The importance of contours for visual object recognition and discrimination

A importância do contorno para o reconhecimento e discriminação visual de objetos

La importancia del contorno para el reconocimiento y la discriminación visual de objetos

J. Farley Norman[1]

[1] Western Kentucky University. *E-mail*: farley.norman@wku.edu *ORCID*: https://orcid.org/0000-0001-5877-1914

Informações do artigo:

J. Farley Norman
farley.norman@wku.edu

Recebido em: 21/09/2020
Aceito em: 22/09/2020

ABSTRACT

In contrast to many machine vision systems, we human observers can readily recognize solid objects and visually discriminate their 3-D shapes even under changes in viewpoint and variations in object orientation and lighting. While the importance of binocular disparity has been known since the 1830's, the importance and perceptual informativeness of visual contours for object recognition and discrimination is not adequately appreciated. This article will review those scientific contributions that demonstrate that visual contours and their deformations over time (in response to object or observer motion) carry as much or more information about object shape than other forms of visual information.

KEYWORDS:

Visual system; Object Recognition; 3-D.

RESUMO

Em contraste com muitos sistemas artificiais de visão, nós, observadores humanos, podemos reconhecer prontamente objetos sólidos e discriminar visualmente suas formas 3-D, mesmo sob mudanças no ponto de vista e variações na orientação e iluminação do objeto. Embora a importância da disparidade binocular seja conhecida desde a década de 1830, a relevância e a riqueza perceptiva das informações fornecidas pelos contornos visuais para o reconhecimento e discriminação de objetos não são devidamente apreciadas. Este artigo revisará as contribuições científicas que demonstram que os contornos visuais e suas deformações ao longo do tempo (em resposta ao movimento do objeto ou do observador) proporcionam tanto ou mais informações sobre a forma do objeto do que outras formas de informação visual.

PALAVRAS-CHAVE:

Sistema visual; Reconhecimento de Objetos; 3-D.

RESUMEN

A diferencia de los sistemas de visión artificial, los observadores humanos pueden reconocer fácilmente objetos sólidos y discriminar visualmente su forma 3D, incluso cuando se producen cambios en su punto de vista, en la orientación o en la iluminación. Aunque la importancia de la disparidad binocular es conocida desde 1830, la relevancia y la riqueza de la información provista por los contornos visuales para el reconocimiento y discriminación de los objetos no ha sido apreciada adecuadamente. En este artículo se revisan las contribuciones científicas que demuestran que los contornos visuales y sus deformaciones a lo largo del tiempo (en respuesta a los movimientos tanto del observador como del objeto) proporcionan tanto o más información sobre la forma del objeto que otros tipos de información visual.

PALABRAS CLAVE:

Sistema visual; Reconocimiento de Objetos; 3-D.

The human ability to visually perceive the solid shape of environmental objects has been scientifically studied for almost two hundred years (e.g., Wheatstone, 1838). A surprising and significant advance in our understanding of what visual information is actually needed to recognize objects occurred in 1954; this year saw the publication of a unique article by Fred Attneave (also see De Winter & Wagemans, 2008 and Norman, Phillips, & Ross, 2001). In his article, Attneave used a figure now known as "Attneave's cat" (see his Figure 3) to demonstrate that it was possible to eliminate nearly all of the information from a visual image and the depicted object would still be recognizable to human observers. To illustrate Attneave's process, consider Figure 1; this figure contains a photograph of one of my cats (whose name is Sylvester). The original

photograph of Sylvester has a resolution of 1,606 x 1,278 pixels. This photograph, therefore, contains more than two million pieces of information (i.e., more than two million color/intensity values). From his original image, Attneave created a drawing of his cat that consisted of boundary contours (contours located where the cat occluded its background and where parts of the cat occluded other parts of its body). From these continuous contours, Attneave located the points along the contours that curved the most (i.e., curved the most in either a concave or convex fashion). He then connected these 38 points of maximal curvature with a ruler or straightedge to create the famous illustration of Attneave's cat (Figure 3 of Attneave, 1954).

