Ellen Tucker
EECS 349 - Status Report

My task is to classify images based on the number of discrete objects it contains. It is inspired by research on the development of the natural number concept in infants, which is tested by presenting groups of objects to infants and determining if they can distinguish between groups of different sizes (cardinalities). My inputs are images that I have generated, each of which contains a grouping of between 1 and 4 objects (as per the infant experiments). I should note that the spatial configuration is randomized, as well as the scale of each object to control for sheer amount of pixels in a scene. For instance, an image of two squares may include two large squares and an image of four squares may contain one large and three smaller squares, and so forth.

My task is interesting both from the perception and developmental areas of psychology: infants as young as 3 months are able to exhibit quantitative information about sets of up to 3 objects - for instance, they are able to distinguish between sets of 1, 2 and 3 items; however, they are unable to distinguish between 4 and 5 items. The studies suggest that this is due to working memory constraints in the infants, and that they are only to attend to up to 3 individuals (for adults, the number is typically 4, and in my project, I'm going to experiment with groups of size 4). This is interesting in that I am able to experiment with different representations of these sets based on the bitmaps I've generated: one such representation is simply the flattened, normalized bitmap vector, another representation I'm working with is contour information, and a third is representing qualitative spatial relationships between the centers of mass of the objects.For example, infants can represent when objects are to the left, right, aligned, above, below, etc. of another object. I am very curious about how these visual relations may give rise to discrete quantitative concepts.

2) The dataset I'm working with so far has 3,000 bitmaps of each of 4 types {1, 2, 3, & 4 objects}. Each bitmap is 60x60 pixels. Additionally, for each of these bitmaps I have done various forms of preprocessing. For each bitmap, I also have a file containing the contour vectors (obtained using openCV library) as well as binary pairwise spatial relations for the center of mass for each object. For example, let c_1 be the center of mass for one object and c_2 the center of mass for a second object (from a set of two objects). In our bitmap, **c_1 is above and to the right of c_2**. Then we have a feature *(c_1 leftOf c_2)* for which we have a value of 0 (since c_1 is to the right of c_2). However, for the feature *(c_2 leftOf c_1)*, we would have a value of 1. Analogously, we would have *(c_1 above c_2) = 1, (c_2 above c_1) = 0.* And so on. The feature vector would contain the leftOf, above, relations pairwise for *c_1, c_2, c_3, c_4* and *horizontalAligned* and *verticalAligned* for each non-ordered pair. Note that I am using the working memory constraints for representing only up to 4 objects (in adults) and the zero values simply represent the case when a relation is not perceived. In other words, the vector for an image is

x = *<(c_1 leftOf c_2), (c_1 leftOf c_3), (c_1 leftOf c_4) (c_2 leftOf c_1), ...*
, (c_1 above c_2), (c_1 above c_3), ..., (c_1 horizontalAlign c_2), ..., (c_3 horizontalAlign c_4),
(c_1 verticalAlign c_2), (c_1 verticalAlign c_3), ..., (c_3 verticalAlign c_4)>.

*Note I am not encoding *rightOf* or *below*, as these can be deduced from the *leftOf* and *horizontalAlign* data respectively (for instance, if c_1 and c_2 are in the image, then *one and only one* of *(c_1 leftOf c_2)*, *(c_2 leftOf c_1)* or *(c_1 horizontalAlign c_2)* will be 1 and the others 0; additionally, since *horizontalAlign* and *verticalAlign* are symmetric relations, I do not encode this relation pairwise, which would be redundant (e.g. *(c_1 horizontalAlign c_2) = (c_2 horizontalAlign c_1)*). Note also that in some images c_3 and c_4 do not exist, thus all relations involving them will simply be 0 (e.g. in a set with only 2 objects, for any relation R, $(c_i \; R \; c_j) = 0$ for $j > 2$. What I am modeling with this representation is the cognitive *capacity* to represent the spatial relations between objects, and when this relation is encoded between two attended-to objects, the feature will take on a value of 1.

So far I have been putting my effort into constructing my dataset, extracting the various spatial and segmentation features from it (using openCV contours methods) to encode each example in the ways I've described. I am currently using the scikit neural network API to construct multilayer perceptron neural networks. I am using this because I wish to be able to use the same approach across my different representations (flattened normalized bitmap values vector), contours, and the binary pairwise spatial relations between objects.

My plans for the remainder of the quarter are to compare classification performance across the 3 representations I've described and across different ML techniques. My main concern is in comparing pixel-based classification with the qualitative spatial representation classification and the contour representation classification tasks across models. I think neural networks will be a very interesting common technique, especially since they seem to deal with missing information in a potentially more elegant way than other approaches (send in a zero). Since I am interested in the role of working memory's limits on potential encodings, I am using this to advantage with my spatial relation representations. I am interested also in how nearest neighbor and decision tree techniques will perform with the spatial relations representation and with the contours representations. The pros of my work at this point is that I believe I have very interesting alternative representations and I am now ready to experiment with different techniques and compare representations and have a discussion about how the representations relate both to perception and to the nature of the classification task, and also to the interesting philosophical question about numbers as an abstract concept and as a perceptual entity and "knowable" by preverbal infants.