Mid Level Vision

March 18, 2004

Josh McDermott

What is mid level vision?

- Perception involves inferring the structure of the world from measurements of energy generated by the world (in vision, this is light).
- We can think of early vision as a set of useful measurements (made with receptive fields).
- In V1, for instance, the various kinds of receptive fields give us local measurements of orientation, contrast, disparity, color, spatial frequency, etc.
- Mid level vision is where the visual system begins to make inferences about the world based on those measurements.

Local measurements are ambiguous

(Image removed due to copyright considerations.)

We need to combine information across space to infer what is out there in the world.

What is mid level vision?

- Mid level vision probably involves taking local measurements made by the receptive fields early in the visual system and combining them in various ways.
- We have already talked about several aspects of mid level vision…

Lightness perception as revealed by lightness illusions

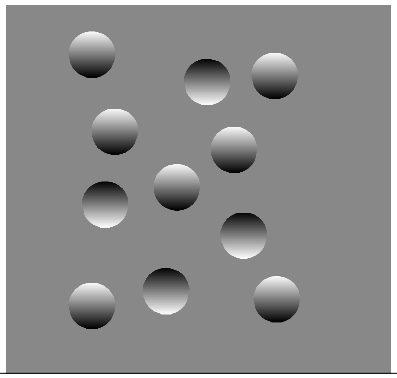
(Image removed due to copyright considerations.)

Combining luminance measurements can help to infer lightness:

(Image removed due to copyright considerations.)

Aside: what aspect of this story doesn't make sense?

Shape from shading



What is mid level vision?

- There are many visual areas beyond V1 that probably implement these phenomena.

Image removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 4.15, p. 129

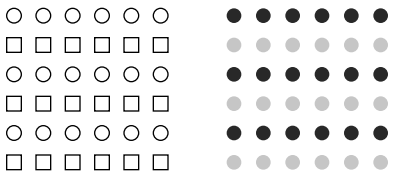
Another aspect of mid level vision:

Perceptual Grouping

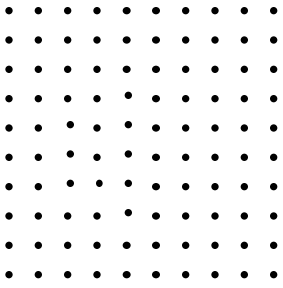
Things that are similar tend to group together.

There are other grouping rules as well, as studied by the Gestalt psychologists.

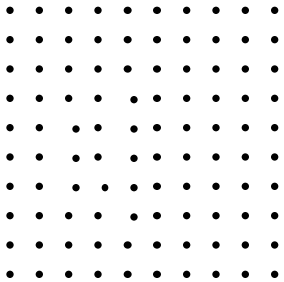
Grouping by similarity



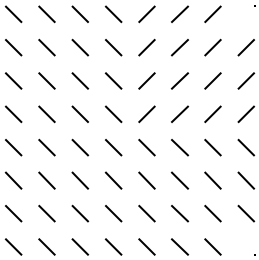
Grouping by motion
(common fate)



Grouping by motion (common fate)

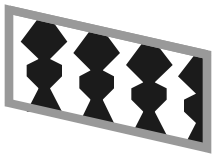


Texture grouping



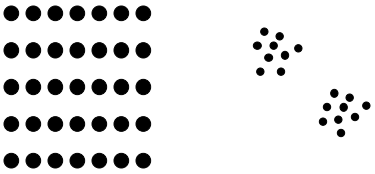
Grouping by symmetry

Black or white can be figure or ground here. Both are symmetrical.

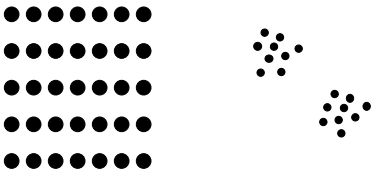


Black is more symmetrical here, and more likely seen as figure.

Grouping by proximity

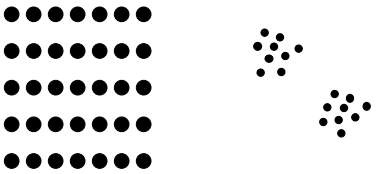


The bottom-up mechanistic approach



We can explain these effects by saying that the brain is hooked up in a certain way; for instance, there are certain neurons in the brain that respond more strongly when their neighbors are also responding, and so you get a strong response when many dots are in a line or in a clump. This is an explanation of the mechanics of the processing, but says nothing about why the brain is set up that way.

