

A computational model that recovers the 3D shape of an object from a single 2D retinal representation

Yunfeng Li, Zygmunt Pizlo and Robert M. Steinman

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-2081, U.S.A.

Key words: 3D shape recovery, *a priori* constraints, symmetry, planarity, maximum compactness, minimum surface

Send all correspondence to: Zygmunt Pizlo
Email: pizlo@psych.purdue.edu
Tel: (765) 494 6930
Fax: (765) 496 1264

Acknowledgment: This research was supported by the National Science Foundation (grant # 0533968), US Department of Energy, and Purdue Research Foundation.

Abstract.

Human beings perceive 3D shapes veridically, but the underlying mechanisms remain unknown. The problem of producing veridical shape percepts is computationally difficult because the 3D shapes have to be recovered from 2D retinal images. This paper describes a new model, based on a regularization approach, that does this very well. It uses a new simplicity principle composed of four shape constraints: *viz.*, symmetry, planarity, maximum compactness and minimum surface. Maximum compactness and minimum surface have never been used before. The model was tested with random symmetrical polyhedra. It recovered their 3D shapes from a single randomly-chosen 2D image. Neither learning, nor depth perception, was required. The effectiveness of the maximum compactness and the minimum surface constraints were measured by how well the aspect ratio of the 3D shapes was recovered. These constraints were effective; they recovered the aspect ratio of the 3D shapes very well. Aspect ratios recovered by the model were compared to aspect ratios adjusted by four human observers. They also adjusted aspect ratios very well. In those rare cases, in which the human observers showed large errors in adjusted aspect ratios, their errors were very similar to the errors made by the model.

Introduction

Human observers perceive shapes of 3D objects veridically, that is, as they actually are “out there”. The percept of their shape rarely changes, when the direction from which the objects are viewed, changes. This fundamental perceptual achievement is called “shape constancy.” Shape constancy is the *sine qua non* of shape. It allows us to know that we are studying shape, rather than the depth or the orientation of surfaces. In other words, shape constancy provides the defining characteristic for the presence of the property called “shape”. The depth and orientation of surfaces depend on viewpoint. Shape does not. Shape constancy is a member of a large group of perceptual constancies that include size, speed, lightness, and color. All of these constancies are defined in the same way, namely: the percept of X is constant despite changes in the viewing conditions that produced changes of the 2D retinal image of X. But, are all perceptual constancies produced by the same kind of underlying mechanism? Until recently, it was commonly assumed that all constancies, including shape, were: They were all achieved by “taking context into account.” Context was assumed to be critical for all of these constancies because the 2D retinal image of a 3D scene is ambiguous. But note that shape is fundamentally different from all other perceptual properties because it is complex. Complexity allows us to perceive shape without making use of context, that is, without making use of such information as the distance or the orientation of the object’s surfaces. What do we mean by complexity? Look at Figure 1. Consider how many points da Vinci had to move to transform the circle into the inscribed figure of a man. Ambiguity remains only at the tips of his fingers and the soles of his feet. It seems unlikely that anyone would confuse one of these 2D shapes with the other regardless of how they are slanted relative to the observer. Obviously, complexity also works with the shapes of 3D objects. A book will never produce the same retinal shape as a teapot, regardless of the direction from which they are viewed. This is actually true of most shapes of real objects “out there.” The fact that different 3D shapes almost never produce identical 2D retinal shapes is fundamental for understanding shape constancy. This fact cannot be explained by an analysis of the depths, corresponding to individual points on the retina. The depths of these individual points are always ambiguous, as the Bishop Berkeley was the first to emphasize in his monograph, “The New Theory of Vision” (1709). But, keep in mind that the shape of a 3D object rarely is ambiguous. This fact is what makes the recognition of the 3D shapes of familiar objects possible in the first place. Note that one never speaks about the recognition of the depth or of local surface orientation in the same way one speaks about the recognition of 3D shape. The complexity of shape makes it possible to recognize and recover the 3D shapes of both familiar and unfamiliar objects. Thus, shape is unique. Shape’s complexity allows constancy to be achieved more effectively than the constancy of any other perceptual property, but it can only be effective *after* the shape of the 2D retinal image of a 3D object is established by figure-ground organization. The uniqueness of shape and the fundamental significance of figure-ground organization for the perception of the 3D shapes of objects “out there” have been overlooked by almost everyone since the Gestalt psychologists called attention to figure-ground organization almost 100 years ago (Rubin, 1915; Wertheimer, 1923; Koffka, 1935). Fortunately, there has been a revival of interest in figure-ground organization during the last decade by psychophysicists (for a review see Kimchi et al., 2003), engineers (for a review see Boyer & Sarkar, 2000) and neuroscientists (Lamme & Spekreijse, 1996; Lamme, Zipser & Spekreijse, 1998; Toot, Kalitzin, Spekreijse, Lamme, & Super 2006). Progress has been slow so far, but our understanding of figure-ground organization, as well as of its neural substrate, has been improving steadily. The work of Prof. Spekreijse and his associates on figure-ground organization in primates is clearly a promising approach to developing a biologically-plausible theory of figure-ground organization, a development that may advance our understanding of this fundamental process a lot.

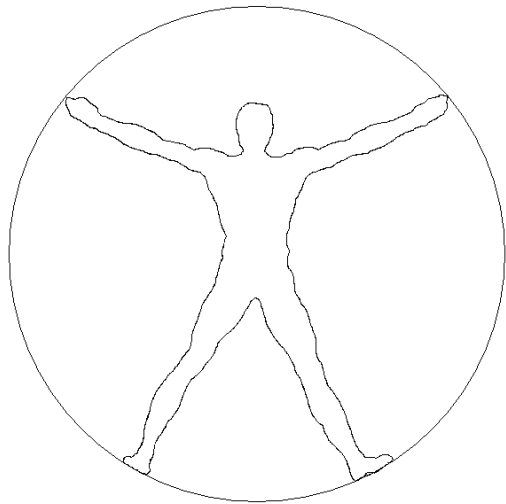


Figure 1. A human silhouette and a circumscribed circle (after Leonardo DaVinci). (from Pizlo, 2008).

The significance of this paper is to show that a *complete* theory of 3D shape perception will be possible as soon as the mechanisms underlying figure-ground organization are understood because we have been able to develop a computational model that recovers the 3D shape of an object “out there” from its single 2D retinal representation. Human beings do this very well. Our model does, too. This paper provides a detailed description of our model, as well as evidence showing that it can recover the 3D shape of complex objects from a wide range of viewing directions. In other words, it shows that the model can achieve a high degree of shape constancy.¹ The model is derived from the traditional Gestalt approach in which the percept of a 3D shape results from the operation of a simplicity principle called the “Law of Prägnanz”. The Gestalt, like our approach, is “nativistic”. It assumes that there are built-in, inherited mechanisms, that abstract the shape of a 3D object from its 2D retinal representation. In this approach, neither learning nor depth perception have a role in the perception of 3D shape. Once one makes these assumptions, as we as well as Gestalt psychologists did, and once one keeps in mind that (i) the human visual system evolved in a 3D environment in which *all* objects have some volume and (ii) most naturally-occurring objects are approximately symmetrical, one expects that a successful functional visual system like ours would both “know” and use these facts to perceive the shapes of objects as they are “out there.” Furthermore, once one adopts this approach, in which the volume of an object and its symmetry need not be *reconstructed* from depth cues, these properties of objects, as well as the shapes of these objects, can be *recovered*² by the application of suitable constraints on a single 2D retinal image. Note that these constraints constitute the new simplicity principle employed in our model. Almost all previous theories of shape perception were built on very different assumptions. They assumed that built-in mechanisms only dealt with the 2D spatial properties of the retinal image, and that the three-dimensionality of objects and the three-dimensionality of the environment itself, must be learned and/or *reconstructed*.

Our model’s capacity to recover 3D shapes is illustrated in two demos attached to this paper. DEMO 1 is introduced in Figure 2, which shows a snapshot of the demo and provides a link to it.

¹ Preliminary reports of this work were presented in Pizlo, Li & Steinman (2006), Li & Pizlo (2007) and Pizlo, Li & Steinman (2008).

² Our use of the term “recovery” here, and throughout this paper, rather than the term “reconstruction”, was done to call attention to a major difference between our approach and most prior approaches, many of which were based on Marr’s (1982) paradigm, in which 3D *surfaces*, not 3D *shapes*, are *reconstructed* from depth cues.

