Chapter 2
Symmetry Is the *sine qua non* of Shape

Yunfeng Li, Tadamasa Sawada, Yun Shi, Robert M. Steinman, and Zygmunt Pizlo

2.1 Introduction

“Shape” is one of those concepts that seem intuitively obvious, but prove to be surprisingly difficult to define. In this paper, we propose a solution of this seemingly insoluble definitional problem. Our definition of shape is based on a fundamentally new first principle. By starting from scratch, we avoided what had been an insurmountable problem inherent in the traditional way of thinking about shape. In our definition, shape is characterized by a *similarity of the object to itself not to other objects* as had always been done previously. This new characterization is done by specifying how spatial features of the object are transformed, spatially or temporally, to its other spatial features. Such transformations, which are called *symmetries*, are the object’s *self-similarities*. In order to anticipate objections of some readers that our definition is too narrow because it excludes objects that are completely asymmetrical from the class of objects having shape, we can point out that our definition explains what is surely the most fundamental perceptual phenomenon of shape called, “shape constancy”.

By the way of reminder, *shape constancy refers to the fact that the perceived shape of a given 3D object is constant despite changes in the shape of the object’s 2D retinal image. The retinal image changes when the 3D viewing orientation changes*. Conventional wisdom holds that our perceptual systems always strive for perceptual constancy and it also accepts empirical results showing that perceptual constancy in general, and shape constancy in particular, is never fully achieved. Constancy always falls far short of perfection. But note that if shape is not defined properly, a putative study of “shape constancy” is likely to produce failures of constancy simply because shape was not actually being studied. It would be completely unreasonable to expect that the observer’s visual system is able to achieve shape constancy when what is meant by “shape” changes from study to study often in ad

Y. Li · T. Sawada · Y. Shi · R.M. Steinman · Z. Pizlo (✉)
Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, USA
e-mail: pizlo@psych.purdue.edu

S.J. Dickinson, Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*, Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5195-1_2,

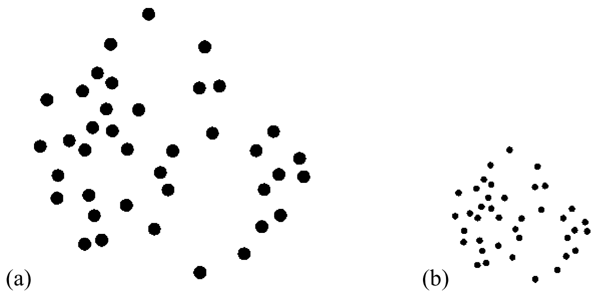


Fig. 2.1 (**a**) An irregular set of scattered points. (**b**) Is a scaled down version of (**a**). According to most conventional definitions, (**b**) has the same shape as (**a**). It is not clear, however, what this shape actually is, or whether either of these two dot patterns actually possesses the property we mean when we say a visual stimulus has “shape”

hoc, arbitrary ways. A better way of dealing with this confusion is to determine how shape *should* be defined so as to make it possible to show in the laboratory, what commonsense tells us happens in everyday life where shape constancy is perfect. This is what *we* did. We started by accepting that shape constancy is the *sine qua non* of shape, without shape constancy there is no shape. By starting this way, we were able to define shape *operationally* [16]. This worked well for planning shape experiments and evaluating their results but it was less than ideal because one cannot know whether an object has shape until shape constancy with the stimulus used was verified by viewing it from more than one direction.

Using an operational, rather than analytical, definition presented us with two problems, namely: (i) it can be argued that our definition was circular, and (ii) this, like all, operational definitions did not provide any *analytical* tools that could be used to formulate a mathematical or computational model of shape constancy. The first problem can be partially circumvented by pointing out that our operational definition, at the very least, allows identification of the class of objects that satisfy the shape constancy criterion. Recall, that for centuries common wisdom believed that shape constancy could *never* be achieved with *any* object. Our operational definition made it possible for us to show convincingly that shape constancy could be achieved with many objects. The second problem made it clear that an analytical definition of shape was needed. This chapter explains how this was done by proposing that *there is as much shape in an object as there is symmetry* (*regularity*) *in it*. Note that the complete *failure* of shape constancy will *never* be observed once you accept our new definition of shape. In fact, when our new definition is used, shape constancy is *almost always perfect*, and when shape constancy does fall short of perfection, we know why it does and we can explain the extent of the failure in every case. Should you worry about excluding objects that have no regularities in them from a definition of shape? The answer is “no” because our definition of shape applies to *all* natural objects important to human beings, including, animal bodies and plants, as well as to the tools we use.

Our new definition questions whether *all* objects and *all* patterns exhibit the property called “shape”. Does the spatial arrangement of the points in Fig. 2.1a

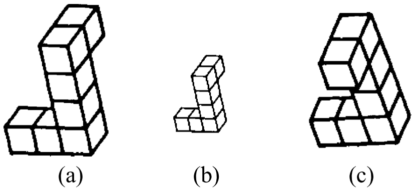


Fig. 2.2 The object in (**b**) is identical to the object in (**a**) except for its overall size. The object in (**c**) was produced by computing a 3D reflection of (**a**). According to the conventional definition of shape, all three objects have the same "shape" (after [22])

have shape? According to conventional definitions it does. If the pattern of points in Fig. 2.1a has shape, is it the same visual quality as the shape of, say, a butterfly or of an airplane? No matter which conventional definition of shape you prefer, your *commonsense* will tell you that the answer to this question is a resounding "No". The quality of shape inherent in a butterfly or in an airplane is nothing like any shape you can make out in the dotted pattern shown in Fig. 2.1. So, if we want to include *all* patterns and *all* objects in a comprehensive discussion of shape, some objects will surely have more shape than others, and there will even be *amorphous* objects without any shape, whatsoever. Bent wires and crumpled papers will fall on, or near, the amorphous end of this continuum. Prior definitions of shape will be reviewed before our new definition of shape is explained.

2.2 Prior Definitions of Shape

Most contemporary shape theorists agree that the property we have in mind when we refer to some visual arrangement as having shape refers to some aspect of this arrangement that is "invariant under transformations". Consider first, an example of what is probably the most appropriate transformation that can be used when we try to define shape. This transformation is produced by the rigid motion of an object within a 3D space. Pulling a chair away from a table is a good example. The position of the chair within the room has changed (this is what we mean by the "transformation"), but the chair, itself, did not. We call this kind of transformation a "rigid motion" because all of the geometrical properties of the chair (what the conventional definition calls the chair's "shape") stay the same. These properties are "invariant." The size of the chair stays the same, as well as all the distances and angles between the individual parts that made it up. The legs are not broken or bent, and the individual parts are not stretched by this kind of transformation. It follows that if there are two identical chairs in the room, we would say that they have the same shape.

