Image Retrieval From the World Wide Web: Issues, Techniques, and Systems

M. L. KHERFI AND D. ZIOU

Université de Sherbrooke

AND

A. BERNARDI

Laboratoires Universitaires Bell

With the explosive growth of the World Wide Web, the public is gaining access to massive amounts of information. However, locating needed and relevant information remains a difficult task, whether the information is textual or visual. Text search engines have existed for some years now and have achieved a certain degree of success. However, despite the large number of images available on the Web, image search engines are still rare. In this article, we show that in order to allow people to profit from all this visual information, there is a need to develop tools that help them to locate the needed images with good precision in a reasonable time, and that such tools are useful for many applications and purposes. The article surveys the main characteristics of the existing systems most often cited in the literature, such as ImageRover, WebSeek, Diogenes, and Atlas WISE. It then examines the various issues related to the design and implementation of a Web image search engine, such as data gathering and digestion, indexing, query specification, retrieval and similarity, Web coverage, and performance evaluation. A general discussion is given for each of these issues, with examples of the ways they are addressed by existing engines, and 130 related references are given. Some concluding remarks and directions for future research are also presented.

Categories and Subject Descriptors: I.3.2 [**Computer Graphics**]: Graphics Systems—*Distributed/network graphics*; I.4.7 [**Image Processing and Computer Vision**]: Feature measurement; I.4.9 [**Image Processing and Computer Vision**]: Applications; I.5.3 [**Pattern Recognition**]: Clustering—*Similarity measures*

General Terms: Algorithms, Design, Human Factors, Performance

Additional Key Words and Phrases: Image-retrieval, World Wide Web, indexing, search, similarity, feature extraction and selection, relevance feedback, crawling

1. INTRODUCTION

What we now know as the World Wide Web began in 1991 at CERN (the European Organization for Nuclear Research) as an organization-wide collaborative environment for the sharing of research documents in nuclear physics. Since then,

The completion of this research was made possible thanks to NSERC and Bell Canada's support through its Bell University Laboratories R & D program.

Authors' addresses: M. L. Kherfi and D. Ziou, Department of Computer Science, University of Sherbrooke, Sherbrooke, Qc, J1K 2R1, Canada; email: {mohammed.lamine.kherfi,djemel.ziou}@usherbrooke.ca; A. Bernardi, Bell University Laboratories, 2020 University Street, Montreal, Qc, H3A 2A5, Canada; email: alan.bernardi@bell.ca.

it has grown to encompass such diverse information resources as personal home pages, online digital libraries, virtual museums, product and service catalogs, government information for public dissemination, and research publications [Gudivada et al. 1997]. The quantity of information on the World Wide Web (also called the WWW or simply the Web) increases daily in a spectacular manner; however, this information varies considerably in relevance, ranging from precise official information to inaccurate statements coming from unknown sources. The great paradox of the Web is that the more information there is available about a given subject, the more difficult it is to locate accurate, relevant information. To help mitigate this problem, many information retrieval engines (or systems) have appeared in recent years. The need for such tools is clearly shown by the success they have achieved. However, despite the visual nature of the Web, few engines have focused on retrieving visual information such as images and videos. Indeed, while there has been some success in developing search engines for text, search engines for other media on the Web (images, audio, and video) are still rare and not as powerful. In this article, we are concerned with the retrieval of visual information from the Web. Visual information is published both embedded in Web documents and as stand-alone objects. It takes the form of images, graphics, bitmaps, animations, and videos [Smith and Chang 1997]. People frequently need to locate such material to use it as a source of information or for illustration. There is thus a vital need for Web image search engines. Such engines are useful for many applications, including law enforcement, image copyright protection, filtering of inappropriate mature content, criminal tracking, home entertainment, education, and training. Existing commercial engines include Google image search (http://images.google.com/), Lycos (http://multimedia.lycos.com/), and AltaVista photo finder (http://image.altavista.com/). These engines use text to look for images, without considering image content. Throughout this article, we will show that both text and image content can contain useful information that should be exploited in indexing images. Some research prototypes that have appeared in the last years, such as ImageRover [Sclaroff et al. 1997; Sclaroff 1995; Taycher et al. 1997; La Cascia et al. 1998], WebSeek [Smith and Chang 1996a, 1997], and Atlas WISE [Kherfi et al. 2003a, 2003b], have investigated the use of image content for retrieval.

The aim of this article is to discuss the usefulness of tools for retrieving images from the Web and to provide an overview of the main issues that must be addressed by someone who wants to develop such tools. We will look at the techniques adopted by existing systems, as well as some open issues and promising avenues for research. We are aware that a survey on general information retrieval from the Web, or on image retrieval in a general context, goes beyond the scope of this article. The reader can find more details about textual information retrieval from the Web in Gudivada et al. [1997] and Kobayashi and Takeda [2000]. For a survey on audio retrieval in general, see Foote [1999]. More details on general-purpose image retrieval techniques and systems can be found in Eakins and Graham [1999], Veltkamp and Tanase [2000], Smeulders et al. [2000], and Rui et al. [1999]. The principal motivation for the current work is that while there have been many surveys on systems and techniques used for image retrieval, surveys on the important issues of specialized image retrieval are rare. Although some papers, such as Aslandogan and Yu [1999, 2000a], have referenced some Web image retrieval engines, to our knowledge, there is no survey that specifically focuses on the issues, techniques, and systems for image retrieval from the Web.

Web users need tools for automatic image retrieval; however, when designing and implementing such tools, people are faced with a number of problems which are related to two factors. The first of these has to do with aspects of the image retrieval field, such as the quantity of information contained in an image and its surroundings. The second concerns the nature of the Web, that is, its huge size, lack of structure, and staggering scale, which limits the retrieval and indexing algorithms that can be used. In this article, we will address each of these issues and try to provide some answers and future directions. The article is organized as follows: in Section 2, we discuss the need for tools for image retrieval from the Web, their possible applications, and the services people may expect such tools to provide. Section 3 gives an overview of some examples of existing Web image search engines. In Section 4, we examine each of the issues that must be considered in designing a Web image retrieval engine. The issues discussed are data gathering, image and data digestion, similarity and matching, indexing, query specification, retrieval and refinement, Web coverage, and performance evaluation. Some concluding remarks and directions for future research are given in Section 5, and the overall conclusions of the article are summarized in Section 6.

2. THE NEED FOR IMAGE RETRIEVAL FROM THE WEB

xThe World Wide Web contains a great quantity of image and other visual information such as videos, movies, and comic strips, that may belong to structured collections (e.g., museum collections) or be independent (e.g., images found in Web pages in the form of individuals' photographs, logos, and so on). Tools for effective retrieval of this information can prove very useful for many applications. In the current section, we try to show why such tools are indispensable for users, what applications people may need them for, and what services users may ask them to accomplish.

2.1. Why Develop Web Image Retrieval Engines?

Since its appearance, the World Wide Web has grown very considerably in size, and the number of Web pages is now evaluated at several billion pages. As an example, the number of pages referenced by Google has grown from approximately 2,000,000,000 to more than 3,300,000,000 in about 1 year. The availability of such a quantity of diverse information means that many subjects may have thousands of documents that are related to them to a greater or lesser extent. However, this is not necessarily an advantage, because it is often difficult to locate the most relevant information among all those related to the same subject. Therefore, the main question today is no longer "Is there any information related to this subject on the Web?" but rather "How can I locate the information most closely related to this subject in the shortest time?" The answer lies in developing tools that index the Web's content, allowing the user to submit queries and then retrieve documents that are related to those queries. Such tools are information retrieval engines.

In the area of textual information retrieval from the Web, many engines have appeared in recent years and have achieved a great degree of success. This can be clearly shown by the fact that the home pages of engines such as Google and Yahoo! are among the most frequently visited on the whole Web. However, the World Wide Web does not contain only textual information. It also includes millions of images, as well as audio and other media, that can be a good source of information and may prove useful for various applications. We know that building open, generic image and media retrieval engines like those available for text may be infeasible using current technology, but we think that even simpler attempts should be welcome.

If we focus on visual information, the World Wide Web contains several kinds of images and other visual information, such as videos, movies, and comic strips, in various formats such as JPG and GIF for still images, and MPG, AVI, and RAM for moving images. Both types of visual information (still and moving) are used to depict a range of types of scenes and objects too wide to enumerate in this article. The following examples are intended as an illustration rather than an exhaustive enumeration: photographs of people, museum collections and paintings, medical libraries, erotic and sex images and videos, maps and charts, star photos and movie shots, advertising tapes, greeting cards, logos, sports images, humorous images, and comic strips. Leaving this enormous store of visual information unorganized, without retrieval tools, would appreciably reduce its utilizability by users and prevent them from benefiting from it. Hence, it is vital to possess tools for image retrieval from the Web and to index the visual content of the Web into a searchable catalog that helps users navigate across various subjects. This could help them quickly find the images they seek, which can then be used for many purposes. For example, an image may be used as a source of information (e.g., a tourist who uses a map found on the Web to get an idea about the country he or she will visit) or to accomplish some task (e.g., a student who uses a photograph of downtown Toronto to illustrate a research report). Developing image retrieval engines for the Web may be useful for many professional applications, such as crime prevention and intellectual property protection, as well as for other purposes such as home entertainment, education, and training. Some of these applications will be discussed next.

2.2. Applications of Web Image Retrieval

Gudivada et al. [1995] have identified many important applications of general-purpose content-based image retrieval: art gallery and museum management, architectural and engineering design, interior design, remote sensing and management of earth resources, geographic information systems, scientific database management, weather forecasting, retailing, fabric and fashion design, trademark and copyright database management, law enforcement and crime prevention, picture archiving and communication systems, the military, biomedical imagery, and home entertainment. Other applications include journalism and advertising, cultural heritage preservation, face recognition, and online shopping. Some of these general-purpose applications, such as home entertainment, remain valid for image retrieval from the Web. Other applications are more specific to the Web image retrieval context, including intellectual property, mature content filtering, and tourism and travel applications. Hereafter, we will detail some important applications of image retrieval from the Web.

2.2.1. Intellectual Property. One application of Web image retrieval to intellectual property can be seen in image copyright protection. Image copyright protection on the Web is vital, especially because unauthorized copies of images can easily be transmitted over the Internet. One existing technique for image copyright protection is watermarking, which involves adding the creator's or distributor's identity to each image. However, this technique is vulnerable to image processing, geometric distortions, and subterfuge attacks [Chang et al. 1998]. In addition, working with a large variety of watermark schemes increases detection time exponentially. Content-based image retrieval from the Web can offer a good alternative for solving this kind of problem. This requires essentially a query-based retrieval functionality, where the requested image is compared with suspect ones available on the Web. An example of a content-based image retrieval system designed to detect replicated images on the Web is given in Chang et al. [1998].

2.2.2. Filtering of Inappropriate Mature Content. There are now a large number of Web sites containing pornographic images and videos that are available for free download. Accessing of such images and videos by children is a problem that many parents are becoming concerned about. In addition to this, in some cultures, such images are not tolerated even for adults. This renders it necessary to develop inappropriate mature content filtering tools for the World Wide Web. Old solutions include pornography-free Web sites, which try to prevent children from accessing undesirable sites, and software filters such as NetNanny, Cyber Patrol, and CyberSitter [Wang et al. 1998]. Such tools have proven ineffective because they cover only a very small part of the Web. Furthermore, despite the fact that pornography sites often contain many images and little text, such tools check only the text or the IP addresses. In recent work, researchers have explored the use of computer vision techniques to automatically identify pornographic images. Fleck et al. [1996] and Forsyth et al. [1996] used an algorithm that tries to identify naked people in images. This algorithm uses a skin filter based on texture and color. Chan et al. [1999] developed a similar system that identifies images of naked humans by skin tone and limb shape. Integrating such tools in Web image retrieval can help considerably in identifying suspicious sites and thus in filtering them. This can be done by submitting suitable sample queries to a query-based retrieval system, which tries to identify all Web pages that correspond to them. Wang et al. [1998], for example, designed a system called *IBCOW* that classifies Web sites into "objectionable" and "benign" based on image content.

2.2.3. Law Enforcement and Crime Prevention. Content-based image retrieval has been applied for many purposes in law enforcement and crime prevention, such as fingerprint recognition [Ratha et al. 1996], face recognition [Martinez 1999], DNA matching, shoe sole impressions [Gerardts and Bishold 2002], and surveillance systems [Eakins and Graham 1999]. Web image retrieval can be applied in additional areas of law enforcement. Many criminals use the Internet as a means of promoting their illicit goods and services, such as illegal weapons, drugs, and pedophilia. Because many such sites contain little text and more visual information, relying exclusively on text-retrieval techniques is often ineffective in locating them. Possessing tools for both content-based and text-based image retrieval from the Web is thus clearly vital in order to be able to locate such sites. Other kinds of illegal operations include the use of Web sites to call for violence, racism, and Nazism. Tools for query-based image retrieval from the Web support police in fighting such operations by helping them locate these sites.

