Neuron, Vol. 24, 919–928, December, 1999, Copyright ©1999 by Cell Press

Stereoscopic Surface Perception

Barton L. Anderson*
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

Physiological, computational, and psychophysical studies of stereopsis have assumed that the perceived surface structure of binocularly viewed images is primarily specified by the pattern of binocular disparities in the two eyes' views. A novel set of stereoscopic phenomena are reported that demonstrate the insufficiency of this view. It is shown that the visual system computes the contrast relationships along depth discontinuities to infer the depth, lightness, and opacity of stereoscopically viewed surfaces. A novel theoretical framework is introduced to explain these results. It is argued that the visual system contains mechanisms that enforce two principles of scene interpretation: a generic view principle that determines qualitative scene geometry, and anchoring principles that determine how image data are quantitatively partitioned between different surface attributes.

Introduction

One of the primary goals of vision science is to understand how the properties of surfaces are recovered from the structured light that projects to our two eyes. The difficulty in image analysis arises because the visual system has to somehow "undo" the image formation process and infer the multiple causes that act collectively to generate the image data. Variations in reflectance (lightness and/or color), opacity, texture, and three-dimensional shape all contribute to the pattern of luminance that falls on the two eyes from a region of visual space. To accurately recover scene geometry, the visual system must correctly partition the image into the different causes that generated the image data. The discovery of the large number of areas in the brain devoted to vision provided suggestive evidence that the visual system employs a "divide and conquer" strategy, where different surface properties might be computed by distinct mechanisms (Fellemen and van Essen, 1991). Stereoscopic vision is a rather striking example of a research domain that has been dominated by this view. Since the invention of the random dot stereogram (Ashenbrenner, 1954; Julesz, 1960), stereoscopic vision has largely been studied as a system whose primary purpose was to provide information about depth. The primary carrier of this depth information was the positional differences of corresponding features in the two eyes' views: binocular disparity. From this perspective, understanding stereopsis reduces to the problem of discovering the physiological mechanisms that detect these interocular positional shifts and discovering how these local signals are integrated into a coherent representation of surface structure.

Here, a series of novel stereoscopic phenomena are presented that demonstrate a striking dissociation between the pattern of positional signals specified by binocular disparity and perceived surface structure. These phenomena show that stereoscopic mechanisms do much more than simply provide information about depth. Indeed, it is shown that dramatic transformations in perceived lightness, depth, and opacity can be induced without any concomitant changes in the positional signals generated by the binocular disparities present in the two eyes. To explain these findings, a novel theoretical framework is introduced that articulates principles utilized by the visual system to infer surface properties from binocular image data. More specifically, it is argued that the visual system contains mechanisms that enforce two principles of scene interpretation: a generic view principle that determines qualitative surface properties, and principles of anchoring that determine how image data are quantitatively mapped onto a specific representation of a surface's depth, lightness, and opacity.

Results

Stereograms were constructed by viewing a class of textures through apertures placed on a homogeneous background. In all of the experiments described here, the textures contained a uniform disparity and were therefore predicted to appear as a coherent surface in a single depth plane by all extant models of stereopsis. This stimulus configuration allowed the disparity of the texture to be shifted relative to the edges of the aperture, introducing a disparity difference between the aperture boundaries and the texture. The textures could be shifted relative to the aperture boundaries in one of two directions in the two eyes, causing the texture to have a disparity consistent with a surface either behind the aperture boundaries ("far" disparity) or in front of the aperture boundaries ("near" disparity). We studied the effects of this simple manipulation on a broad class of textures. One example is the "one-dimensional" texture generated by a sinusoidal variation in luminance depicted in Figure 1. When the grating was given a far disparity relative to the aperture edges, the grating simply appeared as a flat surface behind the aperture, as predicted by extant stereo models (Marr and Poggio, 1977, 1979; Poggio and Poggio, 1984; Pollard et al., 1985; Jones and Malik, 1992). Any perceived depth variations of the grating were attributable to a tendency to interpret the grating as a shaded 3D surface, which was visible monocularly as well as stereoscopically. However, when the grating was given a sufficiently large disparity that placed it in front of the aperture edges, a strikingly different percept emerged. In this configuration, the grating appeared to split into two layers: a near layer containing a transparent surface that varied

*E-mail: bart@psyche.mit.edu.