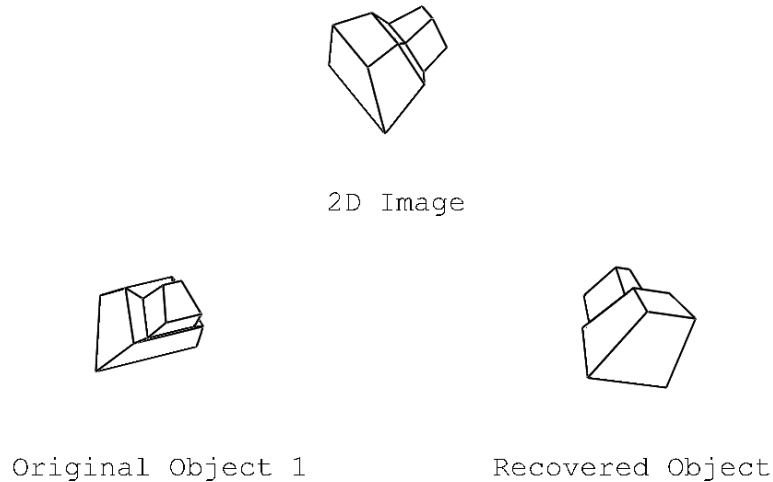


Figure 2. Snapshot of the DEMO 1 (<http://viper.psych.purdue.edu/~pizlo/li-demo2.exe>). A 3D Original Object is shown on bottom left. One of its 2D Images is shown on top and a 3D Recovered Object is shown on bottom right. Use the mouse to rotate the Original or the Recovered Object. The 3D shape of the Recovered Object is almost identical to the shape of the Original Object despite the fact that it is seen from a different viewing direction. This means that the model can achieve shape constancy. (Exit the demo by pressing the ESC key).

Right click the mouse on DEMO 1 to see the contents of Figure 2. A line-drawing of a 3D polyhedral object (Original Object) is shown on the bottom left. Put your cursor on this 3D Object, depress the left button, and move it slightly. This will rotate the Original Object. Rotating the 3D Original Object allows you to view it from different viewing directions. Note that this 3D Original Object is opaque despite the fact that only its contours are shown. The surfaces enclosed by these contours have not been filled in, that is, shown explicitly. In other words, the back part of the object is not visible, its hidden edges have been removed. The 2D drawing on top (2D Image) represents a “retinal” image of this 3D object (an actual retinal image will obviously be smaller than this line drawing, but its 2D shape will be the same). Note that when you look at this 2D Image you actually perceive it as the 3D Original Object shown on the bottom left, viewed from a different direction. This observation is much more important than it appears to be because people rarely realize that the visual system is actually *recovering* a veridical percept of a 3D shape from a single 2D retinal image under these circumstances. Keep in mind that this recovery takes place despite the fact that there are no depth cues, whatsoever, in the display on your screen. Until very recently, no theory could explain this remarkable perceptual achievement. Our model can. Now click the right mouse button. A new, randomly-chosen 2D Image of the same object will appear. It will be obvious that this new 2D image represents the same 3D object out there. This observation is important. It demonstrates that you have achieved shape constancy. You perceived the same shape despite changes in the shape of its 2D Image.

This demo also allows you to visualize how well our model can recover the 3D Original Object’s shape from the 2D Image shown on top. The Recovered Object is shown on the bottom right. Use your mouse to rotate the Recovered Object so you can see it from different viewing directions. This will allow you to compare the model’s recovery of the 3D shape (Recovered Object) to the Original Object’s 3D shape shown on the bottom left. Clearly, the Recovered Object’s 3D shape is almost identical to the Original Object’s 3D shape. Note also that the entire shape was recovered, including the back part of the object that was not visible in the 2D Image used to recover the 3D shape. Hitting the ENTER key rotates both the Original and Recovered Object’s shapes, making your comparison of their shapes easy. Now, right-click in order to see the recovered 3D shape from a different 2D image of the same object. This observation allows you to evaluate how well the model can achieve shape constancy. If constancy were perfect, its 3D shape would not change at all. Observers almost always report that the shape does not change when it is recovered from a different 2D image. In other words, the model’s shape constancy

seems to be close to perfect. Right click a couple of more times to see the recovery for two more 2D images of the same 3D shape. Right-clicking once more will repeat the demo with a new polyhedral object. This demo contains three different 3D Original Objects and each Original Object was recovered from six different 2D Images. You can exit the demo by using the ESC key. Now that you have seen how well the model recovers the 3D shape from a single 2D image, and that it can do this from a variety of viewing directions, how was this done? An overview of the model will be presented first. This will be followed by a description of the model's mathematical and computational details.

The model uses an already "organized" 2D image of the 3D shape for its input. In other words, figure-ground organization is provided to the model because it cannot establish figure-ground organization by itself. Specifically, the model is given information about which: (i) points in the image form edges, (ii) edges and vertices in the image form contours of faces "out there", (iii) edges and vertices represent symmetric edges and vertices "out there", and (iv) edges and vertices define volume "out there". It is necessary to provide this information because the *a priori* constraints that will be applied by our model are *shape* constraints. They are called "symmetry, planarity, maximum compactness and minimum surface." Symmetry refers to the mirror-symmetry of the object. Planarity refers to the planarity of the contours of the object. Compactness is defined as V^2/S^3 where V is the object's volume and S is the object's surface area. Minimum surface is defined as the minimum of the total surface area. Note that depth cues, including binocular disparity, are not used in the recovery. They are not needed. Also note that the symmetry and planarity constraints had been used to recover 3D shape before, but the maximum compactness and minimum surface constraints are entirely new. They have never been used in a shape recovery model. Maximizing compactness is the same as maximizing the volume of an object, while keeping its surface area constant. It is also the same as minimizing surface area, while keeping the object's volume constant. Minimum surface is the same as minimizing the thickness of the object. Basically, our model's recovery of 3D shape is accomplished by choosing a 3D shape that is as compact and, at the same time as thin, as possible, from the infinitely large family of 3D symmetrical shapes with planar contours consistent with the given 2D shape. In other words, our recovery of 3D shape is based on a compromise between maximum compactness and minimum surface. As such, this model belongs to the class of regularization models that solve inverse problems (Poggio et al., 1985).

Mathematical and Computational Details

The application of mirror symmetry and planarity of contours to shape recovery

Let the X-axis of the 3D coordinate system be horizontal and orthogonal to the camera's (or eye's) visual axis, the Y-axis be vertical, and the Z-axis coincide with the visual axis. Let the XY plane be the image. Let the set of all possible 3D shapes consistent with a given 2D orthographic retinal image be expressed as follows:

$$\Theta_I = \{p(O) = I\}, \quad (1)$$

where O and I represent the 3D shape and the 2D image, respectively, and p represents an orthographic projection from the 3D shape to the 2D image.³ There are infinitely many 3D shapes (O) that can produce the same 2D image (I) because translating any point on the surface of a 3D shape along the z axis does not change its 2D orthographic image. Consider a subset of Θ_I , in which all 3D shapes are mirror symmetric and their contours are planar:

$$\Theta_I' = \{O \in \Theta_I : O \text{ is symmetric and its contours are planar}\}. \quad (2)$$

³ In this paper we use orthographic images of 3D symmetrical shapes. When perspective images of symmetrical shapes are used, the recovery problem is more constrained, and thus, easier. Specifically, a single perspective image leads to a unique shape recovery (e.g., Rothwell, 1995). Despite the mathematical uniqueness, constraints will still be needed because recovery is likely to be unstable in the presence of visual noise.

Following Vetter & Poggio (2002), we will show how symmetry can be used to restrict the family of 3D interpretations of a given 2D image, but their restriction will not produce a unique 3D shape. In order to recover a unique 3D shape, additional constraints will be needed. Given a 2D orthographic image P_{real} of a transparent mirror-symmetric 3D shape, and assuming that the correspondences of symmetric points of the 3D shape are known, Vetter & Poggio showed how to compute a virtual image $P_{virtual}$ of the shape:

$$P_{virtual} = D \cdot P_{real}, \quad (3)$$

$$D = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Under this transformation, for any symmetric pair of points $P_{real} = [X_L \ Y_L \ X_R \ Y_R]^T$ in the 2D real (given) image, their corresponding pair of points in the 2D virtual image is $P_{virtual} = [-X_R \ Y_R \ -X_L \ Y_L]^T$. The virtual image is another orthographic image that could be produced by the same 3D shape from another viewing direction. Figure 3 shows an example of a 2D real and virtual image of a symmetric wire (transparent) shape. The virtual image is usually different from the real image. This is not true in degenerate cases, where 2D real image is itself mirror symmetric. For a symmetric 2D image, the 2D virtual and the real images are identical (up to a 2D translation) and Vetter & Poggio's method cannot be applied.

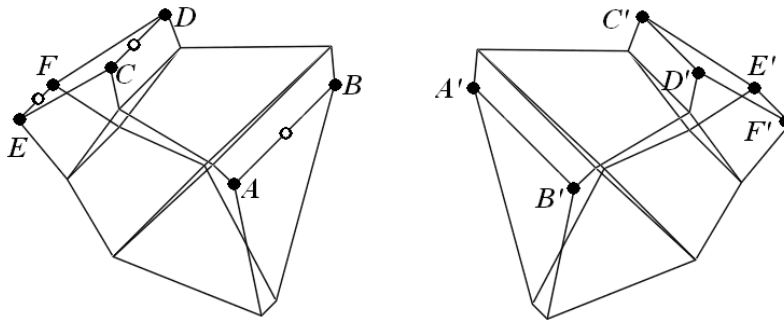


Figure 3. The real (left) and virtual (right) images of a 3D symmetric shape. A, B are images of a symmetric pair of points a, b in the 3D shape. A' and B' are the corresponding points in the virtual image. Note that when the virtual image was produced, A' was obtained (computed) from B. But in the 3D representation, a' is produced after a 3D rigid rotation of a . C, D and E, F are images of other two symmetric pairs of points, c, d and e, f . C' , D' , E' and F' are the corresponding points in the virtual image. The three open dots in the real image are the midpoints of the three pairs A B, C D, and E F that are images of three pairs ab, cd and ef of symmetric points in the 3D shape.

