Level 2

Vision: Form Perception

Donald D Hoffman, University of California, Irvine, California, USA
Manish Singh, Rutgers University, New Brunswick, New Jersey, USA

CONTENTS

Human vision constructs the perceived two-dimensional and three-dimensional shapes of visual objects and scenes from retinal images that are inherently ambiguous. In the process it consults a variety of sources of information, including motion, shading, texture, occlusion, binocular disparity and image contours.

INTRODUCTION

0629.001 The human eye focuses an image onto a light-sensitive sheet of neural tissue called the retina (Dowling, 1987). This image is captured by a discrete array of cells in the retina, called photoreceptors. Each photoreceptor generates a signal which varies in time with the discrete number of photons of light that the photoreceptor catches. This discrete array of time-varying signals is the starting point of vision. The only information this array makes explicit is the varying number of photon catches at each individual photoreceptor. It does not make explicit lines, curves, two-dimensional regions, three-dimensional shapes or any other aspect of the visual forms of objects and their environments. The perception of visual forms is the consequence of sophisticated processes of construction which engage literally billions of neurons and trillions of synaptic connections between neurons. Every line, curve, 2D region or 3D shape that we see is a construction of our visual system, created on the fly starting with just the photon catches at the retina. Vision researchers have made substantial progress in describing the constructive processes underlying the perception of visual form.

CONTOUR DETECTION

0629.002 The construction of contours occurs early in the flow of visual processing. In many mammals the primary visual cortex, which is the first stage of cortical visual processing, begins the construction of contours with neurons called ‘simple cells’, which construct oriented line segments (Hubel and Wiesel, 1959, 1962; Hubel, 1995). These cells construct lines such as the short black lines shown in Figure 1. Computational theories of vision also place the construction of contours early in the flow of visual processing. In the influential theory of David Marr, the flow of visual processing leads to a sequence of visual representations: the primal sketch, the 2½D sketch, and the 3D model (Marr, 1982). The primal sketch makes explicit structure and grouping in the 2D image, the 2½D sketch represents the surface geometry of objects relative to the viewer, and the 3D model represents the volume of objects in coordinates centered in the objects.

The primal sketch contains contours as a key element of the representation. In Marr’s theory these are constructed by first linearly filtering the retinal image with a spatial filter whose shape is defined mathematically as the Laplacian of a 2D Gaussian (Marr and Hildreth, 1980). This filter is circularly symmetric, and its shape resembles a Mexican hat. Edges correspond to those places where the values of the filtered output pass through zero. Marr and Hildreth proposed that simple cells in primary visual cortex are in fact detectors of these zero crossings. Their computational theory of edge detection has been superseded by later approaches that filter the image with spatial filters that are not circularly symmetric, but are designed to optimize the signal-to-noise ratio in the edge construction process (Canny, 1986; Deriche, 1987). 0629.003

Linearly filtering an image is a key first step in the construction of contours. The steps that come after filtering are complex and highly nonlinear. This is also illustrated in Figure 1: in square 1, one sees not only the short black lines, but also a square with clear edges. The square appears to be slightly 0629.004

brighter than the background. In fact, measurement of actual light intensities would indicate that the white region inside the square is identical to the white region outside the square, and that no edge has been drawn around the square. The brightness of the square, and the clear edges that mark its border, are entirely constructs of your visual system. The square region is called a subjective (or illusory) surface and its edges are called subjective (illusory) contours (Petry and Meyer, 1987). Square 2 also appears to have a subjective square region surrounded by clear subjective contours. If we superimpose square 1 on square 2 so that their subjective squares are perfectly aligned, we obtain the image labeled ‘1 + 2’. Here the subjective square and subjective contours are weak or missing altogether; this is the opposite of what one would predict if the construction of contours were an entirely linear process (Albert and Hoffman, 2000).

0629.005 Neurons that signal subjective contours have been found in the primary visual cortex, area V1, of the macaque monkey (Grosof *et al*., 1993) and also in area V2, the next stage of cortical visual processing (von der Heydt *et al*., 1984). Several artificial neural network models have been proposed to explain the properties of these cells (von der Heydt and Peterhans, 1989; Francis and Grossberg, 1996).

