

Image Information Retrieval: An Overview of Current Research

Abby A. Goodrum

College of Information Science & Technology, Drexel University

goodruaa@drexel.edu

Abstract

This paper provides an overview of current research in image information retrieval and provides an outline of areas for future research. The approach is broad and interdisciplinary and focuses on three aspects of image research (IR): text-based retrieval, content-based retrieval, and user interactions with image information retrieval systems. The review concludes with a call for image retrieval evaluation studies similar to TREC.

Keywords: Information Science, Image Retrieval, CBIR,

Introduction

Interest in image retrieval has increased in large part due to the rapid growth of the World Wide Web. According to a recent study (Lawrence & Giles, 1999) there are 180 million images on the publicly indexable Web, a total amount of image data of about 3Tb [terabytes], and an astounding one million or more digital images are being produced every day (Jain, 93). The need to find a desired image from a collection is shared by many groups, including journalists, engineers, historians, designers, teachers, artists, and advertising agencies. Image needs and uses across users in these groups vary considerably.

Users may require access to images based on primitive features such as color, texture or shape or users may require access to images based on abstract concepts and symbolic imagery. The technology to access these images has also accelerated phenomenally and at present surpasses our understanding of how users interact with visual information. This paper provides an overview of current research in image information retrieval and provides an outline of areas for future research. The approach is broad and interdisciplinary and focuses on three aspects of image research (IR): text-based retrieval, content-based retrieval, and user interactions with image information retrieval systems. The review concludes with a call for image retrieval evaluation studies similar to TREC.

Material published as part of this journal, either on-line or in print, is copyrighted by the publisher of Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Editor@inform.nu to request redistribution permission.

Text-Based Image Retrieval Research

Most existing IR systems are text-based, but images frequently have little or no accompanying textual information. The solution historically has been to develop text-based ontologies and classification schemes for image description. Text-based indexing has many strengths including the ability to represent both general and specific instantiations of an object at varying levels of complexity. Reviews of the literature pertaining primarily to text-based approaches include Rasmussen (1997) Lancaster (1998) Lunin (1987) and Cawkell (1993).

Long before images could be digitized, access to image collections was provided by librarians, curators, and archivists through text descriptors or classification codes. These indexing schemes were often developed in-house and reflect the unique characteristics of a particular collection or clientele. This is still common practice and recently Zheng (1999) and Goodrum & Martin (1997) have reported on the hybridization of multiple schemas for classifying collections of historic costume collections. Hourihane (1989) has also reviewed a number of these unique systems for image classification. To date, very little research has been conducted on the relative effectiveness of these various approaches to image indexing in electronic environments.

Attempts to provide general systems for image indexing include the Getty's Art and Architecture Thesaurus (AAT), which consists of over 120,000 terms for the description of art, art history, architecture, and other cultural objects, and the Library of Congress Thesaurus of Graphic Materials (LCTGM). The AAT currently provides access to thirty-three hierarchical categories of image description using seven broad facets (Associated Concepts, Physical Attributes, Styles and Periods, Agents, Activities, Materials, and Objects). The ap-

proach in many collections, particularly general library environments, has been to apply an existing cataloging system like the Dewey Decimal System to image description using the LCTGM, or ICONCLASS.

Assignment of terms to describe images is not solved entirely by the use of controlled vocabularies or classification schemes however. The textual representation of images is problematic because images convey information relating to what is actually depicted in the image as well as what the image is about. Shatford (1986) posits this discussion within a framework based on Panofsky's approach to analyzing iconographical levels of meaning in images (1955). For example, an image may be *of* a glass of wine, but be *about* the Christian mass. Shatford-Layne (1994) extended this discussion by proposing a theoretical model for analyzing the subject of an image and suggests that it may be necessary to determine which attributes will result in useful groupings of images and which attributes should be left to the user to identify. Turner (1994) extended this model by examining the term assignments given to both still and moving images by groups with the aim of discovering appropriate ways to index images. Subjects in this experiment used a few terms often, with the vast majority of terms used only once.

