Constructing and representing visual objects

Manish Singh and Donald D. Hoffman

The objects we see are not given in the images at the eyes, but must be constructed by the human visual system. Indeed, damage to specific brain regions often leads to specific impairments of visual abilities (for example, the perception of shape, color or motion). Human vision constructs the various properties of visual objects, not independently of each other, but in a highly coordinated fashion. The construction of one visual property strongly influences the constructions of other properties. Visual shape is an important construction for successfully recognizing objects. There is growing consensus that human vision represents shapes in terms of component parts and their spatial relationships. These parts and their spatial relationships provide a powerful first index into one's visual memory of shapes.

Seeing objects and their properties is a remarkable achievement of human vision. Indeed objects are not given in the images in the eyes, but must be constructed by the visual system from these images. A striking example of this is the case of Mr S., who suffered diffuse damage to his cerebral cortex from accidental poisoning by carbon monoxide. After the accident he had normal acuity and color vision, and he could see motion. However, he could not see objects, even though he could identify them by sound or touch. He could not, from visual cues alone, name letters, numbers, or common objects, or even recognize himself and family members. In other words, Mr S. was unable to put together his experiences of edges, colors and motions, into experiences of visual objects. He was diagnosed as having visual form agnosia[1].

Another class of patients first discovered in 1909 by Balint[2] and now called dorsal simultanagnosics, can see only a part of an object or sometimes a single small object, at a time and have difficulty holding even that in attention[3,4]. Their condition usually follows bilateral damage to the parietal and occipital lobes, and is typically characterized as an attentional deficit, since they often have full visual fields. These patients can put together their perceptions of edges, colors and motions, but only for one object or part of an object at a time. They can see parts of a visual scene, but never the whole scene. These cases, and many others like them, suggest that human vision constructs visual objects and their properties.

The image available at the retina of the eye is discrete. It consists of a set of photons captured by an array of photoreceptors. But what we perceive are recognizable objects localized in three-dimensional (3D) space, having continuous surfaces, boundaries and shapes, as well as specific colors and motions. In the patients considered above cerebral damage has led to the selective impairment of processes that are responsible for constructing these visual objects, or for allocating attention to them. Indeed, there is growing evidence that visual attention is allocated not to regions of space, but to parsed objects and their parts[5]. For instance, one 69-year-old patient with damage to his right hemisphere was able to segment the visual world into objects, but then could not pay attention to the left half of each object, no matter where the object happened to be in the visual field.

Among the many properties of visual objects that human vision constructs are shape (both in 2D and 3D), motion, color, surface properties (such as transparency, opacity and texture), location in 3D space and illumination. As Barlow and others have suggested, for some of these properties specific brain regions are critical for their construction[6,7]. For example, lesion studies of visual area V4 of the macaque and position emission tomography (PET) studies of the human lingual and fusiform gyri of prestriate cortex suggest that these areas are crucial for the perception of color[8,9]. Damage to this area in humans leads to cerebral achromatopsia, a complete loss of color sensation, despite normal functioning of the retinal cones[10–12]. Similarly, area V5 of the macaque, and a region of cortex at the junction of the temporal, parietal and occipital lobes in humans, have been identified as being crucial for the perception of visual motion[13]. Damage to the area in humans leads to a peculiar condition called akinetopsia in which the patient can see and recognize objects, but is unable to see their motions: objects seem frozen in time, and appear to jump suddenly from one location to another.

Coordinated construction of visual objects

That different cortical regions are necessary for different functions does not entail however, that these regions function independently of each other. Although it is, at times, a

M. Singh and D.D. Hoffman are at the Department of Cognitive Science, University of California, Irvine, CA 92697, USA.

tel: +1 714 824 6795
fax: +1 714 824 2307
e-mail: ddhoff@uci.edu

1364-6613/97/$17.00 PII: S1364-6613(97)01024-3

useful strategy to study them separately as modular systems, psychophysical evidence suggests that in fact the regions interact to a high degree[14,15]. Human vision constructs the various properties of visual objects in a highly coordinated fashion – so that a specific interpretation of one visual property will strongly affect how the other properties are interpreted.

