Week 9: BINOCULAR VISION

1) Disparity and depth
2) Projection theories
3) The correspondence problem
4) Matching algorithms
5) Prazdny’s algorithm
6) Grimson’s wedding cake
7) Kaufman’s stereogram: Rivalry
8) Occlusion, depth, and da Vinci stereopsis
9) Modal and amodal perception (NOT transparency)

CHARACTERIZATION OF STEREOPSIS

Classical formulation of stereopsis:
Salvation through projective geometry

Disparity -- find light rays with same origin in world --> different coordinates in two eyes.

Matching algorithms
What gets matched? (Coarse vs. fine features; emergent features)

Non-matches
What if light from a surfaces reaches only one eye? (occlusion)

Multiple spatial scales
Simultaneous fusion and rivalry in the same region

STEREOPSIS AND FUSION

Stereopsis (classical statement):
From: two “2-D” images
To: single “3-D” percept

CLAIM: The perceptual phenomena of stereopsis are not well captured by the phrase “shape from disparity.”

Note: Greek root “stereo” refers to “solid,” in the sense of volume, (not to two of anything, such as audio speakers).

Question 1: Can we have “fusion” of information from two eyes without “stereopsis”?
Question 2: Can we have stereopsis without fusion?!

RELATIVE DISPARITY

Consider vergence angle:
Formed when two eyes fixate same point.

Suppose two points, A and B, are at different depths.

Retinal distances between pairs of points, (A_L -- B_L) and (A_R--B_R), differ in the two eyes.
**SUFFICIENCY OF RELATIVE DISPARITY**

Vergence angle: Gotten from efferent motor signals (?)

Sperling (1970): Vergence information can influence perceived depth, but . . .

Disparity alone at times suffices to induce impression of depth. (Julesz random dot stereograms.)

Note: For many random dot stereograms, we can only sense that two surfaces are at “different” depths, separated by an almost arbitrary scale factor.

The depth separation may seem very small, as when the edge of one sheet of paper lies on another sheet.

**KEPLER’S PROJECTION THEORY**

Perceived depth corresponds to *locus of ray intersections* for matching targets.

Kaufman Fig. 8-9.

**ISSUES IN PROJECTION THEORY**

*What* should we match?

That is, *what are the units* of disparity?

I.e., must we first find edges, blobs, etc? Can we match “raw” luminance information?

What should we do about “false matches”?

Kaufman, Fig 8-19

**SPERLING’S NBF**

NBF: The “neural binocular field” is said to “mirror” spatial relationships in the world.

Note that in this diagram, from Sperling, 1981, objects are represented “behind” the eyes (in the brain!).
MARR’S VERSION

Corresponding diagram from Marr (1982, Fig 3-5). Note that objects are represented in front of the eyes (in the world!)

```
<table>
<thead>
<tr>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>i4</th>
</tr>
</thead>
<tbody>
<tr>
<td>l1</td>
<td>l2</td>
<td>l3</td>
<td>l4</td>
</tr>
</tbody>
</table>
```

“Of the sixteen possible matchings, only 4 are correct (filled circles); the remaining 12 are false targets (open circles). Without further constraints based on global consideration, such ambiguities cannot be resolved.”

What assumptions are tacit in the above quote?

STEREOPSIS DESIGN PROBLEM

Sperling, 1981:

Depth information is “by its very nature” local.

Q: How local is it?

What are the units to be matched?

GENERAL
- luminance
- zero-crossings
- peaks, edges

SPECIFIC
- complex patterns
- form (e.g. T-junctions, forks, ...)
- emergent segmentations

Results
- Good: Denser depth info (when correct)
- Bad: More false matches

Results
- Good: Fewer false matches
- Bad: More resulting empty (ambiguous) regions

HOMOGENEOUS IS AMBIGUOUS

Homogeneous regions carry no useful information for depth-from-disparity; all interior positions match equally well.

Similarly, a dense, uniform texture provides many “false matches.”

Given that we see the interior at some depth, local disparity alone cannot explain our perception.

Suggestion: Match edges and “fill in” the resulting depth information to interiors. (Sound familiar?)

NOTE: “Filling-in” is here intended in a functional (i.e. vague!) way; the mechanism need not be diffusion.

DISPARITY VARIATIONS FROM A FLAT SURFACE

Imagine a textured surface:

While not to scale, the diagram on the right indicates that flat surfaces can generate disparity gradients.

Compare the above geometry with the standard method of constructing stereograms, in which blocks of data are shifted.

