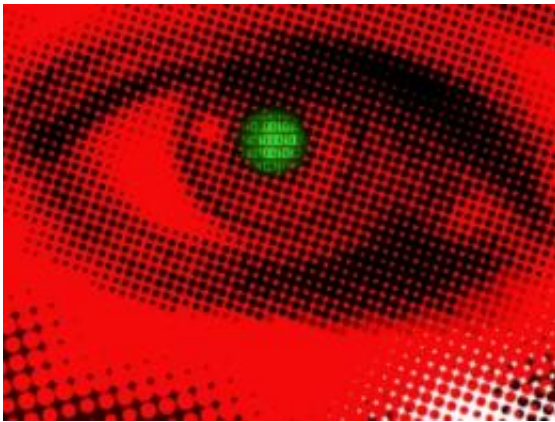


Seeing the forest for the trees: Object recognition system breaks images into ever smaller parts

May 7 2010, by Larry Hardesty



Graphic: Christine Daniloff

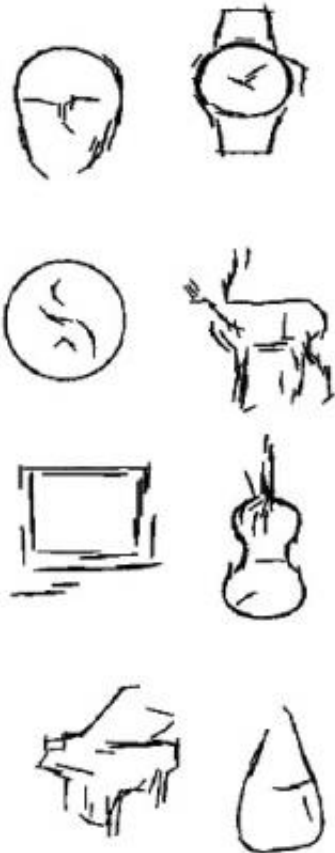
Object recognition is one of the core topics in computer vision research: After all, a computer that can see isn't much use if it has no idea what it's looking at. Researchers at MIT, working with colleagues at the University of California, Los Angeles, have developed new techniques that should make object recognition systems much easier to build and should enable them use computer memory more efficiently.

A conventional object recognition system, when trying to discern a particular type of object in a [digital image](#), will generally begin by looking for the object's salient features. A system built to recognize

faces, for instance, might look for things resembling eyes, noses and mouths and then determine whether they have the right spatial relationships with each other. The design of such systems, however, usually requires human intuition: A programmer decides which parts of the objects are the right ones to key in on. That means that for each new object added to the system's repertoire, the programmer has to start from scratch, determining which of the object's parts are the most important.

It also means that a system designed to recognize millions of different types of objects would become unmanageably large. Each object would have its own, unique set of three or four parts, but the parts would look different from different perspectives, and cataloguing all those perspectives would take an enormous amount of [computer memory](#).

In a paper that they'll present at the Institute of Electrical and Electronics Engineers' Conference on Computer Vision and [Pattern Recognition](#) in June, postdoc Long (Leo) Zhu and Professors Bill Freeman and Antonio Torralba, all of MIT's Computer Science and [Artificial Intelligence](#) Laboratory, and Yuanhao Chen and Alan Yuille of UCLA describe an approach that solves both of these problems at once. Like most object-recognition systems, their system learns to recognize new objects by being "trained" with digital images of labeled objects. But it doesn't need to know in advance which of the objects' features it should look for. For each labeled object, it first identifies the smallest features it can — often just short line segments. Then it looks for instances in which these low-level features are connected to each other, forming slightly more sophisticated shapes. Then it looks for instances in which these more sophisticated shapes are connected to each other, and so on, until it's assembled a hierarchical catalogue of increasingly complex parts whose top layer is a model of the whole object.



A new object recognition system developed by researchers at MIT and UCLA looks for rudimentary visual features shared by multiple examples of the same object. Then it looks for combinations of those features shared by multiple examples, and combinations of those combinations, and so on, until it has assembled a model of the object that resembles a line drawing. Images: Long (Leo) Zhu

Economies of scale

Once the system has assembled its catalogue from the bottom up, it goes through it from the top down, winnowing out all the redundancies. In the parts catalogue for a horse seen in profile, for instance, the second layer from the top might include two different representations of the horse's rear: One could include the rump, one rear leg and part of the belly; the

other might include the rump and both rear legs. But it could turn out that in the vast majority of cases where the system identifies one of these “parts,” it identifies the other as well. So it will simply cut one of them out of its hierarchy.

Even though the hierarchical approach adds new layers of information about digitally depicted objects, it ends up saving memory because different objects can share parts. That is, at several different layers, the parts catalogues for a horse and a deer could end up having shapes in common; to some extent, the same probably holds true for horses and cars. Wherever a shape is shared between two or more catalogues, the system needs to store it only once. In their new paper, the researchers show that, as they add the ability to recognize more objects to their system, the average number of parts per object steadily declines.

Although the researchers’ work promises more efficient use of computer memory and programmers’ time, “it is far more important than just a better way to do object recognition,” says Tai Sing Lee, an associate professor of computer science at Carnegie Mellon University. “This work is important partly because I feel it speaks to a couple scientific mysteries in the brain.” Lee points out that visual processing in humans seems to involve five to seven distinct brain regions, but no one is quite sure what they do. The researchers’ new object recognition system doesn’t specify the number of layers in each hierarchical model; the system simply assembles as many layers as it needs. “What kind of stunned me is that [the] system typically learns five to seven layers,” Lee says. That, he says, suggests that it may perform the same types of visual processing that takes place in the brain.

In their paper, the MIT and UCLA researchers report that, in tests, their system performed as well as existing object-recognition systems. But that’s still nowhere near as well as the human brain. Lee says that the researchers’ system currently focuses chiefly on detecting the edges of

two-dimensional depictions of objects; to approach the performance of the human brain, it will have to incorporate a lot of additional information about surface textures and three-dimensional contours, as the brain does. Zhu adds that he and his colleagues are also pursuing other applications of their technology. For instance, their hierarchical models naturally lend themselves not only to automatic object recognition — determining what an object is — but also automatic object segmentation — labeling an object’s constituent parts.

Provided by Massachusetts Institute of Technology

Citation: Seeing the forest for the trees: Object recognition system breaks images into ever smaller parts (2010, May 7) retrieved 10 April 2024 from <https://phys.org/news/2010-05-forest-trees-recognition-images-smaller.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--