
ELSEVIER
VISION
RESEARCH
An International Journal for Functional Aspects of Vision
Test Contrast
Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control
ISSN 0042-6989 | Volume 50 | Issue 1 | 11 January 2010

Vision Research 50 (2010) 1–11

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Minireview

New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume

Zygmunt Pizlo [a,*], Tadamasa Sawada [a], Yunfeng Li [a], Walter G. Kropatsch [b], Robert M. Steinman [a]

[a] Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA
[b] Institute of Computer Aided Automation, Vienna University of Technology, Vienna, Austria

ARTICLE INFO

Article history:
Received 4 August 2009
Received in revised form 25 September 2009

Keywords:
3D shape
A priori constraints
Simplicity principle

ABSTRACT

This paper reviews recent progress towards understanding 3D shape perception made possible by appreciating the significant role that veridicality and complexity play in the natural visual environment. The ability to see objects as they really are "out there" is derived from the complexity inherent in the 3D object's shape. The importance of both veridicality and complexity was ignored in most prior research. Appreciating their importance made it possible to devise a computational model that recovers the 3D shape of an object from only one of its 2D images. This model uses a simplicity principle consisting of only four *a priori* constraints representing properties of 3D shapes, primarily their symmetry and volume. The model recovers 3D shapes from a single 2D image as well, and sometimes even better, than a human being. In the rare recoveries in which errors are observed, the errors made by the model and human subjects are very similar. The model makes no use of depth, surfaces or learning. Recent elaborations of this model include: (i) the recovery of the shapes of natural objects, including human and animal bodies with limbs in varying positions (ii) providing the model with two input images that allowed it to achieve virtually perfect shape constancy from almost all viewing directions. The review concludes with a comparison of some of the highlights of our novel, successful approach to the recovery of 3D shape from a 2D image with prior, less successful approaches.

1. Introduction

Conventional wisdom holds that the perception of a real object is based exclusively on sensory data representing the physical properties of the object such as its weight, reflectance, rigidity, stiffness and texture. This view has dominated the field called "human perception" for centuries and has influenced the field called "machine vision" since it began a half century ago. This conventional wisdom ignored, or deemed unimportant, the critical *abstract* property of a real object, called its "*shape*". The distinction between the abstract shape of a real object and the object itself was made by Plato about 2500 years ago. Plato not only recognized the importance of the abstract property, shape, he also realized that the *symmetry* of a *volumetric* shape is its defining characteristic. He made this clear by describing what subsequent authors have called "Platonic solids". Beginning with Plato's student, Aristotle, the distinction Plato made between the *abstract* shape of an object and the object, itself, has made trouble for philosophers, as well as for scientists when they took over the study of perception. Plato's distinction between a "*real*", *or* "*physical*", *object* and its *abstract, non-physical, shape* played an important role in the philosophical controversies called *materialism* vs. *idealism* and *empiricism* vs. *nativism.* Why did Plato's distinction lead to so much confusion about the meaning of shape for hundreds of years? The answer to this question can be traced to Aristotle's claim that there is nothing in the mind that was not first in the senses. This claim was taken very seriously by modern empiricists, starting with John Locke and the Bishop Berkeley in the 17th and 18th C., and by most perceptionists ever since. It seemed obvious to almost everyone interested in perception that knowledge about real objects "out there" must be based on obtaining sensory data about the physical properties of the objects. The *shapes* of these objects, by being abstract, obviously provided less reliable knowledge about them than their concrete sensory properties. At present, no one questions the fact that shape is an important property of an object despite its abstract nature. Today philosophers, artists, architects, film directors and even perceptionists accept the "reality" of shape, but the nature of shape remained, until very recently, an elusive characteristic that could only be loosely related to real objects (see Pizlo, 2008, for a detailed historical treatment of 3D shape). Does shape refer to geometrical properties of an object such as its angles, or only to topological properties such as its parts? Is shape characterized by the object's surfaces, or by its contours? Does shape refer to spatially global characteristics, and if it does, what does "global" really

* Corresponding author. Address: Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, USA. Fax: +317 496 1264.
E-mail address: pizlo@psych.purdue.edu (Z. Pizlo).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.09.024

mean? These are but a few of the remaining questions perceptionists have in mind when they think of and talk about shape.

Our recent work has made it clear that the *shape* of a symmetrical 3D object is the property responsible for its being perceived veridically. Technically, *a priori* knowledge about: (i) the symmetries of 3D shapes, (ii) the nature of these symmetries' interactions with the geometrical *complexity* of these shapes, as well as (iii) their interactions with probabilistic properties associated with different 3D viewing directions, are more important for the *veridical* perception of the shapes of real 3D objects, than the sensory data obtained from the objects, themselves. How could *a priori* knowledge be more important than "direct" sensory experience? This claim seems implausible. It is surely anathema for any empricistically-minded scientist or philosopher. It is true nonetheless. Appreciation of this fact encouraged us to develop a new theory of 3D shape perception, a theory that should, once it is established, lead to change within broad areas of visual perception because it can be extended from the perception of 3D shape to figure-ground organization and to the perception of 3D scenes, as well as to visual navigation within these scenes. Our review will: (i) describe the theoretical and empirical developments that led to the formulation of our new theory of 3D shape perception and (ii) include demonstrations that show how well a computational model, based on our novel approach, works. It will begin with explaining how a failure to appreciate the significance of *veridicality and complexity* in perception research prevented progress in the study of shape. This section will be followed by a description of the progress we have made towards understanding 3D shape perception during the last 10 years. This progress depended primarily on our development of an understanding of how critical *symmetry and volume* were in the perception of 3D shape. This section of our review contains demos illustrating the scope and capability of our current computational model. These demos will make it clear that our model can recover the 3D shape of a complex abstract object from a single 2D image and also recover the 3D shapes of a wide variety of natural objects from a single 2D image, including the human body as its posture changes. Our model's recovery of 3D shape is accomplished without using any information about depth, surfaces, or learning. This review concludes with a comparative summary of our novel approach. It highlights how our approach differs from other approaches to the study of shape.

2. Veridicality and complexity

For more than a century scientists engaged in perceptual research have concentrated on studying failures of veridicality rather than on how this important, ubiquitous perceptual property is achieved. This emphasis goes counter to our "common sense". Everyday life experience tells us that we almost always see things as they actually are "out there" when it comes to the perception of 3D shape, arguably the most important visual characteristic within our physical environment. 3D shape is *unique* in perception, because 3D shape is the only visual property that has sufficient complexity to guarantee *accurate* identification of objects. Furthermore, 3D shape perception, when shape is properly defined, is almost always veridical. In our everyday life, we can easily recognize important objects such as a car, chair, or dog on the basis of their shapes alone. The shapes of most real objects are perceived veridically, but, despite what "common sense" tells us, most of the contemporary "scientific wisdom" tells us otherwise. The reason for this discrepancy between the commonsensical and scientific understanding of our perceptions of 3D shape is based upon the fact that contemporary wisdom about 3D shape perception is *not* actually built on studying shape. It is built on studies of perceptual illusions of depth and on the perception of surfaces. This approach assumes that depth and surfaces are the appropriate building blocks for 3D shape perception. This is an unnecessary, as well as an unfortunate, assumption. It is plausible geometrically, but not perceptually. Depth and surface orientation are not sufficiently complex to support the perception of 3D shape. Also, neither the depth nor the visible surfaces of objects posses the essential characteristics of 3D shape called "symmetry" and "volume". Studying depth and surfaces could tell us about the perception of shape if, *but only if*, shape perception were actually based *exclusively* on information contained in the 2D retinal image. The 2D retinal image, itself, does not have any volume and is almost never symmetrical, so, it is not surprising that those students of vision, who assumed that "there is nothing in the mind that was not in the senses, first", completely overlooked the concepts underlying shape called "symmetry" and "volume": these concepts are critical when it comes to studying and understanding 3D shape perception. But, note that trying to derive the perception of 3D shape directly from the 2D retinal image has *not*, despite its prevalence, been a plausible assumption for almost a 100 years, not since the Gestalt psychologists showed that all percepts, and especially the perception of 3D shape, require the operation of a *simplicity principle*, which is much more important than the actual elements contained in the 2D retinal image. Before the Gestalt contribution, it had been commonly-assumed that studying the elements contained in any complex percept could tell us something important about the complex percept, itself. This time-honored, but dangerous, assumption puts one at the top of a slippery slope. For example, once it had been demonstrated that the human's perception of 3D *distance* was not veridical, one could go onto assume that a human's 3D *shape* perception will always be distorted because shape perception was assumed to be built on distance perception. Furthermore, one could also assume that the distortions of 3D shape observed can be predicted from the perceptual errors observed when the 3D distances of the test stimuli were judged.

