic content of the image is rapidly overtaken by cognitive reasoning about the semantic content in the form of objects, activities and scenes. This is followed by high-level, inductive interpretation of the wider semantic context in which the image is located, which brings into play the viewer's subjective belief system. Analysis of user responses to images, whilst revealing 'wildly differing assessments' of particular images, found that

> user assertions about interactions with pictures … form a richer descriptive palette than ordinary indexing [9].

The principles and practice of image indexing by means of which semantic content can be represented have been the subject of comprehensive reviews [3,10,11], together with a variety of other contributions, notably [12–27]. This literature provided the backcloth to the increasingly elaborate conceptual frameworks which came to be built as a means of informing the image indexing process.

The simplest such frameworks recognized three levels, which corresponded with visual primitives (colour, texture, shapes), logical or 'derived' features (objects, activities, events) and inductive interpretation (abstract features) [8,26]. A more developed model, which has figured quite prominently in the literature, rests on the formal analysis of Renaissance art images by the art historian Panofsky [28], who recognized primary subject matter ('pre-iconography') which required no interpretative skill; secondary subject matter ('iconography'), which did call for an interpretation to be placed on the image; and tertiary subject matter, denoted 'iconology', embracing the intrinsic meaning of the image, and demanding of the viewer high-level semantic inferencing.

Shatford [14] was instrumental in generalizing Panofsky's analysis, simplifying the first two modes in terms of 'generic' and 'specific', and amplifying these by distinguishing between what a picture is 'of' and what it is 'about'. The notion of 'generic', 'specific' and 'abstract' semantic content has since figured prominently in the literature, the more developed formulations containing multiple levels, comprising both syntactic or pre-conceptual visual content, to which are added semantic layers of interpretive attributes which invoke the viewer's inferential reasoning about the local object and global scenic content of the image [29,30].

Most recently, the basic level theory expounded by Rosch et al. [31], together with extensions to a facet analysis of the subject attributes of an image [16], have been combined in a more developed form of conceptual model which gives explicit recognition to the combination of semantic content and context in image material [22]. Hare et al. [32] have shown how the keywords allocated by expert indexers to a museum's image collection can be mapped to this rich conceptual model.

Whatever the level of sophistication attained by conceptual models, the manual indexing of images has remained a matter of trying to represent visually encoded semantic content in a verbal surrogate. The problematic nature of this translation process found expression in Markey's observation that individual differences in image perception give rise to 'extraordinary idiosyncrasy' in the assignment of image terms [12], and Hogan et al.'s observation that

> If an image carries a great deal of information for the user which is dependent on contextual and situational factors, the assumption that meaning rests in a pre-defined set of subject terms is of limited utility to control access to the contents of an image base [33].

Besser [34] had already noted that

> Historically, text-based intellectual access systems have been woefully inadequate for describing the multitude of access points from which the user might try to recall the image

and Svenonius [17] went on to state that

> it is useless to attempt to point to unspeakable reality with an index term

because subject indexing presupposes that what is depicted can be named, whereas there are messages addressed to our visual and aural perceptions, the content of which cannot be named.

Arguably, the semantic indexing of film and video poses even greater challenges. The task has to address different levels of semantic structure, from the frame, through shot and scene to the film or video stream as a complete entity, together with other semantically coherent sequences in the form of clips, episodes and news stories [2: pp. 10, 35, 36]. In the absence of exhaustive shot lists the minimalist nature of synopses in standard sources of reference makes them particularly blunt instruments for leveraging the full semantic content of these forms of information object.

The phenomenon of social tagging has brought a new dimension to the representation of the semantic content of visual materials. Exemplified in such products as Flickr (http://www.flickr.com/) and YouTube (http://uk.youtube.com/), the ability to contribute personal tags to image and video metadata challenges the supremacy of professionally sourced, authoratitive subject representation, whilst introducing opportunities for beneficial enhancement of both exhaustivity and specificity in subject indexing.

3. Analysis of user needs

In an attempt to gain insight into effective indexing practice, a rich vein of enquiry was opened up in the 1980s, directed at the analysis of users' needs for images and image sequences. The oft-quoted observation

> The delight and frustration of pictorial resources is that a picture can mean different things to different people [14],

amplified by a recognition that a picture can mean different things to the same person at different times or under different circumstances, provided the platform for these endeavours. Observations such as Falconer's [37], that the subjects most often sought by a particular archive's clients fell into a 'no-man's land of categories' which could not be adequately or precisely classified by any existing system, and Besser's [34], that the retrieval utility of an image is inherently unpredictable, led naturally to the conclusion that the appropriate level of indexing exhaustivity is indeterminate, and that subject indexing is of low utility [19]. Only in those scenarios where the clients are well-defined and their needs well-understood could the negative impact of this unpredictability be lessened.

Jörgensen [3: p. 127] reviewed a number of these user studies, the most widely cited of which analysed some 2700 requests addressed by a variety of client types to the Hulton Deutsch collection – a major, general-purpose picture archive (now part of Getty Images) [4,38]. A preliminary analysis of these requests revealed a very wide variation in subject foci and terminological specificity, and also that the majority of the requests were for specific objects or events, frequently 'refined' by spatial, temporal or other combinations of facets. A number of other studies subsequently confirmed the relatively high incidence of requests for specific, named features [27,39–42], whilst other studies reported quite different user behaviour in which emphasis was placed on more generic or affective visual features [43–48].

The behavioural patterns exhibited in these studies ranged across a number of application areas, including art history [39,42], journalism [18,27], and medicine [43]. In combination, all these user studies contributed to a perception that the further removed the image retrieval scenario is from the scenario of a specialist archival collection, with expert mediation and an experienced user, the lower is the significance of carefully constructed metadata, and the greater is the significance of browsing facilities. Coupled with the latter came a developing appreciation of the significance of relevance feedback, with users able to interact freely with displayed output, using their innate

capacity to perceive at a glance the potential interest of an image [49–51]. Fidel [23] captured the essence of this argument in an important paper which described a continuum of image searching tasks, at the extremities of which were an 'object pole' and a 'data pole'. The former referred to the situation where interest lay in retrieving a specific image identified, for example, by title, whereas the latter denoted the need to retrieve the data or information portrayed by the image. Any particular image might satisfy a number of different requests located at various points along this continuum, but recognition of the two poles carried significant implications for the design and evaluation of image retrieval systems. Towards the 'object pole' relevance becomes more difficult to determine, which lends added emphasis to browsing facilities; conversely, relevance feedback increases in importance towards the 'data pole'. It was towards this latter pole that the image retrieval research community increasingly turned their attention.

