15

Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. pp. 209-228. In Landy, Michael S., and J. Anthony Movshon, eds. Computational models of visual processing. MIT press.

Transparency and the Cooperative Computation of Scene Attributes

Daniel Kersten

Cooperativity and Integration

A major goal of vision research is to understand how the brain constructs a perceptual model of the visual environment from the pattern of changing retinal light intensities. Even in the absence of motion, a pattern of intensities in a natural image is interpreted as due to changes in material, illumination, depth, or viewpoint with no apparent effort. How is this done and what are the computational principles involved? It will be argued that an essential component of perception under natural viewing conditions is the process of finding multiple representations of scene characteristics that must be consistent with each other. The perception of transparency provides a simple example of computing multiple and consistent representations and is the focus of the rest of this chapter.

In general, computational vision research has modularized the problem of computing perceptual scene models from image data. Examples include surface-color-from-radiance (Land, 1959), shape-from-shading (Horn, 1975), and structure-from-motion (Ullman, 1979). A primary result of computational analysis is that scene reconstruction from image data is often underconstrained—there are many solutions that satisfy the data provided by the image. Prior constraints, such as assuming that surfaces or reflectances are smooth, then have to be sought to find a unique interpretation of the environment from the image intensities. Although strong constraints may be required for impoverished viewing conditions, in general one would like to relax prior assumptions without losing uniqueness. It is useful to distinguish two strategies—*integration* and *cooperativity*—that become useful when the image content becomes complex as under natural viewing conditions.

Integration refers to the combination of input information or cues pertaining to a particular scene attribute, such as depth at a point, from a variety of sources such as

motion, stereo, and shading as shown in the top of figure 15.1 (Bülthoff & Mallot, 1987; Chou & Brown, 1988; Terzopoulos, 1986). In cooperative computation, the estimates of two or more scene attributes (such as depth and transparency) are required to be consistent with each other and with constraints on natural imaging (bottom of figure 15.1). A scene attribute can be represented as a spatially indexed map or "intrinsic image" (Barrow & Tenenbaum, 1978). One approach to computing scene attribute maps is to label edges in an image according to the cause in the scene (Poggio, Gamble & Little, 1988). The computer detection of useful edges in natural images has turned out to be more problematic than at first anticipated. In addition to coping with multiple spatial scales and image noise, a useful edge detection must ultimately label edges according to the source in the scene. Changes in image intensity can be due to scene discontinuities such as shadows, surface self-occlusion, occlusion of one surface by another, reflectance change, and texture change. For example, if one primary goal of vision is object recognition, the explicit representation of surface boundaries and orientation discontinuities may be particularly

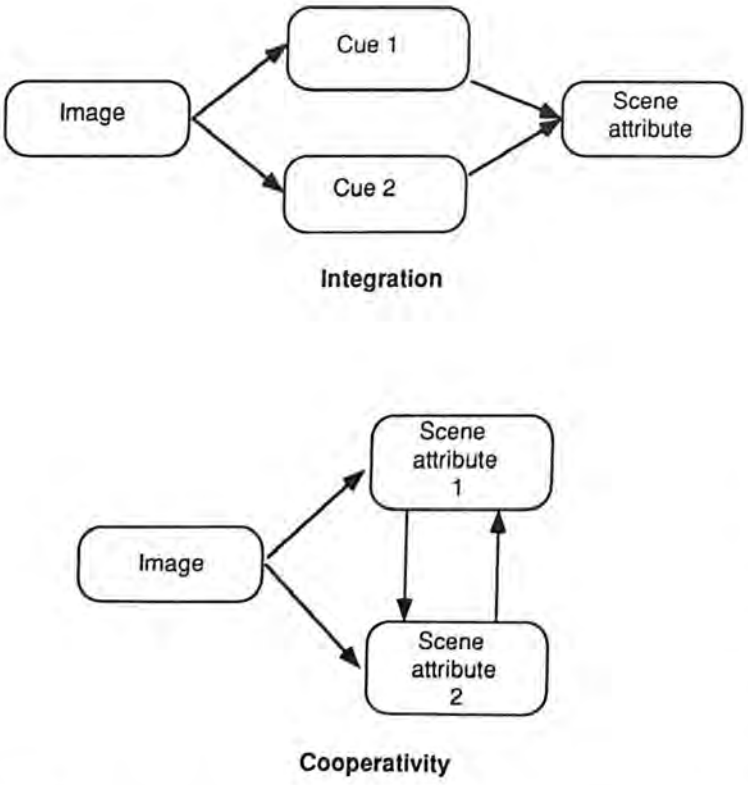


Fig. 15.1
The distinction between integration of cues to compute a single scene attribute representation, and cooperativity—the computation of several scene attribute representations that interact.

critical (Biederman, 1987). But there are many changes of intensity in natural images that are not useful for finding object boundaries. Examine a good line drawing made by an artist—almost all the edges represent object or material boundaries, and very few represent shadows. An artist makes complex inferences in order to filter out uninformative edges while sketching.

A classic example of cooperative computation that may involve edge detection and labeling is the problem of lightness constancy. One approach is to classify edges according to whether they are reflectance or illumination changes (Gilchrist, Delman & Jacobsen, 1983). Although we now have a good understanding of the kinds of algorithms that can filter out slow illumination changes from sharp reflectance changes (Grossberg & Todorović, 1988; Horn, 1973; Land, 1959), the problem of image factoring, when both intrinsic images have discontinuities, is unsolved. There is very little difference in the human estimation of reflectance in complex Mondrians when they are behind a fuzzy shadow vs. a sharp transparency (Plummer & Kersten, 1988). This suggests that at some level, there may be common principles underlying splitting an image into representations of reflectance and illumination, or reflectance and transmittance. Understanding the interaction between two such maps is only a start. Perceptual observations have shown that just two such interacting maps are in general insufficient to account for lightness perception. Ernst Mach showed over a century ago that the perceived surface lightness of a simple folded gray card, placed on a table, depends on the interaction between perceived light source direction, and the bistably perceived geometry of the card, that is, whether it appears convex or concave (Mach, 1886). One needs to take into account the cooperative interaction of shape, illumination, and material reflectance to arrive at a complete account of lightness phenomena.

Studies of lightness perception have been plagued to some extent by ambiguity in both definition and psychophysical results (Beck, 1972). The root of the problem may be that perception of lightness and brightness depends on the extent to which there is an accompanying unambiguous perception of a real surface. This depends on the state of the observer and the naturalness of the image. One way of getting around this problem is to require reflectance estimates, rather than lightness judgments (Arend & Goldstein, 1987). This raises objections that reflectance judgments tap into inferential mechanisms

that are more part of cognition than perception. In this chapter I attempt to circumvent these issues by investigating the perception of surface transparency. Phenomenal transparency may be a less ambiguous percept than lightness by virtue of the fact that it is more tightly coupled to surface perception.

The next two sections of this chapter will treat first the perception and then the computation of transparency. The first section provides perceptual evidence that the problem of transparency, or "seeing through" one cause of image intensity to another, involves cooperative computation. There are several components that interact. These include the various surface representations that have opaque or transparent attributes, the depth of these surfaces derived from other sources, and the relation of the contour shapes to the values of the regions they enclose. The problem of surface transparency, even in its simplest form, poses a number of unsolved computational and perceptual problems. In particular, it shows the power of human vision to incorporate global constraints, even when solving what may be an a priori local problem.

The second section provides a computational analysis of a simplified model of transparency in which the goal of the computation is to arrive at two distinct spatial maps, one representing opaque reflectance values, and one the transmittances corresponding to a transparent component. Here cooperative computation is analyzed in terms of an "ideal image understander." The intention is not to provide a theory of human transparency perception, but to illustrate the role of the statistical approach as a "quantitative computational theory" that encompasses a wide family of possible algorithms. This enables one to distinguish, at least in principle, between the correctness of the statistical model and the correctness of algorithms used to solve it. Purely local constraints are used to specify the statistical problem. Despite the inherent local nature of the constraints, purely local algorithms for computing transparency are inadequate.

Perception of Transparency

Transparency has received relatively little attention compared to other problems of image understanding. This probably reflects both the observation that simple surface transparency is rare in nature, and the anticipation that the computation of transparency is hard. From a general perspective, the problem of transparency may not be all that esoteric. For the purposes of this paper, the word transparency will be used to refer to the "seeing through" of one cause of image intensity to another at the same point. This is phenomenal, as distinct from physical, transparency. Seeing through one cause to another is a general problem that arises continually in everyday viewing. Here are a few examples that can be distinguished on the basis of the physical origins.

Specular transparency is a form of transparency in which a shiny or mirrorlike surface reflects some of the incident light with little diffusion, as with a polished apple. It is usually modeled as a component added to the matte reflectance of surface luminance. It is also commonly seen when looking through a window in which reflections off the glass add to the light coming through the window pane. Identifying specular transparency may be important to infer depth—either by discounting it in establishing stereo correspondence (Marr, 1982), or by using it to infer relative depth (Bülthoff & Blake, 1989).

