

IMAGE MANAGEMENT

FIND PHOTOS VIA COLOR AND SPATIAL RELATIONSHIPS

By Paul Worthington

Unorganized photos on a hard drive can't be enjoyed by anyone. If you can't find the shots you want, photo sharing and printing are all but impossible. And it can be very difficult to find a particular photo amongst the thousands stored on a hard drive: you have to know the date on which a photo was taken... or you must have either manually added tags that describe the picture's contents... or placed every photo, as you copied it to your PC, in an appropriately named folder.

If a search tool could instead use one photo to find others that are similar, it would be easy and intuitive for users of all skill levels — and it would not need the type of "advance work" most photo organization requires.

While most image search tools use text, or, in a few cases, color or texture, the new Xcavator tool recognizes both color and spatial relationships: not just a face's skin tone and a shirt's color, for example, but, when the user clicks on the face and shirt in one image, the distance between the face and clothing — and so it can quickly find all the available images that match both factors.

To work with the "image recognition engine," users pick the key features of interest in a photo that define the type of similarity they want to see in more photos.

As shown in an example video on the company's website, clicking on one red-bronze horizontal beam on the Golden Gate bridge, then the sky showing behind it, and then another beam, finds dozens of GG bridge shots — as only they match those colors in that pattern.

CogniSign says its "intelligent image recognition technology" software uses principles from cognitive psychology to emulate specific human image recognition processes. Their computational model describes "how visual memory and serial attention cooperate in the human brain," which can establish similarity between two non-identical images in a human-like fashion. As when humans view images, the technology is sensitive to some discrepancies in color, shape, and viewing perspective, and is highly tolerant to variation in

position and scale.

The privately held company, based in San Francisco, CA, has been working primarily in the enterprise IT, security, and defense sectors.

The Xcavator photo search and content management tool is its first foray into consumer imaging. CogniSign says Xcavator works on a typical home PC, or with web-based photo management systems and sharing sites. The technology can "easily be configured" to search distributed photo databases across an unlimited number of servers, the company says.

The company will be featured in our look at surveillance, but this month we focus on the consumer imaging angle in our talk with CEO Bryan Calkins.

What is the primary problem CogniSign set out to solve?

There's really two core problems that we solve: One is a task of finding images in a database that are similar to a given image. That's what the Excavator demo does; if we have a picture of this flower we can find stuff like this. If we have a picture of this cabin but we don't like the perspective, we can find other — it really does work. We can find other pictures of the cabin that would give you a better perspective. So the task here is to find images in a target database that are similar to a given image.

And then, using the same core technology in a slightly different way, there is object detection: to find specific objects in complex images. Say there might be satellite surveillance, either still or video; and we've got stored "objects of interest" in an object library — say, trucks, for example — and our technology helps find those trucks in sequences of complex images. This is actually a very difficult project, and our technology moves this whole thing quite a few steps forward as well.

What lets your technology



accomplish this?

Our technology is tolerant to variation in orientation, size, color, and shape. We've tuned it so it's more like a human. For example, in respect to proximity: if you're looking at two objects, and in one image they're right next to one another, and in another image they're quite a ways apart — the human mind will "ignore" that distance — and we can tune our technology to kind of ignore that, just as the human mind does. Or we can tune it to pay attention to that, to say, well, we want these two items together just like they are in the original picture.

The existing technologies are insufficient for emerging applications. Basically there's two major technologies for image search: text-based matching — tags, and metadata — and the space we're in, which is content-based image retrieval, or CBIR. It's image matching based on the content of the image.

I call it kind of a cottage industry, because there's hasn't yet been a powerful enough technology to break out.

What has been done today is images are indexed, reduced to a numerical description or summary of their contents. And that doesn't work very well, for a number of reasons. You end up with color histograms and texture statistics and stuff like that — but what you end up with is a situation where a sunset might look like a monarch butterfly, to the indexing program...

It identifies colors and textures, but not, for instance, geometric shapes?

Right, or their spatial relationship to one another.

That's where we come in — that's the key, actually: geometric shapes and their spatial relationships to one another.

