Special issue: Leading ICT technologies in the Information Explosion

Survey Paper

Image retrieval: Research and use in the information explosion

Masashi INOUE
National Institute of Informatics

ABSTRACT
We surveyed the current research on the retrieval and utilization of images. This paper provides an overview of the difficulties and possibilities of technological or technology-related research topics resulting from the huge amount of images currently available and highlights some of the important research topics. We looked at the ongoing research activities and analyzed them from four aspects, information access and organization technology, the computing infrastructure that enables access to large-scale image resources, issues in human-system interaction and human factors related to using images, and the social aspect of image media. On the technical side, we noticed that as the number of digital images increases, so does the importance of the accuracy and scalability in relation to the image retrieval. The accuracy and scalability are in fact needed to cope with the current explosion in digital images. On the social side, not so long ago, image retrieval technologies were only experimental tools or used by experts within a limited domain. However, now the general public has access to a wide range of digital images, which means that image retrieval technologies are being used by various users in a large diversity of social contexts. Thus, as we show in this paper, the accuracy and scalability are not the only important factors in the era of information explosion, but that researchers must also be concerned with the social aspects of these technologies.

KEYWORDS
Image retrieval, utility, usage, user study, information retrieval, information access

1 Introduction

The emergence of digital cameras and their integration into mobile phones has made digital images more accessible and significantly changed our view of image media. However, people still have difficulty when searching for images, which are typically 'sub-symbolic' in their level of information. The explosion in the number of images available on-line began even before image access technologies had matured.

Digitization was first applied to text media, and for many years now, ordinary people have been generating and storing large amounts of text electronically. Electronic text can easily be transmitted via email and disseminated on the web. In addition, we can easily search for stored text to acquire the relevant information on any topic and even reuse it. Naturally, it is desirable for a similar phenomenon to happen with other types of media.

The increase in the size of stored images in personal storage devices and in cyberspace has ushered in a new era of image retrieval and usage. This large volume may serve either to worsen the situation because of information overload or as a gospel to users because it offers wider choices. In this paper, we review the current research activities surrounding image access from the following aspects, information retrieval and organization technology, the infrastructure that enables large-scale data processing, issues in human-system interaction, and the social issues. These four aspects are all important when it comes to developing systems that can deal with the massive volume of images.

First, we need more advanced image retrieval methodologies to connect user queries with the most relevant images. The information explosion means that there are too many relevant images readily available to meet our cognitive capacity, and that we need to find the most highly relevant images. Researchers are show-

Received September 1, 2008; Revised December 25, 2008; Accepted December 26, 2008.
m-inoue@nii.ac.jp, http://www.nii.ac.jp/˜m-inoue/
DOI: 10.2201/NiiPi.2009.6.2

ing increasing interest in this issue, and there are now several specialized conferences and forums concerned with this issue [1].

Second, such advanced methodologies should be made possible with powerful computer systems. Since images are larger in size than text and much harder to manipulate, an infrastructure should be constructed to manage them. The scalability, reliability, and speed of image retrieval systems are the challenges.

Third, the ways that people use images are also changing. Image media is no longer only for a limited number of experts; it is used by almost everyone. Traditional image retrieval systems did not prepare for this change. We need to study new information behaviors and integrate the findings with the accumulated knowledge from past user studies in designing systems.

Fourth, the continuing popularization of image media will expose various social issues, including legal conflicts and cultural differences. We have to conduct research to clear any potential obstacles to assist in the advancement of image utilities.

Until recently, information retrieval and organization technology and the infrastructure that enables large-scale data processing were topics mainly limited to the computer science field, whereas human-system interaction and social issues were dealt with in the library science field. However, to bring users a more satisfactory experience in the ever-enlarging image environment, the two fields will have to merge their activities especially when it comes to image retrieval research (Fig. 1). We will discuss the research issues in these fields in the following sections.

2 Retrieval Methods

2.1 Automation of Retrieval

In the past, retrieval was done manually by experts. Experts stored and indexed documents and they received requests from information seekers, forcing them to search for the relevant documents by using their own knowledge. Now, due to the information boom we have been experiencing recently, the automation of this process became necessary. The basic idea for automatic retrieval is that documents are automatically indexed according to pre-defined features extracted from documents. Then, given a query that reflects a user's information need, the system ranks the collection according to the estimated relevance to the user's request. The main problem affecting the early attempts in designing image retrieval systems was a lack of rigorous evaluation. The evaluation issue is discussed in Sec. 2.3. Although manual retrieval is not the most effective way of dealing with a huge amount of data, it is still more accurate than today's automated retrieval systems in some domains. In fact, the manual retrieval process can provide us with insights on the users and their search needs. This issue is discussed in Sec. 4.

2.2 Content- and Annotation-based Image Retrieval

There are currently roughly two types of automatic image retrieval technologies, content-based image retrieval (CBIR) and annotation-based image retrieval (ABIR).

In CBIR, the image's signals are analyzed for their visual content, such as the image's colors, textures, and shapes. These colors, textures, and shapes are called low-level *features*. They can be extracted from the image, and the objects present in an image can be identified with them. These objects are called *content*. The images are sought by measuring the similarity in terms of these features between a query image and a set of candidates. Various feature extraction methods and similarity computing methods have been proposed [2]. Although designing a good feature extraction method is difficult and the computational cost is sometimes high, once an extraction algorithm has been devised, it can exploit those low-level features.

ABIR, on the other hand, generally operates on the annotations (textual descriptions) associated with images [3]. The annotations can be carefully chosen keywords, image captions, or entire documents in which the images are embedded. Annotations are intuitive for humans because they can be read. However, we face a chore in preparing them.

CBIR has traditionally been studied by computer science researchers, whereas ABIR has been dealt with by library science researchers (Fig. 1). Furthermore, the two methodologies are somewhat specialized in terms of the aspects of the images they treat. CBIR deals with things that can be seen such as the appearances of objects, while ABIR manages the information associated with the images, such as the names or conceptual impressions. Figure 2 schematically summarizes these differences.