NOTE: The right image contains an element (D) that does not exist in the left image.
MISMATCH MAKER

Occlusion: Some features in one eye’s view may have no match.

GENERAL OR SPECIFIC?

Return to design issue: Match on general or specific features?

Classic tradition in late 20th century stereo literature:
Match on general features and use cooperative-competitive mechanisms to cull out false matches

Sperling (1970)
Julesz (1971) “spring-coupled ‘dipoles,’ in the sense of magnets”
Dev (1975)
Nelson (1975)
Marr & Poggio (1976)

Compare versions of history given by:
Marr & Poggio (1976) and Marr (1982)
Grossberg (1983)
Julesz’s commentary on Grossberg (1983)

Note: Before the invention of random dot stereograms (Julesz), it was not even possible to pose the question above, as all stereograms (e.g. photographs) had matching form features.

da VINCI STEREOPSIS

Barrand (1978), following Gibson’s and Kaplan’s work with accretion and deletion of kinematic texture from occlusion, argues that such non-matching elements may be the most important source of information in binocular vision.

Note: Such elements may at times give you “only” ordinal information about “depth,” though for certain stimuli a fairly precise degree of depth can be obtained. Nakayama and Shimojo, 1990

Cf: S.S.Stevens’s taxonomy of measurement scale types:
(nominal), ordinal, interval, ratio

Note: “da Vinci” stereopsis -- depth from unmatched elements -- has become a growth industry in recent years.

THE STereo CORRESPONDENCE PROBLEM

Which feature in left eye matches which feature in right eye? (This is the “flip side” of the problem of false matches).

ASIDE: Can the correspondence problem be “bypassed”?

Sperling (1970): allow only one match along a given visual direction in the cyclopean field of view; i.e., a “winner-take-all” network in the vertical direction of the NBF, such that “A exclusive-OR B”.

This suggestion was incorporated into Dev’s 1975 algorithm.
**HISTORY, NOT HER STORY**

The following diagram depicts the same situation as in the previous panel. Here the depth axis points diagonally. Marr, 1982, Fig. 3-9

“... if we represent these connections in the same way as in Figure 3-6, it becomes obvious that they implement slightly different constraints. Instead of forbidding double matches down each line of sight, as was the case in Figure 3.6, these connections forbid double matches along the radial out from the viewer. It is incorrect to formulate the stereo correspondence process in this way.”

What does Marr suggest as an alternative?

**MARR & POGGIO’S 1976 INSIGHT**

Marr & Poggio variation on Dev’s (1975) algorithm:
Consider a view of a shallow lake’s water surface and its bottom.

Want: not “one depth per cyclopean direction,” but rather “a visible marking in one eye’s view originates whether from the surface of the water or from the bottom, but not from both.”

**CONTRAST POLARITY AND SPATIAL SCALE**

Recall that for grouping, there is a strong interaction of the effects of spatial scale (separation) and direction of contrast.

(What does this say about the nature an locus of rectification in the visual system?)

**MARR’S STEREO HEURISTICS**

“Rule 1: Compatibility. Black dots can match only black dots.”
True for random dot stereograms; not true for figures such as:

Caution: *What* is being matched here?

Q: What is the key difference between a random dot stereogram and this stereogram?

“Rule 2: Uniqueness. Almost always, a black dot from one image can match no more than one black dot from the other image.”

What if there are no dots? What are the units (!) of matching?

“Rule 3: Continuity. The disparity of the matches varies smoothly almost everywhere over the image.”
Simply not true in general, e.g., looking at a house through bushes.
MARR & POGGIO’S 1976 ALGORITHM

Marr’s description of the M & P, 1976, algorithm is summarized by:

\[
C_{x,y,d}^{t+1} = \sigma \left\{ \frac{1}{\sum_{x',y',d' \neq x,y,d} C_{x',y',d'}^t} \right\} \sum_{x',y',d' \neq x,y,d} C_{x',y',d'}^t - \varepsilon \sum_{x',y',d' \neq x,y,d} C_{x',y',d'}^t + C_{x,y,d}^0
\]

where:

- \( C_{x,y,d} \) -- C = cell; x, y = position, d = disparity, t = time state: 1 = match; 0 = no match
- \( S(x,y;d) \) -- local excitatory neighborhood (Rule 3.)
- \( 0(x,y;d) \) -- local inhibitory neighborhood (Rule 2.)
- \( \varepsilon \) -- inhibition constant
- \( \sigma \) -- threshold function
- \( C_{x,y,d}^0 \) -- initial state (all possible matches)

