Cue Coupling

This section describes models for coupling different visual cues.
The ideas in this section are logical extensions of the ideas in the earlier sections. But we are now addressing more complex aspects of vision and so the techniques and the tools become more complex and more abstract as we begin to reason about surfaces, objects, and their relations.

Vision Modules and Cue Combination

Quantifiable psychophysics experiments for individual cues are roughly consistent with the predictions of the types of models discussed in the previous two sections– see [18, 23] – but with some exceptions [160].

But how are different visual cues combined?

The most straightforward manner is to use a separate module for each cue to compute different estimates of the properties of interest, e.g., the surface geometry, and then merge these estimates into a single representation. This was proposed by Marr [109] who justified this strategy by invoking the principle of modular design.

Marr proposed that surfaces should be represented by a *2 1/2D sketch* which specifies the shape of a surface by the distance of the surface points from the viewer. A related representation, *intrinsic images*, also represents surface shape together with the material properties of the surface.

Cue Coupling from a Probabilistic Perspective

We consider the problem of cue combination from a probabilistic perspective [22].

This suggests that we need to distinguish between situations where the cues are statistically independent of each other and the cases where they are not. We need also need to determine whether cues are using similar, and hence redundant, prior information.

These considerations leads to a distinction between *weak* and *strong* coupling, where weak coupling corresponds to the traditional view of modules while strong coupling considers more complex interactions. To understand strong coupling it is helpful to consider the *causal factors* which generate the image.

Note that there is strong evidence that high-level recognition can affect the estimation of three-dimensional shape. E.g., , a rigidly rotating inverted face mask is perceived as non-rigidly deforming face, while most rigidly rotating objects are perceived to be rigid.

Combining Cues with Uncertainty

We first consider simple models which assume that the cues compute representations independently and then combine their outputs by taking linear weighted combinations.

Suppose there are two cues for depth which separately give estimates $\vec{S}_1^*, \vec{S}_2^*$.

One strategy to combine these cues is by linear weighted combination yielding a combined estimate $\vec{S}^*$:

$$\vec{S}^* = \omega_1 \vec{S}_1^* + \omega_2 \vec{S}_2^*,$$

where ω_1, ω_2 are positive weights such that $\omega_1 + \omega_2 = 1$.

Landy and Maloney [91] reviewed many early studies on cue combination and argued that they could be qualitatively explained by this type of model. They also discussed situations where the individual cues did not combine and "gating mechanisms" which require one cue to be switched off.

Case where weights are derived from uncertainties

An important special case of this model is when the weights are measures of the uncertainty of the two cues. This approach is optimal under certain conditions and yields detailed experimental predictions which have been successfully tested in some cases [69, 32], see [21, 48] for exceptions.

If the cues have uncertainties σ_1^2, σ_2^2 we set the weights to be $w_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ and $w_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$. The cue with lowest uncertainly has highest weight.

This gives the linear combination rule:

$$\vec{S}^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\vec{S}_1^* + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\vec{S}_2^*.$$

Optimality of the Linear Combination Rule (I)

The linear combination is optimal for the following conditions.
(I) The two cues have inputs $\{\vec{C}_i : i = 1, 2\}$ and outputs $\vec{S}$ related by conditional distributions $\{P(\vec{C}_i|\vec{S}) : i = 1, 2\}$.
(II) These cues are *conditionally independent* so that $P(\vec{C}_1, \vec{C}_2|S) = P(\vec{C}_1|\vec{S})P(\vec{C}_2|\vec{S})$ and both distributions are Gaussians:

$$P(\vec{C}_1|\vec{S}) = \frac{1}{Z_1} \exp\{-\frac{|\vec{C}_1 - \vec{S}|^2}{2\sigma_1^2}\},$$

$$P(\vec{C}_2|\vec{S}) = \frac{1}{Z_2} \exp\{-\frac{|\vec{C}_2 - \vec{S}|^2}{2\sigma_2^2}\}.$$

(III) The prior distribution for the outputs is uniform.

