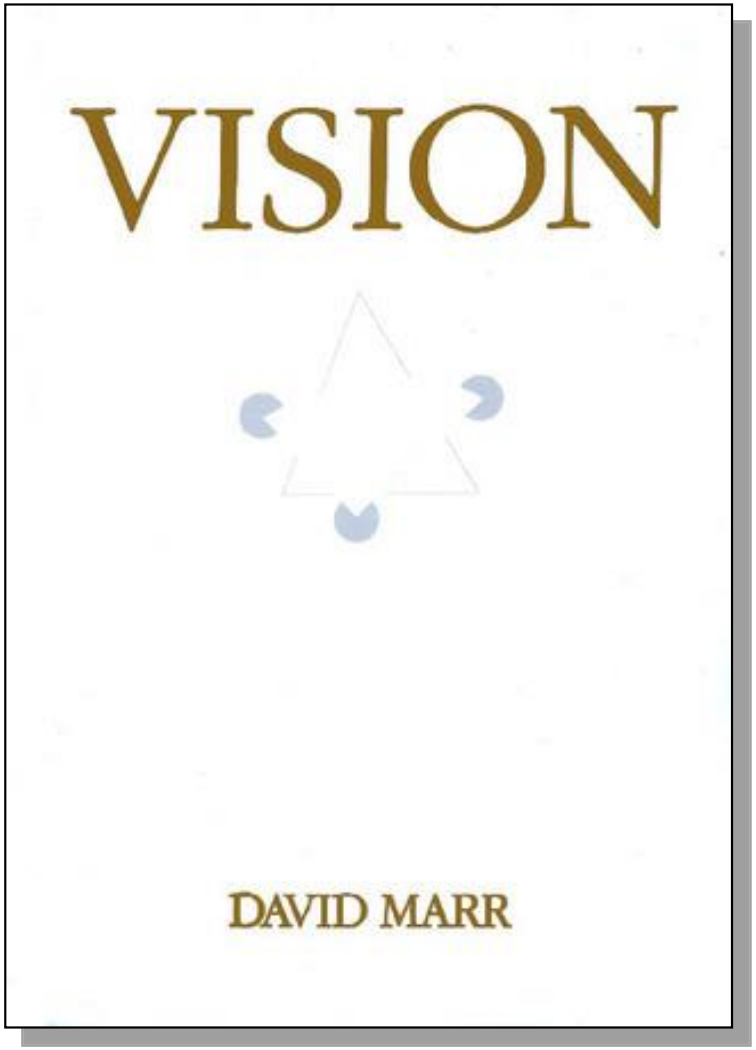
First steps in understanding vision

- The signal (light) and the sensor (eye)
- Processing the signal

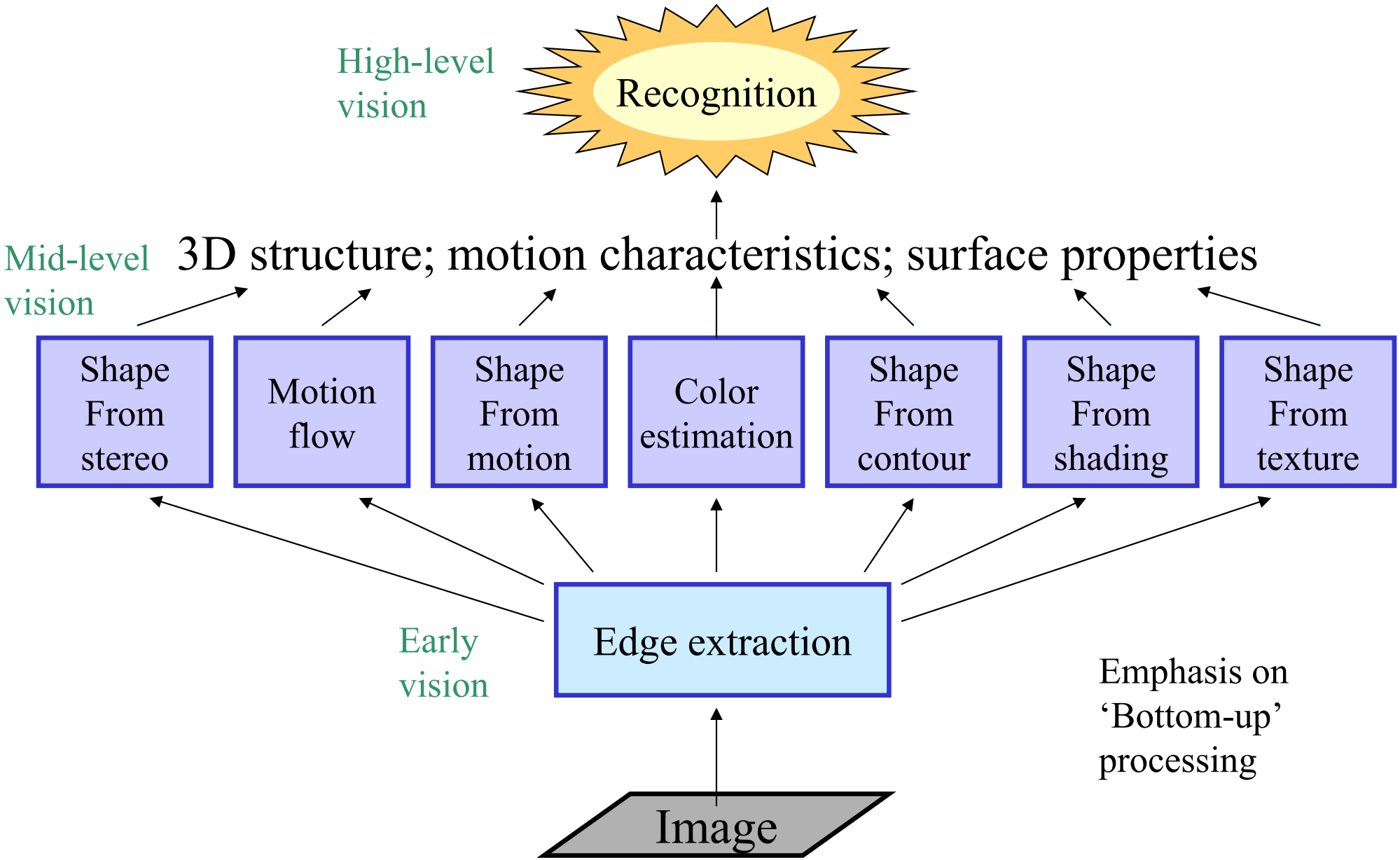
MASSACHVSETTS INSTITVTE OF TECHNOLOG
David Marr
1945-1980



"It was as if vision research suddenly acquired its own Principia Mathematica, or perhaps General Relativity Theory, in terms of the new explanatory power Marr's theories provided. Truly an extraordinary book from an extraordinary thinker in the area of perception, vision, and the brain."

Processing Framework Proposed by Marr

Parallel and Hierarchical

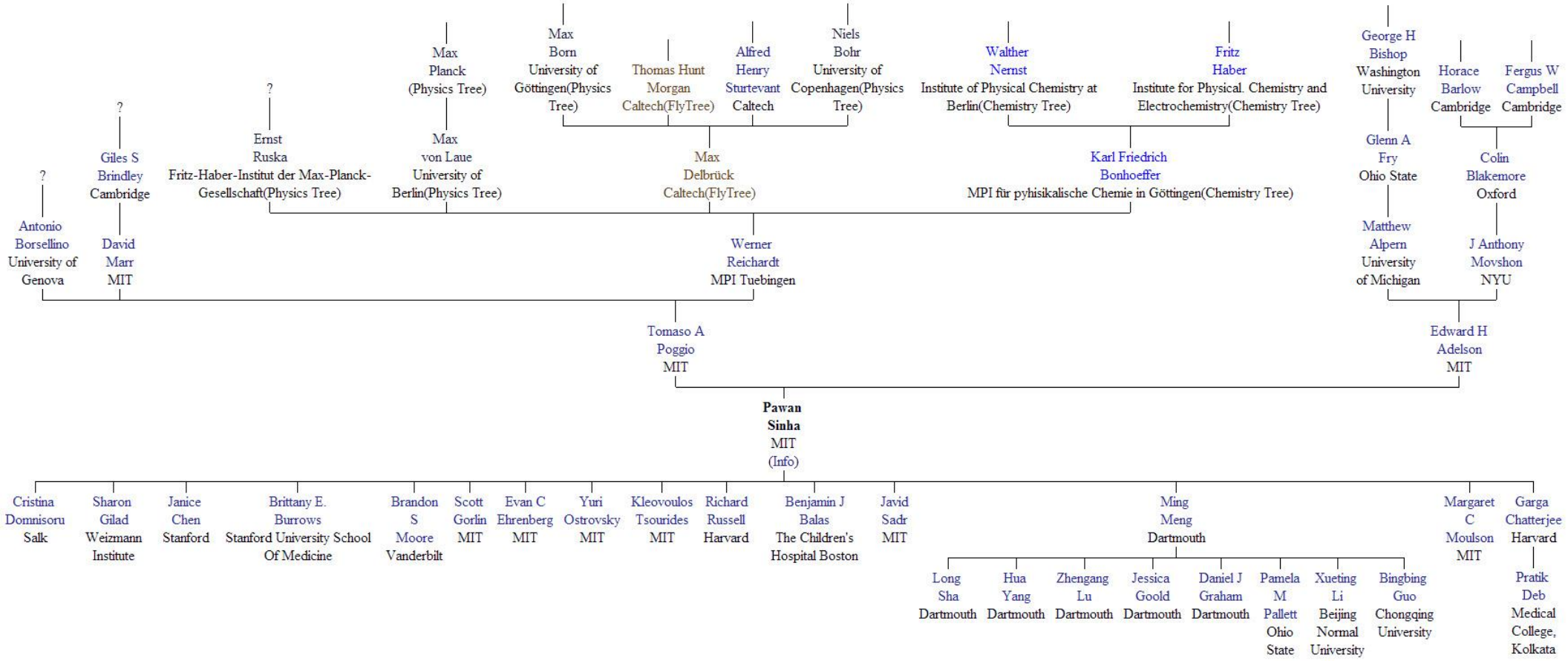


Emphasis on 'Bottom-up' processing

Tangent: Who was David Marr's Ph.D. advisor?

Dr. Giles Brindley

"The Ratio Club" meeting at Cambridge held May 2-3, 1952


Max Planck (Physics Tree)
Max Born University of Göttingen(Physics Tree)
Thomas Hunt Morgan Caltech(FlyTree)
Alfred Henry Sturtevant Caltech
Niels Bohr University of Copenhagen(Physics Tree)
Walther Nernst Institute of Physical Chemistry at Berlin(Chemistry Tree)
Fritz Haber Institute for Physical. Chemistry and Electrochemistry(Chemistry Tree)
George H Bishop Washington University
Horace Barlow Cambridge
Fergus W Campbell Cambridge
?
Ernst Ruska Fritz-Haber-Institut der Max-Planck-Gesellschaft(Physics Tree)
Max von Laue University of Berlin(Physics Tree)
Max Delbrück Caltech(FlyTree)
Karl Friedrich Bonhoeffer MPI für pyhisikalische Chemie in Göttingen(Chemistry Tree)
Glenn A Fry Ohio State
Colin Blakemore Oxford
?
Giles S Brindley Cambridge
?
Antonio Borsellino University of Genova
David Marr MIT
Werner Reichardt MPI Tuebingen
Matthew Alpern University of Michigan
J Anthony Movshon NYU
Tomaso A Poggio MIT
Edward H Adelson MIT
Pawan Sinha MIT (Info)
Cristina Domnisoru Salk
Sharon Gilad Weizmann Institute
Janice Chen Stanford
Brittany E. Burrows Stanford University School Of Medicine
Brandon S Moore Vanderbilt
Scott Gorlin MIT
Evan C Ehrenberg MIT
Yuri Ostrovsky MIT
Kleovoulos Tsourides MIT
Richard Russell Harvard
Benjamin J Balas The Children's Hospital Boston
Javid Sadr MIT
Ming Meng Dartmouth
Long Sha Dartmouth
Hua Yang Dartmouth
Zhengang Lu Dartmouth
Jessica Goold Dartmouth
Daniel J Graham Dartmouth
Pamela M Pallett Ohio State
Xueting Li Beijing Normal University
Bingbing Guo Chongqing University
Margaret C Moulson MIT
Garga Chatterjee Harvard
Pratik Deb Medical College, Kolkata

J. Physiol. (1968), **196**, *pp.* 479–493
With 3 *plates and* 4 *text-figures*
Printed in Great Britain

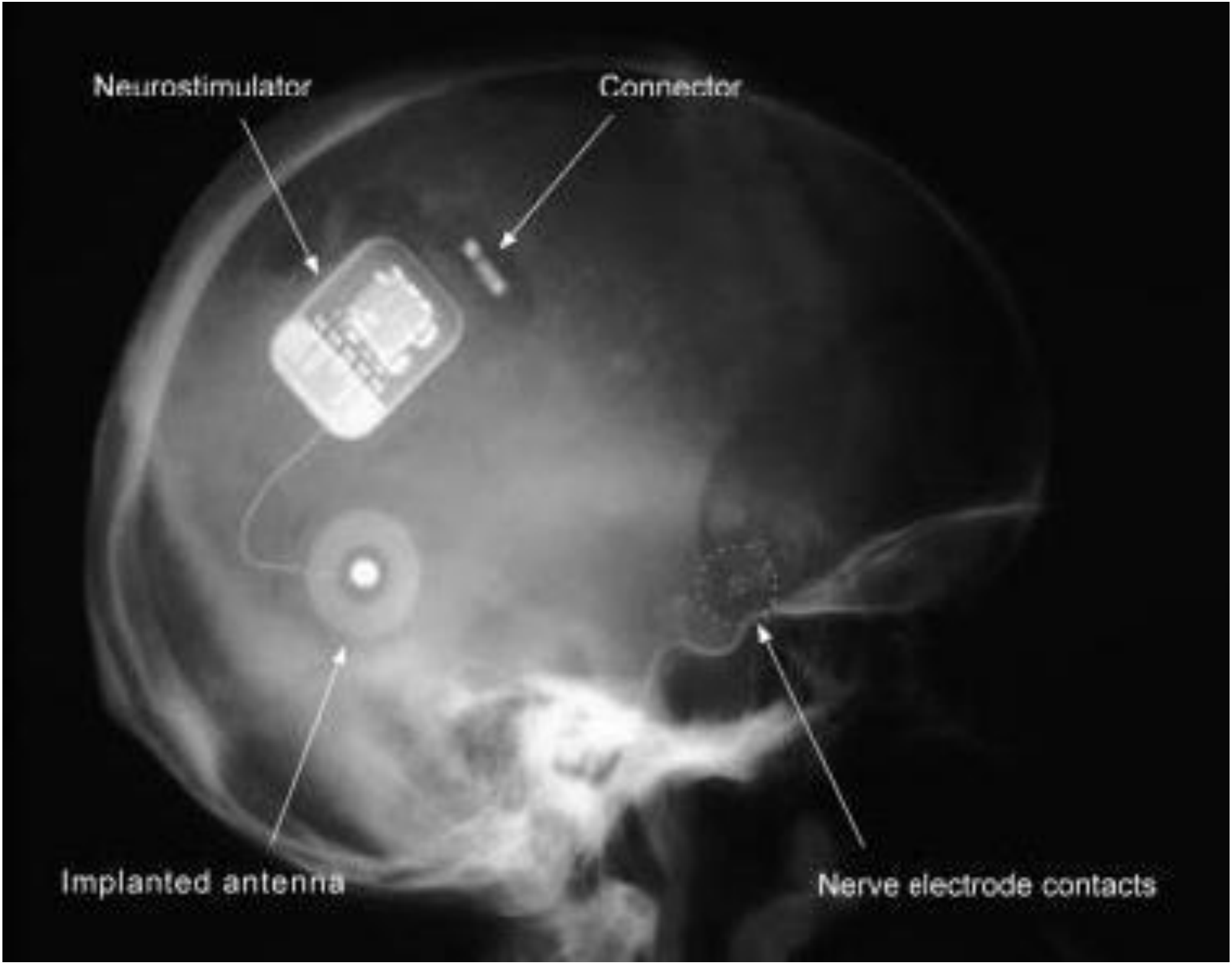
THE SENSATIONS PRODUCED BY ELECTRICAL STIMULATION OF THE VISUAL CORTEX

BY G. S. BRINDLEY AND W. S. LEWIN
From the Physiological Laboratory, University of Cambridge and the Department of Neurological Surgery and Neurology, United Cambridge Hospitals

(*Received* 29 *January* 1968)

SUMMARY

1. An array of radio receivers, connected to electrodes in contact with the occipital pole of the right cerebral hemisphere, has been implanted into a 52-year-old blind patient. By giving appropriate radio signals, the patient can be caused to experience sensations of light ('phosphenes') in the left half of the visual field.


Neurostimulator
Connector
Implanted antenna
Nerve electrode contacts

An Eyewitness Account Of The Most Awkward Urology Lecture Ever

At the 1983 Urodynamics Society conference in Las Vegas, pioneering urologist Sir Giles Skey Brindley delivered what was singularly the most uncomfortable lecture in the history of academia. What began as a presentation on the treatment of erectile dysfunction quickly transformed into an unwanted session of show-and-tell.

HOW (NOT) TO COMMUNICATE NEW SCIENTIFIC INFORMATION: A MEMOIR OF THE FAMOUS BRINDLEY LECTURE

LAURENCE KLOTZ –
Sunnybrook and Women's College Health Sciences Centre, University of Toronto, Toronto, Canada

Accepted for publication 22 July 2005

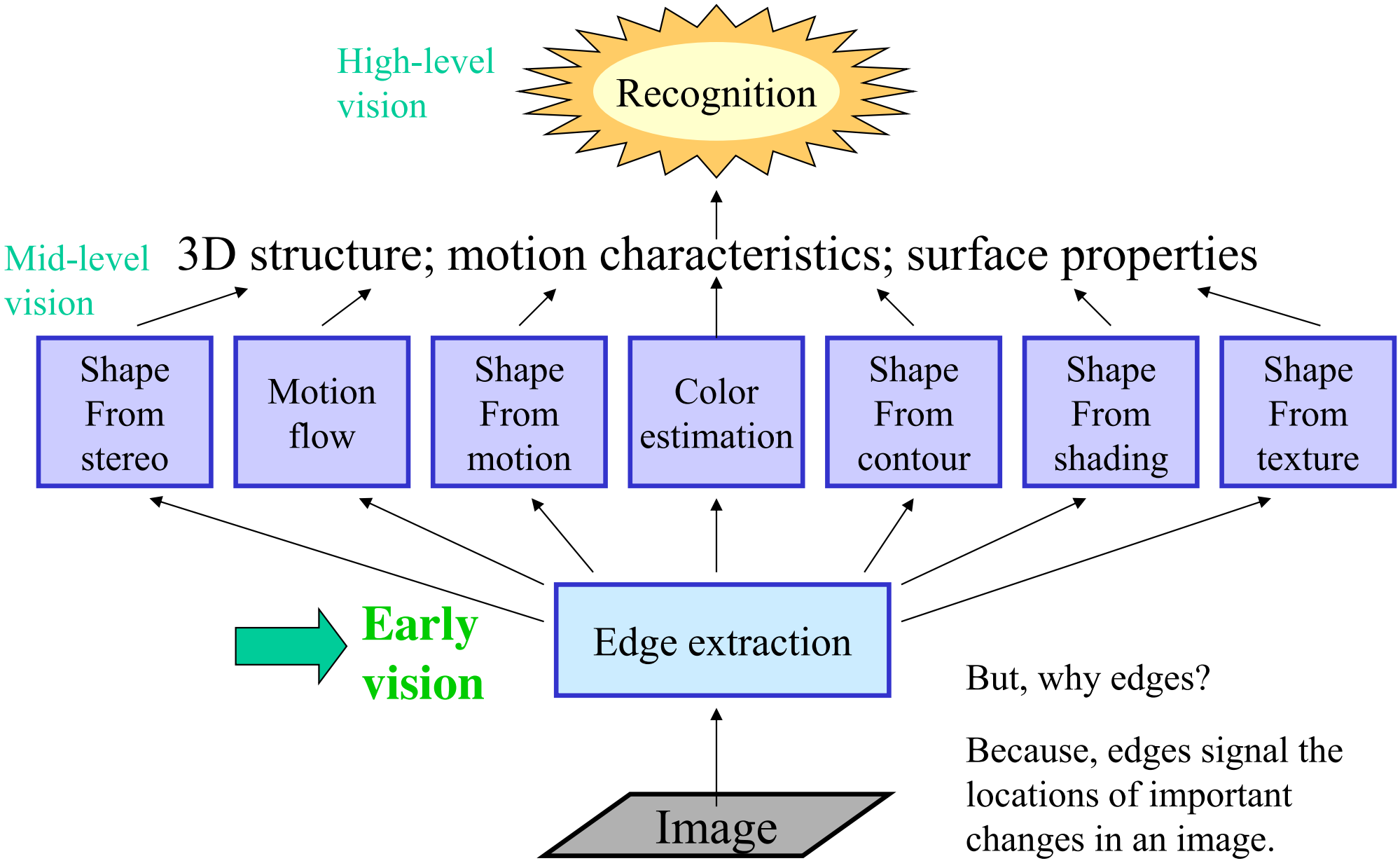
In 1983, at the Urodynamics Society meeting in Las Vegas, Professor G.S. Brindley first announced to the world his experiments on self-injection with papaverine to induce a penile erection. This was the first time that which became crowded, and started examining and ruffling through the 35 mm slides of micrographs inside. I was standing next to him, and could vaguely make out the content of the slides, which appeared to be a

Outline of lectures

Lecture 1	The signal and the sensory organ
	Early vision (edge detection)
Lecture 2	Mid-level vision (Stereo, motion and color)
	High-level influences on early vision
Lecture 3	High-level vision (object recognition)
	Face perception
Lecture 4	Attention and awareness
	Cross-modal perception
Lecture 5	Visual disorders
	Visual imagery and development (student run)
Lecture 6	Hilbert problems in vision

Jacob, Tuan, Sarah, Julia, Olga

Processing Framework Proposed by Marr



But, why edges?

Because, edges signal the locations of important changes in an image.

The significance of edges

Edges signal the locations of important changes in an image.

