CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 20

1 Lecture

Perceptual Organization

To a computer, an image is merely an array of numbers. But when we look at an image, we never see the numbers. We see houses, trees, skies. We can group things. We know how the perceptual units fit together.

There are many rules and principles which affect the way we perceptually group our visual input. Perceptual organization involves two separate concepts: (1) the **laws of grouping** (proximity, feature similarity, connectedness...) and (2) **figure-ground organization** (/segregation; "what doesn't go with what?").

In figure-ground organization, one part of the image is seen as the "figure" (foreground) and the rest is seen as the "ground" (background).

Contour (border) ownership, a very important concept in figure-ground organization, decrees that a contour belongs to either figure or ground, but not both. This can be essential to the perception of shape.

Many computer vision models fail to perform basic perceptual organization tasks that humans are able to do naturally, i.e. in the presence of optical illusions or excessive visual texture.

Gestalt Psychology

> "The whole is other than the sum of the parts." – Kurt Koffka

Gestalt psychology explains such a failure to perform perceptual organization as the idea that every individual element contributes (through complex interactions) to the overall perception, which can be something else entirely.

The holistic perception defines the parts it was composed from, rather than being a secondary quality that emerges from those parts.

From Pixels to Perception

We would like to study how perception of even an unknown environment can be supported by a computational routine. How can we capture some of these perceptual observations through computation?

In a way, segmentation and figure-ground organization are the computational counterparts of Gestalt psychology. We want to turn our ideas into actions – something we can evaluate.

For example, the boundaries of image regions are defined by a number of cues: brightness, color, texture, motion (in video), familiar objects, etc. We can learn the *posterior probability of a boundary* $P_b(x, y, \theta)$ (from local information only!) by computing boundary cues for an image and combining them into a model.

Also, we can perform image segmentation as graph partitioning using **normalized cuts**. We formulate *image pixels* as vertices, and *connections between neighboring pixels* as edges. Then we partition graph nodes so that similarity within individual groups is large (i.e. the group is cohesive), and similarity between different groups is small.

For figure-ground organization, we could cut the graph into two parts: a figure and a ground.