

Developing artificial intelligence systems that can interpret images

December 12 2011, by Helen Knight



Antonio Torralba. Photo: M. Scott Brauer

Like many kids, Antonio Torralba began playing around with computers when he was 13 years old. Unlike many of his friends, though, he was not playing video games, but writing his own artificial intelligence (AI) programs.

Growing up on the island of Majorca, off the coast of Spain, Torralba spent his teenage years designing simple algorithms to recognize handwritten numbers, or to spot the verb and noun in a sentence. But he was perhaps most proud of a program that could show people how the night sky would look from a particular direction. "Or you could move to another planet, and it would tell you how the stars would look from there," he says.



Today, Torralba is a tenured associate professor of electrical engineering and computer science at MIT, and an affiliate of the Computer Science and <u>Artificial Intelligence</u> Laboratory (CSAIL), where he develops AI systems that can interpret images to understand what scenes and objects they contain.

Torralba first became interested in computer vision while working on his PhD at the University of Grenoble in France. "Vision is a really important area," Torralba says. It is also an extremely challenging one, requiring a great deal of computing power. "Around 30 percent of the brain is devoted to or connected to vision," he says.

At that time, most computer-vision researchers were occupied with facial detection and recognition, treating the rest of an image almost as a nuisance. But Torralba was far more interested in recognizing and understanding the other objects within an image. "I wanted to build systems that could put objects into context, to try to understand how different objects relate to each other," he says.

So he began developing systems that used information gathered from the entire image to help identify individual objects. If an image contains an <u>object</u> perched on top of a table, for example, that object is unlikely to be anything very large, narrowing down considerably the number of things it could possibly be.

Ultimately, such systems could be used to annotate all images shown online, making them more easily searchable. They could also allow robots to recognize where they are in a house or office building, based on what furniture and objects they see around them, he says.

Torralba is also attempting to develop systems that can scan a short video clip and predict what is likely to happen next, based on what people or objects are in the scene. To do this, the systems will need to understand



what actions each object or person in the scene is capable of making, and what their limitations are. This will allow the systems to make predictions about what each of these entities is likely to do in the near future.

If AI systems can learn how to predict what will happen next in this way, given all the available information about a particular situation, it should help them anticipate how their actions will influence future events, just as humans can, he says.



Image courtesy of Antonio Torralba

When Torralba is not at CSAIL, he spends his free time producing his own digital artwork by superimposing multiple images together. For one particular image (at right), Torralba took 150 photographs, each of which contained a person in the center of the image, and combined them. The result is a digital image that looks like it was drawn by hand, he says: "The superimposition of all these images gives this quality that looks as if it were produced with a pencil, but these are digital photographs."



Torralba used a similar approach to create a "visual dictionary," which consists of an image made up of thousands of individual pictures, to illustrate his group's work. Each picture represents one of the approximately 50,000 words in the English language that correspond to a visual concept. Clicking on any spot on the image brings up the particular word represented by those individual pictures, and its definition.

While the website is an interesting artistic project in its own right, it also serves a practical purpose, acting as a database of all the images Torralba and his team have collected on which to train their computer-vision systems. "The goal is to develop a system that is able to recognize all those 50,000 objects," he says.

It is a goal that is likely to keep Torralba busy for some time to come.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Developing artificial intelligence systems that can interpret images (2011, December 12) retrieved 10 May 2024 from https://phys.org/news/2011-12-artificial-intelligence-images.html

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.