An area of image classification research for which very little has been published is retrieval by associated metadata. Metadata (data about data) includes attributes such as image creator, image format, date of creation, and simple object descriptions taken from titles or captions. The Dublin Core metadata set, while not a standard, has been adopted for the description of web documents. Similarly, the World-Wide Web Consortium (W3C) is developing a resource description framework (RDF) that will provide not only textual descriptions of the resource but also data such as color histogram or other numeric representations of image content. A recent study by Lawrence & Giles (1999) indicates however that use of metadata tags is still not widespread.

Unfortunately, manual assignment of textual attributes is both time consuming and costly. Manual indexing suffers from low term agreement across indexers (Markey 1984), and between indexers and user queries (Enser & McGregor, 1993; Seloff, 1990). Automatic assignment of textual attributes has been conducted using captions from still images, and transcripts, close captioning, or verbal description for the blind, that accompany many videos (Turner, 1994). While these approaches greatly reduce the labor involved in manual assignment of keywords, it must be remembered that many images are without accompanying text. Furthermore, users' image needs may occur at a primitive level that taps directly into the visual attributes of an image. These attributes may best be represented by image exemplars and retrieved by systems performing pattern matches based on color, texture, shape, and other visual features.

Content-Based Image Retrieval Research

Problems with text-based access to images have prompted increasing interest in the development of image-based solutions. This is most often referred to as content-based image retrieval (CBIR). Content-based image retrieval relies on the characterization of primitive features such as color, shape, and texture that can be automatically extracted from the images themselves. Commercial CBIR systems in use include IBM's Query By Image Content (QBIC) described first by Flickner et al, (1995), Virage's VIR Image Engine (Gupta et al, 1996), and Excalibur's Image Retrieval Ware. On the Web, CBIR image retrieval systems include WebSEEK (Smith & Chang, 1997b), Informedia, and Photobook among others.

Queries to CBIR systems are most often expressed as visual exemplars of the type of image or image attribute being sought. For example, users may submit a sketch, click on a texture palette, or select a particular shape of interest. The system then identifies those stored images with a high degree of similarity to the requested feature. Idris and Panchanathan (1997) discuss in detail the various technologies for image indexing and retrieval based on shape, color, texture, and spatial location. They also examine issues related to the retrieval of moving images, including shot detection and video segmentation. Aigrain et al (1996) provide an overview of approaches to image similarity matching for database retrieval and discuss the difficulty of expressing high-level image needs to low level image features. The following is a brief description of prevailing methods of content-based image retrieval.

Color

Retrieving images based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors) Current research in this area attempts to segment color proportion by region and by spatial relationship among several color regions. (Stricker & Orengo, 1995; Carson et al, 1997)

Texture

Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated. (Tamura et al. 1978) The problem here is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as "silky" or "rough". Ma & Manjanath (1998) have extended work in this area through the development of a texture thesaurus that matches

texture regions in images to words representing texture attributes.

Shape

Queries for shapes are generally achieved by selecting an example image provided by the system or by having the user sketch a shape. The primary mechanisms used for shape retrieval include identification of features such as lines, boundaries, aspect ratio, and circularity, and by identifying areas of change or stability via region growing and edge detection. Of particular concern has been the problem of dealing with images having overlapping or touching shapes.

Several problems remain including retrieval of features based on location within an image, the extension of 2-dimensional features to 3-dimension, and appropriate segmentation of video images. Although research in higher order CBIR is underway, current systems are not capable of retrieving all instances of horses based on the shape, color, or texture of a single instance of a horse in a query. For example, a shape-based query depicting a side view of a horse does not retrieve images of horses from behind or above. Research in object recognition conducted by Forsythe et al (1997) has sought to develop techniques for modeling a class of objects and identifying defining attributes and features for that class. Rorvig has examined the use of human judgments to train the system to recognize patterns of user-defined similarity for automatic identification of image classes. Chang et al, (1998) also utilized users' relevance judgments to refine searches and to assign semantic keywords to images that can be used by subsequent users to query the system.

Although shape, color and texture are undoubtedly important visual features for image representation, there is still little understanding of how best to implement these attributes for image retrieval. An understanding of what constitutes similarity for image retrieval purposes is also needed. The technology for content-based image retrieval is still in its infancy. The focus to date has been primarily on the use of features that can be computationally acquired, but little has been done to identify the visual attributes needed by users for various tasks and collections.