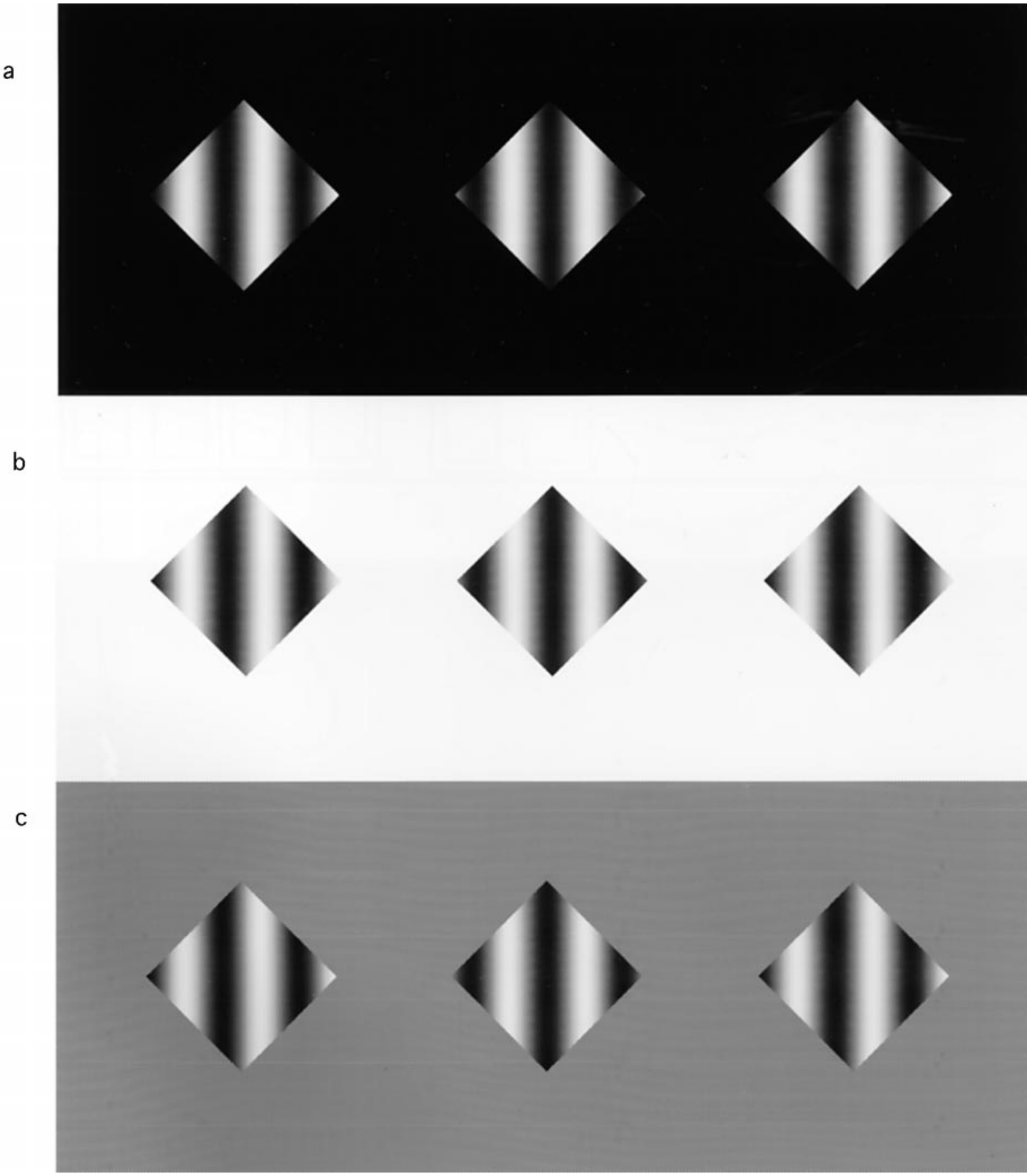

Figure 1. A Depiction of the Stereograms Used in Our Experiments

Identical sinusoidal luminance profiles were viewed through diamond-shaped apertures. Disparity was introduced by shifting the aperture boundaries relative to the gratings. When the right images are cross-fused (or the left two images are fused divergently), the grating appears on a distant surface, visible through a diamond aperture. However, when disparity relationships are reversed (by cross-fusing the left two images or divergently fusing the right two images), the grating appears to split into two depth planes.

(a) The grating region of the pattern appears as a uniform white diamond visible through hazy black stripes. Note that the luminance maxima within the grating appear at the more distant depth layer as part of the diamond, whereas the minima appear as the hazy stripes in front of the diamond.

(b) When the same grating pattern is viewed on a white background, an entirely different percept emerges. The distant layer within the grating now appears as a *black* diamond visible through hazy white stripes. Note that the depth relationships are the inverse of those in (a), despite the fact that the disparity relationships within the grating are identical. The only difference between (a) and (b) is that the luminance of the regions neighboring the diamond apertures was changed from black to white.

(c) When the luminance of the adjacent background fell within the range of luminances present in the sinusoidal grating, the perception of multiple layers is absent or greatly reduced.

in apparent density (or opacity), and a homogeneous diamond-shaped surface at the depth of the aperture edges. This decomposition occurred throughout the texture, despite the fact that all of the binocular regions within the grating contained a single value of disparity.

Remarkably, the apparent lightness of the two layers, the depth attributed to the luminance extrema within the grating, and the apparent opacity of the near layer could be completely inverted by simply varying the luminance of the homogeneous background that neighbored the gratings. When the background was black, the grating appeared as a near layer containing a series of fuzzy black stripes, and a far layer containing a white diamond on a black background (see Figure 1a). The luminance gradients within the grating appeared as variations in the opacity of the near (black) layer. In this configuration, the maxima of the sinusoidal gratings appeared at the depth of the more distant white surface as portions of the apparent diamond, whereas the minima appeared in front. However, when the adjacent background was changed to white (see Figure 1b), the near surface appeared as fuzzy white stripes that varied in opacity, and the far layer appeared as a black diamond on a white background. In this display, the perceived depth of the minima and maxima of the luminance grating reversed: the maxima of the gratings now appeared in front, and the minima appeared as portions of the more distant, black diamond (see Figure 1b). As before, the luminance gradients within the grating appeared as variations in the opacity of the near transparent layer, which now appeared white. When the background luminance was between the extrema of the luminance grating, no coherent percept of two layers was observed (see Figure 1c).

Similar phenomena were observed with a broad class of two-dimensional textures. Figure 2 depicts one example. A uniform disparity texture was viewed through three apertures, and the aperture boundaries were shifted relative to the texture. As in Figure 1, the only difference between the top and the bottom stereo images was the luminance of the background outside the aperture boundaries. When the left two stereo pairs on the top of Figure 2 are cross-fused, the figure appears as three light discs visible through dark-colored mist. Throughout the texture, the mist appears to be approximately uniform in color but varies in its apparent density (or opacity). However, when the adjacent background luminance is changed to light gray, the discs appear dark gray, and the cloudy texture appears as light smoke (Figure 2, bottom). Note that the lightest regions in the top stereo pair in Figure 2 appear as light discs that are unobscured by the dark clouds, but these same image regions appear in the *front* of the disc in the bottom stereo pair (and the dark regions of the texture now appear behind the light mist). The shift in the distribution of perceived depth, lightness, and opacity all arise from a simple change in the luminance of the regions bordering the textured discs. As in Figure 1, no coherent percept of two layers was observed when the background luminance fell between the luminance range within the texture or when the depth relationships between the aperture boundaries and the texture were inverted. These qualitative percepts were confirmed by 53 naive observers that viewed these patterns through a mirror stereoscope.

To understand the surprising quality of these demonstrations, consider the stereograms depicted in Figure 3. These stereograms are identical in structure to those in Figure 2, except that the textured regions are now composed of white noise (random dots). When the dot patterns are given a far disparity relative to the aperture edges, the texture appears in a single plane behind the aperture boundaries, as predicted by all extant stereo models. However, when the disparity relationships are reversed, the texture now appears in a single plane in front of the aperture boundaries, with one caveat: there are thin bands of texture surrounding the central textured regions that appear at the same depth as the circular apertures. These thin textured regions are monocular features (i.e., features seen by only one of the two eyes) that are typically generated along occlusion boundaries. More specifically, when occlusion relationships are viewed binocularly, one eye sees slightly more of the partially occluded surface than the other because it can see around the occluding edge more than the other eye, giving rise to monocular features that are visible in only one of the two eyes (also known as "half occlusions;" see Figure 4a). For occlusion relationships generated along a single depth discontinuity, these features are perceived to lie on the more distant surface in the two eyes and have therefore been described as following a "farthest surface rule" (cf. Julesz, 1964; Nakayama and Shimojo, 1990; Anderson and Nakayama, 1994). Note that this rule correctly predicts that these features should appear at the depth of the central texture when the central texture is the more distant surface. However, when depth is reversed, this rule predicts that monocular features should appear at the depth of the aperture boundaries (since it now is the more distant surface). This accords with observers' reports and is also predicted by extant theories of stereoscopic vision. What is surprising about the percepts experienced in fusing Figures 1 and 2 is that these patterns are not perceived in the same manner as Figure 3, as current theories of stereopsis would predict. Rather, the entire central texture of Figures 1 and 2 appear to split into transparent layers in a manner that depends critically on the contrast polarity of central texture relative to its adjacent background. The remainder of this paper will focus on developing a theoretical framework capable of explaining this striking effect of contrast.

