

Ph.D. in Cognitive Science from the University of California at San Diego
Defense of the thesis *Active Perception* given on November 29, 2010.

Précis of *Active Perception*

by Nicholas John Butko, Ph.D.

Machine perception technologies have the potential to revolutionize the fields of health, individual and public safety, marketing, education, and art. But for all their potential, existing machine perception technologies are insufficient. The aim of *Active Perception* is to contribute to the scientific understanding of perceptual intelligence in humans, and its synthesis in machines. Attempts to synthesize computational intelligence have typically succeeded in inverse proportion to the perceived level of intelligence. For example, winning complex games is difficult for humans but relatively easy for computers; meanwhile, “seeing” is effortless for humans but is astoundingly difficult for computers. This disparity highlights a gap in our understanding of the fundamental nature of human intelligence.

Perception is an active process at many levels: neurons grow synapses and modulate their firing properties; attention sweeps its spotlight from place to place; eyes move; infants explore the effects that they cause; scientists design experiments to understand the world; groups of individuals gather and coordinate information to form a shared understanding. Although there are unique peculiarities at each of these levels, there are also basic commonalities.

Active Perception contains six related research studies that span multiple levels of cognitive science. This breadth is possible because a common computational framework focuses on the invariants of all levels.

Study (1) considers a population of active neurons that modify internal and synaptic parameters to maximize their capacity as information channels. We show that these model neurons exhibit firing rate patterns and receptive field patterns that have been previously reported in empirical studies, and solve non-linear independent components problems. The neural adaptation mechanism is simple and local in time and space, which are necessary for a plausible biological mechanism. This work suggests new algorithms for unsupervised visual feature extraction, a basic building block of visual object recognition systems.

Study (2) explores an information theoretic account for human visual attention, and proposes an efficient computational implementation of attentional systems. This enables deployment of a real-time algorithms. We tested these in a robot, and found that the robot’s attention was directed to regions of its surroundings that contained people even though it had no specific knowledge of the appearance of humans. Our system shares many algorithmic similarities to the popular SIFT key point detection algorithm, while operating at a rate that is about eighteen times faster. Thus it has the potential to significantly increase the speed and reduce the power requirements of many existing computer vision approaches.

Study (3) investigates contingency detection, a process hypothesized to be fundamental in the development of social interaction in infants. At two months, infants are slow to learn that

objects and people respond to them, but by ten months they have learned to actively probe objects for responses and quickly discover new contingencies. We model this as the process of learning to gather information in an efficient manner, *i.e.* learning to learn. We show that a control theoretic reinforcement learning model with information gain as a reinforcing signal accounts for observed infant developmental trajectories. Given the empirical statistics of social interaction, the optimal decision rule can be computed analytically. The behaviors of human infants and the optimal controller are virtually identical. In addition to explaining human development, the optimal controller allows us to build robots that actively probe their environments for contingencies in an effective manner. At the time of this study, it was unknown if information gain was a biologically viable reinforcing signal. Since this study, it was demonstrated in rhesus macaque monkeys that dopaminergic reward systems actively reward behaviors that lead to information gain.

Study (4) uses the same information maximizing framework from study (3) to formalize the problem of visual search. This task is difficult due to a huge branching factor, but we show that a modern reinforcement learning algorithm can use information as a reward signal to optimize parameters of a neural network based control policy. We use this control policy to construct a digital retina that detects faces in images faster than state of the art computer vision algorithms. Our model predicts changes in the optimal search strategy depending on the visual properties of the search target, *e.g.* level of contrast, as well as based on its movement dynamics. Our model, therefore, suggests hypotheses for future psychophysical studies.

Study (5) proposes a method in which a robot learns how to coordinate its eyes and head as it learns to search for objects. This requires the robot to learn about the physical parameters of its own oculomotor system. We show how an optimal observer should arbitrate different sources of uncertainty to discover how its body and senses are organized and relate to each other. We implement our approach on three different robots, and we show that all of them can quickly learn about their distinct sensorimotor capabilities. We propose that multiple gaze components, *e.g.* head and eyes, can be combined based on a principle of maximum accuracy in the presence of signal dependent noise. This principle accounts for saccadic undershooting, a phenomenon reliably observed in humans when the head is fixed and only the eyes are allowed to move; it further predicts that such undershoots are mitigated when the head is free to move as in natural behavior.

Study (6) demonstrates that infants may use contingencies to learn about the appearance of people. We built a robot resembling an infant that actively probed its environment for contingencies using the approach from study (3), and we invited subjects to interact with the robot. When the robot detected acoustic contingencies, it saved images and marked them as “social interaction present.” When it detected no contingencies, it saved images and marked them as “social interaction absent.” Then, the robot applied a visual learning algorithm to discover which visual features distinguished the former class of images from the latter. With less than six minutes of experience, the robot learned to find people in novel images. In addition, it developed a preference for line drawings of human faces, even though it had never been exposed to such schematic drawings before. Thus, previous studies that were thought to provide evidence for innate cognitive modules may actually be evidence for rapid learning mechanisms in a neonate brain that is exquisitely tuned to detect the statistical structure of the world.