Depth discontinuity
(Object border)

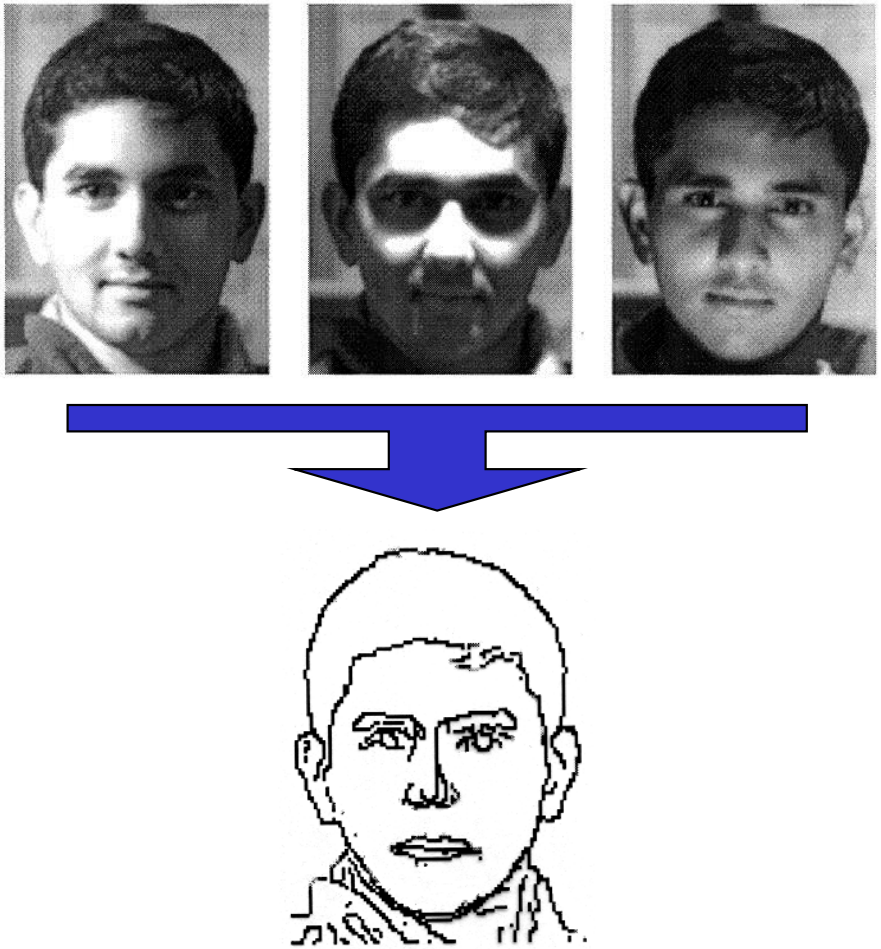
Orientation discontinuity
(Object shape)

Reflectance discontinuity
(Object property)

Edges are believed to play a key role in recognition, because…

The significance of edges (contd.)

Edges are expected to be **invariant** to imaging variations.

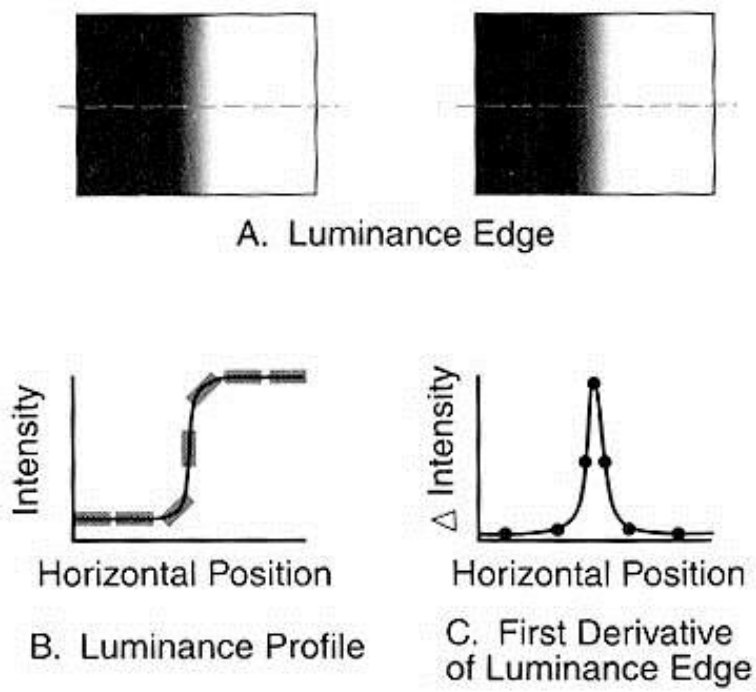


In principle, this can make the subsequent processing very easy.

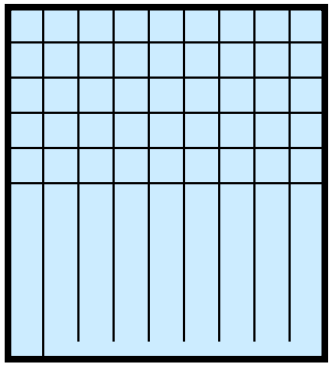
But, what exactly is an ‘edge’ and how can we detect it?…

What is an edge?

- a point at which image luminance (I) changes steeply
- a point at which the first derivative of I has a peak

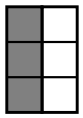


Detecting edges



Image

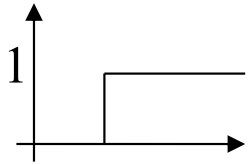
Convolution
(dot-products all over the image)



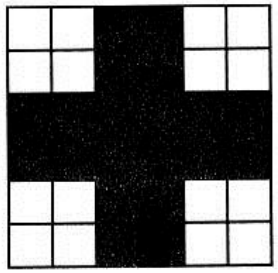
Edge operator

Grid of numbers
Denoting edge
Strength at each
Point in image

Thresholding



Edge map



A. Grayscale Image

10	10	02	02	10	10
10	10	02	02	10	10
02	02	02	02	02	02
02	02	02	02	02	02
10	10	02	02	10	10
10	10	02	02	10	10

B. Image Intensities

-1	+1

C. Vertical Edge Operator

+1
-1

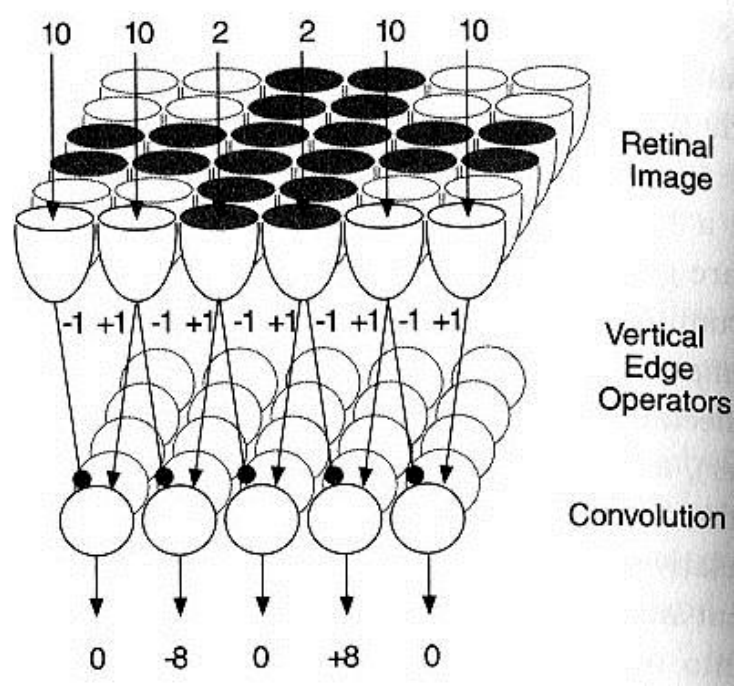
D. Horizontal Edge Operator

0	-8	0	+8	0
0	-8	0	+8	0
0	0	0	0	0
0	0	0	0	0
0	-8	0	+8	0
0	-8	0	+8	0

E. Convolution of Image with Vertical Edge Operator

0	0	0	0	0	0
+8	+8	0	0	+8	+8
0	0	0	0	0	0
-8	-8	0	0	-8	-8
0	0	0	0	0	0

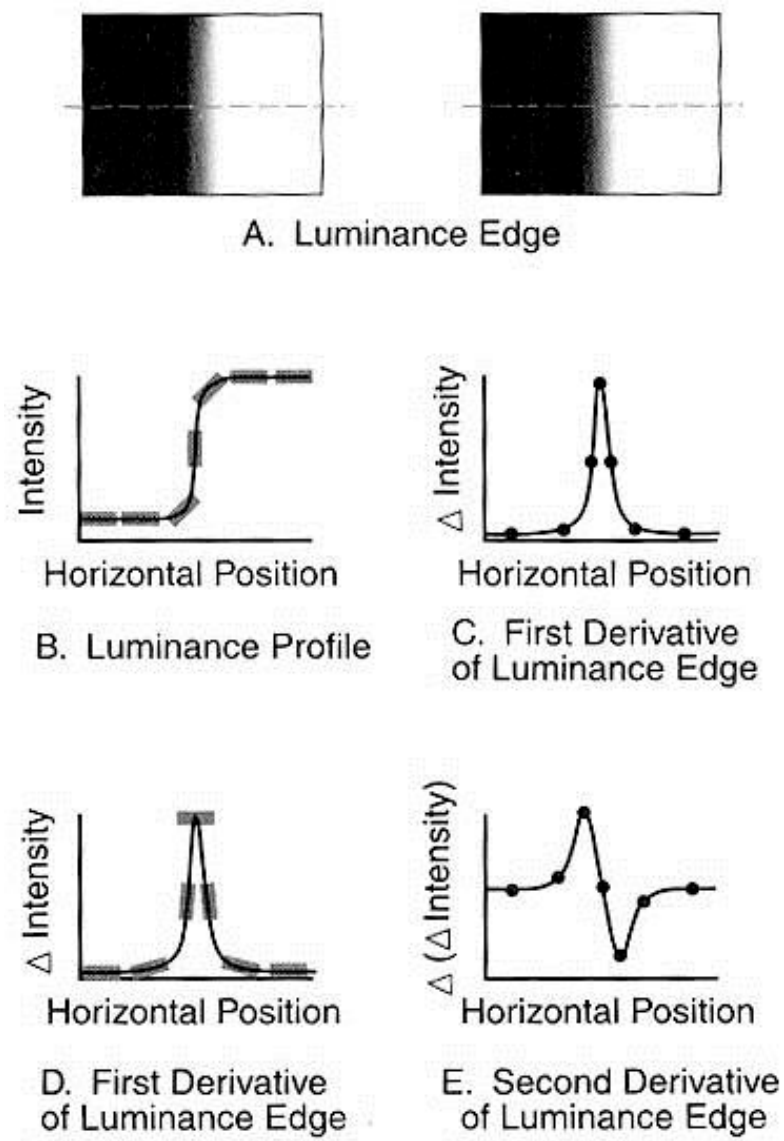
F. Convolution of Image with Horizontal Edge Operator



Network implementation of convolution

What is an edge?

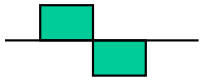
- a point at which image luminance (I) changes steeply
- a point at which the first derivative of I has a peak
- a point at which the second derivative of I has a zero crossing



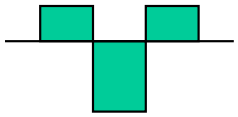
Second order differential operators

Image A B C D

First differences (A-B) (B-C) (C-D)



Second differences (A-2B+C) (B-2C+D)

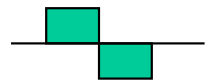


Why would we want to use second order operators rather than first order ones?

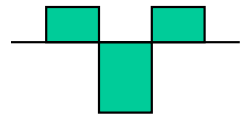
Second order differential operators

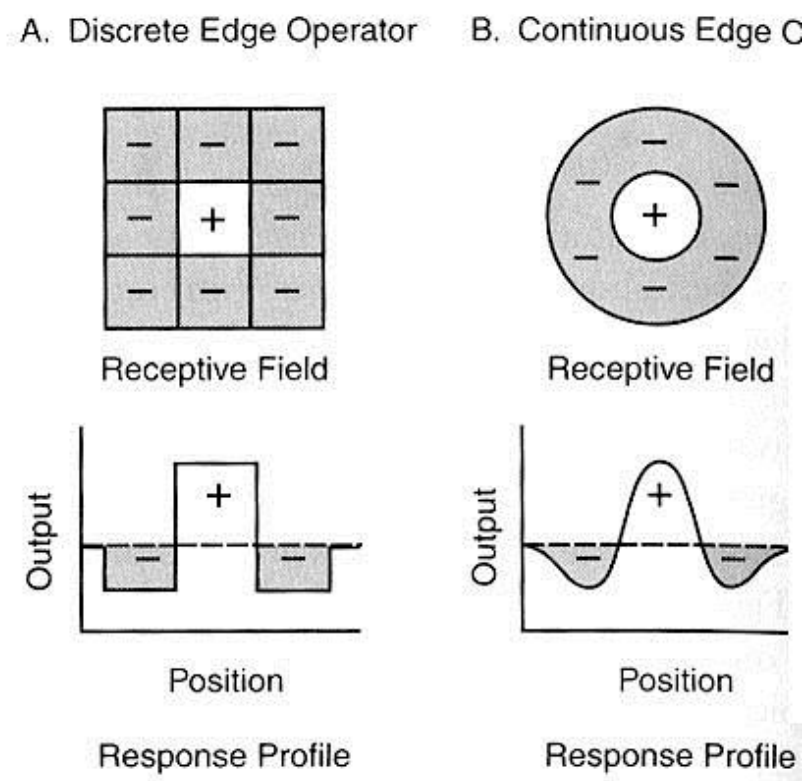
Image A B C D

First differences (A-B) (B-C) (C-D)

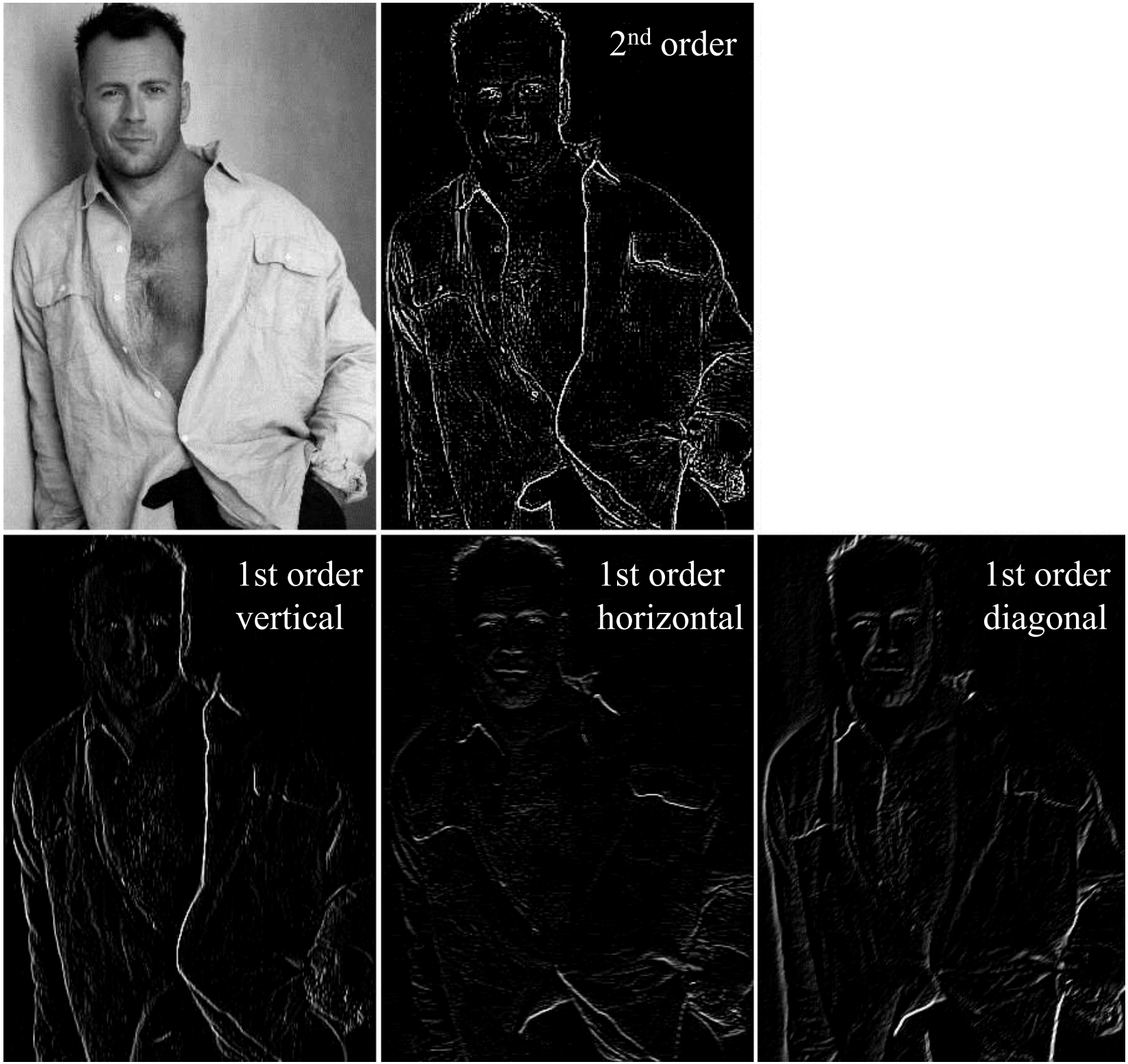


Second differences (A-2B+C) (B-2C+D)

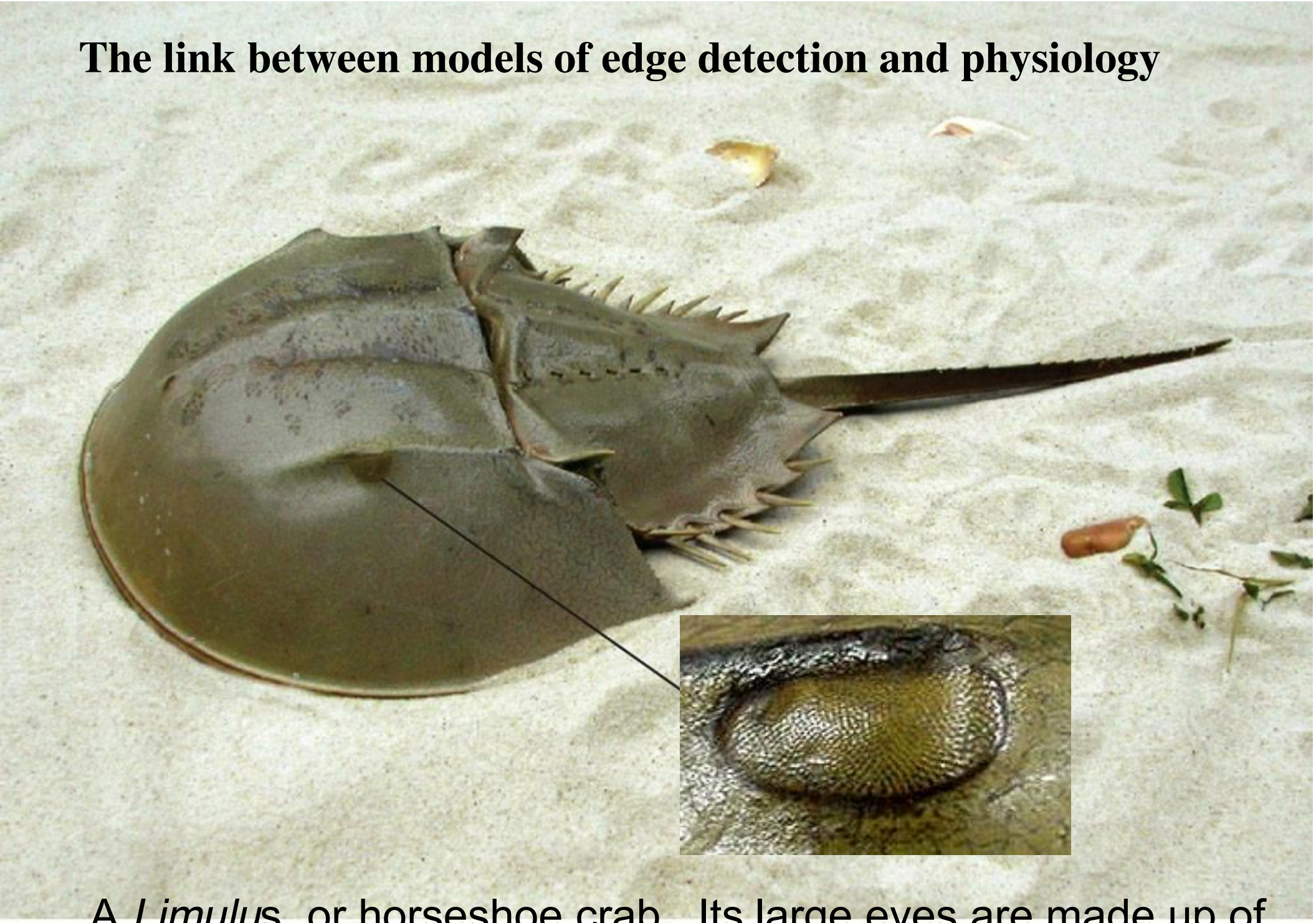




Zero crossings can be detected with Circularly symmetric Filters! (orientation independence)


2nd order
1st order vertical
1st order horizontal
1st order diagonal

The link between models of edge detection and physiology



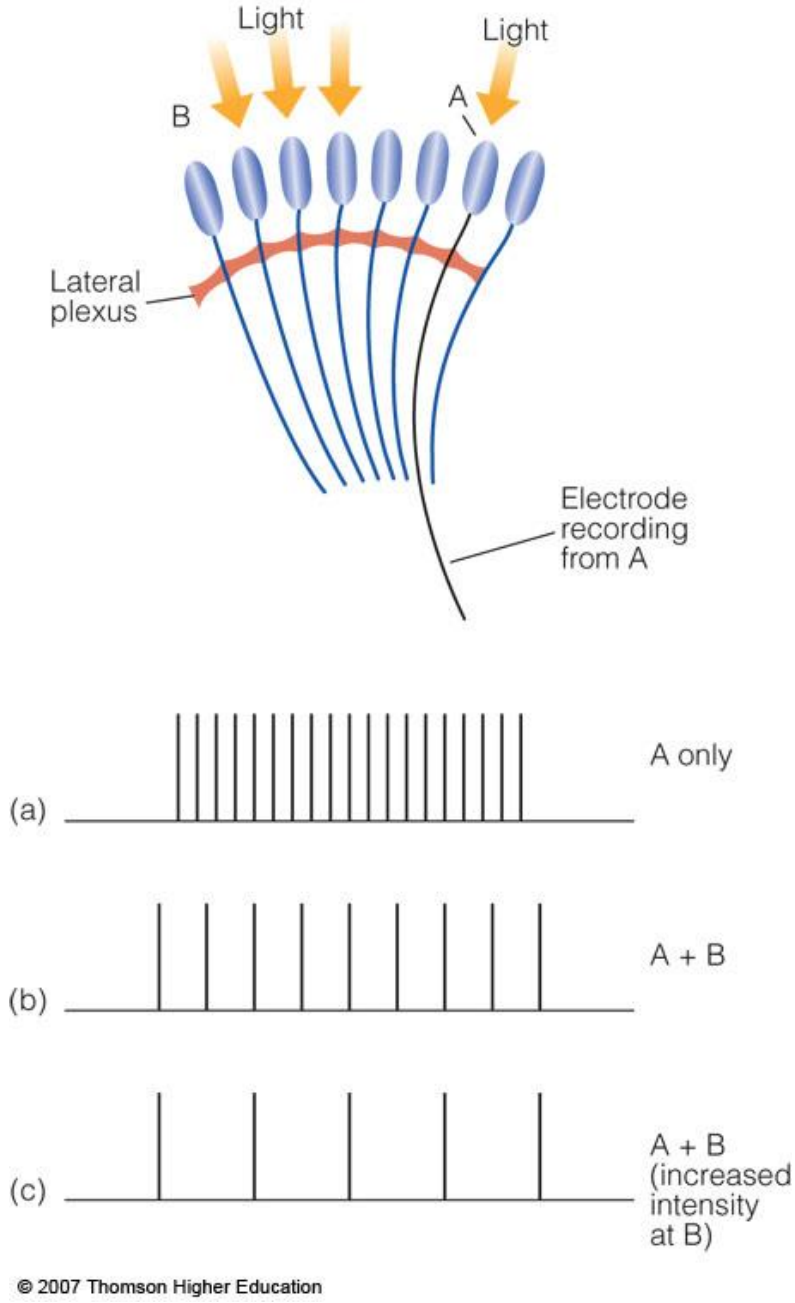
A *Limulus*, or horseshoe crab. Its large eyes are made up of hundreds of ommatidia, each containing a single receptor.