Note that the 2D virtual image is computed directly from the 2D real image, without knowledge of the 3D shape, itself. This means that the original problem of recovering a 3D shape from a single 2D image is transformed into a problem of recovering a 3D shape from two 2D images, real and virtual. Obviously, two images lead to a more restricted family of 3D recovered shapes. This is the main idea behind Vetter & Poggio's method. Next, we explain, how this 3D shape recovery problem is formulated and solved.

The 2D real image can be considered a 2D orthographic image of the 3D shape at its initial position and orientation. The 2D virtual image is a 2D image of the same 3D shape after a particular 3D rigid movement. This movement in 3D space can be expressed as follows:

$$\vec{v}' = R \cdot \vec{v} + \vec{T}. \quad (4)$$

R is a 3x3 rotation matrix and \vec{T} is a 3x1 translation vector. \vec{v}' and \vec{v} are the corresponding points of the 3D shape at two different positions and orientations.

A 3D translation does not affect the shape or size of the 2D image in an orthographic projection. Specifically, translations along the direction orthogonal to the image plane have no effect on the image, and translations parallel to the image plane result in translations of the image. It follows that the 3D translation \vec{T} of the shape can be eliminated by translating the 2D real image or virtual image, or both, so that one pair of the corresponding points in the two images, e.g. A and A' in Figure 3, coincide. Without restricting generality, let G be the origin of the coordinate system on the image plane and the 3D points a and a' whose images are A and A' coincide with G (it follows that both A and A' also coincide with G). Now, the 2D real image can be considered an orthographic projection of the 3D shape at its original orientation and a 2D virtual image can be considered an orthographic projection of the 3D shape after rotation R of the shape around the origin G . This way, the equation (4) takes the simpler form:

$$\vec{v}'_i = R \cdot \vec{v}_i. \quad (5)$$

where $v_i = [X_i, Y_i, Z_i]^T$, and $v'_i = [X'_i, Y'_i, Z'_i]^T$. Equation (5) can be written as follows:

$$\begin{bmatrix} X'_i \\ Y'_i \\ Z'_i \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}. \quad (6)$$

Consider the first two elements of the column vector v'_i :

$$\begin{bmatrix} X'_i \\ Y'_i \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \end{bmatrix} + \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} Z_i. \quad (7)$$

In equation (7), the points $[X_i \ Y_i]^T$ and $[X'_i \ Y'_i]^T$ in 2D real and virtual images are known. Huang and Lee (1989) derived the following relation between $[X_i \ Y_i]^T$, $[X'_i \ Y'_i]^T$ and R :

$$r_{23}X'_i - r_{13}Y'_i + r_{32}X_i - r_{31}Y_i = 0. \quad (8)$$

Let's put the four elements of the rotation matrix R , which appear in equation (8), in a vector $[r_{23} \ r_{13} \ r_{32} \ r_{31}]^T$. The direction of this vector can be computed by applying equation (8) to the three pairs of corresponding points in the 2D real and virtual images (e.g., B, D, F and B', D', F'). The length of this vector can be derived from the constraint that the rotation matrix is orthonormal:

$$r_{13}^2 + r_{23}^2 = r_{31}^2 + r_{32}^2 = 1 - r_{33}^2. \quad (9)$$

Thus, if r_{33} is given, $[r_{23} \ r_{13} \ r_{32} \ r_{31}]^T$ can be computed from two 2D images of three pairs of symmetric points. The remaining elements of the rotation matrix can be computed from the orthonormality of R . It follows that two orthographic images (real and virtual) determine R up to one parameter r_{33} that remains unknown. Note that once the rotation matrix R is known, the 3D shape can be computed using equation (7). This is accomplished by computing the unknown values of the Z coordinate for each image point ($X_i \ Y_i$). Thus, r_{33} completely characterizes the family of 3D symmetric shapes that are consistent with (recovered from) the given image. Usually for each value of r_{33} , two different rotation matrices are produced because if $[r_{23} \ r_{13} \ r_{32} \ r_{31}]^T$ is the solution, $[-r_{23} \ -r_{13} \ -r_{32} \ -r_{31}]^T$ is also a solution. Consequently, two 3D shapes are recovered for each value of r_{33} , and these two shapes are related to one another by depth-reversal.

To summarize, the one-parameter family of 3D symmetric shapes can be determined from four points (A, B, D and F) in the 2D real image and the corresponding four points (A', B', D' and F') in the 2D virtual image. Recall that the virtual points A', B', D' and F' have been computed from the real points $B,$

A, C and E. It follows that the recovery is based on six points A, B, C, D, E and F in the real image that were produced by three pairs of symmetric points a,b,c,d and e,f in the 3D shape. One real and its corresponding virtual point (here A and A') are used to undo the 2D translation. The other three real points (B,D,F) and their corresponding virtual points (B',D',F') are used to compute the rotation matrix (R). Note that the six points a, b, c, d, e and f cannot be coplanar in the 3D shape. To guarantee that these six points forming three pairs of symmetric points are not coplanar in 3D, we only need to verify that the midpoints ($u_1 u_2 u_3$) of the orthographic images of these three pairs of points (the midpoints are marked in blue in the real image in Figure 3) are not collinear:

$$\|(u_1 - u_2) \times (u_1 - u_3)\| \neq 0. \quad (10)$$

In some cases, these three symmetric pairs are not coplanar in 3D, but their midpoints in the image are collinear. This happens when the viewing direction is parallel to the plane of symmetry of the 3D shape. In such a case, the 3D shape is symmetric with respect to the YZ plane, and its 2D image is, itself, symmetric. When this happens, all midpoints of the images of symmetric pairs of points are on the y axis. As a result, the real image and virtual image are identical and the 3D shape cannot be recovered. So, the fact that midpoints in the real and virtual images are not collinear implies that the 3D midpoints are not coplanar and the viewing direction is not parallel to the plane of symmetry of the 3D shape. Note that there is another degenerate case that precludes recovery. This occurs when the viewing direction is orthogonal to the plane of symmetry of the 3D shape. In this case, each pair of 3D symmetric points projects to one 2D point and there is not enough information in the image to perform 3D recovery. Specifically, both r_{13} and r_{23} are zero, and the Z-coordinates in equation (7) cannot be computed.

We will show how Vetter & Poggio's method can be generalized to the shapes of opaque objects before we discuss ways of determining the value of r_{33} . This will be done in the case of polyhedra. Shapes of opaque objects are more difficult to recover because images of such objects provide less information. In extreme cases, information about some parts of a 3D shape may be completely missing from the 2D image, which implies (trivially) that the 3D shape cannot be fully recovered. We restrict discussion to those 2D retinal images that allow full recovery of the 3D shape of an opaque object. How this was done is described next.

As shown above, in order to compute the rotation matrix R , at least three pairs of symmetric vertices of a polyhedron must be visible. Once R is computed, all symmetric pairs whose vertices are both visible can be recovered from Equation (7), e.g. the 3D vertices g, h, m, n and p, q in Figure 4. These two steps are identical to those described above for transparent objects. In the case of the image in Figure 2, there are a total of six pairs of such vertices (the open circles in Figure 4). Recovery fails if both symmetric vertices are invisible. The reason for the failure is that if both $[X_i Y_i]^T$ and $[X'_i Y'_i]^T$ are unknown, Z_i cannot be computed. For pairs of symmetric vertices with one vertex visible and the other occluded, for example, the symmetric pair u and w in Figure 4, a planarity constraint can be applied. In this case, symmetry in conjunction with planarity of the contours of faces is sufficient to compute the coordinates of both of these vertices. For example, the planarity of the face $gmpu$ implies that u is on the plane (s) determined by g, m and p . The vertex u is recovered as an intersection of the face s and the line $[u_x u_y 0]^T + \lambda[0 0 1]$. The hidden counterpart w of u is recovered by reflecting (u) with respect to the symmetry plane of the 3D shape. The symmetry plane is determined by the midpoints of the three recovered pairs. Figure 4 shows a real and a virtual image of an opaque polyhedron that can be recovered completely, that is both the visible front part and the invisible back part can be recovered. On average, about 60% of the 2D images allowed a full recovery of the 3D shapes with the randomly-generated polyhedra we used and with randomly-generated 3D viewing orientations. Interestingly, once the recovery of an opaque object is possible, the recovery is unique for a given value of r_{33} : the depth-reversed version of the 3D shape is excluded by the constraint that the invisible vertex must be behind its visible symmetric counterpart. Recall that for transparent (wire) shapes, there are always two 3D shapes related to one another by depth reversal. So, paradoxically, opaque shapes, which provide less information in the image, are less ambiguous.