Figure 1.
_____

A photograph of a resting cat (Sylvester).
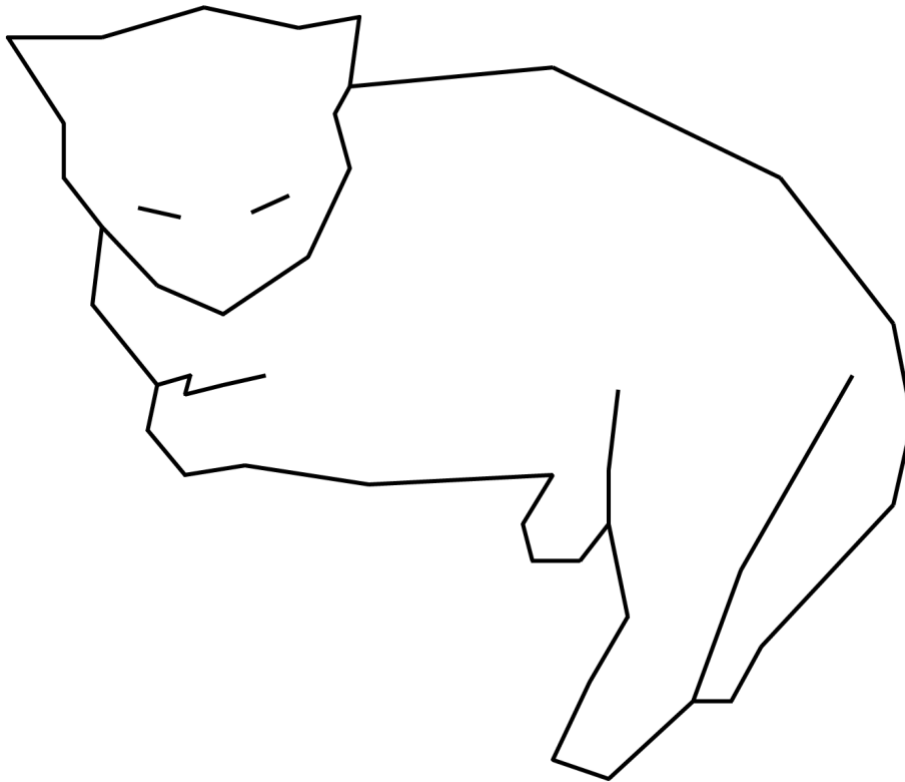
Figure 2.

A drawing of a cat (Sylvester) that is analogous to Figure 3 of Attneave (1954). Forty-five points of maximal curvature (both convex and concave) were identified along the boundary contours of the cat shown in Figure 1. These 45 points were then connected with straight lines. Even though all of the original image intensities (see Figure 1) have been removed and only the 45 connected points of maximal curvature remain, this drawing is still easily recognizeable as a cat.

Fred Attneave, therefore, took an image composed of millions of intensity values and reduced it to a drawing that consisted of only 38 pieces of information, and the object was still visually recognizeable as a cat. If you choose which pieces of visual information to retain, you can delete over 99 percent of the information in an image and the depicted object(s) will still be recognizeable. To see that this is true, consider the current Figure 2. To create this Figure, I performed the same process as Attneave. From the original photograph of Sylvester (Figure 1), I marked 45 points of maximal curvature along the boundary contours of my cat (such as the tips of the ears) and then connected these 45 points with straight lines. Even though nearly all of the original information has been removed (99.998 percent = (2,052,468-45)/2,052,468), the object

