Surveying the Reality of Semantic Image Retrieval

Peter G.B. Enser[1], Christine J. Sandom[1], and Paul H. Lewis[2]

[1] School of Computing, Mathematical and Information Sciences,
University of Brighton
{p.g.b.enser, c.sandom}@bton.ac.uk
[2] Department of Electronics and Computer Science,
University of Southampton
phl@ecs.soton.ac.uk

Abstract. An ongoing project is described which seeks to add to our understanding about the real challenge of semantic image retrieval. Consideration is given to the plurality of types of still image, a taxonomy for which is presented as a framework within which to show examples of real 'semantic' requests and the textual metadata by which such requests might be addressed. The specificity of subject indexing and underpinning domain knowledge which is necessary in order to assist in the realization of semantic content is noted. The potential for that semantic content to be represented and recovered using CBIR techniques is discussed.

1 Introduction

Within the broad church of visual image users and practitioners there has been only a minimal engagement with the endeavours of the research community in visual image retrieval. Correspondingly, those endeavours have been little informed by the needs of real users or the logistics of the management of large scale image collections.

With the developing maturity of these research endeavours has come a realisation of the limitations of CBIR processes in practice, however. These limitations reflect the fact that the retrieval utility of visual images is generally realised in terms of their inferred semantic content. The context for this inferential reasoning process is to be found in the distinction drawn in semiotics between the denotation, or presented form, of the image and the connotation(s) to which it gives rise. It is clear that personal knowledge and experience, cultural conditioning and collective memory contribute towards that reasoning process. The CBIR community has attached the label 'semantic image retrieval' to the formulation and resolution of information needs which engage that intellectual process. Within the three-level hierarchy of perception postulated by Eakins & Graham [1], semantic image retrieval subsumes retrieval by 'derived features' and 'abstract attributes'. The sharply drawn distinction between those retrieval processes and the automatic extraction of low level features from denotative pixel structures is characterised as the 'semantic gap' [2]. A useful listing of research endeavours which have addressed the three levels of perception may be found in [3].

We are engaged upon a project, the aims of which are to provide both research and practitioner communities with a better-informed view of the incidence and

S. Bres and R. Laurini (Eds.): VISUAL 2005, LNCS 3736, pp. 177 – 188, 2005.

significance of the semantic gap in the context of still images; to seek enhanced functionality in image retrieval by bridging that gap; and to design and evaluate an experimental visual image retrieval system which incorporates the construction of such a bridge.

2 A Taxonomy of Images

This paper is concerned with the first of the aims described above. To this end a broad spectrum of operational image retrieval activity has been surveyed in order to take account of the full plurality of types of image and user. This survey has been informed by, and has in turn informed, an image taxonomy, which is shown in diagrammatic form in Figure 1.

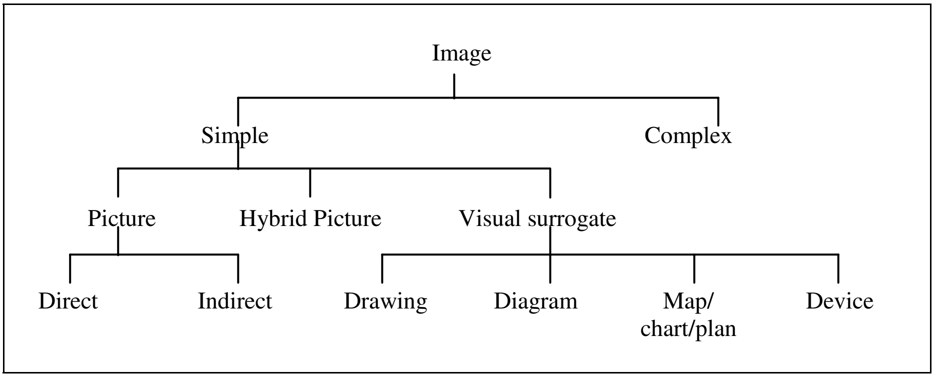


Fig. 1. A taxonomy of still images

The following definitions have been used in the construction of this taxonomy:

Image a two-dimensional visual artefact.
Simple Image an undifferentiated image.
Complex Image an image which comprises a set of simple images.
Picture a scenic or otherwise integrated assembly of visual features.
Hybrid Picture a picture with integral text.
Visual surrogate non-scenic, definitional visual artefact.
Direct Picture a picture, the features of which can be captured and/or viewed within the human visible spectrum.
Indirect Picture a picture, the features of which must be captured and/or viewed by means of equipment which extends viewability beyond the human visible spectrum.
Drawing an accurate representation (possibly to scale) of an object, typical applications being in architecture and engineering.