This way of thinking has a long history. It started at least a 1000 years ago when Alhazen (1083/1989) described shape and size constancy and suggested that they could be explained by "taking slant and distance into account". Alhazen's approach was almost universally accepted by philosophers and perceptionists throughout the 17th–19th centuries. It was also used when the first modern experiments on shape constancy were performed by Thouless (1931a, 1931b). Thouless used only ellipses and triangles to study shape constancy, not knowing that ellipses and triangles are completely unsuitable for studying shape constancy because they are too simple. When an observer is asked to recognize or reconstruct the shape of an ellipse or a triangle, he *must* reconstruct the orientation of the surface first and then *must* take it into account when he judges the shape of ellipses and triangles. Surface orientation must be established first because the shape of an ellipse or triangle is completely ambiguous under *perspective* projection. Perspective projection eliminates all information about the shape of an ellipse and of a triangle. In other words, the shape of both ellipses and triangles and their orientation are confounded in the retinal image, so one cannot judge their shapes without knowing their orientation.[1] The fact that an observer can take orientation into account when judging the shapes of ellipses and triangles does not mean that taking orientation into account is the only, or even the main, mechanism for perceiving the shapes of other more complex objects. With more complex objects

[1] Obviously, a triangle is never confused with an ellipse. Our claim that shape is confounded with orientation refers to the shape of a particular ellipse when compared to any other ellipse, and the shape of a particular triangle when compared to any other triangle.

other mechanisms are more likely to be used because with more complex shapes than ellipses and triangles, the shape of a single 2D retinal image actually can provide a lot of useful information about a 3D shape "out there", considerably more than most people realize.

Consider the image of the opaque cube shown in Fig. 1. Marr (1982) pointed out that you can only see half of the six surfaces of this opaque cube (only 50%), but note that you can easily see seven of its eight vertices, and nine of its 12 edges. So, if shape is represented by vertices and edges, rather than by (or in addition to) surfaces, we often see 70% or even 87% of the shape, appreciably more than 50%, as Marr claimed. So, the perception of a 3D shape from its single 2D image would be much easier than Marr claimed once shape is defined and represented by points, contours and surfaces, rather than by surfaces alone. In fact, Marr's, as well as those who followed his lead, choice of surfaces for the representation of 3D shape seems to have been the worst possible choice. Simply put, relying exclusively on surfaces to represent shape, as most modelers have for more than 20 years, has guaranteed that most of the useful shape information contained in the 2D image would be ignored.

Unfortunately, ellipses and triangles were by far the most popular stimuli used in studies of shape perception for many years after Thouless published his frequently-cited, but misguided, work. Fortunately, not everyone insisted on studying only these simplest of all possible shapes. Gibson (1950) was probably the first to be concerned about the obvious discrepancy between the results of laboratory studies of shape and common sense (described above). Gibson, unlike most of his contemporary perceptionists, accepted the commonsensical fact that human beings perceive 3D shapes and 3D scenes veridically. He recognized that the failure to observe this in the laboratory was a problem for the laboratory scientists not for ordinary people going about their activities in the real world. He went onto suggest that complex 3D shapes, resembling natural objects, should be used to study human shape perception. Unfortunately, Gibson, like Marr, was committed to surfaces being the appropriate building blocks for 3D shapes and Gibson and his intellectual heirs' commitment to this idea probably contributed a lot to the fact that so few actually complex 3D shapes have been used to study shape perception during the 30 years since Gibson's death, and even these shapes have tended to be too simple, e.g., elliptical cylinders and rectangular pyramids. Furthermore, these 3D shapes were not only too simple to be used in studies of shape, they were almost always presented from very special viewing directions, called "degenerate views", that is, views specifically chosen to *remove all 3D shape information*. Here, as was the case in Thouless' and derivative experiments, the observer *must* use information about the depth and orientation of surfaces to make judgments about 3D shape because the shape of these stimuli is completely ambiguous in their 2D retinal images when they are presented in degenerate views. It is not surprising, then, that the shape judgments in these experiments were very variable, as well as biased. They could not be, and hence were not, better than depth judgments made under similar conditions (Johnston, 1991; Norman, Todd, Perotti, & Tittle, 1996; Todd & Norman, 2003). There was no way the results could have come out otherwise.

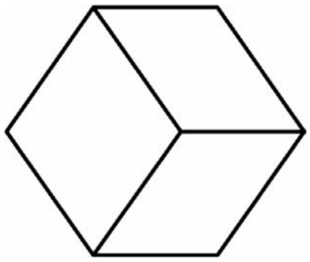

Fig. 1. The image of an opaque cube. Only half of its six faces can be seen. This fact is the main idea underlying Marr's emphasis on what he called the "2.5D sketch", but note that you can see seven of the eight vertices and nine of the 12 edges of this cube. So, once 3D shape is represented by contours or vertices, much more than one half of the 3D shape can be seen.

Eventually, a few more complex 3D shapes were actually used in the laboratory, but here, the researchers put considerable effort into producing what can best be called 3D "degenerate" shapes, specifically, 3D polygonal lines, potatoes, amoebas and crumpled newspapers (e.g., Edelman & Bülthoff, 1992; Rock & DiVita, 1987). These objects are actually amorphous objects because the perception of their shape *can never* be veridical. Shape constancy must fail when they are used, because their retinal images do not contain *any* information about 3D shape. Each of them is amorphous for at least one reason, namely: (i) it is asymmetrical, (ii) it does not have any volume, or (iii) it does not have useful contours. Symmetry, volume and contour are *essential* characteristics of 3D shape. These are the characteristics that allow the application of the effective *a priori* constraints (aka "priors") we use to recover the 3D shape of an object from one of its 2D retinal images. The absence of one or more of these properties has fatal consequences for the study of shape. Without all three there is no 3D shape.

Biederman and Gerhardstein (1993) were the first to demonstrate that 3D shape constancy can be achieved reliably. They used 3D shapes composed of symmetrical 3D parts. Different shapes either had different parts, or had the same parts, arranged differently. So, we have only known for slightly less than 20 years that it is critical to actually use stimuli that have 3D shape to study 3D shape, and that once we do, shape constancy will be observed in the laboratory as well as in everyday life. Once this was appreciated, it became clear that a meaningful definition of shape was needed. Until now, all individuals working on shape perception have implicitly assumed that all 3D objects possess the abstract property called "shape", but this assumption, as implemented by almost all shape researchers has, more often than not, had undesirable experimental consequences. Recently, Pizlo (2008) reviewed the long, confusing history of research on shape and proposed an empirical definition of shape intended to prevent, or at least reduce, the prevailing confusion. His definition is based on the assumption that shape constancy is the *sine qua non* of shape. This makes sense because shape, unlike the slant or the depth of a surface, does not depend on viewing direction. One can only be certain that the abstract property called "shape", rather than surfaces, is being studied when it can be shown that shape constancy was actually achieved with the stimuli used. Employing this empirical definition will not be expensive because it only requires that an experimenter present his 3D shapes from more than one viewing direction. If the chosen stimulus actually has 3D shape, its presence will be manifested from more than one viewing direction. An analytical definition, based on geometrical and topological properties of objects, may also be possible now that we have developed a computational model that recovers 3D shape from a single 2D image. The *a priori* constraints used successfully by this model suggested a basis for an analytical definition.