Consider the phenomenon of neon color spreading[16,17]. Figure 1A displays an example by Redies and Spillmann[18], in which we perceive a transparent red disk in front of intersecting black lines. A desaturated red seems to fill the disk to its edges. Our visual systems construct, in careful coordination, the shape, color and transparency of the disk. As a result, we see red where a photometer would not detect any red at all. Figure 1B shows similar coordination in the neon worm, a simple modification of the Redies and Spillmann figure by Hoffman[19].

Figure 2 displays a stereo example of neon color spreading, similar to the displays of Nakayama *et al.*[20], and Kojo *et al.*[21] Fuse the two sides of Fig. 2A and you will see a subjective surface that curves towards you in three dimensions and floats in front of the black circles. The surface appears transparent, glowing, and a desaturated blue. Interchanging the two sides (as in Fig. 2B) leads to a switch in stereo disparity. The surface now appears to curve away from you behind the white page, and the black circles look like holes through which you see the blue surface. What is remarkable is that the surface now appears opaque, not glowing, and a saturated blue. So in this case, the properties of 3D shape, color and surface quality (transparent or opaque) are all constructed in coordination with relative depth, and all are affected together by a switch in stereo disparity.

Human vision can also construct objects by an interaction of color and motion. This is nicely demonstrated by displays of dynamic color spreading, in which motion induces a spread of color. Figure 3 shows an example by Cicerone and Hoffman[22,23] (see also Shipley and Kellman, Ref. 24). Figure 3A shows a still from a movie. The frame has 900 dots placed at random according to a uniform distribution. Most of the dots are colored red, except for a small set in the center, which are colored green. In this static display, there is little spread of color, and the shape of the green region looks ragged. Figures 3B and 3C show two subsequent frames from the movie. In each frame, the dot positions remain exactly the same. All that changes are the groups of dots that are colored green. But when these frames are set in motion, one perceives a transparent green filter moving smoothly over the red dots. The filter seems to glow and has a perfectly circular shape. In fact, by choosing which dots are colored green in any given frame, the movie can be made so that human vision will construct a glowing 3D object translating and rotating in three dimensions. (See Cortese and

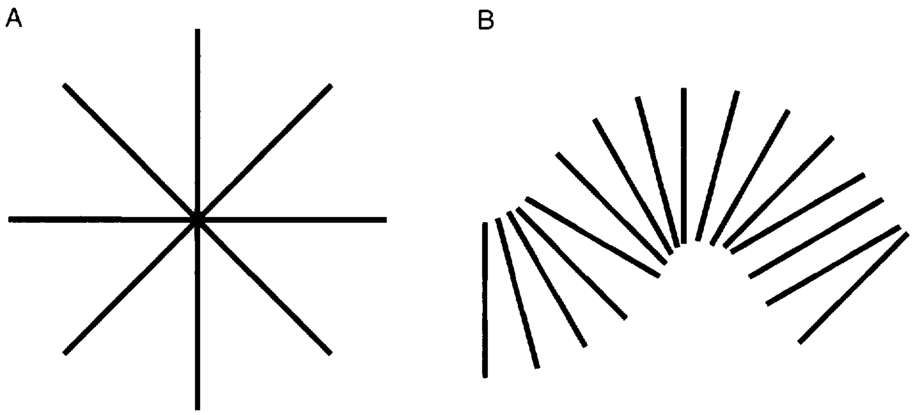

Fig. 1 Neon color spreading. (A) The Redies and Spillmann[18] display of neon color spreading. **(B)** The neon worm by Hoffman[19].

