Computational Vision
U. Minn. Psy 5036
Daniel Kersten
Lecture 22: Cooperative Computation

Initialize

- **Spell check off**

```
Off[General::spell1];
```

Outline

Last time

Statistical texture models & synthesis, MRFs, Gibbs Sampler

Science writing

Today

Integrating perceptual information

Some basic graph types in vision (from Lecture 6)

See: Kersten, D., & Yuille, A. (2003) and Kersten, Mamassian & Yuille (2004)

■ Basic Bayes

$$p[S \mid I] = \frac{p[I \mid S]\, p[S]}{p[I]}$$

Usually, we will be thinking of the **Y** term as a random variable over the hypothesis space, and **X** as data. So for visual inference, **Y = S** (the scene), and **X = I** (the image data), and **I = f(S)**.

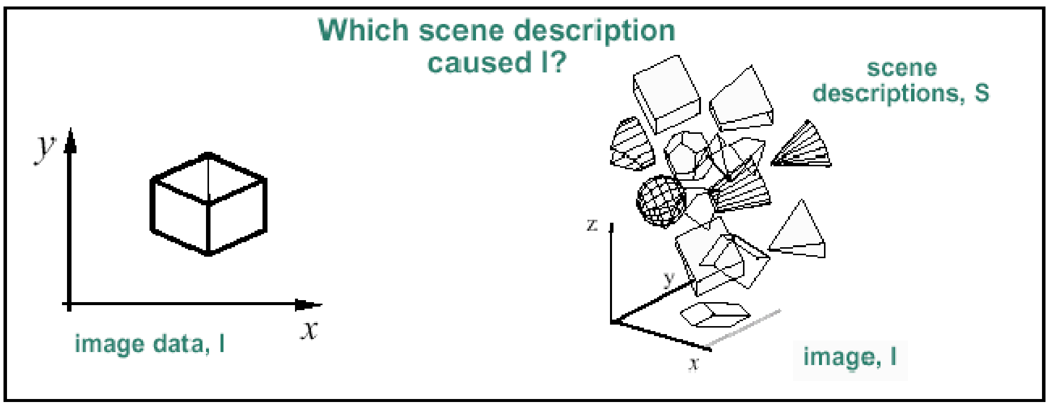
We'd like to have:

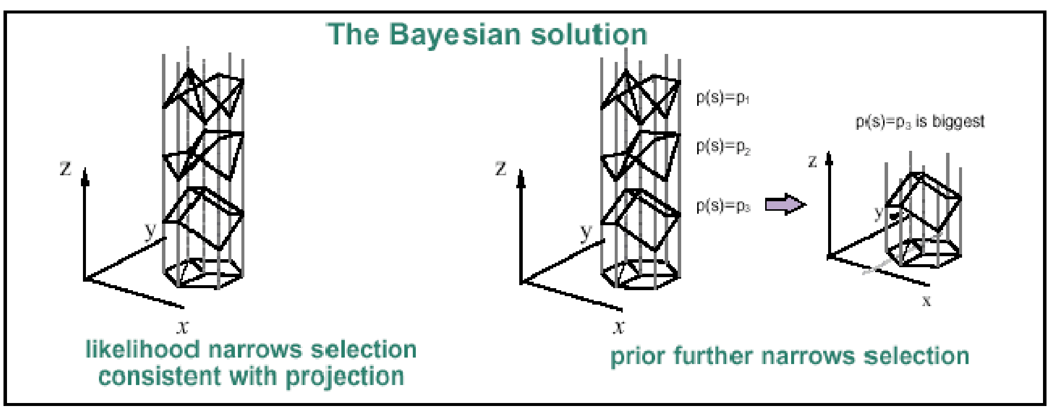
p(S|I) is the **posterior** probability of the scene given the image

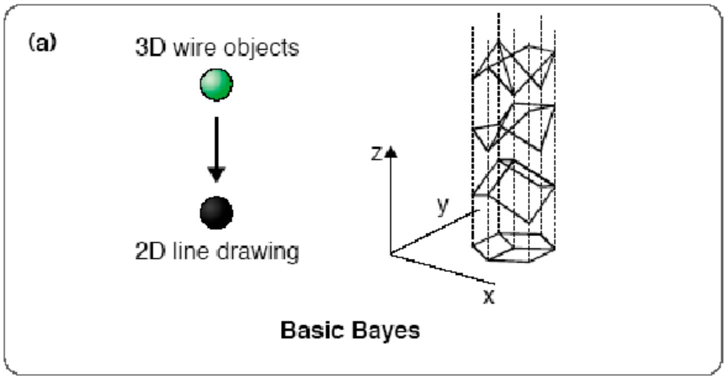
-- i.e. what you get when you condition the joint by the image data. The posterior is often what we'd like to base our decisions on, because as we discuss below, picking the hypothesis **S** which maximizes the posterior (i.e. maximum a posteriori or **MAP** estimation) minimizes the average probability of error.

p(S) is the **prior** probability of the scene.

p(I|S) is the **likelihood** of the scene. Note this is a probability of **I**, but not of **S**.

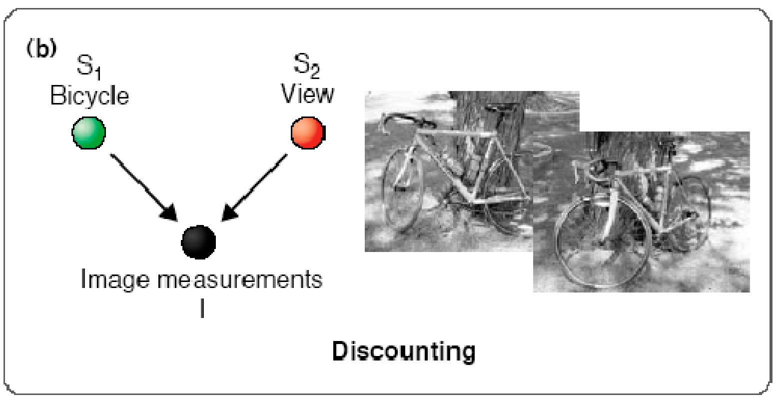






We've seen a number of applications of Basic Bayes, including the algorithms for shape from shading and optic flow.