Whereas earlier user studies involved the collection of image requests recorded manually by image archive staff, and often reflecting some degree of expert mediation by them, the increasing incidence of Web-based, end-user searching of image collections has generated unmediated requests culled automatically from transaction logs. Studies of these have been able to analyse very much larger numbers of requests, as in the case of the analysis by Goodrum and Spink [52] of the transaction logs of over 33,000 image requests submitted to the Excite search engine (http://search.excite.com).

Web-enabled access to digitized image collections, whether through general search engines such as Google (http://images.google.co.uk) and Yahoo! (http://images.search.yahoo.com), specialized image search engines such as Picsearch (http://www.picsearch.com) or collection-specific search engines, brought about a revolution in image retrieval. The factors involved in the design and implementation of web image search engines were discussed by Kherfi et al. [53], who identified a need for more advanced tools to enhance retrieval performance. The user's search behaviour has also been the subject of study: Smeulders et al. [54] proposed a useful categorization which recognized 'target search', 'category search' and 'search by association' (corresponding to the 'text-based', 'subject-based' and 'browsing' labels used in an earlier study [39]). The first of these aims at a specific image, identified by title or other unique identifier, and conforms with Fidel's [23] notion of the 'object pole'. In a 'category search' the client requests images which feature some particular semantic content at the local or global level. In a 'search by association' the client may approach an image collection with no particular semantic content requirement in mind, and is content to browse in order to retrieve images by serendipity.

The evidence available thus far about Web-based searching of image collections points to the increased significance of the 'search by association' relative to the 'category search', which has been the traditional focus of effort in the professional practice of image retrieval [48,52,55]. In reality, little intelligence has been gathered on user interaction with the vast array of visual resources made available, either freely by search engines or in password-protected repositories, in the Web environment. Roddy's [56: p. 48] observation in 1991 that one of the great failures of image access was its inability to provide reliable information on a typical search session was thought by Jörgensen [3: p. 129) to remain true over a decade later, and it seems to the present author that the situation has not changed greatly in the interim.

In comparison with studies of users' needs for still image content, search requests and behaviour in the context of film and video material has received comparatively little attention. Studies involving archival film collections have been reported [35,57,58], but a fully comprehensive study of user interaction with moving images is still awaited.

4. Towards content-based image retrieval

The first milestones along the development path which led to content-based image retrieval (CBIR) were encountered in the late 1970s in the form of databases constructed specifically for picture storage and retrieval [51,54]. Tony Cawkell's [59] detailed analysis of the design factors involved in their construction provides a good insight into the state-of-the-art as it had evolved by the early 1990s. The earliest attempts at image database construction were characterized by the difficulties encountered

in attempting integration of image data and relational database structures [60–67]. By the beginning of the 1990s, however, Besser [34] was able to report on the benefit of clients being able to browse screen displays of thumbnail images without recourse to library personnel and without physical handling of the images themselves, in products such as Imagequery, characterized as the marriage of a standard text-oriented online library catalogue with a powerful image browsing mechanism [33].

Substantial development of image databases followed, usefully surveyed in [68], with digitized images co-located with their metadata, albeit with widely varying levels of adherence to a number of different image metadata standards. Notwithstanding the advances made since the early 1990s in digital visual asset management systems, the traditional paradigm of image retrieval remained that of textual string matching between the client's verbal search request statement and the subject annotations embedded within the image collection metadata. By this time, however, the image retrieval research community had perceived the need to

> relinquish the idea of the utility of using words to index non-verbal understanding … We are looking for alternative ways of image retrieval, ways that are less dependent on familiarity with existing taxonomies and their assigned authorities [33].

There was a complementary wish to reduce dependency on the collection knowledge locked into the heads of the curators of image collections. A compelling case for this was made by the Challenger space shuttle explosion in 1986, in the aftermath of which there was an urgent requirement to retrieve from NASA's huge visual archive all possibly relevant images depicting the Challenger launch sequence and the failed booster rocket. The manual retrieval system, 'highly dependent on the corporate memories of a few dedicated individuals', could not meet this requirement [69]. Such perceptions highlighted the significance of a workshop organized by the National Science Foundation in 1992 to identify major research areas in visual information management systems, with emphasis on interactive image understanding in such applications as medical images and satellite images [54]. Shortly afterwards, the Mosaic Internet browser was released and the manifest difficulties associated with manual indexing of visual images placed in sharp relief the need for indexing tools appropriate for Web-enabled access to digital archives. Thus was fuelled some 15 years of intense research activity directed towards the CBIR paradigm.

The term 'content-based image retrieval' derives from the fact that the CBIR paradigm operates on the explicit content of the digitized image, which is its pixel domain. There are those within the professional image practitioner community who, like Hyvönen et al. [70], have expressed some scepticism about the 'content-based' label, arguing that the content of an image lies in the semantic inferences to which it leads the viewer, and which may be explicitly represented in textual metadata.

In the early stages of CBIR the focus was on syntactic operations conducted on the pixel domain of the digitized image, in order to generate visual feature vectors as surrogates of the image. The elements of these vectors were generated automatically from analysis of the quantifiable attributes, such as colour, texture and geometry, present within the pixel domain. Initially, the feature vectors took the form of global descriptors using relatively simple formulations such as colour histograms [71,72]. The query was similarly surrogated as a picture-by-example, usually a digitized image, although early forms of sketch retrieval system were also reported [73]. Similarity analysis was conducted between the query and the image collection, typically using histogram intersection techniques, leading to the retrieval of candidate images in decreasing order of similarity with the query.

Nurtured by an increasingly engaged research community, feature vectors rapidly grew in sophistication. Colour correlograms captured information about spatial layout of colour that could not be described using colour histograms; combinations of neighbouring pixels ('texels') underpinned textural analysis of images; pixel intensity transformations such as wavelet analysis proved effective at edge detection, as a means of determining an object's shape; and other advanced techniques were developed, capable of segmenting the image into multiple regions, or detecting features from salient regions within an image. Early reviews of these automatic indexing techniques were published by Idris & Panchanathan [74] and Eakins & Graham [36], together with an accessible review of techniques for colour, texture and shape by Forsyth [75] within a special issue of *Library Trends*, edited by Sandore [76], devoted to progress in visual information access and retrieval. Del Bimbo's monograph [2]

provided an authoratitive technical treatment, and was succeeded by other comprehensive reviews [3: pp. 149–154, 54, 77, 78].