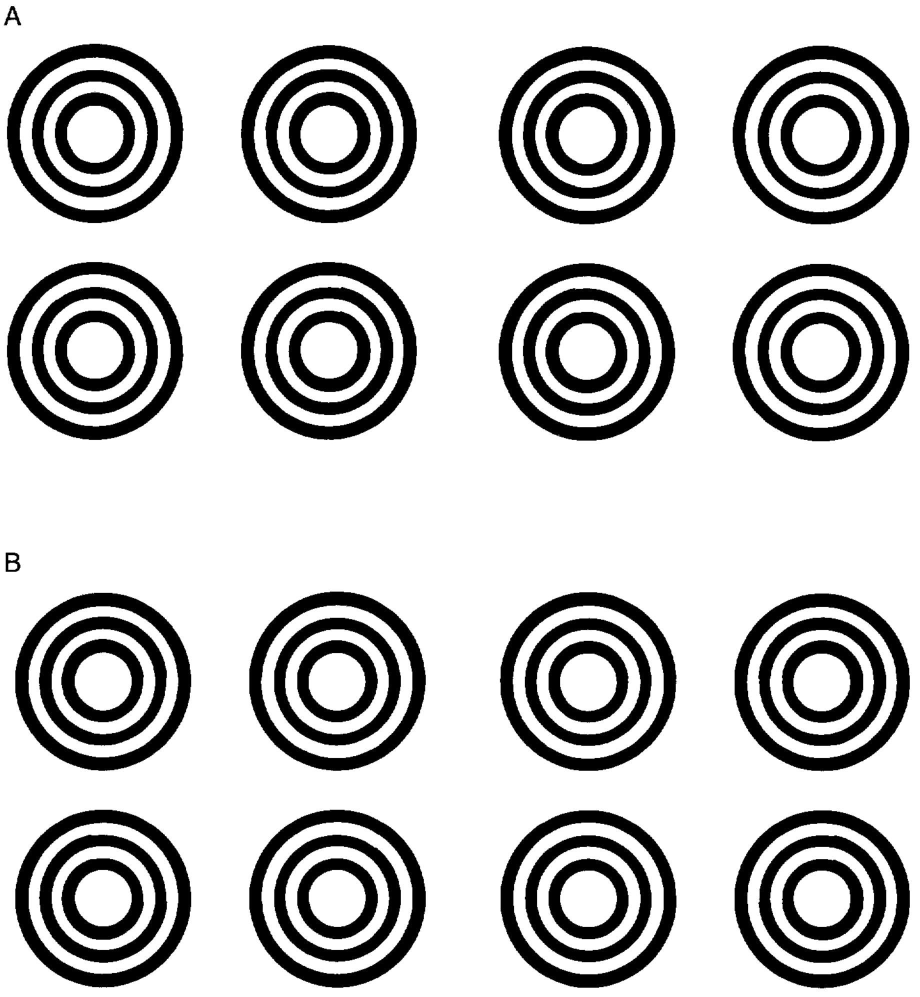

Fig. 2 Stereo displays of neon color spreading. In order to view these figures fuse the left and right sides of each display by crossing your eyes slightly. Both **(A)** and **(B)** have the same elements, but the left and right sides have been reversed, leading to a switch in stereo disparity.

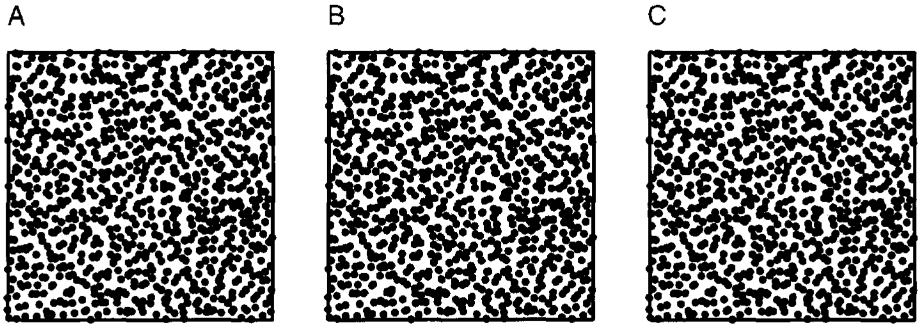

Fig. 3 Dynamic color spreading. Three frames from a motion display of dynamic color spreading. The dot positions remain fixed, but different dots are colored green between one frame and the next. The short film from which these images were taken can be viewed at the following website: http://www.socsci.uci.edu/cogsci/personnel/gstudents/singh/illusions/cfm.html