Discussion

The primary focus of stereoscopic theory during the past century has been to explain how the two views are used to reconstruct depth relationships (see Howard and Rogers, 1995). There are two broad "kinds" of information present in the two eyes that have been shown to contribute to this reconstruction process: matchable features that are visible to both eyes (which generate binocular disparities), and unmatchable features visible to only one of the two eyes. Matchable features arise when surfaces project to both of the eyes, whereas monocular regions occur along occluding contours, generated by either the differential occlusion or camouflage of a surface in the two eyes (see Figure 4). One of the main challenges facing stereoscopic theory is to

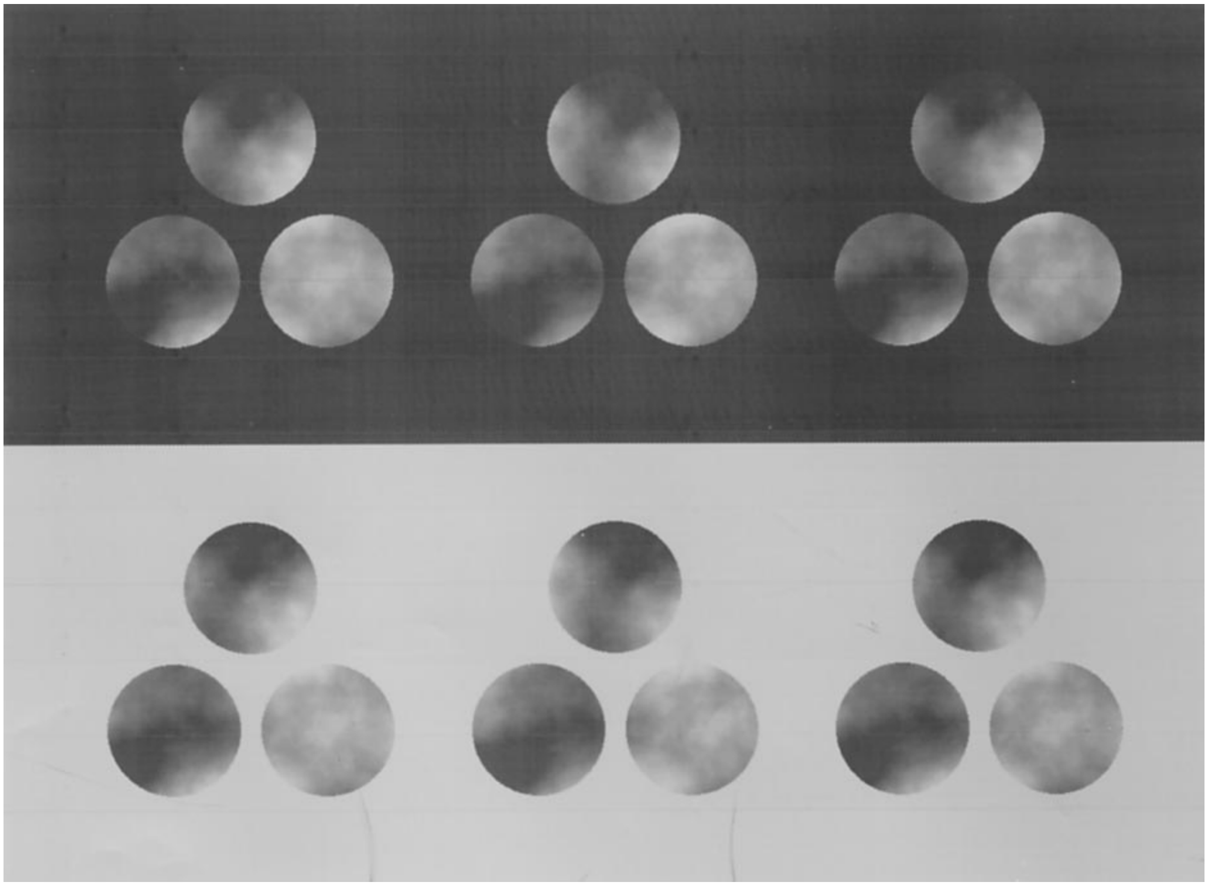

Figure 2. A Stereogram Containing Two-Dimensional Luminance Modulations

As in Figure 1, the texture was identical in the two eyes and was viewed through three circular apertures. Disparity was introduced by horizontally shifting the aperture boundaries relative to the texture. When the right images are cross-fused (or the left two images are fused divergently), the texture appeared on a distant surface, visible through three holes. However, when disparity relationships were reversed (cross-fusing the left two images or divergently fusing the right two images), the texture appeared to split into two depth planes.

(Top) The texture appears as three light discs visible through fuzzy dark clouds. Note that the luminance maxima within the texture appear on the more distant depth plane, whereas the luminance minima appear as portions of the clouds in front of the discs.

(Bottom) When the same texture is viewed on a light background, the distant layer within the texture now appears as three dark discs visible through hazy light clouds. The perceived depth relationships are the inverse of those in the top panel, despite the fact that the disparity relationships within the texture are identical in the two stereograms. Observers also report a compelling completion of the clouds between the gaps of the discs where no texture is present. As in Figure 1, the only difference between the top and bottom panels is that the luminance of the regions neighboring the circular apertures was changed from dark to light.

understand how the visual system correctly determines which image regions have matches—generating binocular disparities—and which do not. Recent psychophysical work has demonstrated that this matching process utilizes the relative contrast of features within each eye to determine whether a given feature is matchable or unmatchable (Anderson and Nakayama, 1994; cf. Smallman and McKee, 1995). For the purpose of the discussion that follows, we assume that this matching process has been successfully accomplished, leading to two sets of matches (a uniform disparity region within the texture and a disparity defined by the contrast of the aperture borders relative to the texture), and the monocular features have been correctly identified (namely, the portions of the texture between the aperture boundaries and the binocularly fused texture). This assumption is reasonable given that virtually all recent stereo models would correctly solve the correspondence problem for these figures. Therefore, the theoretical problem we will focus on here is understanding how the visual system *uses* this pattern of matchable and unmatchable features to infer the underlying surface structure that generated the images.