depicted in Figure 2 is still visually recognizable as a cat. Attneave's Figure 3 and my current example demonstrates that one does not necessarily need the entire content of retinal images (binocular disparity, retinal motion, shading, texture gradients, specular highlights, etc.) in order to successfully recognize visual objects. Sparse visual contours (i.e., line drawings) can be surprisingly informative (also see Kennedy, 1974).
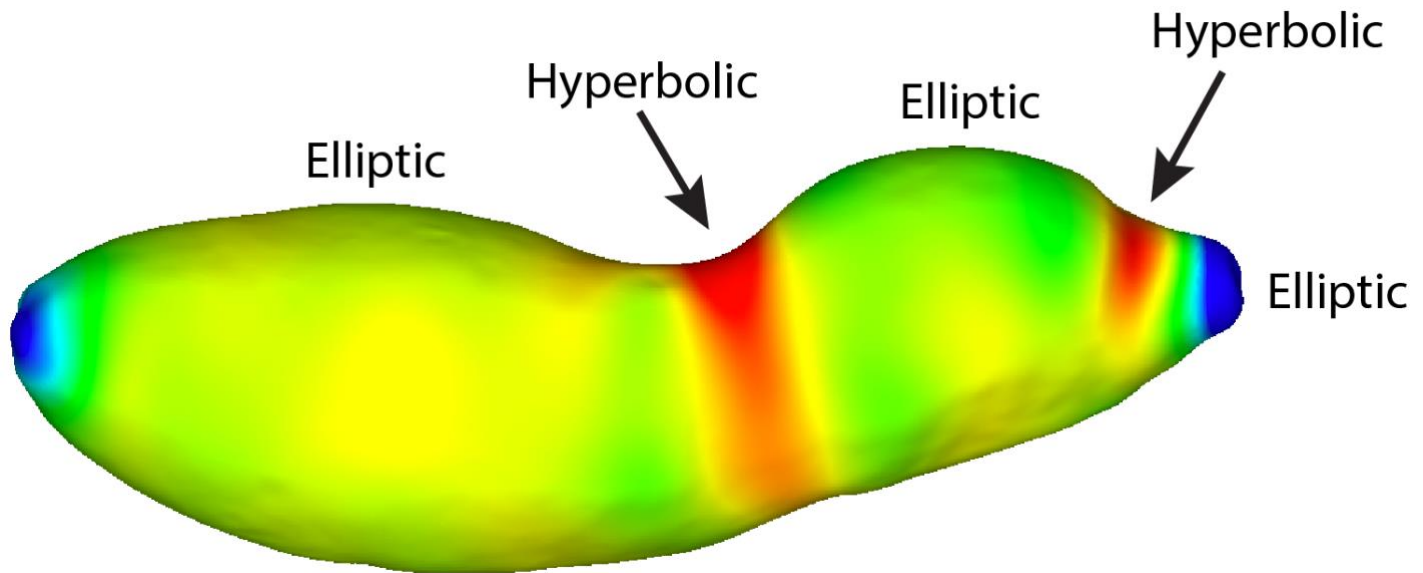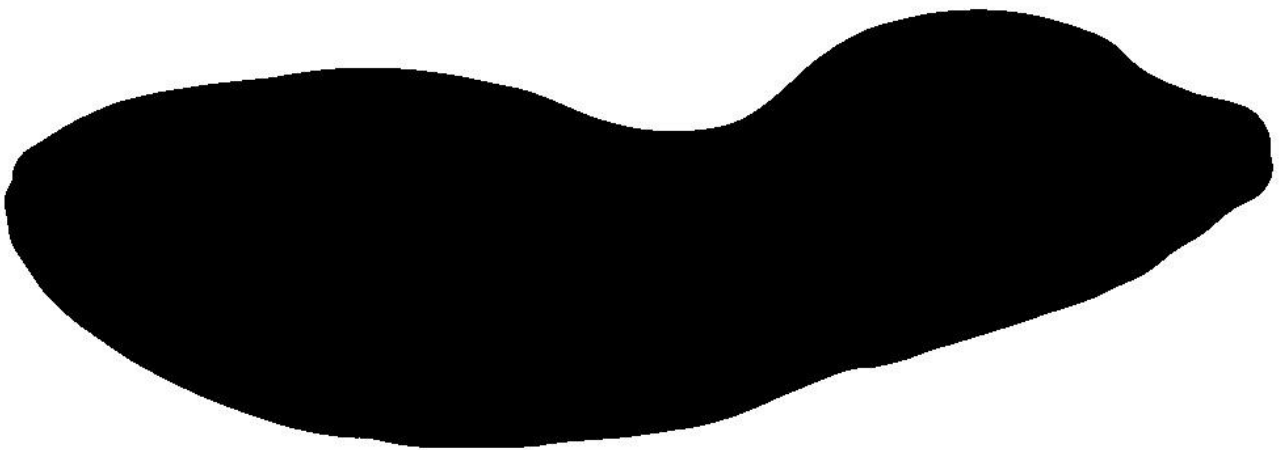

Figure 3.