Film transparency is the familiar form of multiplicative transparency where a clear surface absorbs some fraction of the light without scattering it. This darkens the image of the surfaces beneath it (figure 15.2, left panel). Film transparency will be the prototypical example returned to below in the computational section. Multiplicative trans-

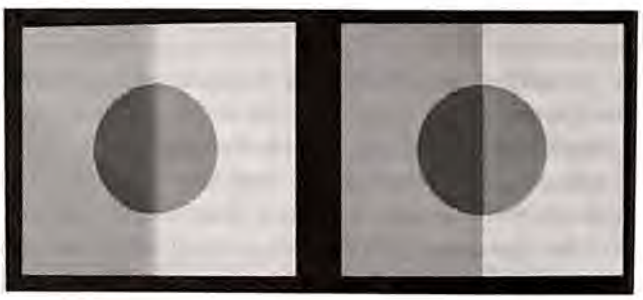


Fig. 15.2
Left, A circular transparent patch over a vertical line. The circular patch is usually seen as transparent. This is an example of multiplicative transparency. Because the image formation constraint is symmetrical, there is no local information at X junctions to bias the edge labeling. Thus a computation based on local information should provide an alternative solution—the circular patch is opaque and behind two abutting transparent rectangles. This interpretation can be seen at times, and demonstrates the multistability of transparency perception. *Right*, Shadow transparency can be seen by the introduction of blur to the vertical edge. This changes the assignment of surface attribute to the circular patch.

parency is a good example of the inherent ambiguity that can arise when two images are combined with a symmetric operation. Although the circular patch in the left panel of figure 15.2 can be seen as transparent and the rectangular regions opaque, sometimes the surface attribute assignments are reversed. In fact, sometimes the four regions appear as they really are—all opaque.

A strong impression of a *depth-induced transparency* can result when two random dot patterns move over each other in different directions. The visual system segregates the dots into two different depth planes, the nearer one appearing to be a perfectly clear, but dotted, transparent surface. An analogous effect can also be produced with a random dot stereogram. The brain seems to interpolate a surface between points of similar depths based on disparity or motion directions and, to be consistent with seeing through the plane, gives the near surface the attribute of transparency. From a physical point of view, this is a form of multiplicative transparency, with only binary valued reflectances and transmittances, which requires depth to be seen.

Diaphanous transparency, or gauzy or sheer transparency corresponds to the case where the holes in a perforated occluder are below the viewer's spatial resolution limit, contributing an additive component to the light shining through. Under everyday viewing, one can have various combinations of film and diaphanous transparency. These combinations should be distinguished from translucency, where scattering hides the spatial structure of the stuff shining through. In a formal sense, additive transparency exists in any low-pass filtered representation of an image of multiple objects. For example, imagine viewing an object through a leafless bush. At low spatial frequencies the object intensities add to those of the bush. Another instance of additive transparency arises because of our limited depth of field. Borders of the images of close objects get smeared over distant objects that are in focus.

For flat surfaces, sharp shadow boundaries are locally identical to multiplicative transparency induced by a dark film overlay (Metelli, 1975). We call this *shadow transparency* to underscore this similarity. Although these cases appear phenomenally quite different—a shadow is an intensity change perceptually attributed to illumination and a dark film is attributed to a surface—some aspects of the underlying computational problems are similar. The simple addition of a fuzzy penumbra to an apparent surface transparency is often sufficient to change the perception to a shadow (figure 15.2). With a more general model of shading where intensity is the vector product of surface normal and illumination direction, computing shadows is quite different from the problem of transparency and requires splitting or factoring the image into vector rather than scalar fields.

From a formal point of view, *occlusion* is the limiting case of zero transparency. There is increasing evidence that occlusion information may be represented fairly early in the visual system (Nakayama, Shimojo & Silverman, 1989). There is also evidence that depth relations inferred from surface transparency must be represented at least at the same level as depth from motion (Kersten, Bülthoff & Furuya, 1989).

Finally, we can see phenomenal transparency even when there is no real or simulated physical cause. We frequently have unfused binocular images when disparity gradients are outside our fusional limits. Most of the time, we do not notice that two different surfaces are actually visible at the same cyclopean point. This *binocular transparency* may be seen when a near surface occludes part of each eye's view. These two surfaces can at various times appear to combine leading to perception of transparency, or to compete for dominance, depending on the circumstances.

Transparency Formation

All of the above cases (except for binocular transparency) can be considered to result from the combination of at least two independent physical processes; for simplicity, we will consider opaque, $R(x, y)$, and transparent, $I(x, y)$ components. For film transparency, $R(x, y)$ and $I(x, y)$ are proportional to reflectance and transmittance, respectively. The image $L(x, y)$ is a function of these two:

$$L(x, y) = f(R(x, y), I(x, y)).$$

This function provides an image *formation constraint*. In general, the image formation constraint is quite complex and requires a complete model of rendering and projection of objects (Magnenat-Thalmann & Thalmann, 1987). An analysis of transparency formation is found in Richards and Witkin (1979). For multiplicative film transparency, the function is the pointwise product of the opaque and transparent components. It is easy to see that in the case of multiplication, the image formation constraint will not

necessarily produce an image that looks transparent to us. Any image can be considered the product of an arbitrary nonzero image, and a suitably chosen cofactor. Phenomenal transparency assumes additional nonpointwise constraints on the nature of the substances that produce it. These are *a priori statistical constraints*. One example follows from the cohesiveness of matter. Clumpiness of similar material often produces smooth self-occluding contours. With a multiplicative image formation function, this assumption would produce "X" junctions at the crossings of the clump boundaries in R and I. The crossing of the two X edges will in general produce four intensities. The relations among these four intensities that give rise to the perception of transparency have been studied by Metelli (1974) and Beck, Prazdny, and Ivry (1984).

One simplification made here, and in the computational analysis below is to assume one transparent and one opaque component. Even a simple image could be composed of two transparencies and an opaque surface. For example, two overlapping squares seen in the bottom of figure 15.7 could both be transparent. The human visual system seems to make the default assumption of opaqueness unless there is evidence to the contrary. Even when the transparency is bistable, the dominant percepts alternate between two cases: If one edge of an X junction is seen as transparent, the other is opaque, and vice versa. It is rare to see both edges as transparent. The conditions under which observers report multiple transparencies have yet to be systematically studied.

Transparency Detection

A first and basic level problem is, given an image, does the evidence favor the hypothesis of a transparent surface? The crossing of two boundaries provides one form of evidence for transparency, and has been the only one considered until recently. Given the right figural relations, to a first approximation transparency is seen if the intensities at X junctions satisfy the constraint that the image was caused by multiplicative and/or additive combination of two source images (Richards & Witkin, 1979). More exact models take into account the observation that the visual system is not sensitive to exact metric relations of intensities: rather, it respects the inequality relations derived from such a constraint (Metelli, 1974). Further, under certain conditions, these relations need to be adjusted to allow for compressive nonlinearity between subjective lightness and physical luminance (Beck et al., 1984)

Can there be local depth information at an X transparency crossing? It is useful to categorize one of the edges in an X junction with respect to qualitative changes of the intensities at the boundary of the other edge (Adelson, personal communication). For concreteness, imagine using a transparent film with a horizontal edge to cover the bottom half of two regions that are separated by a vertical boundary. The horizontal edge is attached to, or "intrinsic" to the bottom region of the image (Nakayama, et al., 1989). One can imagine this bottom film as either preserving or reversing the contrast polarity of the two regions separated by the vertical edge. A contrast-reversing surface does not in general tend to appear transparent, although it is physically realizable by a refractive element. Suppose the horizontal edge is contrast preserving. Then it can either lighten or darken the underling regions, or it could reduce or enhance the contrast at the vertical edge. For example, when the horizontal edge of a neutral density filter crosses the vertical boundary, it darkens the intensity on both sides of this edge. A purely additive transparency lightens both regions that it covers. Of particular interest is an edge that reduces contrast. Specifically, define a contrast-reducing edge to be one that lightens the darker of the two regions it covers, and darkens the lighter without reversing the contrast polarity. If the horizontal edge reduces contrast, there must be a vertical edge that darkens both regions while reversing contrast. Further, the horizontal edge, if considered attached to the top region, is contrast enhancing. This is defined as darkening the darker of two regions it covers, and lightening the lighter without changing contrast polarity. Surfaces attached to contrast-enhancing edges are not likely to be seen as transparent surface discontinuities. This provides a cue to edge attachment. These observations also suggest an ordering cue. In fact, the surfaces attached to contrast-reducing edges, if seen as transparent, tend to be seen in front in a monocular view (see figure 15.7). The above analysis is consistent with how a contrast-reducing edge may be realized physically. The nearer surface may contribute both diaphanous and multiplicative components. Contrast-reducing edges cannot be produced by a simple binary multiplicative or additive operation, but require a combination of both.

In the computational section below, it will be assumed that the combination function is a binary reversible opera-

tion as one would find with a purely multiplicative or additive interaction. This leaves ambiguity at X junctions, and places more demands on global computation to decide which edge is in front. If all junctions had a contrast-reducing edge, the computational problem of simple transparencies would be much easier. As pointed out above, the contrast-reducing edge would be a local sign identifying it with the nearer or transparent surface.

Are sharp edges in an X junction necessary to see transparency? Figure 15.3 shows that surface transparency can be seen when a sharp edged figure is overlayed on a smoothly shaded object. X junctions are eliminated altogether in figure 15.4, where depth information from stereo indicates transparency. Transparency can also be seen, if rather than X junctions, we have ψ junctions. in which there is a discontinuity in the derivative of one of the crossing lines. This can be seen in the bottom two panels of figure 15.8 when the center "book" is seen with its outside edges receding from the viewer.

Is an X junction even sufficient for transparency to be seen? Figure 15.5 shows that transparency is not seen if there is only one X junction. If the same X junction is repeated and arranged appropriately, it provides a consistent interpretation of a square transparent surface demonstrating the global nature of transparency perception.

In the computational analysis, the fact that X junctions are not necessary for the perception of transparency will be ignored. It will be assumed that local scene processes produce X junctions. This is not entirely unreasonable, in that occlusion may be treated as a special case of transparency. It will also be assumed that shapes are flat. Factoring out general shape and orientation representations is part of the problem of splitting images into scene attribute maps. Solving the problem of splitting images into two flat components is a first step. The observation that X junctions are neither necessary nor sufficient for the perception of transparency emphasizes the importance of understanding the global factors that cooperate to compute transparency. Even with X junction information locally available, simple local computations do not work to compute transparency.

