Hannah Small

1. NI and AI vision
   1.1. Three complexities that vision must deal with are the infinite set of possible images, their viewing conditions, and how the scene is composed. The combination of the possibilities for these three aspects, which themselves contain more options like color, distance, lighting, etc., yields an infinite number of images that could be seen.

   Vision must also deal with ambiguities, including local ambiguity of an image and ambiguity in how images are generated from the 3D world. Without context, patches of images are very difficult to parse, since there are an astronomically large number of possible pixel configurations. Ambiguities in images also arise from the combination of lighting and geometry, since there are many ways for a certain image to be created using different geometries and lighting. These complexities and ambiguities make tasks like image segmentation, object recognition, and depth estimation difficult. With only intensity information, complexities and ambiguities are difficult to deal with because the intensity patterns change greatly even with small changes to the scene. Remarkably, the human visual system can deal with this.

   1.2. Failures of human vision are unlikely to help AI. These include change blindness, or the human failure to notice changes in a scene, especially in the periphery, and visual crowding, which is the impairment of discrimination of visual cues when presented in a clutter. It shows a fundamental limit on conscious visual perception that does not necessarily need to be imposed on AI.

   Visual tasks that humans are very good at could help AI. Humans can use their world knowledge to use context to recognize objects. They also use abstraction and domain transfer, for example when they are seeing something completely new or identifying something from a very abstract sketch. We understand the abstract relationships between objects. Finally, humans understand that parts make up objects: they understand the compositionality and that parts are organized in a hierarchy of basic elements. Compositionality enables humans to deal with an enormous number of objects.

   1.3. Three types of theory for object recognition in the visual stream are:
      1.3.1. Marr's theory, a theory that emphasizes that vision can be understood on the computational, algorithmic (or representational), and implementational level.
      1.3.2. Feedforward deep neural networks, which consist of a hierarchy moving from simple features in the lowest levels, to more complex features further up the hierarchy, that are more invariant.
      1.3.3. Analysis-by-synthesis, which is a Bayesian model consisting of high-level areas which generate images and compare to the input image in a feedback process.
      1.3.4. One hierarchy metaphor is the Army. The Army metaphor is that the General has executive summary and contacts the Colonels for information, Colonels contact Major, and so on, down to the Privates. Information flows up and down the hierarchy (combination of feedforward and feedback processing)

2. Explaining Illusions
    2.1. **Neon color spreading** - This illusion is an example of the tendency of the human visual system to interpret images as geometric shapes. The color emphasizes this effect. This involves mid-level vision that knows about geometry, lighting, and that objects can occlude one another.
    2.2. **Motion binding** - This illusion is an example of the visual system trying to solve the ambiguity of motion and form. Since many different motions on the retina can cause the same response in motion-sensitive neurons, the local motion is ambiguous at the low-level. Humans use form to help solve this ambiguity, by establishing distinct objects and integrating their whole motion. In this illusion, the vertices of the square are occluded so there is not enough information on the form. The visual system must solve the motion by assuming motion is slow and smooth. This illusion involves low-level and mid-level vision.
    2.3. **Hollow face illusion** - This illusion is an example of the visual system trying to solve the ambiguity of how the image was generated and the fact that we have a high-level visual area dedicated to processing this particular stimuli: faces. Because the visual system cannot rely on the shadows in this video (due to strange lighting conditions), there is not very much information on the geometry of the mask. While trying to resolve this ambiguity using mid-level vision, the high-level visual region that responds to faces nudges the visual system to interpret the hollow mask as a face.
    2.4. **Dalmatian dog** - This illusion is an example of the visual system trying to solve an extreme ambiguity of low-level cues for edges. Since the only information is seemingly random black and white blobs, it is difficult to detect the edges. However, when you are told that it is a Dalmatian, high-level vision can help you interpret the noisy input and identify the dog. Once you see the dog, you can find the dog very easily.
    2.5. **Ball in a box** - This illusion is an example of the visual system solving for plausible causal relations that can explain visual cues. Specifically, the visual cue of the shadow indicates that the ball is levitating because of prior experience with lighting and shadow motion (which are coupled cues). This illusion uses mid-level vision to process the shapes, shadows, and lighting of this illusion.
    2.6. **Checker shadow illusion** - The visual system interprets the cue of the fuzzy darker shape as a shadow and then solves the luminance (using intensity) differently for the B square than for the A square, resulting in the perception of two different colors. It uses the local contrasts (darker squares around B) as more information in this calculation. This illusion involves mid-level vision.

3. Linear filters

3.1. It has been shown that simple cells can be modeled by Gabor filters. They are sensitive to orientation, frequency, and phase. The Gabor is a combination of a sinusoid, which has perfect frequency localization, and a Gaussian, which has position localization.

Complex cells are sensitive to orientation and frequency, and not as sensitive to the position of the stimuli. This makes them good motion detectors. They have been modeled as receiving input from multiple simple cells, one in particular uses a quadrature pair of Gabors as input to the complex model.

Complex cells are not considered to be a linear filter of its input because each simple cell response is squared and then summed. This results in violating scalability and superposition that is required of linear filters.