A geometric analysis of a naturally-shaped object (a sweet potato, Ipomoea batatas). Elliptic surface regions are colored in green and blue, while hyperbolic surface regions are colored in red. One can also see the parabolic regions (yellow) that separate elliptic and hyperbolic surface regions.

Solid environmental objects (especially those that have a natural shape, not man-made) possess several distinct types of surface regions (e.g., Hilbert & Cohn-Vossen, 1983; Koenderink, 1984a; Koenderink, 1984b; Koenderink, 1990; Koenderink & van Doorn, 1992; Van Effelterre, 1994). These surface regions differ in their local shape. Elliptic surface regions are locally shaped like a bump or a dimple (like the inside or outside of a bowl), while hyperbolic regions are shaped like horse saddles (they are concave in one direction, but convex in a perpendicular direction). Finally, parabolic surface regions are shaped like cylinders (curved in one direction, but not curved in a perpendicular direction). As an example, consider the solid object shown in Figure 3; this is a sweet potato (Ipomoea batatas) that I purchased in a grocery store and then laser-scanned in

3-D (Norman et al., 2019). This sweet potato has three prominent elliptic regions (indicated by green & blue) and two hyperbolic ones (indicated by red). When one views a solid object like this under ordinary conditions, there are many types of visual information that help us to perceive its shape – binocular disparity, shading, etc. However, its outer boundary contour is surprisingly informative, as can be seen from a mathematical analysis published by Jan Koenderink in 1984. In this article, Koenderink (1984b) showed that surface areas shaped like a saddle in 3-D (hyperbolic regions) project (i.e., correspond to) concavities in the object's boundary contour, while surface areas shaped like a bump in 3-D (elliptic regions) project to convexities in the object's boundary contour. Thus, the 2-D shapes of different parts of a boundary contour provide direct visual information about the 3-D shapes of the corresponding regions on the solid object's actual surface. Now consider Figure 4.
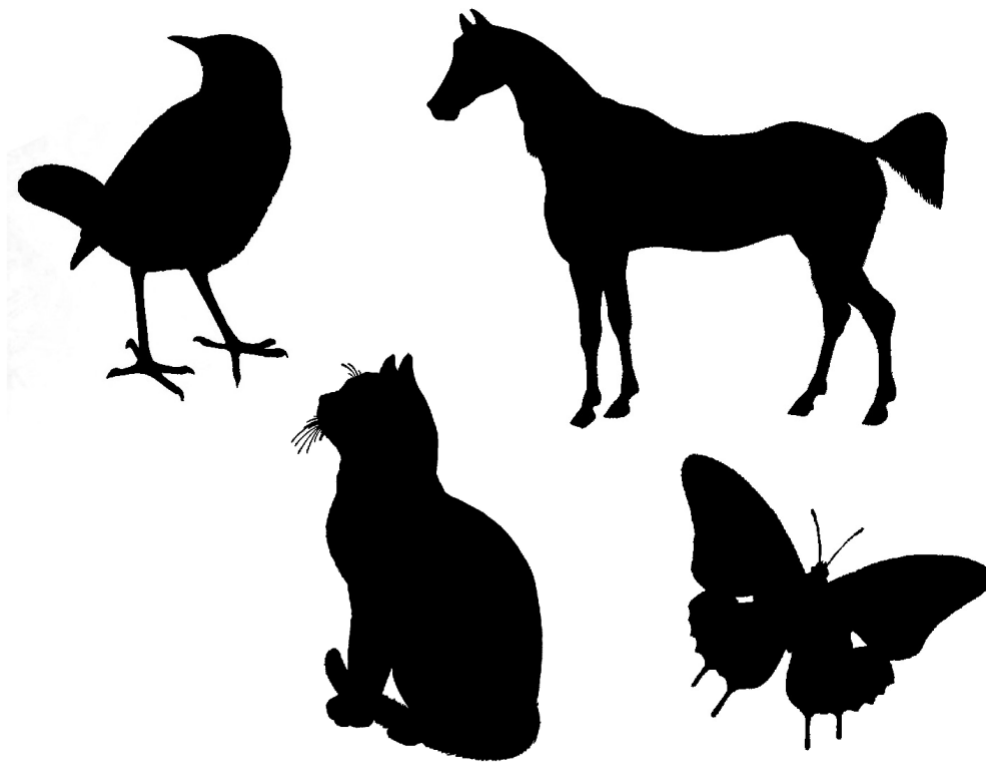

Figure 4.