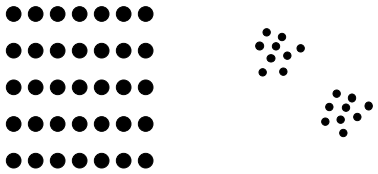
The simplicity principle



Another explanatory principle: our brain is set up to generate simple descriptions.

For instance, on the right, it is simpler to say that there are two clumps of dots with 9 dots in each, than to list the locations of every individual dot. We organize the pattern into parts, which makes it easier to communicate, reason, and remember things about the image. This is an abstract principle, which says nothing about how the neurons are hooked up.

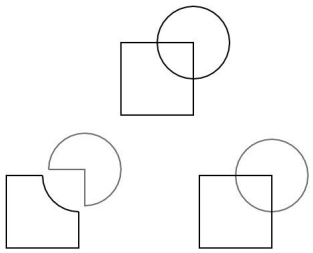
The probability principle.



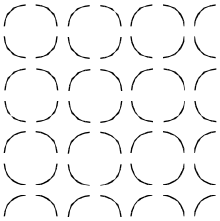
One explanatory principle (Helmholtz). What we see is our best guess as to what is in the world, based on the input data and our prior experience.

For example, when things are close to each other in the image, there is a good chance that they were part of the same object in the world, so we have a tendency to see them as part of the same object. The heuristics of our visual system are based on the probabilities in the world. This is also an abstract explanation that says nothing about the neurons.

Grouping by good continuation

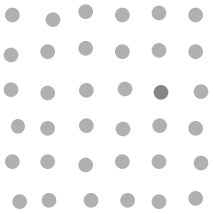


Grouping by Closure



Which is the building block?

Popout: the flip side of grouping



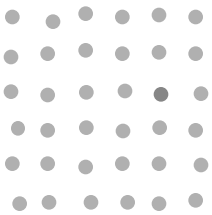
Things that are similar group together.

Things that are different pop out as separate.

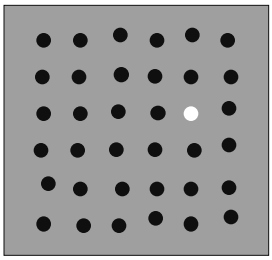
It works for simple, easily computed properties.

(Triesman, Julesz did early work)

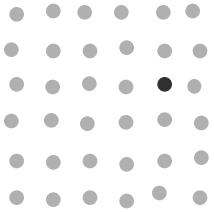
Popout by color



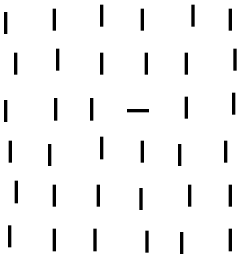
Popout by polarity



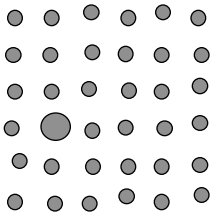
Popout by brightness



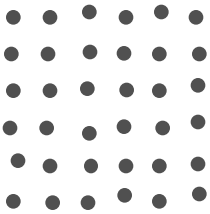
Popout by orientation



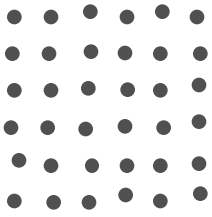
Popout by size



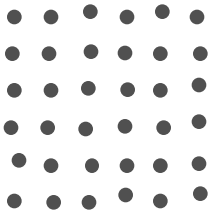
Popout by motion



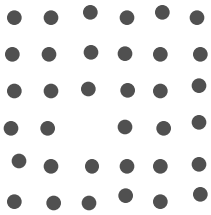
Popout by motion



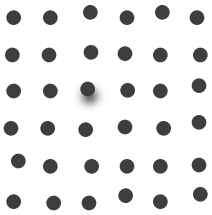
Popout by flicker



Popout by flicker

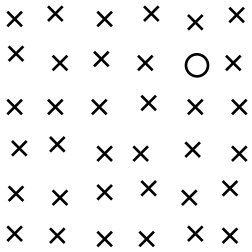


Popout by depth



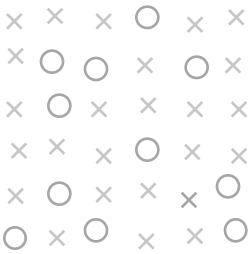
Since we can't show stereoscopic depth here, we are simulating it with a drop shadow.