PRAZDNY’S COHERENCE PRINCIPLE

Quoting from Prazdny, 1985:

The coherence principle is much more general. It recognizes that for transparent surfaces where proximal points on the projection surface may arise from widely separated three-dimensional objects, image proximity does not necessarily imply disparity continuity. While the disparity field may be locally discontinuous, it must (if it is generated by an actual three-dimensional scene obeying the coherence principle) be a superposition of locally smooth disparity fields corresponding to individual three-dimensional surfaces. These smooth variations usually are apparent only when larger image regions are taken into consideration. Locally, the field may be discontinuous due to disparities originating at different depth. In short, a discontinuous disparity field may be a superposition of a number of several interlaced continuous disparity fields each corresponding to a piecewise smooth surface. The coherence principle captures this possible state of affairs and includes continuous disparity variations associated with opaque surfaces as a special case. [emphasis added]

*While truly transparent surfaces may occur infrequently in the nature, semi-transparency (fences, bushes, grass viewed by small creatures against the horizon, etc.) is a general phenomenon.

ALTERNATIVE ALGORITHMS

Marr (1982) criticizes cooperative-competitive algorithms for preattentive vision, claiming that “relaxation time” is too long, and goes on to describe a second algorithm (Marr & Poggio, 1979) that uses instead a “coarse to fine” matching strategy across spatial scales.

I.e., where coarse scales match, look for finer scale matches too.

Prazdny (1985) introduced a
“noniterative, parallel, and local algorithm” (“one-shot”)

No competition, but “facilitation” between similar, possibly somewhat distant, disparities.

Where Marr looks for continuity among adjacent matches, Prazdny looks for coherence in a region.

MORE PRAZDNY

... Two disparities are either similar, in which case they facilitate each other because they possibly contain information about the same surface, or dissimilar in which case they are informationally orthogonal, and should not interact at all because they potentially carry information about different surfaces.

...[seek a] simple scalar function capturing the following three requirements:
1. The disparity similarity function should be inversely proportional to the difference of disparities of interacting points.
2. More distant points should exert less influence while nearby matches should have more disambiguating power.
3. The more distant the two interacting points are the less seriously should their disparity difference be considered because of the inherent uncertainty: steeply-sloped surfaces will generate large disparity differences which should nevertheless contribute to disambiguation.
1) Support from similar disparities is stronger from nearby than from far locations.
2) Support over large distances can occur for large disparity differences.

Note: Physiological studies confirm that cells with relatively large receptive fields can fuse over a wider range of disparities than cells with small receptive fields. (It is not that “large” cells fuse only large disparities.)

Consequences of Prazdny's coherence principle:
Algorithm should consider only the “best evidence” for support among available disparities from one position to another; reward “agreement” without punishing “disagreement.”

Note: This idea also affords computational savings.
ANATOMY OF VISUAL DIMENSIONS

The idea of representing different aspects ("dimensions") of visual experience in separate representational structures is highly resilient, despite evidence of perceptual interactions among dimensions (e.g., neon spreading, contrast polarity and grouping, etc.)

Consider:
- Treisman (1980’s) “feature integration theory”
- Barrow & Tenenbaum (1981) “intrinsic images”
- Marr (1982)
- Livingstone & Hubel (1987) “separate processing streams”

and current successors to these approaches.

Consider also such phrases as: “shape-from-X,” “depth map,” “orientation map,” or “velocity field,” uttered as if all other aspects of stimulation could be abstracted away.

THE ROAD NOT TAKEN

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence;
Two roads diverged in a wood, and I --
I took the one less traveled by,
And that has made all the difference.

Robert Frost (1916)

There he goes again! What has this got to do with vision?

In the “classical” computer vision literature, and persisting to this day, the only role for a binocular vision is construction of a depth map. (Other modules do other jobs, and the stitching together of the results of all the modules is a job for still another module.) We could consider various computational strategies for computing depth maps, including tradeoffs of computational complexity, robustness, sensitivity to various types of image noise, etc.

Instead, we will consider what else primate vision is doing besides (or instead of?) constructing a depth map when using binocular mechanisms.

TWO QUESTIONS

My question:
How can one locate corresponding structure in such “disparate” iconographic schemes?

Your question:
Why should I bother?

Compare THIS ONE to Prazdny, 1985:
CONTINUITY AND DISCONTINUITY

What’s up with the Grimson/Grossberg exchange in the 1983 BBS paper?