Anderson[25] for an achromatic precursor to this.) Thus, from a few dots that change color but do not move, human vision constructs an object with a definite shape (in 2D or 3D), with specific color and surface properties even in the region between the dots, and with a specific motion. This is a far more elaborate set of properties than human vision constructs when presented with the well-known displays of apparent motion studied by Wertheimer and others[26]. As described above, all these properties are constructed in an interactive fashion to yield a consistent interpretation.

Another point that emerges from these and other examples is that color cannot simply be equated with surface reflectances, or even with triples of surface reflectances filtered through the cone sensitivity functions[27]. Color is a complex construction of human vision, and one that is carefully coordinated with the construction of other visual properties, such as shape, motion and depth, and with surface qualities such as transparency and opacity. A striking example of this is illustrated in Fig. 4, which displays a chromatic version of White's illusion[28].

Representing the shapes of visual objects

Shape plays a key role in representing visual objects and this is especially so for the purpose of recognition[29–31]. For example, a large number of objects are recognizable from their silhouettes alone. Indeed, experiments by Biederman and Ju[32] suggest that humans are as fast and accurate in recognizing objects from line drawings (that is from information on shape alone) as from full color pictures. The natural question that arises then is: how does human vision represent shapes? For instance, are its representations viewpoint-dependent or viewpoint-independent? A viewpoint-independent representation of an object is a single canonical (that is standard) model for the object that can be constructed by human vision from almost any view[30,33]. On the other hand, a viewpoint-dependent representation consists of multiple models of the same object, each corresponding to a different set of views. These multiple models can be either two-dimensional[34] or three-dimensional[35].

From a computational perspective, each type of representation presents different advantages and disadvantages. For example, a viewpoint-independent representation requires less memory, since it uses a single model per object. However, it requires more computation because features in the model are not contained in the image directly and must therefore be computed. On the other hand, viewpoint-dependent representations require more memory, since many views need to be stored for each object. But they require less computation because of the greater similarity between the viewed image and (at least some of) the stored views of the object. For viewpoint-dependent representations, novel views of an object can be recognized by linear combinations of stored views[36], by interpolating stored views[34], by mental rotation of stored views[37] or by aligning stored views to the image[38].

The type of representation used depends critically on the purpose it is meant to serve. This purpose might be categorization at the basic level (for example, cat), or recognition of the individual object (for example, Tabby), or motor manipulation of the object (grasping and handling). It is likely that the visual system has multiple representations of the same information in order that the representations can serve different purposes.

Component parts and their spatial relationships

There is a growing consensus among researchers that human vision represents shapes in terms of component parts and the spatial relationships between these parts. Parts provide a computationally useful way of dealing with occlusion, including self-occlusion, and with lack of rigidity (the fact that many objects do not have fixed shapes because they have moving parts). Occlusion and nonrigidity pose serious problems for traditional approaches such as template theories and Fourier models[39]. Furthermore, converging experimental evidence suggests that human vision does parse shapes into parts, and that it does so quickly and automatically[30,40–44]. Indeed, parts are compatible both with viewpoint-dependent and viewpoint-independent representations, and with 2D and 3D representations[43].