Discounting



This Bayes net describes the case where the joint distribution can be factored as:

$p(s_1, s_2, I) = \mathrm{p}(\mathrm{I}|s_1, s_2)\mathrm{p}(s_1)\mathrm{p}(s_2)$

Optimal inference for this task requires that we calculate the marginal posterior:

$\mathrm{p}(s_1|\mathrm{I}) \propto \int_{S_2} p(s_1, s_2 \mid I)\, d s_2$

Liu, Knill & Kersten (1995) describe an example with:

I -> 2D x-y image measurements, s_1 -> 3D object shape, and s_2 -> view

Bloj et al. (1999) have an example estimating s_1 -> surface chroma (saturation) with s_2 -> illuminant direction.

In this lecture, we describe two more simple graphs.

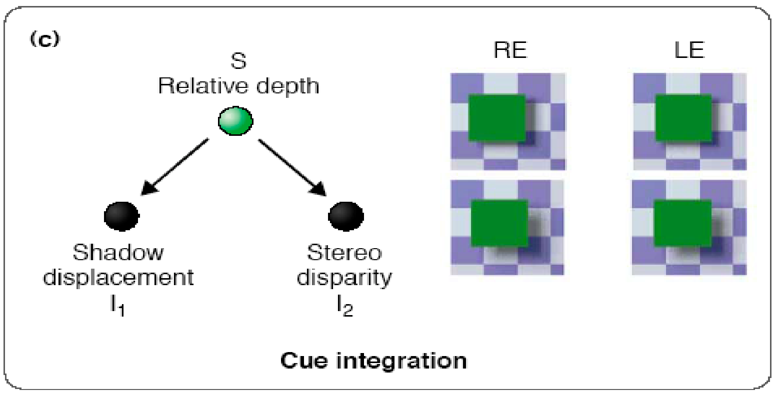
■ Cue integration

■ Perceptual explaining away

Cue integration

■ Strong vs. weak fusion

Clark & Yuille, Landy & Maloney, Knill & Kersten, Schrater & Kersten.



This Bayes net describes the factorization:

p(S,I1,I2) = p(I1|S)p(I2|S)p(S)

We'll change notation, and let x_1 and x_2 be image measurements or cues. The simple Bayes net shown above describes the case where the two cues are conditionally independent. In other words, p(x1,x2|s) = p(x1|s)p(x2|s).

Let's consider the simple Gaussian case where $x_i = \mu_{\text{cue } i} + n_i$. We'll show that optimal combined cue estimate is a weighted average of the cues.

p(s|x1,x2)=p(x1,x2|s)p(s)/p(x1,x2) ∝ p(x1|s)p(x2|s) $= e^{-(x_1-s)^2/2\sigma_1{}^2} e^{-(x_2-s)^2/2\sigma_2{}^2}$

PowerExpand$\left[\text{Log}\left[E^{-(x_1-\mu)^2/(2\sigma_1{}^2)} E^{-(x_2-\mu)^2/(2\sigma_2{}^2)}\right]\right]$

$$-\frac{(x_1-\mu)^2}{2\sigma_1^2} - \frac{(x_2-\mu)^2}{2\sigma_2^2}$$

$$D\left[-\frac{(x_1-\mu)^2}{2\,\sigma_1^2}-\frac{(x_2-\mu)^2}{2\,\sigma_2^2},\ \mu\right]$$

$$\frac{x_1-\mu}{\sigma_1^2}+\frac{x_2-\mu}{\sigma_2^2}$$

$$\text{Solve}\left[\frac{x_1-\mu}{\sigma_1^2}+\frac{x_2-\mu}{\sigma_2^2}==0,\ \mu\right]$$

$$\left\{\left\{\mu\to\frac{x_2\,\sigma_1^2+x_1\,\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right\}\right\}$$

$$\left\{\left\{\mu\to\frac{x_2\,\sigma_1^2+x_1\,\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right\}\right\}\ /.\ \{\sigma_1^2\to 1/r_1,\ \sigma_2^2\to 1/r_2\}$$

$$\left\{\left\{\mu\to\frac{\frac{x_1}{r_2}+\frac{x_2}{r_1}}{\frac{1}{r_2}+\frac{1}{r_1}}\right\}\right\}$$

where $r_i\left(=\frac{1}{\sigma_i{}^2}\right)$, is called the reliability.

$$\mu\to\frac{r_1\,x_1}{r_1+r_2}+\frac{r_2\,x_2}{r_1+r_2}$$

In general, one can show that the combined estimate is the weighted sum of the separate estimates, where the weights w_i are determined by the relative reliabilities :

$$\hat{\mu}_{\text{combined}}=\hat{\mu}_{\text{cue1}}\,w_1+\hat{\mu}_{\text{cue2}}\,w_2=\hat{\mu}_{\text{cue1}}\,\frac{r_1}{r_1+r_2}+\hat{\mu}_{\text{cue2}}\,\frac{r_2}{r_1+r_2}.$$

Perceptual explaining away, Cooperative computation

Modular vs. cooperative computation

Kenneth Craik at Cambridge University in the 1940's (1943) had suggested that perception was analogous to an engineer's construction of model of a ship. It was like the real thing with respect to tests it was subjected to, but left out the inessential details.

The central theme of this course is that a major challenge of vision research is to understand how the brain constructs a model of the visual environment from the pattern of changing retinal light intensities. A brightness change in our eyes is translated to an impression of transparency, shadow, depth, or shape with no apparent effort.

A primary result of computational analysis is that scene reconstruction from image data is often underconstrained--there are many solutions that satisfy the data. Prior constraints then have to be sought to find a unique interpretation of the environment from the image intensities. Regularization theory is one way of modeling the interaction of data, and prior constraints. Bayesian theory takes us step further by showing the importance of marginalization in dealing with secondary variables as a means to reduce ambiguity in the primary variables.

So far, we've primarily studied modular theories of visual estimation, such as, surface-color-from-radiance (Land, 1959), shape-from-shading (Horn, 1975), optic flow (Hildreth, 1983) or structure-from-motion (Ullman, 1979).

