

Photo Retrieval: Multimedia's Chance to Solve a Real Problem for Real People

Ramesh Jain
University of
California, Irvine

Have you noticed that all major search engines list Web, images, and video above the keyword box in that order? And that the text search (wrongly called Web) works quite well but the image search is poor?

Also, have you noticed that most people, if not all, reading this article carry a camera in their pocket wherever they go? All statistics suggest that digital photography has revolutionized the lifestyle of many people and that this trend will continue.

And, I am sure you have noticed that this magazine has published many research articles on image search and image retrieval for the last 13 years. But did you notice that though the nature of photography has changed, the problem of image search and retrieval hasn't changed much? It's really intriguing that research in image search and retrieval continues to use the camera obscura model and that most researchers doing experiments in this area like to test their algorithms using Corel images.

Considering modern cameras

A modern digital camera is no longer just a simple pinhole camera obscura that records optical information about the environment as a projection of intensity values on a chemical surface. A digital camera, even the simplest point-and-shoot camera, has multiple sensors and a flash to capture and modify the environment. Along with the intensity values, it records several additional optical and other parameters. Most importantly, it records the focal length, aperture opening, aperture time, and flash-related information. It also stores the clock and the camera model.

In many cameras, a user can set a specific scene mode to declare the intended nature of the image so the camera can adjust appropriate parameters—and this mode is also recorded with the picture. Already cameras are available with GPS, and

this functionality will become common soon, so you'll know exactly where the picture was taken.

In short, a digital camera is no longer a simple camera; it's a multimedia device that stores the environment where a photo is taken. All this information is explicitly available along with the picture's intensity values. Most digital cameras use standards like the exchangeable image format (EXIF) that stores all this and more information in each digital photo (see <http://www.exif.org> for more about EXIF).

Envisioning better search and retrieval

Most research papers in image retrieval demonstrate search-related solutions as if we can solve the problem using only intensity values. It's a mystery to me that researchers are reluctant to use all of the other obvious information that's easily available and helpful.

If we consider all the information available from digital cameras in digital photos, we can get a significant head start, even before we open the so-called image. One important factor to remember is that all this information—let's call it meta-data in digital photos—is easily available and can be processed several orders of magnitude faster than processing the images themselves (which are becoming increasingly larger in size).

Focusing on the periphery

The most important factor is that even the most powerful visual system known to humans—the human visual system—uses all kinds of meta-data. Also, it's well known that humans first use peripheral vision to form a general idea about what they might see before they really focus on the details and spend valuable perceptual resources on understanding the content of an image captured in their central vision.

continued on p. 111

continued from p. 112

Similarly, we could use the metadata available in digital photos to get peripheral information about the environment in which the picture was taken. What's equally important is that by processing a sequence of digital photographs (rather than only an isolated photo) from a camera and using the time stamps and GPS information from photos, it might be possible to interpret a sequence of photos taken in a context.

Finding a better context

Most photos are usually taken in the context of an event such as a meeting, party, trip, performance, or even the first steps taken by your daughter. At each event, usually we take multiple photos, because digital photos are cheap to take and store. So not only can we use the metadata to get some peripheral environment information, we can also tag the photos as some events, particularly if we have access to a calendar or other information about an image's context. And I'm still not talking about using the rich maps becoming available in cameras that could provide more information—such as using the latitude and longitude (and compass)—about the visual environment in which the picture was taken.

It's truly amazing that we throw away all this information about digital photos and try to solve a problem that we think is difficult. In developing photo management systems, we should use all the information from camera metadata and human-induced tags, as well as any other sources—such as those used by current commercial search engines. Note, though, that human-assigned tags popularized by systems such as Flickr are useful but suffer from many problems that make their application very limited.

Searching for more

I've been working on this problem of image retrieval since before digital images became commonly available to the average consumer. Now digital photos are not only available, they're part of everybody's lifestyle. And in this time, we've not developed any systems that can help organize and retrieve photos. I started doing research in this area around 1990 and cofounded a company called Virage (that really meant Visual Information Retrieval Age) in 1993.

Today, I have more than 15,000 digital photos (increasing by about 1,000 every other month) on my computer, all in different folders. There's no system to help me. And I'm getting

Postscript

As I finished this article, I saw an advertisement of a camera with face detection built into it. It claimed to capture "your digital life."

frustrated with the situation. I need a system that will help me organize my photos. Believe me, I'm not the only person with this situation. This is a serious problem that needs a solution.

I would love to see researchers in the field of multimedia information systems leave their obsession with only using image intensity values and use any and all information that comes from cameras, personal calendars, maps, and any other sources on the Web to help solve the problem. Today's search engines—including Google and Yahoo—use only metadata available on the page on which the image appears. And photo-oriented systems also use very limited, if any, information. I for one have decided to consider photo retrieval a true multimedia problem, and hence use all the metadata from cameras, calendars, maps, the Web, and also intensity values in images. **MM**

Readers may contact Ramesh Jain at jain@ics.uci.edu.

Contact Media Vision editor Forouzan Golshani at golshani@wright.edu.

**The IEEE
Computer
Society**

**publishes over
150 conference
publications a year.**

**For a preview of the latest
papers in your field, visit**

www.computer.org/publications/