New Algorithms from UCSD Improve Automated Image Labeling

March 29, 2007

Daniel Kane

A Google image search for "tiger" yields many tiger photos - but also returns images of a tiger pear cactus stuck in a tire, a racecar, Tiger Woods, the boxer Dick Tiger, Antarctica, and many others. Why? Today's large Internet search engines look for images using captions or other text linked to images rather than looking at what is actually in the picture.

Electrical engineers from UC San Diego are making progress on a different kind of image search engine - one that analyzes the images themselves. This approach may be folded into next-generation image search engines for the Internet; and in the shorter term, could be used to annotate and search commercial and private image collections.

"You might finally find all those unlabeled pictures of your kids playing soccer that are on your computer somewhere," said Nuno Vasconcelos, a professor of electrical engineering at the UCSD Jacobs School of Engineering, and senior author of a paper in the March 2007 issue of the IEEE journal *TPAMI* - a paper coauthored by Gustavo Carneiro, a UCSD postdoctoral researcher now at Siemens Corporate Research, UCSD doctoral candidate Antoni Chan, and Google researcher Pedro Moreno.

At the core of this Supervised Multiclass Labeling (SML **)** system is a set of simple yet powerful algorithms developed at UCSD. Once you train the system, you can set it loose on a database of unlabeled images. The system calculates the probability that various objects or "classes" it has been trained to recognize are present - and labels the images accordingly. After labeling, images can be retrieved via keyword searches. Accuracy of the UCSD system has outpaced that of other content-based image labeling and retrieval systems in the literature. The SML system also splits up images based on content - the historically difficult task of image segmentation. For example, the system can separate a landscape photo into mountain, sky and lake regions.

"Right now, Internet image search engines don't use any image content analysis. They are highly scalable in terms of the number of images they can search but very constrained on the kinds of searches they can perform. Our semantic search system is not fully scalable yet, but if we're clever, we will be able to work around this limitation. The future is bright," said Vasconcelos.

The UCSD system uses a clever image indexing technique that allows it to cover larger collections of images at a lower computational cost than was previously possible. While the current version would still choke on the Internet's vast numbers of public images, there is room for improvement and many potential applications beyond the Internet, including the labeling of images in various private and commercial databases.

The UCSD **Supervised Multiclass Labeling** system "...outperforms existing approaches by a significant margin, not only in terms of annotation and retrieval accuracy, but also in terms of efficiency," the authors write in their *TPAMI* (*IEEE Transactions on Pattern Analysis and Machine Intelligence*) paper.

What does *Supervised Multiclass Labeling* **mean?**

Supervised refers to the fact that the users train the image labeling system to identify classes of objects, such as "tigers," "mountains" and "blossoms," by exposing the system to many different pictures of tigers, mountains and blossoms. The supervised approach allows the system to differentiate between similar visual concepts - such as polar bears and grizzly bears. In contrast, "unsupervised" approaches to the same technical challenges do not permit such fine-grained distinctions. **"Multiclass"** means that the training process can be repeated for many visual concepts. The same system can be trained to identify lions, tigers, trees, cars, rivers, mountains, sky or any concrete object. This is in contrast to systems that can answer just one question at a time, such as "Is there a horse in this picture?" (Abstract concepts like "happiness" are currently beyond the reach of the new system, however.) **"Labeling"** refers to the process of linking specific features within images directly to words that describe these features.

Scientists have previously built image labeling and retrieval systems that can figure out the contents of images that do not have captions, but these systems have a variety of drawbacks. Accuracy has been a problem. Also, some older systems need to be shown a picture and then can only find similar photos. Other systems can only determine whether one particular visual concept is present or absent in an image. Still others are unable to search through large collections of images, which is crucial for use in big photo databases and perhaps one day, the Internet. The new system from the Vasconcelos team begins to addresses these open problems.

To understand SML, you need to start with the training process, which involves showing the system many different pictures of the same visual concept or "class," such as a mountain. When training the system to recognize mountains, the location of the mountains within the photos does not need to be specified. This makes it relatively easy to collect the training examples. After exposure to enough different pictures that include mountains, the system can identify images in which there is a high probability that mountains are present.

During training, the system splits each image into 8-by-8 pixel squares and extracts some information from them. The information extracted from each of these squares is called a "localized feature." The localized features for an image are collectively known as a "bag of features."

Next, the researchers pool together each "bag of features" for a particular visual concept. This pooled information summarizes - in a computationally efficient way - the important information about each of the individual mountains. Pooling yields a density estimate that retains the critical details of all the different mountains without having to keep track of every 8 by 8 pixel square from each of the mountain training images.

After the system is trained, it is ready to annotate pictures it has never encountered. The visual concepts that are most likely to be in a photo are labeled as such. In the tiger photo, the SML system processed the image and concluded that "cat, tiger, plants, leaf and grass" were the most likely items in the photograph.

The system, of course, can only label images with visual concepts that it has been trained to recognize.

"At annotation time, all the trained classes directly compete for the image. The image is labeled with the classes that are most likely to actually be in the image," said Vasconcelos.

One way to test the SML system is to ask it to annotate images in a database and then retrieve images based on text queries. In the *TPAMI* paper, the researchers illustrate some of their image annotation and retrieval results for searches for the following classes: blooms, mountain, pool, smoke and woman.

In their paper, Vasconcelos and colleagues also document the similarities between SML's automated image labeling and labeling done by humans looking at the same pictures.
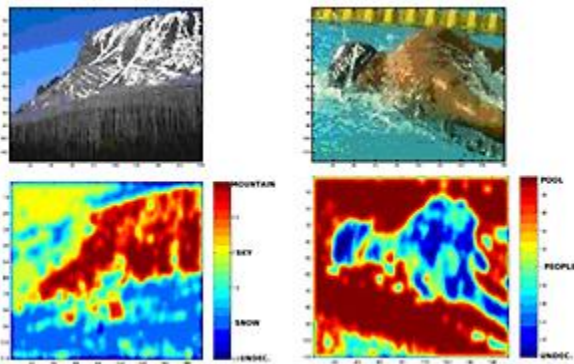
The SML system can also split up a single image into its different regions - a process known as "image segmentation." When the system annotates an image, it assigns the most likely label to each group of pixels or localized feature, segmenting the image into its most likely parts as a regular part of the annotation process.

"Automated segmentation is one of the really hard problems in computer vision, but we're starting to get some interesting results," said Vasconcelos.

The SML project was started in 2004 by Gustavo Carneiro who was then a post doctoral researcher in the Vasconcelos lab. Dr. Carneiro currently works at Siemens Corporate Research in Princeton, New Jersey. Doctoral student Antoni Chan, the second author on the paper, spent a summer at Google testing the system on a cluster of 3,000 state-of-the-art Linux machines. Chan worked under the guidance of Dr. Pedro Moreno, a Google researcher and author on the paper. The results from the Google work indicate that the system can be used on large image collections, Chan explained.

"My students go to Google and do experiments at a scale that they can't do here. The collaboration with Google allows us to use their resources to do things we couldn't do otherwise," said Vasconcelos.

Media Contact: Daniel Kane, 858-534-3262

Human Annotation	sky jet plane smoke	snow fox arctic	sky buildings street cars
SML Annotation	plane jet smoke flight prop	arctic snow polar fox ice	street buildings bridge sky arch
Human Annotation	grass forest cat tiger	bear polar snow tundra	coral fish ocean reefs
SML Annotation	cat tiger plants leaf grass	polar tundra bear snow ice	reefs coral ocean fan fish