Object recognition in the early visual system

Daesu Chung and Laurent Itti

University of Southern California, Computer Science Department

To complement our previous modeling of visual attention and the "where" processing stream in the primate brain, we now investigate the recognition of objects at the attended locations in the "what" visual processing stream. In the work we built an object recognition system of view-tuned units (16×16 pixels, 8-bit gray level scale). The view-tuned units are believed as a critical part for object recognition of human brain, which is found in the inferotemporal cortex. The view-tuned units are constructed using HMAX model that was proposed by M. Reisenhuber and T. Poggio (Hierarchical models of object recognition in cortex. *Nat Neurosci*, 2:1019-1025, 1999). HMAX was based on the hierarchical model of early visual system that uses some amount of abstraction from the incoming information to primary visual system. In our implementation of HMAX, the input image (128×128 pixels) is contracted into 16×16 pixels view-tuned units. As the result, the view-tuned units can be said not only to possess the information of the image, but also some degree of invariance to scale and translation. Here we tested with three types of polygons for the classification(or recognition) problem, i.e. ellipses, rectangles, and triangles with several variations of orientations, positions, shapes, sizes, and even occlusions. With two different approaches, we implemented the successful classification solutions for the view-tuned units. One method was a simple linear method, which is mentioned in the paper by Reisenhuber and Poggio. The point for this work was on the comparison of the simple linear method with a more sophisticated and expensive method, Support Vector Machine. We tested for the improvement in the performance of the trained classifiers. We used three different kernels for the SVM: dot product (linear), degree 2 polynomial, and Radial Basis Function.

At the results, both the linear method and the SVM-based methods used in this project constructed adequate classifiers for the classification of simple grayscale level images. In fact when we used expensive methods of classification such as SVM with linear and polynomial kernel, the model gave slightly worse results compared to the one with simple linear method. From the experiments, we conclude that the preprocessing (HMAX) already provide a clue for linear separation of the data, thus the improvements by more sophisticated methods are insignificant. Further experiments will extend this analysis to a large number of objects, such as the natural environments with complex situation, so the SVM classifier may prove more robust than the simple linear classifier.