Another criterion for classifying automated image retrieval is the query type. If users represent their search needs by using a sample image, the retrieval is called a query-by-example. If they use keywords, it is called a query-by-text. Although these two types of querying can be used in both CBIR and ABIR, for simplicity, we assume that CBIR is usually associated with query-by-example and ABIR with query-by-text.

Automatic image retrieval research started with ABIR as a simple querying of databases that store images as textual records. The motivation to use image content for retrievals emerged with the growth in computational capabilities. The idea of CBIR attracted the attention of computer science researchers because its performance depends on the quality of the feature ex-

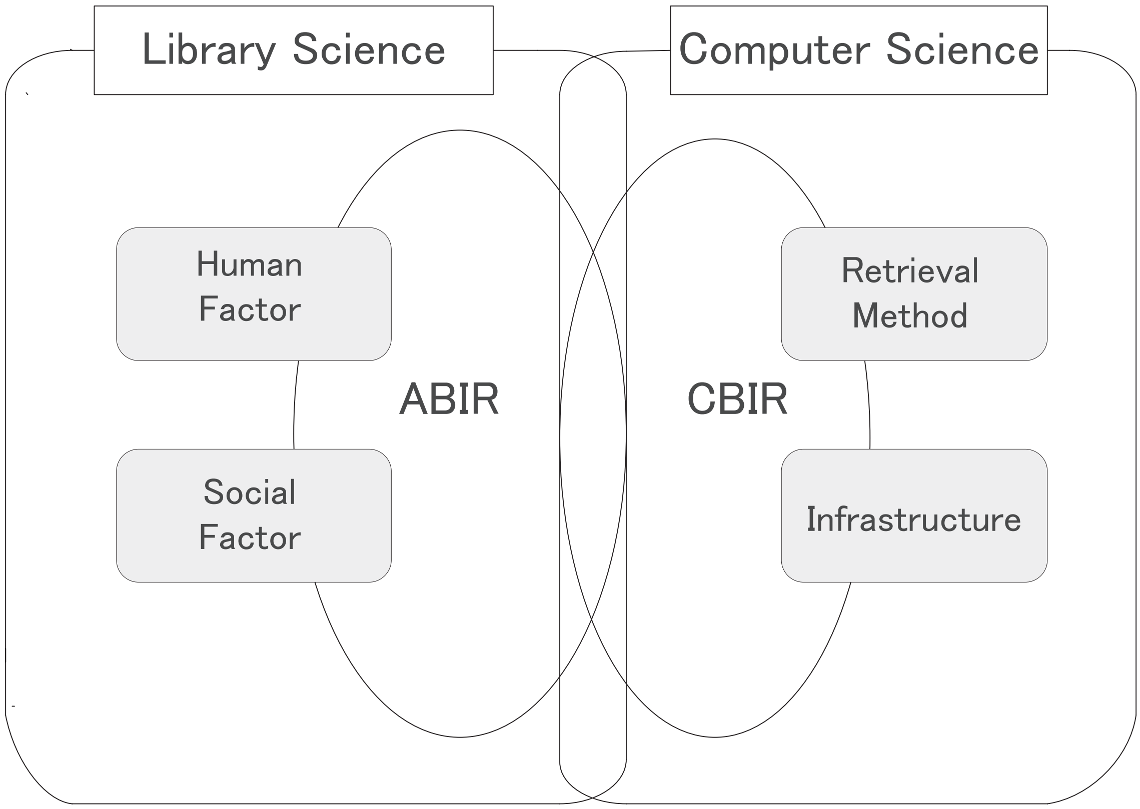

Fig. 1 Research fields and aspects involving image retrieval and usage.

tractors and computer vision algorithms that many researchers are interested in [4]. Also, the automatic labeling of images and objects in images can be used to change a CBIR problem into an ABIR problem by assigning text labels and thus enabling semantic image access [5].

As will be discussed in Sec. 4, studies by library scientists indicate that users often search for images by using abstract semantics rather than by directly inputting low-level features, objects in images, or impressions of images [6]. ABIR is considered the standard means of retrieval because it can handle high-level concepts describing the image content [7]. In particular, ABIR has progressed by incorporating advanced natural language processing (NLP) techniques and external language resources such as thesauruses (Sec. 2.4).

The technologies involved in CBIR, on the other hand, have progressed to the extent that they can offer more concept-level semantic retrievals than before. The traditional method to evaluate CBIR systems treats them as classification systems rather than information retrieval (IR) systems and some progress has been made over the years [8]. Regarding feature extraction, the major contributing factor has been the discovery of robust visual descriptors such as SIFT local descriptors in signal processing research [9]. Such robust features are invariant to the change in scale and the illumination conditions and are effectively used to identify the objects in images. CBIR is used in some specialized domains for identity matching, e.g., for detecting illegally used images and for identifying criminals from fingerprints and iris images.

2.3 Evaluations and Test Collections

The evaluations in the early days of image retrieval research were often subjective. System designers collected their own image collections and decided on queries for themselves. They examined lists of top-ranked output images from the system and judged whether they seemed reasonable. They used subjective procedures because they knew that the construction of the standardized test collections that were similar to those used in text IR would require a tremendous amount of effort. To construct a standardized test collection, first of all, we need public image collections available, but the images themselves are not the only re-