2.2.4. Travel and Tourism. Before visiting a new place, people always want to know more about it. The information sought includes maps of the country they will visit, as well as city maps, which can give them an idea about transportation networks, essential conveniences, monuments, and tourist attractions. A tourist may also be interested in having a basic idea about the country's way of life, such as its markets, characteristic architecture, and traditional dress. What better way than images to provide such information? The World Wide Web is full of such images, but tools for automatic retrieval are needed to help tourists and travelers benefit from them. Furthermore, cataloguing images available on the Web in a browsing structure can help travelers easily reach the sought information.

2.2.5. Education and Training. Education and training are among the most important applications of image retrieval in general [Wactlar et al. 1996; Van der Zwan et al. 1999] and of image retrieval from the Web in particular. Students are regularly asked to do research on particular subjects. They need images for at least two purposes: as a source of information (e.g., a map of Canada in 1890) or to illustrate their ideas (e.g., an image of a family in Saskatchewan in the 1920s). These images, along with videos, can also be used by teachers in preparing their courses, providing them with teaching materials to illustrate and explain their ideas. Many such images are available on the World Wide Web for free access or download; however, locating them is sometimes very time-consuming. Having Web image search engines and navigation catalogs can render this operation much easier.

2.2.6. Home Entertainment. Many people access the Web for entertainment purposes. Visual information used for entertainment includes joke images, caricatures, comic strips, movie shots, and music clips. However, people often have trouble finding the images or visual media they are looking for. Web image and video retrieval tools for such kinds of data appear indispensable in helping users locate the sought images. Furthermore, if the visual content of the Web is catalogued in an index structure that people can use for browsing, access to such material will be easier and navigation more enjoyable.

2.2.7. Fashion, Architecture, and Engineering Design. Graphic, fashion and industrial designers reuse images of previous designs as a source of inspiration [Yang et al. 1994; Bird et al. 1996]. For a similar reason, architects and engineers may need to see plans, images of machines, and other related material when developing new projects. With the advent of the Web, these professionals are no longer limited to their local collections in their quest for inspiration. The Web allows them to access other images that may exhibit similar or different styles, providing richer stimulation for their imagination. Web image retrieval systems and browsing catalogs are thus useful in all of these fields.

2.2.8. Historical and Art Research. Historians, archaeologists, and sociologists use visual data as a source of information or to support their research. When access to the original work of art is restricted (e.g., due to geographic distance, ownership restrictions, or the physical condition of the work), researchers can use surrogates, which may be found on the Web in the form of photographs or pictures of the object, to circumvent this problem. A Web image retrieval engine can save them time in looking for such material. Examples of recent work that attempt to apply image retrieval to art and historical research include Barnard et al. [2001], Jain et al. [1997], and Chen and Wang [2002].

2.3. Services People May Need from a Web Image Retrieval Engine

Now that we have looked at the necessity of developing image retrieval engines for the Web and some important applications for which they could be useful, let us see what services people would expect such tools to provide. The number of users accessing the Internet is increasing daily and their frequency of access varies considerably, ranging from the occasional user who connects purely for fun to the professional or passionate user who spends a large part of his or her time browsing Web sites. The way these users look at the information available on the Web and how they try to locate it varies from user to user, and even for the same user at different times. In fact, the services people want from a Web image search engine are influenced by their needs and the use they will make of the retrieved images. According to Fidal [1997], the use of images lies between the "data pole" and the "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 (e.g., 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 allows us to identify two main services that can be provided by a retrieval system, namely, query-based retrieval and browsing. In addition to these two services, we sometimes want the retrieval system to provide a summary of a set of images or videos. Each of these three services will be described in more detail below. We note that a search engine may offer one, two, or all three of these services.

2.3.1. Query-Based Retrieval. The basic function that users expect from a Web image retrieval system is to allow them to specify queries and then retrieve images that correspond to these queries. To be able to perform retrieval, the system needs to accomplish data gathering, data digestion, and index creation beforehand, as we will see in Section 4. Concerning the queries, they can be text-based, image-based, or a combination of the two. If the query is text-based, it may contain names of objects present in the images sought, a

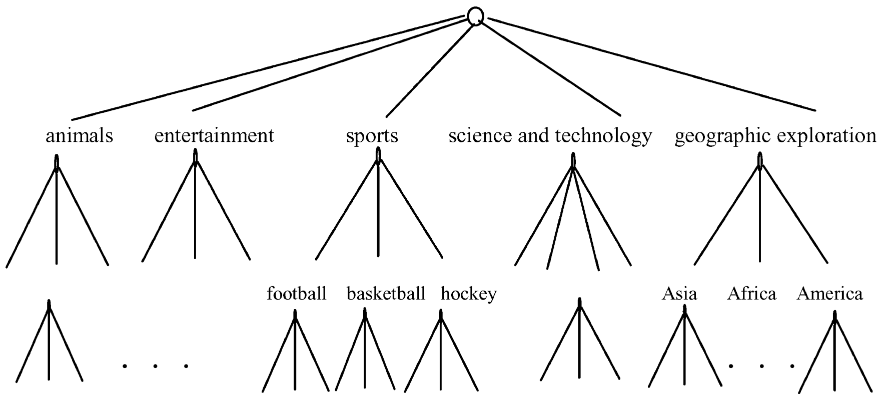

Fig. 1. Example of an image taxonomy by subject.

description of their content, or a citation that can be associated with these images. If the query is image-based, the user may be given a set of sample images or be asked to provide her or his own images. After the query is specified, the system goes through its index to identify the class of images that most likely corresponds to the query. This class of images can be further ranked according to certain similarity measures, in order to return to the user those images which most closely resemble the query. Two problems may be encountered with the retrieved images, namely, noise and miss. These two problems arise because of imperfections at different levels of the process, including query specification, identification of the user's needs, indexing, and similarity measures. To be able to reduce noise and miss, the retrieval system should allow for query refinement through relevance feedback. Query-based retrieval is more appropriate when the user has a precise query in mind that can be easily translated into keywords and/or sample images. All of the Web image retrieval engines that we will cite in Section 3 offer query-based retrieval.

2.3.2. Browsing. Text search engines such as Lycos, AltaVista, and Yahoo! index documents by their textual content. These systems periodically scour the Web, record the text on each page, and, through processes of automated analysis and/or (semi-)automated classification, condense the Web into compact, searchable catalogs. The updating of catalogs must be done periodically because, every day, documents (text, images, etc.) are added to the Web, while existing documents are modified or even deleted [Smith and Chang 1997]. It is very useful for a Web image search system to offer an image taxonomy based on subjects such as art and culture, sports, world regions, technology, etc. Each subject can be further divided into subdirectories; for instance, the directory World Regions can be divided into America, Africa, Asia, Europe, and Australia. The decomposition can go further, to many hierarchical levels. It can propose links between different subjects, constituting a dictionary that enables the user to browse through a number of subjects.

Figure 1 gives an example of an image taxonomy. We can also imagine a catalog similar to the one of Yahoo!. In fact, browsing is more suitable when the user has a broad query in mind. By going deeper into the index levels, she or he can narrow this query and identify what she or he is looking for. Among the systems we will cite in Section 3, WebSeek [Smith and Chang 1996a, 1997] offers an online catalog that summarizes the visual content of the Web by mapping it into a taxonomy with different subjects and subsubjects. It uses image names and directories to build the taxonomy.

2.3.3. Summarization. Another service that a Web image search engine may be asked to offer is the provision of summaries. It may be asked to give a general title that summarizes the content of a set of images collected from the Web, such as "sunset images" or "car images." The system may also give a single image that summarizes the content of a video shot or an image set. Summaries can also be used in combination with the catalog if each category of images in the catalog is represented by some representative words or images that are provided by the summarization service. This could help give users an a priori idea of each category of images, allowing them to decide whether or not it is interesting enough to visit.

In conclusion to this section, we see that the Web contains a great quantity of visual information that may be useful for many applications. However, locating the needed pieces of information remains a challenging task, especially when there is a lot of information related to the same subject. This makes it particularly vital to develop Web image retrieval engines. Such systems could enable users to locate the sought images in various ways, including query-based search and browsing through catalogs. Before presenting the different issues that must be resolved by someone developing a Web image retrieval system, let us first examine some of the more frequently cited existing systems.

3. OVERVIEW OF SOME EXISTING SYSTEMS

A number of prototypes for image retrieval from the Web have been proposed in recent years, including ImageRover, Diogenes, Atlas WISE, WebSeer, WebSeek, ImageScape, PicToSeek, WebMars, and Web-WISE. In this section, we will give some details on some of these systems. It should be noted that all of the reviewed systems are prototypes, not actual engines that allow users to retrieve images from the whole World Wide Web. Some of them do offer online demos, however.

3.1. ImageRover

ImageRover [Sclaroff 1995; Sclaroff et al. 1997; Taycher et al. 1997; La Cascia et al. 1998] was developed at the Department of Computer Science of Boston University in Boston, MA. More details are available at http://www.cs.bu.edu/groups/ivc/ImageRover. It is an experimental Web image search engine that combines textual information and image content. Experiments on its search performance have been carried out for a database of 100,000 images collected from the Web. It uses an LSI (Latent Semantic Indexing) vector to represent the content of the HTML containing the image, in addition to color histograms, orientation histograms, and texture direction to represent image content. Principal Component Analysis (PCA) is used to reduce the dimension of features, and *k-d* trees are used for indexing. ImageRover allows the combination of textual queries with content-based queries. The user must begin by providing a textual query, after which textual queries, content-based queries, or a combination of the two can be used for refinement. Relevance feedback allows the combination of several images in the same query, selection of the appropriate L_m Minkowski distance, and assignment of more importance to concentrated features.

3.2. Diogenes

Developed in the Department of Electrical Engineering and Computer Science of the University of Illinois at Chicago in Chicago, IL, Diogenes [Aslandogan and Yu 2000a, 2000b] is a Web-based image search agent designed specifically for facial images of celebrities. More details about it can be found at http://ranger.uta.edu/~alp/. The user first introduces a textual query consisting of the name of the celebrity to look for. Then the system goes through its base of descriptors and retrieves pages containing images of the person sought. To accomplish this, Diogenes possesses crawlers which initially submit the query to full text search engines such as Google, AltaVista, and Snap. It then merges the returned list of Uniform Resource Locators (URLs) and visits each URL for analysis. A face detection module analyzes each page to detect facial images. If a facial image is detected, a face recognition module analyzes the faces extracted from it, and a text/HTML module analyzes the surrounding text to locate person names and determine their degree of association with each image. This allows Diogenes to associate each image containing a face with a person name. Diogenes already possesses some typical images of famous people.

3.3. Atlas WISE

Atlas WISE [Kherfi et al. 2002, 2003a, 2003b] is a Web image retrieval engine under development at the Department of Mathematics and Computer Science of Université de Sherbrooke in Sherbrooke, P.Q., Canada. It starts the data gathering step by browsing popular index pages, such as Google and Yahoo! and downloading both text and images. The downloaded images and data are preserved until they are digested, after which they are deleted in order to save storage space; only thumbnails and links are kept for later use. Atlas WISE uses both visual and textual features to describe and index images. The visual features used include color histograms and edge-orientation histograms. As for textual features, keywords relevant to images are estimated from their tags and captions, page titles, and the surrounding text.

For visual retrieval, this system relies on a relevance feedback algorithm that combines positive and negative examples for query formulation. In addition to positive example images, the user can add negative example images to a query, thus allowing him or her to express the undesired images or features.

3.4. WebSeer

Developed at the Department of Computer Science of the University of Chicago in Chicago, IL, WebSeer [Athitsos and Swain 1997; Frankel et al. 1996; Swain et al. 1997] is a Web image search engine that uses image content and associated text to index images. This system tries to distinguish between photographs and graphics on the basis of certain tests on image content. In the retrieval stage, the user has to give a textual query coupled with some attributes of the sought image such as its dimensions, file size, type (photograph or graphic), and dominant colors. In addition, if the user is looking for people, he or she can specify the number of faces and the portrait size. WebSeer begins by retrieving every image whose associated HTML contains at least one of the query's keywords. Then the results are sorted by the sum of these words' weight and refined on the basis of image attributes (dimensions, size, type, etc.). WebSeer possesses a crawler that traverses the Web, downloading both HTML pages and images. Each image is indexed by weighted keywords and then analyzed by the *Image processing server*, which attempts to distinguish between photographs and graphics and to detect people's faces.

3.5. WebSeek

Developed at the Image and Advanced Television Lab at Columbia University in New York, NY, WebSeek [Smith and Chang 1996a, 1997] is specialized in retrieving images and videos from the Web. At the beginning, the user can specify only textual queries. WebSeek conducts an initial search by looking in its catalog for images related to the query keywords. A results list is returned to the user, who can manipulate it by ordering textual refinement via logical operations on SQL queries, or by ordering visual refinement by specifying sample images. In the case of textual refinement, the system continues searching in the catalog. However, in the case of visual refinement, the resulting images are sorted by order of similarity to the query image. WebSeek uses color histograms, image attributes such as length and width, and video attributes such as the number of frames. A variant of the query-point movement technique is used for relevance feedback. WebSeek classifies images and videos by subject. This is achieved by analyzing pages' URLs in order to extract the image name and directory which will be considered as *terms* and used, through a combination of automated and semi-automated procedures, to build the catalog, composed of a taxonomy and a dictionary. The taxonomy is a hierarchical representation that catalogs the terms by subject (such as sports or astronomy), and subsubject (such as soccer or football or stars or planets). The dictionary establishes links between terms that are relevant to the same subject. The catalog constitutes a structure that allows users to navigate through different subjects.