Diagram a representation of the form, function or workings of an object or process.

Map/chart/plan a representation (possibly to scale) of spatial data, typical applications being in geography, astronomy, and meteorology.

Device a symbol or set of symbols which uniquely identifies an entity, e.g., trademark, logo, emblem, fingerprint.

3 Representing and Retrieving the Semantic Content of Different Types of Image

We have been fortunate in securing the cooperation of a number of significant picture archives and libraries in the provision of sample requests addressed to their holdings, together with the images selected manually or automatically in response to those requests. We have also collected the metadata associated with each such image.

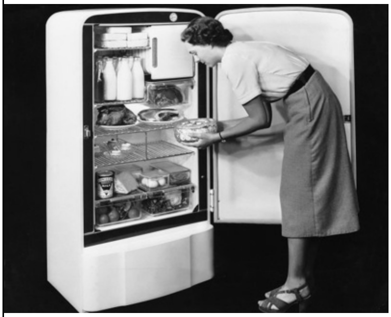
For each of the leaf nodes pertaining to a Simple Image in Figure 1 a sample image drawn from our test collection is shown below. For each image the actual request in response to which the image was selected is shown. The intention is to provide an indication of the way in which real semantic retrieval requirements are expressed, across the spectrum of image types.

Also shown is the textual subject annotation associated with each image, by means of which text-matching with the request was achieved. The specificity of subject indexing (and underpinning domain knowledge) which is necessary in order to assist in the realisation of semantic content is noteworthy. Finally, consideration is given to the potential functionality of CBIR techniques with respect to the type of image, and type of query, shown.

3.1 Direct Picture

Whether captured by photographic process or created by human endeavour, the Direct Picture is that form of image with which the majority of literature concerned with the

Request: *A photo of a 1950s fridge*



Roomy Fridge © Getty Images

Subject Metadata: [8]

Title	Roomy Fridge
Date	circa 1952
Description	An English Electric 76A Refrigerator with an internal storage capacity of 7.6 cubic feet, a substantial increase on the standard model.
Subject	Domestic Life
Keywords	black & white, format landscape, Europe, Britain, England, appliance, kitchen appliance, food, drink, single, female, bending

indexing and retrieval of still images has been concerned. A comprehensive overview of this literature may be found in [4], and further evidence of the predominance of this form of image may be found in Trant's survey of image databases available wholly or in part on the Web [5]. The Corel image data set has been a heavily-used resource of this type in CBIR research projects, [e.g., 6] whilst the Brodatz album [7] has figured prominently in the particular context of textural analysis.

Recovering the Desired Semantic Content

In the image perceptual hierarchy described in section 1 above, the lowest level is thought to involve the engendering of a visual impression by the sensory stimuli, which is first cognitively matched to some form of syntactic equivalence [3]. The two higher levels "… require both interpretation of perceptual cues and application of a general level of knowledge or inference from that knowledge to name the attribute" [9].

The process by which perceptual and interpretive matter in an image is recognised is, as yet, an incompletely understood cognitive phenomenon [3,10]. Following Marr [11], it would seem reasonable to suggest that low-level features within the image might have a role to play, however. Shape may be especially significant, complemented by colour and texture, bringing to bear a previously learned linguistic identifier to generate meaning.

Whatever the perceptual processes involved it would seem to be the case that *identification* is dependent upon the prior existence – and knowledge by the user – of a defining linguistic label. Studies of users of archival and documentary images and footage, in particular, which have revealed the emphasis placed on the recovery of images which depict features (persons, objects, events) uniquely identified by proper name [e.g., 12-15]. Note the example above makes reference to a specific manufacturer and model of the depicted object, whilst enabling requests at the more generic levels of refrigerator or fridge and kitchen appliance to be satisfied. We note also the prevalence of qualification (or 'refinement') [12] in requests, which can only be satisfied by textual annotation; e.g. the request for a 1950s fridge.

Furthermore, the process of identification often involves *context*, recognition of which would seem to invoke high-level cognitive analysis supported by domain and tacit knowledge (see the Subject identifier in the above example). In general, contextual anchorage is an important role played by textual annotation within the image metadata.