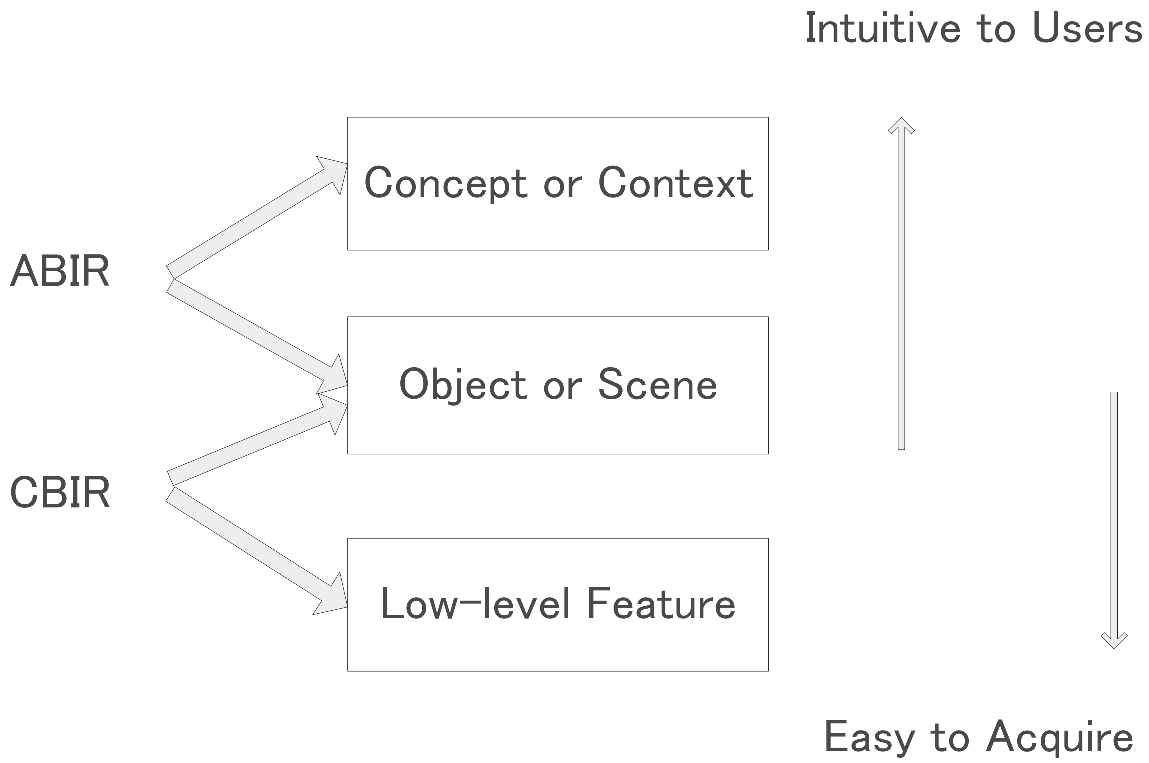

Fig. 2 ABIR and CBIR have different retrieval scopes. They are appearance of images (low-level feature), symbolic representation of visual features (object or scene), their evaluation and interpretations (concept or context). Users search for relevant images in all three of these aspects.

sources needed. To make a more objective evaluation, we need some representative search topics that contain descriptions of simulated user needs or samples of user search logs. We also need relevance assessments for each query; i.e., we need to determine which images in a collection are considered relevant given a query. Although several good test collections now exist in the text retrieval domain, only a few well-designed image test collections actually exist.

In contrast to the tedious work of constructing new IR test collections, researchers in the early days had access to image classification collections that contained ground truth class labels for images. With them, they made some attempts to cast the problem of retrieval as one of classification. That is, users were not assumed to be looking for images with a particular content, but rather they were searching for images in a category, such as animals, buildings, people, etc. The problem setting extended into the classification of many classes and automatic labeling, because, in reality, images are often annotated with several labels. The most frequently used collection was the Corel dataset, a set of images selected from proprietary stock-photo CDs. The set was also used in automatic labeling evaluation. However, as argued in [10], its usage is not standardized and comparisons of reported results are impossible with it. In addition, the dataset's CDs are no longer commercially available. An early attempt to create a publicly available image retrieval test collection was the Benchathlon[1]. It was intended for evaluating CBIR systems [11]. The current initiative is ImageCLEF[2], which is part of the cross-language information retrieval evaluation campaign called CLEF[3]. ImageCLEF consists of several tracks with different tasks. One track is an ad hoc photo retrieval that focuses on ABIR [12]. This track provides a common ground for evaluating ABIRs. Although the objectivity of evaluations has dramatically improved thanks to the preparation of publicly shared test collections, researchers have not reached an agreement on the ideal properties of the test collection. One issue is the characteristics of the images in the collection. For example, ImageCLEF organizers have changed the target collection from mostly

[1] http://www.benchathlon.net/
[2] http://imageclef.org/
[3] http://www.clef-campaign.org

monochrome historical images to colorful tourist photo snapshots, which they believe would be more generic.

Even though standardized test collections are carefully designed, they cannot cover all the domains and usage of images. Automated construction of test collections using pseudo queries generated from the target items by changing them for CBIR [13] or by sampling them for ABIR [14] can be regarded as attempts to widen the variety of low-cost test collections. The queries and annotations are still matters of discussion in terms of their closeness to reality and usefulness for system comparisons. In fact, the languages used in search topics and annotations of ImageCLEF have changed for the same target collection. In addition, the relationship between evaluation criteria and user satisfaction is an issue. A new evaluation criterion to measure the diversity of the results was introduced in 2008 [15]. It assumes that there should be more topics in the presented results [16]. This shows that image retrieval evaluation methods are still an active research topic. In particular, we have to discuss whether there is any difference in the utility criterion between text retrieval users and image retrieval users. Most of the current evaluation frameworks have been borrowed from the text retrieval field, and which aspects should be included or excluded when evaluating image retrieval systems remains an open question.

2.4 Textual Descriptors

The major difficulties in image retrieval lie in the lack of available semantic information sources. When text is used as the key for retrieval, the subjectivity of the annotations becomes a problem. Although many high quality image collections, such as those found in museums, have text annotations supplied by professionals, different annotators may assign different tags to the same items. To alleviate the cost of this mismatch, library scientists and domain experts have developed taxonomies, which are sometimes called thesauruses, to give stable descriptions of the images. For example, the Getty Research Institute has created a taxonomy for the art domain, called the Categories for the Description of Works of Art (CDWA)[4] [17].

The labeling of images is an essential pre-processing component for ABIR when there are no or few keywords assigned to the target images. However, since manual annotation is a tedious and difficult task, there have been many attempts to automate it. Some researchers claim that retrievals involving automatically generated annotations are sometimes as accurate as those involving manual annotations [18].