3.6. ImageScape

Developed at the Department of Computer Science of Leiden University in The Netherlands, ImageScape [Lew 2000; Lew et al. 1997] is a system implemented to find images over intranets and the World Wide Web. Details are available at http://www.wi.leidenuniv.nl/home/lim/image.scape.html and demos at http://skynet.liacs.nl. In addition to keyword-based queries, ImageScape allows for queries based on semantic icons, in which the user can choose an icon and place it on a canvas in the position where it should appear in the target image, and queries based on a sketch paradigm, in which the user creates a query by drawing a rough sketch to describe object edges and contours within the sought images. ImageScape uses global shape matching with invariant moments, and local shape matching by elastic deformation energy. Images are indexed using a *k-d* tree.

3.7. PicToSeek

Developed at the Department of Computer Science of the University of Amsterdam in The Netherlands, PicToSeek [Gevers and Smeulders 1997, 1999; Gevers et al. 2000] is a Web image search engine that relies on image content and relevance feedback. Many visual features can be chosen and weighted explicitly by the user. The system tries to distinguish between photographs and graphics. Details are available at http://www.science.uva.nl/research/isis/pictodeek; and demos at http://zomax.wins.uva.nl:5345/ret_user/. In PicToSeek, the user can give only queries by example, by first choosing an image from the database of typical images or providing a query image URL, then selecting its type (photograph, graphic, or person's face), its important features, and the desired similarity method. PicToSeek uses the correlation function and histogram intersection to find the *k-nearest neighbors*. Relevance feedback is performed by query-point movement. Images gathered from the Web are classified according to their attributes, such as their dimensions, creation date and color depth. Another classification, based on color variation, color saturation, and color transition strength, is done to distinguish between photographs and graphics. The system uses Fisher's linear discriminate method for classification and SR-trees for indexing.

3.8. WebMars

Developed at the Department of Computer Science of the University of Illinois at Urbana-Champaign, IL, and the Department of Information and Computer Science of the University of California at Irvine, CA, WebMars [Ortega-Binderberger et al. 2000] is a multimedia Web search engine that utilizes both textual and visual information for HTML document retrieval. WebMars retrieves the whole HTML document rather than only the images found in it. The user may initially submit a coarse query using one media type (image or text), and then use successive refinements to search for related material in different or composite media. Different kinds of media are combined within the same query using a tree representation whose branches are media types. Each media type is further divided into features (e.g., color and texture for images). Relevance feedback is applied in query definition by performing query-point movement. The proposed approach can be extended to introduce other kinds of media such as audio and video. This could help narrow the gap between the different kinds of media by integrating them into the same framework, thereby enabling cross-media browsing and retrieval.

3.9. Web-WISE

Developed at the Department of Computer Science of Wayne State University in Detroit, MI, Web-WISE [Wei et al. 1998] is a content-based image retrieval system for the World Wide Web. The user can give either sample images or directly estimated values for the features she or he is interested in, and weight each of these features. The system exploits color and texture features extracted directly from the JPEG compressed color images. It returns to the user the set of thumbnails corresponding to the images most similar to the query.

3.10. RIME

Developed at the Department of Computer Science of Stanford University in Stanford, CA, RIME [Chang et al. 1998] is a good example of Web image retrieval applications. It is a replicated image detector whose objective is to detect unauthorized image copying on the Internet. RIME crawls the Web periodically, collects images, digests them, and stores the feature vectors and URLs of the images in its repository. When a copy detection request is received, RIME matches the requested image's feature vector with the vectors stored in the repository and returns a list of suspect URLs. RIME characterizes each image using Daubechies' wavelets, and uses a multidimensional extensible hashing scheme to index these high-dimensional feature vectors. As this system aims to detect other copies of the same image, it uses matching rather than similarity.

3.11. DrawSearch

Developed at the Department of Electrical and Electronic Engineering of the Technical University of Bari, Italy, DrawSearch [Di Sciascio et al. 1999] is a content-based image retrieval system for the Web that uses color, shape, and texture. It allows query by sketch using color and shape, as well as query by texture content. The user can refine the results by relevance feedback that uses query-point movement.

3.12. The Web Image Search Engine from Monash University

Developed at the Gippsland School of Computing and Information Technology at Monash University in Churchill, Australia [Lu and Willams 1999], this system is a Web image search engine that combines text and image content. A description of the system can be found at http://ausweb.scu.edu.au/aw99/papers/lu/paper.html. Keywords related to images are estimated from image attributes, page title, page headings, image URL, alternate text, other words with high frequency of appearance, etc. Image content is characterized with color histograms. The same query is processed using color only, then using keywords only. The images resulting from the two interrogations are then combined and the ones in the top positions are returned to the user.

3.13. Summary

In Table I, we give a summary of the main characteristics of each of the cited systems.

4. ISSUES RELATED TO IMAGE RETRIEVAL FROM THE WEB AND TECHNIQUES ADOPTED TO ADDRESS THEM

We saw in the previous section that there have been some attempts in recent years to build tools for image retrieval from the Web. But what are the main issues that must be considered in developing a Web image retrieval engine? Anyone who wants to design and implement a Web image retrieval engine must think about a number of issues and resolve certain problems. The aim of the current section is to identify and shed light on each of these issues, discuss the important underlying questions, propose some solutions where possible, and give some examples of the techniques adopted by existing systems. Many of the issues around Web image retrieval are related to the tasks the system has to carry out, which include data

Table I. Summary of the Main Characteristics of the Cited Systems

System	Main characteristic	Descriptors	Query style	References
ImageRover	Experimented on 100,000 images	—Pictorial —HTML text	—textual —By example images	Sclaroff [1995] Sclaroff et al. [1997]; Taycher et al. [1997]; La Cascia et al. [1998]
Diogenes	Specific for celebrities' faces	—Human faces —HTML text	—Textual: celebrity's name	[Aslandogan and Yu 2000a, 2000b]
Atlas WISE	Combines positive and negative examples	—Pictorial —HTML text	—Textual —By example images	Kherfi et al. [2003a, 2003b]
WebSeer	Distinguishes photographs from graphics	—Pictorial —HTML text	—Textual enriched with image attributes	Athitsos and Swain [1997]; Frankel et al. [1996]; Swain et al. [1997]
WebSeek	Retrieves images and videos	—Pictorial and image attributes —Page URL	—Textual —By example images	Smith and Chang [1996a, 1997]
ImageScape	Retrieves images from intranets and the WWW	—Image content —Image tag text	—Textual —By icons and sketches	Lew [2000]; Lew et al. [1997]
PicToSeek	Distinguishes photographs from graphics	—Pictorial	—By example images	Gevers et al. [2000]; Gevers and Smeulders [1997, 1999]
WebMars	Retrieves whole HTML documents	—Pictorial —HTML text	—Textual —By example images	Ortega-Binderberger et al. [2000]
Web-WISE	Allows queries by estimated feature values	—Pictorial	—By example images —By estimating features' values	Wei et al. [1998]
RIME	Replicated image detector	—Pictorial	—By requested image	Chang et al. [1998]
DrawSearch	Allows queries by sketch	—Pictorial	—By sketch or by texture content	Di Sciascio et al. [1999]
Monash	Combines images retrieved by text and those by content	—Pictorial —HTML text	—Textual —By example images	Lu and Williams [1999]

gathering, image digestion, index creation, retrieval and refinement, as illustrated in Figure 2.

First, images and related data must be collected from the Web; then this data is analyzed (digested) to compute image descriptors which will be used for both retrieval and indexing. Once this is done, the system will be ready to receive users' queries and to process them using the adequate similarity measures. After that, the retrieved images are shown to users who should be allowed to refine them via relevance feedback. Finally, when the system is designed and implemented, an evaluation of its performances should be carried out. Some other issues are very specific to the Web such as making sure the engine covers it globally. We can summarize the main issues discussed in this section as follows: data gathering, the identification and the estimation of image

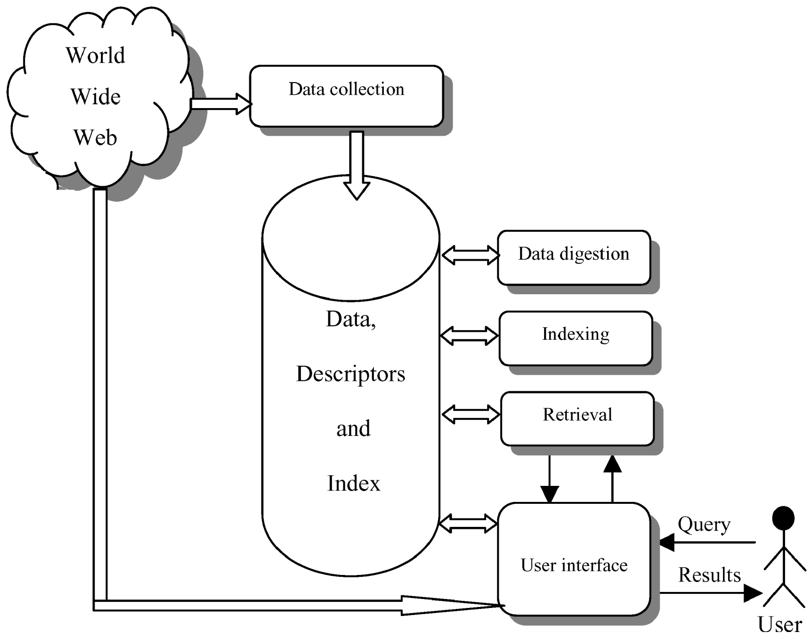

Fig. 2. General structure of a Web image search engine and its main tasks.

descriptors, similarity and matching, indexing, query specification, retrieval and refinement, Web coverage, and performance evaluation.

4.1. Data Gathering

In the context of the Web, operations like image gathering, feature extraction, and index creation should be done off-line, because no system can wait until the user gives the query to start traversing the Web looking for images. Indeed, given the huge size of the Web, this could take several days without giving satisfactory results. The only way to deal with this is to collect data and compute image descriptors before users introduce their queries. Concerning data gathering, a Web image retrieval engine should have a module (generally called the *crawler*) that regularly traverses the Web collecting the needed data [Aslandogan and Yu 2000; Frankel et al. 1996; Kherfi et al. 2003a]. Depending on the descriptors to be used, the data collected may be the images themselves, surrounding text, URL addresses, or other types of data. The main issues related to data gathering that the system designer must consider are Web traversing, dealing with special links, data storage, data gathering time, image validity, image conversion, thumbnail creation, and data updating. Each of these issues is discussed below.

4.1.1. Starting and Traversing the Web. The crawler views the Web as a graph, with the nodes being the objects located at Uniform Resource Locators (URLs). These objects could be HTTPs (HyperText Transfer Protocols), FTPs (File Transfer Protocols), mailto (email), news, telnet, etc. [Hu 2003]. The technique used in traversing the Web may be one of the following [Gudivada et al. 1997]:

—Providing the crawler with a "seed URL" to initiate exploration. The crawler identifies relevant data items in the seed document, copies them, extracts URLs pointing to other documents, then examines each of these URLs recursively. This technique is used by the WebSeek system, in which the crawler starts from seed URLs and follows a breadth-first search across the Web.

—Starting with a set of URLs found in popular sites (such as the home pages of existing retrieval engines) and searching recursively. For example, the system ImageRover contains software robots that regularly traverse the Web and gather images, starting with the Yahoo! directory. The systems PicToSeek and Atlas WISE adopted the same technique.

—Partitioning the Web space on the basis of Internet names or country codes and assigning one or more robots to explore the space exhaustively. We note that this last technique can ensure a better coverage of the Web than the other two; however, it have some drawbacks. First it can be very time-consuming, and second, intrusion detection mechanisms may be triggered and countermeasures activated.

4.1.2. Dealing with Special Links. In addition to ordinary URLs, the crawler may encounter several kinds of special links that cannot be developed automatically without prior processing. In interactive pages, for example, it is sometimes necessary to introduce parameters or to fill out a form to be able to reach a given page. If such parameters are included in the URL itself, as in Get statements, the crawler can follow the link without any problem. On the other hand, if the parameters must be introduced by the user (the crawler in our case), as in the case of a form to be filled out, it is often impossible for the crawler to guess the possible parameters, unless it knows beforehand the values each field can take. JavaScripts are another example of objects that can embed special links that may prove difficult to extract. Most existing text retrieval engines are unable to follow such links. A possible solution could be to develop specialized modules that analyze the source code of the functions corresponding to each JavaScript in order to extract the embedded links. Other special links that the crawler has to deal with can be found embedded in Flash animations, which are becoming increasingly frequent. Fortunately, tools for extracting such links are now available.

4.1.3. Data Storage. In recent years, considerable advances have been made in data storage technology, resulting in greater storage capacity and faster data transfer. However, given the size of the Web, keeping a permanent copy of every image and piece of data collected would be inconceivable. A solution could be to keep a copy of each image and its corresponding data until they are digested, then to delete the image and data [Kherfi et al. 2003a]. In order to be able to recall each image in the future, their URL addresses should be kept within their corresponding descriptors [Lew 2000].

