Questions

1. See the video with the Ames Room illusion in the lecture. What prior $p(S)$ might the visual system be using to account for this illusion?

2. It has been know for hundreds of years that a hill-like proturberance on a surface that is illuminated from one direction can be confused with a valley-like indentation that is illuminated from the opposite direction. More generally, consider a depth map $Z_0 + Z(X, Y)$ where the Z_0 is an offset used to position the surface at some distance from the observer, and the $Z(X, Y)$ describes the variations in depth around the Z_0. Assume the surface is illuminated from direction (L_X, L_Y, L_Z).

   Show that a surface with depth map $Z_0 - Z(X, Y)$ can give exactly the same image according to the $\mathbf{N} \cdot \mathbf{L}$, if one modifies the light source direction accordingly. Hint: what is the normal of the depth reversed surface? Assume the "N dot L" model.

3. Suppose we have a likelihood function $p(I|S)$ and a prior $p(S)$, and that both have a Gaussian shape. The prior, in particular, will be a Gaussian probability function. Further suppose we have an instance $I = i$ and we can compute the likelihood function. How can we combine the likelihood function and prior to estimate the value of S that maximizes the posterior ?

   Hint: the derivation is essentially the same as given at the beginning of the lecture for combining two likelihoods functions in the case of conditional independence.

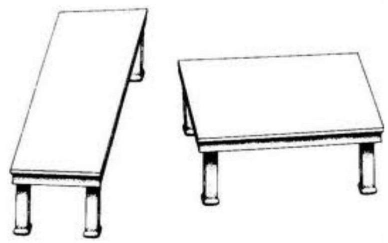
4. What if we have two likelihood functions and a prior, for example, we have texture and stereo cues that give us (conditionally) independent likelihoods of surface slants, and we also have a prior ? How do we combine the weights so that we can compute a maximum posterior probability of S ? Assume conditional independence of the two likelihood functions, so:

$$p(S|I_1, I_2) = \frac{p(I_1|S)p(I_2|S)p(S)}{p(I_1, I_2)}.$$

5. One might expect the prior $p(x, y, d)$ on binocular disparities to have its peak at $d = 0$ for image positions near the center of the fovea(s), i.e. near $(x, y) = (0, 0)$, since the eyes tend to verge so that the center of fovea has zero disparity.

   Would you expect the prior to also have its peak at $d = 0$ for other (x, y) values ? In particular, would you expect the prior for points above the line of sight to have its peak at $d = 0$. Similarly what about below the line of sight?

6. The drawing below (due to Roger Shephard) shows two trapezoids. The 2D shape and size of these two trapezoids appear to be different: the one on the left looks longer and skinnier than the one on the right. But in fact they are the same. How would you explain this illusion in terms of likelihoods or priors?



7. (a) Describe a psychophysical experiment to measure how accurately people perceive the orientation of a line in an image. Mention what happens on each trial, what the task is, how the experimenter determines the threshold, and what is plotted to illustrate the result.

   (b) What would a maximum likelihood model predict about how the plot changes as the image noise is increased? Briefly explain why.

   (c) Natural visual environments tend to have more vertical and horizontal structures than diagonal structure. For example, trees and buildings are vertical and objects lying on the ground are horizontal. These prior probabilities suggest a Bayesian model of image line orientation perception. What specific predictions would such a Bayesian model make about perceived line orientations?

Solutions

1. The prior assumption that the visual system seems to make is that rooms are typically cubes, that is, the walls and ceilings and floors tend to make 90 degree angles to each other.

   In fact the Ames room is not a cube. It is trapezoidal. However, when viewed from a particular position in space, it has the same image projection as a cube. There are many ways to make trapezoidal shaped rooms, but most trapezoidal rooms do *not* appear as cubes.

   What's most interesting about the Ames room is that the prior on the room shape overrides another strong prior that you would think the visual system has, namely that objects (especially people!) do not change size as they move.

2. Ignoring shadows, the "N dot L" model is:

$$I(X,Y) \;=\; \frac{1}{\sqrt{(\frac{\partial Z}{\partial X})^2 + (\frac{\partial Z}{\partial Y})^2 + 1}} \, (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1) \cdot (L_X, L_Y, L_Z)$$

   If we reverse the variations in depth about the Z_0 axis, i.e. $Z_0 - Z(X,Y)$, then the partial derivatives of Z with respect to X, Y change sign. If we also negate the signs of L_X and L_Y i.e. we illuminate the surface from a direction $(-L_X, -L_Y, L_Z)$, then we get the same original image $I(X,Y)$. i.e.

$$I(X,Y) \;=\; \frac{1}{\sqrt{(\frac{\partial Z}{\partial X})^2 + (\frac{\partial Z}{\partial Y})^2 + 1}} \, (-\frac{\partial Z}{\partial X}, -\frac{\partial Z}{\partial Y}, -1) \cdot (-L_X, -L_Y, L_Z)$$

   For example, a hill illuminated from say the left gives the same image as a valley illuminated from the right.