The 1990s saw the launch of a number of experimental CBIR systems, one of the earliest of which, and certainly the best-known, was QBIC (Query By Image Content) [79]. Other systems, including Blobworld, Excalibur, MARS, Photobook and VisualSeek followed; comprehensively surveyed in [53,80], a comparative evaluation was also undertaken [81]. The Benchathlon network (http://www.benchathlon.net/) was established with the aim of developing benchmarking facilities in support of the experimental CBIR environment.

The CBIR paradigm and the experimental systems which it spawned had been responsible for a marked upsurge in the rate of publication about image indexing and retrieval after 1990 [82], and it was with some reluctance that the image retrieval research community responded to the view that an image retrieval paradigm which operates on the low-level, syntactic properties of an image had limited practical value [83]. The information science community, in contrast, had reached that view somewhat earlier, informed by experiments with specific illustration tasks which used similarity perceptions in a real work context [27] and by tests on CBIR features which revealed that users did not find these low-level features either intuitive to search or relevant to their queries [84]. Fidel's [23] concern, that much research effort and financial resources were being invested in improving CBIR without an awareness of the situations in which such retrieval might be useful, was echoed in the view that

> the emphasis in the computer science literature has been largely on what is computationally possible, and not on discovering whether essential generic visual primitives can in fact facilitate image retrieval in 'real-world' applications [3: p. 197].

Indeed, none of the commercial CBIR systems launched in the first half of the 1990s achieved significant market penetration, and all have since ceased to be actively promoted [85].

Typical of the dangers of forsaking semantic integrity in the retrieval of images were observations that a colour-based CBIR algorithm will match busy city scenes containing beige brick backgrounds with scenes of desert sand [86], and a shape-based one might return images of the Statue of Liberty in response to queries seeking images of starfish – the so-called 'rhyming image' phenomenon [80]. Nevertheless, such algorithms were shown to have real value in situations where it is difficult for the perceptual saliency of some visual features to be captured in text, such as the perceptual elements of a texture, the outline of a form and the visual effects in a video sequence [2: p. 4]. In a comprehensive survey of the principles and practice of CBIR towards the end of the 1990s examples were provided of specialized applications – in medicine, fine art and textile design, for example – where the verbalization challenge was so great that CBIR provided the only effective solution [26].

5. Towards semantic image retrieval

For more traditional image retrieval applications, however, 'semantic image retrieval' and the 'semantic gap' began to penetrate the literature from the mid 1990s onwards, with Gudivada and Raghavan [87], in a special issue of *IEEE Computer* devoted to CBIR systems, and Aigrain et al. [88], in a state-of-the-art review of CBIR one year later, introducing a publishing surge which drew the observation

> while content-based image retrieval papers published prior to 1990 are rare, almost certainly obsolete, and of little direct impact today, the number of papers published since 1997 is just breathtaking [54].

The semantic gap is that rift in the image retrieval landscape between the information that can be extracted automatically from a digitized image and the interpretation that humans might place upon the image [54]. Early endeavours to bridge the semantic gap saw effort directed at the automatic identification of objects and scenes, undertaken either as a statistical classification procedure or as a knowledge-based recognition task [77]. The latter approach necessitated the construction of a model for each type of object of interest, which acted as the comparator in searches of each image in the collection,

looking for regions similar to the models. The earliest approaches envisaged digitized reference images which depicted the object in a variety of light conditions, and at different angles, sizes and perspectives. Limited success has been achieved in automatic scene classification and object recognition, although one example of the latter – naked people – has been usefully applied to the automatic detection of pornographic images [89]. In general, however, the domain knowledge/effort involved in building the models is very considerable, leading to some scepticism that the problems of updating and extending complex model-based approaches to cover more than a 'toy subset of object classes' will prove insuperable unless some form of adaptive learning is employed [77].

By the late 1990s, it had become clear that the semantic gap could not be bridged by operations on the pixel domain alone, and that CBIR should be treated as a complement to, rather than a replacement for, text-based image retrieval [51]. Henceforth, the integration of the two paradigms became a significant focus of attention, especially in those application areas, such as investigative medicine, which generate enormous quantities of continuous-process visual data [90]. Automatic annotation of images came to the fore as a means of trying to achieve that integration.

In an overview of automatic annotation techniques, Hare et al. [91] note two basic approaches; one seeking to discover links between regions and words by statistical inference [92], and the other using a supervised learning technique which echoes document vector analysis in text-based information retrieval [93]. In this second case a training set of annotated images is used, each image surrogated as a textual term vector, the elements of which represent the allocation of keywords drawn from the indexing vocabulary. To this is appended a 'visual term' vector, with elements drawn from the image's quantized visual primitives. The 'dimensionality curse' [83] of the matrix formed from these stacked textual-and-visual term vectors called for a data reduction technique, which was found in latent semantic indexing (LSI), a procedure borrowed from the traditional theoretical model of text retrieval [94]. Vectors of visual terms from a test set of un-annotated images are compared with the visual term constituents of the training set, and where a sufficiently high level of similarity is encountered between a pair of images, one drawn from the training set, the other from the test set, the annotation associated with the former is propagated to the latter in the form of automatically assigned object/scene/activity labels.

Typically, experimentation in automatic annotation has been conducted using training sets of images derived from small, ground-truth image databases, where both the exhaustivity and specificity of the indexing has been low [22]. When compared with the rich semantic indexing typical of professionally managed image collections, these limited-vocabulary experimental scenarios appear unrealistic, and the precision of their results has tended to be erratic.

The limited perception of objects and scenes permitted by these highly constrained vocabularies combines with another disadvantage of automatic annotation techniques, which is their dependency on search engines which can only be trained to recognize features actually visible in the image. A peculiarity of visual images, however, is their ability to convey messages independently of visually perceived reality. Some of the facets which contribute to the rich conceptual model of image semantic content described earlier in this paper have no visual presence; they represent 'extrinsic semantics'. This has been shown, for example, in analyses of image perception, where the majority of the terms viewers used to describe the contents of a set of images were not visibly present in the images [8], and in user studies within the practitioner environment where requests very frequently incorporated non-visible facets [22]. Experimentation has failed, thus far, to provide any reliable evidence that automatic annotation can span the very considerable conceptual distance between object/scene/activity labelling and the high-level reasoning which situates those objects, scenes or activites appropriately within the user's sociocognitive space.

