Teaching machines to see: New smartphone-based system could accelerate development of driverless cars

December 20 2015

An example of SegNet in action: the separate components of the road scene are all labelled in real time. Credit: Alex Kendall

Two newly-developed systems for driverless cars can identify a user's location and orientation in places where GPS does not function, and identify the various components of a road scene in real time on a regular camera or smartphone, performing the same job as sensors costing tens of thousands of pounds.

The separate but complementary systems have been designed by researchers from the University of Cambridge and demonstrations are freely available online. Although the systems cannot currently control a driverless car, the ability to make a machine 'see' and accurately identify where it is and what it's looking at is a vital part of developing autonomous vehicles and robotics.

The first system, called SegNet, can take an image of a street scene it hasn't seen before and classify it, sorting objects into 12 different categories—such as roads, street signs, pedestrians, buildings and cyclists - in real time. It can deal with light, shadow and night-time environments, and currently labels more than 90% of pixels correctly. Previous systems using expensive laser or radar based sensors have not been able to reach this level of accuracy while operating in real time.

Users can visit the SegNet website and upload an image or search for any city or town in the world, and the system will label all the components of the road scene. The system has been successfully tested on both city roads and motorways.

For the driverless cars currently in development, radar and base sensors are expensive - in fact, they often cost more than the car itself. In contrast with expensive sensors, which recognise objects through a mixture of radar and LIDAR (a remote sensing technology), SegNet learns by example—it was 'trained' by an industrious group of Cambridge undergraduate students, who manually labelled every pixel in each of 5000 images, with each image taking about 30 minutes to

complete. Once the labelling was finished, the researchers then took two days to 'train' the system before it was put into action.

"It's remarkably good at recognising things in an image, because it's had so much practice," said Alex Kendall, a PhD student in the Department of Engineering. "However, there are a million knobs that we can turn to fine-tune the system so that it keeps getting better."

SegNet was primarily trained in highway and urban environments, so it still has some learning to do for rural, snowy or desert environments—although it has performed well in initial tests for these environments.

The system is not yet at the point where it can be used to control a car or truck, but it could be used as a warning system, similar to the anti-collision technologies currently available on some passenger cars.

"Vision is our most powerful sense and driverless cars will also need to see," said Professor Roberto Cipolla, who led the research. "But teaching a machine to see is far more difficult than it sounds."

As children, we learn to recognise objects through example—if we're shown a toy car several times, we learn to recognise both that specific car and other similar cars as the same type of object. But with a machine, it's not as simple as showing it a single car and then having it be able to recognise all different types of cars. Machines today learn under supervision: sometimes through thousands of labelled examples.

There are three key technological questions that must be answered to design autonomous vehicles: where am I, what's around me and what do I do next. SegNet addresses the second question, while a separate but complementary system answers the first by using images to determine both precise location and orientation.

The localisation system designed by Kendall and Cipolla runs on a similar architecture to SegNet, and is able to localise a user and determine their orientation from a single colour image in a busy urban scene. The system is far more accurate than GPS and works in places where GPS does not, such as indoors, in tunnels, or in cities where a reliable GPS signal is not available.

It has been tested along a kilometre-long stretch of King's Parade in central Cambridge, and it is able to determine both location and orientation within a few metres and a few degrees, which is far more accurate than GPS—a vital consideration for driverless cars. Users can try out the system for themselves here.

The localisation system uses the geometry of a scene to learn its precise location, and is able to determine, for example, whether it is looking at the east or west side of a building, even if the two sides appear identical.

"Work in the field of artificial intelligence and robotics has really taken off in the past few years," said Kendall. "But what's cool about our group is that we've developed technology that uses deep learning to determine where you are and what's around you - this is the first time this has been done using deep learning."

"In the short term, we're more likely to see this sort of system on a domestic robot - such as a robotic vacuum cleaner, for instance," said Cipolla. "It will take time before drivers can fully trust an autonomous car, but the more effective and accurate we can make these technologies, the closer we are to the widespread adoption of driverless cars and other types of autonomous robotics."

The researchers are presenting details of the two technologies at the International Conference on Computer Vision in Santiago, Chile.