The matching process generally works

with only a few points, and it looks at images in an entire target database. The key to this technology is the spatial relationship. It's the secret sauce, that the



CogniSign CEO Bryan Calkins

spatial relationship is continuously maintained and analyzed and updated during the point-selection process. That's actually really hard to do in a way that's meaningful for image recognition.

In the cabin photo example: How does the system not rule out photos where, from a different perspective, those points would have a different spatial relationship?

It's tolerant to differences in those spatial relationships. That's the key. For example, where there are white pillars on the cabin porch — they're just a slightly different distance apart.

The other thing about the use of spatial relationships is that it creates a "lighter-weight" image recognition technology — it is less computationally intensive because you're just looking at points and areas, instead of the whole image. It's more scalable. We think we can easily manage 20,000 images in a typical high-end computer for the purpose of our application. Maybe even more.

Other technologies are like carving it into pieces: "this is an object, this is an object..." And that is incredibly computationally intensive.

Will that lead to use in other fields?

Yes: We had a meeting with one of the major chip makers recently, and they're interested in this technology for object avoidance systems in cars.

They're already putting video processors in cars, taking pictures, and they want the car to be able to immediately highlight a deer or a telephone pole or a stroller, or whatever. And we believe that our technology can do all that.

Other than the spatial relationship analysis, what else sets this apart from a user perspective?

Most enterprise internet search applications are both single-query and text-based; that's what you see on Google images.

Desktop photo management is also pretty much text-based, but they've become a little bit more interactive with these visual rolls of film, a metaphor for visual collections [such as used by Picasa and iPhoto]. It's a huge step forward compared to PC folders and file names.

CBIR is image-based, and is indexing the entire image. It's still single-query, and it's incredibly static. The system only has one way of indexing this photo, usually, and it can only look at that image in that singular kind of way.

A key element of our process is that it's interactive: image parts can be selected intelligently, to better define the search. It's a natural and engaging process for humans, and you get much better results. Our UI is basically picking sequences of points.

You mean the user chooses, not the algorithm?

Right — but that's an interesting point, because the algorithm can be extended to handle stored searches.

For example: 20th Century Fox, with "The Simpsons," constantly has incoming content — not just new episodes, but also merchandise photos and promotional stuff... They can do stored searches of say, Mr. Burns or Bart or Lisa, and then that content can be automatically categorized and labeled by our system.

The capability for such stored searches really extends this technology.

And there are examples in our patent applications where the computer generates the sequence of points for the search.

Can consumers use your system to automatically find and label who is in a photo?

Facial recognition is an exciting example, because the points, the key

features, are all in the same place: there are automatic location elements, so it doesn't even require stored searches.

We can deploy a technology like that eventually, where you just have an input group of photos of one identified face, like your son or your daughter or your mom, and that future photos, as they're put into a target database, can be indexed that way. Face recognition is relatively easy, because it is highly structured.

How does it work on the Net?

We can launch a much more refined Internet search platform. There's a big opportunity in Internet searches. Current applications [such as Google and Yahoo] are text-based. So when you type in, say, "Dali," for Salvador Dali, if there's a particular piece of artwork you're interested in, you're going to get pictures of Dali himself, and all his artwork [and perhaps pictures of the Dalai Lama...]

For a demonstration, we have interfaced with Flickr: we search through a Flickr keyword set of 4,000 images.

[Calkins demonstrates the Web application.] So — we just start picking points, and some pretty sophisticated color recognition is going on. We've got a lot of matches with a color combination [of two points on a particular woman's outfit]. There's another one. Is that her?

Yes, it is. Same outfit.

Then I start picking blue... and they all come up.

That's very cool.

And the technology here is not optimized for a desktop application, obviously: this is working in a browser, looking at the Flickr site through its open API.

Our technology works great on the Web, but it also works great on a desktop application. It's not like we have to have 20 servers back at the office processing this Web version.

Years ago we reported on a similar thesis: Web searches can be better with images. That was probably before its time...

Well, based on what we're seeing in the market, we're a little bit early still — but I'd rather be a little bit early and be patient, than arrive late. ■