_____

A silhouette showing only the outer boundary contours of the same sweet potato that is shown in Figure 3.

This figure shows a silhouette of the same sweet potato that was depicted in Figure 3. Note that in Figure 4, there is no inner detail at all; the only information one has about the object is provided by the shape of its outer boundary contour. Nevertheless, given the analysis by Koenderink (1984b), we can obtain a knowledge of much of its 3-D shape from the convexities and concavities of this object's boundary contour

(i.e., a convexity of the boundary contour proves that the corresponding 3-D surface region on the actual solid object is shaped like a bump, while a concavity of the boundary contour proves that the corresponding 3-D surface region on the actual solid object is shaped locally like a horse saddle).

If a 3-D object rotates in depth relative to an observer (or we move relative to the object), then that movement will create even more information about solid object shape. If an object rotates in depth, its visible boundary contour will change (i.e., deform, see Experiment 1 of Wallach & O'Connell, 1953) and new areas on the 3-D surface of the object will project to its boundary contour. If a solid object makes a complete rotation in depth, eventually all of its constituent surface regions will project to the outer boundary; one could then theoretically derive a representation of the entire object's solid shape (Norman, Lee, Phillips, Norman, Jennings, & McBride, 2009; also see Pollick, Giblin, Rycroft, & Wilson, 1992).

So far in our review, we have seen that boundary contours contain a lot of potential information to support the human perception (e.g., recognition and discrimination) of solid object shape. What scientific evidence exists to show that this potential information is actually used by human observers? In an interesting study by Wagemans et al. (2008), these researchers made silhouette versions of a wide variety of object stimuli that were originally developed by Snodgrass and Vanderwart (1980). The silhouette versions of the 260 objects were presented for 5 seconds each, and the participants' task was to identify (i.e., provide a verbal label) for each object. Of the 260 silhouettes presented, about half (128) had excellent identifiability, while a further 25 percent (64) had moderate identifiability; only a minority of 68 silhouettes out of the 260 produced poor identifiability. It is clear that the boundary contours of silhouettes can be very informative for the human perception of object shape; see Figure 5 for some readily identifiable examples.
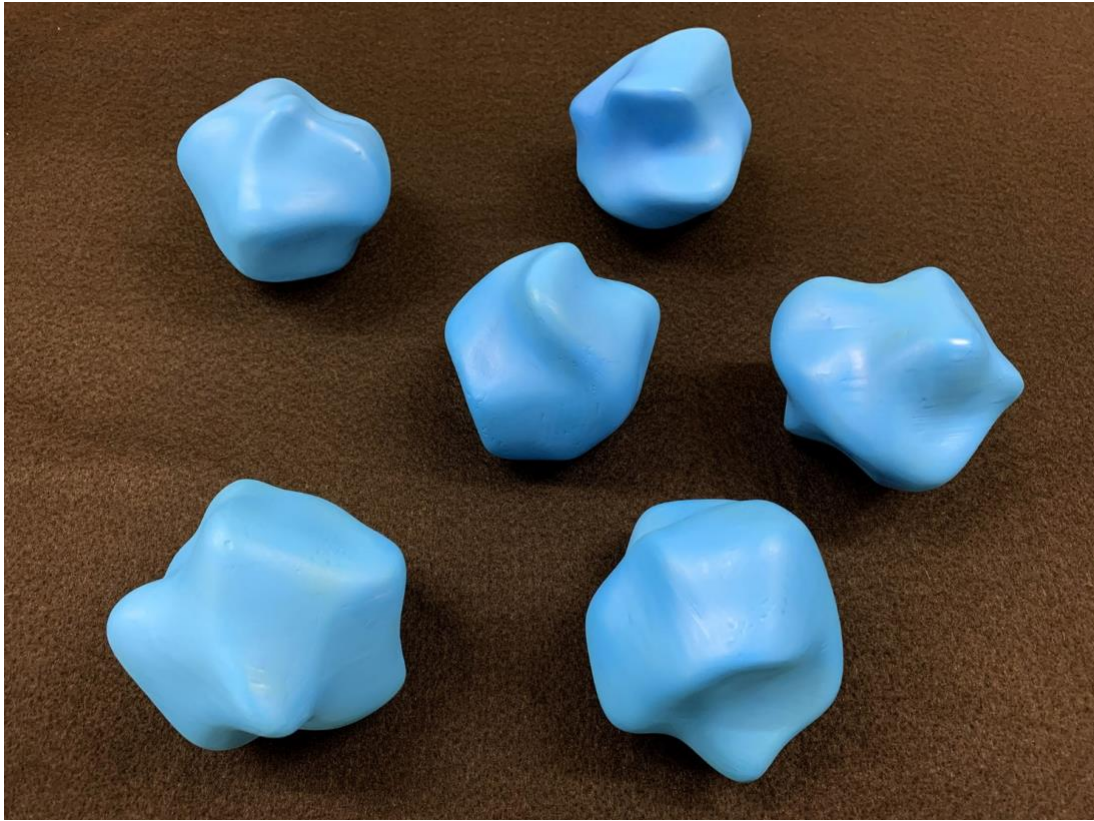
Figure 5.