Just as in retrieval, there are a number of test collections for automatic labeling. These collections are used to evaluate the systems that label images by analyzing their content. Although existing taxonomies discussed above are well-designed, few of them are used in automatic labeling for the computer science field. One reason is that the scopes of existing taxonomies are sometimes different from the target domain. As will be discussed later in Sec. 4, typical users of image retrieval were once art scholars or historians who had their own specific views on a subject. The difference between the selected materials found in libraries and museums and more random entities such as user-generated content or systematic entities such as scientific images that are often the target in the computer science field may be the cause of the limited popularity of existing resources. In contrast, certain linguistic taxonomies created for generic purposes are sometimes used in developing automatic retrieval systems. For example, the WordNet ontology that addresses the generic relationships among English words is used to deal with the problem of subjective annotation [19].

Besides the manual annotation done by professionals and automatic annotation provided by machines, another solution has recently emerged. The paradigm is called folksonomy or collaborative tagging. Once someone uploads a photo on-line, other users who have viewed the data can add annotations as keyword tags. Although these tags are initially noisy and unreliable, they can be improved by subsequent people who add their own tags and this constitutes a rich semantic context of publicly shared images. In addition to the tags, comments about the image can also be used as if they are textual annotations.

2.5 Visual Descriptors

To represent images, various descriptors based on the color, texture, and shape of the objects in the images have been developed and tested. The difference between textual and visual descriptors is evident when documents are being indexed and compared with queries. For text, automatic indexing is relatively straightforward if the text can be decomposed into words or characters depending on the chosen document representation. However, such a pre-defined symbolic representation does not exist for images, and system designers have to choose representations balancing on their effectiveness and efficiency. Although improvement of the visual descriptors used to find identical objects or similar images would help to improve the quality of automatic image retrieval, a survey of the visual descriptors is beyond the scope of this paper. Readers who are interested in this topic may refer to survey papers (e.g., [20]).

[4] http://www.getty.edu/research/conducting_research/standards/cdwa/

2.6 Combining Visual and Textual Information

As discussed in Sec. 2.2, ABIR is a reliable way to meet the needs of users. However, textual information is scarce or completely missing from many images. Thus, there have been many attempts to supplement the shortage of annotations by combining visual and textual information. One approach is enriching the document descriptors by concatenating both the visual and textual descriptors and performing feature matching in a reduced feature space (e.g., [21]). Alternatively, it is also possible to take a *find-similar* approach that gathers all visually related images on the basis of the CBIR concept after an initial ABIR retrieval. That is, the visual proximity of the images can be used as post-processing information to manipulate the ranking [22]. Enrichment using external resources is also possible. Yanai used a web search engine to identify the *visualness* of several keywords by retrieving images from the web [23]. This method assumes that the web is a socially constructed medium, and thus, it should reflect the average human perception of images.

2.7 User Feedback

Since the initial retrieval results for image retrieval are usually unsatisfactory, a technique that incorporates a feedback function on the relevance of the displayed images assessed during the interaction process would be helpful. The inclusion of user feedback for improving the retrieval results is called relevance feedback (RF) (it was first used in text retrieval), and it has been studied as it relates to image retrieval for many years [24]. The difficulty with the RF method is that users usually do not want to spend time giving their feedback on the systems. There are some techniques that use active learning to cope with the small number of user feedback behaviors. They suggest the initial images to users as possible candidates that are the most useful for estimating the user's intention if these images are used in the feedback. Semi-supervised learning frameworks that work with only a small number of labeled data, which are the users' relevance assessments, have also been tested in interactive image retrieval. In addition, RF operates within a combined visual and textual features space, such as in [25]. Moreover, implicit relevance feedback given by eye movement or mouse clicks has also been tested. Although the implicit feedback is not necessarily as positive as explicit one, it likely has some correlation with relevant data.

3 Devices, Systems, and Infrastructure

3.1 Image Acquisition Devices

The most important change that brought about the explosion of readily available images was the advent of the digital camera. Digital cameras have replaced the analog cameras of a decade ago and have reduced the cost of taking photos. Approximately 100.37 million digital cameras were shipped out in 2007; in contrast, only approximately 0.79 million film cameras were shipped [26]. Statistics indicate that the average number of photos taken by a household in a year is from 400 to 700 depending on the family structure [27]. As a result of these changes, there are now too many photos to skim through, and it is believed that many of these photos will never be viewed again afterwards.

Another big change is that contextual information can now be used to organize images. Besides tags, various types of contextual information can be helpful for certain image retrieval applications. For example, EXIF[5)] is a standard metadata format for images taken by digital cameras that indicate the camera conditions when a photo was taken. Moreover, sensors attached to cameras can now generate various associated information. For instance, a GPS functionality can give the place the photo was taken, and this location information can be stored as image metadata. The circumstances affecting imaging devices are also dramatically changing. Many commercial products are being developed to integrate image acquisition devices with networking functionalities. In particular, mobile phones with cameras are powerful tools for supplying contextual information. There has even been a study on uisng sensors that log encountered users and their contacts to supplement location information provided by GPSs [28].

3.2 Evolution of Software and Communications Platform

These days, taking photos is a casual form of communication. In the past, people took photos only on important occasions and they had to meet face-to-face to share their experiences. People can now upload images on-line or send photos on mobile phones. The evolution of software infrastructures certainly has motivated users to generate more images. Similar changes may happen with the activities of drawing and painting. People have more chances to get feedback on-line if they upload their work to a relevant site, and the feedback may motivate them to create more.

3.3 Efficiency

The difference between images and text is in their sparseness when represented as feature vectors. Textual documents consist of words or characters. The vocabulary is large, but a document will usually contain only a small portion of it. Therefore, it is possible to make an index for text using only a small amount of memory, and we can quickly search for it. On the other

5) http://www.cipa.jp/exifprint/index_e.html

hand, images are not symbolic in nature. They are represented as high-dimensional vectors with continuous feature values. Even after quantization, images are often represented as dense vectors. Researchers in the database field are working on these challenging issues in order to construct a more efficient form of indexing for dense feature vectors.

