Abstract

Selective visual attention provides an effective mechanism to serialize perception of complex scenes in both biological and machine vision systems. In extension of previous models of saliency-based visual attention by Koch & Ullman (*Human Neurobiology*, 4:219–227, 1985) and Itti et al. (*IEEE PAMI*, 20(11):1254–1259, 1998), we have developed a new model of bottom-up salient region selection, which estimates the approximate extent of attended proto-objects in a biologically realistic manner.

Based on our model, we simulate the deployment of spatial attention in a biologically realistic model of object recognition in the cortex and find, in agreement with electrophysiology in macaque monkeys, that modulation of neural activity by as little as 20 % suffices to enable successive detection of multiple objects.

We further show successful applications of the selective attention system to machine vision problems. We show that attentional grouping based on bottom-up processes enables successive learning and recognition of multiple objects in cluttered natural scenes. We also demonstrate that preselection of potential targets decreases the complexity of multiple target tracking in an application to detection and tracking of low-contrast marine animals in underwater video data.

A given task will affect visual perception through top-down attention processes. Frequently, a task implies attention to particular objects or object categories. Finding suitable features can be interpreted as an inversion of object detection. Where object detection entails mapping from a set of sufficiently complex features to an abstract object representation, finding features for top-down attention requires the reverse of this mapping. We demonstrate a computer simulation of this mechanism with the example of top-down attention to faces.

Deploying top-down attention to the visual hierarchy comes at a cost in reaction time in fast detection tasks. We use a task switching paradigm to compare task switches that do with those that do not require re-deployment of top-down attention and find a cost of 20–28 ms in reaction time for shifting attention from one stimulus attribute (image content) to another (color of frame).