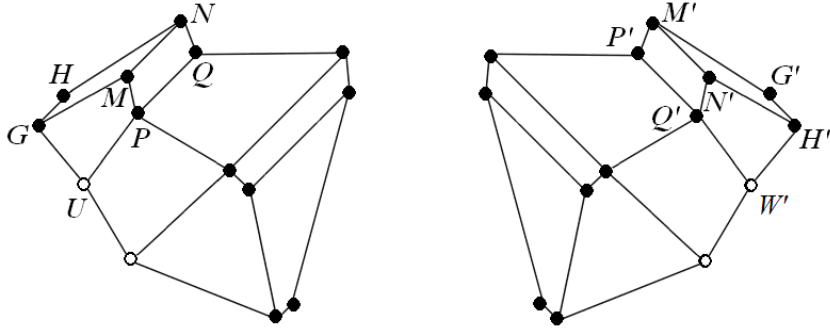


Figure 4. A real (left) and a virtual (right) image of a 3D symmetric opaque polyhedron. Points G, H, M, N, P, Q and U are images of the 3D vertices g, h, m, n, p, q and u , respectively. The symmetric pairs gh, mn, pq can be reconstructed from equation (7) once the rotation matrix R is known since both points of these pairs are visible. There are six pairs of such vertices. These pairs are marked by solid dots. The vertex u , which resides on the plane determined by vertices g, m and p , is reconstructed from the planarity constraint. The invisible symmetric counterpart w of vertex u is obtained by reflecting u with respect to the symmetry plane. There are two such vertices, whose reconstruction used both symmetry and planarity constraint. These vertices are marked by open dots.

Up to this point, we described how the one-parameter family Θ_I' of 3D shapes is determined. This family is characterized by r_{33} . For each value of r_{33} , one, or at most two, shapes are recovered. All 3D shapes from this family project to the same 2D image (the real image). All of them are symmetric and the contours are planar. Because r_{33} is an element of a rotation matrix, it is bounded:

$$\Theta_I' = \{O = g_I(r_{33}) : -1 \leq r_{33} \leq 1\}. \quad (11)$$

Next, we describe two shape constraints, called “maximum compactness” and “minimum surface” that are used to determine the value of the unknown parameter r_{33} . These constraints are new; they have never been used to model 3D shape recovery.

The application of the maximum compactness constraint

A 3D compactness C of shape O is defined as follows:

$$C(O) = \frac{V(O)^2}{S(O)^3}, \quad (12)$$

where $V(O)$ and $S(O)$ are the volume and surface area of the shape O , respectively. Note that compactness is unit-free, and, thus independent of the size of O . Its value depends only on shape. Applying the maximum compactness constraint recovers a unique 3D shape. Specifically, selecting the maximally compact 3D shape from the one-parameter family of 3D shapes recovered by the method based on Vetter and Poggio’s (2002) algorithm, leads to a unique 3D shape. Note that we do not have a proof of our claim that the result of the recovery is always unique. But, the result was always unique in our simulations with several thousands of 3D shapes.

Maximizing $C(O)$ corresponds to maximizing the volume of O for a given surface area, or minimizing surface area of O for a given volume. Compactness defined in equation (12) is a 3D version of the 2D compactness constraint used in the past for the reconstruction of surfaces (e.g. Brady & Yuille, 1983). The 2D compactness of a closed contour is defined as a ratio of the surface’s area enclosed by the contour to the perimeter, squared. The circle has maximal compactness in the family of 2D shapes. The sphere has maximal compactness in the family of 3D shapes. Recall that the Gestalt psychologists considered the circle and the sphere to be the simplest, and therefore, the “best” shapes (Koffka, 1935).

They were the simplest because they were the most symmetric of all shapes. The relation between symmetry and compactness was established formally by the Steiner symmetrization operation (Polya & Szego, 1951).

Note that maximum 3D compactness is a generalization of the minimum variance of angles constraint used previously to recover the shapes of polyhedra (Marill, 1991; Sinha, 1995; Leclerc & Fischler, 1992; Chan et al., 2006). The maximum compactness constraint, like the minimum variance of angles constraint, “gives” the 3D object its volume. The minimum variance of angles constraint is very limited, it only applies to polyhedra. The maximum compactness is much less confined. It can be applied to almost *any* 3D shape.

The application of the minimum surface constraint

This constraint is quite straightforward. It chooses the 3D object whose total surface area $S(O)$ is minimal. In other words, the model maximizes the expression $1/S(O)$. If there were no other constraint, the resulting 3D object would be flat, it would have no volume. Recall, however, that this constraint will always be applied to objects that have some volume. This means that the minimum surface constraint will produce the thinnest possible object, the object with the smallest range in depth. We already know that maximizing compactness is useful. Why is making an object as thin as possible, less than maximally compact, useful? It is useful because it will allow the veridical recovery of shapes, they way they are “out there.” Said technically, recovering a 3D shape, which has the smallest range in depth, is desirable because it minimizes the sensitivity of the 2D image to rotations of the 3D shape. This makes the 3D recovered shape most likely. Combining a maximum compactness with a minimum surface constraint will lead to the best recovery of 3D shapes.

How should these two constraints be combined? Several combination rules were tried, and the following seems to be optimal:

$$V(O)/S(O)^3 \quad (13)$$

That is, our model recovers the 3D shape that maximizes the ratio defined in eq. (13). Note that this ratio is the geometrical mean of V^2/S^3 (compactness) and $1/S^3$ (minimum surface).

Robustness in the presence of image noise

The model (described above) assumes that the retinal (or camera) image has no noise, but real images always contain some noise. How can such image-noise be handled? This becomes an important question as soon as one wants the model to recover the 3D shapes of real objects in real environments from their 2D retinal images. Noise can be handled at three different stages of the model. First, it can be verified whether pairs of symmetric points form a set of parallel line segments in the image. In the absence of noise, they must be parallel because the parallelism of these lines is invariant in an orthographic projection (Sawada & Pizlo, 2008). If they are not parallel because of noise and/or because of uncertainty in the figure-ground organization, their positions can be changed to make these line segments parallel. Obviously there will always be some ambiguity about how this change should be made but it should be possible to do so. For example, a natural constraint that can remove such ambiguity is to minimize the sum of squared distances that represent the change of the positions of the points. In other words, the points should be moved as little as possible to satisfy the parallelism constraint. An alternative way to make the line segments connecting pairs of symmetric points parallel is to apply a least-squares approximation at the stage the one-parameter family of 3D symmetrical shapes is produced. Note that a least-squares correction that makes the line segments parallel will not ensure the planarity of the faces of the 3D polyhedron. Planarity can be restored at the very end of the recovery by changing the depths of individual points. We performed preliminary tests of these three methods for correcting noise with synthetic images and found that our 3D shape recovery model was quite robust.

Testing the model

This section describes a simulation experiment that tested the model's capacity to recover 3D shape from a single randomly-chosen 2D image. Current common sense holds that no existing machine vision system can "see" shapes as well as humans (Pizlo, 2008). Furthermore, most published studies of human shape perception concluded either that humans do not achieve shape constancy, or that their shape constancy is far from perfect. It follows from these current commonly-held claims that a computational model of 3D shape recovery either would not demonstrate shape constancy, or that shape constancy would be poor if it were manifested at all. The reader, who has used the demo, already knows that neither claim can be true. The next section describes a formal evaluation of the model that confirms the reader's compelling, informal, subjective observations.

Stimuli, 2D images (line drawings) of randomly-generated 3D abstract shapes like those shown in Figure 4 were used. Abstract shapes, rather than shapes of common objects, like chairs, couches or animal bodies, were used to make it possible to compare our model's performance with the performance of human observers. Human observers must be tested with abstract shapes to avoid familiarity confounds (Pizlo & Stevenson, 1999; Chan et al., 2006). Obviously, our model, which has no provision for "learning", is not subject to this problem. For the model all stimuli are novel, including those familiar to humans. Common objects could be used with the model, but this would make it impossible to compare human's and the model's performance. The shapes were 2D orthographic images of opaque 3D symmetric polyhedra (hidden edges were removed). Only images allowing complete recovery of each 3D opaque polyhedron, were used. Sixty percent of the 2D images, produced from randomly-generated viewing directions, satisfied this requirement.

Every polyhedron had 16 vertices. Their positions were randomly-generated in 3D space with the following constraints: (i) the object had planar faces, (ii) it had one plane of symmetry, (iii) the "front" part of the object was a box smaller than the box in the "back", and (iv) these boxes had a pair of coplanar faces. The simulation used 100 randomly-generated polyhedra, whose aspect ratios varied between 1/5 and 5. For each polyhedron a randomly-chosen viewing orientation was used and its orthographic image was computed. Viewing orientation was random subject to one constraint, namely the slant of the plane of symmetry of the 3D object had one of the following five values: 15, 30, 45, 60 and 75 deg. Each slant was used 20 times for a total of 100 images. The value of slant was controlled to allow the model's shape constancy to be evaluated.