A yet more pressing need for supporting textual metadata occurs when the *significance* of some visual feature is at issue. Significance is an attribute which is unlikely to have any visual presence in the image. Often reflecting the reason for the image having been created in the first place, and recording the first, last or only instantiation of some object or event, it is frequently encountered in both indexing and querying of image collections [12-17], and involves interpretive processing completely removed from CBIR functionality.

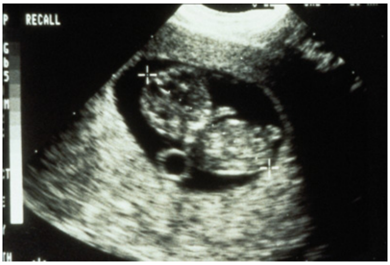
When the focus of interest lies with the abstract or affective content of the image, the client wanting images of suffering or happiness, for example, CBIR techniques might offer limited potential - colour can be an effective communicator of mood, for example – but the appropriate cognitive response may be dependent on the presence within metadata of an appropriate textual cue which conditions our *interpretation* of the image.

Overall, a realistic assessment of the potential of CBIR techniques with Direct Pictures, in operational as opposed to artificial, laboratory-based environments, lends support to the image research community's developing interest in the integration of text-based and CBIR indexing and retrieval strategies, as exemplified in techniques which are designed to uncover the latent correlation between low-level visual features and high-level semantics.

3.2 Indirect Picture

Reported usage of the Indirect Picture in the image retrieval literature is most frequently encountered in the field of medicine, where variants of this form of image include x-rays and ultrasound/MRI/CRT scans. Other domains in which Indirect Pictures may be sought include molecular biology, optical astronomy, archaeology and picture conservation. Our research to date has shown that demand for the retrieval of Indirect Pictures held within organized collections is most frequently encountered in publishing and educational contexts. In medical practice, for example, such images are most frequently stored as adjuncts to a specific patient's record, and it is the latter – as opposed to an image - which is the subject of a retrieval request

Request: *We are trying to source images for a new pregnancy timeline that we are developing and are finding it very difficult to get any decent ultrasound images.*



Wellcome Photo Library

Subject Metadata: [18]

Title	Fetal development
Short Description.	Fetal development
Description of image content	Fetal development Ultrasound scan, showing fetus

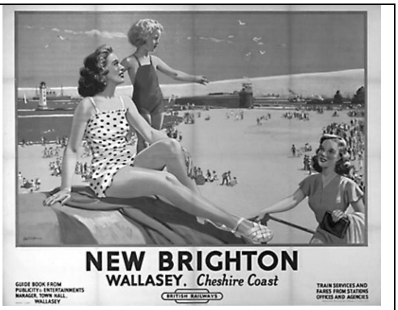
Recovering the Desired Semantic Content

The above example demonstrates the need for mediated domain knowledge in forging a relationship between 'fetal [*sic*] development' and 'pregnancy'. In terms of the CBIR potential, however, such images do offer significant opportunities for the semi-automatic detection of certain conditions or objects based upon colour, texture or spatial properties; a number of applications in the medical domain have been reported [e. g. 19].

3.3 Hybrid Picture

Pictures with integral text are frequently encountered in the form of posters and other advertisements, and in the form of cartoons. Importantly, the integral text identifies a significant component of the semantic content of the image and may be the focus of a request for which the image is deemed relevant.

Request: *A British Rail poster circa 1955. Advertising New Brighton, Wallasey and Cheshire Coast. Woman in foreground - Mersey in background*


NEW BRIGHTON
WALLASEY. Cheshire Coast
GUIDE BOOK FROM PUBLICITY & ENTERTAINMENTS MANAGER, TOWN HALL, WALLASEY
BRITISH RAILWAYS
TRAIN SERVICES AND FARES FROM STATIONS OFFICES AND AGENCIES

NRM – Pictorial Collection

Subject Metadata: [20]

Title	'New Brighton - Wallasey, Cheshire Coast', BR (LMR) poster, 1949.
Subject	PLACES > Europe: UK > Merseyside
Caption	Poster produced for British Railways (BR), London Midland Region (LMR), promoting rail travel to the beaches of New Brighton - a district of Wallesey - on the River Mersey estuary. A woman and child are shown sunbathing, with the beach, sea and Mersey seen in the background. Artwork by George S Dixon. Printed by London Lithographic Co, London, SE5. Dimensions: 1010mm x 1270mm
Keywords	1940s, 20th Century, Ads, Advertisements, Advertising, Art, Bathing, Bathing costumes, Beaches, Brighton, British, British Railway, Children, Coast, Costume, Design, Dixon, Geo S, Fashion, Fashion, 1940s, Girls, Graphic, Graphic design, Happiness, Holidaymakers, Holidays, Leisure, Merseyside, New, New Brighton, Poster, Poster art, Railway, Railway poster, Recreation, Resort, Sea, Seaside, Social, Summer, SUNBATHING, Swimming, Swimming costumes, Swimsuits, The 1940s (1945-1949, Tourism, Tourist, United Kingdom, Wallasey, Woman, Women

