Perception of 3D Symmetrical and Nearly-Symmetrical Shapes

Vijai Jayadevan, Aaron Michaux, Edward Delp, Zygmunt Pizlo
Purdue University, West Lafayette, Indiana, USA

A 2D orthographic image of a 3D mirror-symmetrical shape determines a one-parameter family of 3D symmetrical shapes. When the 2D orthographic projections of the 3D symmetry lines of the 3D shape are set parallel to the x-axis on the image plane, any pair of 3D shapes in this one parameter family are related to one another by the following 3D affine transformation:

$$\begin{bmatrix} X \\ Y \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\cos(2\alpha_1)-\cos(2\alpha_2)}{\sin(2\alpha_2)} & 0 & \frac{\sin(2\alpha_1)}{\sin(2\alpha_2)} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z_1 \end{bmatrix} \tag{1}$$

where α_1 and α_2 are the slants of the symmetry planes of the two 3D shapes. It is obvious that all 3D shapes in this family are consistent with the same 2D orthographic image, represented by coordinates (X,Y). Li et al.[1], showed that the monocular percept of a 3D mirror-symmetrical shape is the 3D shape in this one-parameter family that maximizes a modified 3D compactness measure, V/S^3, where V and S are the volume and the surface area of the 3D object. Because monocular 3D shape perception is under-constrained, the 3D percept is not guaranteed to be veridical, but the departure from veridicality agrees closely with the operation of the compactness constraint (aka prior). The binocular percept, on the other hand, *is veridical.* This was explained in [1], by the use of a Bayesian posterior that combines a compactness prior with a likelihood function that represents the perceived depth-order of the feature points

The present study generalized the experiment and the results reported in [1] in two ways. First, our 3D shapes were not always symmetrical as was the case in [1]. Our asymmetrical shapes were produced by stretching and shearing symmetrical shapes. These transformations represent the general 3D affine group that corresponds to a 3×3 transformation matrix. Second, our subject's task was to recover the 3D shape viewed monocularly or binocularly by adjusting 3, rather than only 1 of the parameters of the 3D affine transformation of the 3D shape that leaves the 2D orthographic image unchanged:

$$\begin{bmatrix} X \\ Y \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z_1 \end{bmatrix} \tag{2}$$

Transformation (2) is more general than transformation (1), which means that the subject in our experiment could produce (recover) a mirror-symmetrical, or an asymmetrical 3D shape, from the image of a mirror-symmetrical shape. Note that we included a pure stretch in the depth direction in (2); this stretch is represented by a_{33}. It turns out (Sawada - personal communication) that when a 2D orthographic image is produced by a 3D asymmetrical shape, which itself is an affine distortion of a 3D mirror-symmetrical shape, transformation (2) always contains a one-parameter family of mirror-symmetrical shapes. This means that our subjects could produce (recover) a mirror-symmetrical, or an asymmetrical 3D shape from the 2D image of an asymmetrical shape.

The viewing distance in our experiment was 1m. The stimuli were displayed using a head-mounted display. The subject adjusted the 3D shape of a rotating object to match the 3D percept produced by a single monocular, or a binocular, view of a 3D reference shape. Our main result was that the *binocular percept was always nearly veridical with both symmetrical and asymmetrical shapes.* This generalizes the veridicality result obtained in [1], in which only symmetrical shapes were used.

The following cost function accounts for monocular and binocular performance of 3 subjects:

$$E(X) = 0.4 * asymmetry(X) + 0.2 * compactness(X) + 0.4 * stereoacuity(X) \tag{3}$$

Specifically, the 3D shape X that minimizes E is what the subjects perceive. In (3), $asymmetry(X)$ is the mean absolute difference of all corresponding angles, $compactness(X)$ is V/S^3 of X, and $stereoacuity(X)$ measures the difference in depth order of the binocularly perceived shape and the depth order of X, when X has the 3D orientation at which it produces the same cyclopean orthographic image as the reference shape. In monocular viewing, $stereoacuity(X)$ disappears from the cost function. The results of optimization in the 3-parameter space of the 3D shapes represented by transformations (2) are sensitive to the starting point, suggesting the presence of multiple local minima. Good fits to our binocular results were obtained by starting with 3D shapes that satisfied the "perceived" depth-order of a subset of pairs of the vertices of the object. The model's binocular shape recovery is close to the true shapes which means that the model "perceives" 3D shapes veridically. It is also close to the shapes recovered by the binocular subjects. The model's monocular shape recovery of symmetrical and asymmetrical shapes is also close to the shapes recovered by the monocular subjects: both agree when the percept is veridical, as well as when the percept departs from veridicality.

References

[1] Li Y, Sawada T, Shi Y, Kwon T, Pizlo Z (2011) A Bayesian model of binocular perception of 3D mirror symmetrical polyhedra. *Journal of Vision* 11(4):1-20