Chapter 1

Introduction

When we open our eyes, we experience a world abound with visual information. What is mostly an exhilarating experience for us is a signal processing nightmare for our brains – a continuous flow of visual information bombarding our retinas needs to be processed to extract the small portions of information that are important for our actions.

Selective visual attention provides the brain with a mechanism of focusing computational resources on one object at a time, either driven by low-level image properties (bottom-up attention) or based on a specific task (top-down attention). Moving the focus of attention to locations one by one enables sequential recognition of objects at these locations.

What may appear to be a straight-forward sequence of processes (first focus attention to a location, then process object information there) is in fact an intricate system of interactions between visual attention and object recognition. How, for instance, can we move the focus of attention from one object to the next if object recognition only proceeds *after* the shift of attention? Or what does it actually mean for the object recognition system that attention is shifted, i.e., how is attention deployed to it? Can we use existing knowledge about a target object in the recognition system to bias attention from the top down? Does re-deploying attention to a new task come at a measurable cost?

Machine vision systems face a similar problem: a flood of visual information streaming into the system needs to be scanned for the task relevant parts. How can we transfer the concept of selective visual attention from biological to machine vision systems? In what way does the concept of spatial selection need to be adjusted to the specific machine vision system? Is there a measurable benefit from using the concept of attention in machine vision?

In this thesis we attempt to address these questions with a combination of computational modeling, human psychophysics, and machine vision.