Since the disparities in the images specify two depth planes, we begin by introducing a general model of image formation generated by surfaces lying at two depths:

$$L(x,y) = [1 - \alpha]\, I_n + \alpha I_f \qquad (1)$$

In this equation, L(x,y) is the total luminance reaching an eye, I_n and I_f are the luminances projected from the near and far layers (respectively), α is the proportion of luminance of the far surface that is actually transmitted by the transparent layer (the proportion of the transparent layer that is "holes"), and $[1 - \alpha]$ is the proportion of light coming from the near layer (the proportion of the transparent layer that is filled by "particles"). In general, all of the terms in this equation can be functions of position (i.e., $\alpha = \alpha(x,y)$, $I_n = I_n(x,y)$, etc.), which means

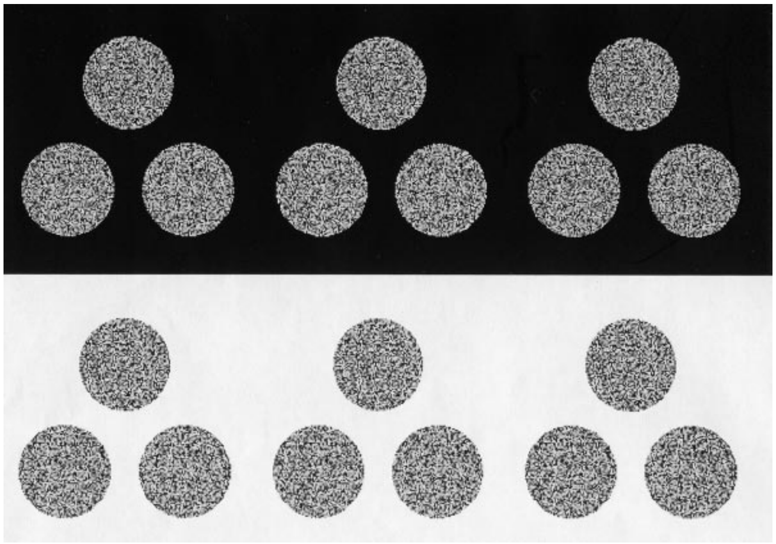

Figure 3. Stereogram Generating a Percept Predicted by Extant Stereo Models

A stereogram containing two-dimensional luminance modulations that does not generate a perception of separate layers. As in Figures 1 and 2, the texture was identical in the two eyes and was viewed through three circular apertures. Disparity was introduced by horizontally shifting the aperture boundaries relative to the texture. When the right images are cross-fused (or the left two images are fused divergently), the texture appeared on a distant surface, visible through three holes. However, when disparity relationships were reversed (cross-fusing the left two images or divergently fusing the right two images), the central region of the texture appears as a single, opaque surface, and the thin bands of monocular features separating the texture and the aperture boundaries appear at the depth of the aperture boundaries, consistent with an interpretation of these features as half occluded. Unlike Figures 1 and 2, this percept is unaffected by the color of the adjacent background.

that this equation can describe any arbitrary image generated by the sum of two layers. Equation (1) generalizes the physical model of transparency introduced by Metelli (1974), which only considered cases in which α and the luminances I_n and I_f were scalar constants. Note that equation (1) can describe both instances of occlusion or transparency, since occlusion would simply correspond to the special case in which $\alpha = 0$ (i.e., when the transmittance of the near layer is zero). Here, we restrict attention to a model containing two depth planes because we are considering stereograms containing only two disparities, and we assume that these have been correctly identified by mechanisms that establish binocular correspondence.

To develop an intuitive understanding of the role played by the different terms in equation (1), consider the percepts achieved when viewing Figures 1 and 2. In the top stereo pair of Figure 2, observers report the appearance of dark clouds floating in front of light gray discs. The apparent variation in the density of the clouds corresponds to variations in α, while the differences in the "color" of the two layers in the top and bottom stereo pairs corresponds to changes in the perceived luminance projected by the near and far surfaces (I_n and I_f, respectively). More generally, I_n and I_f in equation (1) can be written as products of surface reflectance and illumination, but since observers were not required to distinguish between these two dimensions in our experiments, we collapse this term into what Gerbino et al. (1990) refer to as a surface's "effective luminance."

Any cogent theory of the percepts experienced when viewing Figures 1 and 2 must explain why the textures split into two layers, as well as the specific patterns of depth, lightness, and opacity perceived when viewing these figures. Since there are an infinite number of ways that equation (1) can be satisfied physically, the problem confronting the visual system is to determine the most *likely* cause of the stereo images in Figures 1 and 2. This suggests that the perceived interpretation of these images should be that which entails the fewest number of improbable "coincidences." From this perspective, the theoretical problem is to determine what kinds of coincidences the visual system attempts to minimize in its efforts to infer surface structure from binocular images. In keeping with recent theoretical efforts, we will assume that viable image interpretations require that the local image properties are assumed to be stable under a change in viewpoint (i.e., that the observer is situated in a *generic* or *nonaccidental* viewing position) (Koenderink and van Doorn, 1979; Binford, 1981; Malik, 1987; Nakayama and Shimojo, 1992; Freeman, 1993). The intuition behind this principle is that the qualitative relationships within the image data that support a particular scene interpretation should be stable over some range of viewing positions. Although it will be shown that this principle can provide some understanding of why images like those in Figures 1 and 2 appear as multiple layers, it is *not* sufficient to understand the particular pattern of perceived lightness and transparency in these (or any) images. An additional principle is needed that describes how the specific patterns of depth, lightness, and opacity are quantitatively distributed between the near and far surfaces.