But the assumption of modularity depends on the task (Schrater and Kersten,). Further, if there are several classes of primary variables, then they all should be estimated in such a way that they are consistent with the image data. Can one relax prior assumptions, about specific domains, without losing uniqueness? The answer pursued here is to go beyond simple modularity, and look at how interacting modules or representations of the scene determine image information.

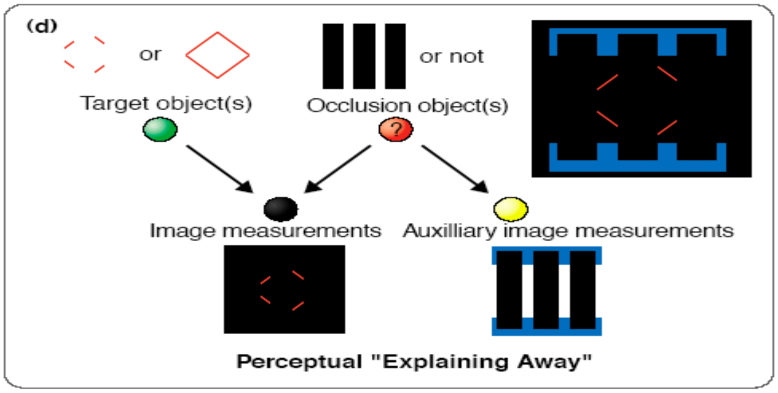
In contrast to the modularity theories of vision, it is phenomenally apparent that visual information is eventually integrated to provide a strikingly singular description of the visual environment. Visual ambiguity is the exception, rather than the rule. By looking at how human perception puts integrates scene attributes, we may get some idea of how vision modules in the brain interact, and what they represent.

Cooperative computation: multiple estimations of scene attributes, the estimates of which satisfy an internal model of consistency

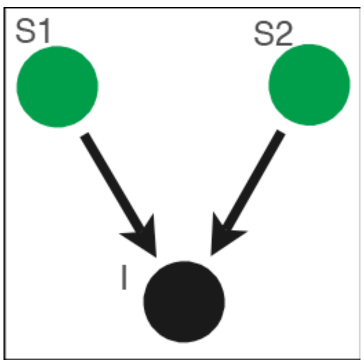
■ Perception as puzzle solving

Rock, I. (1983). *The Logic of Perception*. Cambridge, Massachusetts: M.I.T. Press.

■ Perceptual explaining away



Both causes S1 and S2 can be primary variables.



The above Bayes net describes the factorization:

p(S1,S2,I) = p(I|S2,S2) p(S1)p(S2)

If we average over I, S1 and S2 are independent. However, knowledge of I makes S1 and S2 conditionally dependent. The two causes S1 and S2 can behave like competing hypotheses to explain the data I.

In general, "explaining away" is a phenomenon that occurs in probabilistic belief networks in which two (or more) variables influence a third variable whose value can be measured (Pearl, 1988). Once measured, it provides evidence to infer the values of the influencing variables.

Imagine two coins that can be flipped independently, and the results (heads or tails) have an influence on a third variable. For concreteness, assume the third variable's value is 1 if both coins agree, and 0 if not (NOT-XOR). If we are ignorant of the value of the third variable, knowledge of one influencing variable doesn't help to guess the value of the other—the two coin variables are independent. (This is called marginal independence, "marginal" with respect to the third variable.)

But if the value of the third variable is measured (suppose it is 1), the two coin variables become coupled, and they are said to be *conditionally dependent*. Now knowing that one coin is heads guarantees that the other one is too.

The phrase "explaining away" arises because coupling of variables through shared evidence arises often in human reasoning, when the influences can be viewed as competing causes. Suppose that the evidence is 0. If our interpretation is that "heads" in either coin can cause such a "suppression" of the NOT-XOR output, then which coin did the suppressing? One of the coins is heads and one tails, but not both. Any auxiliary evidence that tips the balance toward one coin being "to blame", reduces our belief that the other caused the observed 0. The other coin's possible influence is explained away by the new evidence supporting the true-culprit coin’s value of heads. Human reasoning is particularly good at these kinds of inferences.

“Explaining away” is also a characteristic of perceptual inferences, for example when there are alternative perceptual groupings consistent with a set of identical or similar sets of local image features.

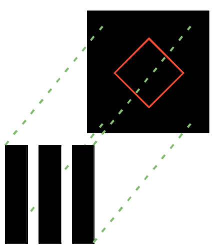
Demonstrations of cooperative computation in perception

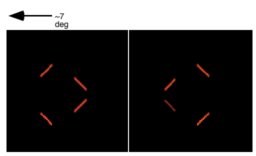
Several perceptual phenomena that we've seen before can be interpreted as "explaining away".

Occlusion & motion: Lorenceau & Shiffrar, Sinha

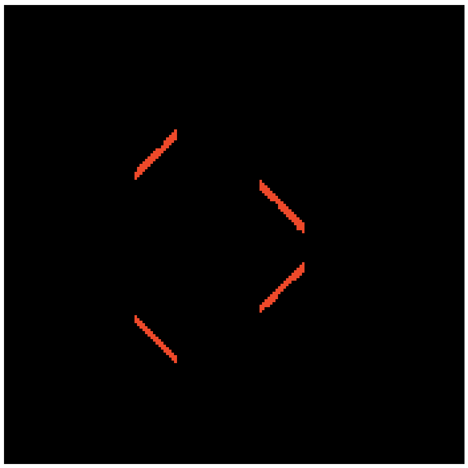
Recall translating diamond used to illustrate the aperture problem.

When the diamond is seen as coherently translating, one often also interprets the vertices as being covered by rectangular occluders.