Efficiency is an important topic in the field of networking. For example, the implementation of a peer-to-peer (P2P) based system for image retrieval is different from other key-value based implementations because it involves a comparison of the dense feature vectors rather than key-matching. An efficient indexing and data representation has been developed for conducting CBIR over a P2P network [29].

A number of factors should be taken into account when collecting images: ease of access, usage, and cost to keep them. In particular, the amount of storage required to maintain images is a very important consideration. Some applications such as casual image collections for mobile devices can use high compression formats and thus are less costly to keep. Others, like archiving, ideally require a file format to have loss-less characteristics. For example, the Dutch Royal Library estimated that they need 650 TB of storage to store the next four years of data if they use the current file formats [30].

3.4 High-performance Computers and Networks

Improvements to hardware and associated system software technology have tremendously increased the availability of digital content. These advances have influenced how digital images are managed. At the same time, explosion of images require more advanced systems. Single image files are usually larger than single text files. Therefore, if there are an equivalent number of images as there are texts in a collection, image access systems have to deal with much larger set of data than text retrieval systems. However, there have been only a few studies that directly address the issue of image access.

An issue that makes image access less attractive research target is the proportion of images among many different media. As we can see on the web, the number of images is usually smaller than the number of available text documents. Although there are important exceptions to this trend: visual information sources are crucial and plentiful sources in the field of computer assisted design (CAD)s [31], people create text documents more often than they do visuals. Another issue that makes image access less attractive as a target of research for developing efficient and high-performance computing systems is the rise of video data. Video is a mixture of visual and audible information, and some text may be associated (e.g., closed captions and subtitles) with it. Images can be considered as the subset of videos. Further, single video files are often far larger than single image files and challenging. Therefore, it is sometimes more reasonable to use videos rather than images to assess the scalability of information systems In addition, the recording costs of video are as low as those for photos in terms of the amount of human operation. Therefore, there will be as many video files as image files. However, images are a good target for high-performing computing system research when videos are too large to mange with the available systems even at the experimental level. One ongoing initiative, Content-based Photo Image Retrieval (CoPhIR)[6], uses high-performance computers to do research on images. Its researchers have started to construct a publicly available image retrieval test collection [32]. They used the European Enabling Grids for EsciencE (EGEE) computer GRID[7] consisting of 73 computers in several European countries to retrieve more than 60 million images from the Flickr photo-sharing service[8]. The goal is to test the capability of high-performance computing in a previously computationally impossible task. Most of the computing power was spent on extracting seven MPEG-7 visual features from the images. This task is also challenging from a system management perspective, because the hardware and software within the grid are heterogeneous and are not accessible all the time.

4 Human-factors of Image Access

4.1 Image Media

It is important to learn how we recognize the signals from images and what messages these signals can uniquely convey. Jörgensen summarized findings on the human perception of images [33] . Some psychological research focuses on the subjectivity of perceptions, such as color and photo quality [34]. In addition to learning how humans perceive images, we need to understand how images act as social media. Images sometimes have advantages over text [35]. For example, irony can be conveyed by carefully shooting photos as they let viewers interpret the images; such images may have a more striking effect that text [36]. In other words, images can serve as a handy tool to generate high-quality metaphors. In addition, text can send information as intended by the sender, whereas images are sent as a whole with redundancy. The receivers of a photo may find something interesting but unintended in the background of a photo, for example. These properties can be either beneficial or harmful, and users need to know how to effectively use image media. Images

[6] http://cophir.isti.cnr.it
[7] http://www.eu-egee.org/
[8] http://flickr.com

also have advantages over videos. If a video shot and a photo contain the same information, photos are a more efficient medium because it takes longer to go through a video shot. For example, instructional videos are easy to create, but it is often time-consuming for users to watch the entire video. If the same instruction is given by photos for key points, sometimes it is quicker to understand the message. This is because images can be regarded as carefully edited and summarized key-frames from videos.

There have been attempts to use image media in intelligent information systems, whereby images are retrieved for purposes other than the conventional ones. For example, retrieved images can be used to enrich encyclopedia entries by providing word sense disambiguation clues [37]. Images are also helpful in question-answering systems [38] [39]. The way of sharing experiences in remote locations can be enhanced by using a real-time collaborative photo annotation system [40].

4.2 Usability

To improve the user-experience, we have to improve the system's usability and utility. The term usability concerns the system's behavior, most notably the user interfaces (UIs). Usability can be examined from a psychological perspective. Utilities, as we shall see in the next section, concern the behavior of users, and it can be examined from an ethnographic perspective.

The basic output from information retrieval systems is a list of documents ranked according to their estimated relevance. In addition to the quality of the list itself, the presentation of the list affects the user satisfaction. Images are currently often listed using thumbnails and in certain situations, such presentation is not the most usable for many users. One way to assist users in finding an image is the addition of an advanced browsing feature. A user study has indicated that browsing is still a major means of image access for casual users [41]. As the number of images increases, browsing will one day cease to be an efficient information access method. However, until that day, a functionality for more efficient browsing is essential. For example, a zooming functionality would be helpful. It would enable the placement of a large number of thumbnail photos on a screen; detailed images can be accessed by the user zooming in on an image or set of images. A zooming functionality can also include a time aspect [42]. Browsing by means of color has also been studied [43]. It can help users to organize images from different perspectives. Categorized by one's own criteria also has some educational effect. Browsing tools are more effective than retrieval systems for such purposes [44]. One of the most frequently used methods is sorting while browsing. Images are often sorted according to the date they were created (or accessed) and then browsed. Graham et al. reported that their browser which summarizes images using the time information, could assist users in efficiently finding images [45].

We can assess the usability of working systems or prototypes to identify their problematic elements by asking human subjects to use them and then monitoring their operations. In particular, some usability studies use advanced recording tools for tracking the behaviors of users. An example is using eye-tracking technology to determine where users are looking while they are searching or browsing. Moreover, some software can keep records on how people use pointers and where people click on their screen.