One common approach to the problem of object parts has been to postulate that human vision stores in memory a set of basic shape primitives which it looks for in images. By finding these primitives in objects, it not only parses the objects into parts, but also represents them in terms of the primitives. Shape primitives that have been studied include generalized cylinders and cones[33,45], superquadrics[46] and geons[30]. To recognize a complex shape, like grandmother,

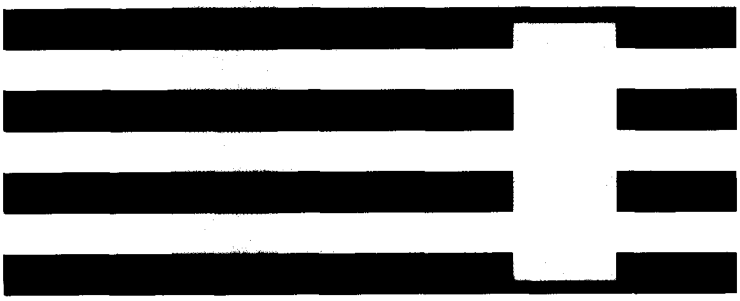

Fig. 4 White's illusion. A chromatic version of White's illusion[28]. The two sets of light blue bars are, in fact, indistinguishable to a photometer.

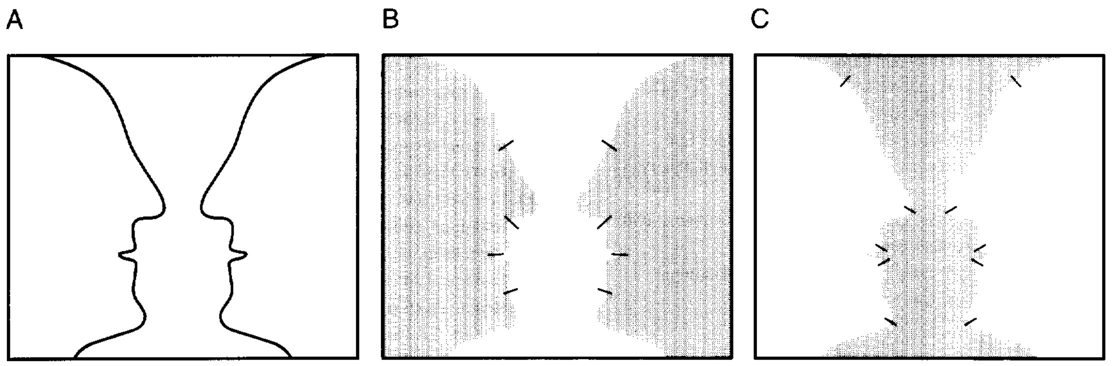

Fig. 5 The face–goblet illusion. (A) The image. The minima rule predicts that we perceive different parts when we see **(B)** the faces as figures, than when we see **(C)** the goblet as figure.

one first tries to find these simpler shape primitives. Each of these schemes works well on a special class of object shapes. However, each misses parts whose shapes are not in the predefined set of primitives. Hence, none can account for the variety of part shapes that we see and recognize.

A different approach is to separate the issue of finding parts from the issue of describing them. Hoffman and Richards[39] have argued that parsing shapes into parts is carried out by low-level mechanisms that operate regardless of what descriptions the parts eventually get. In other words, human vision makes no prior assumptions about what shapes it will encounter, but uses general computational rules to parse objects into parts. These rules are based on geometrical properties alone. Hence, they apply quite generally.

The minima rule[39] is an important step in this direction. For a silhouette, the minima rule uses negative minima of curvature to define boundary points on the contour of the silhouette. These correspond to points in concave regions where the magnitude of curvature is locally maximal. For 3D shapes, the minima rule uses negative minima of the principal curvatures along lines of curvature to define boundary curves on the surface[39,43].