THREE-DIMENSIONAL SHAPE FROM IMAGE CONTOURS

0629.006 The visual system not only constructs 2D contours, it also constructs the 3D shapes of objects and represents these 2D contours and 3D shapes in terms of parts and their spatial relationships. In the process of constructing 3D shape, human vision consults a variety of sources of information, including shading, texture, motion, occlusion, binocular disparity, and 2D contours.

0629.007 Figure 2a illustrates the construction of 3D shapes from 2D contours. The figure uses just a few contours to depict a doughnut shape. The contours are called ‘occluding contours’ because they depict points where the visible portions of the 3D shape just begin to occlude the hidden portions. Yet these few occluding contours are sufficient to trigger the visual system to construct the smooth 3D shape of the doughnut. Human vision uses several geometric facts and assumptions in the process. The qualitative shape of a 3D object can be described at each point as being convex, concave, cylindrical, or saddle-shaped. Convex regions are shaped like the outside surface of an egg, concave regions like the inside surface of an egg, cylindrical regions like a cylinder, and saddle-shaped regions like the surface of a horse’s saddle. Convex regions of a doughnut are denoted by plus signs in Figure 2a, and saddle-shaped regions by minus signs. As can be seen from Figure 2a, the outer occluding contours are in convex regions, whereas the inner occluding contours are in saddle-shaped regions. Jan Koenderink showed that human vision can, in principle, infer the qualitative shapes of smooth 3D surfaces from their projected occluding contours (Koenderink, 1984). The visual system also assumes in this case that the surface varies smoothly between the occluding contours. Since the qualitative shape changes from convex to saddle between these contours, it must be cylindrical at the boundary between convex and saddle. The cylindrical points are indicated by the dashed circle on top of the doughnut. This gives a complete qualitative description of the 3D shape of the doughnut, which Koenderink has shown can be derived entirely from its occluding contours and some built-in knowledge of geometry and projection.

Human vision organizes 2D and 3D shapes into 0629.008
parts as an aid to recognition, manipulation, and naming. It often uses concave regions, and especially points of highest magnitude of curvature within these concave regions, to divide shapes into parts: this is called the ‘minima rule’ (Hoffman and Richards, 1984; Hoffman and Singh, 1997; Singh *et al*., 1999). This is illustrated in Figure 2b, which shows the well-known face–goblet ambiguous figure. If the faces are taken as the object, then the extrema of curvature within the concave regions of the faces are used by human vision as the boundaries between parts. These extrema are indicated by the short dashes on the right-hand side of the figure, and divide the faces into parts corresponding to the forehead, nose, lips and chin. If the goblet is taken as the object, then the regions of the curves that are concave and convex are reversed from when the faces are taken as the object. Since only concave regions are used to create part boundaries, human vision creates a new set of part boundaries for the goblet that are different from the part boundaries of the faces. These are shown on the left-hand side of Figure 2b, and divide the goblet into a lip, bowl, stem and base. Neurophysiological studies of visual area V4 have found many cells that signal the extrema of curvature of 2D contours (Pasupathy and Connor, 1999).

SHADING

0629.009 Consider Figure 3a. It contains a pattern of gray-level intensity values distributed on a flat sheet of paper. However, it is quite difficult to see this figure as simply a two-dimensional pattern. One cannot help but see a light-colored wrinkled surface extended in three dimensions, lit from the left. The ability of the visual system to construct a 3D surface shape from 2D shading information is a striking computational feat. When viewing a surface, the light intensity at any given location on the retina is a combined function of (at least) three different variables: the reflectance properties of the surface, the local orientation of the surface, and the position and intensity of the light sources. In order to construct the percept of a 3D surface, the visual system must somehow decompose the pattern of luminance values into the separate contributions of surface reflectance, surface shape, and illumination.