Recovering the Desired Semantic Content

For image retrieval purposes, the integral text in such an image will normally be transcribed into the metadata, as shown in the above example. The need for this indexing effort could be mitigated by the use of automatic detection of embedded text [e.g., 21], but in general the same limitations of CBIR processing hold for this type of image as for the Direct Picture.

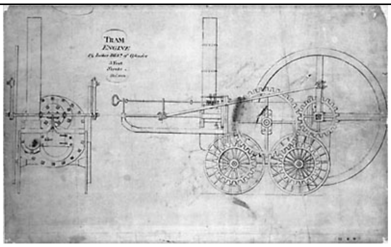
3.4 Drawing

At this point in our research we have only encountered this type of image in the context of museum or archival collections; thisis clearly reflected in the textual annotation associated with the example image.

Recovering the Desired Semantic Content

Indications are that requests for images of this type reflect the need for visual information about highly specific objects, often identified by creator name or created object; the identification has to be resolved by means of textual metadata. However, given the absence of the foreground/background disambiguation problem in this type of image, the potential for recovering (the generic content of) such images by means of shape detection techniques has been recognized; [e.g., 22].

Request: *John Llewellyn's drawing of a Trevithick locomotive*



National Railway Museum

Subject Metadata [20]

Title	Trevithick's tram engine, December 1803.
Subject	TRANSPORT > Locomotives, Rolling Stock & Vehicles > Locomotives, Steam, Pre-1829
Caption	Drawing believed to have been made by John Llewellyn of Pen-y-darran. Found by FP Smith in 1862 and given by him to William Menelaus. Richard Trevithick (1771-1833) was the first to use high pressured steam to drive an engine. Until 1800, the weakness of existing boilers had restricted all engines to being atmospheric ones. Trevithick set about making a cylindrical boiler which could withstand steam at higher pressures. This new engine was well suited to driving vehicles. In 1804, Trevithick was responsible for the first successful railway locomotive.
Keywords	19th Century, Drawing, Engine, F, Industrial Revolution (1780-18, John, Llewellyn, Llewellyn, John, Locomotive, Locomotives, Steam, Pre-1829, Menelaus, William, Menelaus, P, Pen-Y-Darran, Pre-1829, Richard, Smith, Smith, F P, Steam, Tram, Tram engines, Trevithick, Trevithick, Richard, United Kingdom, Wale, William

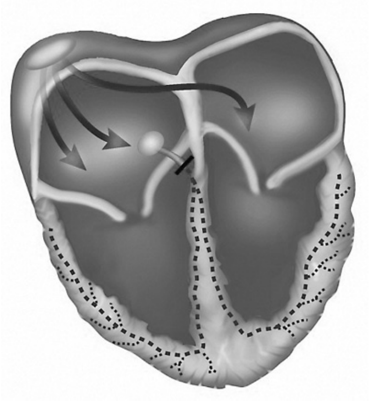
3.5 Diagram

Diagrams may be encountered in a number of formats and applications, and may incorporate textual or other symbolic data. The semantic content will often be reflected in the title and will frequently be the form in which the request is cast.

Recovering the Desired Semantic Content

The example above illustrates the need for mediation in order to establish the pertinence of a specific entitled diagram to some condition or focus of interest. It would seem clear that CBIR techniques incorporating textual support, possibly in the form of ontologies or embedded text detection, could have some potential to address retrieval of this type of image.