To understand why the texture appears to split into two layers when its disparity is nearer than the aperture border, but not when the texture is behind the aperture boundary, consider the depth information within the texture in Figures 1 and 2. In both images, the textured regions and the aperture borders intersect, yet lie in different depth planes. The texture contains both monocular and binocular features. The binocularly visible regions generate two disparities: one within the texture, and another along the aperture boundaries. This displacement also generates bands of monocular texture that are situated between the aperture boundaries and the binocularly visible texture (see Figure 4). As mentioned above, monocular features can be generated in two ways: by the differential occlusion that allows one eye to see around an occluding edge more than the other (Gillam and Borsting, 1988; Nakayama and Shimojo, 1990; Anderson, 1994; Anderson and Julesz, 1995); or by the camouflage of a near surface against

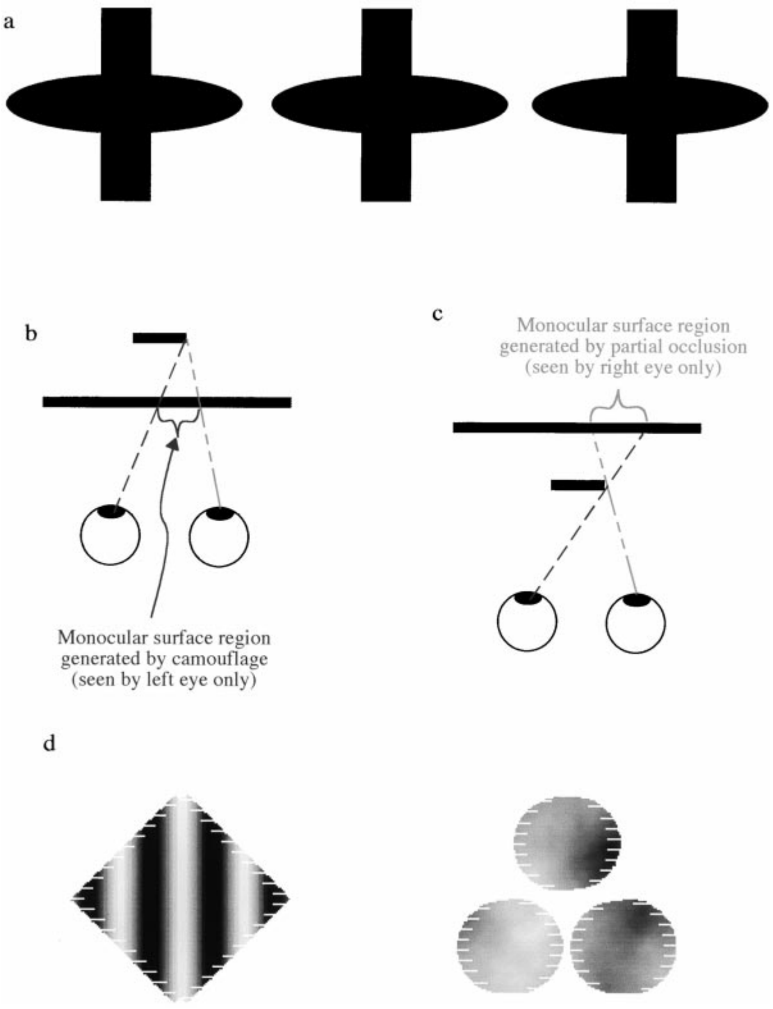

Figure 4. A Schematic of the Two Ways that Monocular Features Can Be Generated during Binocular Viewing

(a) When the two left images are cross-fused (or right two are fused divergently), an oval-shaped surface appears in front of a vertical black bar. When the right two images are fused, the opposite depth pattern results.

(b) When the oval appears in in front of the bar, some portions of the oval are seen by only one eye because of the differential camouflage of the oval by the bar in the two eyes.

(c) When the depth relationships are reversed, the portions of the oval that are seen by only one eye arise from the differential occlusion of the oval in the two eyes rather than camouflage.

(d) A schematic depicting the locations of the unmatched monocular features in Figures 1–3. In these images, the binocularly matchable portions of the gratings (a) or 2D textures (b) were aligned in a single image so that the disparity of the aperture boundaries relative to the texture could be seen clearly. The hatched regions indicate portions of the texture that are visible in only one of the two eyes.

a distant background (Von Szily, 1921; Anderson and Julesz, 1995; Anderson, 1997). When the texture appears behind the aperture boundaries, both the disparity relationships and the unmatched monocular features support the same interpretation: a single surface visible behind occluding apertures. When the disparity relationships are reversed, however, the interpretation of the unmatched monocular features as due to occlusion competes with an interpretation that attributes these features to camouflage. For this depth configuration, the camouflage interpretation wins in Figures 1 and 2: observers report that portions of the texture appear to disappear into the background and can even appear to generate a form of textural completion. The question is why this interpretation wins over the putatively “simpler” interpretation of the binocular visible texture forming a single opaque surface, with the monocular bands appearing at the depth of the aperture boundaries (such as that experienced when fusing Figure 3). Clearly, the answer to this question must lie in the properties of the textures used, since this is the only property that distinguishes Figure 3 from Figures 1 and 2.

To gain insight into the critical differences between these textures and their strikingly different perceptual outcomes, consider what the interpretation of the near texture as a simple occluding surface would entail. This interpretation requires splitting the monocularly continuous texture within each eye into two different surfaces, placing the binocular regions of the texture in the near depth plane, consistent with its disparity, and placing the monocular regions at the position of the more distant surface (following a “farthest surface rule” for half occlusions). There are a number of problems with this interpretation for the textures used in Figures 1 and 2. Intuitively, the textures in Figures 1 and 2 contain monocularly conspicuous large scale structure, whereas in Figure 3, this large scale information is not perceptually salient (i.e., the textural variation appears to exist primarily on a fine scale; cf. Field and Brady, 1997). A simple occlusion interpretation would entail “breaking” the monocular continuity of this large scale structure without any local, small scale monocular information that would support the presence of such discontinuities. Indeed, it is essentially impossible to generate abrupt depth discontinuities in such low frequency textures without generating monocularly conspicuous edges in at least one of the two eyes. This can be seen readily by viewing Figure 5. In this figure, a region of the texture has been shifted horizontally by the magnitude of the disparity used in our displays. Note that a vivid monocu-

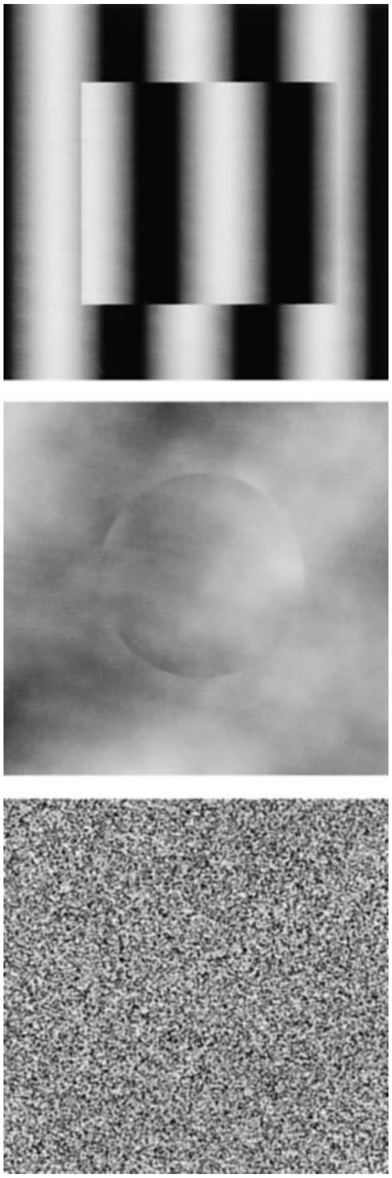

Figure 5. A Schematic Demonstrating the Consequences of Placing Statistically Similar Texture on Two Different Depth Planes