3. Towards an analytical definition of 3D shape

The success of our computational 3D shape recovery model rests on only four *a priori* constraints, namely, maximal 3D symmetry, maximal planarity of contours, maximal 3D compactness and minimal surface area.[2] These four constraints operate on the three properties that contribute to shape, namely, volume (maximal 3D compactness), surface (minimal surface area) and contours (maximal 3D symmetry and maximal planarity of contours). These three properties can be used to predict whether an object actually has 3D shape, or whether it is what we will call an "amorphous object". Without volume, surfaces or contours, our constraints cannot be applied and when they cannot be applied, 3D shape cannot be recovered. These three properties, viz., volume, surfaces and contours, form the basis of a concept called a "combinatorial map" in topology (Alexandrov, 1947/1998; Brun & Kropatsch, 2003). A combinatorial map describes how the space, containing the shape, is partitioned, that is, which volumes fill the 3D space, which 2D surfaces bound these volumes, and which 1D contours bound these surfaces. In a simple case, such as a cube residing in an empty 3D space, a combinatorial map represents; (i) a single volumetric object and its background (the rest of the 3D space), (ii) the decomposition of the bounding surface into six sides of the cube, (iii) the decomposition of the contours, bounding each side, into four edges, and (iv) eight corners, each corner bounding three edges.[3] A *geometrical* interpretation of such a cube is produced from the *topological* partition, represented in the combinatorial map, by adding information about the planarity of the surfaces, the straightness of the contours, the equality of the angles and the line segments, and the positions of the endpoints of the line segments. Combinatorial map descriptions of 3D shapes can range from very simple (the cube described just above) to very complex, where many (in principle, an infinite number of) tiny elements of volume, surface and contour are present. *We propose that a 3D object can be said to possess the abstract property called "shape" only when its geometry can be represented by a combinatorial map of bounded complexity, providing that the object and/or its combinatorial map is symmetrical. If the 3D object cannot be represented this way, this object is, and should be called, "amorphous".* The requirement in this analytical definition called "bounded complexity" guarantees that spatial details, which do not contribute to shape, are excluded.[4]

The combinatorial map representation is obvious with polyhedral shapes like the one shown in Fig. 2b. The reader *always* sees a 3D symmetrical shape when he views the 2D image in Fig. 2b, despite the fact that this particular 3D shape is both abstract and unfamiliar. One can easily see in this single 2D image, which surfaces (faces) encompass the volume, which contours encompass the faces and where the symmetry is located. Note that in this case, both the object, itself, as well as its combinatorial map, are symmetrical. The reader *never* sees a 3D symmetrical shape if the contours are left out as they are in Fig. 2a. *N.B.* that eliminating *only one* of the four levels of representation in this combinatorial map completely eliminated the perception of 3D shape in Fig. 2a. If the analytical definition we are proposing were to be adopted, one would not expect a 3D shape to be perceived with the kind of stimulus shown in Fig. 2a. Note that even adding short edges around all of the vertices of the polyhedron, as was done in Fig. 2c, does not make it easy to see the polygonal 3D shape. The 3D shape is only likely to be perceived when the observer can interpolate the polyhedron's edges perceptually to form closed contours encompassing its faces (see Pizlo, Li, & Francis (2005), for a direct test of the interaction between this kind of interpolation and a 3D symmetry constraint).

Now, we can ask whether adopting our empirical and analytical definitions of 3D shape will provide classifications of natural 3D objects that are consistent with our common sense? Animal and human bodies are arguably the most important 3D shapes humans deal with in everyday life. We recognize them reliably from almost any viewing direction. Shape constancy is achieved very easily with such objects. This satisfies our empirical definition of 3D shape, but can such important objects also be handled by our analytical definition? Can these objects be represented by a combinatorial map? Natural objects like these have clearly visible surfaces and volumes, but how about their contours? Animal and human bodies can always be represented faithfully by simple line-drawings. This fact, which has been known to human beings for quite a while, is illustrated in Fig. 3. The line-drawing of the young elephant, which was produced with a modern computer graphics' technique called "suggestive contours", is shown on the left (DeCarlo, Finkelstein, Rusinkiewicz, & Santella, 2003). The line-drawing of the horse shown on the right was found on the wall of a cave near Lascaux in France. This line-drawing was made by visually-guided human hands more than 10,000 years ago. The 2D contours

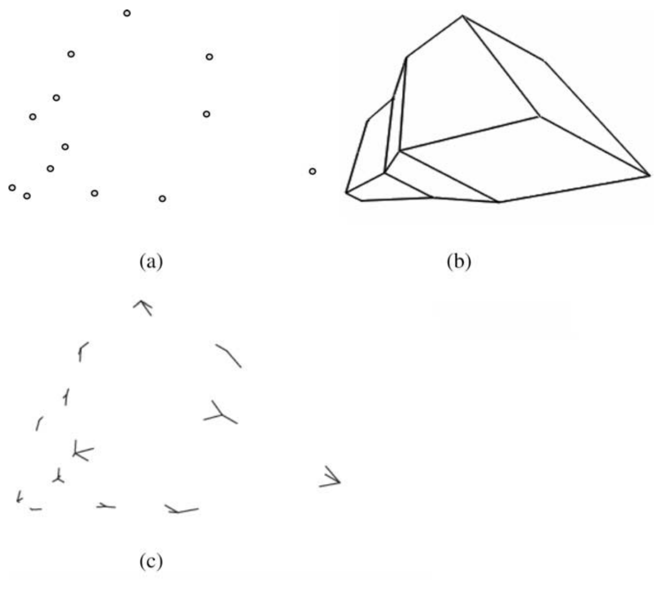

Fig. 2. You will never see the 3D symmetrical shape seen in (b) when only the dots representing the vertices of the polyhedron in (b) are provided as they are in (a). Note that even adding short edges around all of the vertices of the polyhedron, as was done in (c) does not make it easy to see the polygonal 3D shape. The 3D shape is only likely to be perceived when the observer can interpolate the polyhedron's edges perceptually to form closed contours encompassing its faces.

[2] "Symmetry", as used here, means mirror symmetry or translational symmetry. Mirror symmetry is present when one half of an object (e.g., one half of a human head) is a mirror image of the other half. Translational symmetry is defined and illustrated in Fig. 4. Topological mirror symmetry is present when a graph representing an object is mirror symmetric. Consider a human body. Let each important feature of a human body (the head, joints etc.) be represented by a node of a graph. Let two nodes be connected by an edge when the corresponding features are physically connected. The positions of the nodes and the orientations of the edges in the graph should be ignored. Only the relations among nodes and edges are important, which means that changing the articulation of arms and legs of a human body does not change anything in the graph of the body. Once this is appreciated, it becomes clear that the graph representation of a human body will always be symmetrical regardless of postural changes.

[3] There are several equivalent combinatorial data structures, e.g., abstract cellular complexes, plane graphs in 2D, hypergraphs in higher dimensions, generalized maps, descriptions used in algebraic topology, i.e., homology and co-homology groups, etc. A combinatorial map is a formalism expressing all the important topological relations explicitly.

[4] The volume, surfaces and contours of the 3D shape of a stimulus should all have spatial scales similar to the scale of the entire 3D shape. This has not always been the case in modern experiments purporting to be on shape. For example, the surface and volume of the thin wire from which Rock and DiVita (1987) made their wire objects had a much smaller spatial scale than the length of the wire they used to make their stimulus. This meant that the volume and surface of the wire could contribute little, if anything, to the 3D geometry of their wire object, and could, therefore, have been left out of the wire object's representation. With the volume and surface greatly minimized, or even removed, from the wire's representation, a combinatorial map cannot be constructed and the wire object will be amorphous, i.e., have no 3D shape.