Contour Binding

A second problem is how to integrate the local evidence to arrive at a global description of each of the multiple

Fig. 15.3
Transparency can still be seen when sharp edges overlap smooth intensity variations.

sources. This can be viewed as a process of perceptual organization in which one tries to globally bind pieces of the image that belong to one object based on a surface attribute. One important factor that contributes to binding is the contour shape. Figure 15.6 shows some examples adapted from Kanizsa (1979) in which the organization into transparent and opaque parts depends on how we segregate contours. A model has to take into account statistical knowledge of how contours typically bend to arrive at an account of the likelihood of the various stable percepts. Of particular importance is the tendency, at an X junction, to assume that a discontinuity in a single physical cause is more likely to be straight than curved. Minimizing an integrated measure of curvature is one way to express this constraint (Kass, Witkin & Terzopoulos, 1987). One problem that arises immediately is deciding whether to bind the four segments at an X junction together, at the cost of high curvature, or to bind pairs of segments (e.g., reflectance and transmittance changes) into lines of low or zero curvature. In the computational section below, curvatures are assigned discrete local probabilities on a fixed lattice using Markov random fields. Multiple lines at a point are allowed only if they have different causes and are thus represented in different intrinsic images. This problem underscores the need to per-

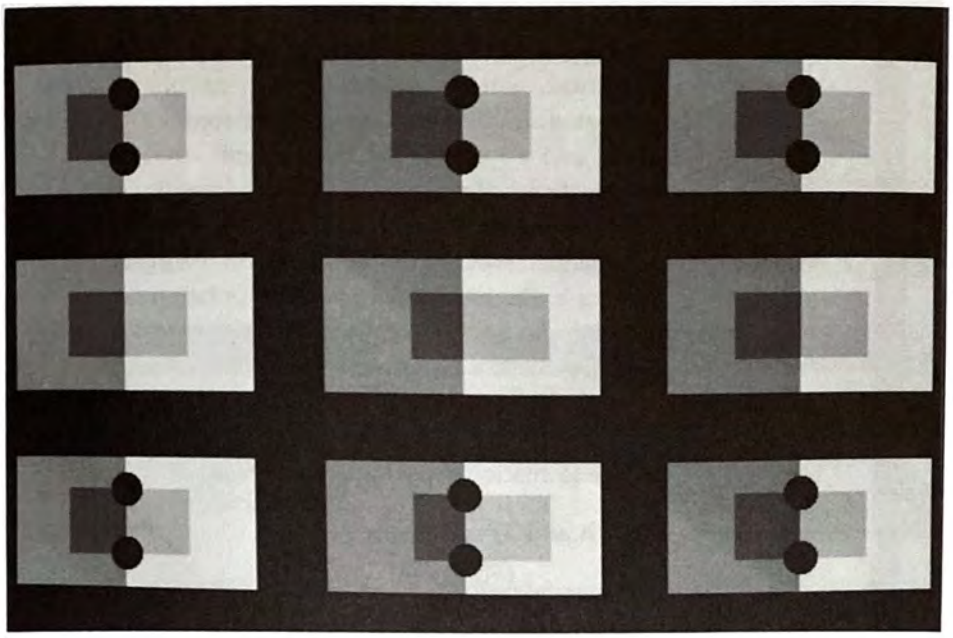


Fig. 15.4
Depth from stereo is sufficient to induce perception of transparency even when the X crossings are covered by dark patches.

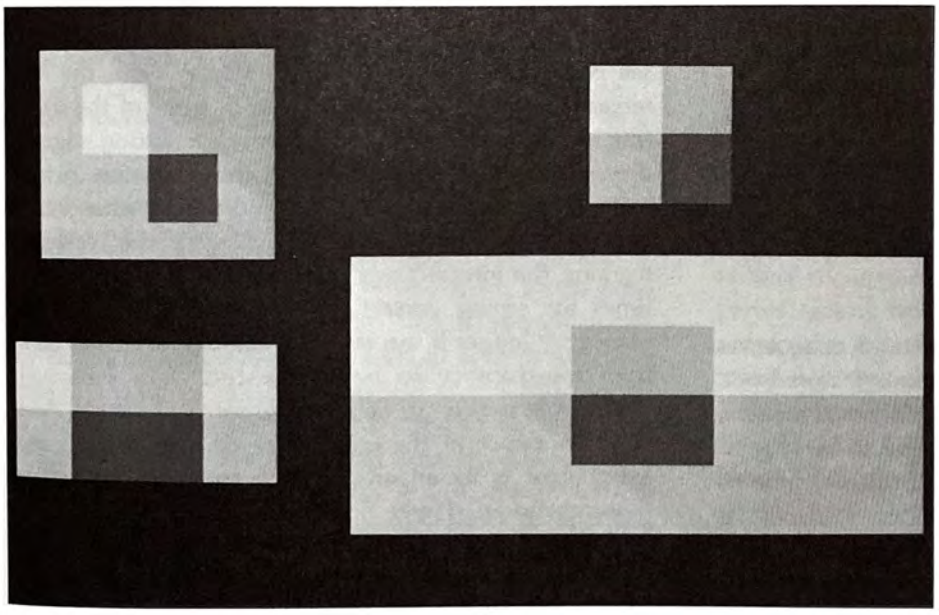


Fig. 15.5
One X junction does not necessarily induce phenomenal transparency. If a single X junction is repeated and arranged appropriately, an interpretation of a square transparent surface emerges. This illustrates the global nature of transparency perception.

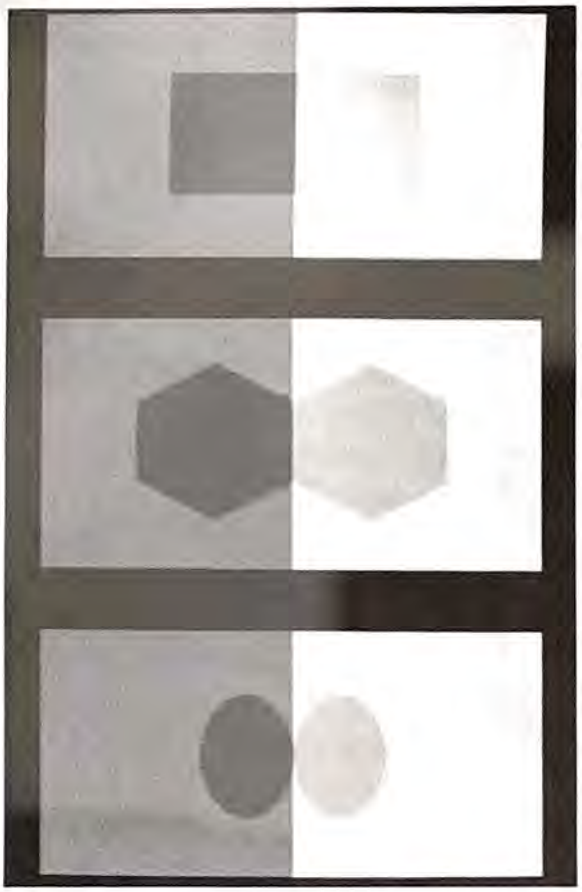


Fig. 15.6
Contours have a perceptual orientation momentum that determines how they bind across junctions. These demonstrations, adapted from Kanizsa (1979), show how the strength of apparent transparency in the presence of these grouping factors. There may also be an effect of symmetry.

mit multiple edge directions at a point in models of early edge detection (Zucker et al., 1988).

Attribute Attachment

A third problem is to attach the attribute of opaqueness or transparency to the surface components factored out. Sometimes the evidence at an X junction is informative; for example, contrast-reducing edges tend to be seen as transparent. If the combination resulted from a symmetric or commutative operation, and if the local evidence favors transparency, one has a more ambiguous choice. We have already seen in figure 15.2 that the perception of transparency can be multistable. Observers report seeing either the circle as transparent and the two rectangles as opaque, or the reverse, although there is a strong bias to seeing the circle as transparent. Observers sometimes see the entire pattern as it really is—an opaque figure in fairly uniform illumination—but seldom, if ever, see both circle and rectangles as transparent at the same time. When shadow transparency is introduced by simply blurring the vertical edge, the contour of the circle is seen as an opaque reflectance change on the background, rather than as a film transparency. A computational model must take into account both attribute assignment and its multistability. In the computational section, the opaque and transparent components are represented by distinct spatial maps. The ambiguity is represented by a symmetric image formation equation.

Depth and Transparency

Perceived depth affects surface attribute attachment. Similar to Mach's observations of the dependence of lightness on depth, perceived transparency depends on depth. If X junctions are covered, surface transparency can be seen with a stereo cue to depth (see figure 15.4). This suggests that depth, rather than intensity changes, should be the input to transparency computation; however, figure 15.7 shows that intensity relations can also affect the ability to see depth from stereo. In a monocular view, the light rectangular ring is invariably seen in front of the darker ring due to the presence of a contrast reducing edge.[1] If disparity cues are used to force the dark ring in front (seen by fusing the two right-hand panels with crossed eyes), observers report either rivalry or the rectangles marking the intersections as appearing opaque. Reaction times for seeing correct depth relations between two planes are longer if the observer initially perceives depth from transparency to be inconsistent with subsequent depth from motion or stereo (Kersten et al., 1989). Additional evidence of the importance of early transparency assignment is its apparent affect on motion coherence (Ramachandran, 1989). Two superimposed square-wave gratings moving in different directions tend to be seen as moving in one coherent direction if the superposition is not seen as transparent.

1. A monocular version of this figure was designed by Edward Adelson and demonstrated informally at the Cold Spring Harbor Workshop.