Partly in recognition of this, the CBIR research community has demonstrated a rapidly developing interest in semantic web technologies in general, and ontologies in particular. Schreiber et al. [95], working in the medical domain, were among the earliest proponents of ontologies for image annotation and retrieval, and interest has extended to experimentation in the generation of semantic inferencing rules, formulated by medical domain experts, that link low-level visual features to domain concepts [96,97]. Other applications have been reported in the cultural heritage sector [98–100].

As a result of the adoption of ontologically supported experimental image retrieval processes, tools which are well-established within the professional image management environment, such as the *Union List of Artist Names (ULAN)*, *Thesaurus of Geographic Names (TGN)*, *Art and Architecture Thesaurus (AAT), ICONCLASS* and *WordNet* are beginning to penetrate the research environment, where they are treated as quasi-ontologies [98]. They make a welcome appearance – one which would not have been foreseen a few years ago – at the research frontiers of image retrieval.

Nevertheless, the challenge of semantic image retrieval remains daunting. Ontology construction – albeit assisted by the adoption of standard knowledge organization and representation tools – is technically demanding, and ontologies tend to be domain-specific. One approach to enhancing functionality within ontology-supported semantic image retrieval systems has been lexical expansion through the harnessing of multiple vocabularies [99,101]. Such approaches had their origin in query expansion using thesaural relations in text retrieval. Hollink [102: pp.90–94] has shown how diminishing returns set in under different combinations and degrees of propagation of such relations: there may be no counterpart in the visual image to the semantic relationships which link terms at the lexical level, leading to the danger of automatically adding wholly inappropriate terms to an image's subject metadata.

It seems clear that the widest reaches of the semantic gap cannot be spanned using current techniques. At the present time, most attempts at bridging the semantic gap have faltered at the very broad separation between object labelling and the high-level reasoning which situates those objects appropriately within the viewer's sociocognitive space. In effect, the semantic gap is a two-part fracture, and the focus of attention has been on the first part alone [91].

6. Content-based video retrieval

Although the still image was the early focus of attention among the CBIR research community, once digitization and transmission of the moving image became a viable proposition the realization was reached that a spatio-temporal distribution of blobs was an easier target for syntactic analysis of the pixel domain than the spatial distribution offered by the still image:

> Video comes as a sequence, so what moves together most likely forms an entity in real life, so segmentation of video is intrinsically simpler than a still image [103].

Automatic segmentation of a video stream into shots using shot boundary detection techniques was an early focus of attention, a detailed technical treatment of which was provided in [2: pp. 203–264]. Each automatically detected shot makes available a set of frames, from among which keyframes are selected on some consistent basis to act as surrogates of the shot. Other useful operations on video sequences followed from the developing robustness of automatic shot boundary detection. A chronological ordering of keyframes enabled 'storyboards' to be formulated, which acted as surrogates for the entire film or video sequence [86]. Where a set of keyframes representative of every shot would generate too much data for efficient analysis, and clips, scenes and episodes formed significant semantic units, video segmentation into shot aggregates was developed [2: pp. 224–229].

For the searcher, cataloguer, programme compiler and editor, storyboarding techniques offered considerable savings in time, especially in those cases where the fast detection of highlights is valued, as in sports and news broadcasts. A further advantage for the searcher was the high probability that shots within a storyboard which were adjacent to a shot which had been deemed relevant to a query would have a close temporal relationship with that shot, and would be likely also to be relevant.

Rapid strides were made in the development of techniques for visual feature extraction, indexing, searching, browsing and summarization in video, with comprehensive reviews by Naphade and Smith [104] and Smeaton [105] making significant contributions to the literature. Initially, these techniques operated only on the visual content of video; more recently, research effort has been directed towards the full audiovisual content of this visual resource. Capabilities in automatic

speech recognition (ASR) have advanced to the stage where the audio channel can be harnessed as a means of generating textual annotations to the complementary video channel. The best recognizers, trained for broadcast news, currently have a word error rate of about 15% on studio recorded anchor speech; naturally, performance degrades as constraints on identified speaker and comprehensiveness of vocabulary are relaxed [106].

The Informedia digital video library project, begun in 1994, is the most widely reported example of this approach [107]. Operating on news stories from television broadcasts, this landmark project seeks to make such material searchable by means of ASR-enabled transcripts, and the integration of speech recognition, natural language processing, image analysis and information retrieval [108]. Automatically generated metadata and indexes to multiple terabytes of video are continuously available online to local users. A very similar system, called Físchlár-News, automatically analyses the nightly Irish broadcast television news [109]. Both systems enable the user to inspect keyframes, play the associated video, and conduct other browsing and retrieval operations. Analytic functions also include speech/music discrimination, programme start/end identification, TV advertisement detection, and automatic detection of anchorperson shots. The outputs of these analyses are fed into a trained statistical classifier which segments the broadcast into discrete news stories which are then available as units of retrieval.

The Informedia project, in particular, has generated a wealth of experimental results which point to speech transcripts providing the single most important clue for successful video retrieval [86]. More advanced techniques reflecting research in computer vision have not, as yet, proved robust enough to be usable, and Hauptmann's recent advice to the research community is

> give up on general, deep understanding of video – that problem is just too hard for now [106].

Instead, he has argued that a few thousand high-level semantic concepts that have reasonably reliable detection accuracy can be combined to yield high-accuracy automatic video retrieval. Retrieval experiments, using sets of rich intermediate semantic descriptors derived from a standard lexicon and taxonomy, have provided support for such an approach [110].

The research effort in content-based video retrieval has been characterized, and stimulated, by the emphasis placed on the evaluation of techniques through the medium of the annual TRECVID benchmarking event. This video track offshoot of the Text REtrieval Conference began in 2001, and provides participating organizations with a large video test collection, embracing corpora which range from documentaries to advertising films and broadcast news [111–113]. A further stimulant has been the development of the MPEG-4 and MPEG-7 multimedia representation standards. MPEG-4 was designed to provide technological elements which enabled the production, distribution and content access paradigms of interactive multimedia, mobile multimedia, interactive graphics and enhanced digital television to be integrated [114]. Its provision of shape-based encoding of natural scene video has excited the interest of the image processing research community [105], but MPEG-4 offers limited capability for the interpretation of semantic content. MPEG-7 was born of a realization of that limitation, and has the ability to describe both low-level features and high-level semantics, together with structural aspects of any multimedia document or file. Modalities may include still pictures, graphics, 3D models, audio, speech, video, and composition information about how these elements are combined in a multimedia presentation [114]. The Físchlár-News system, to which reference was made earlier, automatically analyses and structures the broadcast into an MPEG-7 annotation [109].