Haldan K. Hartline

Nobel Prize, 1967

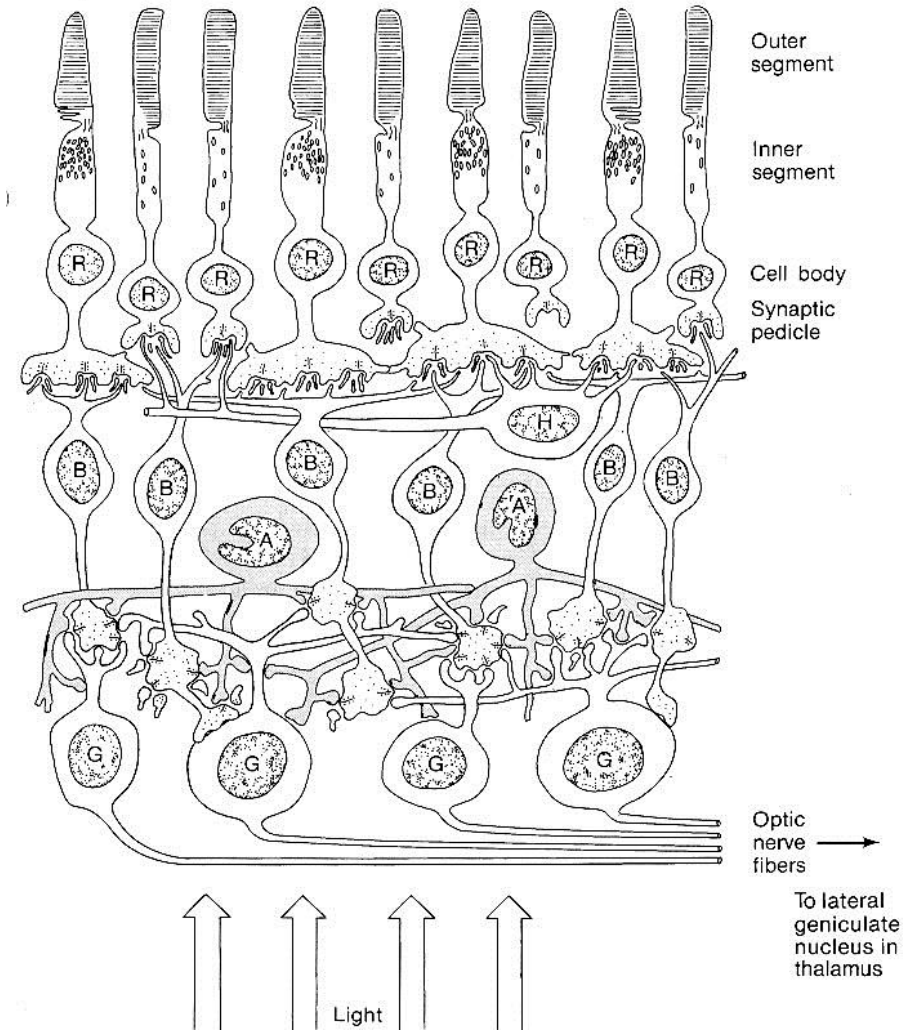
Lateral Inhibition of Neurons

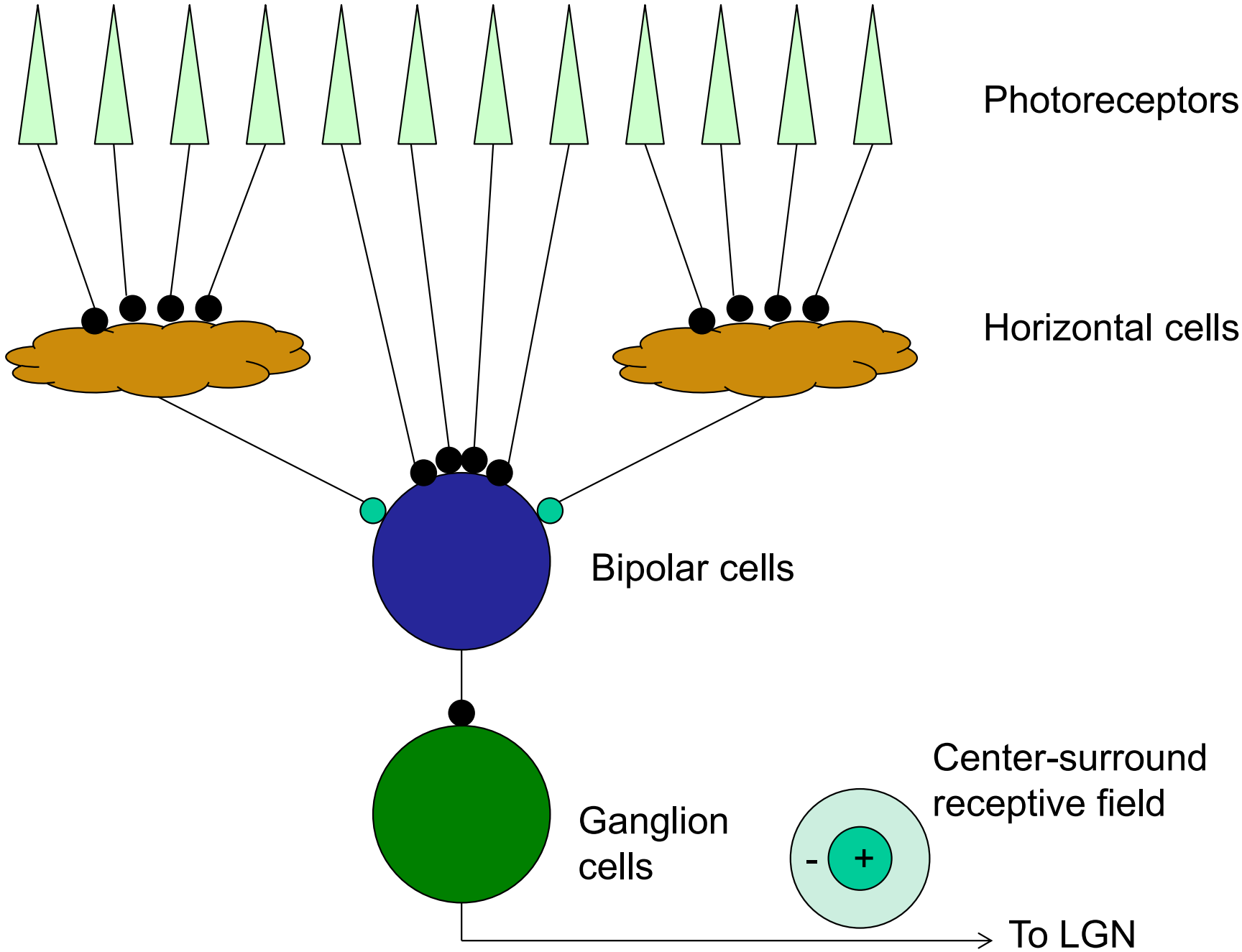
- Hartline and Ratliff's experiments with eye of *Limulus*
  - Ommatidia allow recordings from a single receptor
  - Light shown into a single receptor led to rapid firing rate of nerve fiber
  - Adding light into neighboring receptors led to reduced firing rate of initial nerve fiber



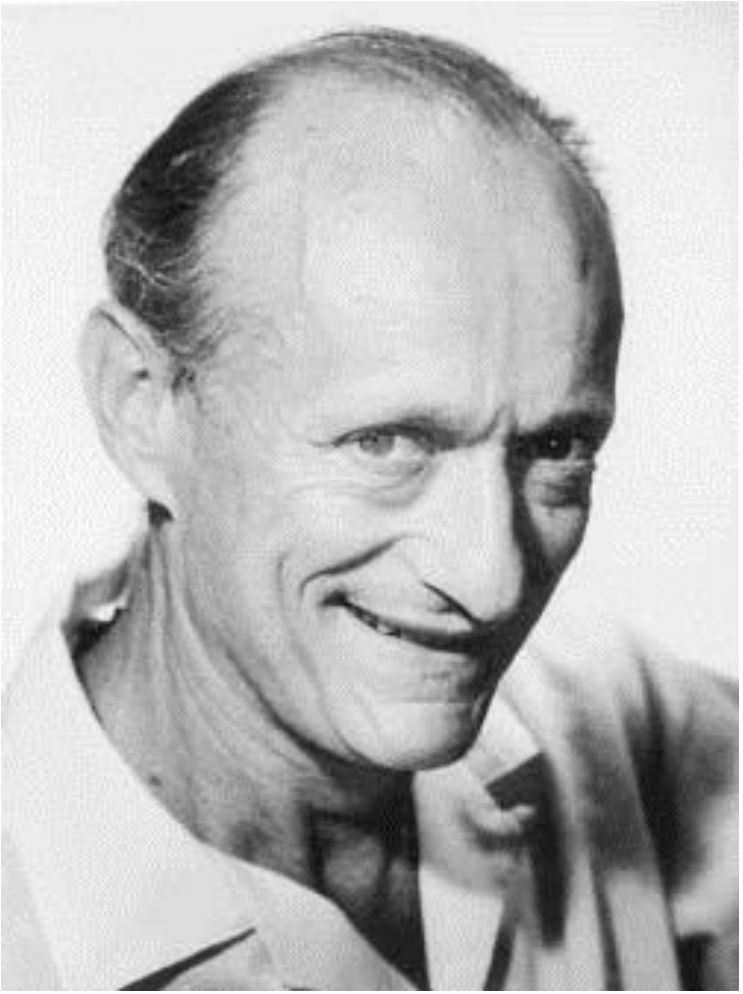
Lateral inhibition is also observed in mammalian retinas

Mammalian retina




Photoreceptors
Horizontal cells
Bipolar cells
Ganglion cells
Center-surround receptive field
-
+
To LGN

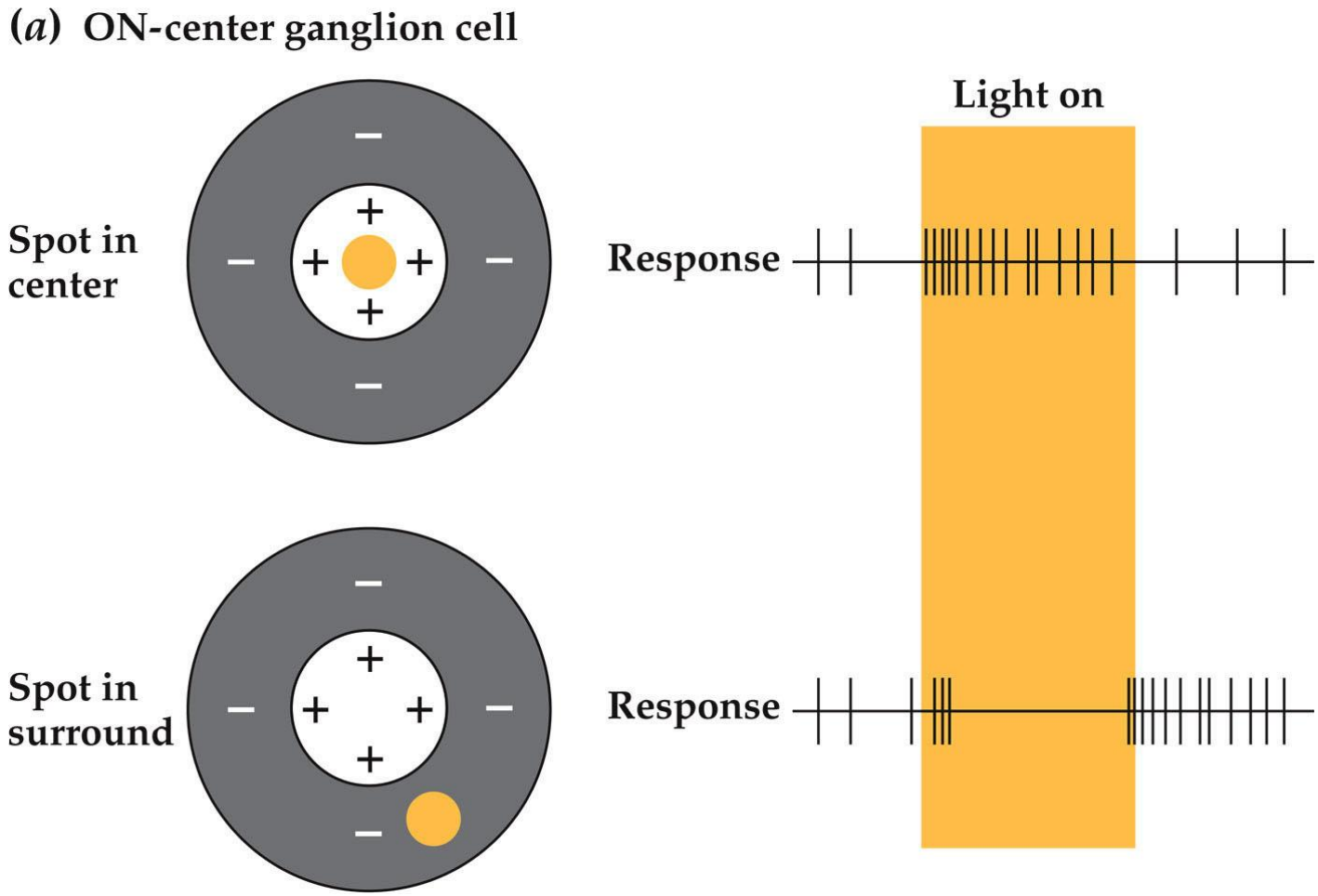
Response properties of ganglion cells



Stephen W. Kuffler
1913-1980

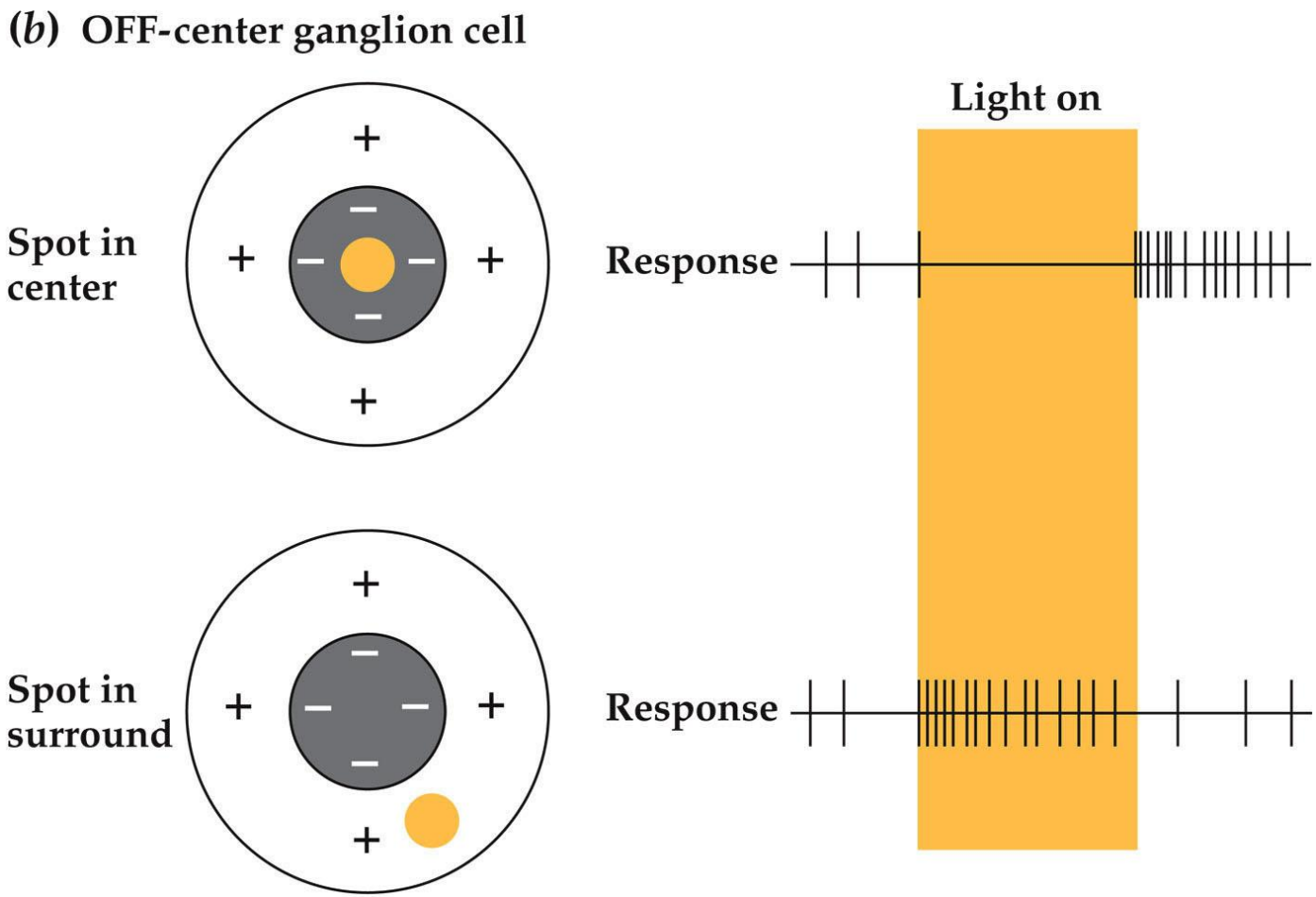
- Kuffler mapped out the receptive fields of individual retinal ganglion cells in the cat and found two kinds of cells
- ON-center ganglion cells
- OFF-center ganglion cells

Retinal ganglion cell receptive fields



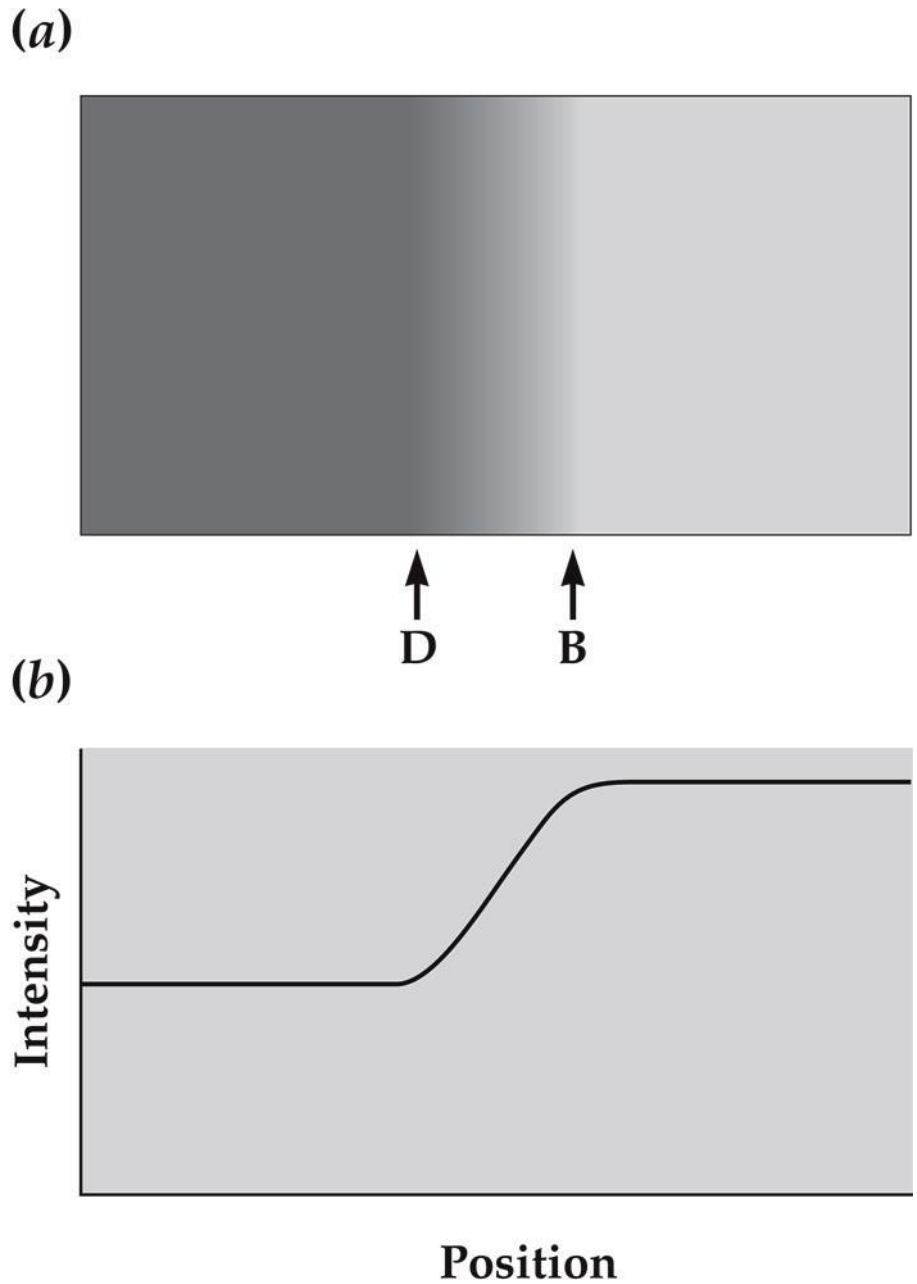
Excited by light falling on center, inhibited by light falling on surround