4.1.4. Data Updating. The World Wide Web is changing rapidly, and every day data is added, modified, or even deleted. To stay up to date, the Web crawler should traverse the Web regularly to update the collected data, so that the stored descriptors and the index can be kept up to date [Smith and Chang 1997].

4.1.5. Data Gathering Time. To reduce data gathering time, many copies of the same program (usually called *robots*) can be launched, each traversing a different part of the Web [La Cascia et al. 1998].

4.1.6. Image Conversion. Images are present on the Web in different formats, including JPG, GIF, and TIF. On the other hand, a digestion module is generally implemented to deal with a single format. Hence, the different formats of images must be converted to the format(s) supported by the digestion module before it can digest them [Kherfi et al. 2003a].

4.1.7. Thumbnail Creation. As the images are deleted after being digested, the system can keep a compact copy of each image (a thumbnail) requiring minimal storage space, in order to be able to present the retrieval results to the user [Smith and Chang 1997; Lew 2000; Wei et al. 1998]. Of course, the system should allow users to see the image in its original state by sending them to its original location. In WebSeek for example, after each image or video is digested, a specialized module creates an icon that represents the image or a shot that summarizes the video.

4.1.8. Image Validity. Not all the images found on the Web are valid for a given application. First, there is matter of the image's dimensions [Sclaroff et al. 1997; Sclaroff 1995]. In the ImageRover system, for example, to be gathered, an image must not be under 64×64 pixels. A more complicated criterion is whether an image's purpose is valid for the current application. Indeed, in addition to content images (those associated with the page's subject), images on the Web may be there for advertisement, decoration (e.g., buttons, balls, rules, masthead), information (e.g., under construction, warnings, what's new), navigation (e.g., arrows, back to home, image maps), or identification (e.g., a company's logo) [Paek and Smith 1998]. For some applications, only content images are of interest, and hence the others must not be gathered. However, deciding whether a given image is a content image sometimes requires sophisticated image analysis algorithms, as discussed in Section 4.2.1. Thus, the image gathering module can extract all the images it finds, and leave the job of differentiating between valid and invalid images to another specialized module.

The size of the Web, which does not cease increasing, is among the most important challenges that will be facing data gathering in the future. Indeed, in order to minimize gathering time and storage space, techniques should be developed to identify, among all the data available on the Web, only those items which have undergone a change and to update the collected images and data.

4.2. Image Descriptors

After images and data have been collected from the Web, the system designer has to think about the "digestion" of these data in order to extract features (or descriptors) that will be used to index images and to retrieve them. These features can be extracted from the images themselves or from the surrounding text. Depending on where the features are extracted from, we can distinguish two main image retrieval approaches that are valid for the Web case. In the first approach, images are first annotated with text, then text retrieval techniques are used to perform image retrieval. This approach, known as *text-based image retrieval*, dates back to the late 1970s and is due to the database management community [Rui et al. 1999]. Some commercial systems, such as Google image search and AltaVista photo finder, use this technique. The second approach involves using image content, such as color and shape. This approach, known as *content-based image retrieval*, was proposed in the early 1990s and comes from the computer vision community. Examples of systems that adopted this approach are PicToSeek, Web-WISE and DrawSearch. Both techniques can be combined in the same framework, as in ImageRover, Diogenes, WebSeer, WebSeek, or Atlas WISE. Image attributes other than text and visual content can also be exploited for retrieval and refinement. In WebSeer for example, image dimensions, color versus gray-scale, file type (JPEG, GIF, etc.), file size, and file date are used. In WebSeek, the image and video attributes used include length, width, number of video frames, and data type. A fundamental question when designing a Web image retrieval system is what features to use. It is well known that, to classify images, people may use high-level semantics, image content, or both [Eakins and Graham 1999]; and that high-level semantics can be better extracted from text, while image content is better described by visual features. Hence, we think that, to be able to do effective image retrieval from the Web, the system designer must understand the needs of the ultimate users of the engine and the way they categorize images. In the remainder of this section, we will discuss some issues related to textual features and pictorial features in the context of Web image retrieval.

4.2.1. Textual Features. Users may be interested in high-level attributes of an image, such as the depiction of a given emotion, that cannot be directly derived from the image content using current techniques. In many cases, such attributes can be derived from the text accompanying the image, such as its caption and surrounding text. Many words found in a Web page may be related to images found or referenced in the same page, as well as images pointing to this page. Some systems such as Google image search extract keywords related to images from the Web page text and the image caption. In fact, these keywords can be found in many locations. In WebSeer, for example, keywords are extracted from the image tag, the page title, the image file name, the image caption, the alternative text, the image hyperlinks, and the body text titles. Each keyword is given a weight according to where it was encountered in the page. WebSeek analyzes the URL of the page containing an image to retrieve the image name and directory. ImageScape extracts keywords from the image tag text. Atlas WISE exploits image tags, captions, and page titles. Diogenes gives more significance to titles, words that are part of the image name or URL, and words that are enclosed in the same tags as the image or in the alternate text. In addition to special words, the body text of a Web page can be exploited, and words relevant to the image can be identified based on criteria such as the frequency of occurrence of the word on the page, its position with respect to the image, and its style (bold, italic, etc.). In Diogenes, for example, the importance of a word is estimated according to its absolute frequency, that is, within the Web page, and relative frequency, that is, its frequency on the page relative to its frequency on the Web as a whole: if a rare word appears frequently on a page then it may be very significant for that page. In ImageRover, words relevant to an image are identified based on their frequency of occurrence in the HTML text, their position with respect to the image, and their style. In fact, making effective use of textual keywords can render image retrieval by high-level concepts more efficient, and makes it possible to apply existing text retrieval techniques to image retrieval. The key issue is: given a Web image together with its surrounding text, how can relevant words be distinguished from irrelevant ones? Rowe and Frew [1998] developed a system dedicated to locating words related to images in Web pages. In the first step, their system has to decide whether an image is a photograph or not, on the basis of certain criteria such as height/width ratio, colors, and whether the name of the image file contains certain common words. This distinction is useful because a photograph is more likely to be annotated than a nonphotograph. The second step involves identifying keywords by examining the text near each image reference. Keywords are extracted from marked text which is identified according to the font family (e.g., Times Roman), font style (e.g., italics), font size (e.g., 12 point), text alignment (e.g., centering), text color (e.g., blue), text state (e.g., blinking), and text significance (e.g., a page title). The authors also considered alternative text, names of Web pages pointed to by the image, and the name of the image file itself. Conversely, someone may be interested in deciding whether an image is related to the textual content of the page. Peak and Smith [1998] investigated some issues related to this question. To be able to classify images into content images (related to page text) and noncontent images, they use a classification algorithm based on a decision tree. The learning is performed manually, and then the classification is performed on the basis of text keywords extracted from the image URL and alternate text, and image attributes such as format, dimensions, colors, and the occurrence of certain labels such as advertisement and decorative labels.

Even though the extracted keywords are related to the images, their use is not without problems. The first concerns their subjectivity. Indeed, textual descriptions are generally added manually by the Web page creators. Such manual annotation depends greatly on the annotator, because two different people may associate

Table II. Features Used by Existing Systems

System		Image Rover	Diogenes	Atlas WISE	WebSeer	WebSeek	ImageScape	PicToSeek	WebMars
Keywords		*	*	*	*	*			*
Face detection			*		*		*		
Layout		*					*		*
Shape	Fourier descriptors						*		*
	Elementary description						*		
	Angles between edges, and cross ratios of them							*	
Texture	Wavelet, Fourier transform								*
	Edge-orientation histogram	*		*					
	Local binary patterns						*		
	Automatic texture features	*					*		*
Color	Eigen image								
	Dominant colors						*		
	Region histogram								
	Fixed subimage histogram	*							*
	Farthest neighbor histogram				*				
	Global histogram	*		*	*	*		*	*
	Laplacian						*		

different keywords with the same image. In fact, keywords sometimes say more about the person who assigned the keywords to an image than they do about the image [Jain 1995]. Tools for automatic annotation of images and standards for metadata description can help solve this problem. The second problem concerns the exclusive use of these keywords, which may prove insufficient, especially when the user is interested in the visual components of an image. In most images, there are several objects that can be referenced, and each object has a long list of attributes. In addition, there are many image properties that defy description in words: textures, composition, unknown objects, etc. In this situation, the use of visual features seems to be indispensable to reinforce textual descriptors. These features are discussed in Section 4.2.2.

4.2.2. Pictorial Features. Using current computer vision techniques, many features related to image content can be directly derived from images. These features include color, orientation, edges, texture, shapes, etc. Numerical descriptors are generally assigned to pictorial features in the form of real vectors of variable dimension. Several good surveys on pictorial features are available in the literature, such as Schettini et al. [2001] and Gevers and Smeulders [1996] for color, Tuceryan and Jain [1993] and Randen and Husoy [1999] for texture, and Marshall [1989] and Mehtre et al. [1997] for shape. Many of these features have been used for image retrieval purposes. However, we think that the main underlying task is to choose the appropriate features that best fit the target application, because this can considerably improve retrieval efficiency. For some specific applications of image retrieval, the choice of features is made easier by the homogeneity of the images to be processed. In a fingerprint-recognition application, for example, texture features could be sufficient; for a face retrieval application, shape descriptors may do the job. In the context of Web retrieval, however, the choice of features is more difficult because of the great heterogeneity of the images to be processed. A multitude of features have been used by existing Web image retrieval systems, as illustrated in Table II, which was derived from the summary given in Veltkamp and Tanase [2000] but tailored to the systems described in Section 3.

In order to be able to perform retrieval in the highly heterogeneous collection of images present on the Web, the system designer may think that combining the maximum number of features will make it possible to cover the different kinds of Web images. However, this can cause a serious problem known as *the curse of dimensionality*. Indeed, the great number of features and the dimension of each of them render most existing indexing and retrieval techniques unusable, as we will see in Section 4.4. Hence, the system designer has to find a good compromise consisting of a set of features that is at once large enough to describe the multitude of image categories that are present on the Web, but not so large as to degrade the system's efficiency and performance. The designer must also consider using some techniques for reducing feature dimension and performing some kind of feature selection. In addition to the number of features and their dimensions, other factors can intervene in the choice of image descriptors. For example, depending on the target application of the Web image retrieval engine, the designer has to choose between local and global features. For some applications in law enforcement, region-of-interest-based retrieval is more suitable, for example, looking for all images containing a given weapon, whatever the background. In such applications, descriptors have to be region-based, that is, each image is first segmented into a set of regions, then pictorial descriptors are computed for each region. For other applications, such as image copyright protection over the Web, global-image-based retrieval seems to be more suitable. In such a case, descriptors can either be global or region-based. Global descriptors are extracted from the image without any prior segmentation; while region-based descriptors are extracted from each region separately. It has been noted that, in general, region-based descriptors perform better for images of distinctive objects while global features outperform for images of distinctive scenes [Kherfi et al. 2003b; Carson et al. 1999; Jing et al. 2002b].

In short, meticulous attention should be paid to the choice of the set of pictorial features to be used in a Web retrieval system, and techniques for dimension reduction and feature subset selection may prove essential. Furthermore, more research is needed in order to be able to automatically extract semantic concepts from image content and surrounding text and data, which would allow retrieval by high-level concepts.

4.3. Similarity and Matching

When a user submits a query, the system goes through the database containing features that represent the images gathered from the Web in order to retrieve all images that correspond to this query. If the images are indexed, retrieval time will be reduced because the system will not have to traverse the whole database but only its index. The first question the system designer has to answer is: "should the retrieval system use matching or similarity to find the desired images?" In traditional databases, matching is a binary operation that consists of comparing an item with the query and deciding whether the item satisfies the query or not [Santani and Jain 1996]. In a similarity-based search, on the other hand, items are ordered with respect to their similarity to the query, given a fixed similarity criterion; and the results consist of the items most similar to the query. In the context of Web image retrieval, if the retrieval is text-based, then the techniques used with text databases can be applied to compare keywords. If, on the other hand, retrieval is content-based, then both similarity and matching are possible. The choice between similarity and matching depends on the target application and on the needs of its users. In some cases of target search, such as Web image copyright protection, the system may be asked to find other copies of the same image. In such cases, matching can be applied because the goal is to identify other copies of the same image rather than other similar images. However, requiring robust matching can lead to some target images being missed because there may be a small difference between the target image and the query due to some accident, imperfection, or manipulation. We think that in this situation matching can be applied but with some tolerance for imperfections. On the other hand, for category search, as in the case of a user who wants to find more images on the Web that are related to the same subject as his or her query, matching will not be useful because the compared images may be similar but not necessarily identical, or they may even be identical but one of them may have undergone some geometric transformation or change in illumination. In this case, it is more suitable to use similarity rather than matching.