0629.010 In principle a unique solution to the shape-from-shading problem can be derived under restrictive conditions: the reflectance function of the surface is known, the surface is smooth, and the illuminant is a single point source (Horn, 1977). These assumptions allow variations in image luminance to be attributed entirely to gradual changes in surface orientation. Humans, however, are able to perceive shape from shading in many situations where the reflectance is unknown, the surface contains tangent discontinuities, and is lit by more than one light source. This suggests that human vision brings to bear other constraints in computing 3D shape from shading information – although it is fair to say that the precise constraints used by human vision remain largely unknown.

0629.011 Koenderink and colleagues have probed psychophysically the human perception of surface relief from gray-level images (these typically contain both shading and occluding contours). One technique uses a gauge figure – the projection of a circular disk oriented in 3D, with a short line segment sticking perpendicularly out of its center (Figure 3b). In a typical experiment, a gauge figure is superimposed at a large number of locations on a shaded image, and the observer adjusts its perceived 3D orientation using a trackball, in order to match the perceived local orientation of the surface (Koenderink *et al.*, 1992). These local measurements can then be integrated to construct the global surface structure perceived by the observer (this presupposes internal consistency in the observer's responses, which is typically the case). Although this perceived surface structure is qualitatively similar to the depicted surface, it is not quantitatively the same. In particular, surface reliefs seen by different observers differ by depth scalings (i.e. flattenings and elongations in depth), and similarly, surface reliefs seen by the same observer under different illumination conditions differ by depth scalings. Moreover, different parts of the depicted surface are often seen depth-scaled by different amounts (Todd *et al.*, 1996), providing further evidence that human vision represents surface shapes in terms of parts.

Qualitatively similar surface reliefs are obtained 0629.012
using a different technique in which pairs of points are presented at different locations on a shaded image, and observers are asked to indicate which of the two points appears closer in depth (Koenderink *et al.*, 1996). This method, however, yields results that are less precise than those obtained by the gauge figure method by an order of magnitude. This, in itself, is a significant finding: it suggests that a depth map – a pointwise specification of relative depth at each image location (e.g. Marr, 1982) – is not the primary way human vision represents surface structure, from which it then derives local surface orientation. Rather, surface orientation itself appears to be a perceptually fundamental variable. Moreover, the fact that quantitatively different surface reliefs are obtained with different experimental methods suggests that the visual system may not have a single representation of surface shape, but rather may invoke distinct representations depending on the task at hand (Koenderink, 1998).

CONCLUSION

Human vision starts with just the shower of 0629.013
photons that hit the retina of each eye, and proceeds to construct two-dimensional contours and three-dimensional forms by consulting various sources of information, including shading, texture, motion, occlusion, binocular disparity and image contours. In the process it uses many rules of construction, some of which are based on laws of geometry, reflectance, lighting and projection.

References

Albert MK and Hoffman DD (2000) The generic-viewpoint assumption and illusory contours. *Perception* **29**: 303–312.

Canny JF (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**: 769–798.