PERCEPTION:
Continuity of the perceived surface depth coexists with the discontinuity of the featural identity of the dots (brightness, color) relative to their background.

REVENGE OF THE WEDDING CAKE

Interpolation of depth-from-disparity -- vs. “lifting the full featural landscape” -- whatever THAT means.

Q1: Which surface best meets specified variational constraints (smoothness, flatness) while passing through certain coordinates specified by disparity information?

Q2: How do discrete pockets of activity ever coalesce* to form a coherent (continuous) surface in a medium in the first place?

Note: The answer to Q2 must keep the featural identity of inducing elements (dots) intact; their color does not generally spread.

Q3: Why are rivalry and fusion alternate (i.e, incompatible) modes?

* In some other stereogram (e.g., “leaves swirling in the air”), the disparities of individual elements might not fuse to a surface.

WARPED MIND

Grossberg’s claim: We need a deformable space, in order to sharply fuse disparate segmentations.

“Normal” peak summation is too sloppy, yielding graded or blurred responses.

Segmentations from left eye must be fused with those from right eye OR one eye’s segmentation must suppress the other’s -- nothing in between these outcomes can be tolerated!

Cf. allelotropia -- “displacement”

left eye  binocular  right eye

2 3 4  2 3 4  2 3 4

56 7  5 6 7  5 6 7

STEREOPSIS AND FUSION

Stereopsis ≠ Fusion

You can have an impression of “solidness” (depth from stereo) based on information across two eyes’ views, without fusion, and vice versa.

Rivalry is the (alternating) suppression of data going to one eye by (incompatible) data from the other eye.

No consensus exists about the mechanisms of stereopsis, fusion, and rivalry, or their relation to one another.

Grossberg’s work on 3D figure/ground segmentation depends “muchly” on the observation that one can experience simultaneous fusion and rivalry from a given portion of a scene, and on the premise that this implicates multiple spatial scales, because fusion and rivalry are alternate modes within a scale.
KAUFMAN'S (1974) STEREOGRAM

The edges of the emergent square region are at different disparities from those of the frame.

The square is seen in depth.

Claim: The emergent segmentations are what gets fused, and stereopsis results from the differing relative disparities of inner square and outer frame.

The thin lines in the square region (and in the background region) are rivalrous, because they are at perpendicular orientations in the two eyes' views.

FUSION AND EMERGENT SEGMENTATION

What is the relation of emergent segmentation to binocular vision (stereopsis, fusion, rivalry)? Which comes “first”?

Claim: Illusory edges of the inner square are what get matched.

Warning: One cannot conclude from this example that disparity matching occurs only after emergent segmentation, as fusion and stereopsis can themselves be segmentation procedures -- even the only such procedures in a scene, as in random dot stereograms.

RIVALRY ISSUE 1

How are signals from non-dominant eye suppressed? What is suppressed and where? (Cf. “What are the units . . .?”)

Key data:
Rivalry “survivors” group into spatially coherent regions composed of elements from the same eye; the shape and size of these regions fluctuate over time. (Blake, 1988)

<table>
<thead>
<tr>
<th>left eye</th>
<th>middle eye</th>
<th>right eye</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ / / / /</td>
<td>/ / / /</td>
<td>/ / / / /</td>
</tr>
<tr>
<td>/ / / /</td>
<td>/ / /</td>
<td>/ / / /</td>
</tr>
<tr>
<td>/ / / /</td>
<td>/ / /</td>
<td>/ / / /</td>
</tr>
<tr>
<td>/ / /</td>
<td>/ / /</td>
<td>/ / /</td>
</tr>
<tr>
<td>/ /</td>
<td>/ /</td>
<td>/ /</td>
</tr>
</tbody>
</table>

Local winner is not independent of neighbor, so . . .

Think about: spatially linked, oriented gated dipoles

Piers Howe, recent CNS grad, notes:

“Be careful here. People have claimed that at the edges of the inner square the tips of the V shaped intersections look like diamonds at a suitable small scale. These diamonds form the matching elements in the two stereo half images. Therefore, perhaps, it is not necessarily the emergent segmentations that are fused. See Howard and Rogers for in depth discussion of this and related stereograms.”

I.P. Howard & B.J. Rogers, Seeing in depth: Vol. II. Depth perception.
Toronto: Porteous Publisher.
RIVALRY ISSUE 2

Rivalry is highly nonspecific in orientation.
I.e., rivalrous line segments do not need to be perpendicular.

| + | = fusion
| + / = fusion (with slant in depth)
| + / = rivalry
| + / = rivalry
| + / = rivalry

Since rivalry concerns “visibility”, FCS must be involved; how?