User Interaction

Users seeking images come from a variety of domains, including law enforcement, journalism, education, entertainment, medicine, architecture, engineering, publishing, advertising, and art. Most of the published research in this area has focused on specific collections, or specific groups of users. For example, Ornager, (1997), examined the use of newspaper image archives, Keister, (1994) analyzed queries submitted to the image archive at the National Library of Medicine, and Markey, (1988) and Hastings (1995) explored the use of im-

ages by art historians

Most of the research in visual information seeking behavior and use has been conducted in non-digitized collections with written or verbal queries. The seminal work in this area was conducted by Enser (1993) who analyzed nearly 3000 written requests from 1000 request forms at the Hulton Deutsch archive. Results indicated that queries for visual materials exhibited a greater level of specificity than requests for textual materials, and that the majority of requests were for specific instances of a general category ("London Bridge" rather than the generic "Bridges"). Armitage and Enser (1997) extended this research by categorizing requests across seven picture archives. Their work resulted in a framework for queries with 4 main categories (who, what, when, where) and 3 levels of abstraction (specific, generic, abstract). Similarly, Keister's (1994) analysis of query logs at the National Library of Medicine demonstrated that most queries were structured using both abstract concepts as well as concrete image elements. She concluded that the aesthetic and emotional needs of the user are highly subjective and do not lend themselves to indexing.

Research examining users' interactions with electronic image retrieval systems is still quite sparse. Goodrum & Spink, (1999) analyzed 33,149 image queries made to EXCITE, a major search engine on the Web. They found that users input very few terms per query and that most query terms occurred only once. The most frequently occurring terms appeared in less than 10% of all queries. They also noted the presence of terms that modified a general request such as "girls" into a specific visual request such as "pretty girls."

Several studies have demonstrated that when unconstrained from a retrieval task, users tend to create narratives to describe images. O'Connor (1999) found that when image descriptions were elicited, subjects created short narratives or stories for images that went well beyond describing the objects depicted. Jorgensen (1995) also demonstrated that story or narrative attributes are commonly assigned to images by users outside of a retrieval task.

Research investigating the effect of task and image use on user's interactions with image information has identified two ends of a continuum: focused specific searching and looser searching or browsing (Goodrum, 1997). Fidal (1997) described a continuum of use between a "Data Pole" and an "Objects Pole." At the data pole, images are used as sources of information; at the objects pole images are defined in terms of some task (to be used in the creation of an advertisement, book jacket or brochure). At the data pole, users want the smallest set that can provide the information needed; at the objects pole, users want to be able to browse larger sets of retrieved items. This continuum of search activity may also be related to the type of attributes users seek. For example,

Image Information Retrieval

browsing tasks may call for image attributes and visual examination of images of interest, while search tasks may require the specificity of text (Goodrum, 1997). Rorvig's (1988) examination of the use of the NASA Visual Thesaurus indicated that given a choice most users reverted to text rather than search using images as input. In Mostafafa's studies (1994, Mostafa & Dillon, 1996) visual queries were more likely to be used in conjunction with conceptual information needs, but users input more verbal queries overall.

If users interactions with visual IR systems is determined in some part by their tasks, by the type of images in the collection, or by the users' domain, then interfaces will have to be developed to capitalize on these facets of use. Research addressing the design of interfaces for visual information retrieval systems is remarkably thin. While most systems utilizing CBIR provide some mechanism for users to query by visual exemplar, research is needed to identify at what point in their interaction users want or need to express a query using such tools (Lee et al, 1994). Research is also needed that examines interface support for browsing, query reformulation and iterative searching. This is problematic for many reasons, not the least of which is a lack of research examining the range of attributes associated with images that might prove most useful within different retrieval contexts. Ding et al, (1997) and Tse et al, (1998) have explored the relationship between users' tasks and the presentation of video key frames in interfaces to support video browsing. Further research is needed to identify appropriate presentation of retrieved images in a variety of contexts and domains.