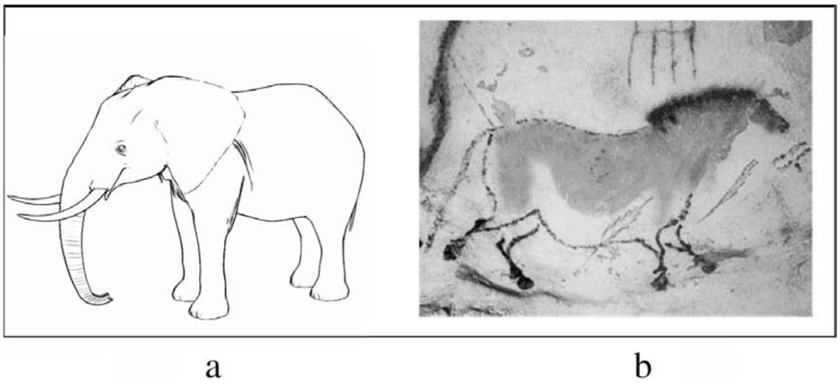

Fig. 3. Line-drawings of animals. (a) Young elephant (from DeCarlo:http://www.cs.princeton.edu/gfx/proj/sugcon/models/ with kind permission of the author). (b) The line drawing of a horse found on the wall of a cave near Lascaux in France .

of both animals, new and not so new, are sufficient to produce the perception of opaque, clearly recognizable 3D objects with clearly-defined surfaces and volume. The surfaces and volumes of their bodies are perceived as clearly as the surfaces and volumes of the polyhedron in Fig. 2b. What makes this possible? The simplest explanation seems to be that the lines in both drawings provide enough visual information to allow *a priori* 3D shape constraints to operate, and when they do, they produce a 3D shape percept that includes not only the 3D contours, but also the surfaces and volume of the 3D object represented in the drawing. We are not the first to call attention to this possibility. Recall that Biederman (1987) proposed that animal and human bodies can be represented by simple volumetric objects he called "geons" (basically, generalized cones; Binford, 1971; Marr & Nishihara, 1978). An important characteristic of such objects is their special kind of symmetry called "translational symmetry". The term "translational" refers to the process by which these generalized cones are constructed. Consider the triangular top face of the generalized cone shown on the bottom left of Fig. 4. This triangle has been "swept" downwards in 3D space along a curve called the object's "axis". This sweeping process consist of nothing more than taking a copy of the triangle and pasting it infinitely many times along this axis. This operation is called "translational symmetry" (Weyl, 1952). The size of the "cross section" of the generalized cone (the triangle in the bottom left example) need not be constant, and the contour of the cross section might be polygonal or smoothly curved. One only needs a pair of contours, the contour representing the cross section and the curve representing the axis, to form a generalized cone. Its surfaces and the volume are produced by the sweeping process. This fact, alone, suggests that generalized cones can be represented by a combinatorial map.

So, once generalized cones, which are characterized by translational symmetry, can be used to represent the limbs, torso, neck and head of an animal, it is reasonable to propose using a combinatorial map to represent human and animal bodies as well as polyhedra. Note that the configuration of body parts need not itself be symmetrical as is the case when the pose of the arms, legs and/or torso changes. Note that the overall configuration is almost always topologically symmetrical because all normal human and animal bodies have two arms and two legs. In other words, when the geometrical symmetry of the body is perturbed and when its pose is arbitrary, its topological symmetry remains intact. This means that a combinatorial map of a human and animal body will always be symmetrical regardless of the arrangements of its parts.

4. Three-dimensional (3D) shape recovery model (2007 version)

Our novel computational 3D shape recovery model will be described next. Our approach to modeling 3D shape is entirely

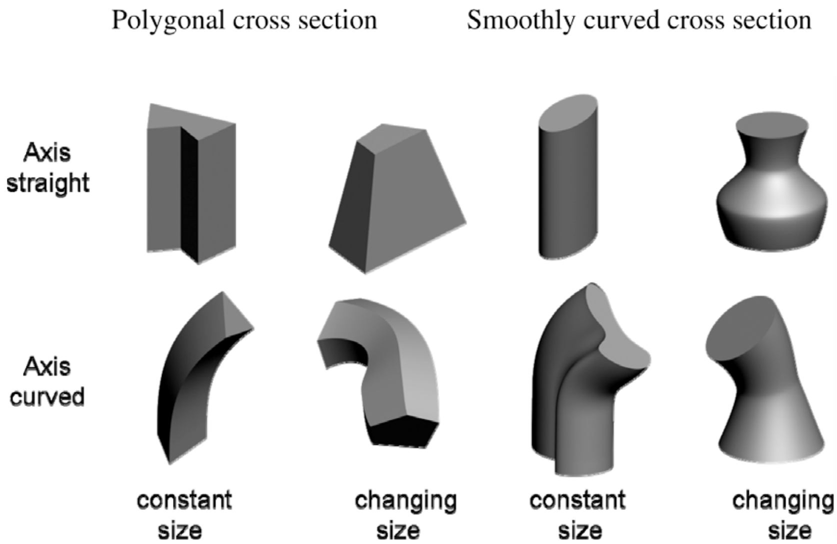

Fig. 4. Some generalized cones (from Pizlo, 2008 – Fig. 5.4).

different from Marr's (1982), which set the stage for much of the contemporary research on shape perception. Marr tried to *reconstruct* 3D shape from depth and surfaces. We *recover* 3D shape from a limited number of *a priori* constraints and do not use depth, surfaces or learning. Once the importance of both veridicality and complexity in 3D shape perception is appreciated, the significance, as well as the nature of the *a priori* constraints (our "simplicity principle") that can be used to achieve veridicality comes into play. Once they do, two questions arise: (i) why is a simplicity principle essential for establishing the veridical perception of 3D shape and, if it is, (ii) what will this simplicity principle be like? The answer to the first question is obvious. A single 2D retinal image simply does not provide enough information to produce a veridical percept of a 3D shape. In fact, a single 2D image is completely ambiguous. It could have been produced by *infinitely many* 3D shapes "out there". As pointed out above, this ambiguity can be reduced by reconstructing depth and 3D surfaces, but when only *depth* or 3D *surfaces* are used to reconstruct 3D shape, the reconstructed shapes are never veridical. The answer to the second, more interesting, question, requires considerable elaboration because the only way to disambiguate the perceptual interpretation of a 3D shape associated with a given 2D retinal image is to impose *a priori* simplicity constraints on the family of all possible 3D interpretations. In plain English, the visual system can only do this if it knows something about the nature of 3D shapes on the basis of its inheritance.[5] In other words, it "learned" about important properties of 3D shapes during its evolution. It might, for example, know that the shapes of objects "out there" are almost always symmetrical and that they always have some volume. Such knowledge will obviously be helpful, but will knowledge about symmetry and volume be *sufficient* to permit the veridical perception of 3D shape? It turns out that it will.

Li, Pizlo, and Steinman (2009) published a computational model that shows how this can be done, and a DEMO illustrating the model's performance should now be viewed at: http://www1.psych.purdue.edu/~sawada/minireview/.

Start DEMO 1 to see how well our model recovers a complex, abstract 3D polyhedral shape. This synthetic polyhedron simulates a 3D polyhedral object that could be "out there". Check the box labeled "simulated shape". This will make the simulated 3D shape rotate. It is rotated to allow you to view it from different viewing directions. Note that this "simulated shape" is opaque, that is, only its front, visible, contours are shown. The five 2D line-drawings on top are 2D images of the same 3D simulated shape obtained from five different viewing directions. When you look at *any* of these 2D images, you perceive the same 3D "simulated shape". Try all five by clicking on them one at a time. Now note that there is much more to perceiving stimuli like these than is usually appreciated. People rarely realize that they are *recovering* a veridical percept of a 3D shape from a single 2D retinal image when they look at such 2D stimuli. Note also that your veridical recovery of this 3D shape is taking place despite the complete absence of depth cues. There are none in your display. The fact that you perceived the same 3D shape from each and every one of these 2D images is critical because it provides an unambiguous demonstration of your ability to achieve "shape constancy". The shape of the 2D image changed, but you continued to perceive the same 3D shape "out there".

