Teaching a computer to perceive the world without human input

September 23 2013, by Marlene Cimons

Humans can see an object—a chair, for example—and understand what they are seeing, even when something about it changes, such as its position. A computer, on the other hand, can't do that. It can learn to recognize a chair, but can't necessarily identify a different chair, or even the same chair if its angle changes.

"If I show a kid a chair, he will know it's a chair, and if I show him a different chair, he can still figure out that it's a chair," says Ming-Hsuan Yang, an assistant professor of electrical engineering and computer science at the University of California, Merced. "If I change the angle of the chair 45 degrees, the appearance will be different, but the kid will still be able to recognize it. But teaching a computer to see things is very difficult. They are very good at processing numbers, but not good at generalizing things."

Yang's goal is to change this. He is developing computer algorithms that he hopes will give computers, using a single camera, the ability to detect, track and recognize objects, including scenarios where the items drift, disappear, reappear or when other objects obscure them. The goal is to simulate human cognition without human input.

Most humans effortlessly can locate moving objects in a wide range of environments, since they are continually gathering information about the things they see, but it is a challenge for computers. Yang hopes the algorithms he's developing will enable computers to do the same thing, that is, continually amass information about the objects they are

tracking.

"While it is not possible to enumerate all possible appearance variation of objects, it is possible to teach computers to interpolate from a wide range of training samples, thereby enabling machines to perceive the world," he says.

Currently, "for a computer, an image is composed of a long string of numbers," Yang says. "If the chair moves, the numbers for those two images will be very different. What we want to do is generalize all the examples from a large amount of data, so the computer will still be able to recognize it, even when it changes. How do we know when we have enough data? We cannot encompass all the possibilities, so we are trying to define 'chair' in terms of its functionalities."

Potentially, computers that can "see" and track moving objects could improve assistive technology for the visually impaired, and also could have applications in medicine, such as locating and following cells; in tracking insect and animal motion; in traffic modeling for "smart" buildings, and improved navigation and surveillance in robots.

"For the visually impaired, the most important things are depth and obstacles," Yang says. "This could help them see the world around them. They don't need to see very far away, just to see whether there are obstacles near them, two or three feet away. The computer program, for example, could be in a cane. The camera would be able to create a 3-D world and give them feedback. The computer can tell them that the surface is uneven, so they will know, or sense a human or a car in front of them."

Yang is conducting his research under a National Science Foundation Faculty Early Career Development (CAREER) award, which he received in 2012. The award supports junior faculty who exemplify the role of

teacher-scholars through outstanding research, excellent education and the integration of education, and research within the context of the mission of their organization. He is receiving $473,797 over five years.

Yang's project also includes developing a code library of tracking algorithms and a large data set, which will become publicly available. The grant also provides for an educational component that will involve both undergraduate and graduate students, with an emphasis on encouraging underrepresented minority groups from California's Central Valley to study computer sciences and related fields. The goal is to integrate computer vision material in undergraduate courses so that students will want to continue studying in the field.

Additionally, Yang is helping several undergraduate students design vision applications for mobile phones, and trying to write programs that will enable computers to infer depth and distance, as well as to interpret the images it "sees."

"It is not clear exactly how human vision works, but one way to explain visual perception of depth is based on people's two eyes and trigonometry," he says. "By figuring out the geometry of the points, we can figure out depth. We do it all the time, without thinking. But for computers, it's still very difficult to do that.

"The Holy Grail of computer vision is to tell a story using an image or video, and have the computer understand on some level what it is seeing," he adds. "If you give an image to a kid, and ask the kid to tell a story, the kid can do it. But if you ask a computer program to do it, now it can only do a few primitive things. A kid already has the cognitive knowledge to tell a story based on the image, but the computer just sees things as is, but doesn't have any background information. We hope to give the computer some interpretation, but we aren't there yet."