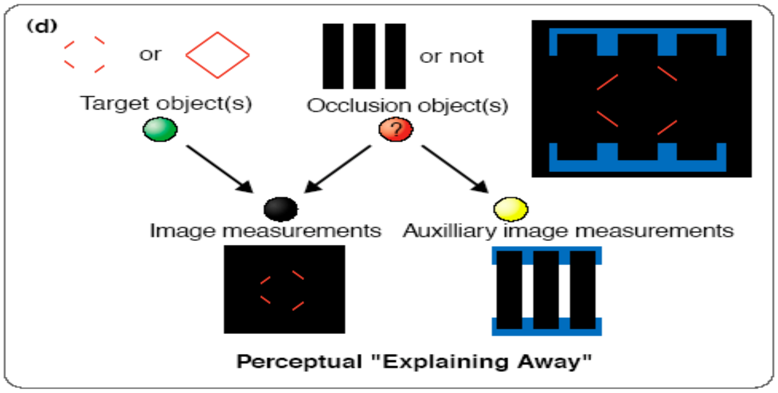




■ **Translating diamond with "occluding occluders"**

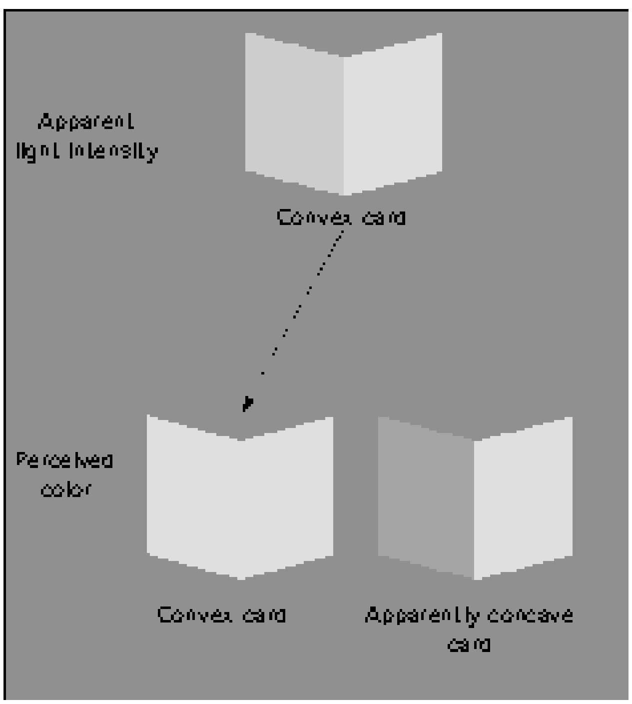


Here is one way of viewing occlusion as explaining away:



Lightness & surface geometry

■ Mach card

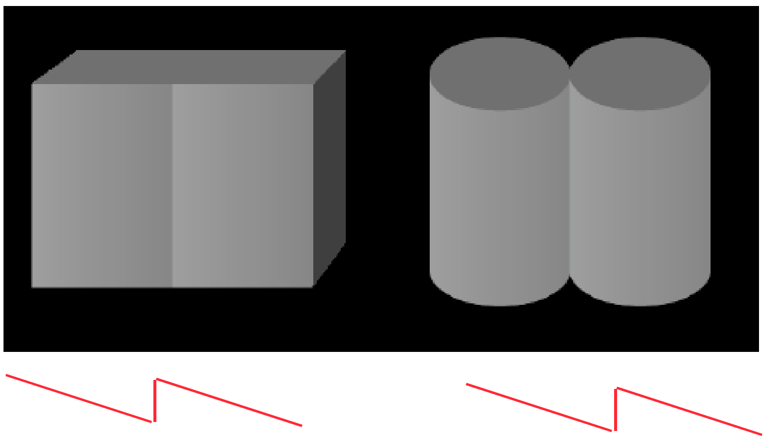


■ Lightness and shape

Recall the lightness demonstration that is similar to the Craik-O'Brien-Cornsweet effect, but difficult to explain with a simple filter mechanism (Knill, D. C., & Kersten, D. J., 1991). The idea is that the lightness of a pair of luminance gradients on the left of the figure below look different, whereas they look similar for the pair luminance gradients on the right. The reason seems to be due to the fact that the luminance gradients on the right are attributed to smooth changes in shape, rather than smooth changes in illumination.

http://vision.psych.umn.edu/www/kersten-lab/demos/lightness.html

These demonstrations suggest the existence of scene representations in our brains for shape, reflectance and light source direction.



Draw a diagram to illustrate the above illusion in terms of "explaining away"

■ Dependence of lightness on spatial layout

Gilchrist:

In the 1970's, Alan Gilchrist was able to show that the lightness of a surface patch may be judged either dark-gray, or near-white with only changes in perceived spatial layout! (Gilchrist, A. L. (1977). How did he do this? What is going on? Interpret lightness as reflectance estimation.

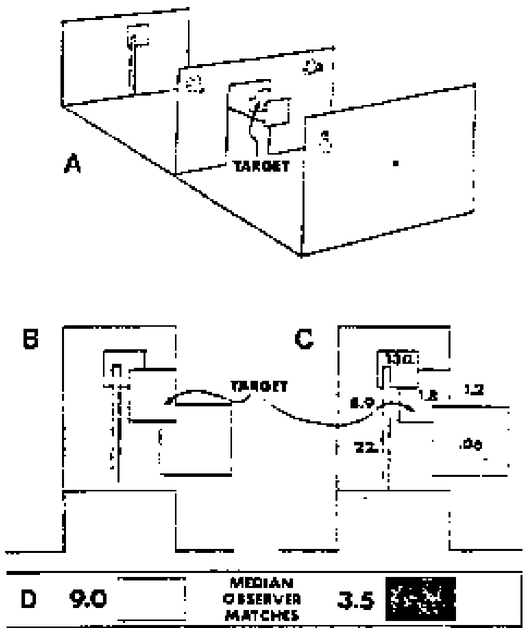


Figure 5. (A) Perspective view of the parallel planes display, showing hidden light bulbs. The display (as seen through the pinhole) in which the target appeared to be located either (B) in the near plane or (C) in the far plane, with luminances shown in foot-Lamberts. (D) The average match from a Munsell chart for the two displays.

