Researchers teach computers to perceive three dimensions in 2-D images

June 13 2006

This composite image shows a photograph and three 3-D reconstructions derived from it.

We live in a three-dimensional world but, for the most part, we see it in two dimensions. Discerning how objects and surfaces are juxtaposed in an image is second nature for people, but it's something that has long flummoxed computer vision systems.

Now, however, researchers in Carnegie Mellon University's School of Computer Science have found a way to help computers understand the

geometric context of outdoor scenes and thus better comprehend what they see. The discovery promises to revive an area of computer vision research all but abandoned two decades ago because it seemed insoluble. It may ultimately find application in vision systems used to guide robotic vehicles, monitor security cameras and archive photos.

Using machine learning techniques, Robotics Institute researchers Alexei Efros and Martial Hebert, along with graduate student Derek Hoiem, have taught computers how to spot the visual cues that differentiate between vertical surfaces and horizontal surfaces in photographs of outdoor scenes. They've even developed a program that allows the computer to automatically generate 3-D reconstructions of scenes based on a single image.

"The technique provides an approximate sense of the scene, a qualitative grasp of the structure of a scene," said Efros, assistant professor of computer science and robotics.

In their latest work, to be presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 17–22 in New York City, the Carnegie Mellon researchers will show that having a sense of 3-D geometry helps computers identify objects, such as cars and pedestrians, in street scenes.

Identifying vertical and horizontal surfaces and the orientation of those surfaces provides much of the information necessary for understanding the geometric context of an entire scene. Only about three percent of surfaces in a typical photo are at an angle, they have found.

Using 300 images gleaned from a Google search, Hoiem showed the computer numerous examples of vertical and horizontal surfaces, allowing a machine learning program to develop statistical associations between certain shapes, shadings and other characteristics typical of

each orientation.

The program also takes advantage of the constraints of the real world -- skies are blue, horizons are horizontal and most objects sit on the ground.

"In our world," noted Hebert, a professor of robotics, "things don't just float."

To demonstrate the utility of this technique, the researchers have designed a graphics program to automatically generate 3-D reconstructions by "cutting and folding" along vertical and horizontal lines in an image.

"It's like a children's pop-up book," Efros said.

"The amazing thing they did was show that it was actually possible," said computer vision pioneer Takeo Kanade, the U.A. and Helen Whitaker University Professor of computer science and robotics at Carnegie Mellon. "I would say it's a breakthrough."

A Longstanding Problem

Inability to understand the geometric context of a scene has limited the ability of computers to recognize objects. Though researchers have had some success at identifying objects, such as faces or cars, the lack of context results in preposterous mistakes, such as faces seen in clouds, or cars perched in treetops.

Scientists have struggled since early times to understand how people visually perceive three dimensions. Ancient Greeks reasoned that the eyes must emit rays that bounce off objects, measuring distances much like today's laser rangefinders. By the 19th century, scientists realized

that a pair of eyes gives humans binocular vision, allowing them to perceive depth. But stereoscopic vision is useful at distances of no more than 50 meters. Even then, the mind often overrides binocular vision, such as when watching a football game on television.

Vision was an early problem that artificial intelligence researchers tried to tackle and "context-based" outdoor scene analysis was a favorite subject during the 1970s.

Researchers found they could describe the geometry of an object, such as a chair, but matching the description with actual pixels proved a herculean task. Statistical learning tools were limited then and research computers were about 100 times less powerful than a typical laptop today. By 1980, most had concluded that the feat was either impossible or, if possible, computationally impractical.

An Unexpected Advance

Even when Efros and Hebert assigned Hoiem to use machine learning techniques to teach visual context to a computer two years ago, they regarded it primarily as a learning exercise for their student. "We didn't believe it would work," Efros said.

To their surprise, Hoiem found the computer often discerned which surfaces were vertical or horizontal, and whether a vertical surface faced left, right or toward the viewer. Based on the examples it was shown, the computer identified each feature in an image and assigned to it a probability that it had a horizontal or vertical orientation.

In their latest work, the researchers have used the geometric context information to improve the ability of computer programs to recognize objects within the scene. And improved object recognition, they note, should ultimately provide feedback to further improve understanding of

the geometric context.

"If you can find a car," Hebert explained, "you know it is on a flat surface."

Source: Carnegie Mellon University