Deriche R (1987) Using Canny's criteria to derive an optimum edge detector recursively implemented. *International Journal of Computer Vision* **2**: 167–187.
Dowling JE (1987) *The Retina: An Approachable Part of the Brain*. Cambridge, MA: Harvard University Press.
Francis G and Grossberg S (1996) Cortical dynamics of form and motion integration: persistence, apparent motion, and illusory contours. *Vision Research* **36**: 149–173.
Hoffman DD and Richards WA (1984) Parts of recognition. *Cognition* **18**: 65–96.
Hoffman DD and Singh M (1997) Salience of visual parts. *Cognition* **63**: 29–78.
Horn BKP (1977) Image intensity understanding. *Artificial Intelligence* **8**: 201–231.
Hubel DH (1995) *Eye, Brain, and Vision*. New York: Freeman Press.
Hubel DH and Wiesel TN (1959) Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology* **148**: 574–591.
Hubel DH and Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* **160**: 106–154.
Koenderink JJ (1984) What does the occluding contour tell us about solid shape? *Perception* **13**: 321–330.
Koenderink JJ (1998) Pictorial relief. *Philosophical Transactions of the Royal Society of London Series A* **356**: 1071–1086.
Koenderink JJ, van Doorn AJ and Kappers AML (1992) Surface perception in pictures. *Perception and Psychophysics* **52**: 487–496.
Koenderink JJ, van Doorn AJ and Kappers AML (1996) Pictorial surface attitude and local depth comparisons. *Perception and Psychophysics* **58**: 163–173.
Marr D (1982) *Vision: A Computational Investigation into The Human Representation and Processing of Visual Information*. San Francisco: WH Freeman.
Marr D and Hildreth EC (1980) Theory of edge detection. *Proceedings of the Royal Society of London Series B* **207**: 187–217.
Pasupathy A and Connor CE (1999) Responses to contour features in macaque area V4. *Journal of Neurophysiology* **82**: 2490–2502.
Petry S and Meyer GE (1987) *The Perception of Illusory Contours*. New York: Springer.
Singh M, Seyranian GD and Hoffman DD (1999) Parsing silhouettes: the short-cut rule. *Perception and Psychophysics* **61**: 636–660.
Todd JT, Koenderink JJ, van Doorn AJ and Kappers AML (1996) Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception and Performance* **22**: 695–706.
Von der Heydt R and Peterhans E (1989) Cortical contour mechanisms and geometrical illusions. In: Lam DM and Gilbert CD (eds) *Neural Mechanisms of Visual Perceptions*, pp. 157–170. Woodlands, TX: Portfolio Publishing.
Von der Heydt R, Peterhans E and Baumgartner G (1984) Illusory contours and cortical neuron responses. *Science* **224**: 1260–1262.

Further Reading

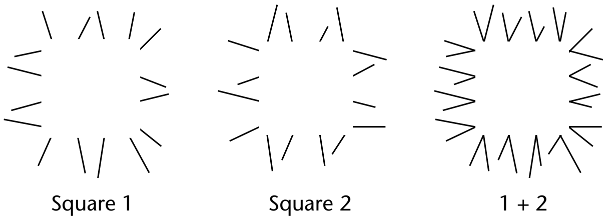
Farah MJ (2000) *The Cognitive Neuroscience of Vision*. Malden, MA: Blackwell Publishers.
Gregory RL (1997) *Eye and Brain*, 5th edn. Princeton: Princeton University Press.
Hoffman DD (1998) *Visual Intelligence: How We Create What We See*. New York: WW Norton.
Horn BKP and Brooks MJ (1989) *Shape from Shading*. Cambridge, MA: MIT Press.
Kellman PJ and Arterberry ME (1998) *The Cradle of Knowledge: Development of Perception in Infancy*. Cambridge, MA: MIT Press.
Koenderink JJ (1990) *Solid Shape*. Cambridge, MA: MIT Press.
Marr D (1982) *Vision: A Computational Investigation into The Human Representation and Processing of Visual Information*. San Francisco: WH Freeman.
Palmer SE (1999) *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.
Regan D (2000) *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*. Sunderland, MA: Sinauer.
Wandell BA (1995) *Foundations of Vision*. Sunderland, MA: Sinauer.
Zeki S (2000) *Inner Vision: An Exploration of Art and the Brain*. Oxford: Oxford University Press.

Glossary

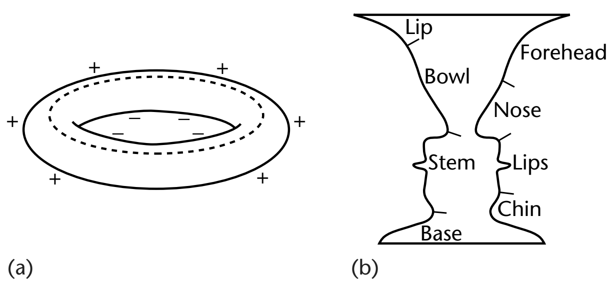
Extrema of curvature Points on a curve or surface where the magnitude of curvature is locally the greatest.
Occluding contour Points where the visible portions of a three-dimensional form just begin to occlude the invisible portions.
Photon The smallest unit of light.
Subjective contour A contour seen by human observers but not by measuring devices such as photometers.
Visual cortex Regions of the cortex of the brain that are involved in processing visual information.