DEMO 1 also allows you to see how well our model can recover 3D shape from the 2D image shown at the center. Check the box labeled "original image" to see the 2D image used for this recovery; then, check "recovered shape". This will make the "recovered shape" at the center rotate, allowing you to see it from different viewing directions. Now, check "simulated shape", again. Toggling between "simulated" and "recovered" shapes allows you to compare the model's 3D shape recovery to the shape that had to be recovered. As you can easily see, the recovered shape is virtually identical to the simulated shape. Furthermore, the entire 3D shape was recovered, including its invisible back part. Clicking on another 2D image on top will allow you to see the 3D shape recovered from a different viewing direction, a direction that produced a different 2D image of the same 3D shape. Simply, toggle among the buttons labeled "original image", "simulated shape" and "recovered shape". This allows you to see how much shape constancy was achieved by our model. Constancy would be called "perfect" if the 3D recovered shape appeared to be the same when all five of these 2D images were used. Most people, who have viewed this DEMO, reported that the 3D shape remained the same despite all of the changes of its 2D image produced by changing the viewing direction. Simply put, by this criterion, our model's shape constancy is almost perfect. You can convince yourself about how good it is by clicking on all of the 2D images and toggling between the "simulated" and "recovered" shapes. When satisfied, you can exit DEMO 1 by closing its window. Do not close the link, itself, because you will be asked to view DEMO 2 and 3 later.

The model used for DEMO 1 recovered the entire 3D shape, both its front visible, as well as its back invisible part from a single 2D image of an opaque 3D polyhedral object. Depth cues and learning were not needed for the recovery. Recall that a 2D image of a symmetrical 3D shape is almost always *asymmetrical*. It is important to note that the model must be given an *organized* image in the sense that the 2D contours in the image have already been extracted. It must also be given the correspondence of the symmetrical points in the 2D image. In other words, the model, at this time (July 2009), must be provided with what the Gestalt psychologists called "figure-ground organization". Figure-ground organization is arguably the most important unsolved problem in both human and machine vision. We, as well as many other groups around the world are actively working on it.[6] Finally, note that our model recovers 3D edges, represented by points, rather than surfaces. Points and edges are sufficient when synthetic polyhedra are used because they define surfaces and volume uniquely. This will not be the case when real objects are used (see below). With real objects, the interpolating surfaces will be added once the 3D points and edges have been recovered. This is an important point: in our model, 3D surfaces follow, rather than precede, 3D shape recovery. Our recovery order is opposite to Marr's.

It is important to keep in mind that *a priori* constraints can only be effective when they are applied to a *complex* perceptual characteristic such as shape because *a priori* constraints are likely to be effective only if they are applied to relational properties of stimuli, and only complex stimuli manifest meaningful relationships. Surface orientation is not complex; it is represented by only two numbers: slant and tilt. There are not many relations that can be built from these two numbers, and the relations that can be built, are not likely to represent anything critical about a physical object.

[5] Characteristics of 3D shape such as 3D symmetry, volume and a combinatorial map, could, in principle, be learned during one's life through the observer's interaction with his environment. We prefer to think that these characteristics were "learned" during evolution and are present at one's birth. One of our reasons for saying this is that it would be very difficult to learn the concept of 3D symmetry through an interaction with symmetrical objects because we do not have direct sensory access to the 3D symmetry of objects "out there". The retinal images of 3D symmetrical objects are never 3D, nor are they symmetrical. Haptic perception can convey the three-dimensionality of an object, but haptic sensory input is symmetrical only when a 3D symmetrical object is grasped symmetrically. Furthermore, neither vision nor haptics can convey the volume of a 3D object because volume is hidden inside the object. See Pizlo, Li, and Steinman (2008) for a more detailed discussion of our treatment of the nature of *a priori* constraints.

[6] The interested reader is encouraged to examine the work of the speakers invited to the first two workshops devoted to research on shape: http://bigbird.psych.purdue.edu/events.html – the research groups, led by these speakers, represent much of the contemporary work on figure-ground organization.

For example, neither the product nor the ratio of slant and tilt refers to anything important about the geometry of a surface. In contrast, the mirror symmetry of a 3D shape represents something fundamentally important about the object "out there". Consider also that one surface orientation cannot be simpler than another, the same way that a cube is simpler than a randomly-shaped polyhedron. A cube is simpler than a randomly-shaped polyhedron because it has multiple symmetries. Symmetry probably provides the simplest, as well as the best, example of the power of a simplicity principle to allow the perception of 3D shape, and once one realizes that most, if not all, important objects in our natural environment are symmetrical, one would expect symmetry to play a major role in the perception of 3D shape.[7] It surely does. Our everyday life experience tells us, and has told all of our ancestors, that all objects have three dimensions and all 3D objects reside in a three-dimensional environment. Despite these obvious facts little use has been made of symmetry or volume in attempts to understand 3D shape perception despite their obvious potential. Volume has almost never been used in prior studies of shape, and symmetry has been used infrequently. When it was, only a relatively unimportant aspect of symmetry was considered. Specifically, symmetry was only used to reduce the amount of memory required for sensory coding.

Clearly, memorizing a symmetrical pattern requires only half as much memory as memorizing an asymmetrical pattern, but saving some memory does not seem to be critical, because even if it were critical, our retinal images are almost never symmetrical even when the 3D shape "out there" is symmetrical. The retinal image of a 3D symmetrical shape "out there" will only be symmetrical for a very restricted set of viewing directions. Specifically, the 2D retinal image will only be symmetrical when the visual axis is located on the symmetry plane of the 3D shape. Once one realizes that the retinal image of a symmetrical shape is almost never symmetrical, it follows that one has to recover the 3D symmetrical shape *before* any memory can actually be saved. It obviously follows that the recovery of 3D shape takes precedence over saving memory. Once this is understood one can ask "how can symmetry be used to recover 3D shape?" As pointed out above, a single 2D image of 3D object is not sufficient for a unique, let alone a veridical, recovery of the shape of the 3D object whose 2D image is present on the retina. It is not sufficient because the depth of each 3D point is unknown, when only a single 2D image of the 3D shape is available. If the 3D shape has N points, the recovery of the 3D shape is characterized by N unknowns. But, if the shape of the 3D object is symmetrical, as most natural objects are, the recovery of an object's 3D shape can be characterized by only one, or even no, unknown! This means that the shape of a 3D symmetrical object can be recovered veridically, even when individual points on the object's surfaces cannot be recovered at all!

The constraints called "symmetry" and "volume" provide very powerful tools in shape perception because, as you have seen in our 3D shape recovery DEMO, employing these constraints as a simplicity principle produced a computational algorithm that can recover 3D shape veridically. This fact compels one to take symmetry and volume very seriously in perception research and once one does, they can be used to solve a centuries' old problem in visual perception, namely, the recovery of a 3D shape from a 2D retinal image.

Technical details of our model, which was used for the DEMO you just viewed, were described in Li et al. (2009). Here, we will only explain how the symmetry constraint, used in this model, works because symmetry is the model's most valuable constraint. It is valuable because it reduces the ambiguity inherent in the recovery of 3D shape from infinitely many unknown parameters to just one! Consider Fig. 5b which illustrates how a pair of 3D symmetrical points is recovered from the 2D orthographic projection in a retinal image. This kind of recovery, when it is done for all pairs of points of a given 3D shape shown in Fig. 5a, produces a 3D *symmetrical* shape. There are infinitely many different 3D symmetrical shapes corresponding to different orientations of the symmetry plane, but the recovery is computationally very simple when the symmetry plane forms an angle of 45 deg with the x-axis. This simplest-case-recovery is shown in Fig. 5a and b. Points in the 2D retinal image are shown on the x-axis. The equation of the symmetry plane is $z = x$ (the y-axis is not shown). This plane bisects the angle ZOX and is parallel to the y-axis (i.e., the plane of symmetry forms a 45 deg angle with the XOY plane). If the symmetry plane is not parallel to the y-axis, one can always rotate the coordinate system (without restricting generality) such that this assumption is satisfied. In this case, only the x-axis needs to be considered, so the symmetry plane can be substituted by a symmetry line (see Sawada, Li, & Pizlo, submitted for publication, for details). The depths of the two symmetric points are recovered trivially by setting the z-coordinate (depth) of the first point to be equal to the x-coordinate of the second point, and vice versa. This implies that the two recovered points shown in Fig. 5b as filled dots are mirror symmetric with respect to the symmetry line. If the symmetry line forms an angle other than 45 deg with the x-axis, the recovery is also possible, but the computation of the depths of the points is a little more complicated. Each orientation of the symmetry plane leads to a different 3D symmetrical shape. All of these shapes form a one-parameter family. Three members of this family are shown in Fig. 5c. Once symmetry has been used to establish this one-parameter family, a *unique* 3D shape is recovered by applying the additional three *a priori* constraints, namely, maximal 3D compactness, maximal planarity of contours and minimal surface area.