Request*: the adverse health effects of space travel, specifically long periods of zero gravity … weakening of the heart*



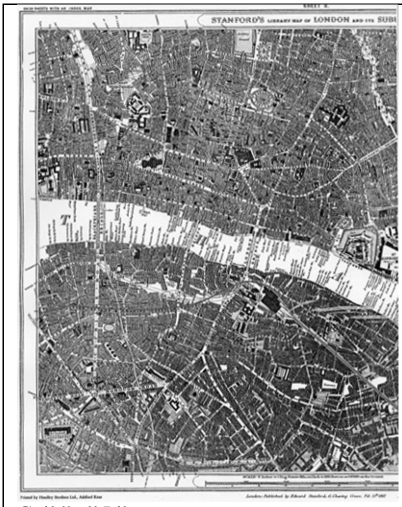
Wellcome Photo Library

Subject Metadata: [18]

Title	Heart block
Description	Heart block Colour artwork of cut-away heart, showing right and left ventricles with diagrammatic representation of a right bundle block, usually caused by strain on the right ventricle as in pulmonary hypertension
ICD Code	426.9

3.6 Map/Chart/Plan

To date, collections of this type of image to which we have gained access focus on historic spatial data, and this is reflected in the types of request addressed to such collections. The requests are necessarily cast in the form of linguistic search statements, to which textual matching with metadata and, as the example below shows, intellectual mediation must be applied.



Guildhall Library

Request: …. *a map of central London before 1940. I wish to discover where was Redcross Street, Barbican E.C.1. The current London A-Z does not list any similar address in the E C 1 area.*

Subject Metadata [23]

Title	Stanfords Library Map of London and its suburbs/ Edward Stanford, 6 Charing Cross Road
Physical description	1; map; line engraving; 1842 x 1604 mm
Notes	Extent: Crouch End – Canning Town – Mitcham – Hammersmith. Title in t. border. Imprint and scale in b. border. Hungerford and Lambeth bridges shown as intended. Exhibition buildings shown in Kensington.

Recovering the Desired Semantic Content
Consideration of CBIR potential in this case would again indicate the possible value of embedded text detection, possibly coupled with shape recognition of cartographic symbols; also colour and textural analysis of image features. Some CBIR applications were reported in [1].

3.7 Device

Such images may integrate a number of visual features, but are most likely to be requested by the identifying title of the device or by picture example.

Recovering the Desired Semantic Content
As in the case of drawings, such images tend to be context free and, as such, may lend themselves to shape recognition techniques. Significant applications in trade mark matching have been reported [1,24], as has experimental work in fabric design pattern matching [1].

Request: CRESTS: London, Brighton and South Coast Railway

National Railway Museum

Subject Metadata [20]

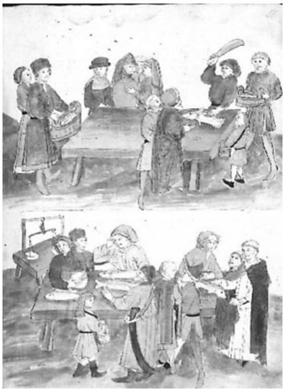
Title	Coat of arms of the Southern Railway on a hexagonal panel, 1823-1947.
Subject	TRANSPORT > Railway Heraldry > Coats of Arms
Caption	The coat of arms of the Southern Railway features a dragon and a horse on either side of a shield.
Keywords	20th Century, Arm, Coat, Coat of arms, Coats of arms, Dragon, Horse, Industrial Revolution (1780-18, LOCO, Rail travel, Railway, Railway coat of arms, Shield, Southern, Southern Railway, SR, Train, Unattributed, United Kingdom

3.8 Complex Image

Our research has uncovered many examples of images which are composites or sequences of simple images, where the latter may take any of the forms described above. In such cases the focus of interest would normally lie with the composite image as a single entity.

Recovering the Desired Semantic Content
For retrieval purposes, complex images would not appear to raise either challenges or opportunities different from those encountered with simple pictures.

Request: *Fish being sold*



f.57 Two scenes depicting fishmongers (w/c) New York Public Library, USA / The Bridgeman Art Library

Subject Metadata [25]

Title	Two scenes depicting fishmongers (w/c)
Keywords	merchant scales fish monk buying selling business Medieval Garramand fishmonger

4 Conclusion

The semantic gap is now a familiar feature of the landscape in visual image retrieval. The developing interest in bridging the semantic gap is a welcome response to the criticism directed at the visual image retrieval research community, one expression of which is 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" [4, p.197].

The project reported in this paper seeks to contribute towards this effort by locating 'semantic image retrieval' in the real world of client requests and metadata construction within commercially managed image collections. An analysis across the broad spectrum of image types has identified some of the challenges and opportunities posed by a CBIR-enabled approach to the realisation of semantic image content.

Acknowledgements

The 'Bridging the semantic gap in visual information retrieval' project is funded by the Arts and Humanities Research Council (MRG-AN6770/APN17429) whose support, together with that of our various contributors, is gratefully acknowledged.