Keywords: (Check)

extrema of curvature; occluding contour; shading; subjective contour; surface orientation



0629f001 **Figure 1.** Constructing visual contours is not a simple linear process. In square 1 and square 2 we see a square with clear edges. When these two figures are linearly superimposed, as in right-hand image, we no longer see a square with clear edges: this is the opposite of what one would expect from a linear process



0629f002 **Figure 2.** (a) Human vision can construct 3D shapes from simple 2D drawings of contours, as illustrated by the doughnut in this figure. (b) Human vision has rules for carving 2D curves and 3D shapes into parts. In the face–goblet ambiguous figure these rules divide the faces into parts corresponding to the forehead, nose, lips and chin; they divide the goblet into a lip, bowl, stem and base

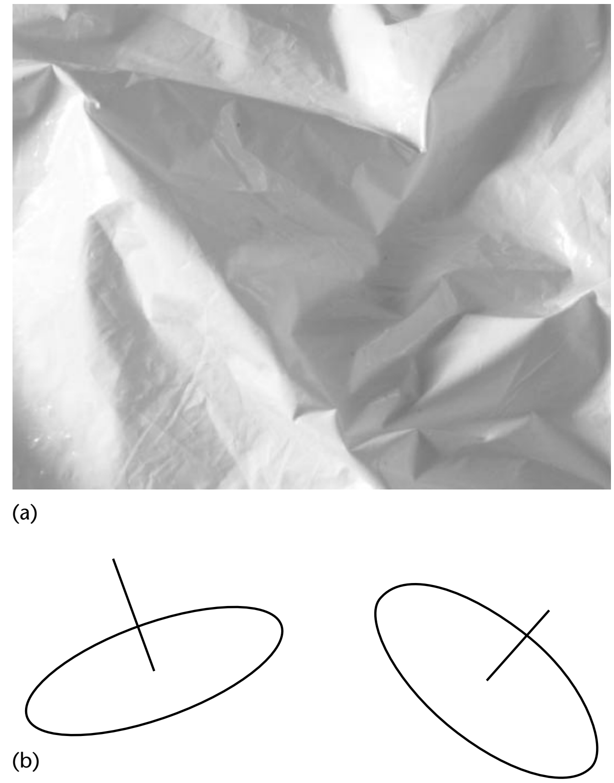


Figure 3. (a) Human vision can construct three-dimensional surface shape from two-dimensional images depicting shading information. To do so, it must separate the contributions of surface shape from those of surface reflectance and illumination (Photograph by Marc Talusan). (b) The gauge figure is used to probe the surface structure perceived by observers (Koenderink *et al.*, 1992). Observers adjust the 3D orientation of the gauge figure until it 'fits' a depicted surface locally 0629f003

***ECS* author queries**

Article 629 [Hoffman & Singh]

Introduction

second to last sentence '...created on the fly...' - a bit idiomatic: please suggest another wording (or just omit?)

`...created rapidly...'

Contour detection

paragraph 1, sentence 3 'These cells construct lines...' - is 'construct' OK, or would 'respond to' (or similar) be better?

`construct' is OK, but so is `respond to'

paragraph 2, sentence 2 'Laplacian', 'Gaussian' - what? filter? function? Or is it OK like this?

OK as it is

paragraph 4, sentence 1 'Grosof et al 1993' - not in reference list.

GROSOF DH; SHAPLEY RM; HAWKEN MJ.
MACAQUE-V1 NEURONS CAN SIGNAL ILLUSORY CONTOURS.
NATURE, 1993 OCT 7, V365 N6446:550-552.

Three-dimensional shape

paragraph 2, sentence 7 'Convex regions are shaped... saddle' - suggest omit this sentence: not sure it really increases understanding.

OK to omit sentence

Figures

Please confirm any necessary permissions have been obtained for reproduction in both print and electronic formats (particularly Figure 3).

Permissions have been obtained