4.3 Utility

Usability used to be the main concern of human factors researchers. Gradually, as the use of images expanded to include various human activities, utility became the most important concern. Utility is heavily influenced by the social context; thus, it sometimes lies outside the realm of studies on technologies. For example, Jansen found that the cost of retrieval is the determining factor for the query reformulation among casual users of web image retrieval systems [46]. That is, they change queries if their initial queries yield only proprietary materials. This is not often the case for web text retrieval, where most documents are likely to be free. Other social factors will be discussed in Sec. 5.

Here, let us consider another important factor: the nature of the user's task. Regarding the role of image access in the entire work process, some insights have been gained from observing professional users. Markkula and Sormunen studied people who do editing work at newspapers and found that they tend to search images using queries of abstract concepts [47]. Another example is the queries for art images; they were classified into four categories [48]. The queries of art historians who do not do the searches themselves, but rather ask librarians for the images they need, were analyzed [49]. Queries of historical images in the medical domain were also studied [50]. These studies revealed the subjects the experts searched for, how long it took to fulfill their search needs, and the main obstacles that prevented them from obtaining satisfactory materials.

To understand the nature of image access, we have to take into account that information has a life cycle. Images are created by humans or acquired by sensors. Then, they are modified and stored. Sometimes they are sent to other locations. Finally, they are archived or discarded. How people organize their personal photo collections reveals a lot about the life cycle of the images [51]. Kirk et al. have started to investigate the life cycle of digital photos taken by casual photographers under the concept of photowork [41]. Shneiderman et al. contrasted the behavior of casual users with that of professional users and suggested system requirements for each type of user [52]. Cox et al. interviewed eleven amateur photographers and users of Flickr and analyzed their responses [53]. Their up-to-date findings show that the Flickr user group can be situated somewhere between casual users and professionals.

Distinguishing between retrieval styles makes the outcome clearer. Roughly speaking, there are two search types: known-item and exploratory. In a known-item search, users already know which images they want. A known-item search is a common way for people to search for books, music, and videos because people often search for commercial products such as pop songs, films, and best-seller books; it is a less common way of searching for images. One example of a known-item search is to obtain higher resolution images of the image at hand. Another example is to retrieve particular images in one's personal collection that the user knows is there because it had been created or added to the collection by the user. In contrast, exploratory searches are activities to decide which image is relevant for a certain purpose during the retrieval. For example, users usually search the web for images without assuming that the objects they are looking for are unique. Any objects described by the same information are good enough for a user's generic need. Therefore, some browsing is often involved in exploratory searches.

5 Social Factors Affecting Image Access and Use

5.1 Copyright and Privacy

For many office workers, information searches on the web are the quickest and most reliable option because the web is often regarded as the largest information resource. Images are sometimes included in the targets of the web resources because information receivers may find it easier to understand visual representations rather than textual ones (Sec. 4.3). However, copyright problems often reduce the utility of images. If a user is seeking particular knowledge, it suffices to assimilate any important facts textually described on a particular web page and then to use it later. Citing a part of text is also an acceptable usage. In contrast, a citation of images is more difficult than text because they are usually holistic, and the copying and pasting of images may constitute a copyright violation.

Copyright rules, either formal or informal, vary from region to region. For example, the responsibility of obtaining the copyright holder's permission is either in the writer's or the publisher's hand depending on the whether a book is published in Japan or in the US [54]. Such differences make it harder to use images legally and we can infer that this reduces the motivation of potential users.

Creative Commons (CC) [9] is an important initiative to address copyright issues. It allows creators, such as photographers or illustrators, to assign tags so that others can use their works without worrying about copyright violations. Although this scheme has yet to be established globally, it can create a large pool of *searchable and useable* image materials in contrast to *searchable but blocked* content. The increase in available images is a major driving force encouraging users to use image retrieval systems. The increase in need may consequently lead to greater technological sophistication.

Besides copyright regulations, even when the images are photos taken by the users themselves, there are problems of publicity and privacy. You cannot freely use a photo if it includes an identifiable person. For example, the University of Arizona has guidelines on photo usage[10]. Although such guidelines are useful, we must also be aware that laws and perceptions vary from region to region.

5.2 Cultural and Regional Aspect

Besides regional differences in copyright regulations, the culture of the user affects how he or she will use images. Different societies are responding to the explosion of digitized images in different ways. One factor is the socially dominant service for images. For example, Flickr is currently one of the largest on-line photo sharing services. Many studies are currently being conducted on this platform. For example, Negoescu and Gatica-Perez have analyzed how users having the same interests construct groups in Flickr [55]. Furthermore, research testbeds such as iCLEF have been created on it to evaluate interactive image retrieval [56] (Sec. 3.4). The problem is that the findings on a particular platform are not necessarily globally applicable. For example, in Japan, people are reluctant to publicly share photos on-line, but they actively exchange photos on mobile phones [57]. Similarly, in some European countries, people prefer to use password-protected on-line repositories to share photos rather than openly display them [58]. Therefore, if the research is based on realistic data and user models, the implications of the findings may be less than universally applicable.

On the positive side, an important utility characteristic that distinguishes images from text is their language independence. Although even pictograms are not universal and the use of iconic entities is culturally dependent, images can often deliver their message across lan-

[9] http://creativecommons.org/

[10] http://uaweb.arizona.edu/people.0.html

guage barriers. This property can make cross-language image retrieval a useful tool [59].

6 Conclusion

We overviewed the current research being conducted on image retrieval and its use in the information explosion. The greatest changes affecting the research currently being conducted on image retrieval and its use has been the changes in the users. People have become used to various visual entities, and the acquisition and distribution of images has become a common activity.

The use of digitized text information is growing because of its searchability. People already know how to use text information. In contrast, the use of digitized images has only come about with the recent emergence of certain technologies. Images used to be too expensive for the general public to use, but now anyone can use images and people are finding new ways of using them. Therefore, it is natural for there to be some confusion about image access technologies. This paper was an attempt to clarify the situation surrounding image access research in the first decade of this century.