5. Recent additions to our model

5.1. Natural objects

Once we had shown that the model can recover the 3D shapes of abstract symmetrical polyhedral objects very well, we turned our attention to the 3D shapes of natural objects, such as animals and natural man-made objects. This generalization was important for three reasons. First, natural objects, unlike polyhedra, are better represented by contours than by distinctive points. Would our shape-recovery algorithm work on contours? Second, information about 3D contours is not sufficient to uniquely identify the surfaces that are encompassed by the contours, or to uniquely identify the volume that is encompassed by these surfaces. Surfaces and volume can be used to implement the maximal 3D compactness and the minimal surface area constraints despite the fact that surfaces and volume are not explicitly recovered by the model. Note that with synthetic polyhedral objects, we, as well as the algorithm, knew the location of the surfaces and the volume. But, would our shape-recovery algorithm be able to recover 3D contours without accurate knowledge about the object's surfaces and volume? Finally, would the model be robust in the presence of the noise and uncertainty that would always be present with real 2D images of real objects? The answer to all three questions is "yes". Examples of the model's recovery of the 3D shape of six natural objects can be seen by going back to the DEMO page: http://www1.psych.purdue.edu/~sawada/minireview/ and clicking on DEMO 2.

DEMO 2 shows how well our model can recover six "realistic" natural objects. DEMO 2 starts with the image of a "truck" (the

[7] Such objects as trees are never exactly symmetrical, but their global structure is usually approximately symmetrical, e.g., a Christmas tree.

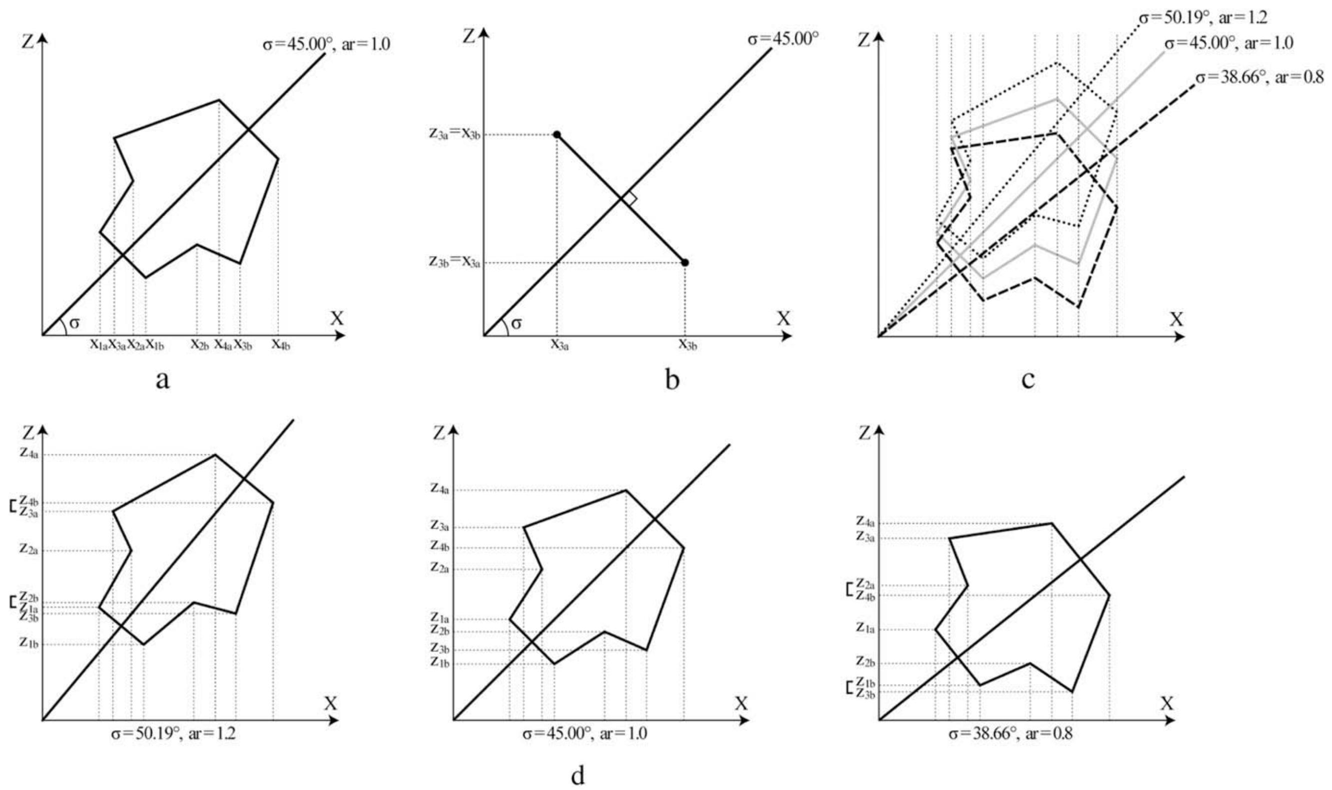

Fig. 5. Recovery of a symmetrical shape. The points on the *x*-axis are points in the 2D retinal image. Only the *x*-axis of the 2D image is shown in these graphs. The symmetry line in (a) and (b) forms a 45 deg angle with the image's axis. The filled circles in (b) are the recovered symmetrical points. Note that the *z*-coordinate (the depth) of the first point is equal to the *x*-coordinate of the second point, and vice versa. It is easy to verify that the two recovered points (x_1, x_2) and (x_2, x_1) are symmetrical with respect to the symmetry line. Three members from the one-parameter family of symmetrical shapes are shown in (c). These shapes have different aspect ratios (*ar*) and the slants of their symmetry planes are also different (σ = 38.66, 45.00, 50.19 deg). The fact that different shapes from this one-parameter family are likely to have a different order of points in depth is illustrated in (d). The three panels show the three shapes from (c). Consider the left and the middle panel. When the slant of the symmetry plane is changed from 45.00 to 50.19 deg, two pairs of points change their order in depth. The *z*-coordinates of these points are marked by two brackets on the *z*-axis of the graph on the left. Similarly, when the slant of the symmetry plane is changed from 45.00 to 38.66 deg, two pairs of points change their order in depth (see the graph on right).

images of the truck and the spider were copied from: Shilane et al., The Princeton Shape Benchmark, http://shape.cs.princeton.edu/benchmark/). You can change to another image by clicking on any of the 2D images on top. You cannot rotate these objects, as you did in DEMO 1, because the digital models of these four real animals in their natural environment, needed to do this, were not available. The 3D shape recoveries shown in DEMO 2 were based on contours in the 2D image extracted by an unskilled human hand. Check "drawn contour" to see these contours. The 3D shape recovered from these hand-drawn contours can be seen by checking "recovered shape". You can avoid clutter by unchecking the "original image" and "drawn contour" boxes. Note that all of these 3D shapes were recovered very well from a single 2D image despite the fact that the contours provided to the model were rather crude. This is important because it shows that our model's recovery of 3D shape is quite robust in the presence of noise and errors in the 2D image provided to it. The human visual system does a much better job extracting contours than the unskilled human, who drew the contours used for 3D recovery in this DEMO. The important message conveyed by DEMO 2 is that the spatially-global aspects of the 2D image (its 2D shape) are the important determinants of 3D shape recovery. Spatial details, such as exact positions of points and magnitudes of curvatures of contours, are irrelevant, allowing us to claim that the whole is not only different from its parts, it is more important than its parts.