We believe that, for most Web image retrieval applications, similarity is more suitable than matching. In the event that similarity is adopted, another question emerges: "how can similarity be measured?" In fact, similarity is a cognitive process. Understanding it requires knowledge of image analysis as well as other background knowledge from other areas of science. A good place to look for the characteristics of human similarity is in the psychological literature. Because similarity perception is a complicated activity, it is difficult to give it a unique characterization. Several similarity measures have been proposed in the literature for content-based image retrieval, including those by Santani and Jain [1996], Thurstone [1927], Ennis and Johnson [1993], Tversky [1977], Ashby and Persin [1998]. Amsaleg et al. [2000], and Candan and Li [2001]. Most of these measures remain valid for the context of image retrieval from the Web. However, choosing one or more of them depends on the ultimate end for which the Web image retrieval engine has been designed as well as the needs and specific characteristics of its users.

4.4. Indexing

After the image descriptors are computed, they must be saved on a permanent storage device. The great number of Web images, the size of each image descriptor, and the use of a storage device make it necessary to use effective indexing techniques to facilitate access and thereby reduce the retrieval time. Indeed, using an index can help to minimize the retrieval time dramatically because the system will not have to go through the whole database of features when it looks for images, but can simply retrieve the most similar images according to the index. Existing multidimensional indexing techniques include the bucketing algorithm, k-d trees, priority k-d trees, quad-trees, K-D-B trees, hB trees, R-trees, R^- trees, R^* trees, and X-trees [Smeulders et al. 2000; Rui et al. 1999; Berchtold et al. 1996]. Clusterering and neural networks can also be used in indexing [Duda et al. 1973; Zhang and Zhong 1995]. More details about indexing methods can be found in Gong [1997]. Existing Web image retrieval systems use some of these techniques: for instance, ImageRover and ImageScape use k-d trees, and PicToSeek uses SR trees. The most serious challenge in Web indexing is how to deal with the huge number of images on the Web and the heterogeneity of these images. Indeed, in order to be able to locate the different kinds of images available on the Web, a combination of several features of high dimensions may be required. However, increasing the number of features and their dimensions, added to the great number of images to be indexed, can render most existing usual indexing techniques unusable. Weber et al. [1998] showed that when the number of dimensions exceeds 10, sequential search becomes better than any other index navigation method. Amsaleg et al. [2000] cited the following problems that arise when comparing images: first, the great dimension of feature vectors makes it difficult to compare them with the query; and second, the use of similarity rather than matching renders the situation more difficult because the system will have to explore the space of descriptors rather than searching for a unique point. Some recent indexing methods such as VA-File [Weber et al. 1998] and Pyramid-Tree [Berchtold et al. 1998] have been proposed specifically to circumvent the problem of dimensionality.

However, Amsaleg et al. [2000] showed that, in large databases of high-dimension descriptors, sequential retrieval still outperforms these techniques in terms of retrieval time.

Other solutions entail prior manipulation of the image descriptors. Before indexing data, the great dimension of feature vectors can be attenuated by performing dimension reduction. Two main approaches have emerged in dimension reduction, namely, the Karhunen-Loeve Transform (KLT) and column-wise clustering [Rui et al. 1999]. In addition to reducing the dimension of feature vectors, the system can totally eliminate some features through what is called *feature selection*, that is, eliminating redundant features or those that provide little or no predictive information. Amsaleg et al. [2000] suggested, in the retrieval stage, not exploring neighbor cases in a descriptor space; truncating the judgment and accepting an image when a sufficient number of its descriptors correspond to the query; using the fact that one descriptor may provide more information than another about a given image; using the results of previous interrogations; and managing many small-dimension indexes rather than a single large index.

We conclude that more work is needed before it will be possible to index the entire visual content of the World Wide Web, especially because of its size and diversity.

4.5. Query Specification

Query specification is an issue for information retrieval in general that remains valid for the Web image retrieval problem, and the system designer must take it into account. Interaction with users in information retrieval systems has been studied in [Pao and Lee 1989], for example. In what concerns fuzzy queries in multimedia databases, some interesting results are given in Fagin [1998]. In the context of Web image retrieval, if the retrieval is text-based, the user has to introduce a textual query consisting of keywords or a description of the sought images. Queries of this type are supported by many Web image retrieval systems, such as ImageRover, Diogenes, Atlas WISE, WebSeer, WebSeek, ImageScape, and WebMars. In the Diogenes system, for example, the user must enter the name of the celebrity she or he is looking for. If, on the other hand, retrieval is image-content-based, query formulation may be more difficult and a number of approaches have been proposed in the literature. Early CBIR (content-based image retrieval) systems such as QBIC [Flickner et al. 1995] ask the user to select image features such as color, shape, or texture. Other systems like BLOBWORLD [Carson et al. 1999] require the user to provide a weighted combination of features. This approach has been adopted by some Web image retrieval systems such as Web-WISE, in which, in addition to queries based on sample images, the user can directly estimate values for the features he or she is interested in and weight each of these features. This method of direct estimation of features' values and importance has proven to be limited, however, because it is generally difficult to directly specify the features needed for a particular query, for several reasons. First, not all users understand the image vocabulary (e.g., contrast, texture, color) needed to formulate a given query. Second, even if the user is an image specialist, it may not be easy to translate the images he or she has in mind into a combination of features. Another approach in query specification involves allowing the user to specify the features and their corresponding weights implicitly via a visual interface known as *query by example* [Kherfi et al. 2003b; Ishikawa et al. [1998]. The user can choose the images that will participate in the query and weight them according to their resemblance to the images sought. This approach has been adopted by some Web systems, such as ImageRover and WebSeek. As for Atlas WISE, with it we investigated the combination of positive example images and negative example images in query formulation.

Adopting query by example, however, leads to the emergence of another difficulty known as the *page zero problem*: that is, finding a good query image to initiate the retrieval. To avoid falling into this problem, the system must provide the user with a set of candidate images that are representative of the entire visual content of the Web, for example, by choosing an image from each category. If none of these images is appropriate, the user should be allowed to choose to view more candidate images. Another solution to the page zero problem, used in the PicToSeek system, involves asking the user to provide an initial image; however, it is not always easy for users to find such an image. ImageRover and WebSeek ask the user to introduce a textual query at the beginning, after which both textual and visual refinement are possible. In Kherfi et al. [2003b], we showed that the use of a positive example together with a negative example can also help to mitigate the page zero problem.

Other approaches have been proposed and used for query specification, including the two approaches used in ImageScape. In the first approach, queries are based on semantic icons. The system offers some representative icons such as *the blue sky* and *the beach*, then the user chooses an icon and places it on a canvas in the position where it should appear in the target image. This allows the user to explicitly create queries for images of people on the beach, for example. In the second approach, queries are based on sketches. The user creates a query by drawing a rough sketch to describe object edges and contours in the sought images. This allows the user to directly specify which part of the image is important.

Before closing the discussion of query specification, we note that it is often useful to allow the user to combine text with visual image content in formulating the query. This might allow the user to simultaneously retrieve based on high-level semantics and the low-level pictorial content of images, thereby narrowing what is called *the semantic gap* (to be discussed in Section 5.2). Many of the existing Web image retrieval systems, such as ImageRover, WebSeek, Atlas WISE, ImageScape, and WebMars, allow for the combination of text-based queries with content-based ones. This subsection would not be complete without a word about the simplicity of the user interface. Users interact with the system via its interface, using it to specify their queries and then to view the retrieved images. Hence, the interface must be as easy to use as possible, both at the query specification stage and at the result presentation stage. Finally, we note that it is of great interest to perform studies on users and, the style of interaction they prefer with the retrieval system, because this could allow users' needs to be met by designing interfaces that allow them to specify queries in an easier and more accurate way.

4.6. Retrieval and Refinement

Retrieval concerns how the system goes about locating images that respond to a given query. As discussed in Section 4.3, except in some special applications such as detection of replicated images on the Web, systems for Web image retrieval operate on the basis of similarity rather than matching, and as a result two interrogation techniques are commonly used [Amsaleg et al. 2000]. The first is the retrieval of the *k-nearest neighbors* of each image descriptor and the second involves searching all descriptors whose distance from the query descriptor is less than a given ε. The first technique involves some problems in the general case, such as the choice of k and the fact that its results are not necessarily close to the query. Moreover, in the Web context, another serious problem emerges. The large size of the Web and the possible use of several features of high dimensions render the search for the nearest neighbors very time-consuming because it may be necessary to explore a large portion of the retrieval space to locate them. The second interrogation technique also suffers from a drawback when applied to the Web context. Although it does ensure that the returned results are near the query, it may give no result at all or—and this is the main problem—a very large number of results. Indeed, even though the Web may contain several thousand images that respond to certain queries, it is inconceivable to return all of these images to the user. The system designer must decide to truncate the results at a given rank, and is thereby brought back to the k-nearest-neighbors technique.

After the results of a given query are obtained, iterative refinement may be necessary to give the user the chance to increase the relevance of the retrieved images. This can be done through relevance feedback (RF), which is defined in Rui et al. [1997] as the process of automatically adjusting an existing query using information fed back by the user about the relevance of previously retrieved documents. Many image retrieval researchers have been interested in relevance feedback [Kherfi et al. 2003b; Rui et al. 1997; Ishikawa et al. 1998; Vasconcelos and Lippman 1999; Zhang et al. 2001], demonstrating how it aids in capturing the user's needs by helping to identify the ideal query that is in his or her mind, weight features, and define the corresponding similarity measures.

In the context of image retrieval from the Web, relevance feedback techniques make at least two other contributions. The first relates to retrieval time. Techniques such as query-point movement, in which the ideal query is moved toward positive examples and away from negative examples, can help identify the ideal query that is in the user's mind. Techniques such as axis reweighting can help determine which features are most important to the user. Identifying the ideal query and the most important features can considerably reduce the number of iterations required to retrieve the desired images, thereby reducing retrieval time. Given the great size of the Web, retrieval can be very time-consuming, and, hence, any solution that can help reduce this time should be welcome to the system designer. The second contribution has to do with the number of features and their dimensions. Feature weighting that is achieved thanks to relevance feedback can be applied to perform feature selection, by retaining only those features which are important enough and eliminating the rest. Furthermore, relevance feedback makes it possible to attribute weights to the various components of a given feature or transform the feature's space into another of lower dimension, thereby reducing the dimension of some features. By eliminating some features and reducing the dimension of others, many problems can be solved or at least alleviated. In the indexing stage, *the curse of dimensionality* can be surmounted because indexing techniques work better in low-dimensional spaces than in high-dimensional ones. At the retrieval stage, dimension reduction helps to improve retrieval performance by defining better similarity measures and reducing retrieval time.

Many improvements have been achieved recently by information and image retrieval researchers in the field of retrieval efficiency and refinement [Kherfi et al. 2003b; Jing et al. 2002a]. These results can be applied in the future to image retrieval from the World Wide Web.

4.7. Web Coverage and Image Metasearch on the Web

The Web continues to increase in size and thousands of pages are added every day. Because of this, the relative coverage of Web search engines is decreasing and becoming partial, with the result that search engines tend to excel in certain classes of queries but be less effective for other classes [Lawrence and Giles 1999]. In addition to the huge size of the Web, its lack of structure makes it almost impossible to reach certain pages. To overcome this problem, the first thing the system designer must do is to choose a good starting point from which to crawl the Web collecting images and data (cf. Section 4.1). Rather than starting with randomly chosen URLs, the system can start gathering with URLs found in popular Web sites such as the index pages of existing engines. Another solution that could ensure a better coverage of the Web would be to partition the Web space on the basis of Internet names or country codes and assign one or more robots to explore the space exhaustively. The drawback of this technique, however, is that it can be very time-consuming.

The problem of Web coverage can be handled differently by adopting what is called a *metasearch* approach, that is, combining the results of multiple search engines in order to give better results. A metasearch engine queries one or multiple search engines on the Web, then combines the results and presents them to the user [Beigi et al. 1998]. It serves as a common gateway, linking users to multiple search engines in a transparent manner. For an image metasearch engine, the different search steps can be summarized as follows. First, the user submits a query, which may be text-based, image-based, or a combination of the two. Second, a module called the dispatcher selects the search engines to be involved. If the dispatcher is manual, the selection is performed by the user; otherwise, the dispatcher selects the target search engines, either randomly or according to criteria such as the accuracy and speed of each engine during previous queries. Third, each query translator converts the query into the format accepted by its corresponding search engine. And finally, after the results are returned by the search engines, the user interface merges these results and displays them to the user. The ranking and number of images taken from each engine can be chosen randomly or according to some criterion such as the general accuracy of the engine.

Many metasearch engines for text retrieval have been proposed by researchers [Dreilinger and Howe 1997; Howe and Dreilinger 1997; Selberg and Etzioni 1997]. In the field of image retrieval, however, few metasearch engines have been cited in the literature; two that have are MetaSEEK [Beigi et al. 1998] and Inquirus [Lawrence and Giles 1999]. MetaSEEK is an image metasearch engine that supports text-based and image-content-based queries. The query dispatcher selects the target engines automatically according to the type of query submitted (image-based or text-based) and the score that each engine has obtained for the same images, or for similar images if there is no score available for the same images. MetaSEEK allows for search by example images, URLs, or keywords. The user can specify a value for maximum waiting time. The other system, Inquirus, queries multiple image search engines on the Web, downloads the actual images, and creates image thumbnails for display to the user. Inquirus handles image search engines that return direct links to images, and engines that return links to HTML pages. For engines returning HTML pages, Inquirus analyzes the text on the pages in order to predict which images are most likely to correspond to the query.