Conclusions

Although the CS research community has been primarily concerned with content-based approaches focusing on retrieval of visual attributes such as shape, texture and color, there is increasing interest in the development of ontologies to support and aid content-based IR (Jorgensen et al, 1999). This provides the first clear indication that cross-disciplinary approaches utilizing both text and image features for retrieval are being sought.

Fundamental questions remain in areas such as indexing and classification, vocabulary control, user needs, relevance, similarity measures, granularity of indexing, economies of scale and presentation of retrieval results. There is strong indication that the combined use of both text and image features may result in improved VIR system effectiveness, but when should textual attributes be present: at the time of indexing, at the time of searching, during query reformulation. or all of these?

Finally, in contrast to the evaluation of systems for text retrieval, which have been conducted for more than 30 years (Cleverdon et al, 1966) image IR has suffered from the lack of research analyzing the effectiveness of various systems. This

state of affairs is compounded by the lack of a large image test bed and disagreement on what constitutes effective image retrieval and how to measure it. An image version of the TREC text retrieval experiments has been called for in the multimedia research community, and several test beds have been proposed (Schmidt & Over, 1999; Slaughter & Marchionini, 1999). For image retrieval effectiveness to be studied we need to establish large test collections of images and benchmark queries, and the adoption of a set of evaluation measures such as the pooling methods used in the TREC experiments.

References

- Aigrain, P et al (1996) Content-based representation and retrieval of visual media: A state-of-the-art review . *Multimedia Tools and Applications* **3**(3), 179-202.
- Armitage, L and Enser, P G B (1997) Analysis of user need in image archives. *Journal of Information Science*, **23**(4), 287-299.
- Carson, C. et al, (1997) Region-based image querying, in *Proceedings of IEEE Workshop on Content-Based Access to Image and Video Libraries*, San Juan, Puerto Rico, 42-49.
- Cawkell, A. (1993). Indexing collections of electronic images: A review. *British Library Research Review*, *15*.
- Chang, E. et al, (1998) RIME: A replicated image detector for the WWW, in *Multimedia Storage and Archiving Systems III*, (Kuo, C. et al, eds.), Proc SPIE 3527, 58-67.
- Cleverdon, C. et al (1966). Factors determining the performance of indexing systems. Cranfield College of Aeronautics.
- Enser, P. (1995) Pictorial information retrieval. *Journal of Documentation*, **51**(2), 126-170.
- Enser, P. & McGregor, C. (1993). Analysis of visual information retrieval queries. *British Library Research and Development Report*, 6104.
- Fidal, R. (1997). The image retrieval task: Implications for the design and evaluation of image databases. *The New Review of Hypermedia and Multimedia*, *3*, 181-199.
- Flickner, M. et al (1995) Query by image and video content: the QBIC system. *IEEE Computer* **28**(9), 23-32.
- Forsythe, D. et al (1997) Finding pictures of objects in large collections of images, in *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing* (Heidorn, P. and Sandore, B. eds.), 118-139.
- Goodrum, A. (1997). Evaluation of Text-Based and Image-Based Representations for Moving Image Documents. *Unpublished dissertation*. University of North Texas, Denton, TX.
- Goodrum, A., & Martin, K. (1999). *Bringing fashion our of the closet: Classification structure for the Drexel Historic Costume Collec-*

