Visual Attention and Object Recognition Algorithms for Fast Decision and Control

Visual attention and eye movements in primates have been widely shown to be guided by a combination of stimulus-dependent or 'bottom-up' cues, as well as task-dependent or 'top-down' cues. Understanding the mechanisms of attention can give rise to numerous applications in machine vision, interactive video games, human-computer interfaces, and assessment of human cognitive state and intentions under task pressure. Both the bottom-up and top-down aspects of attention and eye movements have been modeled computationally. Yet, it is not until recent work which I will describe that bottom-up models have been strictly put to the test, predicting significantly above chance the eye movement patterns, functional neuroimaging activation patterns, or most recently neural activity in the brains of monkeys inspecting complex dynamic scenes. In recent developments, models that increasingly attempt to capture top-down aspects have been proposed. In one system which I will describe, neuromorphic algorithms of bottom-up visual attention are employed to predict, in a task-independent manner, which elements in a video scene might more strongly attract attention and gaze. These bottom-up predictions have more recently been combined with top-down predictions, which allowed the system to learn from examples (recorded eye movements and actions of humans engaged in 3D video games, including flight combat, driving, first-person, or running a hot-dog stand that serves hungry customers) how to prioritize particular locations of interest given the task. Pushing deeper into real-time, joint online analysis of video and eye movements using neuromorphic models, we have recently been able to predict future gaze locations and intentions of future actions when a player is engaged in a task. Finally, employing deep neural networks, we show how neuroscience-inspired algorithms can also achieve state-of-the art results in the domain of object recognition, especially over a new dataset collected in our lab and comprising ~22M images of small objects filmed on a turntable, with available pose information that can be used to enhance training of the object recognition model.