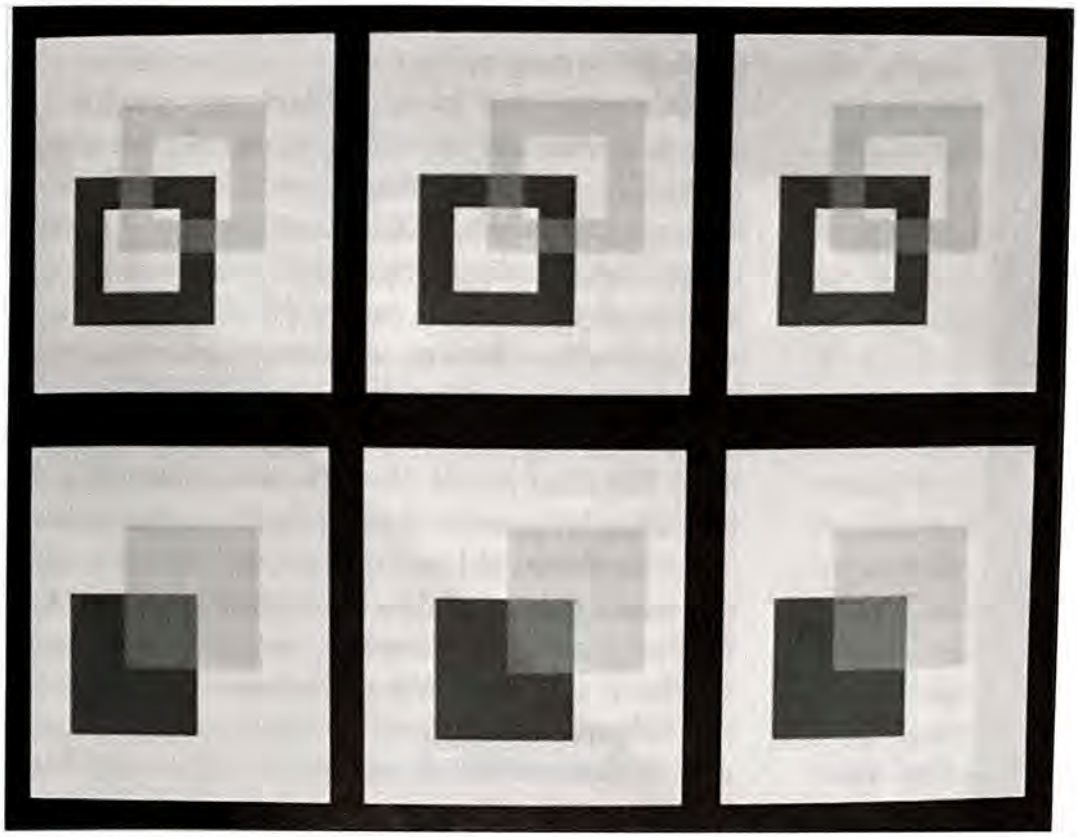


Fig. 15.7
Top, Intensity relations can affect the ability to see depth from stereo. In a monocular view, the light rectangular ring is invariably seen in front of the darker ring. If disparity cues are used to force the dark ring in front (seen by fusing the two right hand panels with crossed eyes, or the left two with uncrossed), one either sees rivalry or the rectangles marking the intersections as opaque. For some observers, the transparency information overrides the disparity information. The effect is not as easily seen in the bottom panel, in part due to the observation that the depth relation between the surfaces is more ambiguous in the monocular view.

Another example of how perceived depth affects attribute attachment is seen in figure 15.8. In the top panel, the central rectangle is easily seen as transparent. The same luminance values when arranged so that they appear to derive from a folded card give rise to an alternative interpretation in the second panel from the top. Here, all the edges seem to be reflectance edges, except for the central vertical edge, which now appears to be due to a change in orientation. Note too how the apparent contrast of the two central regions differs in the two top panels. The apparent contrast is much less for the folded than for the flat card. This is another example of how the estimation of one scene attribute affects the perception of another. The bottom two panels show that if the X crossings are changed to ψ crossings, the central "book like" smaller patches can appear to be either behind transparent larger rectangles, or as an opaque book coming out from the page.

Computation of Transparency

In this section, we will look at the computational problems involved in computing the simplest transparency case—multiplicative combination of just two piece-wise constant components. The goal is to understand computational requirements for estimating the opaque and transparent factors from a single image. It has been argued previously that the Bayesian approach to computing scene descriptions from images provides a quantitative "computational theory" (Kersten, 1987; Marroquin, 1985). The Bayesian estimator can be thought of as an "Ideal Image Understander." It is an extension of the Ideal Observer for detection and discrimination used in early vision (Geisler, 1989) to scene estimation. The Ideal Image Understander makes optimal use of both hard constraints on image formation and soft constraints on the prior likelihoods of scene characteristics.

Bayesian estimation requires three steps: One must specify (1) *the posterior* (or a posteriori) probability of a scene representation conditional on the image data, (2) a statistic representing what one would like to estimate

Fig. 15.8
In the top panel, the central rectangle can be seen as transparent. The same luminance values when arranged so that they appear to derive from a folded card give rise to an alternative interpretation in the second panel. Here, all the edges appear as reflectance edges, except for the central vertical edge, which now appears to be due to a change in orientation. This is another example of how the estimation of one scene attribute affects the perception of another, thereby affecting the edge labeling. The third and fourth panels show that the result is not simply a matter of changing the X to ψ crossings. The central patches can be seen as going behind larger transparent rectangles, or as open opaque books.

(e.g., mode or mean) on this distribution, and (3) an algorithm to find this estimate. The posterior probability function embodies statistical knowledge about the world, and the model of how the image was caused. The model will tell us how probable any particular solution is. In particular, it should give low values to those scene interpretations we never see and high and similar values to the (possibly multiple) perceptual interpretations we do see. In our case, we want to find probable scenes, that is, opaque and transparent components, conditional on the image luminance. Below, the largest mode or peak of the posterior distribution is estimated. This is called *maximum a posteriori* or MAP estimation. MAP estimation is optimal in the sense that it minimizes the probability of error. The Ideal Image Understander developed below is defined to be a MAP estimator of a scene representation conditional on the image.

What is an Ideal Image Understander good for? In general, questions of optimality have to take into account functional goals of the animal or machine, biological or hardware constraints, and processing time. The Ideal Image Understander is only limited by the uncertainty in the information provided and by the simple goal of minimizing classification error without regard for time. It provides an upper bound on the estimation performance of observers or algorithms attempting to achieve the same goal. The Ideal Image Understander makes the useful distinction between the *statistical model* and the *algorithm*. The statistical model is a quantitative statement of the constraints and goal of the computation (parts 1 and 2). In principle, the statistical model can be evaluated as to whether it is right or wrong independent of the specifics of the algorithm. The algorithm is the particular technique used to compute the answer based on the model. There are, in general, many algorithms that can find the MAP estimate for a given problem. But there are countless other algorithms that are suboptimal in the sense that their probability of error is higher. In principle one could make algorithm-independent predictions regarding what are the likely human interpretations of a scene given image data. Modes of the posterior probability should correspond to stable perceptual states. If they do not, then either the brain's algorithm does not locate these modes, or its perceptual model of the scene properties is different from the ideal's statistical model.

The first step of specifying the posterior probability is handled more conveniently if Bayes' rule is used to split the probability into two parts: $p(image|scene)$ represents the image formation constraint discussed above, and $p(scene)$ specifies the *prior* (or a priori) conditions on the scene parameters that we would like to estimate. *Scene* and *image* can be thought of as two very long vectors with the parameters describing a potential scene and its image. The posterior probability is

$$p(scene|image) = \frac{p(image|scene)p(scene)}{p(image)}.$$

The probability also depends on $p(image)$, but this is constant for a given image. One can think of the prior probability model as a procedure that draws sample scene descriptions with the likelihoods specified by the model.

In principle, one could independently try to verify the prior model based on a sample population drawn from the real world. In practice, this may be difficult. Statistical models of scene parameters from real-world data have been sought in only a few cases (e.g., spectral reflectances, Maloney & Wandell, 1986). Suitable mathematical modeling tools are needed to specify the prior model. Markov random fields (MRFs) provide one such tool for specifying prior probabilities of scene parameters represented on a spatial lattice. One of the strengths (and limitations) of MRFs is that they are defined in terms of local, and thus manageable constraints. An MRF is defined in terms of local conditional probabilities on neighborhoods. A neighborhood, N_i, of a site i, is a set of sites not containing the point itself and, further, if the point i is a member of some neighborhood N_j, j must be in the neighborhood of i. An MRF is defined by the rules that (1) the probability of any field (e.g., reflectance map) is positive, and (2) the probability of a site value (e.g., reflectance) conditional on all the other values is equal to the probability conditional on only those values in its neighborhood. Bayesian estimation on MRFs provides a general statistical framework for regularization (Poggio, Torre & Koch, 1985). Many neural net algorithms are doing MAP estimation over MRFs (Golden, 1988). Their limitations are essentially those of regular grammars (Miller, Roysam, Smith & Udding, 1990).

The posterior probability, over an MRF, can be expressed in terms of an energy or cost function E_G:

$$p(scene \mid image) \propto e^{-E_G/T},$$

where the energy, E_G, is the sum of terms corresponding to the prior and image constraints (Besag, 1972). The energy is the sum of small local interactions or potentials. Finding modes is equivalent to finding minima of E_G. The T, or temperature, parameter is useful in some applications for expanding and compressing the distribution in a search for modes.