- tion. *Bulletin of the American Society for Information Science*, Volume 25, Number 6, pp 21-23, August/September, 1999.
- Goodrum, A., & Spink, A. (1999). *Visual Information seeking: A study of image queries on the World Wide Web. Proceedings the 1999 Annual Meeting of the American Society for Information Science*, October 31-Nov 4, 1999, Washington, DC.
- Gupta, A. et al (1996) The Virage image search engine: an open framework for image management in *Storage and Retrieval for Image and Video Databases IV*, Proc SPIE 2670, pp 76-87.
- Hastings, S. (1994). Query categories in a study of intellectual access to digitized art images. *American Society for Information Science, Annual Meeting*, Chicago, IL October, 1995.
- Hourihane, C. (1989). A selective survey of systems of subject classification. *Computers and the History of Art*. 117-129.
- Idris, F and Panchanathan, S (1997a) Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation* 8(2) 146-166.
- Keister, L. (1994). User types and queries: Impact on image access systems. In *Challenges in Indexing Electronic Text and Images*. (Fidel, R. et al, eds.), ASIS 1994, 7-22.
- Lancaster, F. (1998) *Indexing and abstracting in theory and practice*. 2nd ed. Library Association, London.
- Lawrence, S. & Giles, L (1999) Accessibility of information on the web. *Nature*, vol.400, pp. 107—109.
- Lee, D.,et al. (1994). Query by image content using multiple objects and multiple features: User interface issues. *IEEE* pg76-80.
- Lunin, L. (1987). Electronic image information. *Annual Review of Information Science & Technology*, 179-224.
- Jain, R (1993) Workshop report: NSF workshop on visual information management systems. in *Storage and Retrieval for Image and Video Databases* (Niblack, W R and Jain, R C, eds), Proc SPIE 1908, 198-218.
- Jorgensen, C. (1996). The applicability of existing classification systems to image attributes: A selected review. *Knowledge Organization and Change*, 5, 189-197.
- Jorgensen et al, (1999). Considerations in prototyping an image retrieval testbed. *Proceedings of the Multimedia Indexing and Retrieval Workshop presented in Conjunction with the 22nd Annual International ACM SIGIR Conference*, Berkeley, CA, August 19, 1999.
- Ma, W. & Manjanath, B. (1998). Netra: A toolbox for navigating large image databases. *Proceedings of IEEE International Conference on Image Processing (ICIP97)*, 1, 568-571.
- Markey, K. (1988). Access to iconographical research collections. *Library Trends*, 37(2), 154-174.
- Mostafa, Javed. Digital Image Representation and Access. *Annual Review of Information Science and Technology*, Vol 29, 1994.
- O'Connor, B. , O'Connor, Mary K. and Abbas, June M. (1999). User Reactions as Access Mechanism: An Exploration Based on Captions for Images, *Journal of The American Society For Information Science*, Vol 50, No 8, 1999.
- Ornager, S. (1997), Image retrieval: Theoretical and empirical user studies on accessing information in images. *ASIS 97: proceedings of the 60th Annual Meeting of the American Society for Information Science*, vol 34, 202-211.
- Panofsky, E. (1962). *Studies in Iconology*. New York, Harper & Row.
- Rasmussen, E. (1997). Indexing images. *Annual Review of Information Science and Technology*, 32, 169-196.
- Rorvig, M. et al, (1988) The NASA image collection visual thesaurus. *Proceedings of the American Society for Information Science 17th Mid-Year Meeting*, Ann Arbor, MI.
- Seloff, G A (1990). "Automated Access to the NASA-JSC Image Archive." *Library Trends*, 38(4), 682-696.
- Shatford, S (1986) Analyzing the subject of a picture: a theoretical approach. *Cataloging and Classification Quarterly*, 6(3), 39-62.
- Shatford-Layne, S (1994) Some issues in the indexing of images. *Journal of the American Society of Information Science*, 45(8), 583-588.
- Smith, J. & Chang, S. (1997) An image and video search engine for the World Wide Web, in *Storage and Retrieval for Image and Video Databases V* (Sethi, I. And Jain, R. eds.), Proc SPIE 3022, 84-95.
- Stricker, M and Orengo, M (1995). Similarity of color images, in *Storage and Retrieval for Image and Video Databases III* (Niblack, W R and Jain, R C, eds), Proc SPIE 2420, pp 381-392.
- Svenonius, E (1994), Access to nonbook materials: the limits of subject indexing for visual and aural languages. *Journal of the American Society of Information Science*, 45(8), 600-606.
- Tamura, H et al (1978) Textual features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics* 8(6), 460-472.
- Turner, J (1990). Representing and accessing information in the stock-shot database at the National Film Board of Canada. *The Canadian Journal of Information Science* v. 15 p. 1-22.
- Zheng, M. (1999). Metadata elements for object description and representation: A case report from a digitized historical fashion collection project. *Journal of the American Society for Information Science*, 50(13), 1193-1208.