Since the Web size continues increasing and its structure does not cease becoming complicated, we think that more effort is needed to find traversing techniques that are efficient but at the same time ensure a good coverage of the Web.

4.8. Performance Evaluation

The last issue we will discuss in this section is that of performance evaluation. After the system is designed and implemented, its designer has to think about evaluating its performance to make sure it fulfills the functions it was intended for. In this subsection, we review the main techniques for evaluating information retrieval systems in general, examine how these techniques are applicable to the evaluation of Web image retrieval systems, and give some techniques that are specific to our context. In general, an information retrieval system evaluation test-bed consists of a collection of documents, a set of benchmark queries, a set of ground-truth relevance scores for the benchmark queries, and a set of evaluation metrics [Smith and Li 1998]. In the field of image retrieval, two image collections have been used by numerous research groups: images from the State of California Department of Water Resources [Ogle and Stonebraker 1995] and a professional stock of photographs from Corel. We also note the role of the MPEG-7 community in collecting test data for system performance evaluation. In the Web context, finding an image collection is not really a problem since the test can be performed on images collected directly from the Web. However, it is really difficult to compare the performance of two different systems because each of them may cover a different part of the Web. Concerning ground-truth acquisition, human judgment has generally been used to assess the relevance of images retrieved. This is applicable in the case of Web image retrieval systems.

As for evaluation measures, many measures used in information retrieval such as recall, precision, and their rank-adjusted alternatives [Harman 1995] have been adopted for image retrieval evaluation [Smith and Li 1998]. *Recall* is the number of relevant documents retrieved, relative to the number of relevant documents in the whole database. *Precision* is the number of relevant documents retrieved, relative to the total number of documents retrieved. Another parameter that can be used instead of precision is noise, that is, the number of irrelevant retrieved documents relative to the total number of retrieved documents. Some authors have drawn up the precision-recall curve $Pr = f(Re)$ [Smith and Li 1998]. A more effective system will yield higher precision for all values of recall. However, it has been observed that this measure is less meaningful in the context of image retrieval since recall is consistently low [Rui and Huang 2000; Huang et al. 1997]. This can be replaced by the precision-scope curve $Pr = f(Sc)$, where the scope is the number of images retrieved [Kherfi et al. 2003b].

In the context of information retrieval from the Web, conducting such experiments is more difficult for several reasons. First, Web search engines operate on different indexes, and the indexes differ in their coverage of Web documents. Second, it is extremely difficult to compute evaluation measures such as recall because, even if the total number of images on the Web could be estimated, the total number of images relevant to a given query is impossible to compute or estimate. Gudivada et al. [1997] proposed to compare Web retrieval effectiveness in terms of qualitative statements and the number of documents retrieved. However, we think that the number of retrieved documents is not necessarily expressive, because the problem with current systems is that they retrieve too many documents, of which only a small fraction are relevant to the user's needs. Furthermore, the most relevant documents do not necessarily appear at the top of the list of retrieved documents. According to Smith and Chang [1997], text-based Web search engines are evaluated on the basis of the size of the catalog, search speed and effectiveness, and ease of use. Hu 2003 identified the following requirements for a Web search system: effective and efficient location and ranking of Web documents, thorough Web coverage, up-to-date Web information, unbiased access to Web pages, an easy-to-use user interface, expressive and useful search results, and a system that adapts well to user queries. We can define other evaluation measures for retrieval performance:

—*Relevance and ranking of retrieved images*. Since humans are the ultimate users of the system, they should be convinced by the relevance of the retrieved images. In addition to this, close attention must be given to image ranking, so that the most relevant images appear in the top positions.

—*Refinement and number of iterations*. It is very useful for any system to give the user the opportunity to refine the search results. However, the number of iterations needed to locate good results must be as low as possible.

—*Retrieval time*. Retrieval must be as quick as possible in order to return the results in an acceptable time. Some image retrieval systems measure performance based on cost/time to find the right images [Smith and Chang 1996b].

—*Ease of use*. The visible part of a system is its interface. It must be easy to use in all stages: formulating text queries, selecting sample images that will participate in a query, and specifying regions of interest in these images, as well as viewing retrieval results.

If the Web image retrieval system offers a catalog for browsing, then this catalog can also be evaluated. Such evaluation can be based on characteristics like the total number of indexed images, the diversity of the subjects in the catalog and the subdivision of each level into sublevels, the possibility of moving from one subject to another, the relevance of indexed images to the titles they are coupled with, and the ease of use. Some of these measures are objective; for instance, the number of indexed images can be computed directly from the catalog. Others, such as the relevance of images to subjects, are subjective and can be estimated using human judgment as ground truth. A system that offers the summarization service can be evaluated in the same manner by asking a number of representative people to use it and then give their opinion about the summary the system provides for each video or image set.

We end this section with some remarks related to image retrieval in general that are valid for the Web. First, the test-bed must be large in scale to assess scalability, and balanced in image content to test the effectiveness of image features and the system's overall performance [Rui et al. 1999]. Second, a distinction should be made between the evaluation of overall system performance and that of a part of the system [Smeulders et al. 2000].

5. DISCUSSION, FUTURE DIRECTIONS, AND OPEN ISSUES

Despite the huge quantity of visual information on the Web, search engines dedicated to this kind of data are extremely primitive. We believe that the need to develop such tools is more vital than ever, to help users locate desired images in a reasonable time with acceptable precision. However, many issues need to be addressed and many research avenues remain to be explored. In this section, we give a brief discussion on future directions in this field.

5.1. Understanding Users and Queries

Users of a Web image search engine express their needs through queries which describe the images they want. Understanding these queries and analyzing the behavior of those using the search engine can be of significant help in meeting users' needs. Furthermore, by studying users of Web image search engines, it may be possible to find their different profiles and group them into categories. If such a segmentation is possible, and if it can be reliably established that different types of users require different styles of interaction with retrieval systems, then the system designer's task will be made considerably easier. Many authors have studied the queries and behavior of people looking for textual information on the Web [Goodrum 2000; Goodrum and Spink 1999; Jansen et al. 2000]. At the other end of the scale, studies have also been conducted on users of general-purpose image retrieval systems, in both nondigitized collections [Hastings 1995; Keister 1994] and digitized ones [Eakins and Graham 1999; Smeulders et al. 2000; O'Connor et al. 1999; Jorgensen 1999; Lawrence and Giles 1999; Gudivada and Raghavan 1995; Turner 1995]. However, few studies have focused on users of Web image retrieval systems and their queries. Smith and Li [1998] analyzed the queries submitted to their Web image retrieval system WebSeek. They concluded that users tend to search for images mainly at a higher semantic level. Jansen and Pooch [2000] analyzed the behavior of users looking for multimedia information on the Web. They found that people look more frequently for images than for audio or videos. More research is needed in the area of users and queries of Web image search engines, not only to get a better understanding of how and why people use these images, but also to design more effective retrieval systems. To achieve this goal, we believe there is a vital need to perform a detailed analysis of the submitted queries and to analyze the users' behavior when using these engines. Such studies could yield more detailed information about users. Here are some of the things it could allow us to do: categorize users according to criteria such as their profile or the precision of their queries, and identify the preferences of each category; understand the way people formulate their queries and what they expect of the system; identify the cases in which content-based retrieval is preferred and those in which text-based retrieval is more suitable; in the case of text-based queries, analyze the number of terms per query, and identify some characteristics of terms such as the most frequent ones and the correlated ones; in the case of content-based queries, analyze the number of images per query and the preference for region-based queries or global ones; get feedback about the degree of appreciation users assign to retrieved images, which would aid in evaluating the system's performance; and analyze the relationship between users preferences and their profiles (age, sex, occupation, place, education, citizenship, language, etc.), which are recovered with their assent, of course. All of this could be of great help in meeting users' needs.

5.2. Narrowing the Semantic Gap

In addition to low-level features such as color, people use high-level concepts to categorize and identify images. This has led to the emergence of two levels in image retrieval: low-level retrieval and high-level retrieval. Low-level retrieval comprises retrieval by primitive features such as color, texture, shape, and spatial location of image elements. High-level retrieval comprises retrieval of named objects and persons, and retrieval by abstract attributes such as emotions that can be associated with a given image. Current computer vision techniques allow for the automatic extraction of low-level features from images, with a good degree of efficiency. However, it is still difficult to extract objective high-level concepts either from images or from their surrounding data. The lack of coincidence between the information that can be extracted from the visual data and the interpretation that a user assigns to the same data in a given situation is known as *the semantic gap* [Smeulders et al. 2000]. As in general-purpose image retrieval, the semantic gap remains a challenging task in Web image retrieval. The key issue is how to derive high-level concepts automatically from the image content and its surrounding text and data. An initial avenue that has been explored by many researchers is to try to derive semantics from the text found in the same Web page as the image, first, because the text encodes high-level semantics better than the image content, and, second, because many words and expressions appearing in the same page as a given image may be related to it. However, the use of text involves some problems. First, there is generally a lot of text in a Web page and the system has to decide which words are related to images and which are not. Some techniques [Paek and Smith 1998; Rowe and Frew 1998] have been developed to address this issue, as discussed in Section 4.2.1. Second, even if there is a certainty that a given word is related to an image (for example, a description that the page creator added to the image), this word may be subjective because two different people may give two different descriptions and interpretations to the same image. In order to mitigate the problem of subjectivity, some image retrieval researchers have resorted to the automatic annotation of images, that is, developing tools that automatically extract keywords from images and then using these keywords for retrieval. Unfortunately, except for extracting the names of some objects in simple images, such tools remain primitive and are not yet sophisticated enough to extract semantic concepts from images. A similar solution that has been explored involves deriving high-level concepts without necessarily coupling these concepts with keywords [Barnard and Forsyth 2001; Naphade and Huang 2000; Vasconcelos and Lippman 1998], but this remains a research topic that requires further exploration. In short, more work needs to be done in order to be able to automatically extract objective semantics from Web pages, which would allow the retrieval of images on the basis of high-level concepts. Adopting metadata standards and using ontological structures to give meaning to textual metadata seems to be a promising solution in this regard.

5.3. The Need for Standardization in Image Description

In the Web context, there is a vital need to define standards for image and data descriptors. This will allow the description of images and other kinds of data to be standardized, making it possible for search engines to identify relevant attributes with more precision, effectiveness, and objectivity. Some standards on metadata definition (not necessarily limited to images) have appeared in recent years. They include the Resource Description Framework (RDF) for metadata, proposed by the World Wide Web Consortium W3C (http://www.w3.org/Metadata/). XML [Morrison et al. 1999; Dick 2000; Harold 2000] is the preferred language for writing RDF schemas (http://www.w3.org/XML/). It is an extensible markup language that is today taking the place of HTML [Graham 1997; Musciano and Kennedy 2000; Schwartz 1996] in creating Web pages. Another standard on metadata definition is the Dublin Core metadata set (http://dublincore.org/), in which metadata about a resource include its creator, title, subject keywords, type, and format. Standardization in the markup languages used to create Web pages should be addressed further and XML seems to be the promising language in this area. Finally, we mention the role of Mpeg-7 (http:// www.chiariglione.org/mpeg/) in standardizing the representation of multimedia information content for search, filtering, management, and processing.

5.4. Mitigating the Problem of High Dimension

One of the biggest challenges facing Web image retrieval is the curse of dimensionality. Indeed, as we discussed in Section 4, the number of images on the Web and their diversity may require the use of a multitude of features of high dimensions. However, this poses a serious problem for both retrieval and indexing. In the indexing stage, most existing techniques become unusable when the number of dimensions exceeds a certain threshold or when the number of images becomes very large, which is precisely the case with the Web. Retrieval suffers from the problem of dimensionality because high dimension causes overlapping between image classes in the search space, and traversing it thus becomes time-consuming and not particularly precise. Ways of avoiding this problem include dimension reduction and feature subspace selection. To deal with it when it cannot be avoided, special indexing and retrieval techniques must be developed specifically for collections of very large amounts of data with many features of high dimension. This is still an open issue that must be addressed by researchers to be able to perform an efficient image retrieval from the World Wide Web.

5.5. Offering Browsing Functionality

It is important for a Web image retrieval system to offer a catalog that categorizes images by subject and allows users to navigate between the different subjects. Such a service is very handy for users who do not have a clear idea in mind about the images they are looking for, or who want to retrieve more images belonging to the same category. Most Web text retrieval engines, such as Google, Yahoo!, Lycos, and AltaVista, propose such catalogs. However, in the field of retrieving images and visual media from the Web, few systems offer this service and the proposed catalogs cover only a limited part of the Web. We believe more work is needed in this area in order to offer users a structure that organizes the visual content of the Web in a browsing catalog.