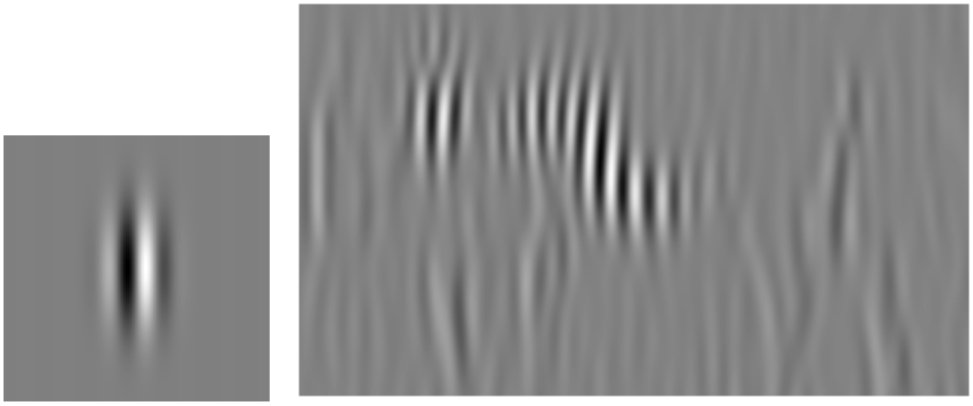
3.2. The first derivative of a Gaussian filter captures edge information (orientation and magnitude) from an image. The second order derivative is also an edge detector, but does not retain the edge orientation information. The first derivative is more susceptible to noise because you have to use a threshold, whereas you can just use zero-crossings in the second derivative to detect edges.

Center surround cells, found in the retina and LGN, can be modeled by the second derivative of the Gaussian filter, or the Laplacian of the Gaussian (LoG). Multiple center surround cells are used as input to simple cells, found in V1 and modeled by Gabor filters. A quadrature pair is a pair of neighboring simple cells that are tuned 90 degrees out of phase. Complex cells, also found in V1, can be modeled as having a quadrature pair (sine Gabor and cosine Gabor) as input, where the response from each simple cell is squared and then summed to compute the complex cell's response.

The order of convolving an image with a Gaussian filter and a derivative of a Gaussian filter does not matter because you can apply convolutions in an arbitrary manner. The order of convolution does not matter.

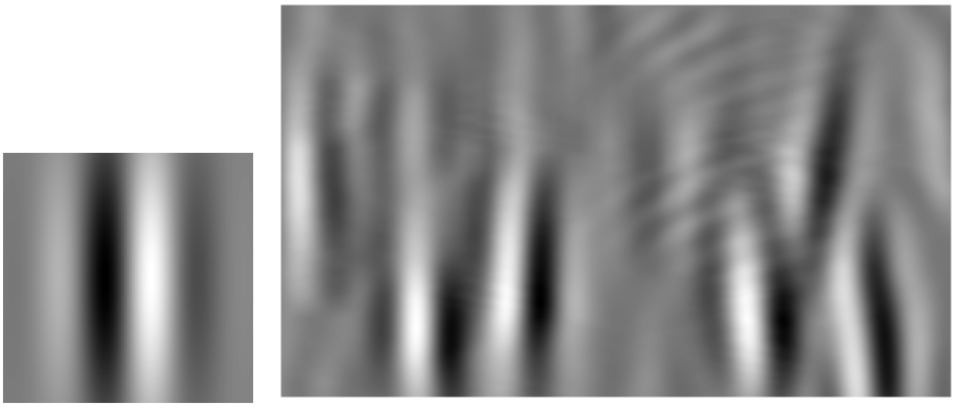
4. Jupyter notebook

4.1. Vertical stripes: sinFilterBank[12] -- omega=0.32, theta = 0



Horizontal stripes: sinFilterBank[14] -- omega = 0.32, theta = 1.5707963267948966

Legs: sinFilterBank[4] -- omega = 0.12, theta = 0



4.2. The simple cell responds to the transition between dark and light, so the highest responses are at edges from dark to light, which happens every other bar in the stimulus.

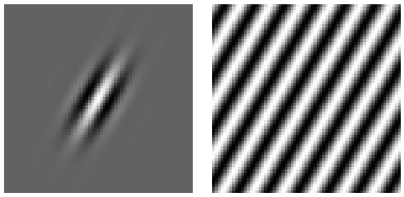
The complex cell combines the response of the sine and cosine Gabor, which are complementary since the sine Gabor responds to edges going from light to dark and the cosine Gabor responds to edges going from dark to light, so it is responsive at all points.

The complex cell uses the magnitude of the simple cells' response, so it only is sensitive to the frequency information and not the phase information, resulting in an unmodulated full response.

I was surprised that the simple cells responded this much when the Gabor filter was not at the same orientation. I adjusted the Gabor to match perfectly and found that indeed the response was higher. Since complex cells are sensitive to frequency, the full response makes sense.

4.3. omega= 0.7000000000000001, theta = 0.5236000000000001, rho = 0

Strength of response:[[69.51876024]]



4.4.

The maximum response occurs at $\omega = 0.07$ We know that the cell responds best to frequencies with

$$|\omega^2| = \frac{2}{\sigma^2}$$

so we have

$$0.0049 = \frac{2}{\sigma^2}$$

so

$$\sigma = 20.203050891$$