Consider the following energy function for splitting an image into just two components, a transparent and an opaque part:

$$E_G = V_I(\mathbf{I}, \mathbf{I}^{\mathrm{I}}) + V_R(\mathbf{R}, \mathbf{I}^{\mathrm{R}}) + \lambda_1 V_L(\mathbf{L}, \mathbf{R}, \mathbf{I}) + \lambda_2 V_E(\mathbf{I}^{\mathrm{L}}, \mathbf{I}^{\mathrm{R}}, \mathbf{I}^{\mathrm{I}}, \mathbf{I}^{\mathrm{D}}).$$

The first two energy terms capture prior statistical assumptions about the cohesiveness of like matter, and include specific terms to constrain contour binding. Only the second two terms depend on the data. The energies can be summed because the probabilities, expressed as exponentials, are independent, and thus multiply. V_R and V_I represent energies for the prior statistical models of the opaque $\mathbf{R}$ and transparent $\mathbf{I}$ vectors, respectively. The opaque and transparent vectors contain the reflectance and the transmittance of the corresponding components. $\mathbf{I}^{\mathrm{R}}$ and $\mathbf{I}^{\mathrm{I}}$ represent *line processes* and are binary vectors representing the discontinuities in the opaque and transparent factors (Geman & Geman, 1984). In the implementation below, horizontal and vertical line processes are placed between vertical and horizontal pixel pairs, respectively. A value of one indicates the presence of a line, and zero its absence. Knowing V_R means that we can assign a global probability to any sample two-dimensional opaque map in our model.

V_L represents the luminance or image constraint. The image luminance vector, $\mathbf{L}$ is a function of $\mathbf{R}$ and $\mathbf{I}$. Each luminance value is determined by the point wise application of the image formation equation. Minimizing V_L encourages the estimate of the image, which is in turn a function of the estimates of the transparent and opaque components, to agree with the measured image data. The exponential of its negative is proportional to the probability of a given image conditional on a scene vector. V_E is an edge data constraint that allows conditional probabilities relating scene discontinuities to image discontinuities (Poggio et al., 1988), and cooperatively couples scene discontinuity estimates from other modules, such as depth. $\mathbf{I}^{\mathrm{L}}$ and $\mathbf{I}^{\mathrm{D}}$ are line process vectors marking image and depth continuities.

λ_1 controls the weight given to the multiplicative image constraint. If $\lambda_1 = 0$, then the product of opaque and transparent components does not have to equal the image luminance to produce a low energy. The noise level in the imaging process determines λ_1. In the simulations, it is assumed that there is no imaging noise, and λ_1 is used for constraint relaxation, and as such it is not a free parameter, but starts small and tends to infinity. λ_2 controls the weight given to the interaction between the discontinuities. In the simulations, it is set to either one or zero, depending on whether the constraint is used or not.

Let us return to the modeling of the prior terms, V_R and V_I. The opaque component is modeled in terms of local conditional probabilities of reflectance values and the invisible line processes between them. Recall that by the

definition of an MRF, the conditional probability of a reflectance value, R_i at site i depends only on the values, R_j, of its neighbors, where j is a site in the neighborhood of i. A neighborhood system where each site i has as its neighbors the four nearest pixels is shown in figure 15.9. In particular, the clumpiness of "like matter" would require the central reflectance to have a value near its neighbors. The neighborhood can be broken if a discontinuity, marked by a line process being on, exists between two sites. In simulations, the neighborhood consisted of the four nearest neighbors when there are no line processes turned on.

The prior energy term for the reflectance is determined by the sum of local potentials which are in turn determined by the local conditional probabilities

$$\mathbf{V_R}(\mathbf{R}, \mathbf{l^R}) = \sum_{i,j \in N_i} V_R(R_i, R_j)(1 - l_{ij}^R) + \sum_{C_l} V_{l^R}(l_{ij}^R, l_{kl}^R).$$

l_{ij}^R represents the binary line process between pixels i and j. If equal to one, it knocks out the contribution of the jth neighbor. N_i is the four nearest neighbors of i. In general, a global prior potential is determined by summing local energies for each *clique*. A clique, C, is a single site or a set of sites such that each member is a neighbor of the others. For the four nearest neighborhood system, cliques are either single sites, or nearest vertical or horizontal pairs.

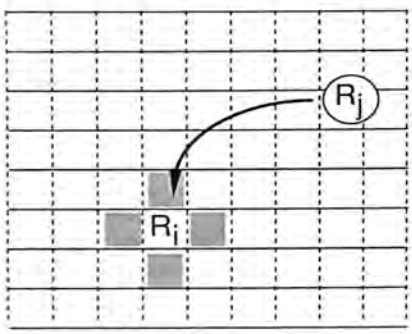


Fig. 15.9
Site i in a Markov random field. Its four neighbors are shown shaded and make up the neighborhood N_i. A smoothness constraint assumes that the probability of the value, R_i, conditional on the four nearest neighbors is higher when it is close to the values of its neighbors. An example of a clique is the pair of sites i and j.

The first sum above is over pairwise cliques. The second sum represents the cost of various contour arrangements. The cliques for line processes (C_l), are slightly more complicated than those for the field (i.e., reflectance or transmittance) values and are discussed below.

The form of the prior potential is based on heuristic judgement. One convenient form is the quadratic potential which encourages smoothness by giving increasingly low probabilities to large differences of reflectance between neighbors[2]:

$$V_R(R_i, R_j) = (R_i - R_j)^2/\sigma^2.$$

For the simulations, the transparency problem is approximately symmetric in both image formation and priors, so the transparent component is similar to the opaque reflectance term

$$\mathbf{V_I}(\mathbf{I}, \mathbf{l^I}) = \sum_{i,j \in N_i} V_I(I_i, I_j)(1 - l_{ij}^I) + \sum_{C_l} V_{l^I}(l_{ij}^I, l_{kl}^I).$$

$V_I(I_i, I_j)$ is also a quadratic potential which, together with the line processes, captures the piecewise smooth characteristic of transparent overlays. In the simulations, $\sigma = 1$ for both reflectance and transmittance potentials.

The line processes represent physical discontinuities in the reflectance (or transparency) function and break the neighborhood determining the local potential as one goes from one type of clump to another. Line processes are also modeled as MRFs. In the simulations here, there are only vertical lines and horizontal lines. Each line has six neighbors. Each line site has a neighborhood structure that enables one to characterize the probabilities of various contour configurations in terms of just local interactions (figure 15.10). For example, two lines of the same orientation are more likely than ones at right angles. One can discourage "T" junctions for transparencies, but not for opaque functions, and so forth. For the clique cases: no lines, a collinear pair, one line, a pair at right angles, three lines and four lines, the values of the potential, V_{l^R}, used for the reflectance line processes were determined from configuration energies: 0.0, 0.0, 2.0, 1.8, 1.8, and 2.0 respectively. The values for the configuration energies of the transmittances, V_{l^I} were: 0.0, 0.0, 3.0, 1.8, 3.7, 3.7. The

2. Another example of a local potential is a "smoothed Ising potential"

$$V_R(R_i, R_j) = \frac{-\beta}{1 + (R_i - R_j)^2/\alpha}$$

which, like the quadratic potential, gives high probabilities to similar intensities, but on the other hand, gives nonvanishing probabilities to very large differences in intensity between two neighboring reflectances. It encourages piecewise constant patterns even without line processes.

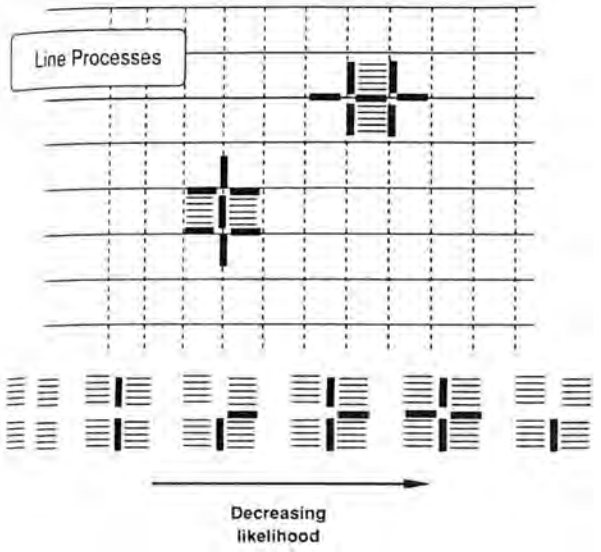


Fig. 15.10
The line processes represent physical discontinuities in the modeled surface property. They have binary values and, when turned on, mark the break in the neighborhood as one goes from one value of surface property to another. Line processes are modeled as Markov random fields with their own prior probability structure. The bottom of the figure shows decreasingly likely configurations as one goes from left (no lines) to right (an isolated line).

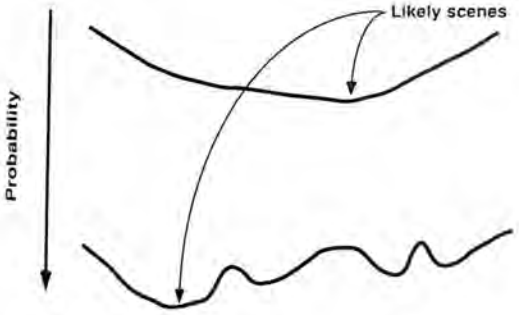


Fig. 15.11
Examples of hypothetical one-dimensional energy functions that are inversely related to some corresponding posterior probability. The horizontal axis indexes the scene attribute map under consideration. *Top*, A convex energy function. Gradient or other descent methods could find the bottom and thus the most likely scene interpretation. *Bottom*, An example with local minima. Each minimum represents either one of several stable perceptual interpretations, or states that may never be "seen" by our hypothetical visual system if the particular algorithm chosen never settles there.

exact values are not critical. The relative weighting is more important. The additional energy for three and four line configurations penalized this sort of junction more heavily for the transparent than for the opaque surface and provided the only asymmetry in the simulations described below.

An additional advantage of making the scene discontinuities explicit is that one can model the interactions between image and scene discontinuities expressed in V_E. Image edges can be coupled to the scene discontinuity estimates and to each other. Shadow edges are transparent and intersect with each other less frequently than occlusion edges. Edges due to shadows or transparency rarely coincide with the other types of discontinuity. On the other hand, edges due to occlusion often coincide with reflectance or texture change. Weights given to these constraints can be embodied in V_E (see tables 15.1 and 15.2).