Optimality of the Linear Combination Rule (II)

In this case, the optimal estimates of the output $\vec{S}$, for each cue independently, are given by the maximum likelihood estimates:

$$\vec{S}_1^* = \arg\max_S P(\vec{C}_1|\vec{S}) = \vec{C}_1, \quad \vec{S}_2^* = \arg\max_S P(\vec{C}_2|\vec{S}) = \vec{C}_2.$$

If both cues are available, then the optimal estimate is given by:

$$\vec{S}^* = \arg\max_{\vec{S}} P(\vec{C}_1, \vec{C}_2|\vec{S}) = \arg\max_{\vec{S}} P(\vec{C}_1|\vec{S})P(\vec{C}_2|\vec{S})$$

$$= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\vec{C}_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\vec{C}_2,$$

which is the linear combination rule by setting $\vec{S}_1^* = \vec{C}_1$ and $\vec{S}_2^* = \vec{C}_2$.

Optimality of Combination Rule: Illustration

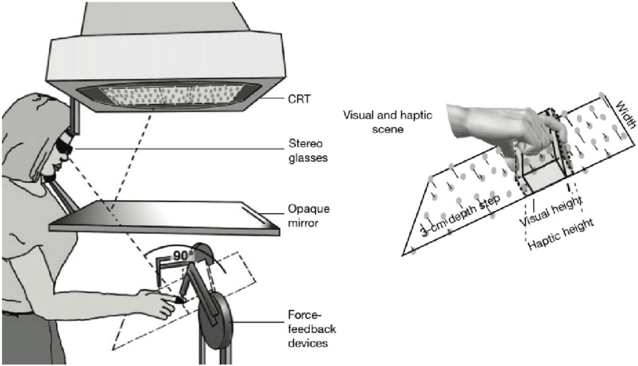

Figure 33: The work of Ernst and Banks shows that cues are sometimes combined by weighted least squares where the weights depend on the variance of the cues. Figure adapted from [32]. The interactive demo (5) illustrates how the cues coupling result depends on the means and variances of each cue.

Bayesian Analysis: Weak and Strong Coupling

We now describe more complex models for coupling cues from a Bayesian perspective [22][185] which emphasizes that the uncertainties of the cues are taken into account and the statistical dependencies between the cue are made explicit.

Examples of cue coupling, where the cues and independent, are called "weak coupling" in this framework. In the likelihood functions are independent Gaussians and if the priors are uniform, then this reduces to the linear combination rule.

By contrast, "strong coupling" is required if the cues are dependent on each other.

The priors: avoiding double counting (I)

Models of individual cues typically include prior probabilities about $\vec{S}$. For example, cues for estimating shape or depth assume that the viewed scene is piecewise smooth. Hence it is typically unrealistic to assume that the priors $P(\vec{S})$ are uniform.
Suppose we have two cues for estimating the shape of a surface which both use the prior that the surface is spatially smooth. Taking a linear weighted sum of the cues will not be optimal, because the prior would be used twice. Priors introduce a bias to perception so we want to avoid doubling this bias.
This is supported by experimental findings [18] where subjects were asked to estimate the orientation of surfaces using shading cue, texture cues, or both combination. If only one cue, shading or texture, was available than subjects underestimated the surface orientation. But human estimates were much more accurate if both cues were present, inconsistent with double counting priors [185].