Popout by shape



Popout is weak for complex properties

Conjunctions may not popout.



Element grouping and popout are usually based on simple properties.

Simple properties can be computed by the machinery of area V1, and involve comparisons between filter outputs.

The filters are usually weighted sums on retinal inputs.

Examples: color, size, orientation, flicker.

How might grouping and popout be implemented in the brain?

"Association fields" for contour integration (good continuation)

(Image removed due to copyright considerations.)

Main idea: have neurons that would be stimulated by the same contour excite each other, so as to enhance the representation of that contour.

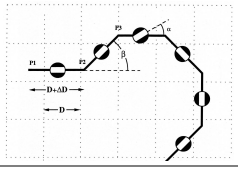
Solid lines represent excitatory interactions, dashed lines represent inhibitory interactions.

How might one study grouping experimentally?

Path integration: Detect the aligned snake in the field of gabors

(Image removed due to copyright considerations.)

Task is easy if the orientation varies smoothly from one Gabor to the next, hard if it varies less smoothly.



(Images removed due to copyright considerations.)

(Images removed due to copyright considerations.)

(Images removed due to copyright considerations.)

(Images removed due to copyright considerations.)

Another aspect of perceptual organization: Figure/ground organization: the face/vase figure.

When a face is seen as figure, it seems to lie in front of the background. It also "owns" the contour; i.e., it is the cause of the contour and is attached to it.

When the vase is seen as figure, it appears in front, and now it owns the contour.

A version of this figure is available at <http://www.maxdstandley.com/art_info/i/fv.gif>.

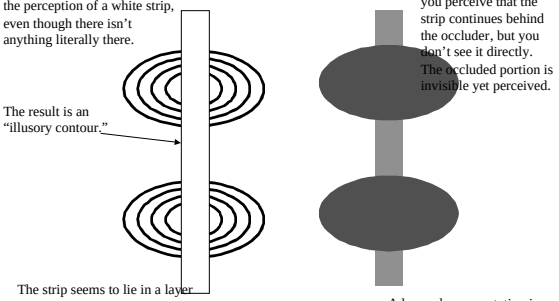
Important concept: border ownership.
If an edge is due to occlusion, only one side can "own" it.
Our perceptual representations seem to embody this constraint.

Modal, amodal completion

"Modal" completion: you have the perception of a white strip, even though there isn't anything literally there.

The result is an "illusory contour."

"Amodal" completion: you perceive that the strip continues behind the occluder, but you don't see it directly. The occluded portion is invisible yet perceived.



The strip seems to lie in a layer above the background; it is revealed by its occlusion.

A layered representation is also evident here.

Relatable contours

Kellman and Shipley found that certain rules predict whether contours would complete behind an occluder.

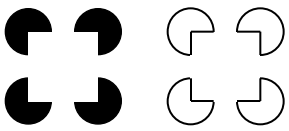
Basically, the contours should be "aiming" toward the same place behind the occluder, and they should not be at too steep an angle. K&S called this *relatability*.

K&S also argued that amodal completion obeys the same rules as modal completion, and thus is likely the result of the same mechanism. Others, including our own Bart Anderson, have argued against this notion.

Kanisza figures

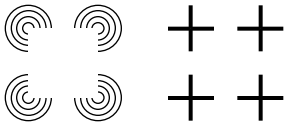
Kanisza developed figures that give a strong sense of an occluding (but illusory) square. Modifications help reveal the heuristics the visual system is using.

The standard Kanisza square.



The effect is weakened when the "pac-man" figures are outlines.Why?

Using concentric rings enhances the effect. Multiple aligned endpoints gives a strong effect.



This version is really weak. Interpreting it as four crosses makes more sense than inferring an occluding square

Why are the illusory contours weaker in (b) than in (a)?

(Image removed due to copyright considerations.)

Neon color spreading

Brain infers presence of colored filter when arrangment of local image data supports this:

(Image removed due to copyright considerations.)