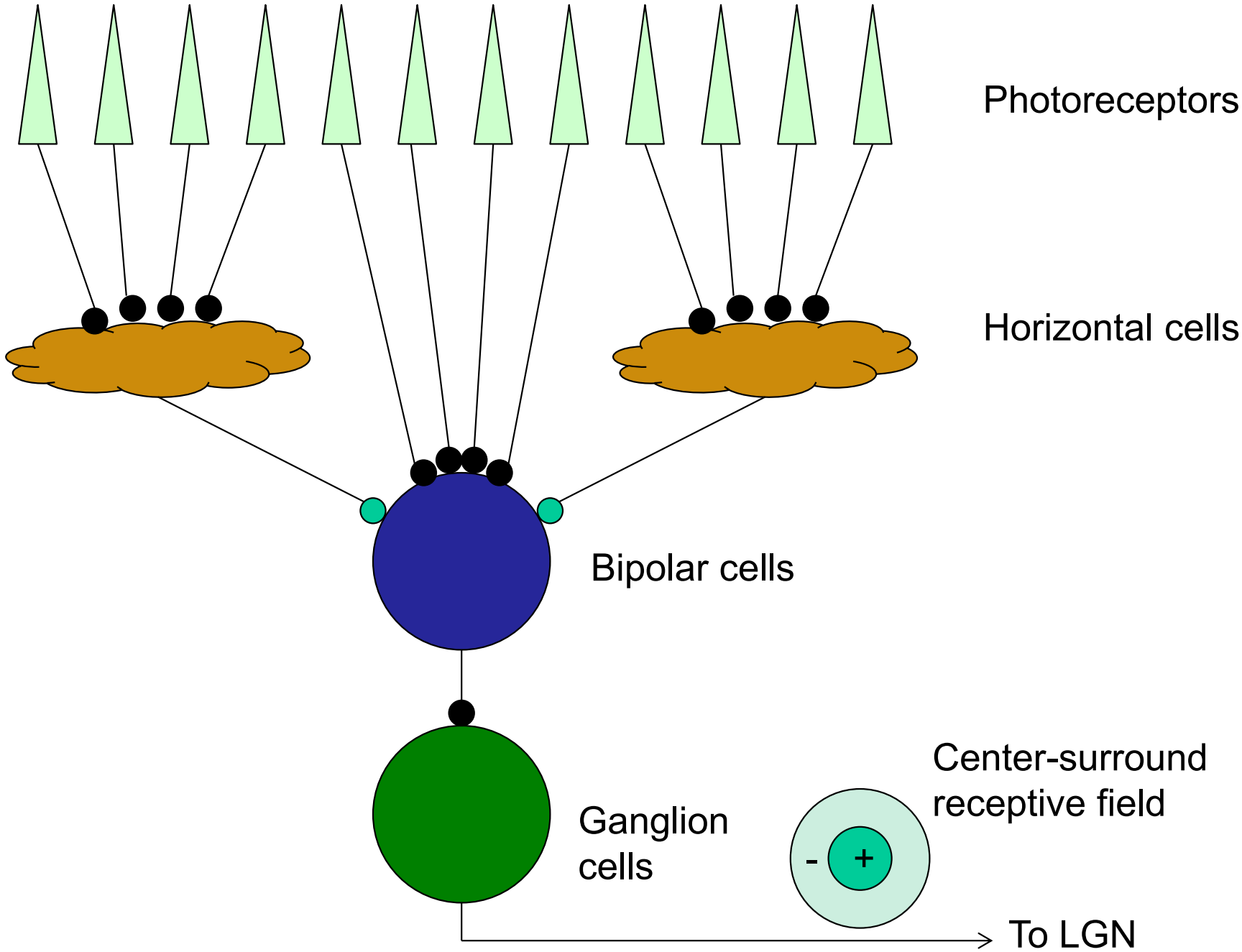
Retinal ganglion cell receptive fields



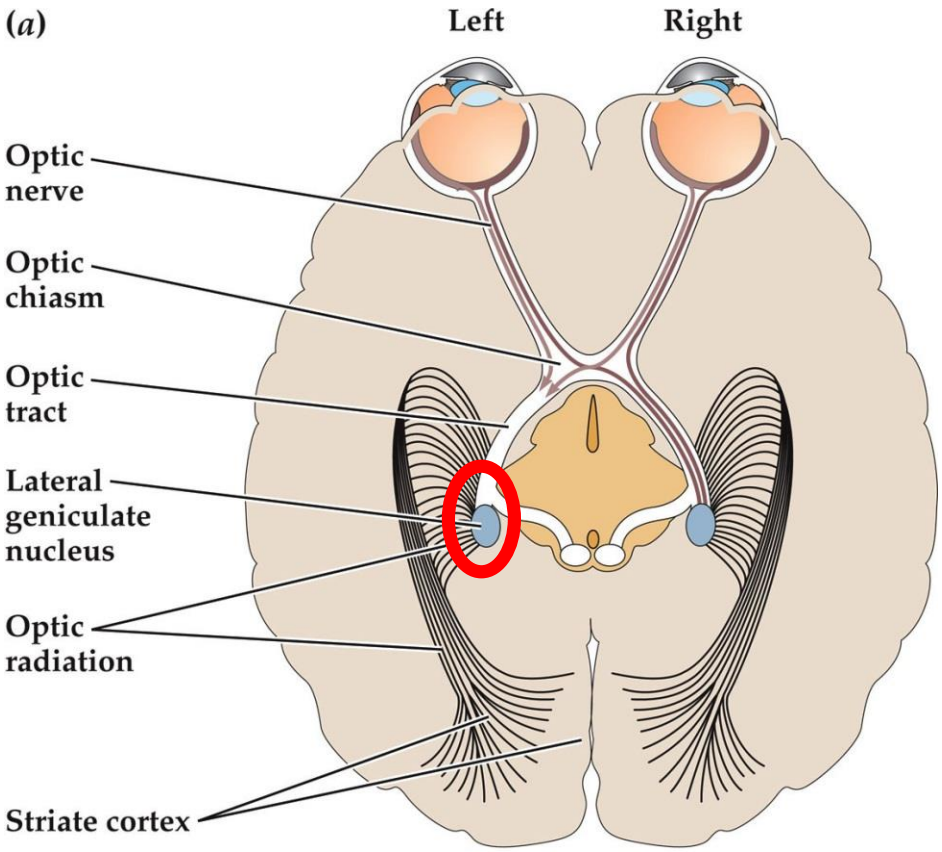
Inhibited by light falling on center, excited by light falling on surround

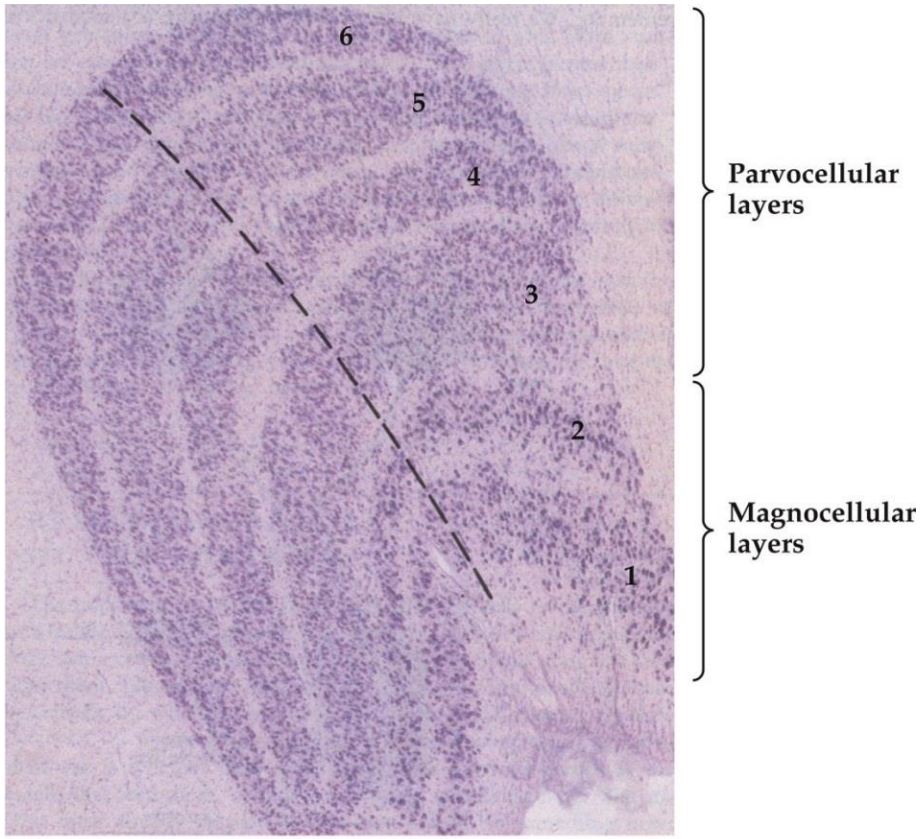
Understanding Mach bands




Photoreceptors
Horizontal cells
Bipolar cells
Ganglion cells
Center-surround receptive field
-
+
To LGN

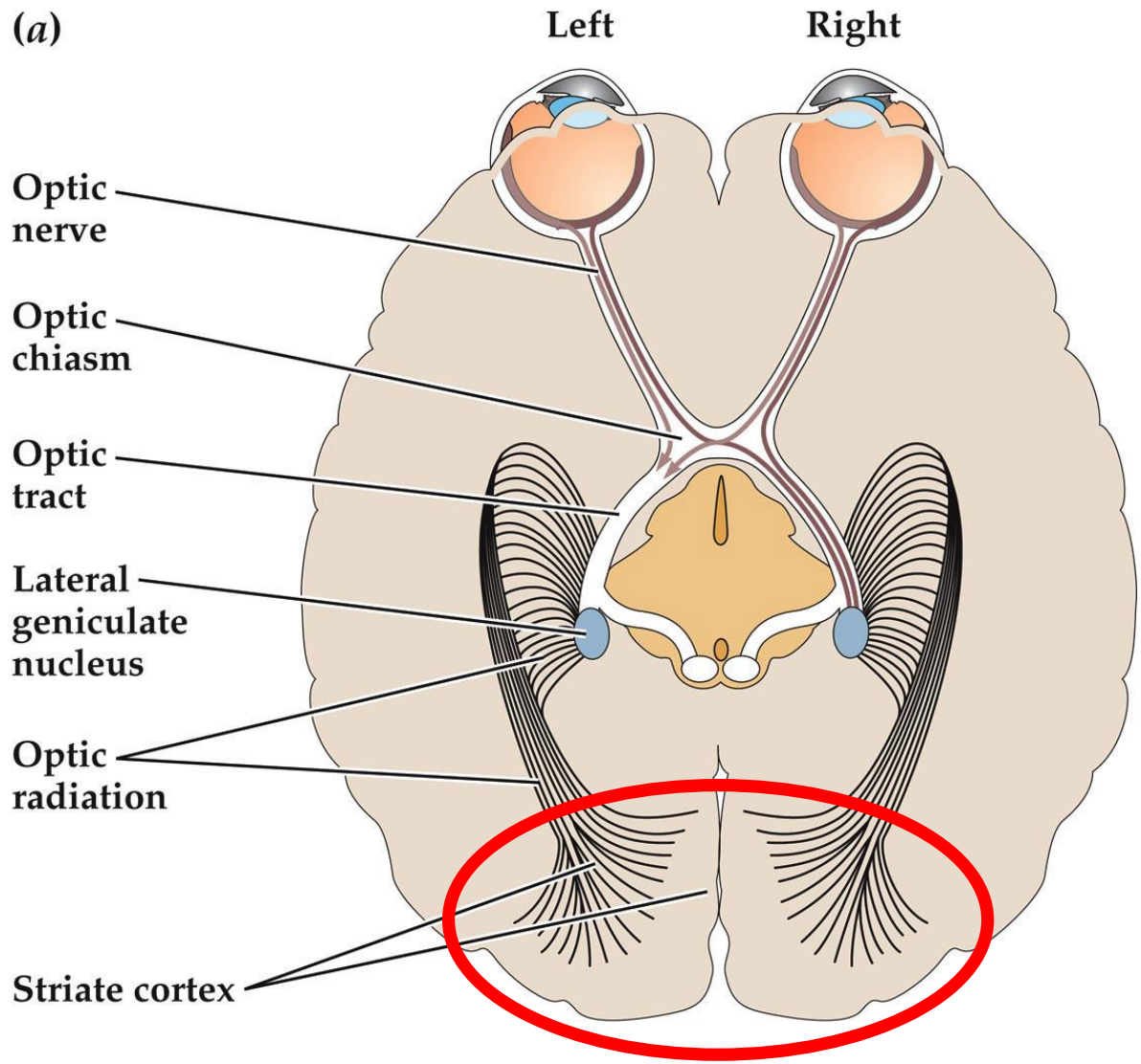
Central visual pathways





The LGN is partly a relay station and partly a location for high-level influences on inputs from the eyes

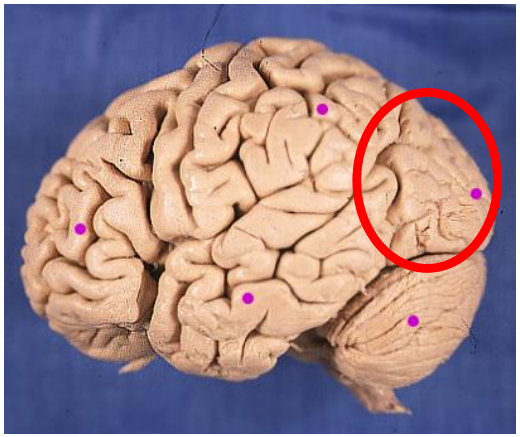
Central visual pathways



200 million cells!

Striate Cortex

- Striate cortex: Also known as primary visual cortex, area 17, or V1
- A major transformation of visual information takes place in striate cortex
  - Circular receptive fields found in retina and LGN are replaced with elongated “stripe” receptive fields in cortex



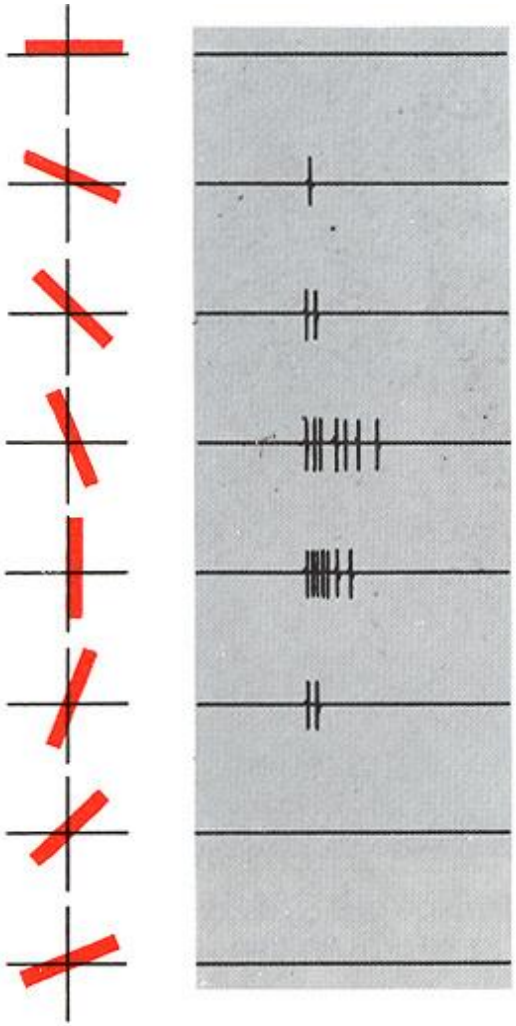
Primary
Visual
Cortex

Oriented bar and edge detector neurons

VIDEOS

https://www.youtube.com/watch?v=IOHayh06LJ4

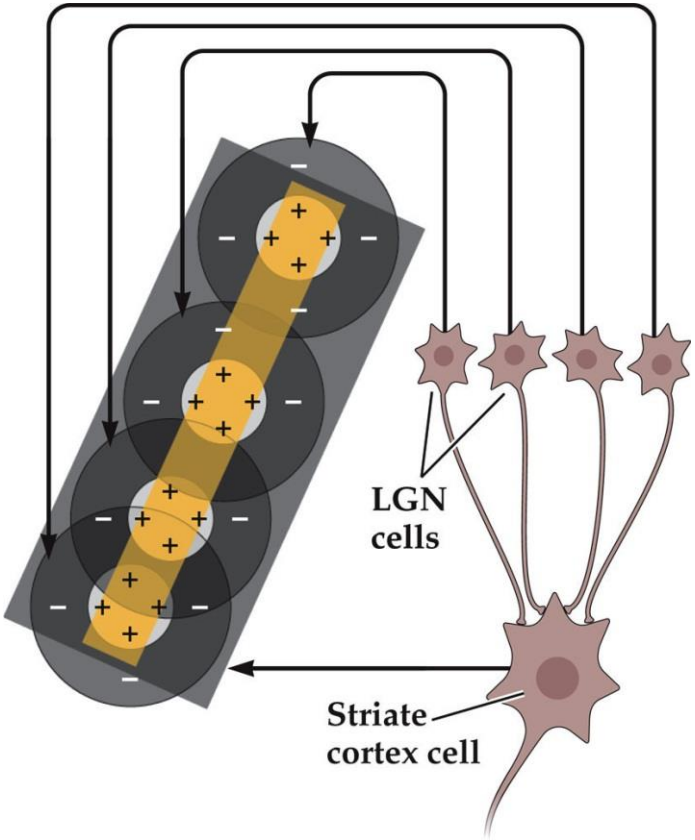
https://www.youtube.com/watch?v=Cw5PKV9Rj3o



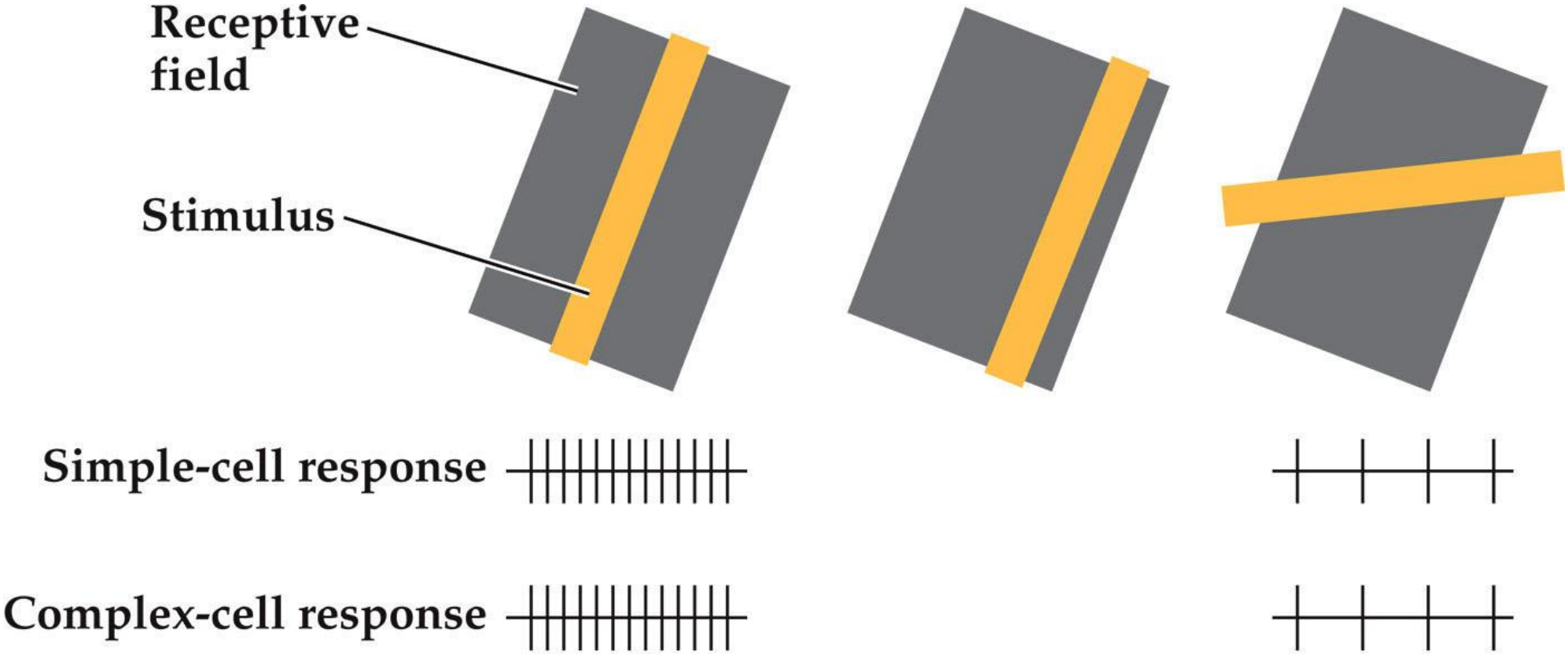
Hubel and Wiesel (1977)

How can orientation tuned cells be created from circularly symmetric inputs?