Analysis A quantitative measure of 3D shape was needed to compare the recovered 3D shape with the original 3D shape. In order to derive this measure, we first needed to establish the number of parameters that were required to characterize both the original and recovered shapes. The shape of each original polyhedron was determined by 16 vertices, each vertex having three coordinates. Only half of the vertices were needed because the polyhedron was mirror-symmetric. This leads to 24 parameters (8 x 3). The other half of the object required three parameters to specify the symmetry plane. But, since 3D position, orientation and size do not affect 3D shape, the 3D shape of the original polyhedron was characterized by only 20 parameters (24+3-7). The actual number of parameters for all original polyhedra was smaller (15) because of the planarity constraint. Now, consider the 3D shape of the recovered polyhedron. This polyhedron was also characterized by 15 parameters because it had the same overall 3D structure. Recall that the recovered 3D shape was obtained from a 2D image that was produced by the original 3D shape. It follows that the original and recovered shapes differ with respect to only one parameter, r_{33} . Thus, the 3D shapes, representing the original polyhedron and the recovered polyhedron, can be compared simply. Only one parameter, r_{33} , is needed. But note that this parameter is not ideal because it is abstract; it is an element of a 3D matrix used for computations in the model. Unfortunately, no intuitive interpretation of this parameter is available, one that would refer directly to the 3D shape perceived. Fortunately, there is a perceptually-relevant parameter that can be used in place of r_{33} , namely, one of the three aspect ratios of the polyhedron. Specifically the ratio of its thickness measured in two orthogonal directions. The

“thickness” of a shape along the direction \vec{n} is defined as the maximum difference among all vertices along the direction \vec{n} :

$$T_I^{\vec{n}}(O) = \max(\vec{v}_i \cdot \vec{n}) - \min(\vec{v}_i \cdot \vec{n}) \quad i = 1, 2, \dots, n,$$

where \vec{v}_i is a 3D vertex and n is the number of vertices. The aspect ratio $Q_I(O)$ is defined as the ratio of thicknesses along two directions: one is parallel to the normal of the symmetry plane \vec{n}_s , and the other is parallel to the normal of the base face \vec{n}_b (see Figure 5).

$$Q_I(O) = \frac{T_I^{\vec{n}_s}(O)}{T_I^{\vec{n}_b}(O)}. \quad (14)$$

This ratio specifies the 3D shapes of our polyhedra uniquely.

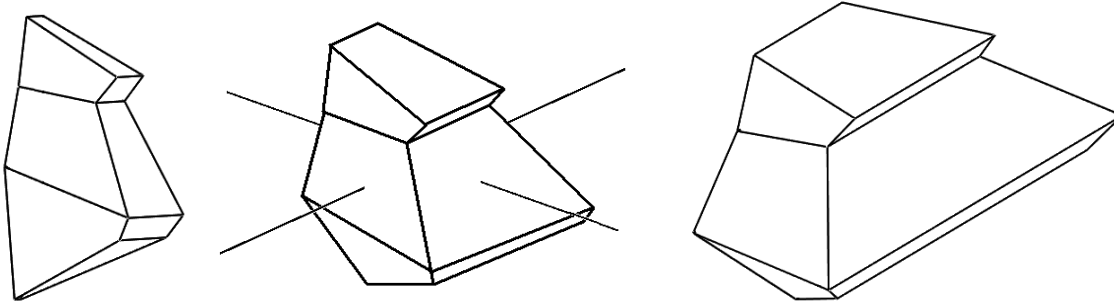


Figure 5. An illustration of how two directions were used to compute a shape’s aspect ratio. The aspect ratio for each shape (from left to right) is: 1/3, 1 and 3.

Once we know how to measure the original and the recovered 3D shapes, we need a way to compare them. We did this by defining their similarity. More exactly, the following formula measures the “dissimilarity” between shapes O_1 and O_2 :

$$L_I(O_1, O_2) = \begin{cases} \frac{Q_I(O_1)}{Q_I(O_2)} & \text{if } Q_I(O_1) > Q_I(O_2) \\ 1 & \text{if } Q_I(O_1) = Q_I(O_2) \\ \frac{Q_I(O_2)}{Q_I(O_1)} & \text{if } Q_I(O_1) < Q_I(O_2) \end{cases} \quad O_1, O_2 \in \Theta_I'. \quad (15)$$

The dissimilarity is simply a ratio of aspect ratios of two shapes, computed in such a way that the result is never less than one. So, if an aspect ratio of the first shape is 2 and that of the second is 1/2, their dissimilarity is 4. The dissimilarity is also 4 when the aspect ratio of the first shape is 1/2 and that of the second shape is 2. When $L_I(O_1, O_2)$ is equal to one, the two shapes O_1 and O_2 are identical.

Results Figure 6 shows a scatter plot of the relation between the aspect ratio of the original 3D shape and the aspect ratio recovered by our model. Different colors represent different values of slant. Two facts stand out in this graph. First, the data points representing individual slant-values form a set of approximately straight lines. This means that there was a high correlation between the recovered and original aspect ratio for the individual slant-values. The correlation coefficients range between 0.92 and 0.97. Second, these straight lines do not coincide. They are shifted relative to one another and stay approximately parallel to the diagonal line. In particular, the data points for slants 30, 45 and 60 deg, are close to the diagonal line, the line representing veridical recovery of the aspect ratio. Note however, that

the data points for the extreme slant-values, 15 and 75 deg, are farther from the diagonal line indicating that there were systematic errors in the recovered aspect ratio. When these extreme slant-values are included, the overall correlation coefficient of the recovered and original aspect ratios is much lower, namely: 0.61.

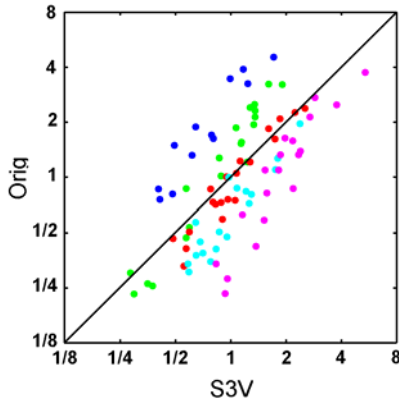


Figure 6. The aspect ratio of the original 3D shape is shown on the ordinate, and the aspect ratio recovered by the model is shown on the abscissa. Colors represent slants of the plane of symmetry: blue 15 deg, green, 30 deg, red 45 deg, cyan 60 deg and pink 75 deg.

The effect of the slant of the plane of symmetry on the systematic error of the recovered aspect ratio is illustrated more clearly in Figure 7. The ordinate shows “dissimilarity” between the recovered and original shapes as defined in equation (15). Recall that dissimilarity equal to one means that the recovered aspect ratio is equal to the original aspect ratio, and that dissimilarity equal to two means that the recovered and original aspect ratios are different by a factor of two. The data points represent individual shapes (there were 20 points for each slant). The continuous line represents the median dissimilarity. The errors were usually small for slant-values 30, 45 and 60 deg. For the extreme slants, 15 and 75 deg, the errors tended to be larger. However, the overall median dissimilarity across all slants was 1.4, which means that for half of the recovered shapes, the errors in the aspect ratio were not larger than 40%. Large errors occur when the 3D object looks like a long rod with its long axis close to the viewing axis. In such cases, the 2D image is compact, and, as a result, the recovered 3D shape is less elongated than the original shape. The same was true when a human observer, rather than the model, recovered the 3D shapes (see below). Note that the model *only* made errors in *one* of the 15 parameters that characterize the 3D shape. This allows us to say that the 3D shape recovered by the model is always quite accurate even when there are errors in the aspect ratios recovered.

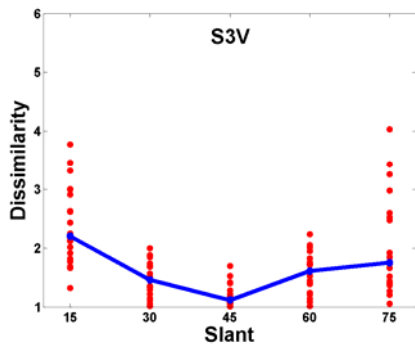


Figure 7. The effect of slant of the symmetry plane on the error in the aspect ratio recovered by the model.

Why does the model make any errors at all when it recovers 3D shapes? The answer is simple. A single 2D orthographic image of a 3D shape with a single plane of symmetry can never be sufficient for completely error-free recovery of this shape. This is why our model needed more than two, already known, useful constraints, namely, symmetry and planarity. Our additional constraints, maximum compactness and minimum surface area achieved a great deal. They made it possible to recover the 3D

shapes quite well. The more interesting question is why is the model performing so well? Apparently, 3D maximum compactness and minimum surface capture some critical aspect of the relation between 3D shapes and their 2D perspective images. Namely, compact 3D shapes, never produce non-compact 2D perspective images. For example, a cube always produces a compact 2D perspective image regardless of the viewing direction.⁴ Conversely, non-compact 3D shapes rarely produce compact 2D perspective images. For example, a long rod “out there” rarely produces very short rod in the 2D image. The important point to remember is that the recovery of the aspect ratio is quite accurate for a wide range of compactnesses and for a wide range of viewing directions. Even more important, the model’s recovery of the 3D shape, itself, was very accurate: It made errors with only one of the 15 parameters used to characterize the 3D shape!⁵

Human observers’ ability to adjust aspect ratios

Our model was tested by examining its recovery of the aspect ratios and it was found to do so very well. The question now arose whether the human observers can do this too.

Method

Two authors (*YL and ZP*) and two naïve observers participated in this experiment. All of them had normal or corrected-to-normal vision. Each observer was tested with the same 100 images that were used to test the model. The stimulus subtended 5.7 deg (5x5 cm at the 50 cm viewing distance).