Note that this conventional definition of shape is often generalized, slightly, by including a 3D reflection of the object and the change of its overall size. This results in a "similarity transformation." Look at Fig. 2.2. According to the conventional definition of shape, all three objects seen in Fig. 2.2 have the same shape. All angles remain the same in a similarity transformation, so an angle formed by two line segments is an invariant of this transformation. If all corresponding angles in two objects are equal, one object can be produced by transforming the other by using a similarity transformation. These two objects are said to have the same "shape" because such a transformation is possible.

This is by far the most commonly used definition of shape. There are several variants of this definition that use more general groups of transformations, leading some shape experts to suggest that shape refers to invariants of an affine transformation:

$$\begin{aligned} x' &= ax + by + cz + d \\ y' &= ex + fy + gz + h \\ z' &= kx + ly + mz + n \end{aligned} \tag{2.1}$$

Affine transformation allows for uniform stretching of an object along an arbitrary direction. As a result, angles, surface areas and volumes are no longer invariant. What is invariant is the ratio of areas of two figures residing on parallel planes or the ratio of the volumes of two objects. According to this definition, any two rectangular boxes, say a shoebox and a pizza box have the same shape. This definition obviously violates our commonsense. Most people would say that a pizza box and a shoebox have very different shapes. Few, if any, people would look for their pie in the shoebox, or try to put their foot in the pizza box. Despite the obvious fact that the affine definition of shape is counterintuitive, this definition has been used in shape perception research and applications for two reasons. The first reason is geometrical. A camera image of a planar figure can be approximated by a 2D affine transformation of the figure [16]. It follows that affine invariants of planar figures will be preserved (approximately) in any camera image. This could serve as a tool for recognizing planar figures in camera (or retinal) images. The second reason was suggested by the results of psychophysical experiments. When an observer is asked to judge depth relations of points on 3D surfaces, the judgments are *always* quite *unreliable*. This poor performance was taken to indicate that metric aspects of depth are not reconstructed by the observer, which has led many, probably most, researchers to conclude that metric aspects of depth are not represented in the visual system. The smallest non-metric group is the affine group, so the observer's failure to judge metric properties led many shape experts to claim that shape is represented by affine invariants in the human visual system. The first reason just described is acceptable to us, but the second is not. We believe that *the definition of shape, including perceived shape, should be based on what the human visual system can do very well, not on what the visual system cannot do*. Very many, quite different, reasons are probably responsible for failures in visual perception, and using the failure of shape perception does not seem to be a good way to derive a useful definition. Affine invariants obviously cannot form the basis of a useful definition of shape, at least not shape as we humans perceive it. A transformation that shows that shoe and pizza boxes have the same shape cannot apply to human shape perception.

The affine group is not the end of the line when it comes to trying to use more and more abstract properties to define shape. Another definition of shape uses a

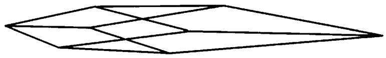


Fig. 2.3 An image of a 3D projective transformation of a cube (from Pizlo [16])

projective group of transformations:

$$x' = \frac{ax + by + cz + d}{px + qy + rz + s}$$
$$y' = \frac{ex + fy + gz + h}{px + qy + rz + s} \quad (2.2)$$
$$z' = \frac{kx + ly + mz + n}{px + qy + rz + s}$$

The motivations for using this group are essentially the same as those used with the affine group. The advantage of using a projective group is that, unlike an affine transformation, a projective transformation provides an accurate description of image formation in a camera or in the human eye (but see [17, 18] for a detailed discussion of the limitations of the projective group as the model of retinal image formation). The disadvantage is that the projective group is larger than the affine group. Comparing them, a 3D *affine group* is characterized by 12 parameters, 5 of which affect the 3D shape as defined by a 3D rigid motion plus size scaling while a 3D *projective group* is characterized by 15 independent parameters, 8 of which affect the 3D shape as defined by a 3D rigid motion plus size scaling. Note that all hexahedra with 8 vertices and 6 quadrilateral faces are valid 3D projective transformations of a cube as long as the planarity of quadruples of points in the cube is preserved. According to the projective definition of shape, the object in Fig. 2.3 should look like a cube. This, obviously, is not the case. The fact that a 3D projective transformation of a cube does not look like a cube is precisely the reason why the Ames's room demo is so striking. According to a projective definition of shape, there is nothing special in Ames's distorted room. Ames's trapezoidal room has, according to this definition, the same shape as a normal rectangular room. So, despite the fact that the projective transformation is an essential tool for describing the relation between the 3D space and the 2D retinal image, the projective group, like the affine group, cannot provide the foundation needed for the study of human shape perception.

Shape is sometimes defined by an even more general group of transformations, namely, the topological group. The topological transformation is a continuous transformation. When used in a 2D space, this transformation is often called "rubber sheet geometry", because the rubber can be stretched arbitrarily without tearing or cutting. The main reason for using a topological group to define shape is that it allows one to identify two different postures of an animal body as the same shape. But the "price" paid for being able to handle non-rigid objects is very high: for example, when a topological definition is used, a needle and a coffee cup have identical shapes! Both are 3D surfaces with one hole. It is obviously the *metric* properties which allow one to use a cup to drink and a needle to sew. Obviously, the topological transformation, like the affine and the projective transformations, is not without its problems when human shape perception is under study.

There is a way to avoid the excessive generality inherent in the topological transformation (just described) while preserving the ability to handle non-rigid and piecewise rigid objects, namely, the shape under consideration can be characterized by geodesics along a surface. Recall that the shortest path between two points on a surface is a geodesic curve of the surface. When an animal changes its posture, all geodesic lines stay the same, or nearly the same. Similarly, when the stem of a flower bends, the geodesics along its surface stay approximately the same. So, geodesic lines are much more attractive than topological properties for describing shapes. There are, however at least two serious shortcomings in using geodesic lines. First, finding geodesic lines is computationally difficult, so using them beyond toy examples is impractical. Second, defining a 3D shape by using lines, which are 1D properties, will not work because geodesic lines do not convey any information about the volumetric aspects of the object. For example, all *origami* shapes (3D shapes produced by folding paper) are identical in a "geodesic" definition of their shape, and they *all* have the same shape as an unfolded, flat piece of paper.