For multiplicative transparency, the following image formation energy function enforces the constraint that the image be the product of the estimates of the opaque and transparent parts

$$\mathbf{V}_\mathbf{L}(\mathbf{L}, \mathbf{R}, \mathbf{I}) = \sum_i (L_i - R_i I_i)^2.$$

It is straightforward to put in alternative image formation constraints.

At this point, one could pause and ask whether the statistical model is any good in the sense that if one plugged in answers, inferred from human perception studies, would these answers be at the global energy minimum? Although one could plug in the solutions that some observer has produced, the remaining problem is to compare these probabilities (or energies) with the Ideal Image Understander's. Unfortunately, it is not always clear when one is at the global minimum. Just because we have an energy function does not mean we understand how to solve the problem (Yuille, 1987). An algorithm is required to find minima.

The Algorithm and Results

To get an intuitive understanding for the nature of the algorithmic problem, figure 15.11 shows hypothetical one-dimensional energy functions to illustrate the kind of problems one encounters when seeking a global minimum. If lucky, the energy function would be like the top panel of figure 15.11, that is, convex. Gradient or other

descent methods could get one to the bottom and thus to the most likely scene interpretation. For the transparency problem as formulated here, we have a problem somewhat akin to trying to find the bottom of the curve in the lower panel of figure 15.11, in which there are multiple minima. On the one hand, we do not want to eliminate multiple minima, which are required to represent multistable perceptions; on the other hand, there may be numerous false solutions that are never seen. Two possible approaches are either to try new representations that do not lead to craggy energy functions, or to find brute force algorithmic techniques that work with the existing statistical model. The easier, but less satisfactory, brute force approach will be taken here. The positive lesson will be an understanding of the limitation of local algorithms to deal with a local statistical model. There are several tricks that can be used to search for both global minima, and ones close to it.

Gibbs Sampler with Gradient Descent and Image Edge Constraint

The Gibbs Sampler was originally developed by Geman and Geman (1981). Initially, all the scene estimates are chosen at random. The idea is to draw a sample from the local conditional posterior probability (calculated from the global posterior distribution) at a randomly chosen site. The local probabilities at site i depend on the image intensity, reflectance, and transmittance at that location and the neighborhood values of the reflectance and transmittance. The sampled values of reflectance and transmittance replace the previous values. The order in which the sites are visited does not matter as long as each one is visited "often enough." In the simulation results shown below, the pixels were visited in a raster order. Figure 15.12 illustrates the local sampling part of the algorithm. If the sampling is done with temperature set to zero, the procedure corresponds to gradient descent on the energy function. Zero temperature compresses the distribution around a single point approaching the mode. Figure 15.13 illustrates some results using gradient descent. The top panel of figure 15.13A represents the input data (thickened lines indicate image edge markings). The bottom left and right panels are the initially randomized representations of the opaque reflectance, and transmittance respectively. The thick gray lines between square pixels indicate line processes that are turned on. Figure 15.13B shows

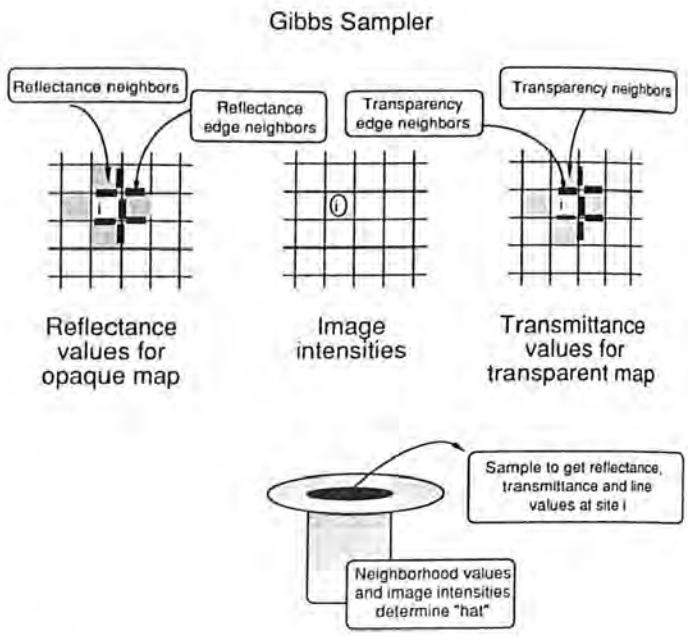


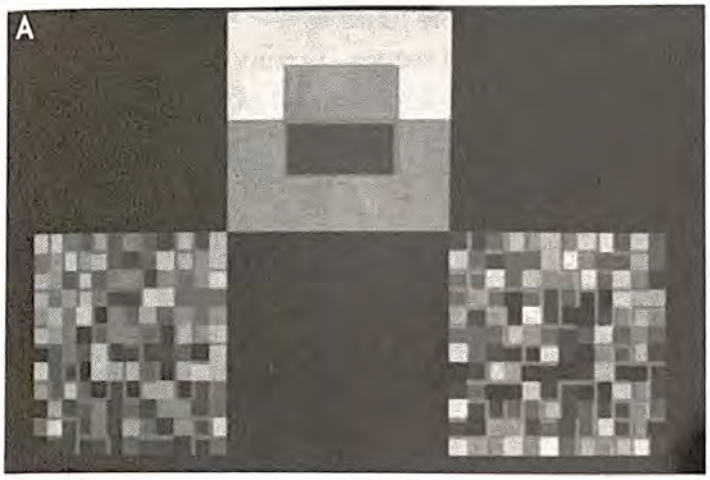
Fig. 15.12
The Gibbs sampling procedure used in the simulations to estimate reflectance and transmittance maps from the data. The opaque and transparent maps are initialized with random values. A site is then selected, a sample is drawn from an appropriate "hat," and used to replace the previous value. Each site is visited in a raster fashion (random asynchronous updating can be used). In general, the algorithm gradually converges to a stable state representing the estimate of reflectance and transmittance consistent with the image intensities. See the text for more details.

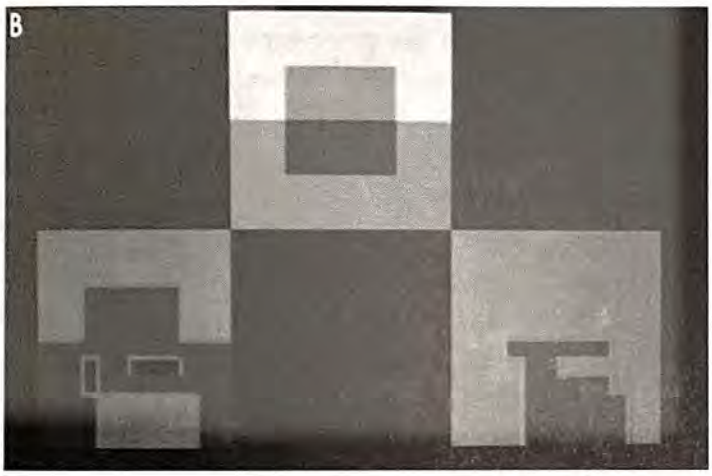
how gradient descent quickly leads to a local energy minimum satisfying the image formation constraint. In this example, $\lambda_2 = 0$; that is, there is no image edge constraint.

One problem with this solution is that lines show up in the solution that do not have support from lines in the image. Intensity changes in the image can usefully constrain the search for opaque and transparency edges. If there is an edge in the image, this increases the likelihood that there is either an edge in the opaque component, or one in the transparent component. Further, it is unlikely that both types of edge coexist at a site. The potential for this constraint is

$$\mathbf{V}_E(\mathbf{l}^L, \mathbf{l}^R, \mathbf{l}^I) = \sum V_E(l^L, l^R, l^I).$$

Table 15.1 shows values used for V_E. Figure 15.13C shows the local minimum arrived at using gradient descent with this image edge data potential. Although coming closer to being reasonable, the solution is clearly unsatisfactory as a perceptual interpretation. The solution





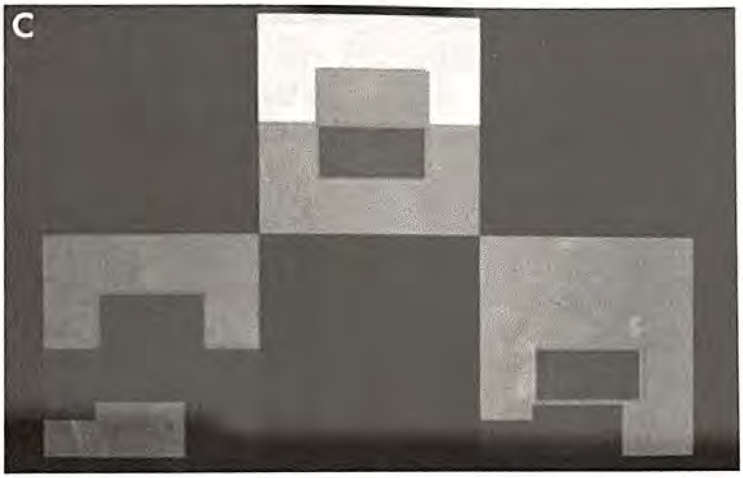


Fig. 15.13
Each of the three panels consists of a triad. The picture at the top of each triad represents the input image data. The original input picture was 12 by 12 pixels, with 64 possible graylevels. The reflectance and transmittance scene maps had 8 levels each. Thick gray lines indicate line processes that are turned on. The bottom left and right pictures of each triad indicate the current estimate of reflectance and transmittance, respectively. (A) The starting configuration for a simulation. (B) The local minimum after 24 iterations of gradient descent. One iteration is completed after all sites have been visited once. (C) A local minimum after 20 iterations, but here scene edges are constrained to have support in the image edges and to not overlap.