5.6. Coverage of the Web

Partial coverage of the Web is a problem encountered by all kinds of Web information retrieval systems. Indeed, the Web contains a huge number of documents and the number is constantly increasing, which means the relative coverage of Web search engines is decreasing every day. Among other things, dealing with this problem will require a deep understanding of the structure of the Web, and the development of effective techniques for traversing this structure. This could allow the design of more intelligent Web crawlers capable of reaching the different parts of the Web but without visiting the same pages many times, thereby ensuring a better coverage of the Web and minimizing scouring time. Another solution for Web coverage which has been explored in the text retrieval field is metasearch. However, in the image field, the effectiveness of existing engines is still limited and metasearch tools cannot rely on them in their current state. This is a promising direction that should be explored further in the future.

5.7. Integration of Different Kinds of Media

There are different kinds of media on the World Wide Web, including text, images, videos, and audio. Unfortunately, most of the existing search engines support only one type of medium. Text retrieval engines are the most widely known. Many examples of image and video retrieval engines have been cited in previous sections. Examples of audio retrieval engines can be found in Foote [1999]. However, very little work has been done on the integration of different kinds of media in the same framework. This could allow for cross-media browsing as well as finding composite media related to the same subject. Because the different media types found in the same Web page tend to be related to the same subject, making effective use of all of the kinds of media belonging to the same page could aid in locating the needed information with more precision and effectiveness. Research in this direction includes WebMars [Ortega-Binderberger et al. 2000], which retrieves HTML documents based on text and images. These two kinds of media (text and images) are combined within the same query, using a tree in which each branch represents a media type. This approach could be extended to support more types of media. We can also cite the AMORE system [Mukherjea 1997, 1999] (http://www.ccrl.com/amore/), which focuses on multiple media types in a single retrieval framework. Other research includes the work of Lewis et al. [1997] and the Informedia project [Wactlar et al. 1996], whose overall aims are to allow full content search and retrieval of video by integrating speech and image processing.

5.8. Other Improvements

Many other improvements are possible in the context of image retrieval from the Web, including the identification and estimation of good and representative features from images and from the surrounding text and data. More studies are also needed in the field of similarity, in order to define measures that correspond more closely to human judgment. Retrieval refinement and relevance feedback is still a research issue in the field of image retrieval in general, aimed at gaining a better understanding of users and their needs. The results of this research need to be utilized and applied to the Web retrieval problem. Additional sophistication can also be achieved in the area of query specification. For example, some approaches have been investigated in the use of acoustical interfaces for text retrieval from the Web [Raman 1996; Mereu and Kazman 1996; Asakawa 1996]. This could help visually disabled or impaired people to use retrieval engines. For image retrieval from the Web, this is a future direction for further exploration.

6. CONCLUSION

In this article, we have presented the most important aspects of image retrieval from the World Wide Web. The main issues related to this question have been discussed, such as data collection and digestion, similarity measures, indexing, query specification, relevance feedback, and Web coverage. Developing tools for the automatic retrieval of images and visual data from the Web can be useful for many purposes and applications. Such tools could help people make effective use of the huge quantity of visual data available on the Web. Although some systems have appeared in recent years, more work needs to be done in this area and a number of questions remain to be addressed. First, more studies must be carried out on users and their queries in order to understand them in a more full manner and to meet their needs. Second, more effort is needed in the area of narrowing the semantic gap in order to allow people to retrieve images with high-level concepts. Third, there is a vital need for standardization in image description, which would allow engines to retrieve images with more precision. Fourth, efficient indexing and retrieval techniques should be developed to deal specifically with the great number of high-dimensional features that we have to handle in the Web context. Other important issues that remain to be addressed include offering catalogs for browsing, ensuring a better coverage of the Web, and integrating of different kinds of media.

REFERENCES

AMSALEG, L., GROS, P., AND MEZHOUD, R. 2000. Mise en base d'images indexées par des descripteurs locaux: problèmes et perspectives. Res. Rep. No. 1316-mars 2000. Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France.

ASAKAWA, C. 1996. Enabling the visually disabled to use the WWW in a GUI environment. Tech. Rep. of the IEICE. HCE, Tokyo, Japan, 39–44.

ASHBY, F. G. AND PERRIN, N. A. 1998. Toward a unified theory of similarity and recognition. *Psych. Rev. 95*, 1, 124–150.

ASLANDOGAN, Y. A. AND YU, C. T. 1999. Techniques and systems for image and video retrieval. *IEEE Trans. Knowl. Data Eng. 11*, 1, 56–63.

ASLANDOGAN, Y. A. AND YU, C. T. 2000. Diogenes: A Web search agent for content based indexing of personal images. In *Proceedings of the ACM SIGIR* (Athens, Greece).

ASLANDOGAN, Y. A. AND YU, C. T. 2000b. Evaluating strategies and systems for content based indexing of person images on the Web. In *Proceedings of the ACM Multimedia* (Los Angeles, CA).

ATHITSOS, V. AND SWAIN, M. J. 1997. Distinguishing photographs and graphics on the World Wide Web. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (San Juan, Puerto Rico), 10–17.

BARNARD, K., DUYGULU, P., AND FORSYTH, D. A. 2001. Clustering art. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Kauai, Hawaii), 434–441.

BARNARD, K. AND FORSYTH, D. 2001. Learning the semantics of words and pictures. In *Proceedings of the 8th International Conference on Computer Vision* (Vancouver, B.C, Canada), vol. 2, 408–415.

BEIGI, M., BENITEZ, A. B., AND CHANG, S.-F. 1998. Metaseek: A content-based meta-search engine for images. In *Proceedings of the Symposium on Electronic Imaging: Multimedia Processing and Applications-Storage and Retrieval for Image and Video Databases VI* (IST/SPIE'98, San Jose, CA).

BERCHTOLD, S., BÖHM, C., AND KRIEGEL, H.-P. 1998. The Pyramid-Tree: Breaking the curse of dimensionality. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (Seattle, WA), 142–153.

BERCHTOLD, S., KEIM, D. A., AND KRIEGEL, H.-P. 1996. The X-Tree: An index structure for high-dimensional data. In *Proceedings of the 22nd VLDB Conference* (Bombay, India), 28–39.

BIRD, C., ELLIOTT, P., AND GRIFFITHS, E. 1996. User interfaces for content-based image retrieval. In *Proceedings of the IEE Colloquium on Intelligent Image Databases* (London, U.K.).

CANDAN, K. S. AND LI, W.-S. 2001. On similarity measures for multimedia database applications. *Knowl. Inform. Syst. 3*, 1, 30–51.

CARSON, C., THOMAS, M., BELONGIE, S., HELLERSTEIN, J. M., AND MALIK, J. 1999. Blobworld: A system for region-based image indexing and retrieval. In *Proceedings of the 3rd International Conference on Visual Information Systems* (Amsterdam, The Netherlands), 509–516.

CHAN, Y., HARVEY, R., AND SMITH, D. 1999. Building systems to block pornography. In *Proceedings of the Challenge of Image Retrieval Workshop* (CIR99, Newcastle, U.K.).

CHANG, E. Y., WANG, J., LI, C., AND WEIDERHOLD, G. 1998. RIME: A replicated image detector for the World-Wide Web. In *Proceedings of the SPIE Symposium of Voice, Video, and Data Communications* (Boston, MA), 58–67.

CHEN, C.-C. AND WANG, J. Z. 2002. Large-scale Emperor Digital Library and semantics-sensitive region-based retrieval. In *Proceedings of the the International Conference on Digital Library—IT Opportunities and Challenges in the New Millennium* (Beijing, China), 454–462.

DICK, K. 2000. *XML a Manager's Guide*. Addison-Wesley, Reading, MA.

DI SCIASCIO, E., MINGOLLA, G., AND MONGIELLO, M. 1999. Content-based image retrieval over the web using query by sketch and relevance feedback. In *Proceedings of the 4th International Conference on Visual Information Systems*, 123–130.

DREILINGER, D. AND HOWE, A. E. 1997. Experiences with selecting search engines using metasearch. *ACM Trans. Inform. Syst. 15*, 3, 195–222.

DUDA, R. O. AND HART, P. E. 1973. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York, NY, 211–249.

EAKINS, J. P. AND GRAHAM, M. E. 1999. Content-based image retrieval: A report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle, Newcastle, U.K.

ENNIS, D. M. AND JOHNSON, N. L. 1993. Thurstone-Shepard similarity models as special cases of moment generating functions. *J. Math. Psych. 37*, 104–110.

ENSER, P. G. B. AND MCGREGOR, C. G. 1992. Analysis of visual information retrieval queries. British Library Research and Development Rep. No. 6104. British Library, London, U.K.

FAGIN, R. 1998. Fuzzy queries in multimedia database systems. In *Proceedings of the 17th ACM Symposium on Principles of Database Systems* (PODS'98, Seattle, WA).

FIDAL, R. 1997. The image retrieval task: Implications for the design and evaluation of image databases. *New Rev. Hypermedia Multimedia 3*, 181–199.

FLECK, M., FORSYTH, D. A., AND BREGLER, C. 1996. Finding naked people. In *Proceedings of the 4th European Conference on Computer Vision* (Cambridge, U.K.), vol. 2, 593–602.

FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., HUANG, Q., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., STEELE, D., AND YANKER, P. 1995. Query by image and video content: The QBIC system. *IEEE Comput. 28*, 9, 23–32.

FOOTE, J. T. 1999. An overview of audio information retrieval. *Multimedia Syst. 7*, 1, 2–11.

FORSYTH, D. A., MALIK, J., FLECK, M. M., LEUNG, T., BELONGIE, S., BREGLER, C., CARSON, C. AND GREENSPAN, H. 1996. Finding pictures of objects in large collections of images. In *Proceedings of the International Workshop on Object Recognition* (Cambridge, U.K.).

FRANKEL, C., SWAIN, M. J., AND ATHITSOS, V. 1996. WebSeer: An image search engine for the World Wide Web. Tech. Rep. No. 96-1. University of Chicago, Computer Science Department, Chicago, IL.

GERARDTS, Z. AND BIJHOLD, J. 2002. Content based information retrieval in forensic image databases. *J. Forensic Sci. 47*, 2, 40–47.

GEVERS, T., ALDERSHOFF, F., AND SMEULDERS, A. W. M. 2000. Classification of images on internet by visual and textual information. In *Proceedings of the Internet Imaging* (SPIE, San Jose, CA).

GEVERS, T. AND SMEULDERS, A. W. M. 1996. A comparative study of several color models for color image invariant retrieval. In *Proceedings of the 1st International Workshop on Image Databases and Multimedia Search* (Amsterdam, The Netherlands), 17–23.

GEVERS, T. AND SMEULDERS, A. 1997. PicToSeek: A content-based image search system for the World Wide Web. In *Proceedings of the VISUAL97 Conference* (San Diego, CA), 93–100.

GEVERS, T. AND SMEULDERS, A. W. M. 1999. The PicToSeek WWW image search system. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems* (Florence, Italy), 246–269.

GONG, Y. 1997. Intelligent image databases: Towards advanced image retrieval. Kluwer Academic Publishers, Dordrecht, The Netherlands.

GOODRUM, A. A. 2000. Image information retrieval: An overview of current research. *Inform. Sci. 3*, 2, 63–66.

GOODRUM, A. A. AND SPINK, A. 1999. Visual information seeking: A study of image queries on the World Wide Web. In *Proceedings of the Annual Meeting of the American Society for Information Science* (Washington, D.C.).

GRAHAM, I. S. 1997. *HTML Source Book.* John Wiley and Sons, New York, NY.

GUDIVADA, V. N. AND RAGHAVAN, V. V. 1995. Content-based image retrieval systems. *IEEE Comput. 28*, 9, 18–22.

GUDIVADA, V. N., RAGHAVAN, V. V., GROSKY, W. I., AND KASANAGOTTU, R. 1997. Information retrieval on the World Wide Web. *IEEE Internet Comput. 1*, 15, 58–68.

HARMAN, D. 1995. Overview of the Fourth Text Retrieval Conference (TREC-4). In *Proceedings of the Fourth Text Retrieval Conference* (TREC-4, Gaithersburg, MD).

HAROLD, E. R., French adaptation by PICARDE, G. 2000. *XML Le guide de l'utilisateur.* Osman Eyrolles Multimedia, Paris, France.

HASTINGS, S. K. 1995. Query categories in a study of intellectual access to digitized images. In *Proceedings of the 58th ASIS Annual Meeting*, vol. 32, 3–8.

HOWE, A. E. AND DREILINGER, D. 1997. SavvySearch: A meta-search engine that learns which search engines to query. *AI Mag. 18*, 2, 19–25.

HU, W. -C. 2003. World Wide Web search technologies. *Architectural Issues of Web-Enabled Electronic Business*, Shi Nansi, Ed. Idea Group Publishing, Hershey, PA.

HUANG, J., KUMAR, S., MITRA, M., ZHU, W.-J., AND ZABIH, R. 1997. Image indexing using color correlogram. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Puerto Rico), 762–768.

ISHIKAWA, Y., SUBRAMANYA, R., AND FALOUTSOS, C. 1998. Mindreader: Query databases through multiple examples. In *Proceedings of the 24th International Conference on Very Large Data Bases* (New York, NY), 433–438.

JAIN, A. K., GOEL, S., AGARWAL, S., MITTAL, V., SHARMA, H., AND MAHINDRU, R. 1997. Multimedia systems for art and culture: A case study of Brihadisvara Temple. In *Proceedings of the Storage and Retrieval for Image and Video Databases V. SPIE 3022*, 249–261.

