The Perceptual Representation of 3D Shape

James T. Todd
Ohio State University
Collaborators
Andrea Van Doorn
Astrid Kappers
Jan Koenderink
Joe Lappin
Baoxia Liu
Farley Norman
Stijn Oomes
Victor Perotti
Flip Phillips
A Miracle Occurs

Physiological Processes
Mathematical Algorithms
Prior Knowledge

The Physical Environment $\Phi$

$\Lambda = f(\Phi)$

The Structure of Light $\Lambda$

Perception of the Environment $\Psi$
The transformation $\Lambda = f(\Phi)$ is a many-to-one mapping

$$(X, Y, Z) \xrightarrow{} (X', Y')$$

Visual information is almost always ambiguous with an infinity of possible 3D interpretations.
The set of all possible 3D structures

Structures that are compatible with a particular pattern of visual stimulation
Relations among distance intervals in different directions are called **Metric Structure**

Metric structure is seldom specified by patterns of visual information.
Relations among distance intervals in parallel directions are called **Affine Structure**.

Affine structure is often specified by patterns of visual information.
Overview of this Presentation

<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Type of Ambiguity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion</td>
<td>Depth Scaling</td>
</tr>
<tr>
<td>Texture</td>
<td>Depth Scaling</td>
</tr>
<tr>
<td>Shading</td>
<td>Depth Scaling + Shear</td>
</tr>
<tr>
<td>Binocular Disparity</td>
<td>Nonlinear Depth Scaling</td>
</tr>
</tbody>
</table>
The perception of 3D structure from motion
In order to determine a unique 3D structure from a sequence of images it is necessary to have at least 3 distinct views of 4 points.
Todd & Bressan (1990)

For a motion sequence composed of just two views, the 3D structure is specified up to an unknown scaling in depth.
Observers

JT  JN  DL  JS  VP  Mean

Tilt  Slant

r

0.0  0.2  0.4  0.6  0.8  1.0

Tilt  Slant

Observers
Observers

JT

VP

FP

Mean

Shape Characteristic

Curvedness

r

0.0

0.2

0.4

0.6

0.8

1.0

Observers

JT

VP

FP

Mean
Shape from Texture
Near-Far Task

Positions of the Depth Extrema from Left Edge of Monitor (cm)

Judged Positions of Depth Extrema (cm)

5  10  15  20  25  30  35  40
Profile Task

Depth Difference Between Adjacent Probe Points (cm)

Judged Depth Difference (cm)
Shape from Shading
Bas-Relief Ambiguity

**Proposition:** For a given pattern of image shading there is an infinite number possible surface interpretations that are all related by an affine transformation.

Belhumeur, Kriegman, & Yuille, (1999)
Bas-Relief Ambiguity
adjust
Affine Structure

\[ Z' = a \cdot Z \]

\[ Z' = a \cdot X + b \cdot Y + c \cdot Z \]
Two Possible Conclusions

1) Our perceptual representations are based primarily on nonmetric properties such as affine, ordinal or topological relations.

2) Our perceptual representations have explicit information about 3D metric structure, but that information is inaccurate.