Clearly, there are multiple problems with all of the conventional definitions in use for describing shape: some are too restrictive and others too general. Recall what we really want our definition to do. We want it to exclude random dot patterns like the pattern shown in Fig. 2.1, but we want it to include non-rigid objects such as walking animals and human beings. Furthermore, if we do not want to exclude *any* objects, whatsoever, can we find a way to assign some degree of shape to all objects, even to objects with very little or even no shape? It can be done but this requires us to adopt an entirely new way of thinking about shape. The way we adopted goes as follows: *If shape is to capture permanent* (*invariant*) *properties of an object's geometry, properties that will allow us to recover the object, recognize it, remember it and identify its function, shape must refer to the object's intrinsic characteristics in a way that does not require comparing one object with other objects. The way to do this, perhaps the only way, is to define shape by object's self-similarities.*

2.3 Explanation of the New Definition and How We Worked It out

Recall that all conventional definitions of shape have assumed that *all* objects have shape. Intuitively, even commonsensically, something seems to be missing from this very strong claim. Namely, there are patterns and objects that actually have no shape at all, or at most, they have very little of this property. Asking someone about the shape of the pattern of randomly generated points like the pattern shown in Fig. 2.1, makes little sense. Commonsense tells us that there is little, if any, shape in Fig. 2.1. We also "know" that shapeless common objects exist in everyday life. A crumpled piece of paper, a bent paperclip, or a rock before it is shaped by a human hand do not

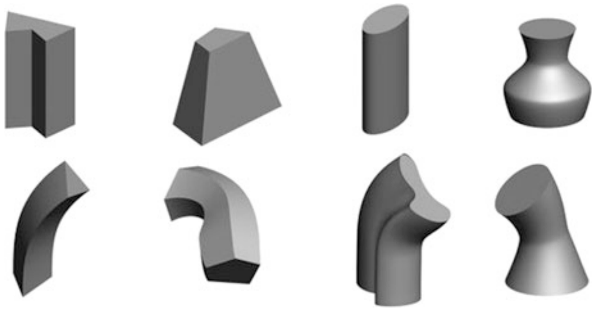


Fig. 2.4 Eight differently-shaped meaningless objects characterized by translational symmetry. The shape of the cross section is constant for each cone, but the size is not necessarily constant. The axis is orthogonal to the cross sections and it is a planar curve or a straight line (from Pizlo [16])

have what we really mean when we refer to an object's shape. All of these objects, as well as random patterns like the pattern in Fig. 2.1, are, and should be, called "amorphous" or "shapeless." Why? They are amorphous simply because they are *completely* "*irregular*."

This observation makes it very clear that the term *"shape" refers to the spatial regularity* (*self-similarity*) *possessed by an object*. We have all had lots of experience dealing with such regularities in our everyday life. The bodies of *all* animals are mirror-symmetrical. By "mirror-symmetrical" we simply mean that one symmetrical half is the mirror image (the reflection) of the other with respect to the animal's plane of symmetry. But there is more to symmetry than mirror symmetry and reflection. Limbs of animals, trunks of trees, and stems of flowers are characterized by what we call "translational symmetry". An object with translational symmetry is produced by taking a planar shape and sweeping it through a 3D space using rigid motion along an axis. During the sweeping process, the size of the cross section may change. Figure 2.4 shows several examples of objects with translational symmetry. They are called "Generalized Cones" (GC) [2, 4].

Take one of the 8 objects in Fig. 2.4, say the second from the left in the top row. All cross sections of this object are *similar* to each other. The technical meaning of *similar* here is that the members of any pair of cross sections in this object are related to each other by a similarity transformation (rigid motion and size scaling). So, *we can use rigid motion, reflection and size-scaling of the "parts" within the object, itself, to define the shape of the object as its "spatial self-similarity"*(*regularity*) instead of using rigid motion, reflection and size-scaling of the entire object in 3D space to define the shape of this object by comparing it to another object. *Put simply, shape is an intrinsic characteristic of an object because it refers to its self-similarity, rather than to the similarity of one object to another*. A small-scale model of an airplane has the same shape as a real airplane not merely because the model is a scaled version of the plane, but because both the model and a real airplane are characterized by the same symmetries.

Self-similarity of biological forms seems to be their inherent characteristic. It is the result of the natural process called "growth" (D'Arcy Thompson [24]). Growth explains why all flowers and plants are characterized by one or more types of symmetry. They have the shape they have because of *how they grow*. All animal bodies are mirror symmetrical because of the *way they move*. A dog without a mirror symmetrical body could not run straight along a straight path. *All biological forms have shape because all of them are symmetrical.* Inanimate objects such as rocks and crumpled papers, which have no trace of symmetry, are obviously shapeless. It is also important to note that many inanimate objects actually do have shape. All objects that serve some useful function, such objects as furniture and tools, have one or more types of symmetry, without which they would probably be dysfunctional.

Symmetry relations among parts of objects imply the presence of invariants of 3D symmetry transformations. These invariants can be represented as the eigenvectors of the 3D transformation matrix. We will analyze their 2D perspective images to derive the perspective invariants of their symmetries after we derive the formulas for the eigenvectors characterizing their 3D symmetries. These invariants are needed for the veridical recovery of 3D shapes. This approach leads naturally to the two essential aspects that are required to characterize shape perception, namely, (i) properties of the retinal image that provide *visual data* about the *invariants* of symmetries, and (ii) the kind of *a priori knowledge* that is needed to produce the 3D shape percept which provides information about the symmetry transformations characterizing the self-similarities of the particular object. The reader should appreciate the fact that our new definition of shape is *richer* than any of the previous definitions because our definition uses *both* invariants and the transformations, whereas all previous definitions only used invariants.

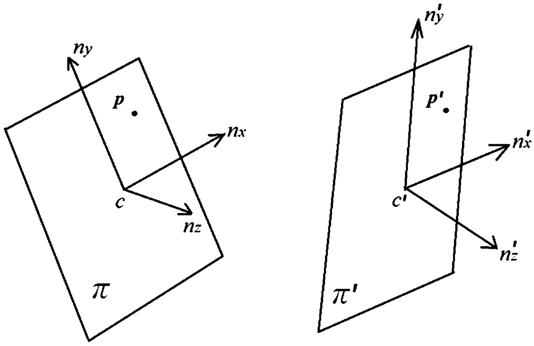
2.4 Symmetry Groups for 3D Shapes, Their Invariants and Invariants of the Perspective Projection

Our analytical definition of shape states that the shape of an object refers to all of its spatially-global symmetries (*its self-similarities*) *as measured by the group of rigid motions, reflections and size-scaling of the "parts" within the object itself.*

Groups of transformations are known to have invariants. Unlike all conventional approaches to shape, we begin *not* with invariants of transformations from one object to another, but with invariants of transformations of one part of an object to another part of the same object. This makes sense because we defined 3D shape as the presence of self-similarity. It is known that a similarity transformation is a linear transformation and that it can be represented by a matrix. Furthermore, it is known that eigenvectors are the only invariant vectors of a linear transformation. It follows that it is natural to look for invariants by analyzing the properties of the eigenvectors characterizing the transformation matrices. Consider the three basic symmetries: mirror, translational and rotational. We begin with a symmetrical shape, whose repeated part is planar, and then extend the results to general symmetrical shapes. Some invariants are limited and exist only for the symmetries with a planar configuration, and the others are general.