Some of the issues have already been dealt with in two rather separate research fields, namely computer and library science. The activities and findings of researchers in these fields are now being merged and are generating practical image retrieval systems. Images used to be treated as special cases in library science and information retrieval research. This was natural because we still have far more textual documentation and we use text more often than we do images. In contrast, although image processing is a central concern of computer science researchers, it is often regarded as a single application task. What was missing from computer science is the information life cycle perspective of library science. The difference between how image retrieval is dealt with in these two fields will not go away anytime soon. Despite this handicap to progress, there will continue to be an ingression of images into our daily lives, and retrieval technologies will remain a key to this trend. Advances in the technologies and the underlying methodologies will enable library and computer scientists to overcome their social barriers so that they can help to usher in new life- and work-styles using images.

References

[1] P. Enser, "The evolution of visual information retrieval," *Journal of Information Science*, vol.34, no.4, pp.531–546, 2008.

[2] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol.2, no.1, pp.1–19, 2006.

[3] M. Inoue, "On the need for annotation-based image retrieval," *Proc. Information Retrieval in Context (IRiX)*, Sheffield, UK, pp.44–46, July 2004.

[4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.12, pp.1349–1380, 2000.

[5] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol.40, no.2, pp.1–60, 2008.

[6] L.H. Armitage and P.G. Enser, "Analysis of user need in image archives," *Journal of Information Science*, vol.23, no.4, pp.287–299, 1997.

[7] P.G. Enser, C.J. Sandom, J.S. Hare, and P.H. Lewis, "Facing the reality of semantic image retrieval," *Journal of Documentation*, vol.63, no.4, pp.465–481, 2007.

[8] N. Vasconcelos, "From pixels to semantic spaces: Advances in content-based image retrieval," *Computer*, vol.40, no.7, pp.20–26, 2007.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol.60, no.2, pp.91–110, November 2004.

[10] H. Müller, S. Marchand-Maillet, and T. Pun, "The truth about Corel evaluation in image retrieval," *CIVR '02: Proceedings of the International Conference on Image and Video Retrieval, LNCS*, vol.2383, pp.38–49, 2002.

[11] H. Müller, W. Müller, S. Marchand-Maillet, T. Pun, and D.M. Squire, "A framework for benchmarking in CBIR," *Multimedia Tools Appl.*, vol.21, no.1, pp.55–73, 2003.

[12] M. Grubinger, P. Clough, H. Müller, and T. Deselaers, "The IAPR TC-12 benchmark: A new evaluation resource for visual information systems," In *Proceedings of International Workshop OntoImage 2006 Language Resources for Content-Based Image Retrieval, held in conjunction with LREC'06*, Genoa, Italy, pp.13–23, 2006.

[13] N.R. Howe, "Using artificial queries to evaluate image retrieval," *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'00)*, Hilton Head, South Carolina, p.5, June 16 2000.

[14] M. Inoue and N. Ueda, "Retrieving lightly annotated images using image similarities," *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, pp.1031–1037, March 2005.

[15] T. Arni, J. Tang, M. Sanderson, and P. Clough, "Creating a test collection to evaluate diversity in image retrieval," *SIGIR 2008 Workshop: Beyond Binary Relevance: Preferences, Diversity and Set-Level Judgment*, Singapore, July 24 2008.

[16] K. Song, Y. Tian, W. Gao, and T. Huang, "Diversifying the image retrieval results," *Proceedings of the 14th Annual ACM international Conference on Multimedia*, Santa Barbara, CA, USA, pp.707–710, October 23 - 27 2006.

[17] P. Harpring, Introduction to art image access: Issues, tools, standards, strategies, ch. The language of images: Enhancing access to images by applying metadata schemas and structured vocabularies, pp.20–39, Getty Publications, 2002.

[18] J. Magalhães, F. Ciravegna, and S. Rüger, "Exploring multimedia in a keyword space," *ACM Multimedia*, Vancouver, Canada, October 2008.

[19] K. Hwang, J. Choi, and P. Kim, "An efficient retrieval of annotated images based on WordNet," *Advanced Communication Technology, The 9th International Conference on*, pp.368–372, Feb. 2007.

[20] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no.10, pp.1615–1630, Oct. 2005.

[21] T.T. Pham, N.E. Maillot, J.H. Lim, and J.P. Chevallet, "Latent semantic fusion model for image retrieval and annotation," *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, Lisbon, Portugal, pp.439–444, 2007.

[22] M. Inoue, "Mining visual knowledge for multi-lingual image retrieval," *DMIR'07: IEEE International Symposium on Data Mining and Information Retrieval*, Niagara Falls, Ontario, Canada, pp.307–312, May 21-23 2007.

[23] K. Yanai and K. Barnard, "Image region entropy: A measure of "visualness" of web images associated with one concept," *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, Singapore, pp.419–422, 2005.

[24] X.S. Zhou and T.S. Huang, "Relevance feedback in image retrieval: A comprehensive review.," *Multimedia Syst.*, vol.8, no.6, pp.536–544, 2003.

[25] M. Ferecatu, N. Boujemaa, and M. Crucianu, "Semantic interactive image retrieval combining visual and conceptual content description," *Multimedia Systems*, vol.13, no.5-6, pp.309–322, Feburary 2008.

[26] Camera & Imaging Products Association (CIPA), "2008-2010 outlook on the shipment forecast by product-type concerning cameras and related goods," January 29 2008.

[27] S. Shankland, "Cameras: Shipments rising, but prices falling, CNET, September 19," 2007.

[28] B. Adams, D. Phung, and S. Venkatesh, "Sensing and using social context," *ACM Trans. Multimedia Comput. Commun. Appl.* vol.5, no2, pp1–27, 2008.

[29] W. Müller and A. Henrich, "Fast retrieval of high-dimensional feature vectors in P2P networks using compact peer data summaries," *MIR '03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, Berkeley, California, pp.79–86, 2003.

[30] R. Gillesse, J. Rog, and A. Verheusen, "Alternative file formats for storing master images of digitisation projects," tech. rep., National Library of the Netherlands Research & Development Department, March 2008.