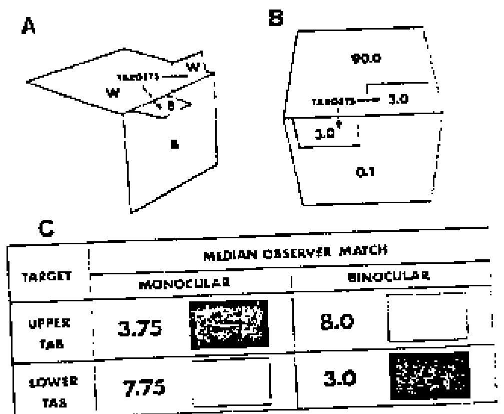


Figure 4. (A) Perspective view of the stimulus display used in the critical test, showing color (B, black; W, white) of each part. (B) Monocular retinal pattern showing luminances in foot-Lamberts. (C) Average Munsell matches for monocular and binocular viewing conditions.

o The Room-in-a-Shoe-Box experiment

o Coplanar card experiment

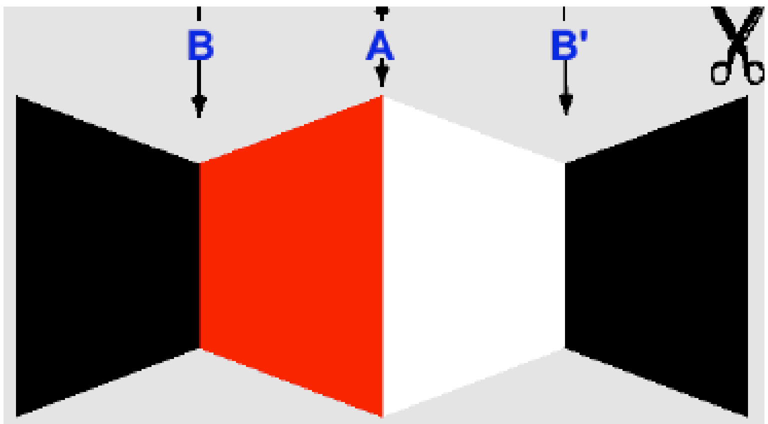
The left and right inner gray disks in the above figure are the same intensity. In classic simultaneous contrast, the brighter annulus on the right makes the inner disk appear darker.

Color & shape

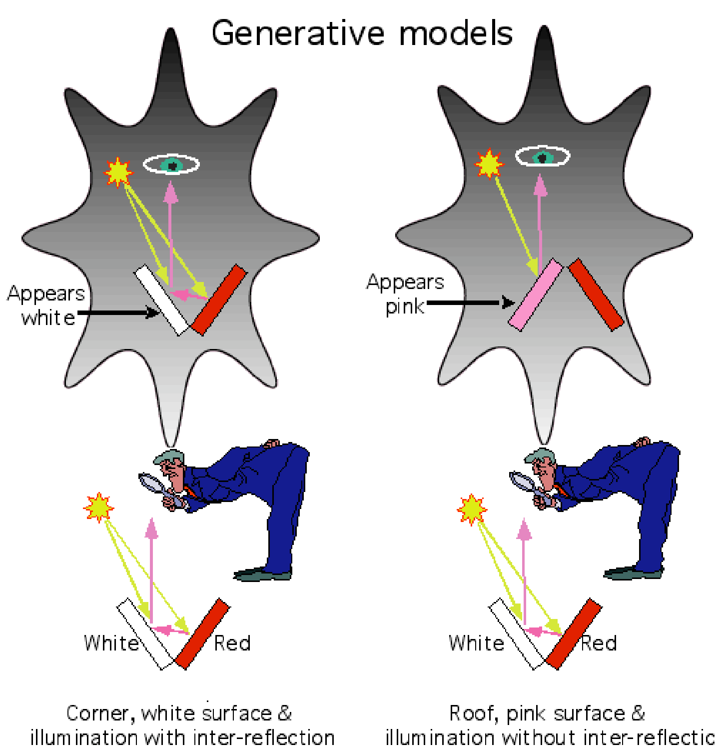
■ Bloj, Kersten & Hurlbert

Demo

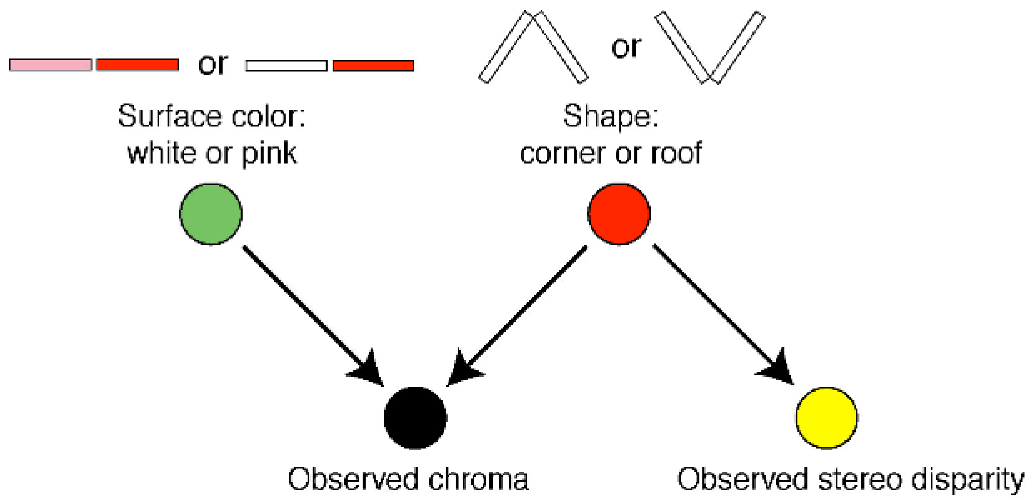
http://vision.psych.umn.edu/www/kersten-lab/Mutual_illumination/BlojKerstenHurlbertDemo99.pdf



Interpretation



Stereo can be used as an auxiliary cue to change the perceived shape from concave to convex.