Assume that c is a point on a plane π, n_X and n_Y are two perpendicular axes in π. The normal of π is n_Z (see Fig. 2.5). c, n_X and n_Y define a 2D Cartesian coordinate system, in which c is the origin and n_X and n_Y are the two axes. Let a

Fig. 2.5 Illustration of a 3D translation of a point from one plane to the other



3×3 matrix A represent this coordinate system

$$A = (n_X \quad n_Y \quad c) \tag{2.3}$$

Then any point in π can be expressed as

$$P = Ap \tag{2.4}$$

where $p = (p_X, p_Y, 1)^{\mathrm{T}}$ in which p_X and p_Y are the Cartesian coordinates of P in π. Assume that π' is the resulting plane after some rigid transformation of π. The normal of π' is n'_Z and the Cartesian coordinate system is expressed as

$$A' = \left(n'_X \quad n'_Y \quad c'\right) \tag{2.5}$$

The resulting point P after the rigid transformation is obtained as:

$$P' = A'p \tag{2.6}$$

Combining Eqs. (2.4) and (2.6), we obtain the transformation from the point P to P'

$$P' = A'A^{-1}P \tag{2.7}$$

which means that the transformation from P to P' is a 3D affine transformation.

Next, we use A and A' to define the three types of symmetries, translational, mirror and rotational, and identify the invariants for those symmetry transformations.

(a) If $n_X = n'_X$, the transformation from π to π' is a translational symmetry (see Fig. 2.6a). The translation axis (the red curve in Fig. 2.6a) is a planar curve and n_X coincides with the normal of the plane containing the axis. If the translation axis is not a planar curve, the transformation is a mixture of a translational symmetry and a rotational symmetry. It is easy to prove that n_X is one of the eigenvectors of the transformation matrix $A'A^{-1}$. Since n_X is constant and is only determined by the plane in which the translation axis resides (see Fig. 2.6a), n_X is an invariant of the projective transformation from one cross section to another.

(b) If $n_X = n'_X$ and cc' bisects the angle formed by n_Z and n'_Z, the transformation from π to π' is a mirror symmetry (see Fig. 2.6b). Compared with the translational symmetry, an additional constraint is added in the mirror symmetry. It follows that a mirror symmetry with a planar configuration is a special case of the translational symmetry. The fact that cc' bisects the angle formed by n_Z and n'_Z is equivalent to the fact that a symmetry plane (the plane in red in Fig. 2.6b) bisects the planes π to π'. The normal of the symmetry plane is $n_Y - n'_Y$. Both n_X and $n_Y - n'_Y$ are the eigenvectors of $A'A^{-1}$.

(c) If $c = c'$, $n_Z = n'_Z$ and $n_X \neq'_X$, the transformation from π to π' is a rotational symmetry (see Fig. 2.6c). It is easy to prove that c is an eigenvector of $A'A^{-1}$. Since c is the rotation center of a planar rotationally symmetrical object and it is a fixed point, c is an invariant of a rotationally symmetric transformation. The other two eigenvectors of $A'A^{-1}$ are $n_X + in_Y$ and $n_X - in_Y$. They are not invariant because n_X *or* n_Y could be an arbitrary direction (or vector) on the plane π. However, their cross product n_Z is. The geometrical application of the cross product (n_Z) will be discussed in the next part.

Up to this point, we characterized the invariants of the three types of symmetries in 3D space. This is a transformation from one part of an object to another. We are also interested in the invariants of 2D perspective images of 3D symmetry relations—the invariants of the transformation from the image of one part of an object to an image of another part of the same object. This will be essential for detecting 3D symmetries in perspective images and for recovering 3D symmetrical shapes from perspective images.

Assume that a pair of symmetric corresponding points P and P' in π and π' are projected to an image through a camera and that the camera matrix is K. A camera matrix is an upper triangular 3×3 matrix, consisting of a camera's intrinsic parameters, such as its focal length and principal point. Then, the images of P and P' are

$$v = KAp \tag{2.8}$$

$$v' = KA'p \tag{2.9}$$

Note that the image points v and v' are expressed in homogeneous coordinates and they are 3-element vectors (refer to [9], for the details of differences between Euclidean coordinates and homogeneous coordinates). Combining Eqs. (2.8) and (2.9), we obtain

$$v' = KA'A^{-1}K^{-1}v \tag{2.10}$$

Equation (2.10) implies that the relation between images of the planes π to π' is a 2D projective transformation. By analyzing the eigenvectors of $KA'A^{-1}K^{-1}$, we look for the invariants for the above three types of symmetries in their 2D perspective images. It is known that an eigenvector has the following property: if m is an eigenvector of $A'A^{-1}$, then Km is an eigenvector of $KA'A^{-1}K^{-1}$. Therefore, it is easy to identify the invariants in the 2D image from the invariants of 3D symmetry transformations. Next, we list the invariants in the 2D image and explain their geometrical meaning.

(a) In the case of translational symmetry, since n_X is an invariant vector of the symmetry transformation in 3D, Kn_X is an invariant of the projective transformation from one perspective image of a cross section to a perspective image of another cross section. Geometrically, Kn_X represents the vanishing point of the lines that are parallel to n_X. This means that for a 2D projective transformation between the images of any two cross sections, the vanishing point is projected to itself (the invariant point under projective transformation). Identifying the vanishing point Kn_X should help recover translationally symmetrical 3D shapes from their images [23].