The minima rule predicts a switch in perceived parts when figure and ground reverse. Consider the face-goblet illusion in Fig. 5A. When we see the faces as the figure, the negative minima of curvature, shown in Fig. 5B, carve the face into a forehead, nose, lips and chin. When we see the goblet as the figure, concave and convex reverse, and the new negative minima, shown in Fig. 5C, now carve the goblet into a lip, bowl, stem and base. Remarkably, these perceived parts correspond precisely to the perceptual units which natural language names with single words. This switch in perceived parts also explains an effect first noted by Mach[47], namely that we are more sensitive to symmetry in a visual pattern than to repetition[41]. Furthermore, it appears that human vision chooses the figure ground, such that figure has the more 'salient' parts[43].

The minima rule comports nicely with Gestalt theories of shape perception. As Wertheimer put it, "The given is itself in varying degrees 'structured' ('gestaltet'), it consists of more or less definitely structured wholes and whole-processes with their whole-properties and laws, characteristic whole-tendencies and whole-determinations of parts"[48]. The whole-determination of parts is an essential feature of the minima rule; the choice of figure and ground of the whole determines what are the minima, and therefore what are the boundary points of parts. This is clearly seen in the face–goblet example, where the parts one perceives depend on the global perception of figure and ground.

Although the minima rule gives precise boundary points on the contour of a silhouette, it says nothing about how to join these points to form cuts. In general, as illustrated in Fig. 6, there are many ways of joining boundary points to form cuts, and each gives a different set of parts[42,49]. There is now evidence that human vision has sophisticated rules for making such cuts. For instance, it prefers shorter cuts to longer, and cuts that have locally symmetric endpoints to those that do not[44,50]. Future work will be needed to decide how human vision describes parts, describes their spatial relationships, and uses these

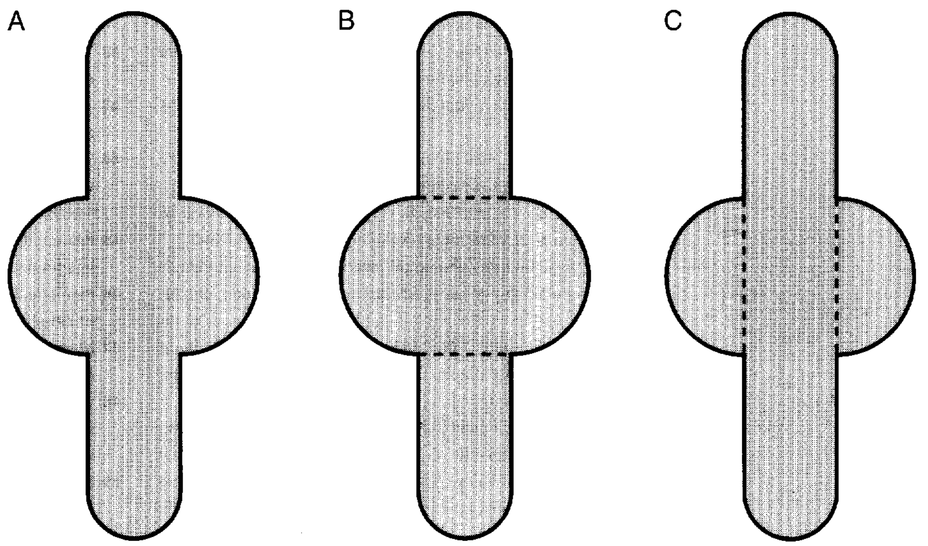

Fig. 6 The boundary points given by the minima rule can be joined in many ways to give part-cuts.

Outstanding questions

- How shall we model interaction between various 'modules' of the visual system; for example, the modules that construct shape (both 2D and 3D), color, motion, location in space and illumination?
- Under what conditions and for which tasks does human vision use viewpoint-dependent versus viewpoint-independent representations, and 2D versus 3D representations?
- How does human vision, for the purpose of recognition, represent the shapes of and spatial relationships between parts?
- How does human vision organize and index its memory of shapes?
- How does human vision judge the similarity of shapes?

descriptions to organize and index into a memory of shapes.