The purpose of this figure is to illustrate that it is essentially impossible to generate depth discontinuities in many textures without generating luminance discontinuities in at least one of the images. In these images, a region of texture has been displaced by a magnitude similar to that used in our experiments in one of the two eyes. Note that this generates a monocularly visible contour in the top two figures but not in the bottom figure.

lar discontinuity is generated by this displacement, whereas no such textural "breaks" are present within the textures in Figures 1 or 2. In contrast, the fine spatial structure present in Figure 3 does generate local contrast variations along the depth discontinuity, but it would not lead to monocularly visible edges because shifts occur in integer multiples of the pixel size (see Figure 4). Thus, the interpretation of the monocular features in images such as Figures 1 and 2 as half occluded is putatively overridden by monocular signals that specify the continuity of the large scale structure present in these images.

The alternative interpretation of the monocular features generating along the depth discontinuities is that they are due to the texture continuing across the aperture edges, disappearing into the background because they are camouflaged (which accords with observers' reports). This interpretation would maintain the monocular continuity of the large scale structure in these textures, but it requires that the entire texture be decomposed into two separate depth planes. There are a number of ways this decomposition could occur, but the only generic way that a near surface could continue into a background while maintaining complete camouflage is if it projected the same luminance as the background. Formally, this implies that the luminance I_n projected by the near surface must equal the luminance of the adjacent background:

$$I_n = I_b \tag{2}$$

where I_b is the luminance of the background adjacent to the apertures. Since $I_f = I_b$ in the region of the background by definition, combining equation (1) with equation (2) gives:

$$L(x,y) = [1 - \alpha]\, I_b + \alpha I_b = I_b \tag{3}$$

which will hold for any value of α. In contrast, if $I_n \neq I_b$ (i.e., if the near surface does not project the same luminance as the adjacent background), the far and near surfaces would not combine to equal I_b, and the conditions for camouflage would not hold [since α is restricted to be in the interval (0,1)]. Thus, the disappearance of the texture along the aperture boundaries can occur generically only if equation 2 holds, which implies that the luminance variations arising within the near texture are due solely to the transmittance $\alpha(x,y)$ of the near surface.

This analysis also provides insight into why coherent percepts of transparency are not perceived when the background luminance falls between the range of luminance values within the texture. The constraints on camouflage imply that if the background is gray, then the transparent surface must also be the same shade of gray. This in turn implies that the regions within the texture that are brighter and darker than this gray value must be due to variations in brightness of the underlying layer. Note, however, that all of the binocularly visible luminance variations within the texture occur in the *near* depth plane. This could only occur if the near layer contained a series of opaque surface patches that just happened to be perfectly aligned with the contrast variations of the underlying surface, causing them to be occluded in both eyes. This clearly involves a highly accidental viewing geometry, since any small perturbation in viewing position would reveal the presence of contrast variations in the distant layer. Thus, for this luminance configuration, both the transparency interpretation and the occlusion interpretation of the near surface are highly improbable, generating an incoherent and unstable surface percept (see Figure 1c). Note, however, that when the disparity relationships are reversed, the disparity information and the monocular features are again consistent with a single, opaque surface appearing behind an aperture, so that this instability should only be present in one of the two depth configurations, which is consistent with observers' reports.

Thus, a generic view principle can provide an understanding of why the images presented in Figures 1 and 2 appear as two layers and why the luminance variations of the texture should be interpreted as variations in surface opacity $\alpha(x,y)$. However, this alone is not sufficient to understand the specific percepts of lightness and opacity achieved with these patterns, since equation (2) will hold for any values of α. This means that there are still an infinite number of possible solutions available to the visual system, involving different combinations of surface lightness, opacity, and depth. The problem, then, is to understand how the visual system partitions the continuous luminance distribution between the two layers [i.e., how it resolves the ambiguity in assigning specific values to the transmittance function $\alpha(x,y)$].

In order to assign transmittance values to an inhomogeneous transparent layer, the visual system must decompose the image luminance between the near and far surface planes. Here, I will focus on how the two endpoints of this mapping are inferred or "anchored": regions of complete opacity, and regions of complete transmittance. We will assume that the mapping of transmittance values between these two anchor points behaves in a simple monotonic manner (which is supported by our demonstrations and data). The anchor point of complete opacity is "natural" in the sense that occluding surfaces cause the contrast of underlying surfaces to vanish. This constraint implies that percepts of complete occlusion should only occur in regions in which the luminance within the texture equals the luminance of the adjacent background, since this is the only luminance that would cause the contrast of the far contour (the aperture boundaries) to vanish. This is consistent with observers' reported percepts (note that the most opaque regions in Figures 1 and 2 occur in regions where the contrast between the aperture boundary and texture are smallest and near zero). The other end of the scale does not have a similar "natural" anchor point that can be derived from the physics of transparency or occlusion. This is because any given contrast could have been generated by an unobscured surface patch of (say) moderate contrast or by a higher contrast surface patch that is partially attenuated by a near transparent layer. However, the images depicted in Figures 1 and 2 reveal how the visual system anchors regions of complete transparency. In particular, the percepts experienced when fusing the images in Figures 1 and 2 demonstrate that the regions of maximal contrast of the more distant contour are treated as regions that are completely transmissive (or unobstructed, i.e., where $\alpha = 1$). When the contrast between the texture and the aperture border is maximal, observers report that *all* of the luminance in these regions appear to arise from the underlying surface (see Figures 1 and 2). Expressed differently, regions of maximal contrast appear unobscured by a transparent layer; they simply appear as holes in an inhomogeneous transparent surface. Note that this is true even though the luminance ranges in Figures 1 and 2 are quite different: Figure 1 ranges from black to (nearly) white, whereas Figure 2 ranges from light to dark gray. Nonetheless, the regions of maximal contrast in both figures appear completely transmissive (i.e., as unobscured "holes"). These anchoring principles correctly predict the shift in apparent depth of the luminance maxima and minima in both Figures 1a and 1b and Figure 2 (top and bottom panels), since the transmittance values of the near surface are predicted to shift from zero to one (and hence, the attribution of luminance shifts from the near to far layer in equation [1], respectively). Regions between these two extremes appear with intermediate values of surface opacity that vary smoothly and monotonically between these two extremes.