(b) In the case of mirror symmetry, n_X and $n_Y - n'_Y$ are the invariant eigenvectors of $A'A^{-1}$. So are Kn_X and $K(n_Y - n'_Y)$ for $KA'A^{-1}K^{-1}$. Kn_X and $K(n_Y - n'_Y)$ are the vanishing points for those lines that are parallel to n_X and $n_Y - n'_Y$, respectively. In particular, $K(n_Y - n'_Y)$ is the vanishing point for those lines that are perpendicular to the symmetry plane. Because $K(n_Y - n'_Y)$ is determined by the normal of the symmetry plane, it is independent of the orientation of π or π'. This means that $K(n_Y - n'_Y)$ can be used with mirror-symmetrical objects whose symmetrical halves are not planar. For example, for the polyhedron in Fig. 2.6e, its lateral side is non-planar and it consists of three planar faces. From the image of each face and of its symmetrical counterpart, we compute a 2D projective transformation matrix. For the three matrices representing the relations between images of the three pairs of symmetrical faces, $K(n_Y - n'_Y)$ is their common eigenvector. In a perspective image, once the vanishing point $K(n_Y - n'_Y)$ is identified and the symmetry correspondences in the image are established, the shape of a 3D mirror symmetrical object can be uniquely determined [14]. Because Kn_X and $K(n_Y - n'_Y)$ are invariant, their cross product $K^{-T}((n_Y - n'_Y) \times n_X)$, representing a line passing through Kn_X and $K(n_Y - n'_Y)$, is also invariant under the projective transformation $KA'A^{-1}K^{-1}$.[1] This means that any point on this line projects onto this line again. The points Kn_X and $K(n_Y - n'_Y)$ are two special points on this line because they project onto themselves.

(c) In the case of rotational symmetry, Kc is the invariant eigenvector of $KA'A^{-1}K^{-1}$. It is the image of c (the image of the rotation center) and it is an invariant point under the projective transformation between the images of a repeated part of a rotationally symmetrical shape. The other two eigenvectors of $KA'A^{-1}K^{-1}$, $K(n_X + in_Y)$ and $K(n_X - in_Y)$ are not invariant. But, their cross product $K^{-T}n_Z$ is and it represents an invariant line. n_Z is the direction of the rotation axis and it is fixed for a rotationally symmetrical shape. $K^{-T}n_Z$ is

[1] The magnitude of a vector is unimportant in a homogeneous coordinate system. So, we can ignore $\det(K)$, which is a constant, from the cross product $\det(K)K^{-T}((n_Y - n'_Y) \times n_X)$ of Kn_X and $K(n_Y - n'_Y)$.

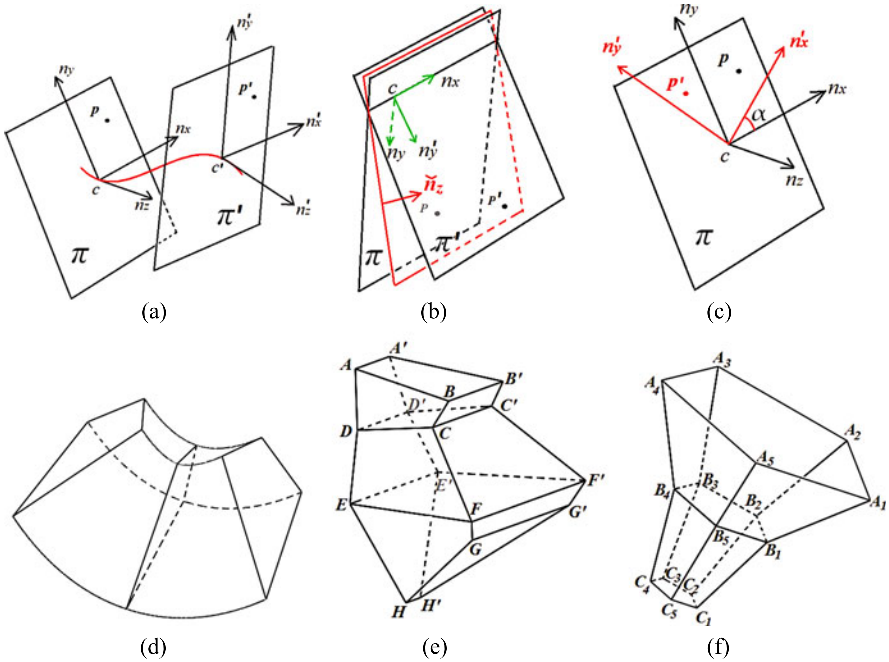


Fig. 2.6 Three types of symmetries and their symmetry transformations. (**a**) Translational symmetry. The *red planar curve* represents the translation axis. (**b**) Mirror symmetry. The *red plane* represents the symmetry plane. (**c**) Rotational symmetry. *c* is the rotation center. (**d**) A translational symmetrical 3D shape with a quadrilateral cross sections. (**e**) A mirror-symmetrical 3D shape that consists of three pairs of mirror-symmetrical planes. (**f**) A rotationally symmetrical 3D shape that consists of three planes

more general than the invariant Kc and it can be applied to the non-planar rotational shape like the one in Fig. 2.6f. $K^{-T}n_Z$ is an invariant line for the plane $A_1A_2A_3A_4A_5$ and it is also the invariant line for the planes $B_1B_2B_3B_4B_5$ and $C_1C_2C_3C_4C_5$.

It is known that at least four points and their correspondences are needed to compute the 2D projective transformation matrix. Therefore, four planar points "out there" and their symmetrical counterparts are needed to identify the invariance in a perspective image. The invariants for the three types of symmetries are listed in Table 2.1. Those invariants representing lines are marked by $*$.

Table 2.1 shows the invariants of symmetry transformations under a perspective projection. Because an orthographic projection is a special case of a perspective projection, these equations can also be applied to the orthographic projection after making two changes in the camera matrix K, and the matrices A and A'. First, in the case of an orthographic projection, the principal point is undefined. So, we set the elements in the camera matrix K that represent the principal points to zero. Second, the last row in vectors A and A' is replaced by $(0, 0, 1)$, which means that the change of Z values of vertices doesn't change their image. As a result, $KA'A^{-1}K^{-1}$ has

Table 2.1 The invariants for the three types of symmetries

Type	Planar configuration	Non-planar configuration
Translation	Kn_X	Kn_X
Mirror	$Kn_X, K(n_Y - n'_Y), K^{-T}((n_Y - n'_Y) \times n_X)^*$	$K(n_Y - n'_Y)$
Rotation	$Kc, K^{-T}n_Z^*$	$K^{-T}n_Z^*$

the same format as A and A', in which the last row vector is $(0, 0, 1)$. It follows that the symmetry transformation under an orthographic projection is a 2D affine transformation, instead of a 2D projective transformation. For an affine transformation, three points and their correspondences are enough to determine the transformation matrix and then identify the invariants. It follows that in the case of an orthographic projection, co-planarity of points or curves is not required.