Circular LGN rfs can be combined to produce orientation sensitivity



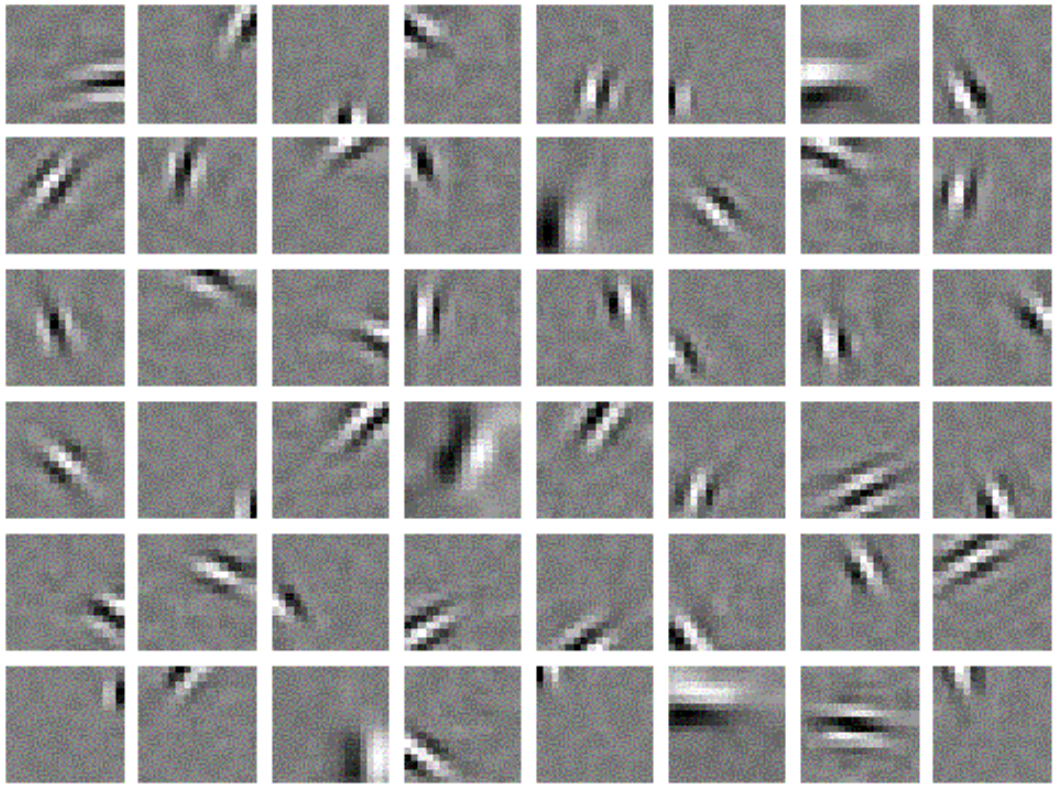
An orientation tuned cell may be a ‘simple’ cell or a ‘complex’ cell



Are such edge detectors learned or are they available innately?

Computational feasibility of learning rf structures

Given many patches from natural images, develop a linear representation which maximizes the fidelity of reconstruction while minimizing the number of bases active at any one time (sparseness).



So, V1 receptive fields can be learned through experience. But, are they really?

Development of spatial receptive-field organization and orientation selectivity in kitten striate cortex

B. O. Braastad and P. Heggelund

Journal of Neurophysiology, **53,** Issue 5, 1158-1178, 1985.

The functional organization of the receptive field of neurons in striate cortex of kittens from 8 days to 3 mo of age was studied by extracellular recordings. A quantitative dual-stimulus technique was used, which allowed for analysis of both enhancement and suppression zones in the receptive field. Furthermore the development of orientation selectivity was studied quantitatively in the same cells. Already in the youngest kittens the receptive fields were spatially organized like adult fields, with a central zone and adjacent flanks that responded in opposite manner to the light stimulus. Our results indicate that the spatial organization of the receptive field is innate in most cells and that visual experience is unnecessary for the organization to be maintained and for the receptive-field width to mature during the first month postnatally.

Thus, even though the rfs can, in principle, be learned through visual experience, the mammalian brain appears to start with the rfs already set up.

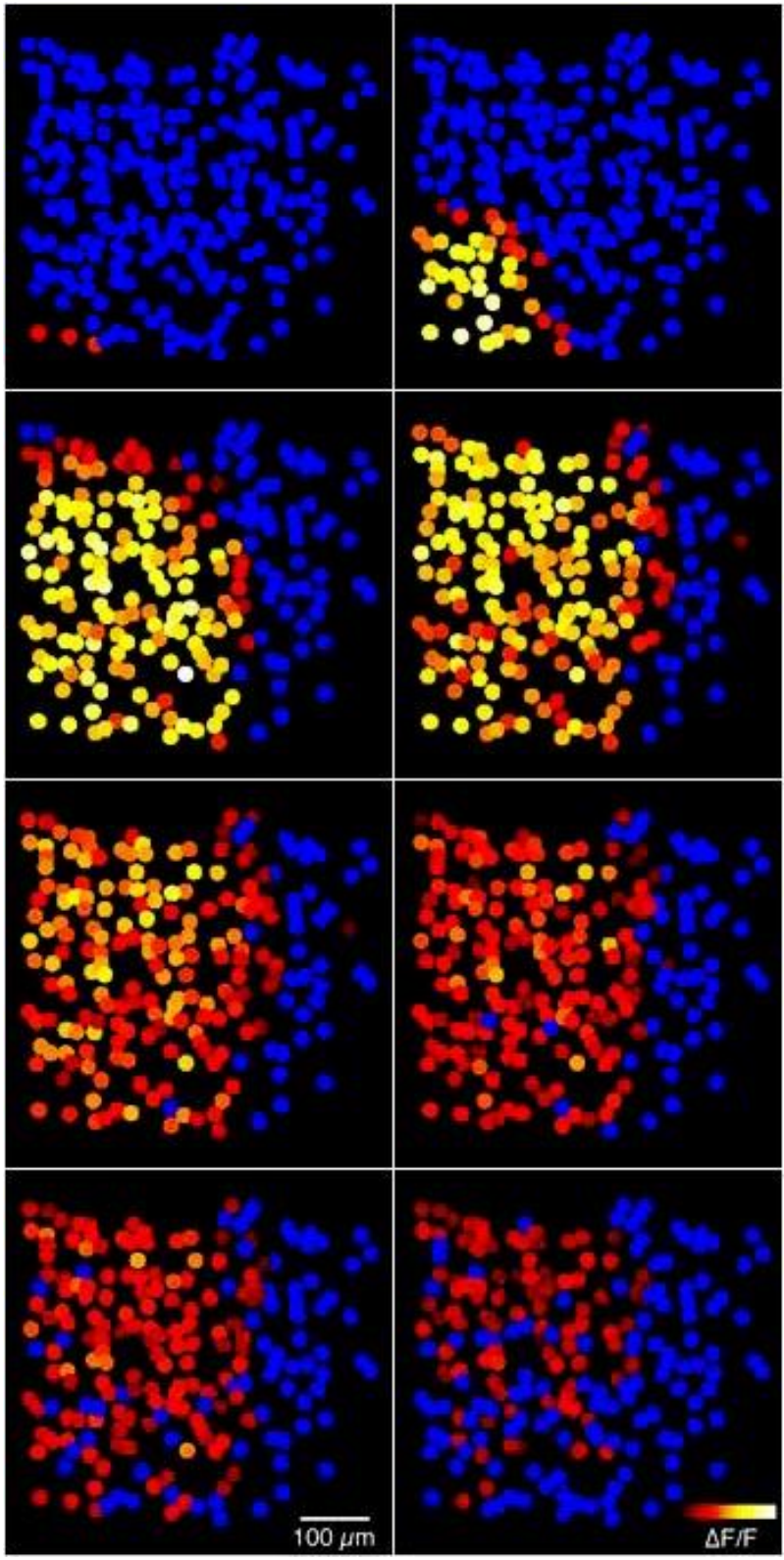
But, does this really rule out a role of visual experience in setting up the rfs?

“Forty percent of the cells were orientation selective before 2 wk, and the fraction increased to 94% at 4 wk.”

-B & H, 1985

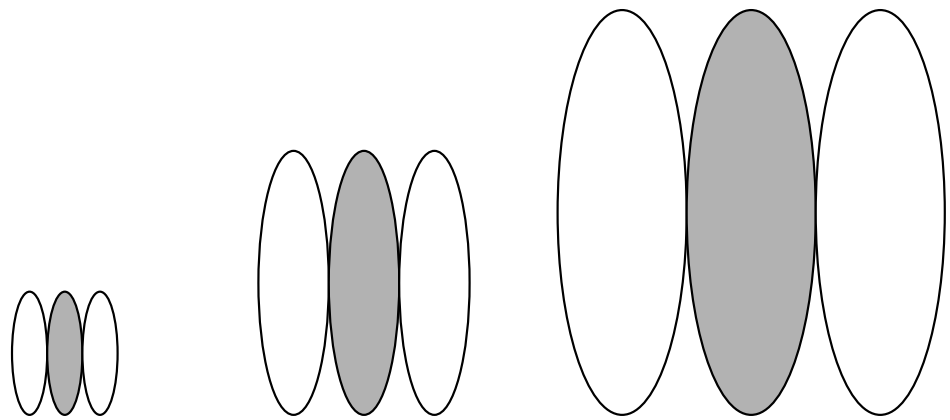
There might be prenatal visual experience in the form of retinal waves.

https://www.youtube.com/watch?v=DGBy-BGiZIM



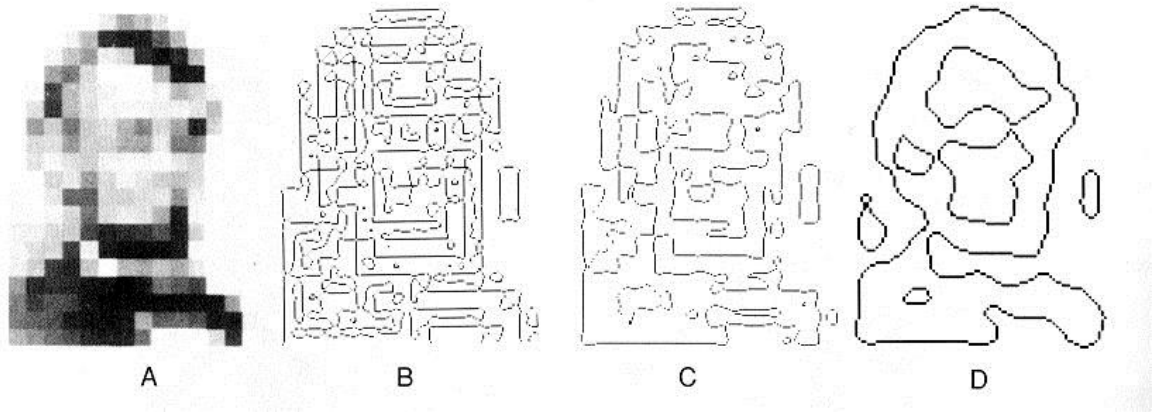
Example of the Propagation of a Spontaneous Wave in an Embryonic Turtle Retina. We mapped the wave with Ca2+ fluorescence. Each disk in the wave is a ganglion cell shown in its actual position and frames are 1-s apart (ordered from left to right and top to bottom). Inactive cells are in blue and relative fluorescence appears in pseudo color (scale on bottom right).

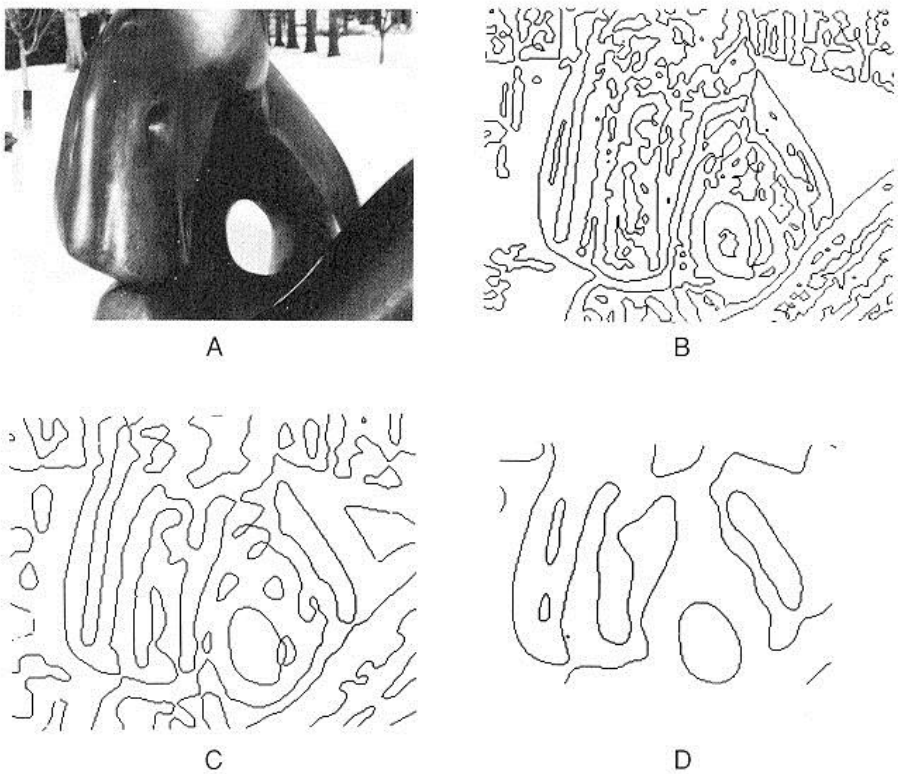
- Oriented receptive fields in V1 likely have both nature and nurture origins
- They come in different sizes



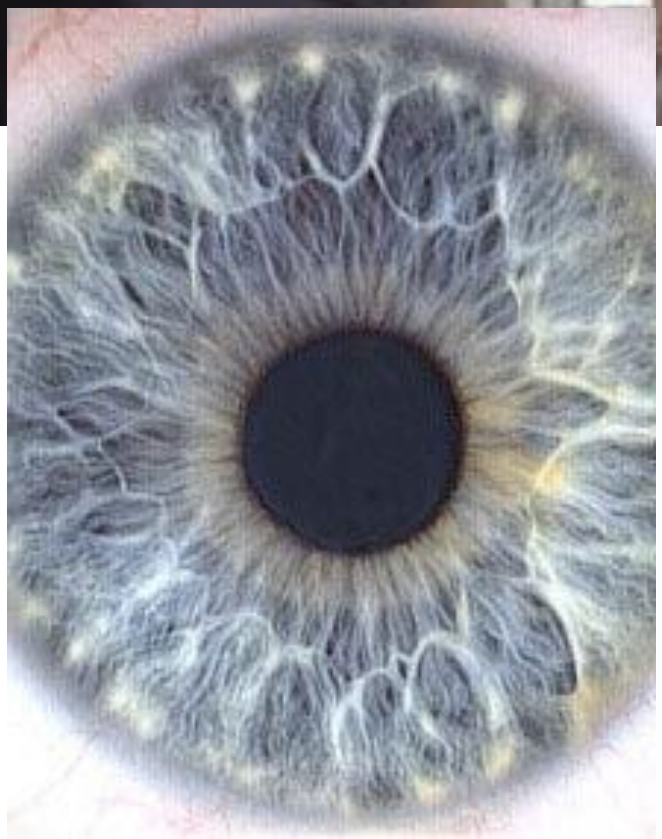
The rfs come in a variety of orientations and also in a variety of sizes to process edges at different scales

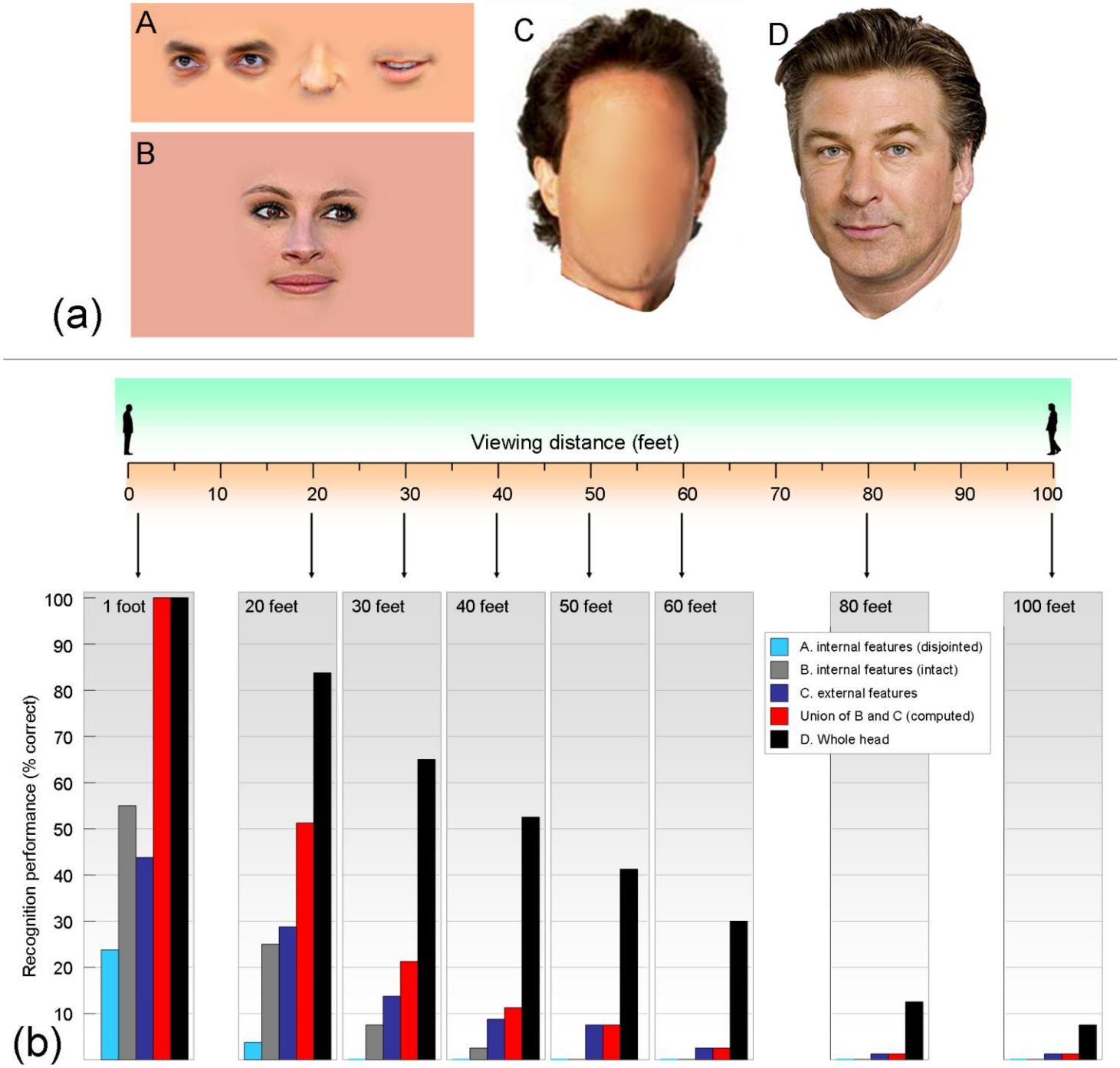
Detecting edges at different scales





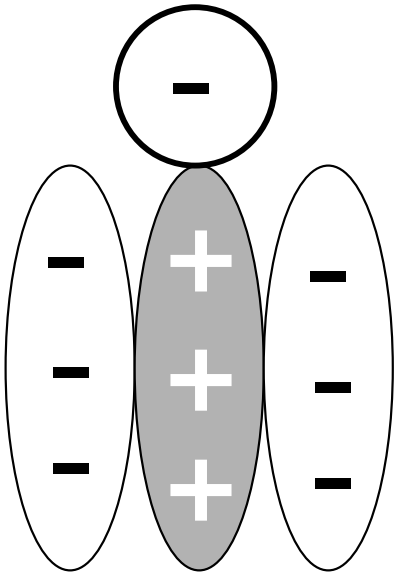
Trite question: Why do we have high-resolution rfs?





Braun, Jarudi & Sinha (in prep)

- Oriented receptive fields in V1 likely have both nature and nurture origins
- They come in different sizes
- Some of the units are ‘end-stopped’



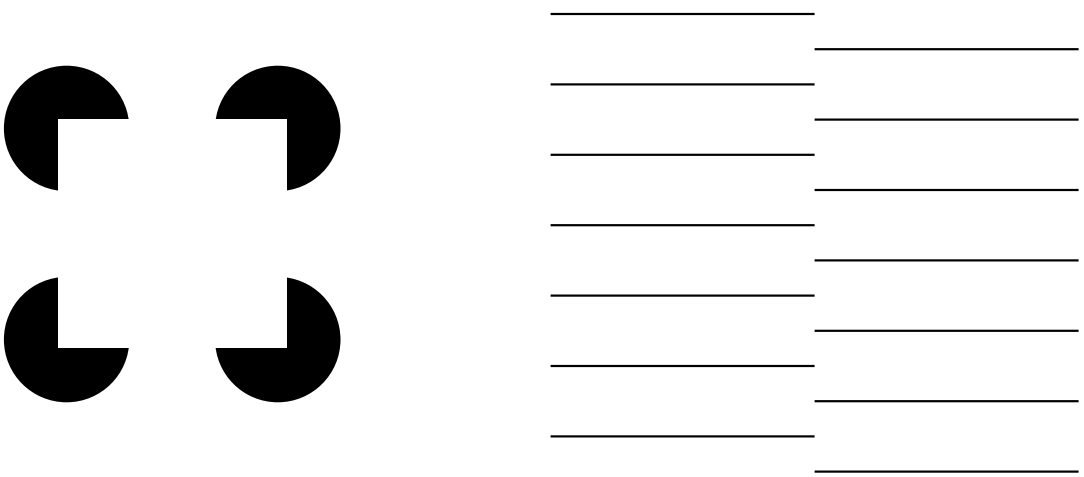
What might be the purpose of such units?

- Oriented receptive fields in V1 likely have both nature and nurture origins
- They come in different sizes
- Some of the units are ‘end-stopped’
- Gabor rfs have limits in their ability to explain boundary perception

Gabor filters detect luminance defined edges.