Table 15.1
Values of the edge potential (inversely related to probability) for various configurations of edge labels

Image edge	Reflectance edge	Transmittance edge	Potential V_E
1	0	0	1
1	1	0	0
1	0	1	0
1	1	1	1
0	0	0	0
0	1	0	1
0	0	1	1
0	1	1	1

Note: High energies are assigned to unlikely combinations.

again corresponds to a perceptual interpretation that is never seen. Because the input image was constructed by multiplying a known reflectance and a known transmittance, the correct answer is known, and the energy can be calculated. Note that this correct answer may not necessarily correspond to the global minimum of the posterior probability. If the statistical model is wrong, then the global minimum of the posterior probability does not correspond to the correct answer. So even a "perfect" algorithm, such as one capable of an exhaustive search, would find the wrong answer. The energy for the two solutions shown so far is much higher than the correct energy. If the statistical model is correct, this would indicate a local minimum. Additional techniques are needed to avoid monotonic descent of the energy landscape.

Bag of Tricks Search

Even for simple transparencies, reasonable solutions in a few iterations have only been obtained by using a combination of techniques to avoid local minima. These tricks

include image edge data (from above), simulated annealing, constraint relaxation, and neighborhood relaxation. The power of each of these techniques was evaluated separately, and none of them provided consistent solutions to the simple transparency shown in the top panel of figure 15.13. The best solutions were obtained when all four techniques are combined into a "bag of tricks." In *simulated annealing*, the temperature parameter, T, is initially set high. For high temperatures, the Gibbs sampler samples from a local distribution with a large variance. This means that sometimes parameters of the scene estimate are chosen that are locally unlikely, but allow escapes from local energy minima. The temperature is gradually lowered until one is doing strict descent on the energy function. In theory, a sufficiently gradual annealing schedule is guaranteed to find the global minimum, but in practice it is too slow (Geman & Geman, 1984). In *constraint relaxation*, one starts off with small values of λ_1, initially ignoring the image formation constraint to find independent samples of opaque and transparent components. This helps to lower extremely high barriers in the energy function caused by hard constraints. λ_1 is gradually raised until the data constraint eventually dominates the energy function.

The principal difficulty is trying to find a global minimum with only local propagation of constraints. The local conditional prior probabilities may in fact be the right ones, from the point of view of the statistical model, but the algorithm is bad because it fails to propagate these local constraints over a wide region, especially with the multiplicity of local minima in the transparency problem. One empirical trick devised for this problem is *neighborhood* relaxation to initially construct the conditional probability for the Gibbs sampler using an expanded neighborhood. In these simulations, the same number of neighbors are maintained (i.e., four field neighbors and six line neighbors), but in the first iterations they are at distant points. As the iterations proceed, the neighborhood size is contracted. Neighborhood relaxation sketches in a global solution, and then fine tunes it. It is in the spirit of coarse to fine multigrid techniques (Terzopoulos, 1984).

Figure 15.14 shows results using the Bag of Tricks descent.[3] Part A shows an intermediate stage in the convergence in which the distance between neighbors is two rather than one pixel. Part B shows a solution corresponding to a common perceptual interpretation—a square transparent film overlaying two opaque rectangles. A less common interpretation is to see the pattern as it is—all the edges are reflectance changes, and there is no transparency. The algorithm also finds this solution, shown in part C. With different random starting configurations and the parameters chosen, the algorithm finds four major interpretations. In addition to the two shown, it also tends to converge to the symmetric solutions, namely, in which all the edges are transmittance changes, or where only the central square is a transparent surface.

Even the bag of tricks does not work well for more complicated transparencies such as the one shown in figure 15.15. After many iterations (figure 15.15B), the descent has traveled down a slope distant to any reasonable perceptual interpretation. One, however, can achieve reasonable solutions by incorporating additional sources of depth information. For example, if an independent stereo process has labeled the edges simply according to whether they are at the same depth plane or not, an improved solution is obtained. This constraint is incorporated into a potential term

$$\mathbf{V}_E(\mathbf{l}^L, \mathbf{l}^R, \mathbf{l}^I, \mathbf{l}^D) = \sum V_E(l^L, l^R, l^I, l^D).$$

Table 15.2 shows values of V_E used in the simulations. Figure 15.15C shows results (after 818 iterations) using the Bag of Tricks with an additional depth constraint. The added complexity of a depth constraint may seem premature in that *we* can see the transparent plane in a monocular view, even if the model has difficulty. Eventually, stereo information has to be incorporated into models of transparency, even when X junctions are visible. Figure 15.16 shows a stereo pair of many overlapping

3. The constraint relaxation schedule was:

$\lambda_1 = \log(2 + i)/\lambda_0$,

and the annealing schedule was:

$T = T_0/\log(2 + i)$,

where i is the iteraction number. Typical values of λ_0 and T_0 were 2 and 4. The neighborhood relaxation started off at a diameter of 12 pixels, and was reduced by one after every four iterations.

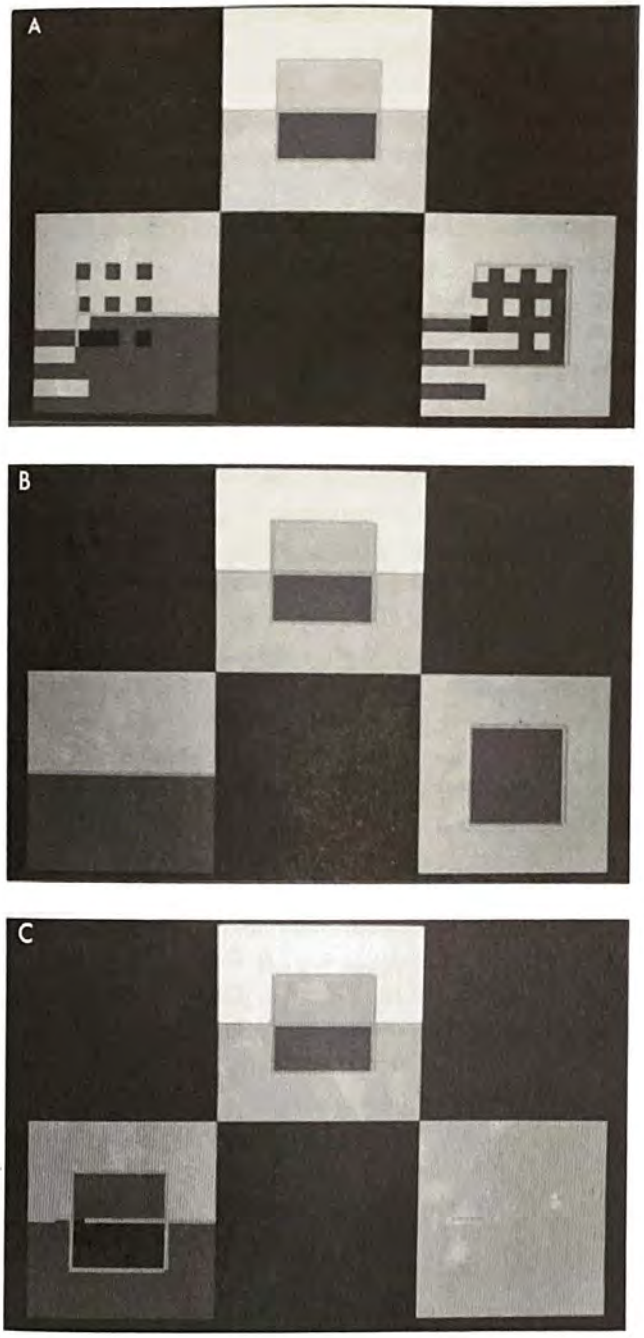


Fig. 15.14
The results of a simulation combining image edge data constraint, an annealing schedule, constraint relaxation, and neighborhood relaxation. (A) The progress of the simulation after 18 iterations. The checkerboard appearance is a consequence of the neighborhood relaxation. Convergence to reasonable solutions is fairly rapid, 20 and 23 iterations for (B), and (C), respectively. Other details are as for figure 15.13.

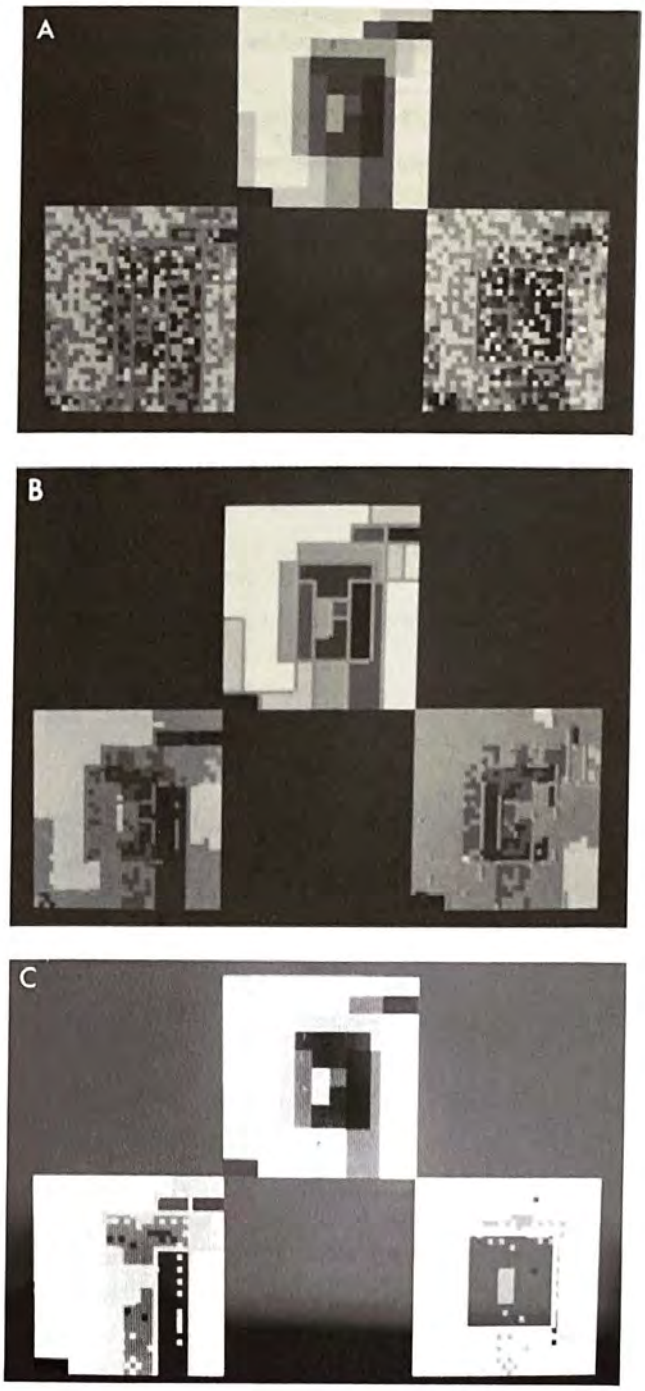


Fig. 15.15
(A) The starting configuration for a more complicated transparency problem. The transparent surface is a square as in the previous example, but the opaque background is Mondrian-like. (B) Results using the "bag of tricks" after 54 iterations. This simulation never recovered after over 1000 iterations steering down a valley distant from anything remotely resembling what we perceive. A considerable improvement obtains when in addition to the bag of tricks, an edge constraint (see table 15.2) is used to group line elements that are in the same depth. (C) The state of the convergence after 818 iterations.

