791DD: Homework 1

John Sweeney

September 12, 2005

1 Can you do recognition with no training data?

You can find correlations between unlabeled examples, aka clustering, aka unsupervised learning. Given enough examples, there should be meaningful correlations in the data that you can extract and use to classify the example, based on the idea that the data you are trying to recognize will produce some measurable effect in the example data. If you wanted to recognize cars, you wouldn't have labeled examples, but just lots of pictures (hopefully mostly of cars, but not required), and a good clustering algorithm, and most likely some sort of feature detector that would be useful for cars. You'd run your feature detector over each image, get a set of features, and do this for all the images, and then use your clustering algorithm to group together all those images whose feature sets were similar. Now you have groups of images that hopefully correspond to different types of cars, and you didn't have to label one training image.

2 Depth perception with one eye?

You can do this by moving the camera in a known trajectory, for example, from side to side, and looking at the motion of the pixels. Objects that are farther away will move faster across the image plane than those closer to the camera. If you were moving the camera horizontally, you would find pixel correspondences between frames, measure their distance in the image space, and there should be a linear function mapping that distance to the depth of that position of the object.

3 What can be done with one pixel?

We assume this pixel can recognize multiple intensities.

If you don't have actuation, you can record the value of the pixel over time and find patterns in the intensity correlated to time. So you might know if it was day or night or something like that.

If you have actuation however, you can do a lot more. Your actuated pixel can become a very low resolution 2D imager. You can record the light gradient as you scan the pixel in space. Assuming you have proprioception (so that you can consistently scan the same spatial location), these maps would somewhat accurate, limited by your scan rate. You can then start to move towards (or away) from the light sources, all the while recording the intensities. You could probably infer the location of light occluding obstacles by noticing sharp "edges" in your light map which correspond to shadows.

The Brandenburg bug vehicles can be thought of as robots with a single pixel sensor; they had a number of behaviors, such as moving towards light, moving away from it, etc.

4 All the ways a baby could recognize its mother.

A baby knows its mothers voice; even in the womb it will be able to hear its mother.

A baby can recognize a particular smell that it associates with its mother (maybe from her preferred shampoo or soap, or maybe from pheremones, etc.).

I imagine there is a sort of primitive "imprinting" process that goes on with a baby and the facial recognition of its mother. (By primitive I mean not learned.) The baby cannot recognize many different faces, but its mother is one that it can.

I don't know much about raising babies, but I will assume that in the case where the baby is breast-fed, the mother is the only one doing the breast-feeding. The baby can infer it is the mother who is feeding it.

A baby could also recognize the particular way in which it is rocked by a person. I imagine each individual has a different way of rocking a baby when they hold it, that combines rocking as well as making noises to it, that are pretty particular. The baby would be able to associate the different way it's being held with a different person, and combined with some of the other ways its can recognize its mom, it will known when its mother has picked it up. For example, the baby was sleeping and wakes up without seeing who picked it up.

5 How many senses do we really have?

Research of the existence of mirror neurons has convinced me that we have a number of "senses" that filter observed behavior. We have a sense for other people's intentions, we have a sense for emotion recognition, and we are very adept at recognizing biologically-generated motion. We are so good at understanding others actions that we even say it's a language: body language. We also have mirror neurons for audition; we hear sounds and try and classify their source.

We also have a visual sense to detect quick movements against a constant background, although this is a lower-level ability.

As noted, we can detect changes in temperature and the motion of the wind. Similarly, we can detect the level of humidity, usually by noting how sweaty we feel. We have a modality for recognizing the viscosity of liquids and the frictional properties of surfaces; related is the ability to judge the hardness of a surface.

We have various proprioceptive sensors for determining the position of our limbs and the contraction of our muscles.

6 How many frames of video needed for speed of object?

It depends on the frame rate (and the speed of the object).Given a constant frame rate, then only a certain range of speeds of objects will cause corresponding pixels to move in image space. If the object is moving slowly with respect to the frame rate, then you will need more than two frames; you will need enough frames to detect a one pixel movement. If the object is moving very fast with respect to the frame rate, then aliasing will occur, and you will get a blurry image. This would actually allow you to infer the speed based on one image, if you measure the length of a blurred pixel, it should correspond to the object's speed.