[31] N. Weber and A. Henrich, "Retrieval of technical drawings in DXF format - concepts and problems," *LWA 2007*, pp.213–220, 2007.

[32] F. Falchi, C. Lucchese, R. Perego, and F. Rabitti, "CoPhIR: COntent-based Photo Image Retrieval," May 30 2008.

[33] C. Jörgensen, *Image Retrieval: Theory and Research*, Scarecrow Press, 2003.

[34] A.E. Stamps III, "Entropy, visual diversity, and preference," *The Journal of General Psychology*, vol.129, no.3, pp.300–320, July 2002.

[35] M. Inoue, "Difficulties in image retrieval," *Contextual and Social Media Understanding and Usage*, Dagstuhl Seminar Proceedings, no.08251, Dagstuhl, Germany, 2008.

[36] B. Scott, "Picturing irony: The subversive power of photography," *Visual Communication*, vol.3, no.31, pp.31–59, 2004.

[37] A. Fujii and T. Ishikawa, "Image retrieval and disambiguation for encyclopedic web search," *The 9th International Joint Conference on Artificial Intelligence (IJCAI-05)*, Aug. 2005.

[38] M. Theune, B.V. Schooten, R.D. Akker, and W. Bosma, "Questions, pictures, answers: Introducing pictures in question-answering systems," *Tenth international symposium on social communication*, Cuba, January 2007.

[39] T. Yeh, J. Lee, and T. Darrell, "Photo-based question answering," *ACM Multimedia*,pp.389–398, 2008.

[40] Y. Sumi, J. Ito, and T. Nishida, "PhotoChat: communication support system based on sharing photos and notes,," *CHI 2008 Extended Abstracts, Works in progress*, Florence, Italy, pp.3237–3242, April 2008.

[41] D.S. Kirk, A.J. Sellen, C. Rother, and K.R. Wood, "Understanding photowork," *Proc. CHI 2006*, Montréal, Québec, Canada, pp.761–770, April 2006.

[42] D.F. Huynh, S.M. Drucker, P. Baudisch, and C. Wong, "Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections," *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, Portland, Oregon, USA, pp.1937–1940, 2005.

[43] T. Kajiyama, N. Kando, and S. Satoh, "Examination and enhancement of a ring-structured graphical search interface based on usability testing," *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, Salvador, Brazil, pp.623–624, 2005.

[44] M. Chang, J.J. Leggett, R. Furuta, A. Kerne, J.P. Willams, S.A. Burns, and R.G. Bias, "Collection understanding," *JCDL'04*, Tuscon, Arizona, USA, pp.334–342, June 7–11 2004.

[45] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd, “Time as essence for photo browsing through personal digital libraries,” *JCDL ’02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, Portland, Oregon, USA, pp.326–335, 2002.

[46] B.J. Jansen, “Searching for digital images on the web,” *Journal of Documentation*, vol.64, no.1, pp.81–101, 2008.

[47] M. Markkula and E. Sormunen, “End-user searching challenges indexing practices in the digital newspaper photo archive,” *Information Retrieval*, vol.1, no.4, pp.259–285, 2000.

[48] S.K. Hastings, “Query categories in a study of intelectual access to digitized art images,” *Proceedings of the 58th ASIS Annual Meeting*, pp.3–8, 1995.

[49] Y. Choi and E.M. Rasmussen, “Searching for images: the analysis of users’ queries for image retrieval in american history,” *J. Am. Soc. Inf. Sci. Technol.*, vol.54, no.6, pp.498–511, 2003.

[50] L.H.Keister, Challenges in indexing electronic text and images, ch. User types and queries: impact on image access systems, pp.7–19, ASIS Monograph Series, Information Today, Inc., Medford, NJ, 1994.

[51] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss, “Requirements for photoware,” *CSCW ’02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA, pp.166–175, 2002.

[52] B. Shneiderman, B.B. Bederson, and S.M. Drucker, “Find that photo!: interface strategies to annotate, browse, and share,” *Commun. ACM*, vol.49, no.4, pp.69–71, 2006.

[53] A.M.Cox, P. Clough, and J. Marlow, “Flickr: a first look at user behaviour in the context of photography as serious leisure,” *Information Research*, vol.13, no.1, 2008.

[54] “Report on the first NCC image use protocol task force meeting,” 2007. IUP Task Force Meeting.

[55] R.A. Negoescu and D. Gatica-Perez, “Analyzing Flickr groups,” *CIVR ’08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, Niagara Falls, pp.417–426, Jul. 2008.

[56] P. Clough, J. Karlgren, and J. Gonzalo, “Multilingual interactive experiments with Flickr,” *Proceedings of Workshop on New Text - wikis and blogs and other dynamic text sources*, Trento, Italy, 2006.

[57] D. Okabe, “Social practice of camera phone in japan,” *First Workshop on Pervasive Image Capture and Sharing, In Conjunction with Ubicomp’05*, 2005.

[58] S. Thieme, “Personal communication (27 Oct. 2008).”

[59] M. Inoue, “The remarkable search topic-finding task to share success stories of cross-language information retrieval,” *New Directions in Multilingual Information Access: A Workshop at SIGIR 2006*, Seattle, USA, pp.61–64, August 2006.

Masashi INOUE

Masashi INOUE received a BA degree in natural science from the International Christian University (ICU), Japan, in 1999 and his MS and Doctoral degrees in information science from the Nara Institute of Science and Technology (NAIST), Japan, in 2001 and 2004, respectively. He joined the National Institute of Informatics (NII), Tokyo, Japan in 2004. He is an assistant professor in the Digital Content and Media Sciences Research Division. His research interests include the problem of data scarcity and the utilization of multiple information sources in pattern recognition and information access, the utility and evaluation issues in user-system interactions, and knowledge extraction from multimodal dialogue and its use. He is a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), the Japanese Society for Artificial Intelligence (JSAI), the Japanese Psychological Association (JPA), the Association for Computing Machinery (ACM), and the Institute of Electrical and Electronics Engineers (IEEE).