Processing Framework Proposed by Marr

Recognition

3D structure; motion characteristics; surface properties

- Shape From stereo
- Motion flow
- Shape From motion
- Color estimation
- Shape From contour
- Shape From shading
- Shape From texture

Edge extraction

Image
Stereopsis

The process of perceiving the relative distance to objects based on their lateral displacement in the two retinal images (aka binocular disparity).

Other potential cues to depth (convergence and accommodation) do not seem to be too important for humans.
Stereopsis

Challenges:
1. Trigonometric calculations
2. Correspondence problem

How is the correspondence problem solved?
The Stereo Correspondence Problem

“During binocular regard of an objective image, each uniocular mechanism develops independently a sensual image of considerable completeness. The singleness of binocular perception results from the union of these elaborated uniocular sensations. *The singleness is therefore the product of a synthesis that works with already elaborated sensations contemporaneously proceeding.*”

- Sherrington, 1906
The Stereo Correspondence Problem

A. Shape-First Theory

- Stereopsis
  - Shape Analysis
    - Left Image
  - Shape Analysis
    - Right Image
Is monocular shape analysis a necessary pre-requisite for stereo correspondence?
Is monocular shape analysis a necessary pre-requisite for stereo correspondence?
Computational theories for solving the correspondence problem:

Given the underconstrained matching problem (100! Possible pairings in an RDS with 100 dots), what assumptions can we bring to bear?

Assumption 1: Epipolar constraint
Marr-Poggio’s network-based formulation of the problem:

Assumptions:

1. Surface opacity / match uniqueness
2. Surface continuity
3. Match compatibility
Sample result of Marr-Poggio’s network:
Enhancing the Marr-Poggio’s model: (Marr-Poggio-Grimson)

Edge-based matching rather than pixel matching.

Advantages:

Disadvantages:
Enhancing the Marr-Poggio’s model:

*Edge-based matching rather than pixel matching.*

Advantages:

1. Edge orientation and polarity provide additional matching constraints

Disadvantages:
Testing for polarity specificity of stereo-matching
Enhancing the Marr-Poggio’s model:

*Edge-based matching rather than pixel matching.*

**Advantages:**
1. Edge orientation and polarity provide additional matching constraints
2. Greater consistency with known physiology (matching begins in V1)

**Disadvantages:**
Depth information is sparse; an additional process of interpolation is needed.
Open questions:

1. How to match stereo pairs where assumptions are violated?

2. Does RDS stereopsis conclusively prove that the shape-first theory is incorrect?
   
   Are RDSs representative of real-world scenes?
   Are RDSs completely devoid of monocular shape cues?
A woman named "Elizabeth," was studied and written about by Charles F. Stromeyer in 1970. She was an artist and teacher at Harvard who could mentally project detailed and exact images onto her canvas and was even able to move her eyes about to inspect the image while the image stayed still. She could also reproduce poems in a foreign language years after having seen the original printed page.

In Stromeyer's tests on her abilities, "Elizabeth" was presented with a 10,000-dot stereogram pattern to one eye for a specified length of time and then was asked to superimpose her eidetic image onto another pattern presented to her other eye. She was able to perform this task with great ease and could see depth and figures in these patterns. Non-eidetikers need a stereoscope to perform this feat.

"Elizabeth" was also capable of projecting her eidetic images onto other images, often obscuring the actual image. Her eidetic images were capable of after-images and movement after-effects just like that of actual visual stimulus, and she is even reported to have been able to see a 10-second section of a movie in complete eidetic detail.

Her only constraint was that she had to move her eyes to scan an eidetic image and generally would create the image in sections rather than as a whole. Also, "Elizabeth"'s images did not just fade, but instead would dim and break apart piece by piece. In any case, "Elizabeth" is the only one of her kind. Since the publication of Stromeyer's paper, no other adult eidetiker of her caliber has been found.
What happens when no correspondence is possible?

Highly mismatched stereo-pairs lead to ‘binocular rivalry’

Open questions:
What is the site of binocular rivalry?
Can rivalry and fusion coexist?
Fun with stereoscopes...

“Although a perfect stranger to you, and living on the reverse side of the globe, I have taken the liberty of writing to you on a small discovery I have made in Binocular vision in the stereoscope. I find by taking two ordinary photos of two Different persons’ faces, the portraits being about the same sizes, and looking About the same direction, and placing them in a stereoscope, the faces blend into One in a most remarkable manner, producing in the case of some ladies’ Portraits, in every instance, a decided improvement in beauty.”

- From a letter to Charles Darwin by A. L. Austin of New Zealand
<table>
<thead>
<tr>
<th>Trait</th>
<th>Measurement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taille 1°</td>
<td>Long</td>
<td>Pied g.</td>
</tr>
<tr>
<td>Voute</td>
<td>Larg</td>
<td>Médius g.</td>
</tr>
<tr>
<td>Enverg 1°</td>
<td>Long</td>
<td>Auric* g.</td>
</tr>
<tr>
<td>Buste 0°</td>
<td>Larg</td>
<td>Coudée g.</td>
</tr>
<tr>
<td>N° de cl.</td>
<td></td>
<td>Coût de l'iris</td>
</tr>
<tr>
<td>N° de cl.</td>
<td></td>
<td>Aur*</td>
</tr>
<tr>
<td>Père</td>
<td></td>
<td>Pér*</td>
</tr>
<tr>
<td>Mère</td>
<td></td>
<td>dep</td>
</tr>
<tr>
<td>Age de</td>
<td></td>
<td>née le</td>
</tr>
<tr>
<td>Age appl</td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