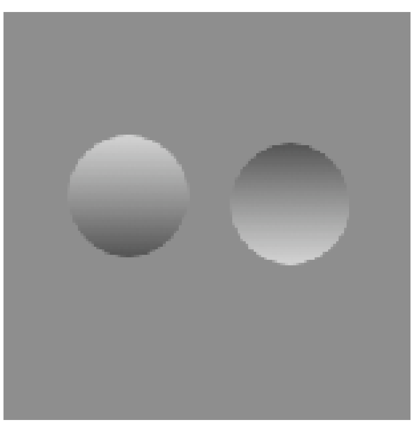


Dependence of shape on perceived light source direction

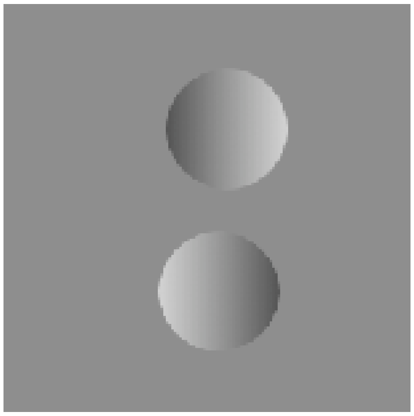
Dependence of shape on perceived light source direction

Brewster (1926), Gibson, Ramachandran, V. S. (1990), crater illusion and the single light source assumption

■ Vertical light direction



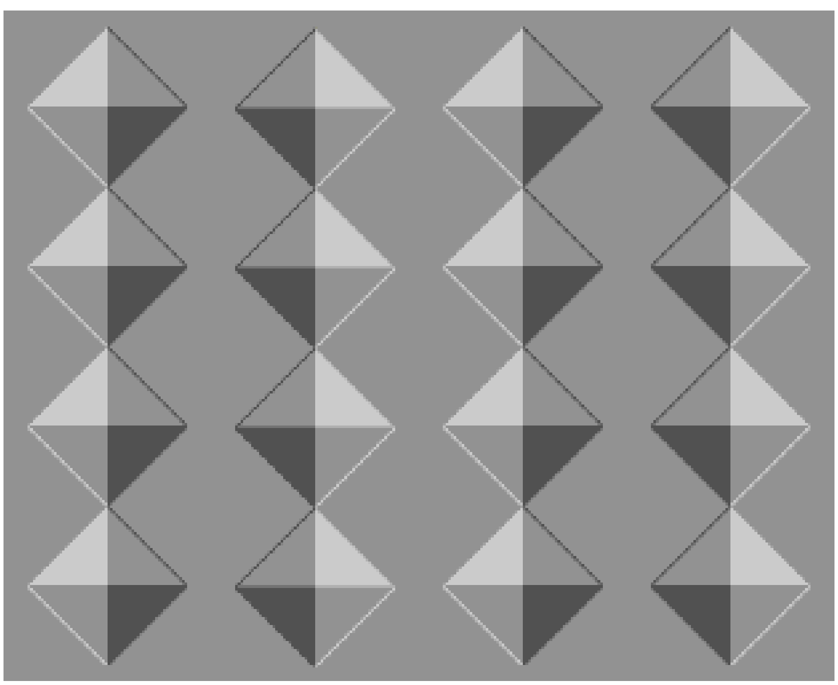
■ Horizontal light direction



Transparency

■ Transparency & lightness

Argyle illusion (Adelson. Also, see Wandell's text).



■ Motion and transparency (Kersten et al., 1992)

Dependence of transparency on perceived depth

Kersten and Bülthoff

o orientation and transparency

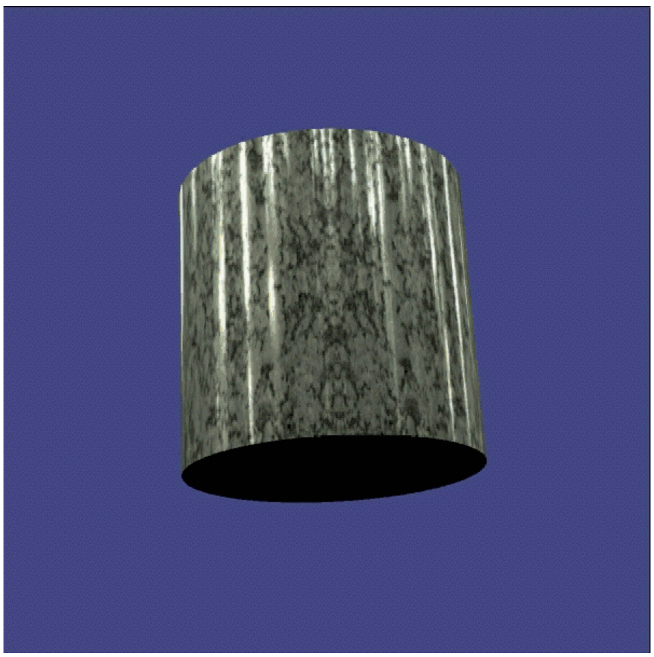
o transparency and depth from motion--computer demo

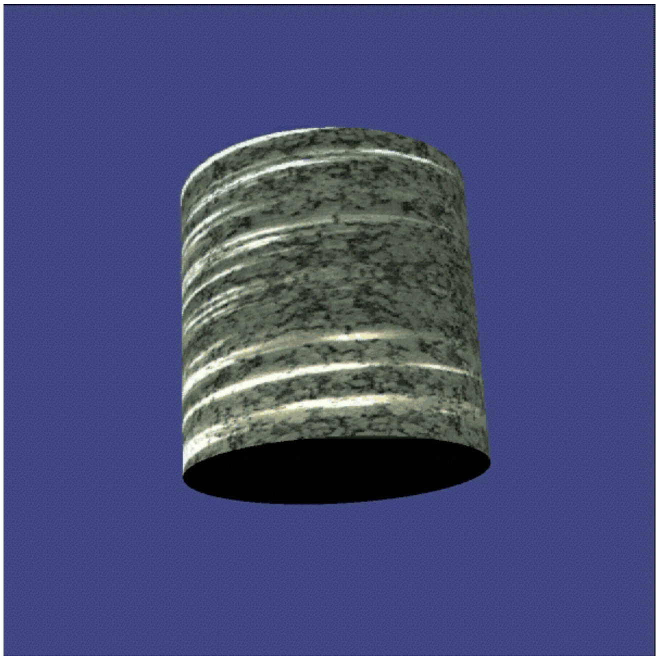
http://vision.psych.umn.edu/www/kersten-lab/demos/transparency.html