Limitations of conventional edge-detectors

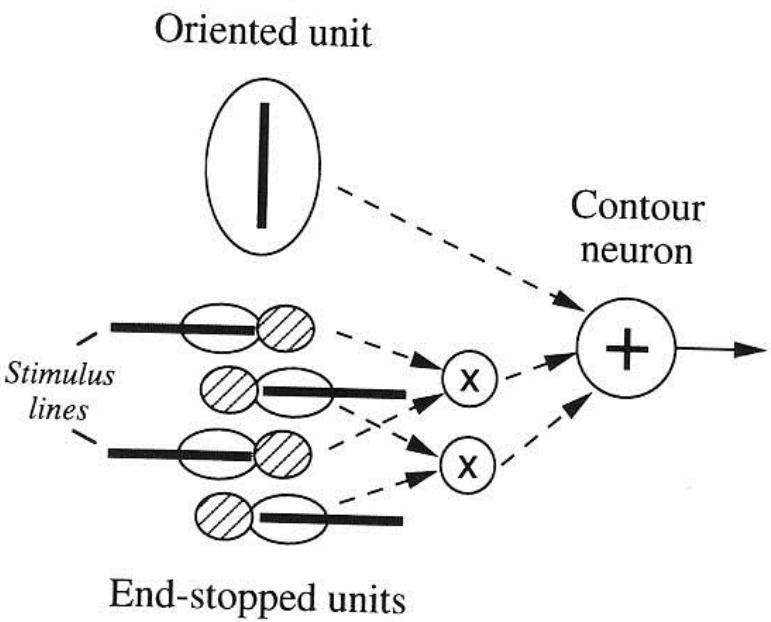
Detecting illusory contours

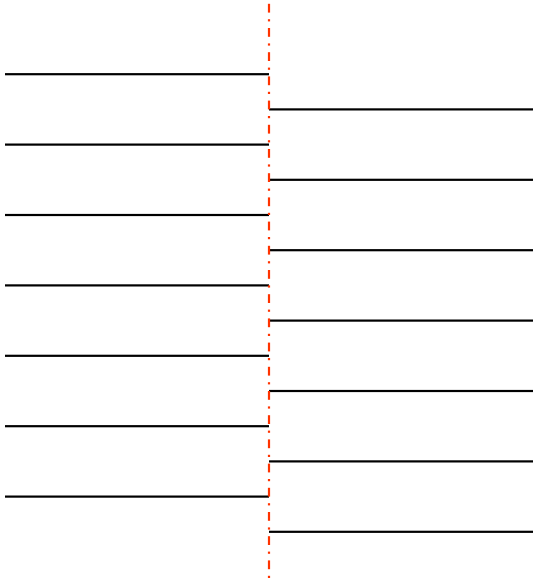


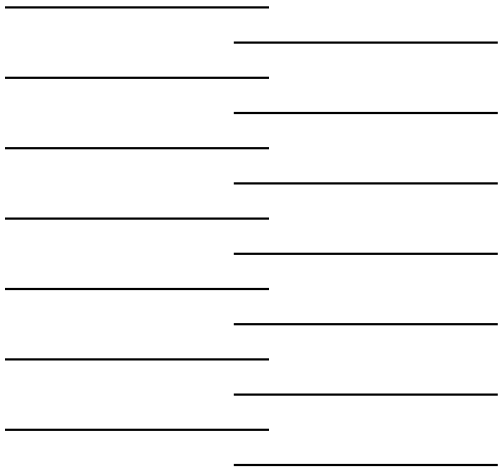
No luminance difference across long sections of the perceived contours

A model for detecting illusory contours

(Peterhans and von der Heydt, 1989)







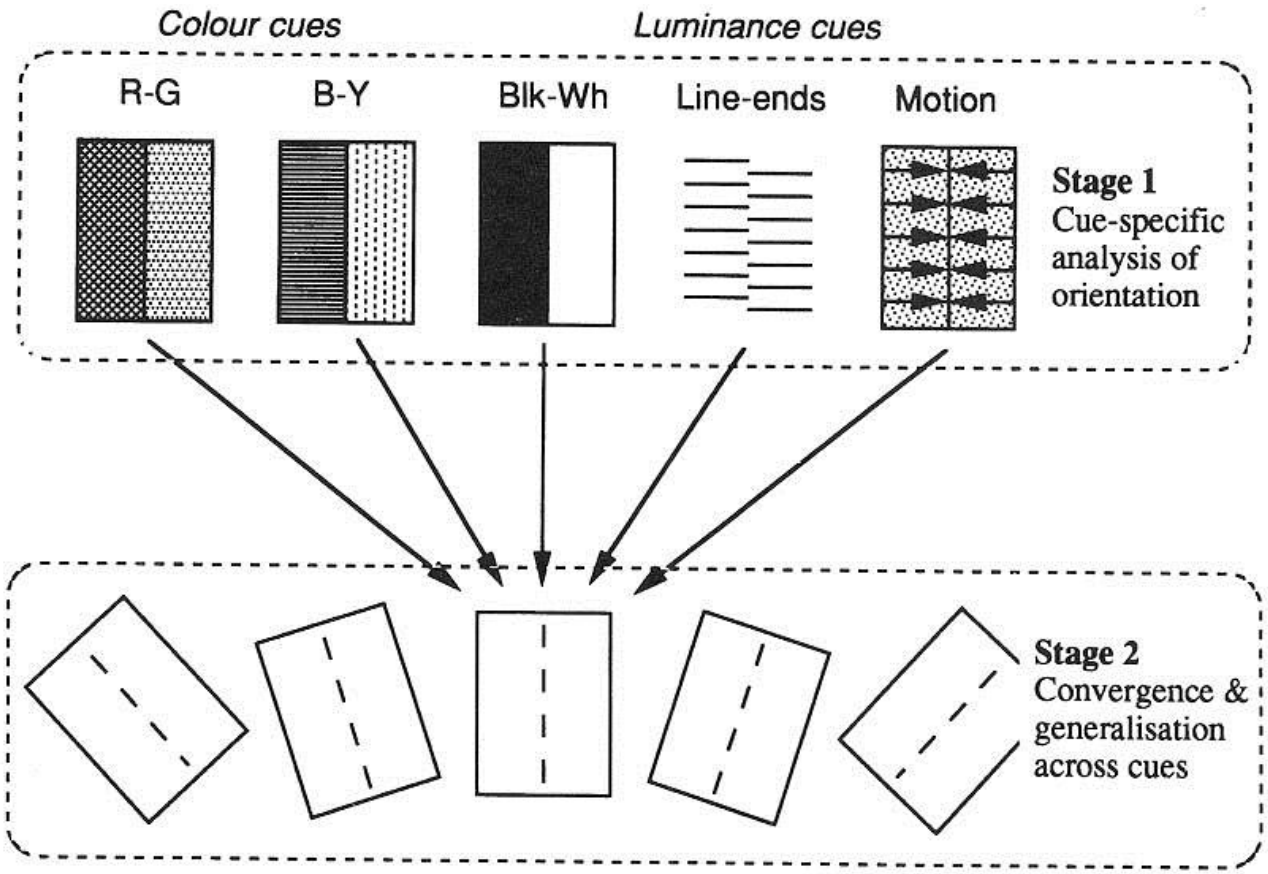
Caveat: Response decreases gradually with increasing intermeshing of gratings. This does not correspond with subjective percepts.

Do real and illusory edge detectors interact? Is there a convergence of cues in contour coding?

Is there a convergence of cues in contour coding?

Cue dependent
Contour coding

Cue independent
Contour coding

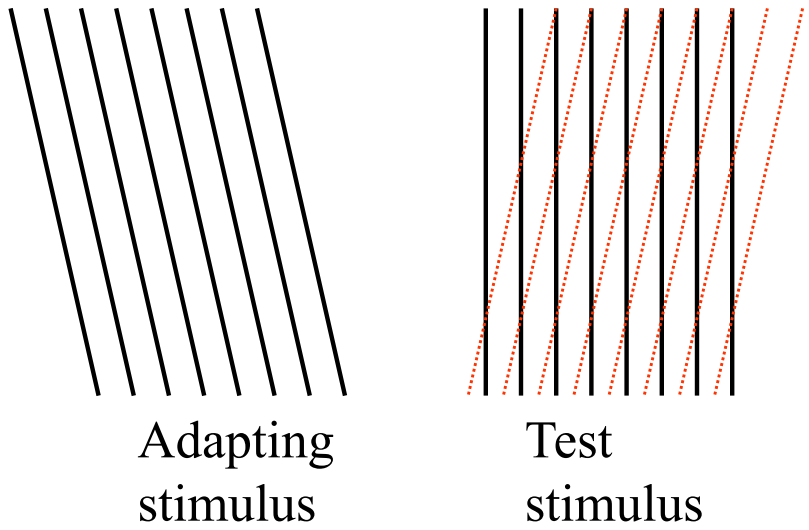


Is there any experimental evidence to support this idea?

What kind of neurophysiological evidence might make this point?
What kind of behavioral evidence might make this point?

Convergence of cues in contour coding

(Studies with Tilt AfterEffect)

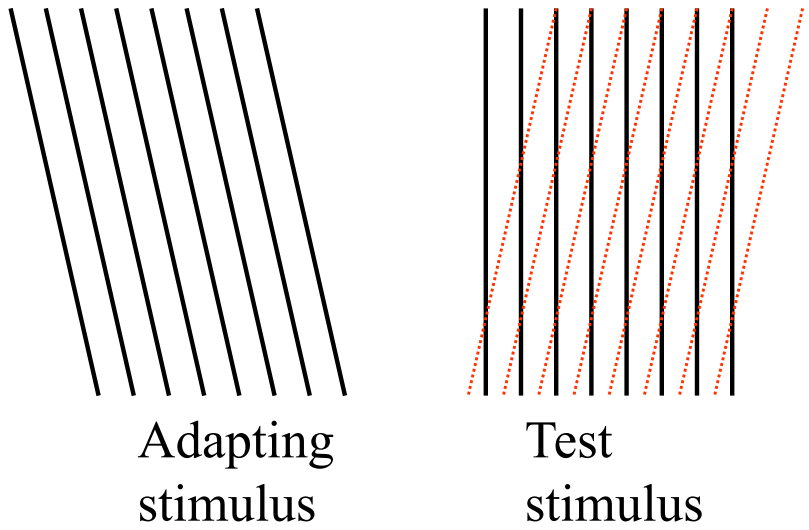


TAE is orientation specific and exhibits inter-ocular transfer.
Therefore, TAE is likely to be of cortical origin.

Does the TAE transfer from one kind of contour (say ‘real’) to another (say ‘illusory’, or ‘motion defined’)?

Convergence of cues in contour coding

(Studies with Tilt AfterEffect)



TAE is orientation specific and exhibits inter-ocular transfer.
Therefore, TAE is likely to be of cortical origin.

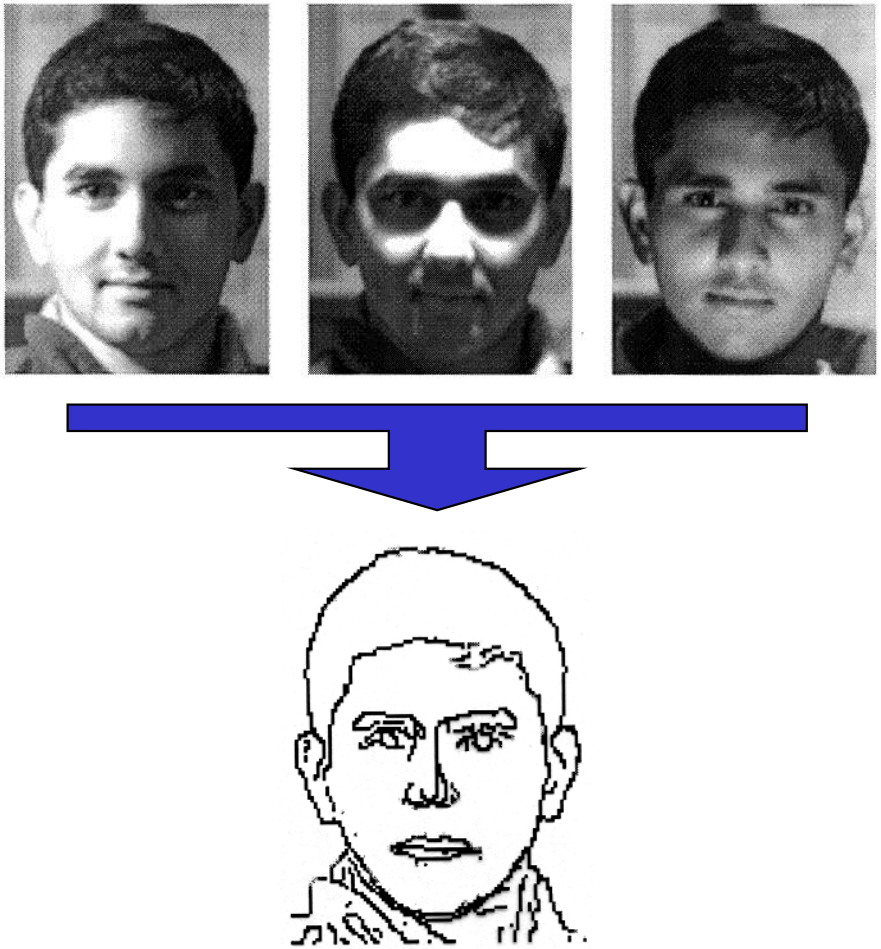
Does the TAE transfer from one kind of contour (say ‘real’) to another (say ‘illusory’, or ‘motion defined’)?

Yes.
(Berkley, DeBruyn & Orban, 1994;
Paradiso, Shimojo & Nakayama, 1989)

Edge-based processing: Parting thoughts…

What underlies researchers' fascination for edges?

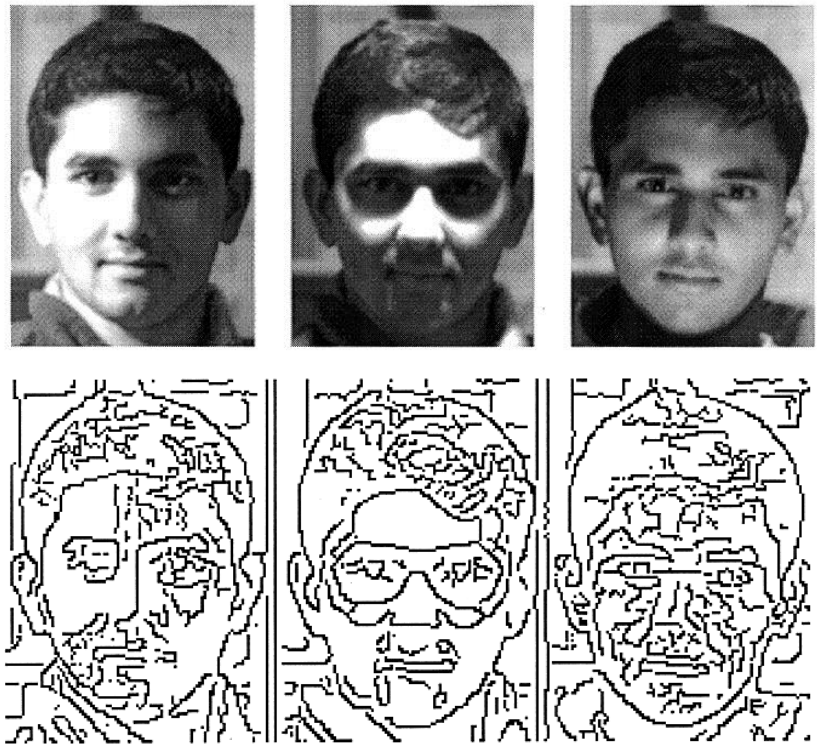
Fine edges are expected to be **invariant** to imaging variations.



In principle, this can make the subsequent processing very easy.

But…

In practice, fine edges turn out to be highly unstable.



Even after more than four decades of research, we have been unable to create a robust edge-based vision system.

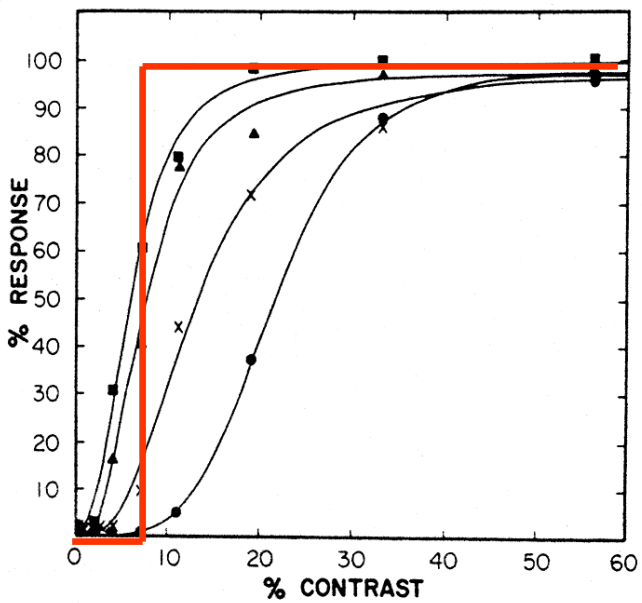
But, doesn't V1 comprise edge-detectors?

A matter of interpretation

Revisiting visual neurophysiology:

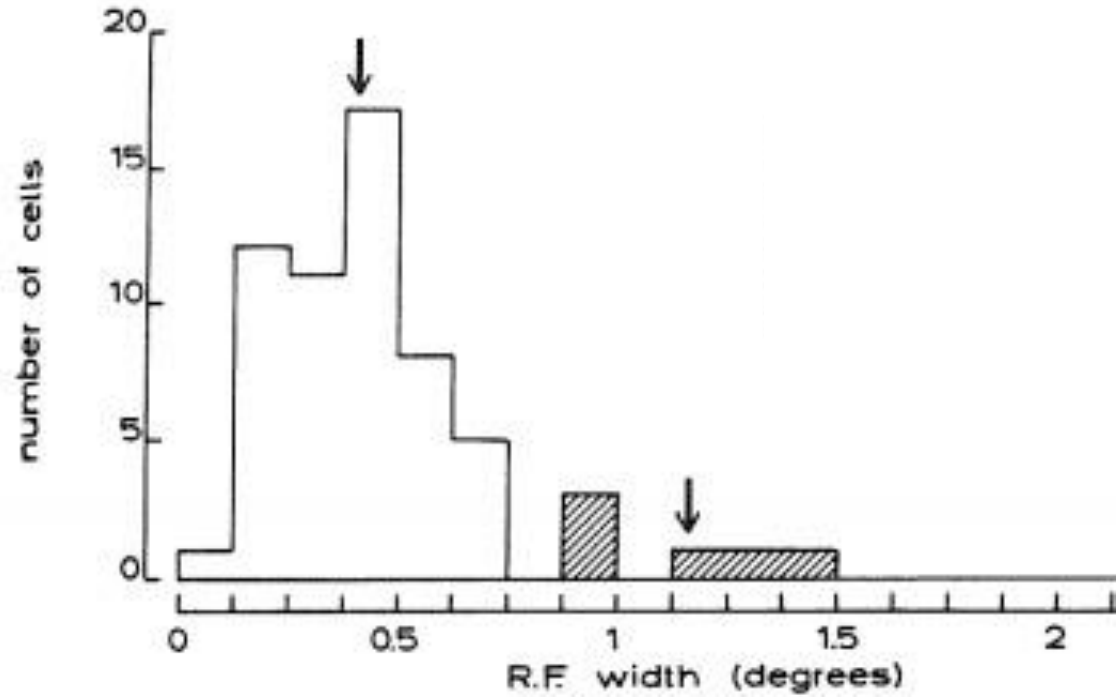
Many neurons in the primary visual cortex have …