2.5 Inferring 3D Shape from a 3D Object

With a real object, its shape (its symmetries) must be inferred (abstracted). The symmetries are not given. The best (perhaps the only) way to do this is by using a Bayesian formalism and a closely-related concept called "Minimum Description Length" [11]. This method will be analogous to the "generative" model formulated by Feldman & Singh [8] and used for their 2D medial axis transform (identification of a "shape skeleton"). The main differences are that our model applies to 3D shapes and it handles several 3D symmetries. We start by formulating the problem as a Bayesian inference [10]. Our task is to estimate the 3D symmetries (we call this the "shape" of the object) that best describe a given 3D "object". This means that we try to maximize the posterior probability, $p(\text{shape}|\text{object})$. Take a generalized cone like the one on the top-left of Fig. 2.4. This 3D object has two possible descriptions, one based on translational symmetry and the other based on mirror symmetry (this object has both symmetries). Translational symmetry seems to capture its 3D structure better, so the maximum of the posterior will probably be higher when translational symmetry serves as the shape description than when the description is based on its mirror symmetry. The planar cross section (pentagon) of this Generalized Cone (GC) is a simple 2D figure whose contour information is fairly low [7]. The same is true with the axis of this GC, which is a straight-line segment. It follows that the prior, $p(\text{shape})$, for translational symmetry will be high in this case. This object does not have any random perturbations, which means that the likelihood, $p(\text{object}|\text{shape})$, will be equal to 1.0. As a result, the maximum of the posterior will also be high:

$$p(\text{shape}|\text{object}) = c \cdot p(\text{object}|\text{shape}) \cdot p(\text{shape}) \tag{2.11}$$

Note that what we call "an object", Feldman & Singh [8] call a "shape", but this difference is only terminological. By taking the negative logarithm of both sides of (2.11), we can express our problem in terms of the description length, DL:

$$DL(\text{shape}|\text{object}) = DL(\text{object}|\text{shape}) + DL(\text{shape}) + c' \tag{2.12}$$

Now we can look for the shortest description length, $DL(\text{shape}|\text{object})$, instead of looking for a maximum of the posterior probability. The "shape" solution is the same.

If we consider the maximum of the posterior, $p(\text{shape}|\text{object})$, for mirror symmetry, we will get a smaller value (a more complex description) because mirror symmetry will lead to lower "compression" of the shape of this object. Mirror symmetry will not "know" about the simplicity of its cross section. The only redundancy represented by mirror symmetry is the fact that one half is the same as the other half. Note that this is less obvious than it sounds because the actual prior, $p(\text{shape})$, in this case, depends on how we describe one half of this mirror-symmetrical object. One could do this by using a large number of points on the surface, or by using straight lines, the object's contours, interpolated by planar surface patches. Mirror symmetry might become a better description of an object like the one on the top-left in Fig. 2.4, when the mirror-symmetrical cross-section becomes less regular. This can be done by introducing random perturbation of the object's contours, while keeping these perturbations mirror-symmetrical. Such perturbations will be counted as random noise in the likelihood, $p(\text{object}|\text{shape})$, when translational symmetry, but not when mirror symmetry is used. This will lower the value of the posterior. It should be obvious that the formalisms (2.11) and (2.12) allow both the object's regularities (symmetries) and random perturbations to be handled naturally. In other words, all objects, no matter how irregular, can be described in this way. Less regular objects will have more complex descriptions and the maximum of the posterior, $p(\text{shape}|\text{object})$, ranging between 0 and 1, can be used as a measure of the object's "shapeness".

In this approach, similarities among different shapes can be evaluated by simply comparing the objects' symmetries. In Sect. 2.2, we discussed how metric symmetries can be generalized to affine and projective groups. Recall that all symmetries are defined by the underlying groups of transformations, where "group" has a specific meaning. Group refers to a set of transformations that satisfies the group's axioms, like closure and associativity. It follows that the change from one "shape" to another (where "shape" means a description of an object's symmetries, using a particular symmetry group) will be represented by a transformation of its characteristics (cross section, axis) by using one of the groups, namely, Euclidean, similarity, affine, projective or topological. At this point it is not clear whether this approach will naturally lead to a one-dimensional dissimilarity metric representing the currently conventional way of thinking about similarity in cognitive psychology (e.g., [1]), or whether it will turn out to be a parameterized (geometrical) measure making explicit use of the concepts of transformation groups. After all, when we compare a pizza box to a shoebox, we may be more comfortable saying that "they have different aspect ratios" than that their "dissimilarity is about 7.4".

2.6 Computational and Psychophysical Implications of the New Definition

Now that we have explained both geometrical and algebraic characteristics of shape based on symmetry, we will discuss several interesting implications of our new definition. Two of these implications were anticipated in our recent papers (Sects. 2.6.1 and 2.6.4), and two are new (Sects. 2.6.2 and 2.6.3).

2.6.1 Veridical Perception of 3D Shapes

Recovering a 3D shape from one or more 2D retinal images is an ill-posed inverse problem [15, 20]. This is the case with all difficult inverse problems, so producing a unique and correct interpretation requires the application of constraints to the family of possible solutions. When a 3D symmetry constraint is applied to a single 2D perspective image of a 3D shape, the 3D interpretation is unique and always very close to veridical! The shape recovered is said to be "veridical" because it is the same as the object's shape "out there". During the last 6 years we provided empirical, both simulation and psychophysical evidence, showing how symmetry leads to veridical 3D shape recovery. This includes the recovery of 3D mirror-symmetrical shapes from: (i) a single image [12], and (ii) a pair of images [13], as well as (iii) the recovery of nearly symmetrical shapes [21] and (iv) 3D shapes characterized by translational symmetry [23]. The claim that 3D shapes can be, and actually are, perceived veridically is completely new [19] and until very recently, the "veridical perception of shape" was considered by most shape researchers to be "science fiction", something that does not exist, never has existed, and never will exist. This conventional "wisdom" was based on hundreds of years of reporting failures to achieve shape constancy in the laboratory. Everyone believed that human shape perception could never be perfect or even nearly so. We now know that all of these reported failures came about because everybody was studying the perception of depth, not the perception of shape [16]. Once shape is defined properly, by it symmetries, this confusion is removed and a "miracle" ensues. Shape perception *is* perfect when the viewing conditions and psychophysical measurements are done correctly. How this can be done was explained in our papers (referenced just above) in which we described computational models that use the mathematical properties of symmetry to recover 3D shape and presented extensive psychophysical data on 3D shape recovery and on shape constancy.