The room was dark and the head was supported by a chin-forehead rest. The observer viewed the stimulus with the preferred eye. In each trial, a 2D orthographic image of a randomly-generated symmetrical polyhedron was shown for 5s near the top of a computer screen. This 2D image led to the percept of a 3D shape. The observer was asked to remember this 3D shape. Then, the 2D image disappeared and a rotating 3D polyhedron was shown in the middle of the screen. This rotating polyhedron was selected from the set of symmetrical 3D polyhedral shapes with planar contours generated by our model from the 2D image that was shown first. In other words, the 2D orthographic image was the image of the 3D rotating polyhedron. This 3D polyhedron was rotating in a random direction at about 90 degrees/second. This allowed many different views of the 3D shape to be seen in a short amount of time. The observer could use a mouse to adjust the aspect ratio of the 3D shape to make it match the percept of the 3D shape produced by the stationary 2D image shown at the beginning of the

⁴ This statement is true in the case of a perspective projection to a spherical retina, like the retina in the human eye. When the “retina” is planar, like those in conventional cameras, this statement is true when the image is only projected to the center of the retina.

⁵ It is natural to talk about shape constancy and shape veridicality when the model or an observer actually looks at a 3D shape. The 3D shape is called the “distal stimulus” and it can be compared to the 3D shape recovered (perceived) from a 2D image. But, the situation is quite different when a single 2D image, shown on a computer screen, is presented to the model or an observer, the situation studied in this paper. Which is the distal stimulus under our conditions, the 2D image, itself, or the 3D shape that was used to produce this 2D image? This question arises because the 3D shape does not exist physically in our experimental conditions. It only exists in the memory of the computer that generated a particular 3D shape and then proceeded to compute its 2D image. The 2D image presented to an observer determines an infinitely-large family of 3D interpretations. Since this is the case how is it possible to evaluate its veridicality by choosing one particular 3D shape from this infinitely-large family of 3D interpretations (this question was raised by Koenderink et al., 2006)? This question can be answered as follows: if it is meaningful to talk about the veridicality of shape when a monocular stationary observer views a real, stationary 3D shape, it must be meaningful to talk about veridicality in our case, when the observer is presented with a 2D image on a computer screen. These two cases are indistinguishable because in both, the observer is presented with exactly the same 2D information. If the 3D shape percept agrees with the 3D (non-physical) shape that was used to produce the 2D image, and if the percept is constant across a number of different viewing orientations, as was illustrated in DEMO 1, shape constancy and shape veridicality are achieved and are meaningful.

trial. Each trial began with the aspect ratio set to a random value. There was no time limit for the adjustment.

Results

Figure 8 shows a scatter plot illustrating the relation between the aspect ratio of the original 3D shape and the aspect ratio recovered by one naïve observer (results of the other three observers were very similar). This scatter plot is quite similar to the scatter plot of the model shown in Figure 6. Specifically, for each individual slant value there was a high correlation between the original aspect ratio and the aspect ratio recovered by the subject (the correlations ranged between 0.70 and 0.79). Second, there is a systematic effect of slant on the recovered aspect ratio. These two observations suggest that the observer's percept would be similar to the model's if the model could "perceive". The scatter plots of all four subjects are shown in Figure 9 to make it possible to evaluate the relation between the aspect ratio recovered by the model and by the observers more directly. These graphs show a strong relation between the model's and the observer's recovery: the correlations shown in these four scatter plots range between 0.76 and 0.87. The correlations between the model's and observer's aspect ratios are very similar to the correlations between the aspect ratios recovered by any two of the observers (these inter-subject correlations range between 0.74 and 0.88). This means that the model can account for an observer's results as well as one observer can account for the results of another observer. In other words, the model can "explain" the observer's percept quite well and it can do this not only when percepts are veridical, but also when the percept was very different from the aspect ratio of the original shape. Large differences between the aspect ratios recovered by the model and by the observers were very rare. They almost never differed by more than a factor of two and the median difference between the model and the observer was equal to a factor of about 1.25 (i.e., 25% difference in the recovered aspect ratio). To the authors' knowledge these results are the very first demonstration in which a computational model performed as well as a human observer in a 3D shape perception task.

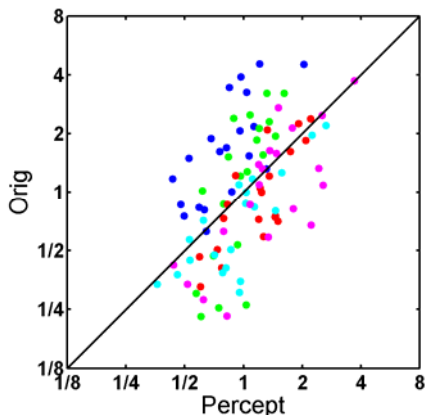


Figure 8. The aspect ratio of the original 3D shape is shown on the ordinate, and the aspect ratio recovered by a naïve subject (ED) is shown on the abscissa. The same symbols are used as in Figure 6.

Finally, we evaluated the effect of slant on the errors in the recovered aspect ratio (see Figure 10). These relations were similar to one another, which means that the recovery produced by all four observers, authors and the naïve observers, were similar. Familiarity with the stimuli and with existing theories of shape perception did not affect their performance at all. Note that the relations shown in Figure 10 are somewhat different from those shown in Figure 7. Specifically, the observers' median dissimilarity for slants 60 and 75 deg was smaller than the model's. We did simulations which showed that the model's performance with these two slants can match the performance of the observers when maximum compactness (eq. 12) is used to recover the 3D shape. This is important because our results suggest that the observers used one constraint, namely the maximum of V/S^3 (eq. 13), for slants 15, 30 and 45 deg, and another constraint, namely the maximum of V^2/S^3 (eq. 12), for slants 60 and 75 deg. In other words,

the human visual system “switched” from using one constraint to the other, depending on the slant of the symmetry plane of the 3D shape. In our experiment, making such a switch would require detecting cases in which the slant of the symmetry plane was close to 90 deg. It is not difficult to imagine how such a detection might be done despite that fact that we have not yet developed a formal model that makes such a detection. This detection would require nothing more than detecting whether the 2D image of a 3D symmetrical shape is nearly symmetrical.

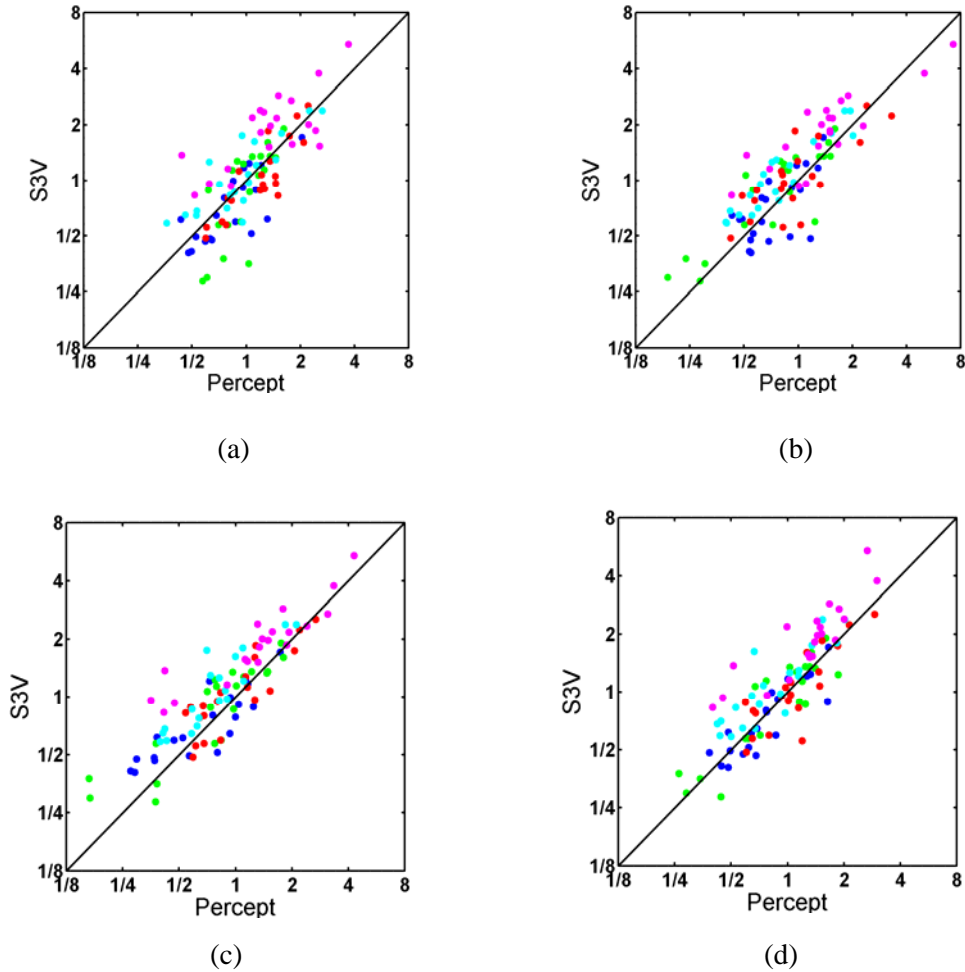


Figure 9. The aspect ratio of the 3D shape recovered by the model is shown on the ordinate, and the aspect ratio recovered by the subject is shown on the abscissa: (a) ED, (b) TJ, (c) YL, and (d) ZP. The same symbols are used as in Figure 6.

