A short explanation of why optimal cue combination can be done with a weighted sum

Suppose the true depth of an object is a random variable D, and we have two depth cues that are also random variables, D_1 and D_2. Bayes' theorem tells us the probability of any possible value of D, given the observed values of the depth cues.

$$P\big(D = d \,\big|\, D_1 = d_1 \,\&\, D_2 = d_2\big) = \frac{P(D_1 = d_1 \,\&\, D_2 = d_2 | D = d)P(D = d)}{P(D_1 = d_1 \,\&\, D_2 = d_2)}$$

If we find the value of d that maximizes the above equation, then we are finding a maximum a posteriori (MAP) depth estimate. Note that we only have to find the value of d that maximizes the numerator, since the denominator doesn't depend on d. So we can drop the denominator:

$$\propto\ P(D_1 = d_1 \,\&\, D_2 = d_2 | D = d)P(D = d)$$

Maximizing the above equation still gives a MAP estimate. If we decide to ignore the prior, then we are left with the likelihood:

$$P(D_1 = d_1 \,\&\, D_2 = d_2 | D = d)$$

If we find the value of d that maximizes this equation, then we're finding a maximum likelihood (ML) depth estimate. That's what we'll do here.

If the two depth cues are statistically independent, then the above equation is equal to:

$$=\ P(D_1 = d_1 | D = d)\ P(D_2 = d_2 | D = d)$$

We're assuming that the cues are unbiased, normally distributed random variables with standard deviations σ_1 and σ_2, so we can evaluate these probabilities using the normal probability density function (pdf), $\phi(x, \mu, \sigma)$, which in R is dnorm().

$$=\ \phi(d_1, d, \sigma_1)\ \phi(d_2, d, \sigma_2)$$

The value of a normal pdf doesn't change if we switch the first and second arguments:

$$=\ \phi(d, d_1, \sigma_1)\ \phi(d, d_2, \sigma_2)$$

Also, a useful fact (not proven here) is that the pointwise product of two normal pdf's is also a normal pdf.

$$= \phi\left(d_1 - d_2, 0, (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}\right) \phi(d, \mu_3, \sigma_3)$$

where $\mu_3 = w_1 d_1 + w_2 d_2$ with $w_1 = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}$ and $w_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}$

and $\frac{1}{\sigma_3^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$

The first ϕ factor in the above equation is a scale constant that doesn't depend on d, so we only need to find the value of d that maximizes:

$$\propto \phi(d, \mu_3, \sigma_3)$$

Obviously this is maximized at $d = \mu_3$, which the definition of μ_3 above shows is $\mu_3 = w_1 d_1 + w_2 d_2$, so the ML depth estimate is this weighted sum of the two depth cues.

Done!