Nakayama, Shimojo (1992)

o transparency and depth from stereo demos, neon color spreading

Dependence on curvature





Computational theory

Overview

■ Main problem: ***There is no single local cue to edge identity***

The studies of lightness perception led some psychologists to suggest that the brain constructed specific representations that were estimated representations of reflectance, light source and shape (Bergstrom, 1977; Gilchrist, 1977). Around the same time, two computer vision researchers presented the idea of computation of multiple "intrinsic images" (Barrow, H.G. and Tenenbaum, J.M., 1978)

The forward problem (or generative model): I(x,y) = f(reflectance(x,y), illumination(x,y), orientation(x,y),...) suggests that ambiguity could be resolved by computing multiple spatial maps: reflectance(x,y), illumination(x,y), orientation(x,y), Various visual cues (derived from I(x,y)) provide soft evidence for the type of edge, and the maps all work together to be consistent with the image I(x,y).

o Edge and surface attribute labelling

-types of edges

reflectance,

illumination (shadow),

depth,

orientation (self-occlusion)

Physiological evidence for surface maps? Not really. Some hint of spatial surface interpolation, but not easy to find in single-unit activity (von der Heydt, 2003). However, there is physiological evidence for feature maps, but these may be sparse--e.g. Swindale. What might be the relationship to efficient coding?

■ *Use* of intermediate-level representation (i.e. of "attribute layers" or intrinsic images)?

Different goals of the organism require different types of edge information to be made explicit. When the edges correspond to object boundaries, this is the segmentation problem.

Examples are:

-Stereo/texture based dense surface reconstruction is more reliable with tokens based on surface markings

(e.g. shadows OK, but not specularities).

-Shape-based object recognition makes use of surface depth and orientation discontinuities

- Object identification and classification makes use of surface attributes, not raw image attributes

(e.g. red vs. green apple; specular or matte).

- Time-to-contact, direction of heading require geometric rather than material estimates.

■ Computational problems

Computational problems?

o Integration vs. cooperativity

Prior constraints

- Expressing prior constraints on interactions

reflectance and depth edges often coincide

reflectance and shadow edges rarely coincide

depth and shadow edges rarely coincide

Problem highly non-linear. Regularization/cost function/prior proability solutions involve convoluted topographies, e.g. when searching for maximum modes in the posterior.

■ Bayesian approaches to cooperative computation

Many of the modular problems of early vision can be interpreted in the framework of Bayesian statistical inference. Problems of cooperative computation can also be framed in this way. Let **s**1, **s**2, ..., **s**n be scene descriptions (e.g. vectors corresponding to shape, reflectance, and illumination) that determine the image, **i**. A MAP approach would be to find **s**1, **s**2, ..., **s**n that account for the image. Again, we would make use of prior as well as posterior constraints to maximize:

The prior probability, which is a joint probability of the various scene attributes is much easier to express if each of the scene attributes are statistically independent.

For example of this approach to transparency perception using Markov Random Fields, see Kersten (1991).

Mixtures of experts. Jacobs et al. , 1991.

Competitive priors. Yuille and Bülthoff (1996).

Graphical models. In general, we can't just assume independence. We've seen how recent work in Bayes nets and graphical models provides methods to analyze the causal or influence relationships between the variables in a complex inference problems (Pearl, 1988; Ripley, 1996).

■ Incorporating higher-level knowledge--Image parsing

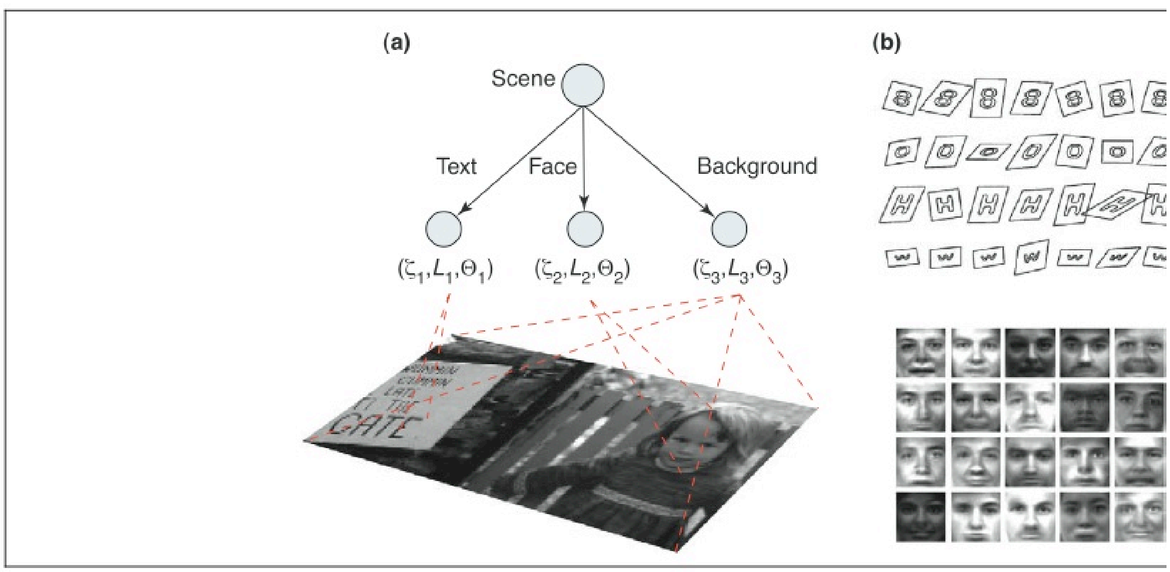


Figure 3. (a) The image is generated by a probabilistic context-free grammar, shown by a two-layer graph with nodes that have prope (z,l,q) in the image. (b) Samples from the face model and the letter model, that is, from pð$_{RaH}$jz;L;qÞ.

For current work in this area applied to segmentation, see: Tu Z, Zhu S-C (2002), Zhu and Tu (2000). For a review, see: Yuille and Kersten (2006).

How complete a reconstruction?

One of the most important topics in computational vision is understanding the short-cuts that vision uses in order to take into account other constraints, such as limited information processing resources in terms of neural and temporal constraints. We've seen examples of change-blindness before, but take a look again. Does human vision compute a rich description of the world at each instant, or does it extract only what it needs from the world out there for the tasks at hand?

