Mental geometry for estimating relative 3D size

Akihito Maruya & Qasim Zaidi
Graduate Center for Vision Research, SUNY

A 3D object seen from different views forms quite different retinal images. Humans are very good at inferring 3D pose by using knowledge of projective geometry (Koch et al PNAS 2018). Shape "constancy" would suggest that we also infer correct 3D relative size/shape despite perspective distortions, but is that true? We presented frontal views of rectangular parallelepipeds (3 lengths) lying on the ground (16 poses). Observers (N=6) adjusted the height of an orthogonally attached narrow cylinder to equate the physical lengths of the two limbs (Fig 1). For a parallelepiped of length, (L_{3D}) the projected length, (L_c) changes with pose as a distorted sinusoid (Viewing elevation=Φ_c, focal length= f_c, distance from the object = d_c):

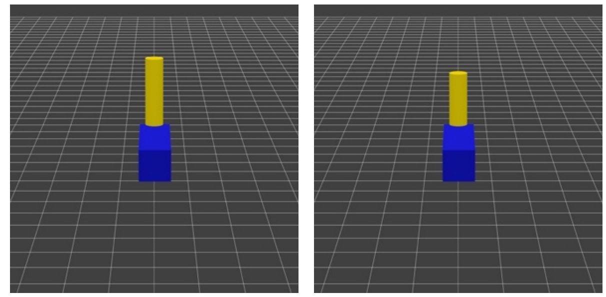

Fig.1

$$L_c = \frac{L_{3D} \cdot f_c \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(\Phi_c)}}{d_c - L_{3D} \cdot sin(\Omega) \cdot \cos(\Phi_c)}$$

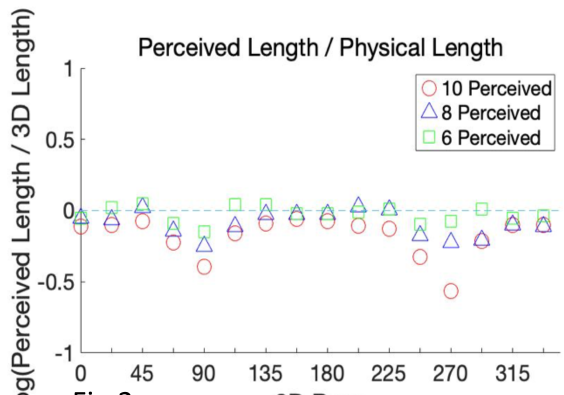

Fig.2

Observer's estimates of parallelepiped length were close to veridical for front-parallel poses but were seriously underestimated for poses pointing at or away from the viewer (Fig 2). The inverse of the function relating the projected length to the pose gives the optimal correction factor for inferring correct physical lengths from retinal images. Observers' correction factors were close to optimal for poses close to fronto-parallel, but seriously low for poses close to line of sight. Interestingly, the underestimation increased with physical length of the parallelepiped. Slant matching measurements revealed that longer objects were seen as slanted down, equivalent to an increase in viewing elevation. Increased viewing elevation requires a smaller correction factor, so we tested a model for estimating $\hat{L}_{3D}$, that adds a free parameter k to the optimal geometrical back-transform, where k>1 indicates overestimates of viewing elevation (focal length of the retina = f_r, distance of pupil from the screen = d_r, projected length on the retina = L_r):

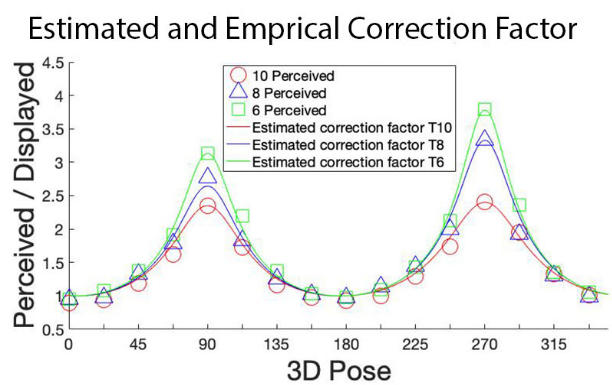

Fig.3

$$\hat{L}_{3D} = \frac{L_r \cdot d_r \cdot d_c}{L_r \cdot d_r \cdot \sin(\Omega) \cdot \cos(k \cdot \Phi_c) + f_c \cdot f_r \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(k \cdot \Phi_c)}}$$

This model explains the underestimation of object length (Fig 3.)

These results show that observers use the optimal geometric back-transform for estimating object length. Since illusory change in relative lengths of limbs describes one class of shape distortion, shape inconstancy results despite correct mental geometry, when retinal images of objects evoke misestimates of viewing elevation.