In all of the examples you saw in DEMO 2, 2D contours, rather than 2D points were used as input to the model. Contours are computationally more difficult than points because of what is called the "correspondence problem". Specifically, when a 2D image of two 3D symmetrical curves is given, the problem of establishing correspondence between points of these curves always has infinitely many solutions. A unique correspondence is chosen by providing the model with the direction of the symmetry line segments in the 2D image. Recall that in an orthographic image of a 3D mirror symmetrical object, the line segments connecting the images of points that are symmetrical in 3D (the line segments called "symmetry line segments") are all parallel to each other. It follows, that once this direction is known, a unique correspondence for all points can be established (see Sawada et al., submitted for publication, for details). Obviously, the correct orientation of the symmetry line segments cannot be known exactly. It follows that errors in establishing this direction will lead to errors in the recovered 3D shape. We were able to show, however, that the 3D shape recovered was quite robust in the presence of small, or even moderate, errors in establishing the direction of the symmetry line segments.

In DEMO 2, four 3D objects were included that might be expected to create problems for our shape-recovery algorithm (i.e., the spider, the fly, the mantis and the bird) because it was designed to recover objects like polyhedra. One might also expect that our algorithm could be applied relatively easily to other natural objects in DEMO 2, such as the truck, because it does not look very different from the polyhedra used to develop the model. One might think that the bird, mantis, fly and spider would present a special problem because they, unlike the truck, do not have much volume. This expectation is unwarranted because all of the natural objects used in this demo present the same challenge to our algorithm.

Even though the truck looks somewhat like a polyhedron, its 3D contours do not actually show where its surfaces and the volume are located. You can see where they are, but the algorithm cannot. The algorithm only used contours drawn by hand in all of the examples included in DEMO 2. The hand-drawn, skimpy contours used to represent all of these natural objects are sufficient for a human reader to see these drawings as three-dimensional, but the model had work to do before it could recover their 3D shape. It recovered their 3D shape by applying the maximal 3D compactness and the minimum surface area constraints to a 3D convex hull (a box) placed around the 3D object, rather than to the object, itself.[8] This worked because the shape of the 3D convex hull is uniquely defined by the shape of the 3D object within it. Finally, while viewing this demo, keep two things in mind, namely: (i) the less-than-perfect contours in the 2D images were drawn by an unskilled human being and (ii) the 3D shapes recovered by the model are quite accurate despite the noise inherent in all of these hand-drawn 2D contours.

5.2. Human bodies and other nearly-symmetrical shapes

We next asked whether the model can be extended to recover approximately symmetrical shapes. This extension is important for at least three reasons. First, animal and human bodies are never exactly symmetrical because they have movable limbs and flexible bodies. This led us to ask whether a global symmetry constraint could be used to recover asymmetrical body postures. Second, some natural objects, which do not have movable parts, are only approximately symmetrical. A human face is one example. A long-used soft-metal garbage can is another. One wonders how large a departure from symmetry can be tolerated when symmetry is used as one of the constraints. Finally, some objects are not even approximately symmetrical. Can the model know whether the 2D image was produced by a symmetrical, nearly symmetrical or a completely asymmetrical object? The completely asymmetrical case is of particular importance because it could help us solve the figure-ground organization problem. Namely, two independent 3D symmetrical objects (two boxes near each other) almost never produce a 3D symmetrical configuration by accident, that is, the planes of symmetry of the two boxes do not coincide. The figure-ground organization problem here is to discover whether we are dealing with one box or two. One way to separate them is to discover that the two boxes taken together cannot represent a single symmetrical 3D object. All three examples can be handled in the same way. We start by correcting the 2D image of an approximately symmetrical 3D shape so that this image could have been produced by a symmetrical 3D shape. Recall that in any orthographic image of a mirror symmetrical 3D shape, the symmetry line segments (i.e., the line segments that connect pairs of points that are images of points symmetrical in 3D) are all parallel to one another, but if the 3D shape is only approximately symmetrical, the symmetry line segments are not parallel. We correct the 2D image by making all symmetry line segments parallel to their average orientation. Once this correction is made, our model recovers a perfectly symmetrical 3D shape and then deforms the recovered 3D shape so that the original 2D image could have been produced by this asymmetrical 3D shape. This deformation is done by performing the smallest rotation of the symmetry line segments in 3D (Sawada & Pizlo, 2008; Sawada, submitted for publication). Once the 3D shape has been recovered, we can evaluate how asymmetrical it is by comparing the two halves of the recovered 3D shape. Three examples of how symmetrical and nearly-symmetrical 3D shapes were recovered with skeletons of human bodies are available in DEMO 3. See this DEMO by going back to the DEMO page: http://www1.psych.purdue.edu/~sawada/minireview/ and clicking on DEMO 3.

DEMO 3 works the same way as DEMO 2. The only difference is that skeletons, rather than contours, were drawn by hand. The image of the sitting person was produced by a 3D model whose posture was mirror symmetrical in 3D. The postures of the two standing 3D models were asymmetrical in 3D. This DEMO shows that our model's recovery was quite good with both symmetrical and asymmetrical postures.

Now that you have seen that our model can recover symmetrical, nearly-symmetrical and asymmetrical 3D shapes, we can ask whether human observers can do this too. Our experiments suggest that they can. In one of these experiments, an observer was shown a single 2D image and was asked whether the 3D shape of the object that produced it was, or was not, symmetrical. The observers' performance was quite reliable and very similar to our model's (Sawada, submitted for publication).

So far we have a model that recovers 3D contours of 3D shapes of symmetrical and nearly-symmetrical objects based on 2D contours that have been extracted and labeled. The model still cannot: (i) find which region in a 2D retinal image represents a 3D shape "out there"; (ii) extract the 2D contours representing the 3D object; (iii) determine the correspondence of 2D pairs of image points and contours that represent 3D pairs of symmetrical points and contours; and finally (iv) add surfaces to the 3D recovered contours. The development of our model will only be complete when it can perform all these four steps without any human intervention.

5.3. The contribution of a second input image

We continued to elaborate our 3D shape recovery model by exploring the benefits that might accrue from adding information from a second image. This seemed to be an important next step because interpreting a 3D scene on the basis of the information present in two, or more, images is much more common in everyday life than interpreting a 3D scene from a single 2D image. Two different images are available whenever an observer views nearby objects with both eyes, and also when the object or the observer move relative to one another even if one eye is covered.[9] We discovered that the way information from a second image is used for the perception of 3D shape was new, as well as completely unexpected. Prior models that tried to reconstruct 3D shape from two images tried to reconstruct the *metric* structure of 3D points (their relative distances) either: (i) by treating the two 2D images as independent views of 3D points (e.g., Longuet-Higgins, 1981) or (ii) by assuming that the two 2D images represent only slightly different viewing directions of two or more points, and then combining the binocular disparity angle (or motion parallax angle) with contextual information such as the viewing distance, interocular distance or the angular speed of rotation (Howard & Rogers, 1995; Rogers & Graham, 1979, 1982). Note that in all of these models, 3D *shape* was not considered or studied. It was completely irrelevant. Only points or lines were reconstructed. The reader surely knows by now that this is a huge mistake because 3D shape perception is not the same as, or derived from, the perception of points or lines. What role might a second input image play in the perception of 3D shape? There will always be more than one image under natural conditions so one might expect additional images to contribute some benefit to the percept if only by increasing its reliability.

[8] A convex hull of a set of 3D points is the smallest 3D convex object that contains all the points in its interior or on its surface.

[9] When the viewing distance is larger than several meters, the two retinal images in a binocular observer are essentially identical, so the second image provided by having two eyes does not add any depth information.

