Optimal Linear Cue Combination

Robert Jacobs
Department of Brain & Cognitive Sciences
University of Rochester
Rochester, NY 14627, USA

August 8, 2008

In this note, we derive an optimal (in a maximum likelihood sense) linear rule for combining information from multiple information sources. The discussion will be simpler if we consider a specific concrete example.

Suppose that you are an observer wishing to estimate the visual depth of an object defined by motion and texture cues. Let d denote a possible depth of a visual object, and let m and t denote the values of the motion and texture cues.

Using Bayes' rule,

$$p(d|m,t) \propto p(m,t|d)p(d). \tag{1}$$

Suppose that all depth values are equally likely, meaning that the prior distribution over depth values is a uniform distribution (i.e., it is constant). In addition, assume that the motion and texture cues are conditionally independent given the depth value d (i.e., each cue is corrupted by its own noise source). In this case,

$$p(d|m,t) \propto p(m|d)\; p(t|d). \tag{2}$$

Also assume that all motion values are equally likely [i.e., $p(m)$ is a uniform distribution] and that all texture values are equally likely [i.e., $p(t)$ is uniform], meaning that $p(m|d) \propto p(d|m)$ and $p(t|d) \propto p(d|t)$. Consequently,

$$p(d|m,t) \propto p(d|m)\; p(d|t). \tag{3}$$

Let's consider the individual distributions $p(d|m)$ and $p(d|t)$. We assume that these are Gaussian distributions

$$p(d|m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\{-\frac{1}{2\sigma_m^2}(d-d_m)^2\} \tag{4}$$

$$p(d|t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\{-\frac{1}{2\sigma_t^2}(d-d_t)^2\} \tag{5}$$

where d_m and σ_m^2 are the mean and variance of $p(d|m)$, and d_t and σ_t^2 are the mean and variance of $p(d|t)$. We can re-write Equation 3 as follows:

$$\begin{aligned} p(d|m,t) &\propto p(d|m)\; p(d|t) && (6)\\ &\propto \exp\{-\frac{1}{2\sigma_m^2}(d-d_m)^2\}\; \exp\{-\frac{1}{2\sigma_t^2}(d-d_t)^2\} && (7)\\ &= \exp\{-\frac{1}{2\sigma_m^2}(d^2-2dd_m+d_m^2) - \frac{1}{2\sigma_t^2}(d^2-2dd_t+d_t^2)\} && (8) \end{aligned}$$

$$= \exp\{-\frac{d_m^2}{2\sigma_m^2} - \frac{1}{2}d^2(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}) + d(\frac{d_m}{\sigma_m^2} + \frac{d_t}{\sigma_t^2}) - \frac{d_t^2}{2\sigma_t^2}\} \quad (9)$$

$$= \exp\{-\frac{1}{2}d^2(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}) + d(\frac{d_m}{\sigma_m^2} + \frac{d_t}{\sigma_t^2})\} \exp\{-\frac{d_m^2}{2\sigma_m^2} - \frac{d_t^2}{2\sigma_t^2}\} \quad (10)$$

$$\propto \exp\{-\frac{1}{2}d^2(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}) + d(\frac{d_m}{\sigma_m^2} + \frac{d_t}{\sigma_t^2})\} \quad (11)$$

where Equations 7 and 11 dropped constants that do not depend on d, and where Equations 8 and 10 have made use of the identity $e^a e^b = e^{a+b}$. As a matter of notation, let

$$d^* = \frac{\frac{1}{\sigma_m^2}d_m + \frac{1}{\sigma_t^2}d_t}{\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}} \quad (12)$$

and

$$\sigma_{m,t}^2 = \frac{1}{\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}} = \frac{\sigma_m^2\sigma_t^2}{\sigma_m^2 + \sigma_t^2}. \quad (13)$$

Using this notation, we re-write Equation 11 as

$$p(d|m,t) \propto \exp\{-\frac{1}{2}d^2\frac{1}{\sigma_{m,t}^2} + d\frac{d^*}{\sigma_{m,t}^2}\}. \quad (14)$$

Next we add into the exponent $-\frac{1}{2}\frac{(d^*)^2}{\sigma_{m,t}^2}$ which is a constant that does not depend on d, yielding

$$p(d|m,t) \propto \exp\{-\frac{1}{2\sigma_{m,t}^2}(d - d^*)^2\}. \quad (15)$$

Because $p(d|m,t)$ is a density that must integrate to one, we get the result

$$p(d|m,t) = \frac{1}{\sqrt{2\pi\sigma_{m,t}^2}} \exp\{-\frac{1}{2\sigma_{m,t}^2}(d - d^*)^2\} \quad (16)$$

meaning that $p(d|m,t)$ is a Normal distribution with mean d^* and variance $\sigma_{m,t}^2$.

This is a remarkable result. Note that the mean of this distribution d^* is a linear weighted average of d_m (the mean depth indicated by the motion cue) and d_t (the mean depth indicated by the texture cue) with weights depending on the variances σ_m^2 and σ_t^2 as indicated in Equation 12. Also note that $\sigma_{m,t}^2 \leq \sigma_m^2, \sigma_t^2$, meaning that the distribution based on both cues is more precise than the distributions based on either cue alone (i.e., cue combination gives us a more precise estimate).