CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 11

1 Lecture

Edges, Templates, Textures

Based on his visual search experiments (find the imposter symbol in a sea of rotated but otherwise uniform symbols!), Julesz conjectured that human vision operates in two distinct modes: pre-attentive vision and attentive vision. *Pre-attentive vision* is parallel and able to cover a large visual field regardless of the number of patterns. It's basically whatever happens in the first 100-200 ms. *Attentive vision* requires a linear search by focal attention in order to locate a certain pattern. It's whatever we get *after* that first look.

Additionally, Julesz proposed the concept of **textons** as units of texture that we can do statistics over. Textons are the fundamental elements of pre-attentive vision, and can vary in many ways that are imperceptible to humans – just not in ways that are pre-attentively noticeable. For reference, pre-attentive vision is sensitive to size, width, and orientation changes.

Anyway, Julesz's conjecture was that *textures cannot be spontaneously distinguished if they have the same first and second-order statistics and differ only in higher-order statistics.* Although this was later proven wrong, in practice it usually ends up being right.

- 1st-order statistics: essentially density. "How much stuff is there."
- 2nd-order: describe the relationship between pairs of elements, e.g. "far enough away from each other."

Inspiration from Biological Vision

Julesz's visual search experiments took a psychophysical approach to understanding vision. In **psychophysics**, one presents subjects with stimuli and quantitatively measures the perceptual response. We might also follow the more invasive approach of **electrophysiology**, which involves sticking probes into brains before presenting stimuli and quantitatively measuring the electrical response. Finally, these days we can also use non-invasive **imaging** (e.g. fMRI) to collect stimuli and response data.

Receptive field: the field of view of a single neuron. The receptive field of a retinal ganglion cell can be modeled as a difference of Gaussians (i.e. the Laplacian of a Gaussian).

The simple cells in the primary visual cortex (V1) are detectors of edges at different orientations. We can emulate this mathematically by taking partial derivatives along different directions (i.e. by sticking a rotation matrix in front of our Gaussian derivatives).

There are also complex cells, which are just simple cells that have some shift invariance built in. (This is the inspiration for max pooling, where e.g. we are looking for an edge but don't quite care where the edge is.) To construct a complex cell, we can take some number of simple cells and combine them into a single cell that is happy if *at least* one of its simple cells fires.

In the visual cortex there are then hypercolumns, which compute similar things but at different orientations and scales. These are like filter banks which are all processing the same receptive field.

We can represent visual information as histograms of filter responses (e.g. for different orientations), perhaps at different scales. This representation is sufficient to classify or discriminate between most textures! And if we match the histograms of an output image (one for every scale) to the histograms of an input image, we can reconstruct a texture. This corresponds to matching first-order statistics (those for individual filters).

Of course, we might also want to match second-order statistics (the statistics of pairs), since neighboring filter responses are often highly correlated as well. In other words, we can match pairwise representations of our filter bank – that is, joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.