References

1. Eakins, John P. and Graham, Margaret E.: Content-based Image Retrieval. A report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle, Newcastle upon Tyne (1999)

2. Gudivada, V.N. and Raghavan, V.V.: Content-based image retrieval systems. IEEE Computer 28(9) (1995) 18-22
3. Greisdorf, H. and O'Connor, B.: Modelling what users see when they look at images: a cognitive viewpoint. Journal of Documentation 58(1) (2002) 6-29
4. Jőrgensen, C.: Image retrieval: theory and research. The Scarecrow Press, Lanham, MA and Oxford (2003)
5. Trant, J.: Image Retrieval Benchmark database Service: a Needs Assessment and Preliminary development Plan: a report prepared for the Council on Library and Information Resources and the Coalition for Networked Information (2004) <http://www.clir.org/pubs/reports/trant04/tranttext.pdf>
6. Lavrenko, V., Manmatha, R. and Jeon, J.: A model for learning the semantics of pictures. (undated) <http://ciip.cs.umass.edu/pubfiles/mm-46.pdf>
7. Brodatz, P.: Textures: a photographic album for artists and designers. Dover, New York (1966)
8. Edina: Education Image Gallery <http://edina.ac.uk/eig/>
9. Jőrgensen, C.: Indexing images: testing an image description template. Paper given at the ASIS 1996 Annual Conference, October 19-24, 1996. <http://www.asis.org/annual-96/ElectronicProceedings/jorgensen.html>
10. Rui, Y, Huang, T.S., Chang, S-F.: Image Retrieval: Current Techniques, Promising Directions, and Open Issues, Journal of Visual Communication and Image Representation, 10(4) (1999) 39-62.
11. Marr, David: Vision. Freeman, New York. (1982)
12. Enser, P.G.B.: Pictorial Information Retrieval. (Progress in Documentation). Journal of Documentation 51(2), (1995) 126-170.
13. Armitage, L.H, and Enser, P.G.B.: Analysis of user need in image archives. Journal of Information Science 23(4) (1997) 287-299
14. Ornager, S.: The newspaper image database: empirical supported analysis of users' typology and word association clusters. In Fox, E.A.; Ingwersen, P.; Fidel R.; (eds): SIGIR 95, Proceedings of the 18th International AGM SIGIR ACM, New York, (1995) 212-218.
15. Enser, P. and Sandom, C.: Retrieval of Archival Moving Imagery - CBIR Outside the Frame? In: Lew, M.S., Sebe, N. and Eakins, J. P. (eds.): CIVR 2002 - International Conference on Image and Video Retrieval, London, UK. July 18-19, 2002, LNCS Series 2383. Berlin: Springer, (2002) 202-214.
16. Markkula, M.; Sormunen, E.: End-user Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. Information Retrieval 1(4), (2000) 259-285
17. Conniss, L.R; Ashford, L.R; Graham, M.E.: Information Seeking Behaviour in Image Retrieval. VISOR 1 Final Report. Library and Information Commission Research Report 95. Institute for Image Data Research, University of Northumbria at Newcastle' Newcastle upon Tyne, (2000)
18. Wellcome Trust: Medical Photographic Library http://medphoto.wellcome.ac.uk
19. Hu, B., Dasmahapatra, S., Lewis, P. and Shadbolt, N.: Ontology-based Medical Image Annotation with Description Logics. In Proceedings of The 15th IEEE International Conference on Tools with Artificial Intelligence, Sacramento, CA, USA. (2003)
20. Science & Society Picture Library. < http://www.scienceandsociety.co.uk>
21. Hauptmann, A., Ng, T.D., and Jin. R. Video retrieval using speech and image information. In Proceedings of Electronic Imaging Conference (EI'03), Storage Retrieval for Multimedia Databases, Santa Clara, CA, USA (2003) <http://www.informedia.cs.cmu.edu/documents/ei03_haupt.pdf>

22. Eakins, J.P.: Design criteria for a shape retrieval system. Computers in Industry 21 (1993) 167-184
23. Corporation of London: Talisweb. http://librarycatalogue.cityoflondon.gov.uk:8001/
24. Eakins, J.P., Boardman, J.M. and Graham, M.E.: Similarity retrieval of trademark images. IEEE Multimedia 5(2), (1998), 53-63
25. Bridgeman Art Library: Bridgeman Art Library – a fine art photographic archive. <http://www.bridgeman.co.uk/index.asp>