What is the visual system like?

The role of computer graphics

Daniel Kersten

http://vision.psych.umn.edu/www/kersten-lab/kersten-lab.html

kersten@tc.umn.edu

Psychology Department, University of Minnesota

Workshop on Rendering, Perception, and Measurement.
April 8-10, 1999
Cornell University
Outline

Interaction between disciplines

Grand challenge: visual information -> performance

Bayesian framework
  Image information, priors, and task requirements
  Task-specific perceptual inference

Examples
  Depth from moving shadows
  Discounting the color of mutual illumination
Quantitative studies of human vision
Relation to Psychology, Computer Science, Neuroscience

Perceptual Psychology

Computer vision
Visual neuroscience
Computer graphics?
Role of computer graphics?

Tools to test theories of visual perception
  Unprecedented ability to control experimental variables

Physically-based models
  Anchor empirical results in human vision to the natural world

Inverse graphics as a metaphor for vision
  – When does it provide insight?
  – Where does it fail?
Primate visual architecture

From: Milner and Goodale, 1995
Visual neural architecture

- Complex system
- Multiple pathways
- Multiple interacting visual areas:
  - feedforward and feedback connections
- Neural architecture reflects task-dependency
Theories of visual perception: Coping with complexity

Levels of analysis

- Qualitative computational/functional theories
- Quantitative theories of statistical inference
- Algorithmic theories
- Neural implementation theories
Our Grand Challenge

Theoretical challenge
- Characterize the limits to visual inference given the complexity of natural images

Empirical challenge
- Testing quantitative theories of visual behavior

Proposed strategy:
- Statistical theories of visual inference to bridge perception and neural theories
Levels of visual processing

Early vision
- Sensory sensitivity, local image measurements, especially those related to surface properties

Intermediate-level vision
- General purpose, global organization processes
  - Surface grouping, smoothing, long boundaries
- Cue integration
- Cooperative computation of surface properties - shape, color & indirect lighting

High-level vision
- Functional tasks
  - Object recognition
- Viewer-object relations - manipulation, navigation
- Object object relations - spatial layout & cast shadows
Problems of ambiguity

Geometric: Many 3D shapes can map to the same 2D image

Photometric: The scene causes of local image intensity change are confounded in the image data
Scene causes of image intensity change

- Attached shadow
- Thin surface edge
- Crease
- Cast shadow
- Highlight
- Material

Question mark at the bottom left corner of the image.
Knowledge required for inference

Specify the visual task:
- Explicit variables: important scene variables for task
- Generic variables: less important for task, but contribute to image data

Image measurements supporting inference
Prior knowledge of scene structure
Bayesian framework for visual inference

Scene variables of interest, $S_e$
Confounding scene variables, $S_g$
Image measurements, $I$

Characterize the posterior probability, $p(S_e | I)$, using Bayes’ rule:

$$p(S_e | I) \propto p(I | S_e) p(S_e)$$

$p(S_e)$ prior probability of $S_e$
$p(I | S_e)$ likelihood term based on model of image formation

Make decision based on prior and likelihood constraints
Image data

Adapted from a figure by Sinha & Adelson
Which scene descriptions are likely to give rise to the image data?
Likelihood selects subset of scene descriptions consistent with image data

Adapted from a figure by Sinha & Adelson
Prior further narrows selection

Priors weight likelihood selections

$p(s)=p_1$

$p(s)=p_2$

$p(s)=p_3$

is biggest

Select most probable

$p(s)=p_3$

Adapted from a figure by Sinha & Adelson
Some scene variables are more important to estimate accurately than others, depending on the task.

Bayesian decision theory
- Loss functions, risk
- Marginalization of the posterior

Task dependence for visual tasks
- Sample taxonomy: recognition, navigation, etc.
Task dependency

- shape
- material
- articulation
- viewpoint
- Relative position
- illumination

(image)
### Task dependency: explicit and generic variables

\[ I = f(\text{shape, material, articulation, viewpoint, relative position, illumination}) \]

<table>
<thead>
<tr>
<th>Object perception</th>
<th>Spatial layout</th>
<th>Entry-level</th>
<th>Subordinate-level</th>
<th>Planning</th>
<th>Reach</th>
<th>Grasp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object-centered</strong> (object recognition)</td>
<td>World-centered</td>
<td>G</td>
<td>G</td>
<td></td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td></td>
<td>E</td>
<td>E</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td><strong>Material</strong></td>
<td></td>
<td>G</td>
<td>E</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td><strong>Articulation</strong></td>
<td></td>
<td>G</td>
<td>E</td>
<td>G</td>
<td>G</td>
<td>E</td>
</tr>
<tr>
<td><strong>Viewpoint</strong></td>
<td></td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td><strong>Relative position</strong></td>
<td></td>
<td>G</td>
<td>G</td>
<td>E</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td><strong>Illumination</strong></td>
<td></td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

**Explicit (E) = Primary**

**Generic (G) = Secondary**
Task dependency

Scene variables that are important to know, $S_m$

Generic variables that contribute to the image, but do not need to be explicitly estimated, $S_g$

$$p(S_e | I) = \int p(S_e, S_g | I) dS_g$$

Perception’s model of the image should be robust over variations in generic variables (Freeman, 1994)

Task specification can reduce ambiguity
Bayesian view of perception

What is visual perception like?