3. Let the likelihood function $p(I = i|S)$ be a Gaussian with mean $S = s_1$ and variance σ_1^2. Let the prior be a Gaussian with mean s_p and and variance σ_p^2. The solution is exactly the same as in the lecture where we maximized $p(I_1|S)p(I_2|S)$. In the current problem, we want to maximize

$$p(I|S)p(S) = \frac{a_1}{2\pi\sigma_p} e^{-\frac{(s-s_1)^2}{2\sigma_1^2}} e^{-\frac{(s-s_p)^2}{2\sigma_p^2}} .$$

   You can crank through the same derivation, or just plug in the new names of the variables. You get:

$$w_1 = \frac{\sigma_p^2}{\sigma_1^2 + \sigma_p^2} \qquad\qquad w_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_p^2}$$

4. We now want to maximize

$$p(I_1|S)\; p(I_2|S)\; p(S) = e^{-\frac{(s-s_1)^2}{2\sigma_1^2}} e^{-\frac{(s-s_2)^2}{2\sigma_2^2}} e^{-\frac{(s-s_p)^2}{2\sigma_p^2}}$$

   so we want to minimize

$$\frac{(s-s_1)^2}{2\sigma_1^2} + \frac{(s-s_2)^2}{2\sigma_2^2} + \frac{(s-s_p)^2}{2\sigma_p^2}.$$

Take the derivative with respect to s and set it to 0 gives

$$\frac{s_1 - s}{\sigma_1^2} + \frac{s_2 - s}{\sigma_2^2} + \frac{s_p - s}{\sigma_p^2} = 0$$

and so

$$s = (\frac{s_1}{\sigma_1^2} + \frac{s_2}{\sigma_2^2} + \frac{s_p}{\sigma_p^2}) \ / \ (\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_p^2})$$

This is the estimate of the maximum of the posterior. Note that this is of the form

$$s = \sum_{i=1}^{3} w_i s_i$$

where

$$w_i = \sigma_i^{-2} / \sum_i \sigma_i^{-2}.$$

5. Recall that there is a prior for floors over ceilings. This is presumably because we walk on the ground and often overhead there is only the sky. Moreover, we often look at the ground e.g. in order to navigate. These observations suggest there is a prior for points above the line of sight to be further away than points below the line of sight. For the disparity prior to be consistent with this slant prior, points above the line of sight should have a disparity prior that peaks at a negative value (not at $d = 0$). Similarly, points below the line of sight should have a disparity prior that peaks at positive value.

6. (a) If you remove the table legs from the drawing and just have the two shapes, then they look identical except they are rotated. So the table legs make a huge difference. They suggest the shapes are the tops of table and are standing on the ground. The prior for floor slants over ceiling slants helps out with this percept, as if you rotate the image 180 degrees then you don't tend to perceive a table glued to ceiling. Because the legs and prior suggest that the table tops are slanted back like a floor, the shapes must be *foreshortened* in the Y direction. This implies the shape on the left is longer than it appears and the shape on the right is wider than it appears.

7. (a) Each trial could consist of two images, each containing a line, and the subject could be asked to judge which image had the more vertical line. For each reference orientation, a psychometric function could be plotted as a function of the test orientation (or the difference in the test and reference orientations). Only orientations very near the test orientation would need to be considered as the task would be trivial for large orientation differences. The threshold could be the orientation difference between reference and test that is needed to get 75 % correct. One would plot the thresholds as a function of the reference orientation.

   (b) If you add noise, then for each reference orientation, the threshod would rise. That is, you would need greater differences in the line orientations in order to obtain 75 % correct.

   (c) The Bayesian model would bias the perceived line orientations toward vertical and horizontal. Specifically, it predicts that any orientation that is less than 45 degrees from vertical would be perceived as cloer to vertical than it really is, and any orientation less than 45 degrees from horizontal would be perceived as closer to horizontal than it really is.

As noise is increased, the bias toward vertial and horizontal would increase since the weight from the noisy image would be lowered relative to the weight from the prior.