CSCI 5822 Assignment 2

Evan Sidrow

February 8, 2018

1 Article Commentary

The article *Motion illusions as optimal percepts* by Yair Weiss, Eero P. Simoncelli, and Edward H. Adelson attempts to reconcile the fact that several deterministic models of human motion tracking vary in performance depending on context. **The main idea behind the article** is to fix this issue by taking a Bayesian, probabilistic approach to the problem which performs well in a more diverse set of contexts. To formulate the model, the authors make two primary assumptions about human motion perception- that humans favor slow motion and that human image measurements are noisy. Perhaps the most important line to remember regarding the usefulness of this model comes from the final paragraph from its "Discussion" section: "Although the details of our model should certainly be refined and extended to handle more complicated phenomena, we believe the underlying principle will continue to hold: that many motion 'illusions' are not the result of sloppy computation by various components in the visual system, but rather a result of a coherent computational strategy that is optimal under reasonable assumptions." This means that the method described in the article can be used to predict future human behavior and allow machines to better emulate humans, while previous methods have less predictive power, since they must be adjusted to fit the observed data.

I am left with several questions after reading the article. One primary question centers around the nature of what it means for human image measurements to be "noisy". The authors themselves admit that they have "have no direct evidence" for the existence or nature of this noise, but the model still fits human data remarkably well. I think that clarifying a physical interpretation of human observational noise could be used to better interpret the results of the study and fit the free parameters used in the model.

I do have a couple of critiques of the article. Once such criticism stems from their small sample size of human subjects. For several experiments, the authors used only three human subjects, which I believe to be too few to make any meaningful conclusions across subjects. It might also be interesting to test a diverse set of subjects and find if their results would be different across culture, gender, etc.

In addition, The researchers claim that qualitative results from their model do not change as a free parameter was varied over two orders of magnitude. I think the authors should have either used a wider range or justified why only two orders was sufficient.

The assumption that noise is independent over spatial location is not very satisfying, but again, the nature of "noise" in human measurement is not well-defined, and the math behind the model would be much more difficult without this assumption.

I have some ideas regarding how the model could be extended. In particular, I think it would be interesting to use a hierarchal Bayesian model with probability distributions for σ (the standard deviation of observational noise) and σ_p (the standard deviation of priors on object speed). It would be interesting to estimate these distributions to gain insight into how they vary across people and what that means practically.

Also, I would be interested to see how prior distributions change across settings or subjects. For example, a person's prior belief about object speed may be different if that person is inside an automobile. It may be interesting to see what other settings change a subject's prior.

Finally, one interesting idea from the article is that humans become more accurate at motion detection after longer exposure. The "ideal" Bayesian observer could match this behavior by integrating information over time. The authors never run any experiments with this idea, but it would be interesting for them to test the performance of the model with integration over time vs humans with prolonged exposure. This experiment may also be helpful in tuning the parameters of the model to match human performance.

2 Derivation of Gaussian Conditional Probability

We can see from equation (8.4.19) in the textbook that if we have a random vector $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})^T$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma_{xx}} & \boldsymbol{\Sigma_{xy}} \\ \boldsymbol{\Sigma_{yx}} & \boldsymbol{\Sigma_{yy}} \end{bmatrix}$$

then:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu_x} + \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}(\boldsymbol{y} - \boldsymbol{\mu_y}), \boldsymbol{\Sigma_{xx}} - \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{yx}}\right) \tag{1}$$

Now, let $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ have components $(x_1, x_2, x_3, ..., x_n)$. We are looking for $p(x_1|x_2, x_3, ..., x_n)$. However, notice that in this case, x_1 is similar to $\boldsymbol{x}$ in equation (1) and $(x_2, x_3, ..., x_n)^T$ is similar to $\boldsymbol{y}$ in equation (1). So:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \boldsymbol{\mu_{2:n}} \end{bmatrix}$$

where $\boldsymbol{\mu_{2:n}} = (\mu_2, ..., \mu_n)^T$, and

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \boldsymbol{\Sigma_{1,2:n}} \\ \boldsymbol{\Sigma_{2:n,1}} & \boldsymbol{\Sigma_{2:n,2:n}} \end{bmatrix}$$

where σ_{11}^2 is a scalar (the variance of x_1), $\boldsymbol{\Sigma_{2:n,1}}$ is an $(n-1)$-by-1 vector, $\boldsymbol{\Sigma_{1,2:n}}$ is a 1-by-$(n-1)$ vector, and $\boldsymbol{\Sigma_{2:n,2:n}}$ is a $(n-1)$-by-$(n-1)$ covariance matrix.

Then, plugging all this in to equation (1), we get that the conditional probability is:

$$p(x_1|x_2, x_3, ..., x_n) = \mathcal{N}\left(x_1|\mu_1 + \boldsymbol{\Sigma_{1,2:n}}\boldsymbol{\Sigma_{2:n,2:n}^{-1}}(\boldsymbol{x_{2:n}} - \boldsymbol{\mu_{2:n}}), \sigma_{11}^2 - \boldsymbol{\Sigma_{1,2:n}}\boldsymbol{\Sigma_{2:n,2:n}^{-1}}\boldsymbol{\Sigma_{2:n,1}}\right) \tag{2}$$

Note that both the mean and standard deviation in equation (2) are scalars, which checks out since this is a probability distribution for just one variable (conditioned on a vector).

Finally, we have:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

SO:

$$p(x_1|x_2, x_3, ..., x_n) = \frac{1}{\sqrt{2\pi\left(\sigma_{11}^2 - \boldsymbol{\Sigma_{1,2:n}}\boldsymbol{\Sigma_{2:n,2:n}^{-1}}\boldsymbol{\Sigma_{2:n,1}}\right)}} \exp\left(-\frac{\left[x_1 - \left(\mu_1 + \boldsymbol{\Sigma_{1,2:n}}\boldsymbol{\Sigma_{2:n,2:n}^{-1}}(\boldsymbol{x_{2:n}} - \boldsymbol{\mu_{2:n}})\right)\right]^2}{2\left(\sigma_{11}^2 - \boldsymbol{\Sigma_{1,2:n}}\boldsymbol{\Sigma_{2:n,2:n}^{-1}}\boldsymbol{\Sigma_{2:n,1}}\right)}\right)$$

where σ_{11}^2 is a scalar (the variance of x_1), $\boldsymbol{\Sigma_{2:n,1}}$ is an $(n-1)$-by-1 vector containing covariances between x_1 and $\boldsymbol{x_{2:n}}$, $\boldsymbol{\Sigma_{1,2:n}}$ is a 1-by-$(n-1)$ vector containing covariances between x_1 and $\boldsymbol{x_{2:n}}$, and $\boldsymbol{\Sigma_{2:n,2:n}}$ is a $(n-1)$-by-$(n-1)$ covariance matrix for $\boldsymbol{x_{2:n}}$.