(Réduction photographique 1/7.)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Measurement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racine (cavité)</td>
<td></td>
<td>Bord. o. s. p. f.</td>
</tr>
<tr>
<td>Dos</td>
<td></td>
<td>Lob. c. a. m. d.</td>
</tr>
<tr>
<td>Haut' Saillie</td>
<td></td>
<td>A. trg. l. p. r. d.</td>
</tr>
<tr>
<td>Haut' Larg'</td>
<td></td>
<td>Pli. f. s. h. E.</td>
</tr>
<tr>
<td>N° de cl.</td>
<td></td>
<td>Cheveux C/T sang*</td>
</tr>
<tr>
<td>Borbe</td>
<td></td>
<td>Car C/uint.</td>
</tr>
<tr>
<td>Sex.</td>
<td></td>
<td>Autres traits caractéristiques :</td>
</tr>
<tr>
<td>Front.</td>
<td></td>
<td>Sig. dressé par M.</td>
</tr>
</tbody>
</table>

M. G. Galton 19.4.93
Composite of 14 criminals’ faces
Composite of 15 women’s faces
Processing Framework Proposed by Marr

3D structure; motion characteristics; surface properties

- Shape From stereo
- Motion flow
- Shape From motion
- Color estimation
- Shape From contour
- Shape From shading
- Shape From texture

Edge extraction

Image
Motion Perception:

-Simple motion detectors
-Extracting 2D motion fields

How can we tell whether this is really a motion selective cell (rather than just an orientation selective one)?

How can we design a simple motion detector?
Motion Perception:

- Simple motion detectors
- Extracting 2D motion fields

How can we tell whether this is really a motion selective cell (rather than just an orientation selective one)?

How can we design a simple motion detector?
Simple motion detectors

Motion as space-time orientation:

A MOTION SEQUENCE is a series of images measured over time. (A) The motion sequence of images can be grouped into a three-dimensional volume of data. (B) Cross sections of the volume show the spatial pattern at a moment in time. (C) Time (t) may be plotted against one dimension (x) of space. When the spatial pattern is one-dimensional, the (t, x) cross-section provides a complete representation of the stimulus sequence.

Desired rf structure to detect oriented patterns in space-time

A: Slow motion detector

B: Fast motion detector

How can such rfs be constructed?
Constructing motion detectors:

Delay and compare networks
Other ways of constructing movement detectors:

A. Delayed summation

B. Delayed inhibition

C. Spatial & Temporal differencing

D. Even & Odd filters

Are there really s-t oriented rfs in the brain?
Is this all there is to determining whether a pattern is in motion?
Accounting for eye-motion

Q. When do we see an object move?
A. When its image moves on the retina.
   Is this really true?
Accounting for eye-motion (contd.)

The corollary discharge model (Teuber, 1960)

Predictions: 1. Pushing on the eyeball would cause the world to

2. A stabilized after-image would appear to when the eye is

   moved voluntarily

3. If your eye was paralyzed with curare and you then attempted to

   move it, you would see the world
Oculoparalytic Illusion: Visual-Field Dependent Spatial Mislocalizations by Humans Partially Paralyzed with Curare

Abstract. In darkness, observers partially paralyzed with curare make large (> 20 degrees) gaze- and dosage-dependent errors in visually localizing eye-level-horizontal and median planes, in matching the location of a sound to a light, and in pointing at a light. In illuminated, structured visual fields visual localization and pointing are accurate but errors in auditory-to-visual matches remain. Defects in extraretinal eye position information are responsible for all errors. The influence of extraretinal eye position information on visual localization is suppressed by a structured visual field but is crucial both in darkness and for intersensory localization if visual capture is prevented.

Shifts of the image at the back of the eye are produced either by eye movements or by displacements of the visual field of view. Although displacements of the visual field are normally perceived to be displacements, in the presence of eye visual field. We have called this the oculoparalytic illusion (OPI) and describe it below.

When the reclining, partially paralyzed observer (Fig. 1g) fixated a single stationary visual target (3) at eye level in

Science, 1982
Interim summary:

We roughly understand how to construct simple motion detectors.

Are such detectors sufficient for estimating the motion of complex patterns in the environment?
From local motion estimates to global ones:

Local motion estimates are ambiguous due to the ‘Aperture Problem’

So, how can we derive the global motion field?
From local motion estimates to global ones (contd):

Theoretically, the ‘aperture problem’ can be overcome by pooling information across multiple contours or by --------------.

What happens if we remove ---------?
Subjective plaids video

Sinha, 1996

Moving grating bars
Stationary background pattern

(b)
From local motion estimates to global ones - physiology:
Motion fields for more complex patterns:

Hildreth (1985): Smoothness of velocity field along the contour

Is there any perceptual evidence for the validity of this idea?
Motion fields for more complex patterns (contd.):

(a) True
(b) Local
(c) Smoothest

(a) True
(b) Local
(c) Smoothest
Motion fields for more complex patterns (contd.):

Ellipse demo

ellipse-plain-fat.mov
Processing Framework Proposed by Marr

3D structure; motion characteristics; surface properties

Shape From stereo | Motion flow | Shape From motion | Color estimation | Shape From contour | Shape From shading | Shape From texture

Edge extraction

Image