■ "Change blindness"

http://www.cs.ubc.ca/~rensink/flicker/index.html

References

■ Cue integration, cooperative computation

Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering Intrinsic Scene Characteristics from Images. In A. R. Hanson, & E. M. Riseman (Ed.), Computer Vision Systems (pp. 3-26). New York: Academic Press.

Bergstrom, S. S. (1977). Common and Relative Components of Reflected Light as Information About the Illumination, Colour and Three-Dimensional Form of Objects. 18, 180-186).

Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science, 4*, 99-108.

Clark, J. J., & Yuille, A. L. (1990). Data Fusion for Sensory Information Processing . Boston: Kluwer Academic Publishers.

Ernst MO, Banks MS, Bulthoff HH (2000) Touch can change visual slant perception. Nat Neurosci 3:69-73.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.

Gilchrist, A. L. (1977). Perceived Lightness Depends on Perceived Spatial Arrangement. Science, 195, 185-187.

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston, MA: Houghton Mifflin.

Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion within, but not between, senses. Science 298:1627-1630.

Humphrey, K. G., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., & Menon, R. S. (1996). Differences in Perceived Shape from Shading Correlate with Activity in Early Visual Areas.,, 1-16.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3, 79-87.

Jacobs RA (2002) What determines visual cue reliability? Trends Cogn Sci 6:345-350.

Kersten, D. J. (1991). Transparency and the Cooperative Computation of Scene Attributes. In M. Landy, & A. Movshon (Ed.), Computational Models of Visual Processing. Cambridge, Massachusetts: M.I.T. Press.

Kersten, D., Bülthoff, H. H., Schwartz, B., & Kurtz, K. (1992). Interaction between transparency and structure from motion. Neural Computation, 4(4), 573-589.

Kersten D, Yuille A (2003) Bayesian models of object perception. Current Opinion in Neurobiology 13:1-9.

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian Inference. Annual Review of Psychology 55:271-304.

Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. Nature, 351, 228-230.

Mach, E. (*1886*, 1959). The Analysis of Sensations . New York: Dover.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. Science, 257, 1357-1363.

Pearl J (1988) Probabilistic reasoning in intelligent systems : networks of plausible inference, Rev. 2nd printing. Edition.

San Mateo, Calif.: Morgan Kaufmann Publishers.

Poggio, T., Gamble, E. B., & Little, J. J. (1988). Parallel integration of vision modules. Science, 242, 436-440.

Ramachandran, V. S. (1990). Visual perception in people and machines. In A. Blake, & T. Troscianko (Ed.), A.I. and the Eye John Wiley & Sons Ltd.

Ripley, B. D. (1996). Pattern Recognition and Neural Networks . Cambridge, UK: Cambridge University Press.

Schrater PR, Kersten D (2000) How optimal depth cue integration depends on the task. *International Journal of Computer Vision* 40:73-91.

Todd, J. T., & Mingolla, E. (1983). Perception of Surface Curvature and Direction of Illumination from Patterns of Shading. *Journal of Experimental Psychology: Human Perception & Performance, 9*(4), 583-595.

Tu Z, Zhu S-C (2002) Parsing Images into Region and Curve Processes. In: Proc. of the 7th European Conference on Computer Vision, p 393 ff. Copenhagen, Denmark: Springer-Verlag, Berlin Heidelberg.

von der Heydt R, Friedman H, Zhou HS (2003) Searching for the neural mechanisms of color filling-in. In: Filling-in: From Perceptual Completion to Cortical Reorganization (Pessoa L, P DW, eds), pp 106-127. Oxford: Oxford University Press.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In K. D.C., & R. W. (Ed.), Perception as Bayesian Inference Cambridge, U.K.: Cambridge University Press.

Zhu S-C, Zhang R, Tu Z (2000) Integrating Bottom-up/Top-Down for Object Recognition by Data Driven Markov Chain Monte Carlo. In: Proc. of Int'l Conf. on Computer Vision and Pattern Recognition. SC.

■ Change blindness

Ballard, D.H., (1996) On the function of visual representation. In K. Akins ed. Perception. New York & Oxford: OUP.

Ballard, D.H., Hayhoe, M.M., Pook, P.K.& Rao, R.P.N. (1997) Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences. 20:4, 723-767.

McConkie, G.W. and Currie, C.B. (1996) Visual stability across saccades while viewing complex pictures. J. Exp. Psychol. Hum. Percept. Perform. 22, 563-581.

McConkie, G.W., & Zola, D. (1979) Is visual information integrated across successive fixations in reading? Perception & Psychophysics, 25, 221-224.

O’Regan, J.K. (1992) Solving the "real" mysteries of visual perception: the world as an outside memory. Canadian journal of psychology. 46: 461-488.

O’Regan, J.K., Rensink, R.A. & Clark, J.J. (1996) "Mud splashes" render picture changes invisible. Invest. Ophthalmol. Vis. Sci. 37 S213.

O’Regan, J.K., Rensink, R.A., & Clark, J.J. (1997) Picture changes during blinks: Not seeing where you look and seeing where you don’t look. Invest. Ophthalmol. Vis. Sci. 38, S707.

O’Regan, J.K., Rensink, R.A., & Clark, J.J. (1999) Change-blindness as a result of ‘mudsplashes.’ Nature 398: 34.

O’Regan, J.K., Deubel, H., Clark, J.J. & Rensink, R.A. (this issue) Picture changes during blinks: looking without seeing and seeing without looking. Visual Cognition.

Rensink, R.A., O’Regan, K. & Clark, J., (1997) To see or not to see: the need for attention to perceive changes in scenes. Psychological Science, Vol 8, No 5, pp 368-373.

Simons, D.J. & Levin D.T., (1997) Change blindness. Trends in cognitive sciences. 1, 7: 261-267.

Simons, D.J., & Levin D.T. (1998) Failure to detect changes to people during real-world interaction. Psychonomic Bulletin and Review. 4:5, 644-649.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci, 10(7), 301-308. http://gandalf.psych.umn.edu/~kersten/kersten-lab/papers/yuillekerstenTICs2006.pdf