JAIN, R. 1995. World-Wide Maze. *IEEE MultiMedia 2*, 2, 3.

JANSEN, B. J., GOODRUM, A., AND SPINK, A. 2000. Searching for multimedia: An analysis of audio, video, and image Web queries. *World Wide Web J. 3*, 4, 249–254.

JANSEN, B. J. AND POOCH, U. 2000. Web user studies: A Review and framework for future work. *J. Amer. Soc. Inform. Sci. Tech. 52*, 3, 235–246.

JING, F., LI, M., ZHANG, H.-J., AND ZHANG, B. 2002. Learning region weighting from relevance feedback in image retrieval. In *Proceedings of the 27th IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP, Orlando, FL).

JING, F., LI, M., ZHANG, H.-J., AND ZHANG, B. 2002b. Region-based relevance feedback in image retrieval. In *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS, Scottsdale, AZ), 365–369.

JORGENSEN, C. 1999. Considerations in prototyping an image retrieval testbed. In *Proceedings of the Multimedia Indexing and Retrieval Workshop Presented in Conjunction with the 22nd Annual International ACM SIGIR Conference* (Berkeley, CA).

KEISTER, L. H. 1994. User types and queries: Impact on image access systems. In *Challenges in Indexing Electronic Text and Images*, R. Field, T. B. Hahn, E. M. Rasmussen, and P. J. Smith, Eds. ASIS, 7–22.

KHERFI, M. L., ZIOU, D., AND BERNARDI, A. 2002. Learning from negative example in relevance feedback for content-based image retrieval. In *Proceedings of the IEEE International Conference on Pattern Recognition* (ICPR, Québec, Canada).

KHERFI, M. L., ZIOU, D., AND BERNARDI, A. 2003a. Atlas WISE: A Web-based image retrieval engine. In *Proceedings of the International Conference on Image and Signal Processing* (ICISP, Agadir, Morocco), 69–77.

KHERFI, M. L., ZIOU, D., AND BERNARDI, A. 2003b. Combining positive and negative examples in relevance feedback for content-based image retrieval. *J. Vis. Comm. Image Rep. 14*, 4, 428–457.

KOBAYASHI, M. AND TAKEDA, K. 2000. Information retrieval on the Web. *ACM Comput. Surv. 32*, 2, 144–173.

LA CASCIA, M., SETHI, S., AND SCLAROFF, S. 1998. Combining textual and visual cues for content-based image retrieval on the World Wide Web. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (Santa Barbara, CA), 24–28.

LAWRENCE, S. AND GILES, C. 1999. Text and image metasearch on the web. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications* (PDPTA99, Las Vegas, NV), 829–835.

LEW, M. S. 2000. Next generation Web searches for visual content. In *IEEE Comput. 33*, 11, 46–53.

LEW, M. S., LEMPINEN, K., AND HUIJSMANS, N. 1997. Webcrawling using sketches. In *Proceedings of the VISUAL97 Conference* (San Diego, CA), 77–84.

LEWIS, P. H., DAVIS, H. C., DOBIE, M. R., AND HALL, W. 1997. Towards multimedia thesaurus support for media-based navigation. *Image Databases and Multimedia Search.* World Scientific, Amsterdam, The Netherlands, vol. 8, 111–118.

LU G. AND WILLIAMS, B. 1999. An integrated WWW image retrieval system. In *Proceedings of the AusWeb99. 5th Australian World Wide Web Conference* (Lismore, Australia).

MARSHALL, S. 1989. Review of shape coding techniques. *Image Vis. Comput. 7*, 4, 281–294.

MARTINEZ, A. 1999. Face image retrieval using HMMs. In *Proceedings of the IEEE Workshop Content-Based Access of Image and Video Libraries* (Fort Collins, CO), 25–39.

MEHTRE, B. M., KANKHANHALLI, M. S., AND FOON LEE, W. 1997. Shape measures for content based image retrieval: A comparison. *Inform. Process. Manage. 33*, 3, 319–337.

MEREU, S. AND KAZMAN, R. 1996. Audio enhanced 3D interfaces for visually impaired users. In *Proceedings of the ACM SIGCHI'96* (CHI'96, Vancouver, B.C., Canada).

MORRISON, M., BROWNELL, D., AND BOUMPHREY, F. 1999. *XML Unleashed.* Sams, Indianapolis, IN.

MUKHERJEA, S., HIRATA, K., AND HARA, Y. 1997. Towards a multimedia World Wide Web: Information retrieval engine. In *Proceedings of the Sixth International WWW Conference* (Apr. 7–11, Santa Clara, CA).

MUKHERJEA, S., HIRATA, K., AND HARA, Y. 1999. AMORE: A World Wide Web image retrieval engine. *WWW J. 2*, 3, 115–132.

MUSCIANO, C. AND KENNEDY, B. 2000. *HTML & XHTML, the Defining Guide.* O'Reilly, Sebastopol, CA.

NAPHADE, M. R. AND HUANG, T. S. 2000. Semantic video indexing using a probabilistic framework. In *Proceedings of the International Conference on Pattern Recognition* (Barcelona, Spain), vol. 3, 83–88.

O'CONNOR, B. C., O'CONNOR, K., AND ABBAS, J. M. 1999. User reactions as access mechanism: An exploration based on captions for images. *J. Amer. Soc. Inform. Sci. 50*, 8, 681–688.

OGLE, V. E. AND STONEBRAKER, M. 1995. Chabot: retrieval from a relational database of images. *IEEE Comput. 28*, 9, 40–48.

ORTEGA-BINDERBERGER, M., MEHROTRA, S., CHAKRABARTI, K., AND PORKAEW, K. 2000. WebMARS: A multimedia search engine. In *Proceedings of the SPIE Electronic Imaging 2000: Internet Imaging* (San Jose, CA).

PAEK, S. AND SMITH, J. R. 1998. Detecting image purpose in World-Wide Web documents. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging: Science and Technology—Document Recognition* (San Jose, CA).

PAO, M. L. AND LEE, M. 1989. *Concepts of Information Retrieval*. Libraries Unlimited, Inc. Englewood, CO.

RAMAN, T. 1996. Emacspeak a speech interface. In *Proceedings of the ACM SIGCHI* (CHI'96, Vancouver, B.C., Canada).

RANDEN, T. AND HUSOY, J. H. 1999. Filtering for texture classification: A comparative study. *IEEE Trans. Patt. Analys. Mach. Intell. 21*, 4, 291–310.

RATHA, N. K., KARU, K., CHEN, S., AND JAIN, A. K. 1996. A real-time matching system for large fingerprint databases. *IEEE Trans. Patt. Analys. Mach. Intell. 18*, 8, 799–813.

ROWE, N. C. AND FREW, B. 1998. Automatic caption localization for photographs on World Wide Web pages. *Inform. Process. Manage. 34*, 1, 95–107.

RUI, Y. AND HUANG, T. S. 2000. Optimizing learning in image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (Hilton Head, SC).

RUI, Y., HUANG, T. S., AND CHANG, S.-F. 1999. Image retrieval: Current techniques, promising directions, and open issues. *J. Vis. Comm. Image Rep. 10*, 1, 39–62.

RUI, Y., HUANG, T. S., AND MEHROTRA, S. 1997. Content-based image retrieval with relevance feedback in MARS. In *Proceedings of the IEEE International Conference on Image Processing* (Santa Barbara, CA), 815–818.

SANTANI, S. AND JAIN, R. 1996. Similarity queries in image databases. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (San Francisco, CA), 646–651.

SCHETTINI, R., CIOCCA, G., AND ZUFFI, S. 2001. A survey on methods for colour image indexing and retrieval in image databases. In *Color Imaging Science: Exploiting Digital Media*, R. Luo and L. MacDonald, Eds., J. Wiley, New York, NY.

SCHWARTZ, J. 1996. *Le Grand Livre HTML.* Micro Application, Paris, France.

SCLAROFF, S. 1995. World Wide Web image search engines. *In Proceedings of the NSF Workshop on Visual Information Management* (Cambridge, MA).

SCLAROFF, S., TAYCHER, L., AND LA CASCIA, M. 1997. Image rover: A content-based image browser for the World Wide Web. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (San Juan, Puerto Rico), 2–9.

SELBERG, E. AND ETZIONI, O. 1997. The MetaCrawler architecture for resource aggregation on the Web. *IEEE Exp. 12*, 1, 8–14.

SMEULDERS, A. W. M., WORRING, M., SANTINI, S., GUPTA, A., AND JAIN, R. 2000. Content-based image retrieval at the end of the early years. *IEEE Trans. Patt. Analys. Mach. Intell. 22*, 12, 1349–1380.

SMITH, J. R. AND CHANG, S.-F. 1996. VisualSEEK: A fully automated content-based image query system. In *Proceedings of the 4th ACM International Conference on Multimedia* (Boston, MA), 87–98.

SMITH, J. R. AND CHANG, S.-F. 1996a. Searching for images and videos on the World-Wide Web. *Tech. Rep. No. 459-96-25*. Center for Telecommunication Research, Columbia University, New York, NY.

SMITH, J. R. AND CHANG, S.-F. 1997. An image and video search engine for the World-Wide Web. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases V* (IS&T/SPIE, San Jose, CA), 84–95.

SMITH, J. R. AND LI, C.-S. 1998. Image retrieval evaluation. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (CBAIVL-98, Santa Barbara, CA).

SWAIN, M. J., FRANKEL, C., AND ATHITSOS, V. 1997. WebSeer: An image search engine for the World Wide Web. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (San Juan, Puerto Rico).

TAYCHER, L., LA CASCIA, M., AND SCLAROFF, S. 1997. Image digestion and relevance feedback in the Image Rover WWW search engine. In *Proceedings of the 2nd International Conference on Visual Information* (San Diego, CA), 85–94.

THURSTONE, L. L. 1927. A law of comparative judgement. *Psych. Rev. 34*, 273–286.

TUCERYAN, M. AND JAIN, A. K. 1993. Texture analysis. In *Handbook of Pattern Recognition and Computer Vision*, C. Chen, L. Pau, and P. Wang, Eds. World Scientific, River Edge, NJ, 235–276.

TURNER, J. M. 1995. Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images. In *Proceedings of the Annual Meeting of the American Society of Information Science* (ASIS, Chicago, IL).

TVERSKY, A. 1977. Features of similarity. *Psych. Rev. 84*, 4, 327–352.

VAN DER ZWAN, R., KUKULSKA-HULME, A., DIPAOLO, T., EVERS, V., AND CLARKE, S. 1999. The Informedia Digital Video Library System at the Open University. In *Proceedings of the CIR-99: The Challenge of Image Retrieval* (Newcastle upon Tyne, U.K.).

VASCONCELOS, N. AND LIPPMAN, A. 1998. A Bayesian framework for semantic content characterization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR, Santa Barbara, CA), 566–571.

VASCONCELOS, N. AND LIPPMAN, A. 1999. Learning from user feedback in image retrieval systems. *Neural Information Processing Systems 12* (Denver, CO).

VELTKAMP, R. C. AND TANASE, M. 2000. Content-based image retrieval systems: A survey. Tech. Rep. No. UU-CS-2000-34. Department of Computer Science, Utrecht University, Utrecht, The Netherlands.

WACTLAR, H. D., KANADE, T., SMITH, M. A., AND STEVENS, S. M. 1996. Intelligent access to digital video: The Informedia Project. *IEEE Comput. 29*, 5, 46–52.

WANG, J. Z., LI, J., WIEDERHOLD, G., AND FIRSCHEIN, O. 1998. Classifying objectionable websites based on image content. *Interactive Distributed Multimedia Systems and Telecommunication Services*, T. Plagemann and V. Goebel, Eds. Lecture Notes in Computer Science, vol. 1483. Springer-Verlag, Berlin, Germany, 113–124.

WEBER, R., SCHEK, H.-J., AND BOTT, S. 1998. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In *Proceedings of the 24th International Conference on Very Large Data Bases* (New York, NY), 194–205.

WEI, G., LI, D., AND SETHI, I. K. 1998. Web-WISE: Compressed image retrieval over the Web. In *Proceedings of the Multimedia Information Analysis and Retrieval, IAPR International Workshop*, 33–46.

YANG, D. S., GARRETT, J. H., SHAW, D. S., AND RENDELL, L. A. 1994. An intelligent symbol usage assistant for CAD systems. *IEEE Exp. 10*, 3, 32–41.

ZHANG, H.-J., CHEN, Z., LIU, W. Y., AND LI, M. 2001. Relevance feedback in content-based image search. In *Proceedings of the 12th International Conference on New Information Technology* (NIT, Beijing, China).

ZHANG, H.-J. AND ZHONG, D. 1995. A scheme for visual feature based image retrieval. In *Proceedings of the SPIE Storage and Retrieval for Image and Video Databases* (San Jose, CA), 36–46.

ZHAO, T. TANG, L. H. Y., HO-SHING, I. H., AND QI, F. 2001. Visual keyword image retrieval based on synergetic neural network for Web-based image search. *Real-Time Syst. 1-2*, 127–142.

Received February 2003; revised January 2004; accepted April 2004