The principles of transmittance anchoring described above make strong predictions about how image luminance is partitioned between the two layers. If the highest contrast regions are interpreted as image locations that provide an unobscured view of the underlying surface, then the apparent luminance of these regions should determine the perceived luminance of the distant layer within the texture. To test this hypothesis, a series of experiments was performed that required observers to match the perceived luminance of the distant surface elicited by the grating patterns depicted in Figure 1 (see Experimental Procedure). The spatial frequency and mean luminance of the gratings were constant across experimental conditions, but the amplitude of the grating was changed in different blocks of trials. Observers performed two sets of experiments. In the stereoscopic depth condition, observers viewed the images depicted in Figure 1 through a stereoscope and adjusted the luminance of a square test patch to match the apparent luminance of the far (diamond-shaped) surface. In the nondepth condition, observers adjusted the luminance of the test patch to match both the darkest and lightest regions in the grating when Figure 1 was viewed without any depth differences within the pattern. The anchoring theory described above predicts that the perceived brightness of the far layer should match the perceived luminance extrema of the grating, since these are the regions that should appear as unobscured "windows" onto the more distant surface layer. Note, however, that the nonlinear transformation of luminance by early visual processing implies that the perceived luminance of the grating's extrema are not expected to be identical to the actual luminance values in the stimulus. We therefore had observers match the perceived luminance of the luminance extrema without any depth differences, so that we could compare these settings to those measured for the stereoscopic settings. Since we found that the decomposition of the texture into layers only occurred when the luminance of the background was outside the range of luminances within the texture, subjects only performed this matching experiment in these luminance regimes (since the task did not have any meaning in the other conditions). The results of this experiment are presented in Figure 6. These data demonstrate that observers' judgements of the luminance of the distant (diamond) surface (Figure 6, top) are essentially identical to the perceived luminance of the luminance extrema within the texture (Figure 6, bottom), providing strong experimental support that the visual system treats the highest contrast regions along the aperture boundaries as regions of 100% transmittance. Similar observations hold for the images depicted in Figure 2. Note that here,

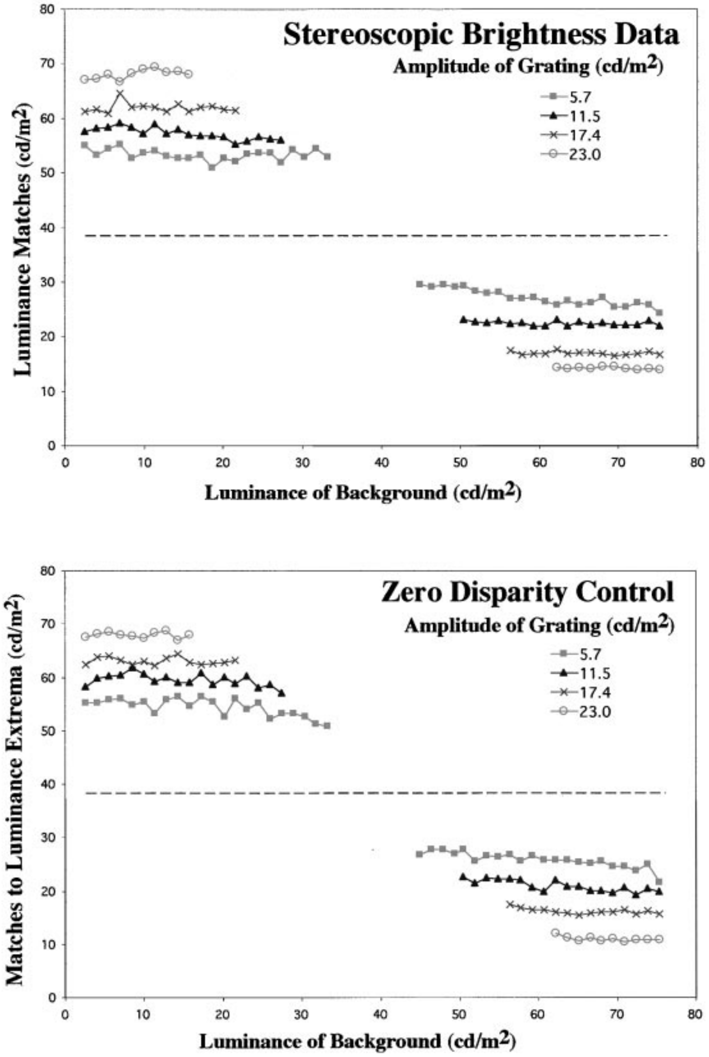

Figure 6. Averaged Data from the Brightness Matching Experiments for the Three Observers

(Top) Each curve represents observers' settings of a small test patch to match the apparent brightness of the perceived diamond. The different curves represent different amplitudes of the sine grating, and each data point represents a match for a given luminance of the adjacent background (see Experimental Procedures). The mean luminance of the grating was constant in all experiments. When the background was darker than the grating (left-hand side of graphs), observers perceived the diamond as a light surface; when the background was darker than the grating (right-hand side of graphs), observers perceived the diamond as a dark surface. The gaps in the middle of the graphs correspond to background luminance values that fell between the luminance values of the gratings, which did not give rise to a coherent percept of two surfaces. Note that there is very little effect of the contrast magnitude on the perceived brightness of the distant layer; the largest determinant of its perceived brightness is the polarity of the grating relative to the background.

(Bottom) A control experiment in which observers adjusted a test patch to match the apparent brightness of the luminance maxima for backgrounds darker than the grating (left-hand side of graph) and the luminance minima of the grating (right-hand side of graph) when the grating and aperture boundaries had the same disparity. Note that the data are essentially identical to the stereoscopic matches in the top panel.

too, the highest contrast regions along the aperture boundaries appear perfectly transmissive.

