

I Google, therefore I am

Robots are going online to escape the narrow confines of their programming and learn about the world

MICHAEL REILLY

SHUT your eyes and picture a Buddha's hand citron. Chances are you can't, unless you happen to know it's an exotic yellow, tentacled fruit that looks a bit like a small bunch of bananas.

The exercise gives a taste of what it's like for robots, which struggle to "picture" almost every word or phrase they come across. That's because until they are switched on, they have never seen anything in the real world.

Now that looks set to change. Just as you might run a Google image search to see what a Buddha's hand citron looks like, so robots, and computer programs, are starting to take advantage of the wealth of images posted online to find out about everyday objects.

When presented with a new word, instead of using the limited index it has been programmed with, which is the conventional method, this new breed of automatons goes online and enters the word into Google. The robot or software uses the resulting range of images to recognise the object in the real world.

"If it works, it's huge. We'll have a robot that understands what it's looking at," says Paul Rybski of Carnegie Mellon University in Pittsburgh, Pennsylvania.

It's easy to program a robot to recognise specific objects. For instance, when given a picture of a chair, a robot can usually find that same chair in the real world using the colours, textures and angles of the image.

This system breaks down, though, when you ask a robot to identify a chair it has never seen before. The great variation in size, colour, shape and sometimes number of legs a chair might have makes it unlikely that the robot can identify any chair based on just the images it has been shown. And the difference between, say,

a small table and a chair can be very slight, making things even more confusing.

Humans don't have this problem because they tend to lump visual information into groups based not on how many legs a chair has, or its exact dimensions, but whether or not we recognise it as being built to sit on. Such an idea is extremely difficult for robots to grasp, says Per-Erik Forssen of the University of British Columbia in Vancouver, Canada, because they don't interact with the world in the same way. To get round this, researchers are focusing on building software that extracts images from the web.

This ability could allow robots to retrieve household objects for visually impaired people or those who have trouble walking. It could also mean robots become capable of teaching themselves about the world. "If you give a robot visual capabilities, it could pretty much do anything," says Alap Karapurkar of the University of Maryland in College Park. "You could tell a robot, 'car', and it could learn what a car looks like, that they're used for driving, then it could download a driver's manual, hop in the car and drive away," suggests Rybski.

To test the idea, last month Rybski, together with colleague

Alexei Efros, organised the first Semantic Robot Vision Challenge at the annual conference of the American Association for Artificial Intelligence in Vancouver. Four teams took part, entering one robot each.

The robots were given a list of 20 objects, including a DVD, a CD case, a banana and a calculator, that would be strewn across tables and chairs in a 6-metre-square area. The robots were allowed one hour to search the internet for images that were relevant to the words on the list and to analyse them. After that, they had to set out in search of the items.

Telltale signs

The first stage of the challenge involved turning the hundreds of images that you get in response to a Google search for, say, "red bell pepper" into a description that could be used to recognise that object in the real world.

To do this, the teams equipped their robots with software that analyses the shading patterns in all of the images brought up by the search and picks out telltale features within them. A large proportion of the images will be of red bell peppers, but there will plenty of others of, for example, dishes containing them or pictures of cookbooks. Assuming that the largest single group of images will be of peppers themselves, the software takes these images as the standard to form a kind of fingerprint. It then compares all the images with this standard. Those that have very different shading and won't help the robot understand what the object looks like are discarded and those that are similar are kept.

Once armed with this knowledge of what their target objects look like, the robots then struck out into the real world. Both the robot built by Forssen's team, Curious George, and a robot built by Karapurkar and other researchers from the University of Maryland, used stereo cameras to identify shapes that stood out from the tables or floor where

PICTURE THIS

Along with helping robots to see, the wealth of labelled images available online is turning computer programs into automatic illustrators.

