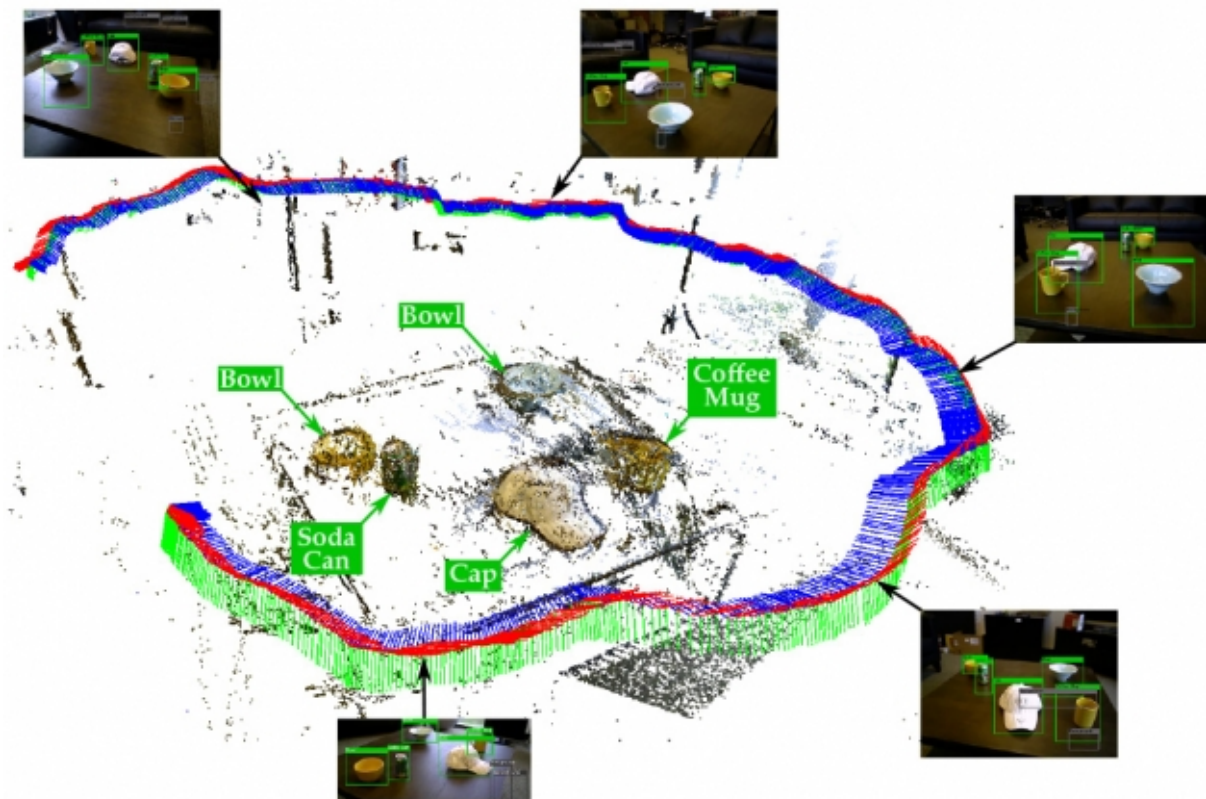# Robots' maps of their environments can make existing object-recognition algorithms more accurate

July 24 2015, by Larry Hardesty

The proposed SLAM-aware object recognition system is able to localize and recognize several objects in the scene, aggregating detection evidence across multiple views. The annotations are actual predictions proposed by the system.

John Leonard's group in the MIT Department of Mechanical Engineering specializes in SLAM, or simultaneous localization and mapping, the technique whereby mobile autonomous robots map their environments and determine their locations.

Last week, at the Robotics Science and Systems conference, members of Leonard's group presented a new paper demonstrating how SLAM can be used to improve object-recognition systems, which will be a vital component of future robots that have to manipulate the objects around them in arbitrary ways.

The system uses SLAM information to augment existing object-recognition algorithms. Its performance should thus continue to improve as computer-vision researchers develop better recognition software, and roboticists develop better SLAM software.

"Considering object recognition as a black box, and considering SLAM as a black box, how do you integrate them in a nice manner?" asks Sudeep Pillai, a graduate student in computer science and engineering and first author on the new paper. "How do you incorporate probabilities from each viewpoint over time? That's really what we wanted to achieve."

Despite working with existing SLAM and object-recognition algorithms, however, and despite using only the output of an ordinary video camera, the system's performance is already comparable to that of special-purpose robotic object-recognition systems that factor in depth measurements as well as visual information.

And of course, because the system can fuse information captured from different camera angles, it fares much better than object-recognition systems trying to identify objects in still images.

Drawing boundaries

Before hazarding a guess about which objects an image contains, Pillai says, newer object-recognition systems first try to identify the boundaries between objects. On the basis of a preliminary analysis of color transitions, they'll divide an image into rectangular regions that probably contain objects of some sort. Then they'll run a recognition algorithm on just the pixels inside each rectangle.

To get a good result, a classical object-recognition system may have to redraw those rectangles thousands of times. From some perspectives, for instance, two objects standing next to each other might look like one, particularly if they're similarly colored. The system would have to test the hypothesis that lumps them together, as well as hypotheses that treat them as separate.

Because a SLAM map is three-dimensional, however, it does a better job of distinguishing objects that are near each other than single-perspective analysis can. The system devised by Pillai and Leonard, a professor of mechanical and ocean engineering, uses the SLAM map to guide the segmentation of images captured by its camera before feeding them to the object-recognition algorithm. It thus wastes less time on spurious hypotheses.

More important, the SLAM data let the system correlate the segmentation of images captured from different perspectives. Analyzing image segments that likely depict the same objects from different angles improves the system's performance.

Picture perfect

Using machine learning, other researchers have built object-recognition systems that act directly on detailed 3-D SLAM maps built from data

captured by cameras, such as the Microsoft Kinect, that also make depth measurements. But unlike those systems, Pillai and Leonard's system can exploit the vast body of research on object recognizers trained on single-perspective images captured by standard cameras.

Moreover, the performance of Pillai and Leonard's system is already comparable to that of the systems that use depth information. And it's much more reliable outdoors, where depth sensors like the Kinect's, which depend on infrared light, are virtually useless.

Pillai and Leonard's new paper describes how SLAM can help improve object detection, but in ongoing work, Pillai is investigating whether object detection can similarly aid SLAM. One of the central challenges in SLAM is what roboticists call "loop closure." As a robot builds a map of its environment, it may find itself somewhere it's already been—entering a room, say, from a different door. The robot needs to be able to recognize previously visited locations, so that it can fuse mapping data acquired from different perspectives.

Object recognition could help with that problem. If a robot enters a room to find a conference table with a laptop, a coffee mug, and a notebook at one end of it, it could infer that it's the same conference room where it previously identified a laptop, a coffee mug, and a notebook in close proximity.

"The ability to detect objects is extremely important for robots that should perform useful tasks in everyday environments," says Dieter Fox, a professor of computer science and engineering at the University of Washington. "This work shows very promising results on how a robot can combine information observed from multiple viewpoints to achieve efficient and robust detection of objects."

  More information: "Monocular SLAM Supported Object

Recognition." [people.csail.mit.edu/spillai/p … ion/pillai_rss15.pdf](http://people.csail.mit.edu/spillai/p)

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