7. Conclusion

In this paper, an attempt has been made to draw, necessarily in a highly selective fashion, on the literature of image and video retrieval in order to outline the development of theory and practice in this absorbing variant of information retrieval. Within the last twenty years, reflecting the rapid burgeoning of interest among the computer vision research community, that body of literature has grown prodigiously, and its character has been described elsewhere as relentlessly abstruse [11]. The effect

has been to create a communication gap between the researcher and professional practitioner communities in image retrieval, a separation which was first surveyed by Cawkell [49] in the form of two minimally linked citation-interconnected clusters derived from an analysis of pre-1991 publications. That separation has widened considerably in the intervening years.

With the intention of providing a forum where members of both research and practitioner communities could become better informed about each other's endeavours and environments, a conference was hosted by the Institute of Image Data Research (IIDR), at the University of Northumbria at Newcastle-upon-Tyne, UK in 1998. This proved to be the precursor to the annual *Challenge of Image and Video Retrieval International Conference (CIVR)* series (http://www.civr.org), now an official ACM conference and generally recognized as a key event in the visual information retrieval research community's calendar, but one which retains the specific brief of bringing that community together with the practitioner community

> to illuminate critical issues and energize both communities for the continuing exploration of novel directions for image and video retrieval [115].

Heralded in 2000 by the inclusion of the term 'video' in the title of what had previously been the *Challenge of Image Retrieval* conference, a shift may be detected in the focus of CIVR, and in the literature of visual information retrieval more generally, towards the end of the period reviewed in this paper, however. Fuelled by the seemingly greater capabilities of CBIR with video than with still images, and by a dawning appreciation of the exceptional difficulty in spanning the wider reaches of the semantic gap, beyond which lie the high-level semantic spaces inhabited by the majority of image practitioners and users, visual information retrieval is being subsumed within multimedia retrieval. Other fora – such as the ACM SIGMM Workshop on Multimedia Information Retrieval (http://www.liacs.nl/~mir), the IEEE International Conference on Multimedia and Expo (ICME) and the International Cultural Heritage Informatics Meeting (ICHIM) – vie with CIVR in a content-based information retrieval environment admirably surveyed by Lew et al [116]. From this paper's brief survey of recent activity in the landscape of visual information retrieval activity, the capture, representation and retrieval of semantic content in the visual medium has presented the practitioner and researcher alike with some difficult terrain. The way ahead, then, amid the broader landscape of content-based multimedia information retrieval, promises to be an exhilarating climb.

Acknowledgements

I am grateful to the referees for their helpful comments.

References

[1] A. Gupta and R.C. Jain, Visual information retrieval, *Communications of the ACM* 40(5) (1997) 71–79.
[2] A. Del Bimbo, *Visual Information Retrieval* (Morgan Kaufmann, San Francisco, 1999).
[3] C. Jörgensen, *Image Retrieval: Theory and Research* (The Scarecrow Press, Lanham, MD, 2003).
[4] P.G.B. Enser and C.G. McGregor, *Analysis of Visual Information Retrieval Queries, Report on Project G16412 to the British Library Research & Development Department* (British Library R&D Report 6104, British Library, London, 1992).
[5] S. Poliakoff, *Shooting the Past* (DVD, BBC, London, 1999).
[6] C. Cochrane, The collection, preservation and use of moving images in the United Kingdom, *Audiovisual Librarian* 20(2) (1994) 122–130.
[7] TASI (Technical Advisory Service for Images), *Putting Things in Order: Links to Metadata Schemas and Related Standards (2006)*. Available at: http://www.tasi.ac.uk/resources/schemas.html (accessed 31 December 2007).
[8] H. Greisdorf and B. O'Connor, Modelling what users see when they look at images: a cognitive viewpoint, *Journal of Documentation* 58(1) (2002) 6–29.

[9] B.C. O'Connor and M.K. O'Connor, Categories, photographs and predicaments: exploratory research on representing pictures for access, *Bulletin of the American Society for Information Science* 25(6) (1999) 17–20.
[10] E.M. Rasmussen, Indexing images. In: M.E. Williams (ed.), *Annual Review of Information Science and Technology 32*, (Information Today, Inc., Medford, NJ, 1997) 169–196.
[11] P.G.B. Enser, Visual image retrieval. In: B. Cronin (ed.) *Annual Review of Information Science and Technology 42*, (Information Today, Inc., Medford, NJ, 2008) 3–42.
[12] 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, (1984) 155–177.
[13] S. Shatford, Describing a picture: a thousand words are seldom cost effective, *Cataloging & Classification Quarterly* 4(4) (1984) 13–30.
[14] S. Shatford, Analyzing the subject of a picture: a theoretical approach, *Cataloguing & Classification Quarterly* 5(3) (1986) 39–61.
[15] A.E. Cawkell, *Indexing collections of electronic images: a review* (British Library Research Review 15, British Library, London, 1993).
[16] S. Shatford Layne, Some issues in the indexing of images, *Journal of the American Society for Information Science* 45(8) (1994) 583–588.
[17] 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.
[18] S. Ørnager, The newspaper image database: empirical supported analysis of users' typology and word association clusters. In: E. Fox, P. Ingwersen and R. Fidel (eds), *Proceedings of the 18th Annual Special Interest Group Conference on Research and Development in Information Retrieval (ACM SIGIR '95)*, (ACM Press, New York, 1995) 212–218.
[19] P.G.B. Enser, Progress in documentation: pictorial information retrieval, *Journal of Documentation* 51(2) (1995) 126–170.
[20] P.G.B. Enser, Visual image retrieval: seeking the alliance of concept-based and content-based paradigms, *Journal of Information Science* 26(4) (2000) 199–210.
[21] R. Hidderley, P. Brown, M. Menzies, D. Rankine, S. Rollason and M. Wilding, 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)*, (Aslib, London, 1995) 79–91.
[22] P.G.B. Enser, C.J. Sandom, J.S. Hare and P.H. Lewis, Facing the reality of semantic image retrieval, *Journal of Documentation* 63(4) (2007) 465–481.
[23] R. Fidel, The image retrieval task: implications for the design and evaluation of image databases, *The New Review of Hypermedia and Multimedia* 3 (1997). 181–199.
[24] M.C. Krause, Intellectual problems of indexing picture collections, *Audiovisual Librarian* 14(2) (1998) 73–81.
[25] F.W. Lancaster, *Indexing and Abstracting in Theory and Practice (3rd edn)* (Facet, London, 2003).
[26] J.P. Eakins and M.E. Graham, *Content-based image retrieval: a report to the JISC Technology Applications Programme.* Available at: http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc (accessed 31 December 2007).
[27] M. Markkula and E. Sormunen, End-user searching challenges indexing practices in the digital newspaper photo archive, *Information Retrieval* 1(4) (2000) 259–285.
[28] E. Panofsky, *Studies in Iconology* (Harper & Row, New York, 1962).
[29] A. Jaimes and S.-F. Chang, A conceptual framework for indexing visual information at multiple levels. In: G. Beretta & R. Schettini (eds), *Proceedings of the First IS&T/SPIE Internet Imaging Conference* (SPIE vol. 3964, SPIE, Bellingham, WA, 2000) 2–15.
[30] C. Jörgensen, A. Jaimes, A.B. Benitez and S-F. Chang, A conceptual framework and empirical research for classifying visual descriptors, *Journal of the American Society for Information Science and Technology* 52(11) (2001) 938–947.
[31] E. Rosch, C. Mervis, W. Gray, D. Johnson and P. Boyes-Braem, Basic objects in natural categories, *Cognitive Psychology* 8(3) (1976) 382–439.
[32] J.S. Hare, P.H. Lewis, P.G.B. Enser and C.J. Sandom, Semantic facets: an in-depth analysis of a semantic image retrieval system. In: N. Sebe and M. Worring (eds), *Proceedings of the Sixth ACM International Conference on Image and Video Retrieval* (ACM, New York, 2007)
[33] M. Hogan, C. Jörgensen and P. Jörgensen, The visual thesaurus in a hypermedia environment: a preliminary exploration of conceptual issues and applications. In: D. Bearman (ed.), *Hypermedia and Interactivity in Museums: Proceedings of an International Conference* (Archives & Museum Informatics, Pittsburgh, PA, 1991) 202–221.

