Popular Science Description

Sight is one of those things that most of us take for granted. Some are of course born blind, others need corrective contact lenses or glasses, and still others need no help at all. The human brain is an amazing thing, with just two eyes it can pinpoint distances, detect animals, colours, orientation, adjust for differences in light, and intuitively understand shadows. This is something extraordinary, and something which is harder than it might seem to those of us born with this machinery in our brain. Computers, on the other hand, have none of these advantages.

Computers need to be taught how to see, a task which fittingly enough is called Computer Vision. This is everything from being able to notice things in pictures, being able to detect a ball in a family photo, to being able to recognise that same ball in a different photo, all the way to being able to create a model of said ball. We might call this first task a form of image analysis, or image segmentation, meaning we segment or divide the image into different categories. The simplest case being "ball" and "not ball".

This final task, creating a model of the ball, is a bit harder than just detecting it. We, as humans, might make some educated guess about the size of this ball relative to everything else in the image, the person holding it, the material the ball is made from (is it filled with air, a football, a golf ball?). We can do this since we have been trained all our life in the great art of estimating things from a distance. Computers can't do this. We have to train them.

The idea behind this training is relatively simple. We take a lot of examples, we write down a lot of the facts we want the computer to learn (what is the ball in this picture, what dimensions does it have, what is its type) and feed this to the computer in an organised fashion. Often, the structure of this learning is based of neural networks, a form of algorithm taking its inspiration from how the neurons that make up our brain learn information. Unlike many other methods which can be very rigorously checked using mathematical theory, for neural networks it's significantly harder to check the results every step of the way. One might think of it closer to trying to give the network an intuition of what it's results should be, rather than it to following a recipe with clear instructions, the same way you don't consciously think of how you know which object is closer to you in a picture.

Now there are many different methods of teaching a computer how to detect and recreate objects. Some need specialised lasers in order to physically measure distances (LIDAR), others require that we take a great number of pictures with the same camera and compare how an object looks from different perspectives (Structure from Motion), and some methods ambitiously try and train neural networks to make educated guesses from just a single picture. Of these methods the first 2 produce good results, but the equipment, time, and amount of pictures can make them difficult to use in many circumstances. The third, on the other hand, if it produced good results, would be both fast, cheap, and easy to use. This is what we want to explore in this report.

Specifically, there is a neural network called PlaneRCNN which estimates flat surfaces in an image. As an example it might mark all the walls in a room, the shelves on a bookshelf, the seat on a chair. It then gives all the details about these flat surfaces (or planes as they're often called in mathematics) to the user.

What we aim to do in this report is to look at how well PlaneRCNN estimates planes in different types of images. Does it perform better outside or inside, is it better in messy rooms, or does it perform better in minimalistic environments with clear boundaries between objects. We do this primarily by using a Structure from Motion system to create really good "correct" planes, and working out ways to compare these correct planes to the planes created by PlaneRCNN. Hopefully this can give us a greater understanding of how to work with and better use PlaneRCNN for further research.