Selective attention integrates sparse and population codes

Albert L. Rothenstein, John K. Tsotsos

York University, Toronto, Canada

How visual information is represented in the brain is the crucial question that drives most research in the field, and answering it will mean significant theoretical and practical advances in the understanding of the brain. Proponents of sparse representations suggest that highly selective, specialized neurons, introduced as cardinal cells by Barlow [1], explicitly code each percept. Energy minimization and information theoretic arguments have been used to justify relatively sparse representations. The main problem with this solution is the combinatorial explosion in the number of units needed to represent all the different possible stimuli. At the same time, one distinct advantage of sparse codes is their capacity to encode multiple stimuli in parallel. Others have suggested distributed/population codes [2]. While these representations have high encoding capacity, they suffer from the source separation problem, not allowing any significant degree of parallelism. Interestingly, there is significant experimental evidence for both representations in the primate brain, and the brain exhibits behavior consistent with both representations under different conditions.

One characteristic of vision is that not all stimuli in the visual field can and need to be represented at the same time to a high degree of accuracy [3]. This implies that some selection needs to be made, constrained by some requirements: important stimuli need to be detected as fast as possible, multiple important stimuli must be represented simultaneously, and non-important stimuli can wait, or can even be ignored. The fact that these requirements and their time course match those of visual attention strongly indicates that attention forms the glue between a sparse, fast, and parallel initial representation that supports object detection and a slow, serial, and detailed representations needed for full recognition. Taking cues from behavioral and physiological studies, we propose that the initial representation is sparse, corresponding to the categoric level, that can be represented sparsely due to the fact that the number of biologically relevant categories is very limited. The Selective Tuning (ST) model of (object based) visual attention [4] can be used to recover the spatial location and extent of the visual information that has contributed to a categoric decision. This allows for the selective detailed processing of this information at the expense of other stimuli present in the image. The feedback and selective processing create the detailed population code corresponding to the attended stimulus. We suggest and demonstrate a possible binding mechanism by which this is accomplished in the context of ST, and show how this solution can account for existing experimental results. We present a number of predictions of this model and suggest experiments to validate them.

References

[1] H. B. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1(4):371–394, 1972.

[2] A. Pouget, P. Dayan, and R. Zemel. Inference and computation with population codes. *Annual Review of Neuroscience*, 26:381–410, 2003.

[3] J. K. Tsotsos. Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13(3):423–444, 1990.

[4] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. H. Lai, N. Davis, and F. Nuflo. Modeling visual-attention via selective tuning. *Artif. Intell.*, 78(1-2):507–545, 1995.