[34] H. Besser, Visual access to visual images: the UC Berkeley image database project, *Library Trends* 38(4) (1990) 787–798.
[35] J. Turner, Representing and accessing information in the stockshot database at the National Film Board of Canada, *Canadian Journal of Information Science* 15(4) (1990) 1–22.
[36] J. Yang and A.G. Hauptmann, Annotating news video with locations. In: H. Sundaram, M. Naphade, J.R. Smith and R.Yong (eds), *Proceedings of the Sixth International Conference on Image and Video Retrieval (CIVR 2006)* (Lecture Notes in Computer Science, Vol. 4071, Springer, Berlin, 2006) 153–162.
[37] J. Falconer, The cataloguing and indexing of the photographic collection of the Royal Commonwealth Society, *The Indexer* 14(1) (1984) 15–22.
[38] P.G.B. Enser, Query analysis in a visual information retrieval context, *Journal of Document and Text Management* 1(1) (1993) 25–52.
[39] S.K. Hastings, Query categories in a study of intellectual access to digitized art images. In: T. Kinney (ed.), *Proceedings of the 58th Annual Meeting of the American Society for Information Science (ASIS'95)* (ASIS, Silver Spring, MD 1995) 3–8.
[40] C. Gordon, Patterns of user queries in an ICONCLASS database, *Visual Resources* 12 (1996) 177–186.
[41] L.H. Armitage and P.G.B. Enser, Analysis of user need in image archives, *Journal of Information Science* 23(4) (1997) 287–299.
[42] H. Chen, An analysis of image queries in the field of art history, *Journal of the American Society for Information Science and Technology* 52(3) (2001) 260–273.
[43] L.H. Keister, User types and queries: impact on image access systems. In: R. Fidel, T.B. Hahn, E.M. Rasmussen and P.J. Smith (eds), *Challenges in Indexing Electronic Text and Images* (ASIS Monograph Series, Learned Information Inc., Medford, NJ, 1994) 7–22.
[44] C. Jörgensen, Image attributes in describing tasks: an investigation, *Information Processing and Management* 34(2/3) (1998) 161–174.
[45] C.O. Frost and A. Noakes, Browsing images using broad classification categories. In: E.K. Jacob (ed.), *Proceedings of the Ninth ASIS SIGCR Classification Research Workshop* (ASIS, Silver Spring, MD, 1998) 71–89.
[46] Y. Choi and E.M. Rasmussen, Users' relevance criteria in image retrieval in American history, *Information Processing and Management* 38(5) (2002) 695–726.
[47] L. Hollink, A.Th. Schreiber,.B.J. Wielinga and M. Worring, Classification of user image descriptions, *International Journal of Human Computer Studies* 61(5) (2004) 601–621.
[48] C. Jörgensen and P. Jörgensen, Image querying by image professionals, *Journal of the American Society for Information Science and Technology* 56(12) (2005) 1346–1359.
[49] A.E. Cawkell, Selected aspects of image processing and management: review and future prospects, *Journal of Information Science* 18(3) (1992) 179–192.
[50] S. Santini and R.C. Jain, Do images mean anything? In: *Proceedings of the IEEE International Conference on Image Processing (ICIP-97*) (IEEE, New York, 1997) 564–567.
[51] Y. Rui, T.S. Huang and S. Mehrotra, Relevance feedback techniques in interactive content-based image retrieval. In: I.K. Sethi and R.C. Jain (eds), *Proceedings of the Sixth SPIE Conference on Storage and Retrieval for Image and Video Databases* (SPIE vol. 3312, SPIE, Bellingham, WA, 1997) 25–36.
[52] A. Goodrum and A. Spink, Image searching on the excite web search engine, *Information Processing and Management* 37(2) (2001) 295–311.
[53] M.L. Kherfi, D. Ziou and A. Bernardi, Image retrieval from the World Wide Web: issues, techniques, and systems, *ACM Computing Surveys* 36(1) (2004) 35–67.
[54] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta and R.C. Jain, Content-based retrieval at the end of the early years, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12) (2000) 1349–1380.
[55] A. Goodrum, M. Bejune and A.C. Siochi, A state transition analysis of image search patterns on the Web. In: E.M. Bakker, T.S. Huang, M.S. Lew, N. Sebe and X. Zhou (eds.): *Proceedings of the Second International Conference on Image and Video Retrieval (CIVR 2003)* (Lecture Notes in Computer Science, Vol. 2728, Springer, Berlin, 2003) 281–290.
[56] K. Roddy, Subject access to visual resources: what the 90s might portend, *Library Hi Tech* 9(1) (1991) 45–49.
[57] C.J. Sandom and P.G.B. Enser, *VIRAMI: Visual Information Retrieval for Archival Moving Imagery* (Library and Information Commission Research Report 129, The Council for Museums, Archives and Libraries, London, 2002).
[58] M. Hertzum, Requests for information from a film archive: a case study of multimedia retrieval, *Journal of Documentation* 59(2) (2003) 168–186.