2.6.2 Shapes of Non-rigid Objects

When shape is defined by self-similarity, rather than by the similarity of one object to another, it becomes much easier to talk about the shapes of "non-rigid" and

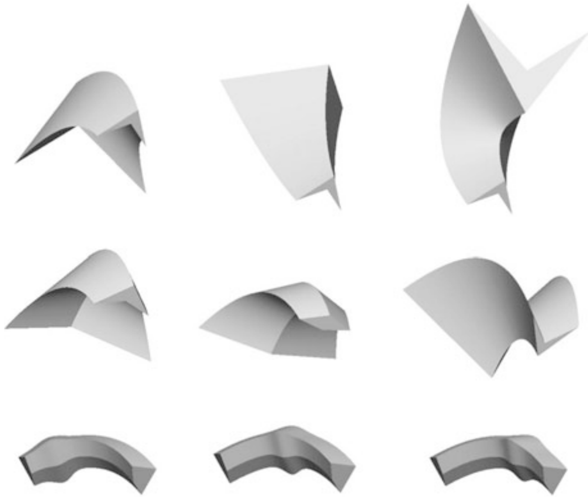


Fig. 2.7 Three snapshots from a range of articulations of non-rigid objects: *Top*—the axis of a GC is changing, but the shape of the cross-section is the same (this looks like a gymnast on uneven bars at the Olympic Games). *Middle*—the shape of the cross-section is changing, but the axis is not (this looks like a flying bird). *Bottom*—the local size of the cross-section is changing, but the shape of the cross-section and the axis of the GC is constant (this looks like a snake that swallowed a large belly-bulging prey)

"piece-wise rigid" objects. If an object is non-rigid, like the stem of a flower, bending the stem does not remove its translational symmetry. All of the flower's cross-sections are still circular. Bending only changes the curvature of the axis of the flower's stem. If an object is piece-wise rigid, like the body of a dog, changes in the articulations of its legs distorts the mirror symmetry of the dog's body, but it does *not* eliminate the symmetry altogether. After all, the dog still has two legs on the right side of his body and two legs on the left side. This obviously applies as well to your body as to your dog's. Our new analytical definition of shape removes the fundamental difficulty inherent in all other conventional definitions of shape. None of them can deal with the non-rigidity of objects, objects that are both common and often very important throughout our natural environment.

Consider some examples (Fig. 2.7). Three snapshots of non-rigid, unfamiliar objects are shown. It is easy to see that the three objects in a given row have something in common. They share symmetries. The objects on top have the same shape of their cross-sections, the objects in the middle have the same axis, and the objects at the bottom have the same axis as well as the same shape of their cross-sections. If an observer is able to see the similarities of the symmetries of an object despite the non-rigidity of this object, he may be able to conclude that the shape of the object being viewed is constant despite its non-rigidity. This is what we mean by perceiving the shape of a non-rigid object.

2.6.3 Symmetry as an Objective, but Informative, Prior

"Objective priors" have a special status in Bayesian methods used to solve inverse problems, probably simply because "objective" sounds more reliable and more scientific than "subjective". But there is another pair of terms for these two types of priors, namely, “uninformative” and “informative”. Uninformative priors are objective in the sense that these priors are derived from some basic statistical and mathematical principles, rather than from some special domain such as knowledge about lung cancer or about earthquakes. Such domain specific knowledge is less interesting because it results in a Bayesian inference method that is specific to a particular domain. Also, it is often difficult to quantify this kind of subjective prior. If the prior is unreliable, the posterior will also be unreliable. The good news in the conventional approach, is that there is an objective way to learn the subjective prior. One begins with an objective, uninformative prior and starts collecting evidence. The posterior computed after the first piece of evidence is acquired is used as a prior for the second piece of evidence. Bayesian inference, including updating priors is optimal in the sense that it extracts all relevant information contained in the data. By the time that the learning has been completed, we have a very good, informative prior that is based on hard data without any “subjective” guessing.

With shape recovery, however, we are presented with a unique situation in which an *objective prior is actually informative*. It seems likely that this unique situation only applies to a symmetry prior. No other prior has this unique characteristic. All other priors in all other inverse problems, can be *either* objective *or* informative. This fact, alone, is responsible both for the special and unique status of shape in visual perception and for the fact that shapes are perceived veridically (see [16], for the uniqueness of shape in visual perception). Once we realize that *all* important objects are symmetrical, the *informative* prior of 3D symmetry becomes an *objective* prior because it refers to mathematical invariants, specifically to invariants of transformation groups. There is no need, whatsoever, to learn group invariants from examples. We can derive them analytically, and once the invariants are derived, we can prove their invariance and examine the necessary and sufficient conditions for them to operate. Symmetries are also informative because they represent the fundamental (permanent, invariant, and intrinsic) characteristics of the 3D objects “out there”. So, once we know that all objects are symmetrical, it makes no sense, whatsoever, to start with any uninformative priors because symmetry, alone, is sufficiently informative, and once symmetry is used as a prior, it also makes no sense, whatsoever, to update it. How could you improve (update) a definition of a mirror symmetry? It simply cannot be done. Note that the symmetry prior can be applied to infinitely many shapes in a finite amount of time, and this includes unfamiliar shapes and even the shapes of non-existent objects.

2.6.4 Shape Constancy: View-Invariant vs. View-Dependent Shape Perception

Note that shape constancy is typically tested with novel (unfamiliar) objects in order to avoid allowing familiarity to influence the shape perceived. All studies of shape constancy prior to ours focused efforts on determining the availability of invariant properties in the 2D image (see [16], for a review). If invariants cannot be extracted reliably from the 2D retinal image, shape constancy fails or, at least, degrades when the size of the change of the viewing direction increases. This result encouraged investigators to accept what is known as the "view-point dependence of shape perception". Before we explain what is missing in this view-point dependent view of shape perception, we will remind the reader about a basic aspect of conventional shape constancy methodology. In a typical experiment of this kind, the subject is shown the same object twice, with the second viewing direction different from the first by an "angle α". The angle α refers to the rotation of the object in depth, that is, a rotation around an axis that is orthogonal to the line of sight. Only then will the shape of the 2D retinal image change, and a change in the shape of the retinal image is the necessary condition for studying shape constancy. When the object is rotated around the line of sight, not orthogonal to it, the 2D retinal shape does not change; only its 2D orientation changes, so such an experiment cannot have any bearing on the shape constancy phenomenon.