Rapidly saturating contrast response functions



Albrecht & Hamilton, 1986

Large receptive fields

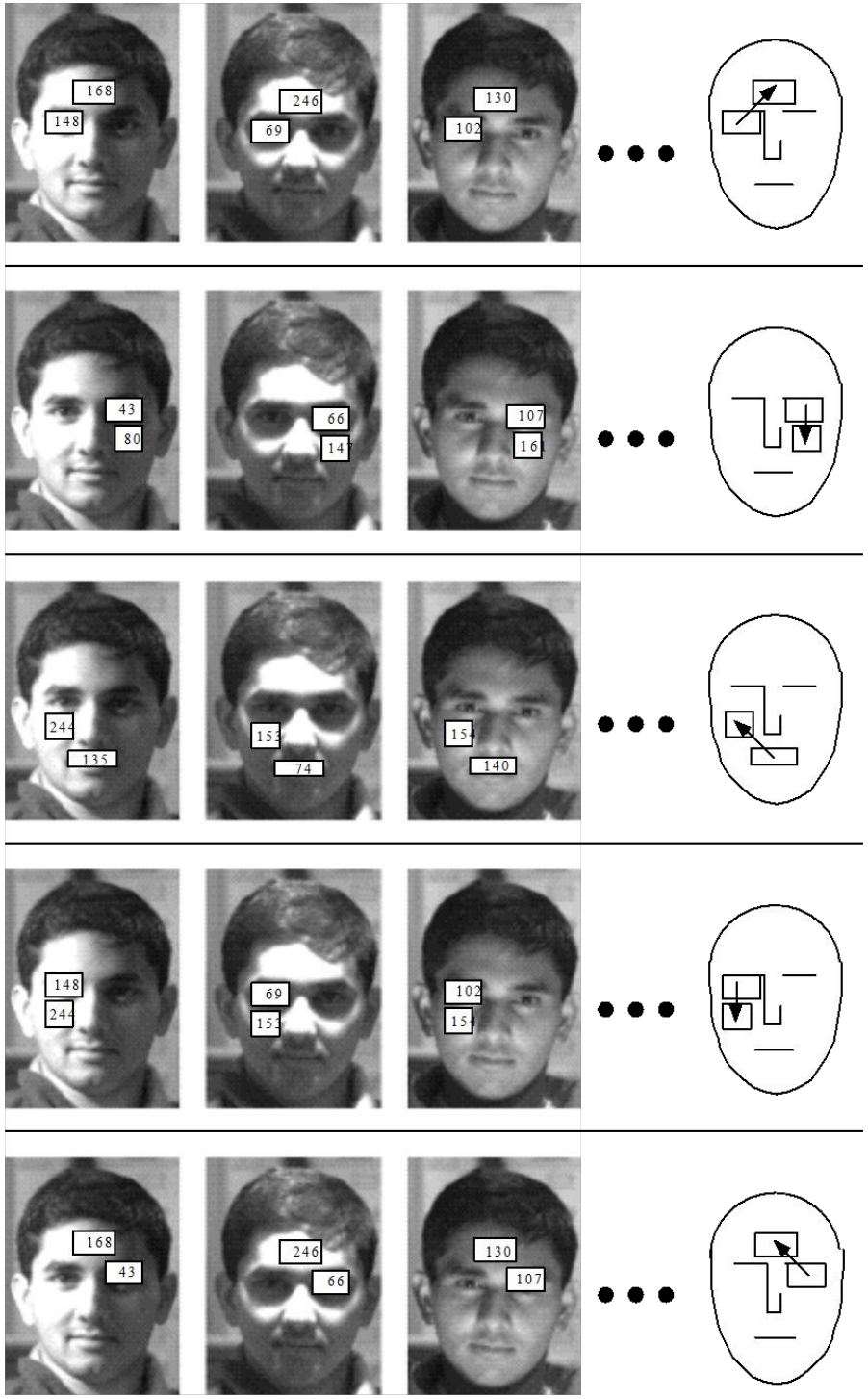


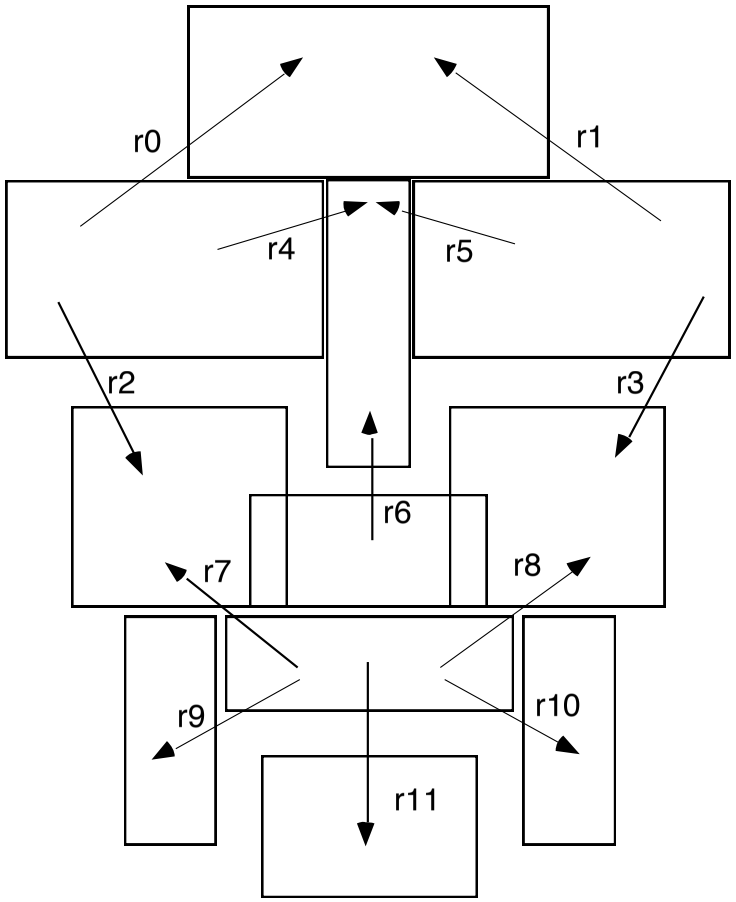
Such cells are well suited to encode

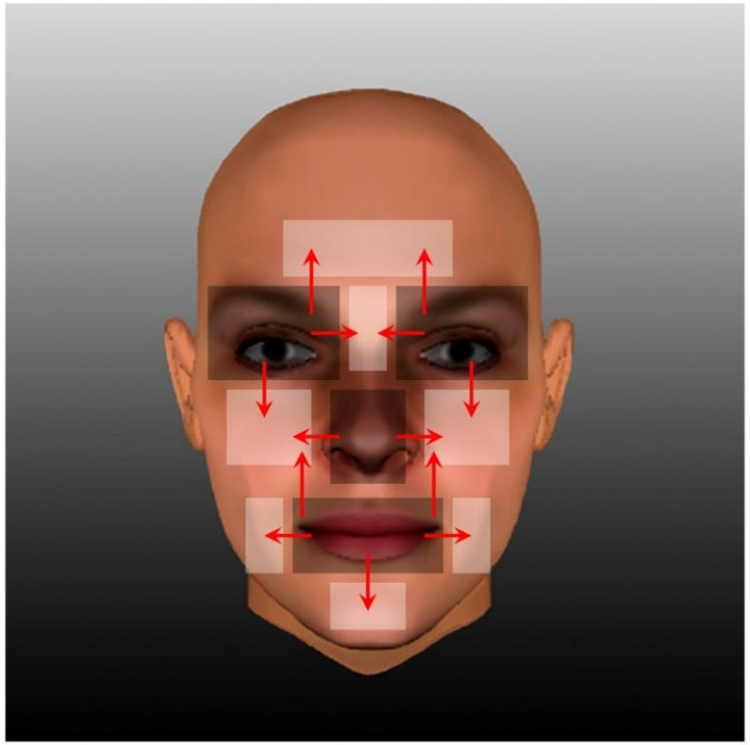
1. In geometry: coarse, low-res structure
2. In photometry: ordinal brightness relations

How can we use this scheme?

Ordinal brightness relationships across several region-pairs stay invariant across a range of conditions.





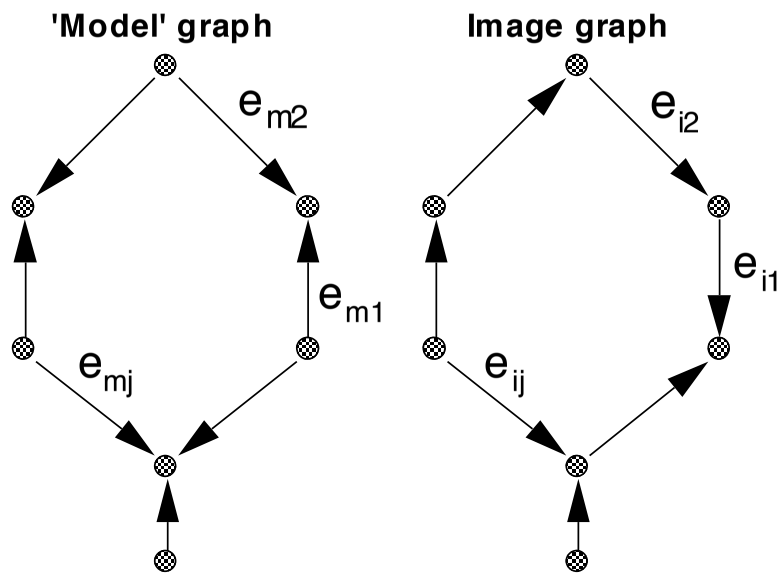


These pair-wise ordinal relations can be conjugated to create a quasi-invariant object signature – the 'Ratio Template'

(Sinha, 2002)

Matching:

A ratio-template can be thought of as a graph on region brightnesses…



$$\text{Match metric} = \sum f(e_{mj}, e_{ij}) * W_j$$

where $f(e_{mj}, e_{ij}) = 1.0$ if $e_{mj} = e_{ij}$

$= -1.0$ if $e_{mj} \circ e_{ij}$

Face detection by ratio-templates

Physiological support for ratio templates

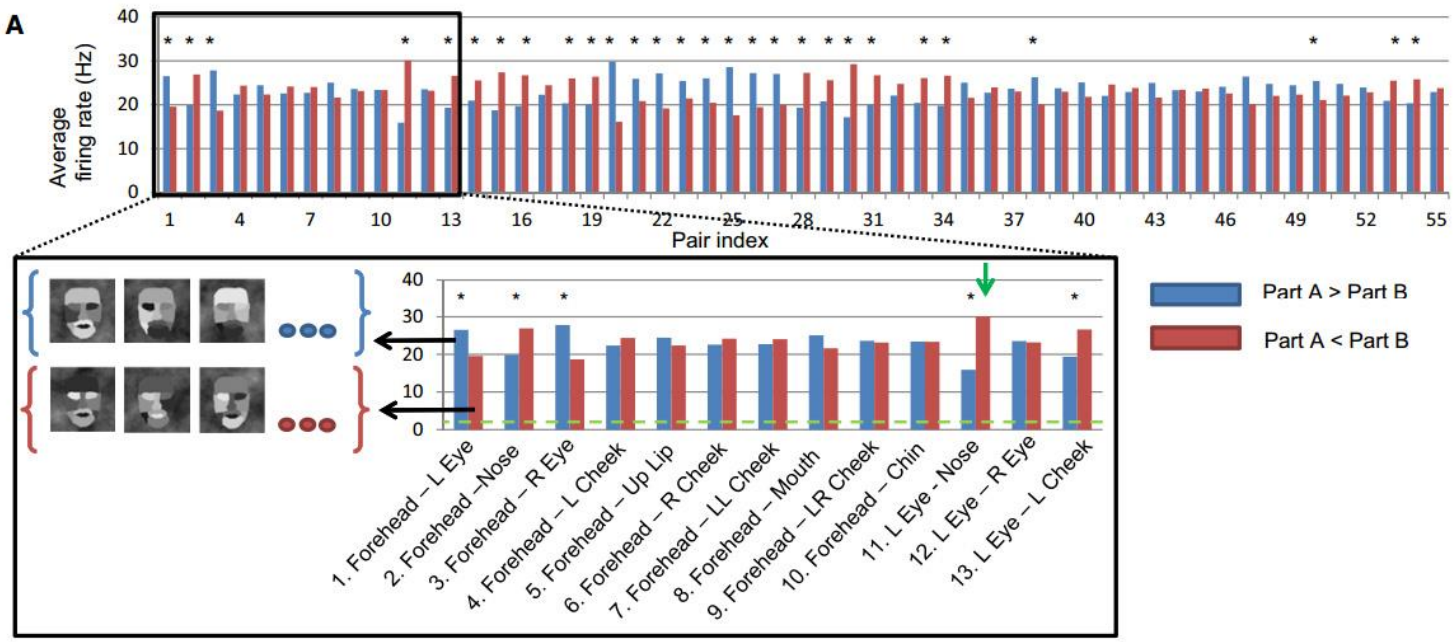
Ohayon S, Freiwald WA, Tsao DY, 2012, What makes a cell face selective? The importance of contrast. *Neuron*, **73**(3)

*The critical stimulus factor determining response magnitude was **contrast polarity**, for example, nose region brighter than left eye. Contrast polarity preferences were consistent across cells, suggesting a common computational strategy across the population, and matched features used by computer vision algorithms for face detection*

Cell PRESS

Neuron

Local Contrast Importance for Face Selectivity



14 May 1994 No1925 Weekly £1·60

NewScientist

Cloning movie stars

How to catch a catalyst in the act

Is ME still a mystery?

The robot that wants to be human

The Cog Project

Imitating Head Nods

Brian Scassellati

MIT Artificial Intelligence Lab

The New York Times

February 13, 2007

"...Dr. Sinha and his colleagues have arrived at unexpected insights into the question of why we sometimes see a cinnamon bun as a cinnamon bun, and other times as the earthly incarnation of a sainted nun."

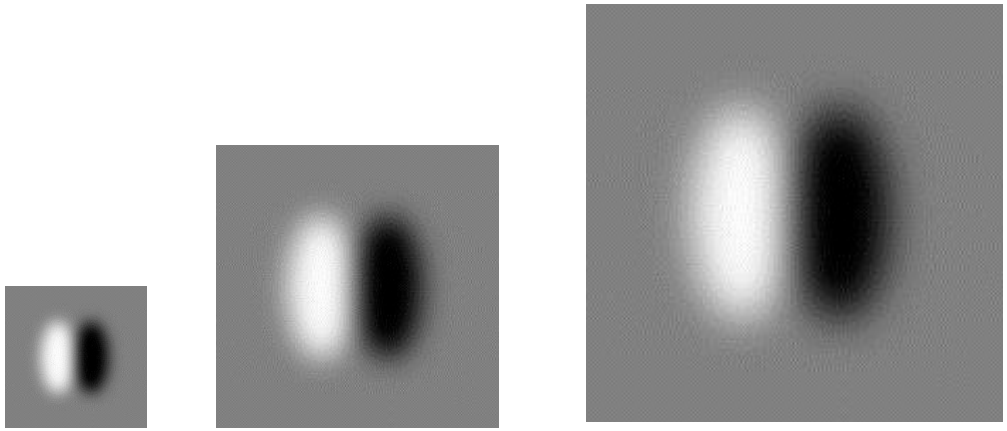
Even further beyond edges…

Edges represent only local relationships between image regions.
Might non-local relationships also prove useful for encoding image structure?

Dissociated Dipoles
aka
'Sticks'

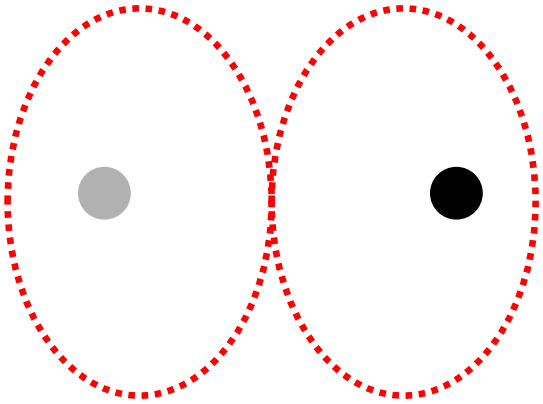
Balas & Sinha (2006) *Neural Computation*

Conventional representation



Good for detecting edges at different scales.

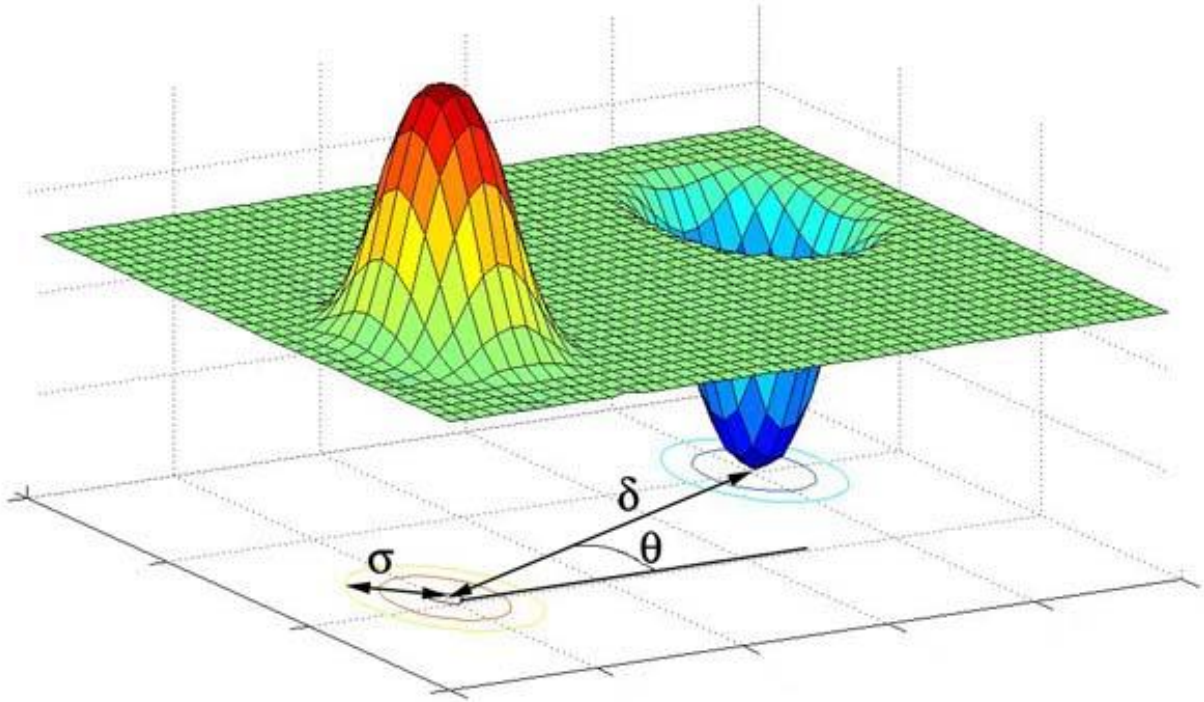
Not so good for comparing small regions across large distances.



<u>Reason for this shortcoming:</u>
Lobe size is coupled with inter-lobe distance.

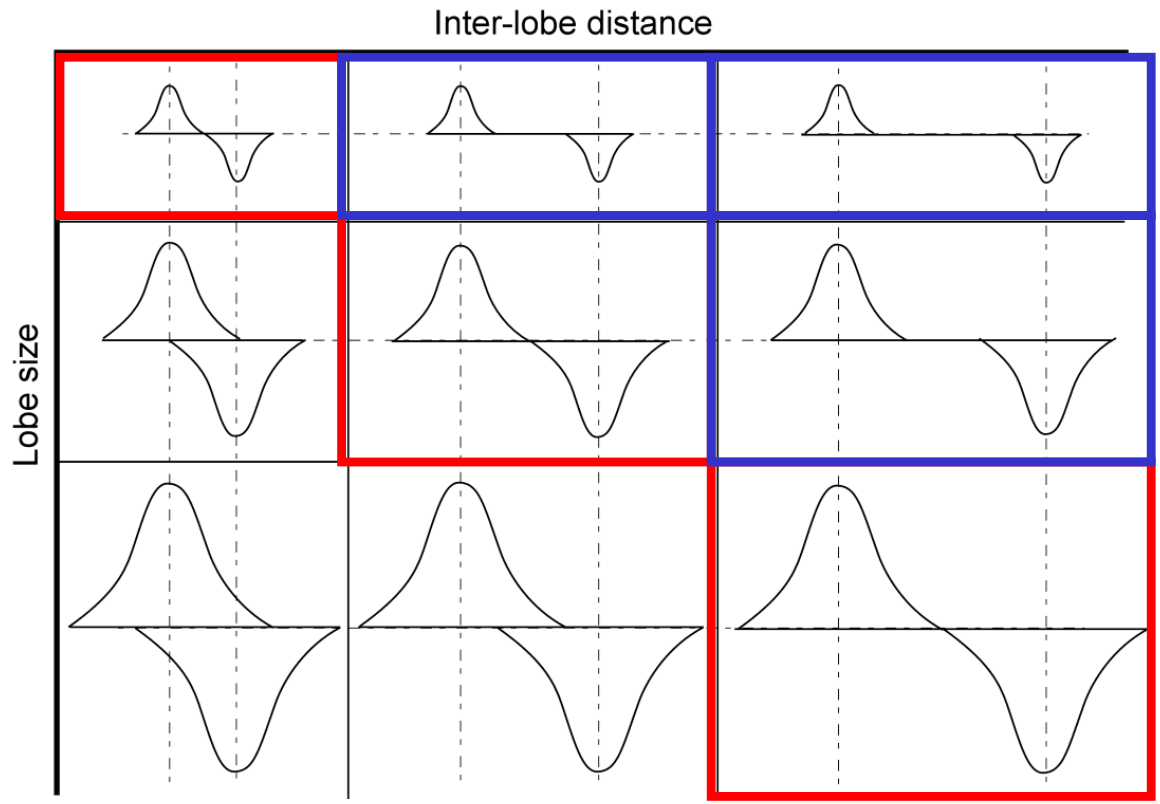
Enter non-local operators…

"Dissociated Dipoles" or "Sticks"



This operator removes the confound between filter span and lobe-size, allowing for sensitive comparisons between distant image regions.

Sticks are the off-diagonal elements in the lobe-size vs inter-lobe distance matrix



Conventional operators

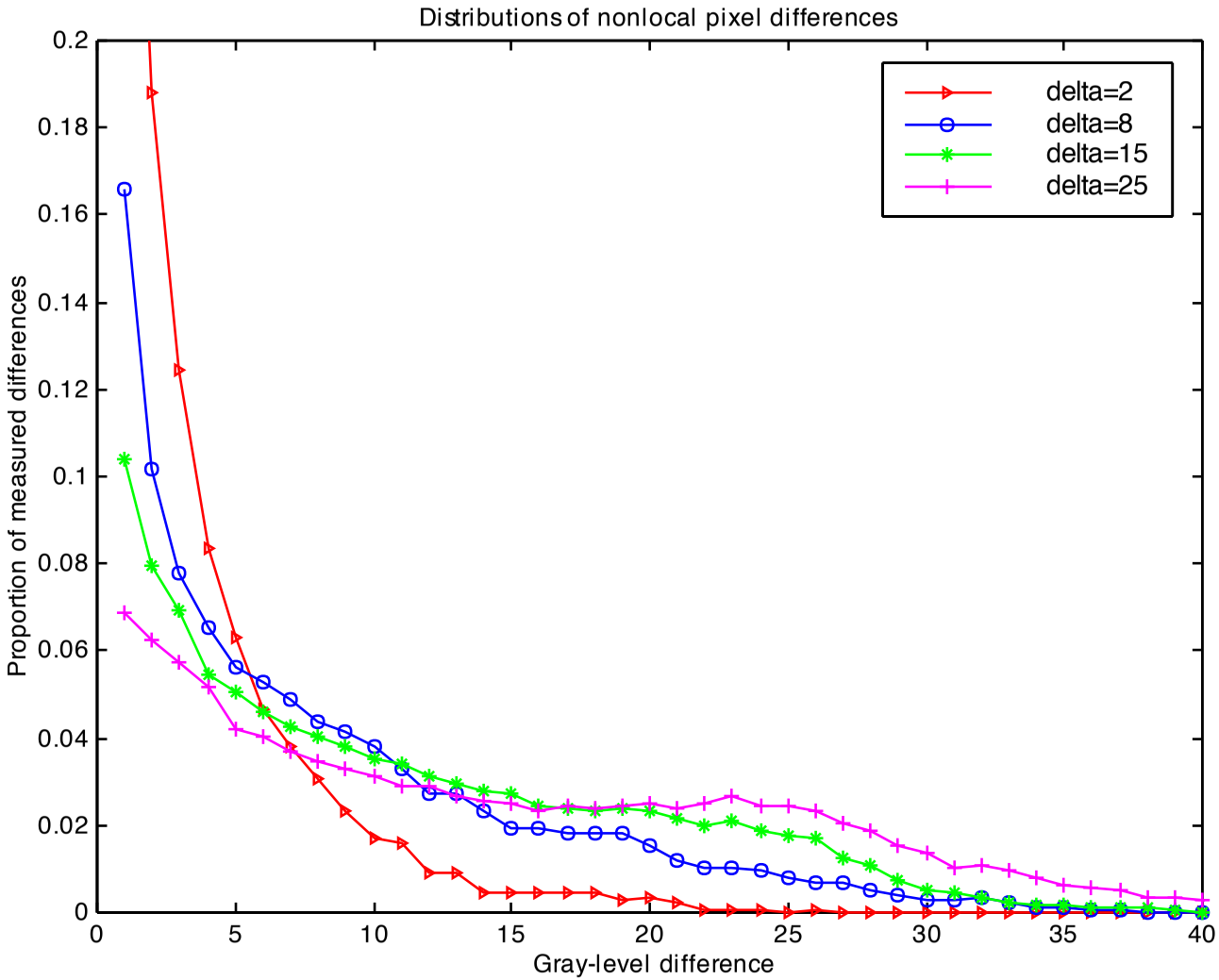
Sticks operators

Sticks provide a way of encoding non-local relationships in an image.