Revisionist Helmholtz

“Perception is (largely) unconscious statistical inference involving unconscious priors and unconscious loss functions”

Recent success stories

Knill (1998) - Perception of surface slant from texture
Weiss and Adelson (1998) - Motion perception
Two examples

Depth from cast shadows
  Qualitative discussion of 3 types of constraints

Discounting the color of mutual illumination
  Quantitative implementation of constraints in a Bayesian model
Depth from cast shadows
Shadow motion vs. object image motion

http://vision.psych.umn.edu/www/kersten-lab/shadows.html

“Square-over-checkerboard”
Summary of results

Light from above is better than from below
Dark shadows are better than light
Extended light sources lead to stronger depth effect
Moving cast shadows can veto other strong cues:
  - Size change
  - Stereo

http://vision.psych.umn.edu/www/kersten-lab/shadows.html
Knowledge required to resolve ambiguity

Piece together a scene model of explicit variables using:

Consistency with image data

Image formation constraints $\rightarrow$ likelihood

Prior probabilities

Task dependence

Robustness over generic variables
Problems of ambiguity

Many 3D shapes can map to the same 2D image

The scene causes of local image intensity change are confounded in the image data
Examples of local image formation constraints

- zero image motion | zero object motion
- edge properties
  - fuzzy edge | shadow penumbra
  - fuzzy edge | surface crease
  - fuzzy edge | attached shadow
- edge junctions
  - “T” | occlusion
  - “T” | accidental alignment
  - “X” | transparent surface
  - “X” | cast shadow
Depth from cast shadows

- zero image motion | zero object motion
- edge properties
  - fuzzy edge | shadow
  - penumbra
- edge junctions
  - “T” | occlusion
  - “X” | transparent surface
  - “X” | cast shadow
Prior constraints

- light source position is above
- no light source motion
- object size is constant
- background is not moving
- background is opaque
Task decisions reduce ambiguity
Depth ambiguity: marginalizing over light direction

\[ p(z \mid x) \propto \int p(x \mid z, \alpha) d\alpha \propto \int e^{-(x-F(z,\alpha))^2 / 2\sigma^2} d\alpha \]

\[ p(z \mid x) \propto \frac{1}{\left| \frac{\partial F}{\partial \alpha} \right|_{\alpha=\alpha_M}} \]

\[ x = F(z, \alpha) = z \tan(\alpha) \]

\[ p(z \mid x) \propto \frac{z}{x^2 + z^2} \quad \text{peaks for} \quad z = x \]
Light source direction: generic

Perception’s model of the image should be robust over variations in generic variables

\[ \Delta x = \Delta \alpha \frac{x^2 + z^2}{z} \]

\[ z = x \]

See too: Shape from shading, Freeman, 1994; Viewpoint as a generic variable: Lowe, 1986; 1987; Nakayama & Shimojo, 1992
Discounting the Color of Mutual Illumination

Marina Bloj, Newcastle University
Anya Hurlbert, Newcastle University
Daniel Kersten, University of Minnesota

Marina
Physical measurements
Chromatic “Mach Card”

Step 1: Cut out

Step 2: Illuminate

Step 3: View
Chromatic Mach Card

“Corner” percept

“Roof” percept

Same retinal stimulus *appears* pinkish
The Phenomenal Result

When seen as concave, the white paper is seen as white

When seen as convex, the white paper is seen as pink

Human visual system uses knowledge of mutual illumination to recover surface color
Chromatic Mach Card: Psychophysics

Pseudoscope
Chromatic Mach Card

“Corner” match

“Roof” match with pseudoscope
Chromatic Mach Card

Roof matches, roof first
Corner matches, corner first

number of matches

saturation
Psychophysics & theory

Assumptions:
- Prior on single white light source
- Independent knowledge of card’s shape (stereo)

Task-dependent assumption:
- Surface reflectance is important (explicit variable)
- Light source direction, $\alpha$, is not (generic)

Geometry states:
- $L =$ corner
  - Indirect plus direct lighting (1 bounce model; Funt & Drew, 1991)
- $L =$ roof
  - Direct lighting only (0 bounce)
Predicted distribution of card matches

\[ p(\rho_1(i) \mid c_{obs}, L) = k \sum_{\alpha, x} \exp \left\{ -\frac{1}{2\sigma^2} \left[ c_L(\alpha, x, i, L) - c_{obs}(x) \right]^2 \right\} \]

\( \rho_1(i) \) : surface reflectance of \( i^{th} \) card (\( i = 1 \) to 23)

\( c_{obs} \) : observed chroma of "white" card

\( c_L(\alpha, x, i, L) \) : predicted chroma of \( i^{th} \) card for illuminant direction \( \alpha \), position \( x \)
Model vs. psychophysics: How well do chroma matches predict shape?

Gaussian approximation:

\[ d' = \frac{\mu_{\text{roof}} - \mu_{\text{corner}}}{\sigma} \]

\[ d'_{\text{human}} \approx 0.82 \]

\[ d'_{\text{model}} \approx 0.99 \]
What is the visual system like?

Complex information processing task is reflected in the sophistication of neural architecture

Grand challenge

- Given image ambiguities, develop quantitative models of how visual information limits visual performance with natural images
- ...not there yet

Treat human vision as processes of statistical inference which depend on task

Role of computer graphics?

- Physically based rendering -> relation of simulated to real
- Statistical prior models: facial surfaces, plants, biological motion, etc..
- Image-based prior models: e.g. textures & density estimation


Collaborators

Marina Bloj, U. Newcastle
Anya Hurlbert, U. Newcastle
Isabelle Bülthoff, Max-Planck-Institut für biologische Kybernetik
Heinrich H. Bülthoff, Max-Planck-Institut für biologische Kybernetik
David C. Knill, University of Pennsylvania
Cindee Madison, U. Minn.
Pascal Mamassian, University of Glasgow
Paul Schrater, U. Minn.
Alan Yuille, Smith-Kettlewell Eye Research Institute, San Francisco