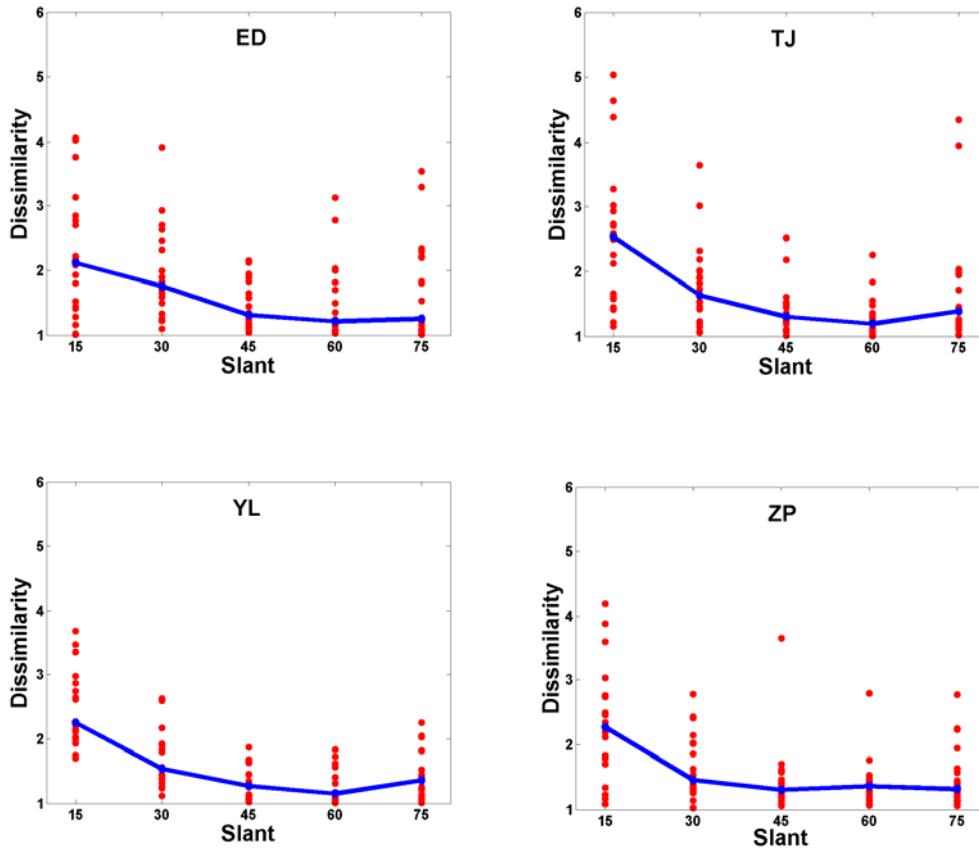


Figure 10. The effect of slant of the symmetry plane on the error in the aspect ratio recovered by the subject.

Concluding remarks

Shape constancy with real images

Now that we have shown that our model can recover 3D shapes from 2D synthetic images of abstract and unfamiliar 3D shapes, we can ask whether the model can recover 3D shapes equally well from real 2D images of real objects in natural environments. We think the answer will probably be yes simply because most objects “out there” are at least approximately symmetrical. Asymmetrical objects, with no planar contours and without clearly defined volume are very rare. Common objects such as animals, buildings, and sofas are almost always symmetrical and they almost always have contours that are approximately planar. They also almost always have surfaces enclosing volume. Chairs and tables have little volume that actually “belongs” to these objects, but it is easy to “see” the surfaces and volume defined by the legs of chairs and tables. The only objects that do not have any of the three properties that have been used in prior studies on shape perception are crumpled pieces of papers and 3D wire objects. Not surprisingly, shape constancy is difficult, if possible at all, to achieve with these objects. They do not have properties that can be processed by our model that uses symmetry, planarity and maximum compactness constraints. They also cannot be processed well, if at all, by human observers.

At this point the reader can get an idea about how well our model can recover “realistic” natural objects by running DEMO 2 which can be accessed from DEMO 1. Right click the mouse to start DEMO 2 and an image of a chair will be visible on the lower left. You can rotate this chair the same way you rotated the 3D polyhedra in DEMO 1, namely, by putting your cursor on the chair, left clicking and moving your mouse. A 2D image of the chair is shown at the center top and a dotted version of the Recovered chair is shown on the bottom right. You can examine the recovered shape by using the mouse

and/or by hitting the ENTER key as you did in DEMO 1. Note that this chair has little volume. Recovering the 3D shape of such a chair by maximizing compactness and minimizing the surface area required using the volume and surface area of the bottom part of the chair that was defined by its four legs (using the convex hull of the entire chair would lead to a very similar recovery). Note that with natural objects, like this, their contours are rarely given explicitly. It follows that with many natural objects the planarity constraint used in our present model would have to be replaced by a constraint of *approximate* planarity and/or a smoothness of surfaces. For the chair recovered in DEMO 2, only the symmetry, maximum compactness and minimum surface constraints were used and the recovery was performed from a 2D image of a dotted 3D chair, circumventing the need for information about the chair's contours.⁶ Right-click the mouse to recover the 3D chair from different 2D views. This demo shows that our model can recover the 3D shape of this chair from its 2D images and it also shows that the model can achieve a high degree of shape constancy.

Once we know that the model can recover 3D shapes of real objects from real 2D images, we can ask whether the model can achieve shape constancy with real images of real objects in natural environments. We think that answer will probably be yes simply because shapes of real objects tend to be very different from one another. Small, or even moderate, errors in the 3D shapes recovered are unlikely to permit confusion of one real object with another. A horse will never look like a couch, or a conch, or a coach from any viewing direction. Furthermore, we already know that humans achieve shape constancy readily with abstract unfamiliar objects. We do not know, as this is written, how well they do this with real images of real objects in natural environments, nor do we know how our model will fare under such conditions. Finding out would surely be worthwhile.

Shape constraints vs. canonical views

Our results showed that views close to what are called “degenerate” or “nearly degenerate” views, lead to 3D recovered shapes whose aspect ratio might be very different from the aspect ratio of the 3D original shape that produced the 2D image. This fact, and what Palmer calls “canonical views” deal with an important issue in shape perception (Palmer et al., 1981). Namely, it recognizes that not all views of a 3D shape are equally informative. Palmer introduced the concept of canonical views (or canonical perspective) assuming that 3D shape perception is based on 2D representations of 3D shapes. The 2D representations (views) are memorized by the observer and subsequently used to recognize 3D shapes. This kind of theory assumes that there are only few representative 2D views that can be used to establish the shape through learning.⁷ It also assumes that one perspective image called “canonical” is prototypical. It is the best representative of the object's shape. These assumptions are incorporated into theories like this because the observer will not be required to memorize very many 2D views before he can recognize the 3D object on the basis of a remembered canonical shape. Our model is completely different in that the perception of a 3D shape reflects the operation of shape constraints that recover the 3D shape from one of its 2D retinal images. These constraints are built-in to the organism. There is no need to store canonical views or to learn anything. Most views of most 3D objects provide enough information to recover the object's 3D shape from its 2D retinal image. We do not believe that the concept of canonical views is either useful or necessary, but we cannot reject the idea completely before we find out how well our model works with real images of real objects in natural environments. Resolving this issue provides an additional reason for finding out how well our model does with “realistic” images of “realistic” objects.

⁶ A model developed to handle such natural objects will only need to “stretch” 3D surfaces along feature points after the 3D shape is recovered from these points. Note that these surfaces will come *after*, not before, the 3D shape is recovered.

⁷ This type of theoretical approach goes back at least to Helmholtz (see Pizlo, 2008, for a detailed treatment of the history of this and other approaches to shape perception).

The role of depth cues in shape perception

What, if any, is the role of cues to depth and surface orientation, such as binocular disparity, motion, shading or texture, in the recovery of 3D shape and shape constancy? Note that when we speak of “3D shape”, we are referring to the spatially global properties of objects. Depth cues are quite different. They are spatially local in the sense that they provide information about the distance of a given point or feature from the observer or about the local orientation and curvature of the surface of an object. So, if depth cues are to be used in shape perception, they will have to provide multiple perceptual measurements at many points on the 3D object, and the measurements from these multiple points must then be integrated if they are going to be used to perceive a 3D shape. How good could such an integration be? In the absence of constraints (priors) for the relations among the points across spatially separated parts of the 3D object, the individual perceptual measurements are likely to be statistically independent. It follows that the integration of depths and surface orientations across multiple points of the object is likely to be less reliable than the percept of depth or surface orientation at a single point of the object. It is known that the percept of 3D distances, angles and aspect ratios are quite unreliable (difference thresholds are large) and subject to large systematic errors (e.g., Johnston, 1991; Pizlo & Salach-Golyska, 1995; Norman et al., 1996; Todd & Norman, 2003). It follows that depth cues, alone, cannot lead to reliable percept of a 3D shape. For example, if an observer tried to recover 3D symmetry from depth cues, the resulting percept will be quite unreliable, and therefore, not actually symmetric, unless the symmetry were used as an a priori constraint. But if symmetry is used as such a constraint, depth cues are superfluous. They are not needed! The only place where depth cues could be of some use is in the correction of the recovered aspect ratio of a 3D shape. Recall that both the model and the observers sometimes made large errors in recovering an aspect ratio. This happened when the object was highly non-compact and the viewing direction was nearly parallel to the long axis of the object (an example of such an object is shown in Figure 11). In such cases, employing the maximal compactness and minimal surface area constraints will lead to a 3D shape that has substantially less depth than the original shape. This kind of error could probably be corrected by using binocular disparity or motion parallax. We have made some informal observations that confirm this prediction. Specifically, binocular viewing of a 3D shape like the one shown in Figure 11, when the viewing axis is parallel to the long axis of this shape, leads to more a veridical percept than monocular viewing.