The results reported here demonstrate that the computation of surface structure from stereoscopic images is performed by mechanisms that infer the opacity of occluding and transparent surfaces from the contrast relationships arising along depth discontinuities. The interpretation of such discontinuities was shown to have a dramatic and nonlocal effect on perceived depth, lightness, and opacity. Whereas previous research has emphasized the modularity of visual processing, the results described herein demonstrate a strong coupling between the diverse computations of depth, lightness, and opacity. I have argued that at least two explanatory principles were needed to understand these phenomena. First, a generic view principle provides an explanation of the luminance and geometric conditions that initiate the decomposition of a stereoscopic texture into more than one surface. However, although this is a necessary component to a theory of scene interpretation (cf. Nakayama and Shimojo, 1992), a generic view principle remains underconstrained and does not specify a unique solution to the interpretation of a stereoscopically viewed scene. Rather, a second principle of *transmittance anchoring* is needed to understand the specific manner in which luminance is partitioned between the different depth layers. This principle not only explains why we see the pattern of inhomogeneous transparency in Figures 1 and 2, but it also explains why we do not always see the world as though we were viewing it through a transparent haze. Without an anchoring principle of this kind, this simple fact of everyday experience cannot be understood, since any given contrast could have been generated by either a single surface or a higher contrast surface visible through a semitransparent medium.

The results and analysis presented here provide novel insights into the rich set of computations employed by the visual system to recover surface structure from binocular images. Indeed, these results demonstrate that the pattern of positional signals of corresponding image points—binocular disparities—do not always provide sufficient information to derive stereoscopic surface structure or even the perceived depth of disparate image regions. These results suggest that any complete theory of stereoscopic surface perception requires understanding the neural mechanisms that enforce these two principles of scene interpretation, rather than simply how binocular positional shifts of corresponding image points are determined.

Experimental Procedures

Three observers with normal or corrected to normal vision participated in the experiments. Two observers were naive as to the purposes of the experiments, and the third was the author (B. L. A.). The stimuli consisted of vertically oriented sinusoidal luminance profiles viewed within a diamond aperture. The major axes of the diamond aperture subtended 2.86°, and the spatial frequency of the grating was 0.8 cycles/degree. The patterns were viewed through a haploscope at a distance of ~40 in. A disparity of 13.4 arc min disparity was introduced to the grating pattern in the nonzero disparity viewing condition. A binocular square test patch subtending 0.72° appeared at the same depth plane as the diamond-shaped aperture boundaries 1.5° below the grating stimuli. To insure that the test patch was visible for all gray scale values, the test patch was placed on a 1.43° black and white checkerboard pattern whose mean luminance matched the mean gray of the monitor (~39.1 cd/m^2). The monitor was calibrated such that the luminance values were a linear function of the 8-bit look-up table values (ranging from 1.7 cd/m^2 to 76.1 cd/m^2). The mean luminance of the grating was fixed for all experiments at 39.07 cd/m^2. There were four different amplitudes tested. The maximum luminances of the gratings were 62.1, 56.3, 50.4, and 44.6 cd/m^2, and the corresponding luminance minima were 15.7, 21.6, 27.4, and 33.2 cd/m^2. The luminance of the homogeneous background adjacent to the diamond apertures was varied randomly from trial to trial.

Each observer participated in four blocks of trials, one block for each amplitude of the sine-wave grating. Within each block, the amplitude of the grating was held constant. The luminance of the homogeneous background was restricted to values outside the range of luminances used within the grating, since pilot work had revealed that it was only in these conditions that the grating would appear to split into two coherent layers. For a given amplitude of the grating, the remaining color table values were divided into equal intervals of 1.46 cd/m^2, which were randomly selected from trial to trial. During an individual trial, the brightness of the small test patch was set to a random value, and observers adjusted the luminance of this patch with a mouse. In the stereoscopic conditions, observers adjusted the luminance of the test patch until it appeared identical to the apparent brightness of the illusory diamond underlying the grating. In the zero disparity control conditions, observers adjusted the luminance of the test square to match the apparent brightness of the luminance minima and maxima. Three observers performed ten matches for each stimulus. The data presented in Figure 3 represent the means of the three observers.

Received June 7, 1999; revised October 15, 1999.

References

Anderson, B.L. (1994). The role of partial occlusion in stereopsis. Nature *367*, 365–368.

Anderson, B.L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. Perception *26*, 419–453.

Anderson, B.L., and Julesz, B. (1995). A theoretical analysis of illusory contour formation in stereopsis. Psychol. Rev. *102*, 705–743.

Anderson, B.L., and Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. Psychol. Rev. *101*, 414–445.

Ashenbrenner, C.M. (1954). Problems in getting information into and out of air photographs. Photogram. Eng. *20*, 398–401.

Binford, T.O. (1981). Inferring surfaces from images. Artif. Intell. *17*, 205–244.

Felleman, D.J., and van Essen, D.C. (1991). Distributed hierarchical processing in primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Field, D.J., and Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. Vision Res. *37*, 3367–3384.

Freeman, W.T. (1993). The generic viewpoint assumption in a framework for visual perception. Nature *368*, 542–545.

Gerbino, W., Stultiens, C.I., Troot, J.M., and de Weert, C.M. (1990). Transparent layer constancy. JEP: HPP *16*, 3–20.

Gillam, B., and Borsting, E. (1988). The role of monocular regions in stereoscopic displays. Perception *17*, 603–608.

Jones, J., and Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. Image Vision Comput. *10*, 699–708.

Julesz, B. (1964). Binocular depth perception without familiarity cues. Science *145*, 356–362.

Koenderink, J.J., and van Doorn, A.J. (1979). The internal representation of solid shape with respect to vision. Biol. Cybernet. *32*, 211–216.

Malik, J. (1987). Interpreting line drawings of curved objects. Int. J. Comput. Vision *1*, 73–103.

Marr, D., and Poggio, T. (1976). Cooperative computation of stereo disparity. Science *194*, 283–287.

Marr, D., and Poggio, T. (1979). A computational theory of human stereo vision. Proc. R. Soc. Lond. B *204*, 301–328.

Metelli, F. (1974). The perception of transparency. Sci. Am. *230*, 90–98.

Nakayama, K., and Shimojo, S. (1990). DaVinci stereopsis: depth and subjective occluding contours from unpaired image points. Vision Res. *30*, 1811–1825.

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

Poggio, G., and Poggio, T. (1984). The analysis of stereopsis. Annu. Rev. Neurosci. *7*, 379–412.

Pollard, S.B., Mayhew, J.E.W., and Frisby, J.P. (1985). A stereo correspondence algorithm using a disparity gradient limit. Perception *14*, 449–470.

Von Szily, A. (1921). Stereoskopische Versuche mit Schattenrissen. Alkbrecht Graefes Arch. Ophthalmol. *105*, 964–972.