Silhouettes of a variety of animals. From upper left proceeding clockwise are 1) a bird, 2) a horse, 3) a butterfly, and 4) a cat.

In 2000, Norman, Dawson, and Raines asked participants to identify naturally-shaped objects (bell peppers, Capsicum annuum). Participants were shown silhouettes and cast shadows of five individual bell peppers (each had a somewhat different 3-D shape); on each trial, one silhouette or cast shadow would be presented and the participants' task was to identify the object (as being object 1, object 2, object 3, etc.). In one condition, stationary silhouettes/cast shadows would be presented, while in another condition, deforming silhouettes/cast shadows were shown (as described earlier, when solid objects rotate in depth, their silhouettes/cast shadows will deform, or change their shape, over time). The participants' object identification performance was nearly perfect for deforming (i.e., moving) silhouettes/cast shadows. The participants' performance for stationary silhouettes/cast shadows ranged all the way from perfect identification performance to chance, depending upon the particular orientation of the object relative to the viewer. For those object orientations that produced prominent convexities and concavities in the silhouette/shadow outer

boundary contour, the participants' identification performance was excellent (see Figures 6 and 7 of Norman et al., 2000). In a very recent similar study, Norman, Dukes, Shapiro, Sanders, and Elder (2020) demonstrated that human observers do not necessarily need to see entire object silhouettes or cast shadows; if moving objects are presented, one can occlude or block over 90 percent of an object's silhouette/cast shadow and the object will remain highly recognizable.

In 2016, Norman et al. presented participants with views of stationary or moving (i.e., rotating in depth) bell peppers (*Capsicum annuum*). The participants' task was to view a single randomly-chosen object on each trial and then identify which of 12 physical bell peppers possessed the same shape (i.e., this was a solid shape matching task). Not surprisingly, the participants' shape matching performance was higher when the stimulus displays contained motion and lower when the objects were stationary. The results showed that in the conditions with motion, the participants' performance for deforming silhouettes was just about as high (i.e., not significantly different) as when the solid object shapes possessed texture or were portrayed by specular highlights (Norman, Todd, & Orban [2004] demonstrated that the specular highlights that occur on shiny surfaces permit shape discrimination at levels that are as high as those that occur for objects defined by any other type of optical information). The results of this 2016 study also demonstrated that vision and the sense of active touch (haptics) were comparable: the ability to perceive solid object shape from deforming boundary contours (i.e., deforming silhouettes) was about as high as that obtained when participants were allowed to haptically explore the object shapes using their hands and fingers.

Figure 6.

_____

A photograph of the type of solid objects used as experimental stimuli by Norman, Bartholomew, and Burton

(2008) and Norman and Raines (2002).