Concluding remarks

Human vision constructs visual objects and their shapes, colors, motions and surface properties. One goal of visual science is to uncover the rules of visual construction[19]. Understanding these rules is key to understanding the biological and computational workings of human vision, and for devising computer-vision systems of practical value for industry, for the visually impaired, and even, some day, for use in the home robot.

References

1 Benson, D.F. and Greenberg, J.P. (1969) Visual form agnosia: a specific defect in visual discrimination *Arch. Neurol.* 20, 82–89

2 Balint, R. (1909) Seelenlaehmung des Schauens optische ataxie, raeumliche Stoerung der Aufmerksamkeit *Monat. Psychiat. Neurol.* 25, 51–81

3 Tyler, H.R. (1968) Abnormalities of perception with defective eye movements (Balint's syndrome) *Cortex* 3, 154–171

4 Farah, M.J. (1990) *Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision*, MIT Press

5 Driver, J., Baylis, G.C. and Rafal, R.D. (1992) Preserved figure-ground segregation and symmetry perception in visual neglect *Nature* 360, 73–75

6 Barlow, H.B. (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1 371–394

7 Barlow, H.B. (1990) What does the brain see? How does it understand? in *Images and Understanding* (Barlow, H.B., Blakemore, C. and Weston-Smith, M., eds), pp. 5–25, Cambridge University Press

8 Zeki, S. (1993) *A Vision of the Brain*, Blackwell Scientific

9 Cowey, A. (1994) Cortical visual areas and the neurobiology of higher visual processes, in *The Neuropsychology of High-level Vision: Collected Tutorial Essays* (Farah, M.J. and Ratcliff, G., eds), pp. 3–31, Erlbaum

10 Verrey, L. (1888) Hemiachromatopsie droite absolue *Archs. Opthalmol.* (Paris) 8, 289–301

11 Zeki, S. (1990) A century of cerebral achromatopsia *Brain* 113, 1721–1777

12 Sacks, O. (1995) *An Anthropologist on Mars*, Vintage Books

13 Zihl, J., von Cramon, D. and Mai, N. (1983) Selective disturbance of movement vision after bilateral brain damage *Brain* 106, 313–340

14 Farah, M.J. (1994) Neuropsychological inference with an interactive brain: A critique of the locality assumption *Behav. Brain Sci.* 17, 43–61

15 Stoner, G.R. and Albright, T.D. (1993) Image segmentation cues in motion processing: implications for modularity in vision *J. Cogn. Neurosci.* 5, 129–149

16 Wallach, H. (1935) Uber visuell wahrgenommene Bewegungsrichtung *Psycholog. Forsch.* 20, 325–380

17 van Tuijl, H.F.J.M. (1975) A new visual illusion: neonlike color spreading and complementary color induction between subjective contours *Acta Psychol.* 39, 441–445

18 Redies, C. and Spillmann, L. (1981) The neon color effect in the Ehrenstein illusion *Perception* 10, 667–681

19 Hoffman, D.D. *Visual Reality: How We Create What We See*, Norton (in press)

20 Nakayama, K., Shimojo, S. and Ramachandran, V.S. (1990) Transparency: relation to depth, subjective contours, luminance, and neon color spreading *Perception* 19, 497–513

21 Kojo, I.V., Liinasuo, M.E. and Rovamo, J.M. (1995) Neon colour spreading in three-dimensional illusory objects *Invest. Ophthalmol. Vis. Sci., ARVO Abstr.* 36, 665

22 Cicerone, C.M. and Hoffman, D.D. (1991) Dynamic neon colors: Perceptual evidence for parallel visual pathways *Univ. Calif. Irvine, Math. Behav. Sci. Memo* 91–22

23 Cicerone, C.M. *et al.* (1995) The perception of color from motion *Percept. Psychophys.* 57, 761–777

24 Shipley, T.F. and Kellman, P.J. (1994) Spatiotemporal boundary formation: boundary, form, and motion perception from transformations of surface elements *J. Exp. Psychol. Gen.* 123, 3–20