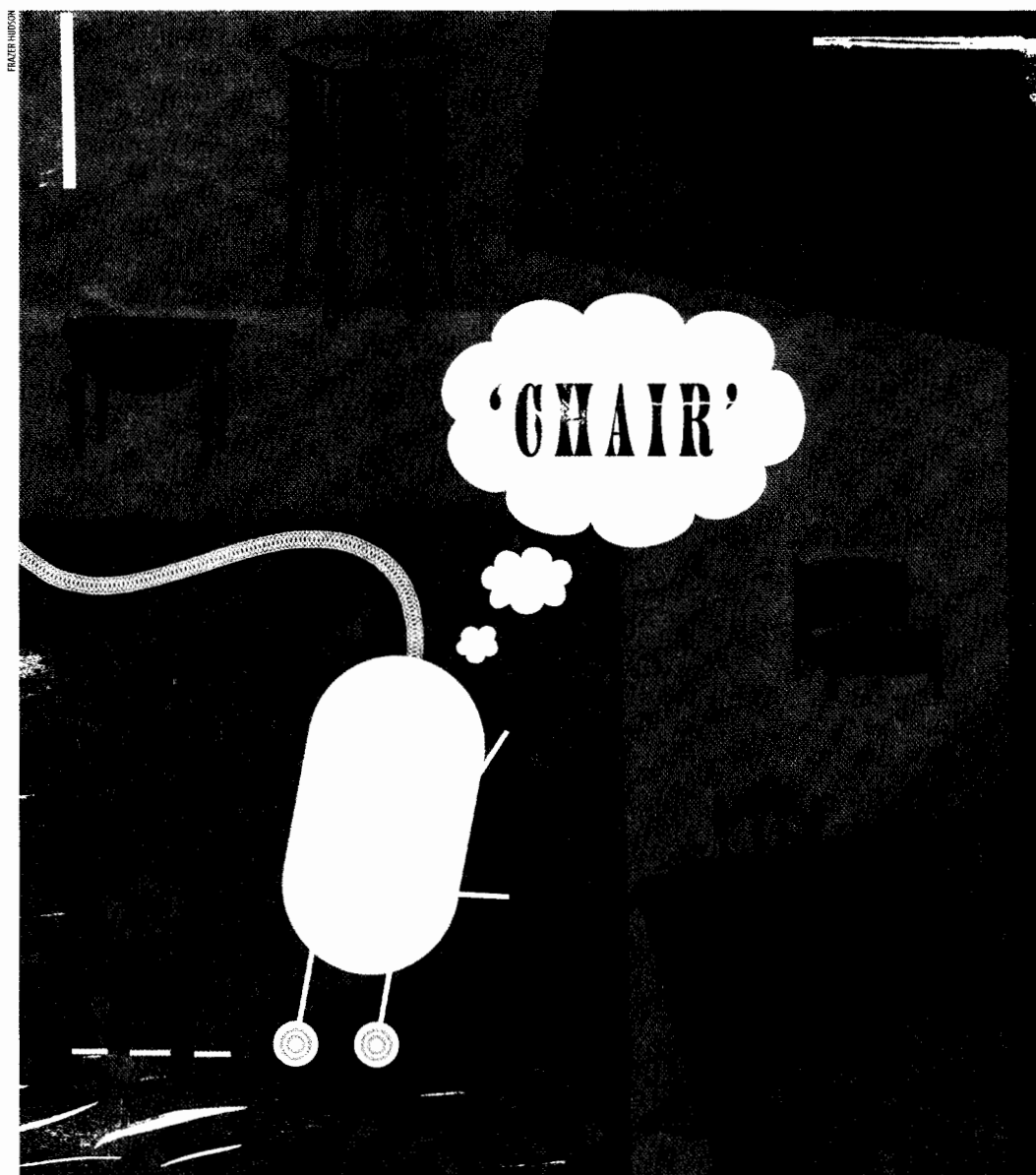
Xiaojin Zhu of the University of Wisconsin-Madison and colleagues have built a program that can read news articles and children's books and then illustrate them with relevant images it has found online. A sentence typically gets represented by a group of images, each one illustrating a word or phrase.

Programs already exist that can convert text into images, but they only work on predefined words and use a limited database of graphics. In contrast, Zhu's bot reads and generates pictures from any text and, because it

retrieves its images from the web, displays a wider variety of illustrations.

To make sure it illustrates the appropriate parts of the story, the program only chooses words that are "picturable", that is, they have a high number of results in a Google image search. To pick the best image possible, Zhu's program looks at the first 20 results thrown up by Google and clusters them based on colour content, shading and pixel patterns. It then picks the image that has the most in common with the others, assuming this is a good representation of the text.

Zhu says the program will help children and people with learning disabilities to learn to read.



they were strewn. The robots then snapped pictures of these objects and compared them to their index of fingerprints. If they discovered a match, that object was declared "found".

Curious George ended up winning, by identifying seven of the 20 objects, including distinguishing between a red bell pepper and a red plastic cup,

which had been deliberately added to cause confusion. This was because it was fitted with an additional, 7 megapixel camera that allowed it to examine objects in detail. In second place was the University of Maryland robot, which identified three things. The other two robots failed to find any of the objects.

Although it is likely to be at

least a few years before these robots find their way into the home as domestic assistants, the type of software they run on has a more imminent application: improving web image searches. "Commercially, we have very unsatisfying image search results," says Fei-Fei Li at Princeton University. For example, a Google image search of "banana" returns mostly pictures of the familiar fruit, but within the top 20 results it also returns a man sitting on a banana-

shaped chair and banana spiders on their web.

This is because the search engine uses the text attached to the pictures to decide which ones are relevant. Programs that understand what objects actually look like, based on analysing characteristics of the images they find online, could better decide whether they were relevant to a particular word. Researchers say this would likely result in more relevant search results.

The Semantic Robot Vision challenge also included a software-only league, allowing programs designed to test this idea to compete as well. Researchers sent a physical robot out to run around the course and video it. They then fed copies of the video to various programs running on stationary computers, which had to find the list of 20 objects using that footage.

The winner was a program called "object picture collection via incremental model learning", or OPTIMOL. Designed by Li, OPTIMOL requires a human to enter about five "seed" images for each word that it must recognise. The program then does a Google search of the word and compares each image that comes back to the seed images. It keeps back the five images that provide the best match and discards the rest, then searches again. The list of seeds grows, increasing the variety of examples the program has for each word but ensuring that images that don't represent that object do not get kept.

Jitendra Malik, a computer vision specialist at the University of California, Berkeley, says that the major search engines may soon use programs like OPTIMOL to do image searches, making the results they serve up far more relevant. "You can bet if they find something that improves search accuracy, it will show up in their next release," he says. He won't give details, but says that several start-up companies are already working to build search engines based on this type of image-searching algorithm. ●

"Humans recognise chairs as being built to sit on, something that is hard for robots to grasp"