In a study of solid shape discrimination, Norman, Bartholomew, and Burton (2008) showed

participants randomly-shaped solid objects like those shown in Figure 6. The 100 stimulus objects were either

presented with a granite texture and shading (relatively full-cue stimuli) or only as silhouettes. Two objects

were presented on each trial for three seconds each; the participants' task was to indicate whether the two

objects had the same shape or possessed different shapes. The task was challenging, because even on the

"same" trials (both objects presented had the same shapes), the two objects appeared different, because their

orientation relative to the participant was different. While the participants' shape discrimination performance

was somewhat better for the full-cue stimuli presented with texture and image shading, the performance for

the silhouette stimuli was still very good (the participants' overall d' values were 3.004 and 2.287 for the

textured/shaded and silhouette conditions, respectively; see Figure 10 of Norman et al., 2008). This study also

found a significant effect of motion; the younger adult participants' (mean age was 21.5 years) shape discrimination performance was about 24 percent higher when the objects rotated in depth (causing deformations of the objects' outer boundary contours).

The boundary contours of silhouettes and cast shadows not only support object recognition and shape discrimination, they also allow human observers to perceive important aspects of the interior shape of single objects. Norman and Raines (2002) used randomly-shaped solid objects like those shown in Figure 6. On any given trial, two surface regions would be highlighted on an object's surface; the participants' task was to indicate which of the two regions was closer to them in depth. From the participants' judgments, Norman and Raines calculated depth difference thresholds (i.e., how much of a difference in depth was needed for the participants to make a reliable judgment about depth order). For conditions with moving objects (i.e., where the boundary contours deformed over time in response to object rotation in depth), the participants' ordinal depth thresholds were 0.54, 0.94, and 1.24 cm for surface regions that were close to the objects' boundary, located at a medium distance from the objects' boundary, and located far from the objects' boundary, respectively. These results show that human observers can perceive local depth variations even inside the interior of silhouettes, where no explicit visible information exists (i.e., inside the interior of a silhouette, there are no variations in image shading, there is no surface texture, no conventional binocular disparity, etc.). In order for human participants to make reliable judgments about the local surface depth of regions inside a silhouette, the information about 3-D shape obtained from the outer boundary contour must obviously propagate inwards from the boundary into the interior of the silhouette (Tse, 2002). These results indicate that human observers possess an amazing ability to recover information about 3-D object shape from boundary contours.

Given the previous review demonstrating the perceptual informativeness of boundary contours, perhaps it is not surprising that researchers in computer vision (e.g., Liang & Wong, 2010; Mendonça, Wong, & Cipolla, 2001; Wong & Cipolla, 2001) have developed algorithms that can recover abitrary solid object shapes from the deforming silhouettes that accompany object rotation in depth. It is exciting that these

methods can extract solid object shape from photographic images taken with ordinary cameras. In the future, we may thus have autonomous robotic workers and companions whose visual ability to perceive shape is facilitated by the deforming boundary contours that accompany object motion in an active world.

Conclusion

The boundary contours (and cast shadows) of solid objects contain large amounts of optical information that are of fundamental importance for the perception, discrimination, and recognition of environmental objects.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183-193. https://doi.org/10.1037/h0054663

De Winter, J., & Wagemans, J. (2008). The awakening of Attneave's sleeping cat: Identification of everyday objects on the basis of straight-line versions of outlines. *Perception*, *37*(2), 245-270. https://doi.org/10.1068/p5429

Hilbert, D., & Cohn-Vossen, S. (1983). *Geometry and the imagination*. New York, NY: Chelsea.

Kennedy, J. M. (1974). *A psychology of picture perception: Images and information*. San Francisco, CA: Jossey-Bass.

Koenderink, J. J. (1984a). The internal representation of solid shape and visual exploration. In L. Spillmann & B. R. Wooten (Eds.), *Sensory Experience, Adaptation, and Perception: Festschrift for Ivo Kohler* (pp. 123-142). Hillsdale, NJ: Erlbaum.

Koenderink, J. J. (1984b). What does the occluding contour tell us about solid shape? *Perception, 13*(3), 321-330. https://doi.org/10.1068/p130321

Koenderink, J. J. (1990). *Solid shape*. Cambridge, MA: MIT Press.