[59] A.E. Cawkell, Picture-queries and picture databases, *Journal of Information Science* 19(6) (1993) 409–423.
[60] S.-K. Chang and T. Kunii, Pictorial database systems, *IEEE Computer Magazine Special Issue on Pictorial Information Systems* 14(11) (1981) 13–21.
[61] S.-K. Chang and S.-H. Liu, Picture indexing and abstraction techniques for pictorial databases, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(4) (1984) 475–484.
[62] H. Tamura and N.Yokoya, Image database systems: a survey, *Pattern recognition* 171(1) (1984) 29–43.
[63] G. Nagy, Image database, *Image and Vision Computing* 3(3) (1985) 111–117.
[64] T.L. Kunii (ed.), *Visual Database Systems* (Elsevier, Amsterdam, 1989).
[65] L.F. Lunin, An overview of electronic image information, *Optical Information Systems* 10(3) (1990) 114–130.
[66] C.H.C. Leung, Architecture of an image database system, *Information Services and Use* 10 (1990) 391–397.
[67] S.E. Arnold, The large data construct: a new frontier in database design, *Microcomputers for Information Management* 7(3) (1990) 185–203.
[68] 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* (Elsevier, Amsterdam, 1997) 11–28.
[69] G.A. Seloff, Automated access to the NASA-JSC image archives, *Library Trends* 38(4) (1990) 682–696.
[70] E. Hyvönen, A. Styrman and S. Saarela, *Ontology-based Image Retrieval* (HIIT Publications Number 2002–03, Helsinki Institute for Information Technology, Helsinki, Finland, 2002) 15–27.
[71] M.J. Swain and D.H. Ballard, Color indexing, *International Journal of Computer Vision* 7(1) (1991) 11–32.
[72] J.R. Smith and S.F. Chang, Querying by color regions using the VisualSEEK content-based visual query system. In: M.T. Maybury (ed.), *Intelligent Multimedia Information Retrieval* (AAAI Press, Menlo Park, CA, 1997) 23–41.
[73] T. Kato and T. Kurita, Visual interaction with electronic art gallery. In: A. Min Tjoa and R.Wagner (eds), *Proceedings of the International Conference in Database and Expert Systems Applications (DEXA'90)* (Springer, London, 1990) 234–240.
[74] F. Idris and S. Panchanathan, Review of image and video indexing techniques, *Journal of Visual Communication and Image Representation* 8(2) (1997) 146–166.
[75] D.A. Forsyth, Computer vision tools for finding images and video sequences, *Library Trends* 48(2) (1999) 326–355.
[76] B. Sandore (ed), *Library Trends* 48(2), 1999, 283–524.
[77] J.P. Eakins, Towards intelligent image retrieval, *Pattern Recognition* 35(1) (2002) 3–14.
[78] R. Datta, J. Li and J.Z. Wang, Content-based image retrieval – approaches and trends of the new age. In: H. Zhang, J. Smith and Q. Tian (eds), *Proceedings of the Seventh ACM SIGMM International Workshop on Multimedia Information Retrieval, (MIR 2005)* (ACM, New York, 2005) 253–262.
[79] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D.Steele and P. Yanker, P., Query by image and video content: the QBIC system, *IEEE Computer Magazine* 28(9) (1995) 23–32.
[80] B. Johansson, *A survey on: contents based search in image databases (2000)*. Available at: http://www.cvl.isy.liu.se/ScOut/TechRep/PaperInfo/bj2000.html (accessed 31 December 2007).
[81] R.C. Veltkamp and M. Tanase, *Content-based Image Retrieval Systems: a Survey (2000)*. Available at: http://citeseer.ist.psu.edu/373932.html (accessed 6 January 2008).
[82] H. Chu, Research in image indexing and retrieval as reflected in the literature, *Journal of the American Society for Information Science and Technology* 52(12) (2001) 1011–1018.
[83] 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* (Elsevier, Amsterdam, 1997) 100–117.
[84] S-F. Chang, J.R. Smith, M. Beigi and A. Benitez, Visual information retrieval from large distributed online repositories, *Communications of the ACM* 40(12) (1997) 63–71.
[85] J.P. Eakins, Content-based image retrieval – what's holding it back? In: *Proceedings of the Eighth Annual Conference of the Advanced School for Computing and Imaging* (ASCI, Delft, 2002).
[86] M.G. Christel and R.M. Conescu, Addressing the challenge of visual information access from digital image and video libraries. In: *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries* (ACM, New York, 2005) 69–78.