We discovered the role of additional input images by examining the relationship between the mirror symmetry of 3D shape and the order of points in depth. The order of points in depth seemed to have considerable potential because human observers are extremely good at judging it. Their judgments are so reliable that this ability is called "hyperacuity". It is called "hyper" because the precision of these judgments is an order of magnitude better than the density of the retinal receptors (Regan, 2000; Watt, 1984; Westheimer, 1975). Hyperacuity, as related to the judgment of the order of points in depth, had never been used in models of 3D shape perception probably because hyperacuity refers to a spatially local feature while shape refers to a spatially global feature.[10] It turns out that 3D symmetry can bring these two features together. Recall that the first step in our 3D shape recovery model is the application of a 3D mirror symmetry constraint to compute a one-parameter family of 3D symmetrical shapes. Different members of this family are 3D shapes having different aspect ratios and different 3D orientations (see Fig. 5c). The changes of these two features, aspect ratio and 3D orientation, lead to changes of the depth order of points of a 3D shape. This fact is illustrated in Fig. 5d. The depth order of at least some pairs of points specifies a very narrow range of 3D shapes in the one-parameter family of 3D shapes. The shape with maximal 3D compactness is always very close to the actual 3D shape. The errors in the recovered aspect ratios of the 3D shape are almost never greater than 20% and are usually less than 10%. These are very small deviations from veridicality. The three shapes in Fig. 5d illustrate a 20% difference in aspect ratio. Specifically, the shape on left is stretched 20% compared to the shape in the center, and the shape on the right is compressed 20% compared to the shape in the center. Clearly, these differences in aspect ratio are rather small. Even more important, *the accuracy of the 3D shape recovery from two input images is much less sensitive (fivefold) to viewing direction than the recovery from a single view. Our model and our human subjects were able to achieve virtually perfect shape constancy under all conditions.* It seems that Gibson was right when he emphasized the importance of studying natural viewing conditions, including binocular viewing and allowing his observers and objects within the environment to move about. The enhanced performance of our model produced by allowing a second input image suggests that it was more important than he, or anyone else, supposed.

6. Comparative summary of our approach to 3D shape perception

(1) The importance of studying and modeling the veridical perception of 3D shape is emphasized in our approach. Almost all prior research emphasized the importance of visual illusions, that is, the non-veridical perception of 3D shape and space. This emphasis was encouraged by a desire to perform psychophysical experiments under laboratory conditions designed to achieve 'scientific purity", but this desire encouraged using impoverished, and highly-unnatural conditions. Using such conditions generated a voluminous literature of questionable significance for understanding the perception of 3D shape in the laboratory much less for understanding shape perception in the natural visual environment.

(2) Symmetry is assigned very special status in our theoretical treatment of 3D shape unlike in all prior theories in which symmetry was only used to simplify sensory coding. In our approach, symmetry serves as a critical *a priori* simplicity constraint; symmetry makes it possible to recover 3D shapes veridically. Furthermore, our symmetry constraint can be used to recover the shapes of both symmetrical and nearly-symmetrical objects. The strategy used to recover nearly-symmetrical shapes by applying a symmetry constraint is new. It is very useful, as well as important, because it allows our model to recover the 3D shape of such ecologically-significant objects as a human body with its head and limbs in different orientations with respect to its torso.

(3) Three additional *a priori* constraints, namely maximal 3D compactness, minimal surface area, and maximal planarity of contours are also used in our approach to permit the recovered 3D shapes to be as veridical as 3D shapes recovered by human observers with the same stimuli. Minimal surface area and maximal planarity have been used in prior computational models of 3D shape, but our maximal 3D compactness constraint is a novel, as well as a particularly effective, addition. Our model recovers 3D shape as well as, and sometimes even better, than a human observer. No prior computational model could even come close to the human beings ability to do anything like this.

(4) Our model can also recover the 3D shapes of natural objects from most of their possible 2D images. Recovery of the 3D shape of a natural object from a single 2D image was, prior to our work, believed to be impossible. We not only showed it to be possible, we also showed that the computational model that does this is surprisingly simple, so simple that it is reasonable to propose that this kind of model could be used by lower animals as well as by human beings. It also can recover the entire 3D shape of an opaque object, including its invisible back. This accomplishment cannot be matched by any model that reconstructs the 3D surfaces of objects.

(5) We explained how the interaction of symmetry with binocular disparity and motion parallax can lead to the almost perfect recovery of a 3D shape. Providing our model with binocular input (with objects within 2 m) makes 3D shape recovery much more accurate than recovery based on monocular input. This result calls attention to a new role for both binocular disparity and motion parallax; both of these visual processes are always present and active under natural viewing conditions. Two input images, provided either by binocular viewing or by motion parallax, enhance the accuracy of 3D shape perception appreciably, making it nearly perfect and allowing a high degree of accuracy *from any viewing direction.*

(6) We showed that the perception of 3D shape is veridical *because* shapes are complex. Specifically, the veridical recovery of the shape of a 3D symmetrical object is possible despite the fact that the veridical reconstruction of individual 3D points is impossible. This computational fact is a good example of the oft-quoted Gestalt claim that "the whole is different from the sum of its parts". This fact also shows that we not only solved Bishop Berkeley's problem by recovering a 3D shape from a single 2D image, it also shows that we understand why this problem remained unsolved so long. Namely, no one realized that recovering a complex 3D shape *before* reconstructing its elements was the correct way to proceed.

Acknowledgment

The authors were supported by grants from the National Science Foundation, Department of Energy and the Air Force Office of Scientific Research.

References

Alexandrov, P. S. (1947/1998). *Combinatorial topology*. NY: Dover.

Alhazen (1083/1989) *The optics. Books 1–3*. (Translated by A.I. Sabra, London: The Warburg Institute).

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

[10] Brookes and Stevens (1989) and Norman and Todd (1998) showed that stereoacuity thresholds are elevated substantially when the points used as stimuli, whose order in depth was judged, were placed on a discontinuous sawtooth surface or on the surface of an irregular object like a "potato". Under these special conditions stereoacuity was not a hyperacuity. The stereoacuity thresholds were probably elevated in these experiments because stereoacuity was put into conflict with monocular cues to 3D depth. Stereoacuity may continue to be a hyperacuity when shape constraints such as symmetry, compactness and planarity are not put into conflict with stereoacuity.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions from three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1162–1182.

Binford, T. O. (1971) Visual perception by computer. In *IEEE Conference on Systems and Control*. Miami.

Brookes, A., & Stevens, K. K. (1989). Binocular depth from surfaces versus volumes. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 479–484.

Brun, L., & Kropatsch, W. G. (2003). Contraction kernels and combinatorial maps. *Pattern Recognition Letters, 24*, 1051–1057.

DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., & Santella, A. (2003). Suggestive contours for conveying shape. *ACM Transactions on Graphics, 22*, 848–855.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research, 32*, 2385–2400.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. NY: Oxford University Press.

Johnston, E. P. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*, 1351–1360.

Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research, 49*, 979–991.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, 293*, 133–135.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B, 200*, 269–294.

Norman, J. F., & Todd, J. T. (1998). Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. *Perception, 27*, 257–272.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 173–186.

Pizlo, Z. (2008). *3D shape: Its unique place in visual perception*. Cambridge, MA: MIT Press.

Pizlo, Z., Li, Y., & Francis, G. (2005). A new look at binocular stereopsis. *Vision Research, 45*, 2244–2255.

Pizlo, Z., Li, Y., & Steinman, R. M. (2008). Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose. *Spatial Vision, 21*, 495–508.

Regan, D. (2000). *Human perception of objects*. Sunderland, MA: Sinauer.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology, 19*, 280–293.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22*, 261–270.

Sawada, T. (submitted for publication). Visual detection of symmetry in 3D shapes.

Sawada, T., Li, Y., & Pizlo, Z. (submitted for publication). Any 2D image is consistent with a 3D symmetric interpretation.

Sawada T., & Pizlo, Z. (2008). Detecting mirror-symmetry of a volumetric shape from its single 2D image. In *Proceedings of the workshop on perceptual organization in computer vision. IEEE international conference on computer vision and pattern recognition*. Anchorage, Alaska, June 23.

Thouless, R. H. (1931a). Phenomenal regression to the real object. *British Journal of Psychology, 21*, 339–359.

Thouless, R. H. (1931b). Phenomenal regression to the real object. *British Journal of Psychology, 22*, 1–30.

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception and Psychophysics, 65*, 31–47.

Watt, R. J. (1984). Towards a general theory of the visual acuities for shape and spatial arrangement. *Vision Research, 24*, 1377–1386.

Westheimer, G. (1975). Visual acuity and hyperacuity. *Investigative Ophthalmology, 14*, 570–572.

Weyl, H. (1952). *Symmetry*. Princeton: Princeton University Press.