Koenderink, J. J., & van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and Vision Computing*, *10*(8), 557-564. https://doi.org/10.1016/0262-8856(92)90076-F

Liang, C., & Wong, K. K. (2010). 3D reconstruction using silhouettes from unordered viewpoints. *Image and Vision Computing*, *28*(4), 579-589. https://doi.org/10.1016/j.imavis.2009.09.012

Mendonça, P. R. S., Wong, K. K. & Cipolla, R. (2001). Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6), 604-616. https://doi.org/10.1109/34.927461

Norman, J. F., Bartholomew, A. N., & Burton, C. L. (2008). Aging preserves the ability to perceive 3-D object shape from static but not deforming boundary contours. *Acta Psychologica*, *129*, 198-207. https://doi.org/10.1016/j.actpsy.2008.06.002

Norman, J. F., Dawson, T. E., & Raines, S. R. (2000). The perception and recognition of natural object shape from deforming and static shadows. *Perception*, *29*(2), 135-148. https://doi.org/10.1068/p2994

Norman, J. F., Dukes, J. M., Shapiro, H. K., Sanders, K. N., & Elder, S. N. (2020). Temporal integration in the perception and discrimination of solid shape. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02031-0

Norman, J. F., Lee, Y., Phillips, F., Norman, H. F., Jennings, L. R., & McBride, T. R. (2009). The perception of 3-D shape from shadows cast onto curved surfaces. *Acta Psychologica*, *131*, 1-11. https://doi.org/10.1016/j.actpsy.2009.01.007

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception, 30*(11), 1285-1294. https://doi.org/10.1068/p3272

Norman, J. F., Phillips, F., Cheeseman, J. R., Thomason, K. E., Ronning, C., Behari, K., … Lamirande, D. (2016). Perceiving object shape from specular highlight deformation, boundary contour deformation, and active haptic manipulation. *PLoS ONE*, *11*(2), e0149058. https://doi.org/10.1371/journal.pone.0149058

Norman, J. F., & Raines, S. R. (2002). The perception and discrimination of local 3-D surface structure from deforming and disparate boundary contours. *Perception & Psychophysics*, *64*(7), 1145-1159. https://doi.org/10.3758/BF03194763

Norman, J. F., Todd, J. T., & Orban, G. A. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, *15*(8), 565-570. https://doi.org/10.1111/j.0956-7976.2004.00720.x

Norman, J. F., Wheeler, S. P., Pedersen, L. E., Shain, L. M., Kinnard, J. D., & Lenoir, J. (2019). The recognition of solid object shape: The importance of inhomogeneity, *i-Perception*, *10*(4), 1-14. https://doi.org/10.1177/2041669519870553

Pollick, F. E., Giblin, P. J., Rycroft, J., & Wilson, L. L. (1992). Human recovery of shape from profiles. *Behaviormetrika*, *19*, 65-79. https://doi.org/10.2333/bhmk.19.65

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement,

image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215. https://doi.org/10.1037/0278-7393.6.2.174

Tse, P. U. (2002). A contour propagation approach to surface filling-in and volume formation. *Psychological Review*, *109*, 91-115. https://doi.org/10.1037/0033-295X.109.1.91

Van Effelterre, T. (1994). Aspect graphs for visual recognition of three-dimensional objects. *Perception*, *23*(5), 563-582. https://doi.org/10.1068/p230563

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*(4), 205-217. https://doi.org/10.1037/h0056880

Wagemans, J., De Winter, J., Op de Beeck, H., Ploeger, A., Beckers, T., & Vanroose, P. (2008). Identification of everyday objects on the basis of silhouette and outline versions. *Perception*, *37*(2), 207-244. https://doi.org/10.1068/p5825

Wheatstone, C. (1838). Contributions to the physiology of vision. – Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371-394. https://doi.org/10.1098/rstl.1838.0019

Wong, K. K., & Cipolla, R. (2001). Structure and motion from silhouettes. *Proceedings of the 8th IEEE International Conference on Computer Vision*, *2*, 217-222. https://doi.org/10.1109/ICCV.2001.937627