Appropriate methodology for performing experiments to test shape constancy introduces a complication that has never been discussed explicitly in the past. For large values of α, shape constancy may be difficult to achieve because some parts of an opaque object that were visible in the first presentation, are not visible after the object is rotated, and new parts may become visible in the second presentation. So, shape constancy, in such cases, may not be perfect for a trivial reason: the relevant information was simply not available to the observer. But if the object is symmetrical, or if it is composed of symmetrical parts, as it was in Biederman & Gerhardstein's [3] experiment, it may be possible to recover the entire 3D shape, including the back, invisible parts. In such cases, shape constancy might be perfect because the entire 3D shape could be recovered correctly in both presentations. This problem has not been studied in the past because there was no computational theory that could predict when an entire shape, back as well as front, can be recovered. We now know that the symmetry of an object is the key concept involved in recovering the invisible backs of 3D objects. These objects must have a sufficient degree of redundancy (regularity and self-similarity) to permit an observer to correctly "guess" (recover) the shape of the hidden part. We already have a computational model that can *usually* recover the entire 3D shape of a mirror-symmetrical object [12]. It can also recover a translationally symmetrical object [23]. However, the entire shape may not be recovered, even if the object is symmetrical, if the object does not have a sufficient degree of regularity. This is precisely what happens with irregular objects like symmetrical polyhedra, whose faces are not planar [5] or with symmetrical irregular "potatoes" and "bell peppers" [6]. It follows that shape constancy is actually much more concerned with invariants in the 3D representation, after the 3D shape is recovered, than with the presence of invariants in the 2D retinal image. For those symmetrical objects, whose entire shape can be recovered, shape constancy will not be affected by the degree of rotation in depth. Put simply, performance will be view-invariant. For objects, like irregular polyhedra, or potatoes and bell peppers, whose back parts cannot be recovered, performance will be view-dependent. This analysis should clarify, once and for all, the apparent controversy between the proponents of both theories. The key to understanding what is going on resides in the recovery of 3D shapes rather than in the presence of cues or invariants in the 2D retinal image.

2.7 Conclusion

In the past, the only thing that everyone agreed about when trying to define shape was that shape refers to the spatially-global geometrical characteristics of an object or a figure. Once one appreciates that all important objects in our natural environment are symmetrical, it follows that any meaningful definition of shape must be based on the concept of symmetry. Imagine how difficult it would be to describe spatially-global geometrical characteristics of a symmetrical object adequately without mentioning its symmetry? It is probably impossible to do this! But using symmetry to describe an object cannot be the whole story because a definition of shape should go beyond a mere description of the object's geometry. The concept called "shape" is used in many ways. We use it to identify objects, we use it to compare similar objects, we use it to remember and to recognize objects, we use it to infer an object's functions, and we use it to identify the permanence of objects in the presence of non-rigidities. We conclude by claiming that *all* of these things can be done *only* when shape is defined by the object's symmetries, as we explained in detail above. Furthermore, all of these things can be done very well, and they can be done in a very principled way because "symmetry groups", with their concepts of transformations and invariants, provide the foundation of large parts of mathematics. By excluding only the very few objects in our natural environment that are completely devoid of symmetries, you can use our new definition of shape to accomplish a great deal more than had been possible before we explained the significance and utility of symmetry in the visual perception of shape. You will have to use experience and learning with irregular rocks and crumpled papers to discriminate their shapes, but with *all* other shapes, you can depend entirely on symmetry.

References

1. Ashby FG, Perrin NA (1988) Toward a unified theory of similarity and recognition. Psychol Rev 95(1):124–150
2. Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94(2):115–147
3. Biederman I, Gerhardstein PC (1993) Recognizing depth-rotated objects: evidence and conditions from three-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform 19(6):1162–1182
4. Binford TO (1971) Visual perception by computer. In: IEEE conference on systems and control, Miami
5. Chan MW, Stevenson AK, Li Y, Pizlo Z (2006) Binocular shape constancy from novel views: the role of a priori constraints. Percept Psychophys 68(7):1124–1139
6. Egan E, Todd J, Phillips F (2012) The role of symmetry in 3D shape discrimination across changes in viewpoint. J Vis 12(9):1048

7. Feldman J, Singh M (2005) Information along curves and closed contours. Psychol Rev 112(1):243–252
8. Feldman J, Singh M (2006) Bayesian estimation of the shape skeleton. Proc Natl Acad Sci 103(47):18014–18019
9. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
10. Knill DC, Richards W (1996) Perception as Bayesian inference. Cambridge University Press, New York
11. Li M, Vitanyi P (1997) An introduction to Kolmogorov complexity and its applications. Springer, New York
12. Li Y, Pizlo Z, Steinman RM (2009) A computational model that recovers the 3D shape of an object from a single 2D retinal representation. Vis Res 49(9):979–991
13. Li Y, Sawada T, Shi Y, Kwon T, Pizlo Z (2011) A Bayesian model of binocular perception of 3D mirror symmetrical polyhedra. J Vis 11(4):1–20
14. Li Y, Sawada T, Latecki LM, Steinman RM, Pizlo Z (2012) Visual recovery of the shapes and sizes of objects, as well as distances among them, in a natural 3D scene. J Math Psychol 56(4):217–231
15. Pizlo Z (2001) Perception viewed as an inverse problem. Vis Res 41(24):3145–3161
16. Pizlo Z (2008) 3D shape: its unique place in visual perception. MIT Press, Cambridge
17. Pizlo Z, Rosenfeld A, Weiss I (1997) The geometry of visual space: about the incompatibility between science and mathematics. Comput Vis Image Underst 65:425–433
18. Pizlo Z, Rosenfeld A, Weiss I (1997) Visual space: mathematics, engineering, and science. Comput Vis Image Underst 65:450–454
19. Pizlo Z, Sawada T, Li Y, Kropatsch W, Steinman RM (2010) New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. Vis Res 50(1):1–11
20. Poggio T, Torre V, Koch C (1985) Computational vision and regularization theory. Nature 317(6035):314–319
21. Sawada T (2010) Visual detection of symmetry of 3D shapes. J Vis 10(6):4 (22pp)
22. Shepard RN, Cooper LA (1982) Mental images and their transformations. MIT Press, Cambridge
23. Shi Y (2012) Recovering a 3D shape of a generalized cone from a single 2D image. Master's thesis, Department of Psychological Sciences, Purdue University, Indiana
24. Thompson DW (1942) On growth and form. Cambridge University Press, Cambridge

Springer

http://www.springer.com/978-1-4471-5194-4

Shape Perception in Human and Computer Vision
An Interdisciplinary Perspective
Dickinson, S.; Pizlo, Z. (Eds.)
2013, XVII, 502 p., Hardcover
ISBN: 978-1-4471-5194-4