Avoiding Double Counting: Experiments

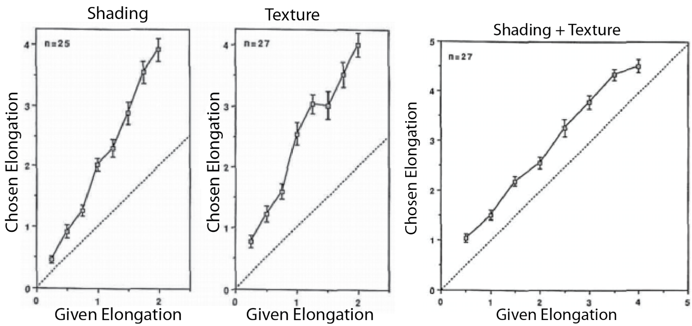

Figure 34: Cue coupling results which are inconsistent with linear weighted average [17]. Left Panel: If depth is estimated using shading cues only then humans underestimate the perceived orientation (i.e. they see a flatter surface). Center Panel: Humans also underestimate the orientation if only texture cues are present. Right Panel: But if both shading and texture cues are available then humans perceive the orientation correctly. This is inconsistent with taking the linear weighted average of the results for each cue separately. Figure adapted from [17].

Avoiding Double Counting: Probabilistic Analysis (I)

We model the two cues separately by likelihoods $P(\vec{C}_1|\vec{S}), P(\vec{C}_2|\vec{S})$ and a prior $P(\vec{S})$. For simplicity ee assume that the priors are the same for each cue. This gives posterior distributions for each visual cue:

$$P(\vec{S}|\vec{C}_1) = \frac{P(\vec{C}_1|\vec{S})P(\vec{S})}{P(\vec{C}_1)}, \quad P(\vec{S}|\vec{C}_2) = \frac{P(\vec{C}_2|\vec{S})P(\vec{S})}{P(\vec{C}_2)}.$$

This yields estimates of surface shape to be $\vec{S}_1^* = \arg\max_{\vec{S}_1} P(\vec{S}|\vec{C}_1)$ and $\vec{S}_2^* = \arg\max_{\vec{S}_2} P(\vec{S}|\vec{C}_2)$.

Avoiding Double Counting: Probabilistic Analysis (II)

The optimal way to combine the cues is to estimate $\vec{S}$ from the posterior probability $P(\vec{S}|\vec{C}_1, \vec{C}_2)$:

$$P(\vec{S}|\vec{C}_1, \vec{C}_2) = \frac{P(\vec{C}_1, \vec{C}_2|\vec{S})P(\vec{S})}{P(\vec{C}_1, \vec{C}_2)}.$$

If the cues are *conditionally independent*, $P(\vec{C}|\vec{S}) = P(\vec{C}_1|\vec{S})P(\vec{C}_2)|\vec{S})$, then this simplifies to:

$$P(\vec{S}|\vec{C}_1, \vec{C}_2) = \frac{P(\vec{C}_1|\vec{S})P(\vec{C}_2|\vec{S})P(\vec{S})}{P(\vec{C}_1, \vec{C}_2)}.$$

Avoiding Double Counting: Probabilistic Analysis (III)

Coupling the cues, using the model in the previous slide, cannot correspond to a linear weighted sum, which would essentially be using the prior twice (once for each cue).

To understand this, suppose the prior is $P(\vec{S}) = \frac{1}{Z_p} \exp\{-\frac{|\vec{S}-\vec{S}_p|^2}{2\sigma_p^2}\}$. Then, setting $t_1 = 1/\sigma_1^2, t_2 = 1/\sigma_2^2, t_p = 1/\sigma_p^2$, the optimal combination is $\vec{S}^* = \frac{t_1\vec{C}_1 + t_2\vec{C}_2 + t_p\vec{S}_p}{t_1+t_2+t_p}$, hence the best estimate is a linear weighted combination of the two cues $\vec{C}_1, \vec{C}_2$ and the mean $\vec{S}_p$ of the prior.

By contrast, the estimate using each cue individually are given by $\vec{S}_1^* = \frac{t_1\vec{C}_1 + t_p\vec{S}_p}{t_1+t_2+t_p}$ and $\vec{S}_2^* = \frac{t_2\vec{C}_2 + t_p\vec{S}_p}{t_1+t_2+t_p}$.