[87] V.N. Gudivada and V.V. Raghavan, Content-based image retrieval systems, *IEEE Computer Magazine* 28(9) (1995) 18–22.
[88] P. Aigrain, H.-J. Zhang and D. Petkovic, Content-based representation and retrieval of visual media: a state of the art review, *Multimedia Tools and Applications* 3 (1996) 179–202.
[89] D.A. Forsyth and M.M. Fleck, Automatic detection of human nudes, *International Journal of Computer Vision* 32(1) (1999) 63–77.
[90] H. Müller, N. Michoux, D. Bandon and A. Geissbuhler, A review of content-based image retrieval systems in medical applications – clinical benefits and future directions, *International Journal of Medical Informatics* 73 (2004) 1–23.
[91] J.S. Hare, P.H. Lewis, P.G.B. Enser and C.J. Sandom, Mind the gap: another look at the problem of the semantic gap in image retrieval. In: E.Y. Chang, A. Hanjalic and N. Sebe (eds), *Proceedings of the 2006 SPIE Conference on Multimedia Content Analysis, Management and Retrieval* (SPIE Vol. 6073, SPIE, Bellingham, WA, 2006) 1–12.
[92] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D.M. Blei and M.I. Jordan, Matching words and pictures, *Journal of Machine Learning Research* 3 (2003) 1107–1135.
[93] G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval* (McGraw-Hill, New York, 1983).
[94] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer and R. Harshman, Indexing by latent semantic analysis, *Journal of the American Society for Information Science* 41(6) (1990) 391–407.
[95] G. Schreiber, B. Dubbeldam, J. Wielemaker and B. Wielinga, Ontology-based photo annotation, *IEEE Intelligent Systems* 16(3) (2001) 2–10.
[96] B. Hu, S. Dasmahapatra, P. Lewis and N. Shadbolt, N., Ontology-based medical image annotation with description logics. In: *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence* (IEEE, New York, 2003) 77–82.
[97] L. Hollink, S. Little and J. Hunter, Evaluating the application of semantic inferencing rules to image annotation. In: *Proceedings of the 3rd International Conference on Knowledge Capture* (ACM, New York, 2005) 91–98.
[98] L. Hollink, A.Th. Schreiber, J. Wielemaker and B.J. Wielinga, Semantic annotation of image collections. In: S. Handschuh, M. Koivunen, R. Dieng and S. Staab (eds), *Proceedings of the K-Cap 2003 Workshop on Knowledge Markup and Semantic Annotation* (ACM, New York, 2003) 41–48.
[99] E. Hyvönen, M. Salminen, M. Junnila and S. Kettula, A content creation process for the semantic web. In: *Proceedings of the 2004 LREC Workshop on Ontologies and Lexical Resources in Distributed Environments* (2004). Available at: http://www.seco.tkk.fi/publications/2004/hyvonen-salminen-et-al-a-content-creation-process-2004.pdf (accessed 12 March 2008).
[100] M. J. Addis, K. Martinez, P. Lewis, J. Stevenson and F. Giorgini, New ways to search, navigate and use multimedia museum collections over the web. In: J. Trant and D. Bearman, (eds), *Proceedings of the 2005 Conference on Museums and the Web* (2005). Available at: http://www.archimuse.com/mw2005/papers/addis/addis.html (accessed on 31 December 2007).
[101] A. Amin, van M. Assem, de V. Boer, L. Hardamn, M. HildeBrand, L. Hollink, van J. Kersen, B. Omelayenko, van J. Ossenbruggen., A.B. Schreiber, R. Siebes, J. Taekema, J. Wielemaker and B. Wielinga, *MultimediaN E-Culture Demonstrator: Objectives and Architecture* (Technical Report BSIK, MultimediaN Project, Subproject N9C "Pilot E-Culture", CWI, Amsterdam, DEN, The Hague, 2006).
[102] L. Hollink, *Semantic Annotation for Retrieval of Visual Resources* (SIKS Dissertation Series No. 2006–24, Vrije Universiteit, Amsterdam, 2006).
[103] N. Sebe, M.S. Lew, X. Zhou, T.S. Huang and E.M. Bakker, The state of the art in image and video retrieval. In: E.M. Bakker, T.S. Huang, M.S. Lew, N. Sebe and X. Zhou (eds), *Proceedings of the Second International Conference on Image and Video Retrieval* (Lecture Notes in Computer Science, Vol. 2728, Springer, Berlin, 2003) 7–12.
[104] M.R. Naphade and J.R. Smith, On the detection of semantic concepts at TRECVID. In: *Proceedings of the Twelfth Annual ACM International Conference on Multimedia* (ACM Press, New York, 2004) 660–667
[105] A.F. Smeaton, Indexing, browsing and searching of digital video. In: B. Cronin (ed.), *Annual Review of Information Science & Technology 38(1)* (Information Today, Inc., Medford, NJ, 2004) 371–407.
[106] A.G. Hauptmann, Lessons for the future from a decade of Informedia video analysis research. In: W-K. Leow, M.S. Lew, T-S.Chua, W-Y. Ma, L. Chaisorn and E.M. Bakker (eds), *Proceedings of the Fourth International Conference on Image and Video* (Lecture Notes in Computer Science, Vol. 3568, Springer, Berlin, 2005) 1–10.
[107] Carnegie Mellon University, *Informedia Digital Video Library (2006)*. Available at: http://www.informedia.cs.cmu.edu/ (accessed 31 December 2007).

[108] M.J. Whitbrock and A.G. Hauptmann, Speech recognition for a digital video library, *Journal of the American Society for Information Science* 49(7) (1998) 619–632.
[109] H. Lee, A.F. Smeaton, N.E. O'Connor and B. Smyth, User evaluation of Físchlár-News: an automatic broadcast news delivery system, *ACM Transactions on Information Systems* 24(2) (2006) 145–189.
[110] A. Hauptmann, R. Tan and W-H. Lin, How many high-level concepts will fill the semantic gap in news video retrieval? In: N. Sebe and M. Worring (eds), *Proceedings of the 2007 ACM International Conference on Image and Video Retrieval* (ACM, New York, 2007).
[111] A.F. Smeaton, Large scale evaluations of multimedia information retrieval: the TRECVid experience. In: W-K. Leow, M.S. Lew, T-S. Chua, W-Y. Ma, L. Chaisorn, L. and E.M. Bakker (eds), *Proceedings of the Fourth International Conference on Image and Video Retrieval* (Lecture Notes in Computer Science, Vol. 3568, Springer, Berlin, 2005) 11–17.
[112] M.G. Christel and R.M. Conescu, Mining novice user activity with TRECVID interactive retrieval tasks. In: H. Sundaram, M. Naphade, J.R. Smith and R. Yong (eds), *Proceedings of the Fifth International Conference on Image and Video Retrieval* (Lecture Notes in Computer Science Vol. 4071, Springer Berlin, 2006) 21–30.
[113] A.F. Smeaton, P. Over and W. Kraaij, Evaluation campaigns and TRECVid. In: *Proceedings of the Eighth ACM International Workshop on Multimedia Information Retrieval*, (ACM, New York, 2006) 321–330.
[114] J.M. Martinez (ed.), *MPEG-7 Overview (version 10) (ISO/IEC JTC1/SC29/WG11) (2004)* (International Organisation for Standardisation, Palma, Mallorca, 2004). Available at: http://www.chiariglione.org/MPEG/standards/mpeg-7/mpeg-7.htm (accessed 31 December 2007).
[115] H. Sundaram, M. Naphade, J.R. Smith and R.Yong (eds), *Proceedings of the Fifth International Conference on Image and Video Retrieval* (Lecture Notes in Computer Science Vol. 4071, Springer Berlin, 2006).
[116] M.S. Lew, N. Sebe, C. Djeraba and R. Jain, Content-based multimedia information retrieval: state of the art and challenges, *ACM Transactions on Multimedia Computing, Communications and Applications* 2(1) (2006) 1–19.