Table 15.2
Values of the edge potential when depth labels are assigned to line locations

Image edge	Reflectance edge	Transmittance edge	Depth index	Potential V_E
1	0	0	0	1
1	1	0	0	0
1	0	1	0	0
1	1	1	0	1
0	0	0	0	0
0	1	0	0	1
0	0	1	0	1
0	1	1	0	1
1	0	0	1	1
1	1	0	1	0
1	0	1	1	0
1	1	1	1	0
0	0	0	1	1
0	1	0	1	1
0	0	1	1	1
0	1	1	1	1

Note: The depth index serves solely to indicate that line segments are at the same depth plane and not their relative depth.

Fig. 15.16
A stereo pair of many overlapping rectangles, some of which are transparent and some opaque. The transparency of many of the rectangles is only apparent when viewed stereoscopically, despite the visibility of X junctions.

rectangles. The observation that some rectangular patches are transparent is only apparent in the stereo view.

Summary

Several new perceptual observations were made that theories of transparency must ultimately have to deal with. The central argument is that perceived transparency is determined by the cooperative computation of distinct scene attributes including depth from stereo, motion or perspective, and the opacity of the surfaces behind. One piece of evidence is the multistability of the perception of the attachment of opacity and transparency attributes to surfaces. A second line of evidence is that not only does the perception of depth affect the perception of transparency, but that perceived transparency can affect depth. The fact that an X junction is neither necessary nor sufficient for the perception of transparency indicates the importance of global computation (depth from stereo is sufficient for the perception of transparency in the absence of X crossings). The prior statistics of the contour shape have a strong affect on the attachment of surface attributes and is explicitly represented in the computation.

The computation section outlined a simplified Ideal Image Understander or Bayesian approach to computing surface transparency. The model was simplified in that it only dealt with multiplicative transparency of flat Mondrian-like surfaces. It was pointed out that a major strength of the Bayesian approach is the separation of the statistical model from the algorithm. Markov random fields provide tools for the statistical modeling of scene attributes that may be useful in the long term. However, because the posterior energy landscape for transparency is extremely rugged when set up in terms of local constraints, local algorithms based on Gibbs sampling may be less useful for computing perceptual models from images—that is, for finding modes of the posterior distribution. Neighborhood relaxation was one new useful technique that helped to overcome this limitation. Incorporating information on depth relations between edges, derived from other sources (e.g., stereo) was another trick that improved the search for modes. There remains a big discrepancy between the Bag of Tricks descent used and human computation of transparency. This gap is left for future research to resolve.

Acknowledgments

This work was supported by NSF grant BNS-8708532 to Daniel Kersten. This research began thanks to time at the Center for Biological Information Processing at MIT, which is supported in part by the Office of Naval Research, Cognitive and Neural Sciences Division, Contract No. N00014-88-K-0164, and in part by the National Sci-

ence Foundation, Contract No. IRI-8719394. The author would like to thank Norberto Grzywacz, Edward Adelson, Michael Landy, David Knill, and Heinrich Bülthoff for many useful comments and suggestions.

References

Arend, L. E. & Goldstein, R. (1987). Simultaneous constancy, lightness and brightness. *Journal of the Optical Society of America A, 4,* 2281–2285.

Barrow, H. G. & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. R. Hanson & E. M. Riseman (Eds.), *Computer Vision System* (pp. 3–26). New York: Academic Press.

Beck, J. (1972). *Surface color perception.* Ithaca, NY: Cornell University Press.

Beck, J. Prazdny, K. & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics, 35,* 407–422.

Besag, J. (1972). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B, 34,* 75–83.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bülthoff, H. H. & Blake, A. (1989). Does the seeing brain know physics? *Supplement to Investigative Ophthalmology and Visual Science, 30,* 262.

Bülthoff, H. H. & Mallot, H. A. (1987). Interaction of different modules in depth perception. *Proceedings of the International Conference on Computer Vision* (pp. 295–305). Washington, DC: IEEE.

Chou, P. B. & Brown, C. M. (1988). Multimodal reconstruction and segmentation with Markov Random Fields and HCF optimization. *Proceedings Image Understanding Workshop,* pp. 214–221. Cambridge, MA: Morgan Kaufmann, San Mateo, CA.

Geisler, W. (1989). Sequential Ideal-Observer analysis of visual discriminations. *Psychological Reveiw, 96,* 267–314.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions or Pattern Analysis and Machine Intelligence, PAMI-6,* 721–741.

Gilchrist, A. L., Delman, S. & Jacobsen, A. (1983). The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception & Psychophysics, 33,* 426–436.

Golden, R. (1988). A unified framework for connectionist systems. *Biological Cybernetics, 59,* 109–120.

Grossberg, S. & Todorović, D. (1988). Neural dynamics of 1-C and 2-D brightness perception: A unified model of classical and recent phemomena. In S. Grossberg (Ed.), *Neural Networks and Natural Intelligence.* Cambridge, MA: MIT Press.

Horn, B. K. P. (1973). On lightness. *MIT Artificial Intelligence Memo,* 295.

Horn, B. K. P. (1975). Obtaining shape from shading information. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 115–155). New York: McGraw-Hill.

Kanizsa, G. (1979). *Organization of vision.* New York: Praeger.

Kass, M., Witkin, A. & Terzopoulos, D. (1987). Snakes: Active contour models. *International Conference on Computer Vision,* London.

Kersten, D. (1987). Statistical limits to image understanding. Symposium on "Vision: Coding and Efficiency" Cambridge, England. To appear in C. Blakemore (Ed.), *Vision: Coding and Efficiency,* 1990. Cambridge: Cambridge University Press.

Kersten, D., Bülthoff, H. H. & Furuya, M. (1989). Apparent opacity affects perception of structure from motion and stereo. *Supplement to Investigative Ophthalmology and Visual Science, 30,* 264.

Land, E. H. (1959). Color vision and the natural image. Parts I & II. *Proceedings of the National Academy of Sciences, 45,* 115–129, 636–644.

Mach, E. (1886). *The analysis of sensations.* Translated by S. Waterlow from the 5th German edition. New York: Dover.

Maloney, L. T. & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A, 3,* 29–33.

Magnenat-Thalmann, N. & Thalmann, D. (1987). *Image synthesis: Theory and practice.* Tokyo: Springer-Verlag.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

Marroquin, J. L. (1985). *Probabilistic solution of inverse problems* (A. I. Technical Report 860). Cambridge, MA: MIT.

Metelli, F. (1974). The perception of transparency. *Scientific American, 230,* 91–98.

Metelli, F. (1975). Shadows without penumbra. In S. Ertel, L. Kemmler & L. Stadler (Eds.), *Gestaltentheorie in der modernen psychologie* (pp. 200–209). Darmstadt: Dietrich Steinkopff.

Miller, M. I., Roysam, B., Smith, K. & Udding, J. T. (1990). On the equivalence of regular grammars and stochastic constraints: Applications to image processing on massively parallel processors. AMS-IMS-SIAM Joint Conference on Spatial Statistics and Imaging (A. Possolo, Ed.). American Mathematical Society.

Nakayama, K., Shimojo, S. & Silverman, G. H. (1989). Depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55–68.

Plummer, D. J. & Kersten, D. (1988). Estimating reflectance in the presence of shadows and transparent overlays. *Technical Digest, 11,* Optical Society of America, Washington, DC., pp. WQ3.

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

Poggio, T., Torre, V. & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317,* 314–319.

Ramachandran, V. (1989). Constraints imposed by occlusion and image segmentation. Conference on *Vision and Three-dimensional Representation*, May 24–26, Minneapolis, MN.

Richards, W. & Witkin, A. P. (1979). *Efficient computations and representations of visible surfaces* (AFOSR Report, Contract Number AFOSR-79-0020). Cambridge, MA: MIT.

Terzopoulos, D. (1984). Multigrid relaxation methods and the analysis of lightness, shading and flow. MIT *Artificial Intelligence Memo, 803*.

Terzopoulos, D. (1986). Integrating visual information from multiple sources. In A. Pentland (Ed.), *From pixels to predicates* (pp. 111–142). Norwood, NH: Ablex.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society London B, 203*, 405–426.

Yuille, A. L. (1987). Energy functions for early vision and analog networks. *MIT Artificial Intelligence Memo, 987*.

Zucker, S. W., David, C., Dobbins, A. & Iverson, L. (1988). The organization of curve detection: Coarse tangent fields and fine spline coverings. Proceedings 2nd International Conference on Computer Vision. Tarpon Springs, FL.