INTERPRETATION OF KAUFMAN STEREOGRAM

With reference to BCS/FCS mechanisms:

1) Why rivalry? -- Cross-orientation inhibition at second competitive stage forces choice of one boundary orientation (from left eye or right eye, but not both).

2) End cut machinery forms inducers for illusory square at corners of the pattern.

3) The completion of the square to be fused and the coherence of rivalry within regions are mediated by cooperative bipoles.

4) BCS “becomes binocular” before FCS; monocular featural signals corresponding to suppressed boundary must be killed.

BINOCULAR THEORY

FACADE Theory:
Form And Color and DEpth

Mismatch of binocularly chosen boundary orientation (path a) and monocular featural information (path b) results in suppression.

THE Bs

Consider: Bregman/Kanizsa Bs:

Seeing occluding object (B or D) helps you recognize the segments of the occluded B’s (C).
Amodal completion (Michotte): Seeing parts as belonging to the same object, behind an occluder. (Cases B and D)
FACADE: FUSABLE DISPARITIES

Size-disparity correlation:
- small
- large

cooperation across scale and position within disparity
to select positions that best match image data

competition within scale and position across disparity
to select the best matched disparity at each position

T-JUNCTION SENSITIVITY

Without T-Junction Detectors

Grossberg, 1994, 1997

The top of the T is assigned to the NEAR depth
The stem of the T is assigned to the FAR depth

Long-Range Cooperation:
(+ ) Bipole Cells
Short-Range Competition:
(- ) Hypercomplex Cells

Simulation of FACADE Figure-Ground Segmentation
from 2D image

from Kelly & Grossberg, P & P, 2000

USE:
- End-gaps at T-junctions
- Boundary grouping (collinear)
- Filling-in of connected regions
- boundary/feature consistency
- Size-disparity correlation
- Near-to-far asymmetry
- boundary combination/suppression (a.k.a. “enrichment/pruning”)

GET:
- Modal and amodal percepts
- Proper depth ordering in percept
**AMODAL PERCEPT**

boundary signals  *amodal* surface percept

| near | near |
| far | far |

**FORMATION OF BINOCULAR SURFACE PERCEPT**

boundary signals to binocular surface stage  filling-in signals to binocular surface stage

| near | near |
| far | far |

**MODAL SURFACE PERCEPT**

binocular surface stage -- final filled-in values  *modal* surface percept

| near | far |

**The Model of Grossberg & Howe (2003)**

Is the convergence of the LAMINART and FACADE series of models

---

Input image
**G & H 2003 Continued**

Can explain in a biologically plausible manner:

- Dichoptic masking
- Panum's limiting case
- Contrast variations of the correspondence problem
- The affect of interocular contrast differences on stereoaucity
- Venetian blind effect
- Stereopsis with opposite-contrast stimuli
- Da Vinci stereopsis
- Various lightness illusions

A total of 20 separate simulations with just one set of parameters!

**FANG AND GROSSBERG**

Random dot stereogram     Julesz, 1971

Simulation results (left to right: near, fixation, far depth planes):

Successor to G & H 2003 model -- compare this simulation with Grimson's wedding cake!

**ARASH Y’S ADDITIONS**

1- Depth can be seen in simple stereograms exposed for only 1 ms!
2- It takes time to process the information in a stereogram after a briefly (1 ms) exposed stimulus. Julesz (1964) obtained an estimate for this processing time: 50 ms.
3- (Some) da Vinci stereopsis can be explained by Panum’s limiting case(1858).
4- Alhazen (1083) described the “Panum's” Fusional Area by a systematic measurement with the final statement that “small differences in visual angle can be tolerated without diplopia.”
5- Note in both da Vinci stereopsis and Classic stereopsis the eye of origin information is critical: the stereopsis (or depth percept from non-matching cues) has to begin not later than V1, where the eye of origin information is still present.

**COMPLEMENTARITY AND CONSISTENCY**

From complementary computations to consistent representations:

BCS/FCS boundary-gated filling-in generalizes in FACADE theory to explain modal and amodal perception of surfaces in depth, figure/ground separation, and more!

**Conclusions:**

There’s much more to binocular vision than “working out the projective geometry” of light entering the two eyes.

There’s more to stereopsis than matching features common to two eyes’ views.

How things appear in depth and “how things appear,” in the sense of color and brightness, are closely linked questions.