25 Cortese, J.M. and Andersen, G.J. (1991) Recovery of 3-D shape from deforming contours *Percept. Psychophys.* 49, 315–327

26 Wertheimer, M. (1912) Experimentelle Studien über das Sehen von Bewegung. *Zeitschr. Psychologie* 61, 161–265

27 Spillmann, L. and Levine, J. (1971) Contrast enhancement in a Hermann grid with variable figure-ground ratio *Exp. Brain Res.* 13, 547–559

28 White, M. (1981) The effect of the nature of the surround on the perceived lightness of grey bars within square-wave test gratings *Perception* 10, 215–230

29 Marr, D. (1982) *Vision*, Freeman

30 Biederman, I. (1987) Recognition-by-components: a theory of human image understanding *Psychol. Rev.* 94, 115–147

31 Ullman, S. (1996) *High-Level Vision*, MIT Press

32 Biederman, I. and Ju, G. (1988) Surface vs. edge-based determinants of visual recognition *Cognitive Psychol.* 20, 38–64

33 Marr, D. and Nishihara, H.K. (1978) Representation and recognition of three-dimensional shapes *Proc. R. Soc. London Ser. B* 200, 269–294

34 Buelthoff, H.H. and Edelman, S. (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition *Proc. Natl. Acad. Sci. U. S. A.* 89, 60–64

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

36 Ullman, S. and Basri, R. (1991) Recognition by linear combinations of models *IEEE Trans. Patt. Anal. Mach. Intell.* 13, 992–1006

37 Tarr, M. and Pinker, S. (1989) Mental rotation and orientation-dependence in shape recognition *Cognitive Psychol.* 21, 233–282

38 Ullman, S. (1989) Aligning pictorial descriptions: an approach to object recognition *Cognition* 32, 193–254

39 Hoffman, D.D. and Richards, W.A. (1984) Parts of recognition *Cognition* 18, 65–96

40 Braunstein, M.L., Hoffman, D.D. and Saidpour, A. (1989) Parts of visual objects: An experimental test of the minima rule *Perception* 18, 817–826

41 Baylis, G.C. and Driver, J. (1995) Obligatory edge assignment in vision: the role of figure and part segmentation in symmetry detection *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1323–1342

42 Siddiqi, K., Tresness, K.J. and Kimia, B.B. (1996) Parts of visual form: psychophysical aspects *Perception* 25, 399–424

43 Hoffman, D.D. and Singh, M. Salience of visual parts *Cognition* (in press)

44 Seyranian, G., Singh, M., and Hoffman, D.D. (1997) Cuts for parsing visual shapes: 2. Experiments *Univ. Calif. Irvine, Math. Behav. Sci. Memo* 97–103

45 Brooks, R.A. (1981) Symbolic reasoning among 3-D models and 2-D images *Artif. Intell.* 17, 205–244

46 Pentland, A.P. (1986) Perceptual organization and the representation of natural form *Artif. Intell.* 28, 293–331

47 Mach, E. (1885) *The Analysis of Sensations, and the Relation of the Physical to the Psychical* (translated by C.M. Williams, 1959), Dover

48 Wertheimer, M. (1922) Untersuchungen zur Lehre von der Gestalt, I *Psychol. Forsch.* 1, 47–58 [in *Source Book of Gestalt Psychology*, Ellis, W.D., ed. (1938) pp. 12–16, Routledge & Kegan Paul]

49 Beusmans, J., Hoffman, D.D. and Bennett, B.M. (1987) Description of solid shape and its inference from occluding contours *J. Opt. Soc. Am.* 4, 1155–1167

50 Singh, M., Seyranian, G., and Hoffman, D.D. (1996) Cuts for parsing visual shapes: 1. Theory *Univ. Calif. Irvine, Math. Behav. Sci. Memo* 96–33