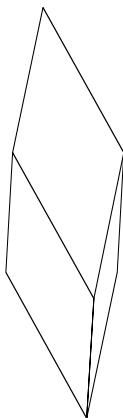


Figure 11. When the visual axis is close to the long axis of this 3D object, the aspect ratio (depth) of 3D shape recovered from a single 2D image will have substantially smaller than the actual aspect ratio.

Comparison with prior models that made use of shape constraints

Our model of shape recovery was motivated, in part, by prior computational models based on shape constraints. This approach started with Biederman's work (1987) in which 3D shape perception was based on volumetric parts (geons) defined by symmetry constraints. This approach was picked up by a number of students of computer graphics and machine vision (Marill, 1991; Leclerc & Fishler, 1992; Sinha, 1995). Surprisingly, they, unlike Biederman, did not use symmetry constraints. They used minimum variance of angles, planarity of contours, and minimum slant constraints instead of symmetry. They were able to recover the shapes of 3D polyhedra that resembled the shapes perceived by human observers with these constraints. Chan, Stevenson, Li & Pizlo (2006) continued this line of research and reported the first psychophysical evaluation of a shape recovery model based on these as well as some other constraints. They reported that shape constancy can be achieved reliably with shapes other than geons. They also reported that depth cues were neither necessary nor sufficient to achieve shape constancy. Finally, they reported that shape constancy depended critically on the operation of several shape constraints, specifically, symmetry, topological redundancy, minimum variance of angles, and planarity of contours. Note that all of these prior models, using shape constraints, had one very important limitation. *They could only be applied to polyhedra.* The model presented in this paper is the first model of shape recovery based on shape constraints that can be applied to wide range of 3D shapes.

An additional reason for doing the research suggested above

We are contemplating a number of experiments on shape constancy now that we have a working model for the recovery of 3D shape from a 2D image. They will address one of the oldest questions in the history of the subject, namely, what is the role, if any, of familiarity in shape constancy? Until now, most perceptionists have assumed that it plays a huge role. As currently configured, our model does not use familiarity to recover 3D shape and it recovers 3D shape very well. It can, therefore, be used to assay the role familiarity plays, or may play, in shape constancy. This can be done by testing both our model and human observers with both familiar and unfamiliar objects. A "familiarity effect" can be measured, as well as demonstrated by comparing the observers' achievement of shape constancy with the model's.

Finally

Now that we have a working model that recovers 3D shapes from a single 2D image, and now that we have specific plans to extend its range of application to real images of real objects in natural environments, it is important to emphasize, that a big gap remains in the knowledge required to accomplish this very ambitious goal. Our, and similar recovery models, can only operate *after* figure-ground organization is established. Human beings do this exceedingly well. Their performance far exceeds our understanding of how they do it. We have been trying to solve this problem for only two years. Prof. Spekrijse recognized the significance of the figure-ground organization problem, and the urgent need to solve it, more than a decade ago, at a time when most vision researchers thought that figure-ground organization was in the dustbin of history where it belonged. Prof. Spekrijse knew better. The research he and his coworkers published during his last years has provided us with insights about the mechanisms underlying figure-ground organization that will help us solve this critical problem. It is sad, as well as unfortunate, that he will no longer be able to help us as we try to work this out.

References

- Berkeley, G. (1709/1910) A new theory of vision. NY: Dutton.
- Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, 115-147.
- Biederman, I. & Gerhardstein, P.C. (1993) Recognizing depth-rotated objects: Evidence and conditions from three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance* 19, 1162-82.
- Boyer, K.L. & Sarkar, S. (2000) Perceptual organization for artificial vision systems. Boston: Kluwer.
- Brady, M. & Yuille, A. (1983) Inferring 3D orientation from 2D contour (an extremum principle). In: Richards, W. (Ed.), *Natural computation* (pp. 99-106), Cambridge, MA: MIT Press.
- Chan, M.W., Stevenson, A.K., Li, Y. & Pizlo, Z. (2006) Binocular shape constancy from novel views: the role of *a priori* constraints. *Perception & Psychophysics*, 68, 1124-1139.
- Huang, T.S. & Lee, C.H. (1989) Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11, 536-540.
- Johnston, E.P. (1991) Systematic distortions of shape from stereopsis. *Vision Research*, 31, 1351-1360.
- Kimchi, R., Behrmann, M. & Oslon, C.R. (2003) Perceptual organization in vision. Behavioral and neural perspectives. Mahwah, NJ: Erlbaum.
- Koenderink, J.J., van Doorn, A.J. & Kappers, A.M.L. (2006) Pictorial relief. In: M.R.M. Jenkin & L.R. Harris (Eds.), *Seeing spatial form*, (pp. 11-33) Oxford: Oxford University Press.
- Koffka, K. (1935) *Principles of Gestalt Psychology*. New York: Harcourt, Brace.
- Lamme, V.A.F. & Spekreijse, H. (1996) Separate processing dynamics for image contours and surfaces in monkey V1. *Progress in Biophysics & Molecular Biology*, 65, Suppl. 1.
- Lamme, V.A.F., Zipser, K., & Spekreijse, H. (1998) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci U S A.*, 95, 3263-3268.
- Leclerc, Y.G. & Fischler, M.A. (1992) An optimization-based approach to the interpretation of single line drawings as 3D wire frames. *International Journal of Computer Vision* 9, 113-136.
- Li, Y. & Pizlo, Z. (2007) Reconstruction of shapes of 3D symmetric objects by using planarity and compactness constraints. *Proceedings of IS&T/SPIE Conference on Vision Geometry*, vol. 6499.
- Marill, T. (1991) Emulating the human interpretation of line drawings as three-dimensional objects. *Int. J. Comput. Vision* 6, 147-161.
- Marr, D. (1982) *Vision*. New York: W.H. Freeman.
- 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 & Performance*, 22, 173-186.

- Palmer, S.E., Rosch, E. & Chase, P. (1981) Canonical perspective and the perception of objects. In: Attention & Performance, Long J. & Baddeley A. (Eds.), vol. 9. Hillsdale, NJ: Erlbaum. 135-151.
- Pizlo, Z. (2008) 3D shape: its unique place in visual perception. Cambridge, MA: MIT Press (in press).
- Pizlo, Z., Li, Y. & Steinman, R.M. (2006) A new paradigm for 3D shape perception. European Conference on Visual Perception. St. Petersburg, Russia. August, 2006.
- Pizlo, Z., Li, Y. & Steinman, R.M. (2007) Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose. Spatial Vision (in press).
- Pizlo, Z. & Salach-Golyska, M. (1995) 3D shape perception. Perception & Psychophysics, 57, 692-714.
- Pizlo, Z. & Stevenson, A.K. (1999) Shape constancy from novel views. Perception & Psychophysics, 61, 1299-1307.
- Poggio, T., Torre, V. & Koch, C. (1985) Computational vision and regularization theory. Nature 317, 314-319.
- Polya, G. & Szego, G. (1951) Isoperimetric inequalities in mathematical physics. Princeton: Princeton University Press.
- Rothwell, C.A. (1995) Object recognition through invariant indexing. Oxford: Oxford University Press.
- Rubin, E. (1915) Synsoplevede Figurer.
- Sawada, T. & Pizlo, Z. (2008) Detection of skewed symmetry. Journal of Vision (in press).
- Sinha P. (1995) Perceiving and recognizing three-dimensional forms. Doctoral dissertation. Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science.
- Tikhonov, A.N. & Arsenin, V.Y. (1977) Solutions of ill-posed problems. New York: John Wiley & Sons.
- Togt, van der, C., Kalitzin, S., Spekrijse, H., Lamme, V.A.F. & Super, H. (2006) Synchrony dynamics in monkey V1 predict success in visual detection. Cerebral Cortex, 16, 136-148.
- Todd, J.T. & Norman, J.F. (2003) The visual perception of 3D shape from multiple cues: are observers capable of perceiving metric structure? Perception & Psychophysics, 65, 31-47.
- Vetter, T. & Poggio, T. (2002) Symmetric 3D objects are an easy case for 2D object recognition. In: Tyler, C.W. (Ed.), Human symmetry perception and its computational analysis. (pp. 349-359) Mahwah, NJ: Lawrence Erlbaum.
- Wertheimer, M. (1923/1958) Principles of perceptual organization. In: D.C.Beardslee & M.Wertheimer (Eds.) Readings in Perception, pp. 115-135. NY: D. van Nostrand.