But, why use non-local relationships at all?

Why use non-local relationships?

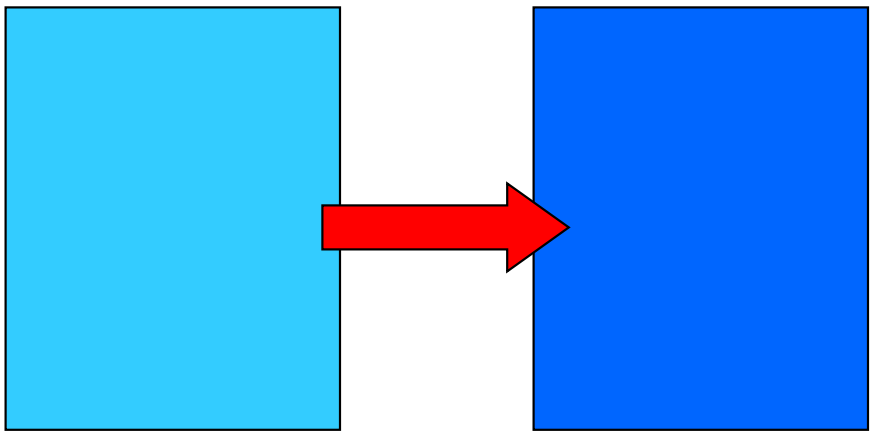
1. Non-local relationships are well-suited to handling the spatial redundancy of natural images



Regions further apart are more likely to be different.

Why use non-local relationships?

2. Non-local relationships provide greater stability to image transformations



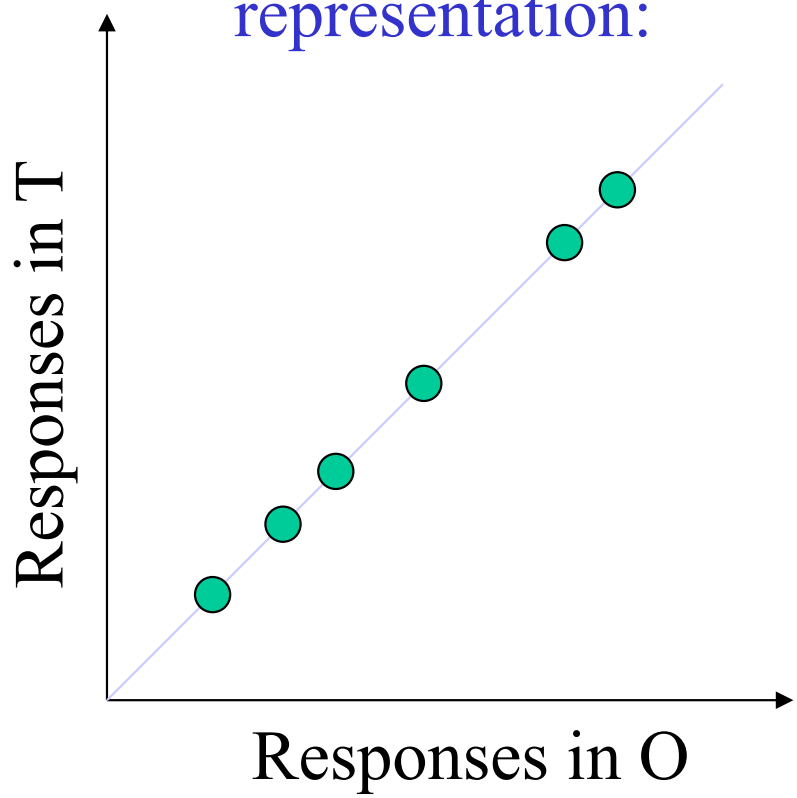
Original image

Transformed image

O = (o1,…, on)

T = (t1,…, tn)

For a stable representation:



Why use non-local relationships?

2. Non-local relationships provide greater stability to image transformations

Example:

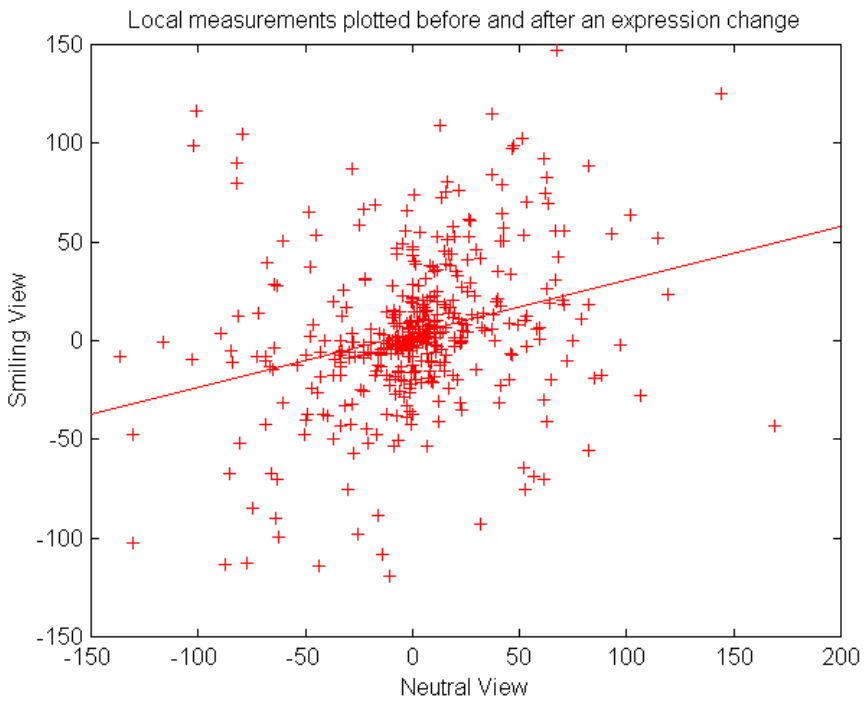
Original

T1: Expression change

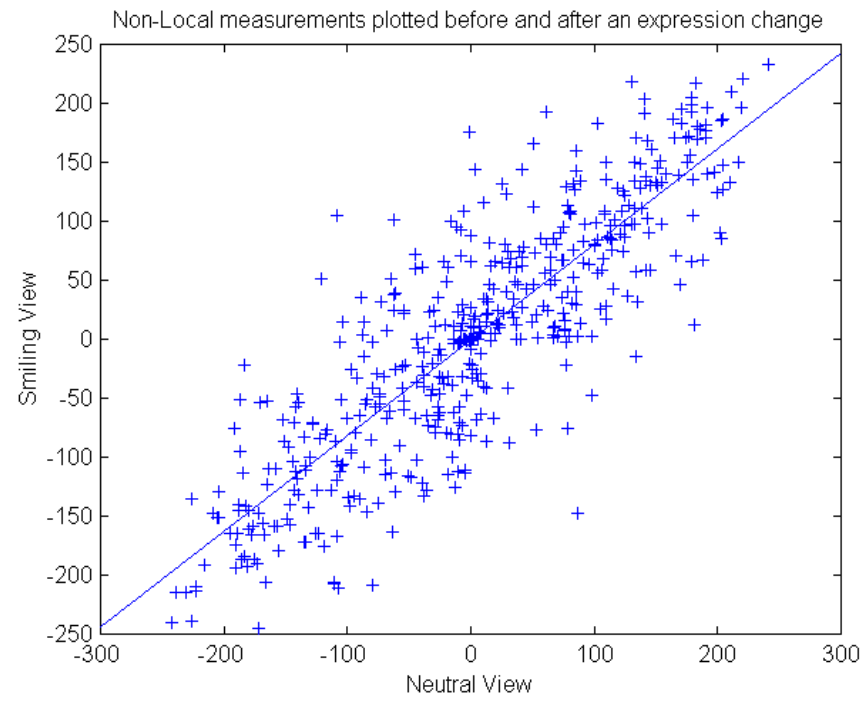
T2: Depth rotation

T3: Translation

Expression change…

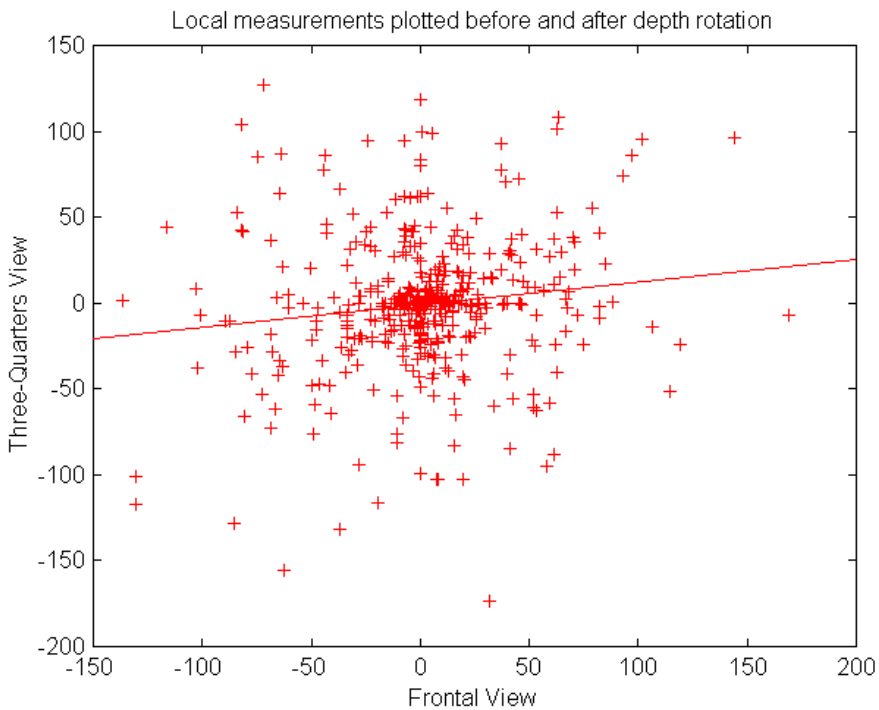


R = 0.28

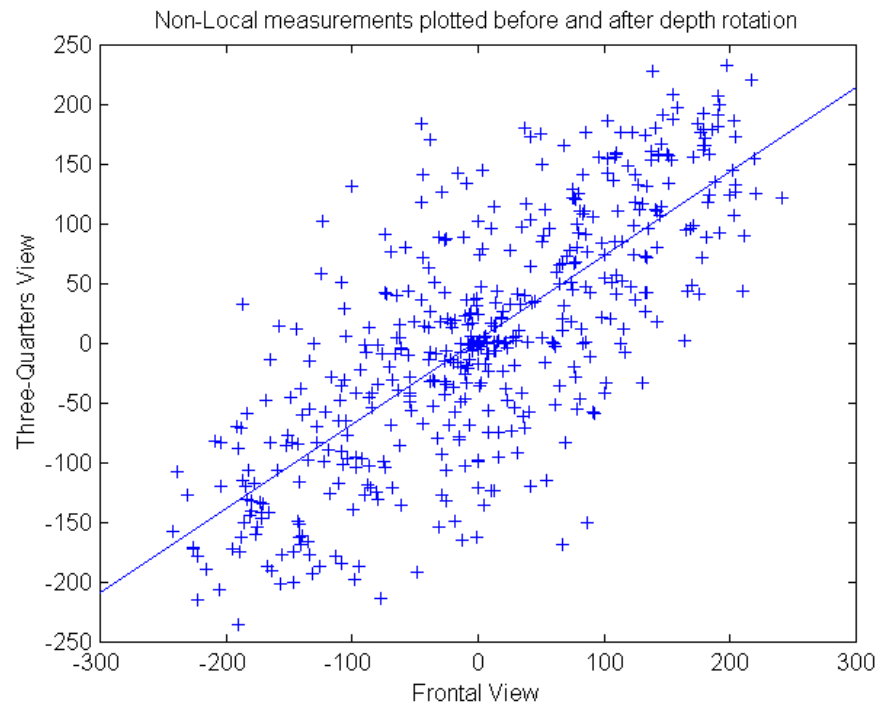


R = 0.85

viewpoint change…

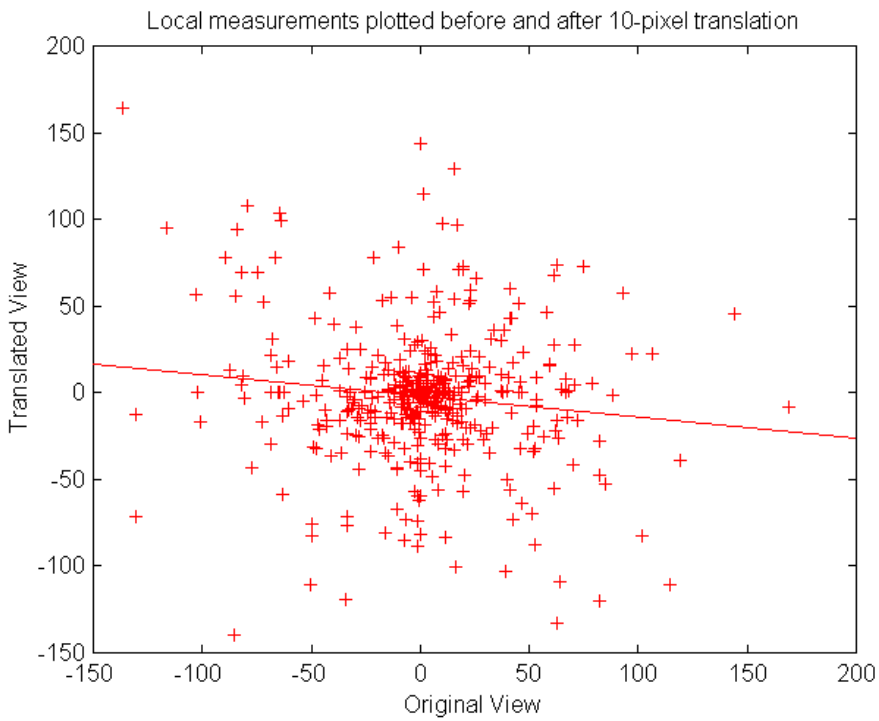


R = 0.13

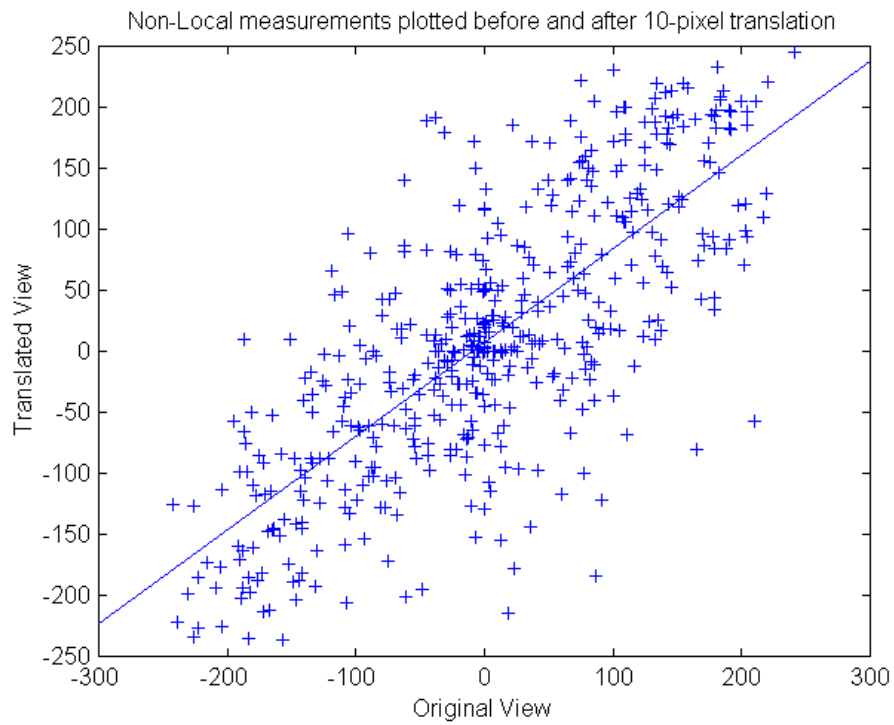


R = 0.75

…and translation.



R = -0.12



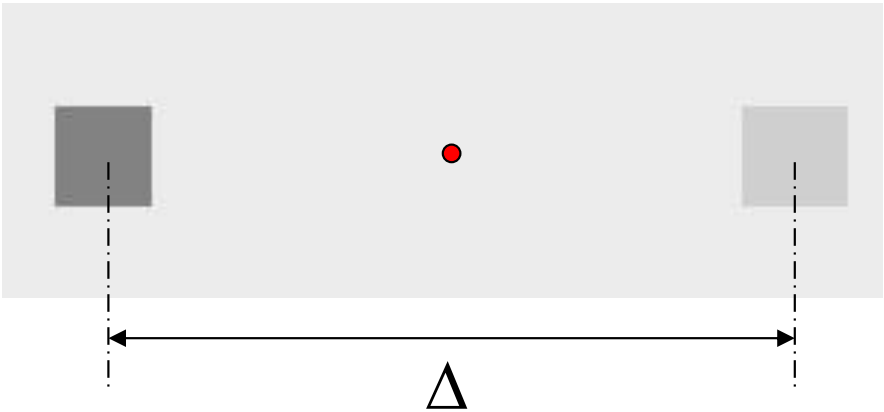
R = 0.88

Given the stability of non-local measurements, we hypothesize that STICKS operators may be useful for complex recognition tasks.

But, first, a reality check on the perceptual plausibility of extracting such non-local measurements...

Are observers any good at making non-local comparisons?

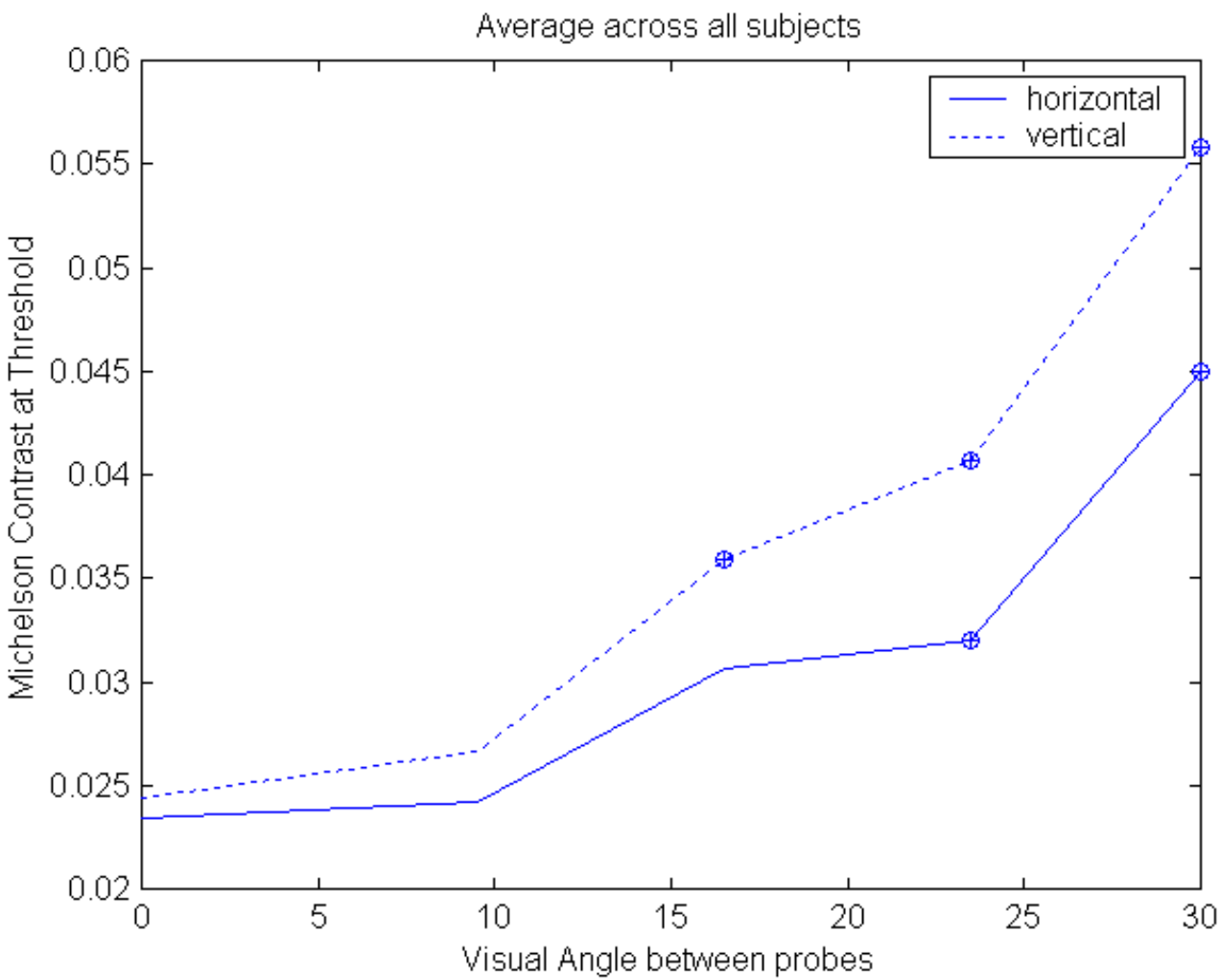
A brief bit of psychophysics…



Subjects' task: Fixate in the center and indicate which probe-patch is brighter

Data analysis: Thresholds of detectable differences as a function of inter-patch separation.

Psychophysics results



Subjects can reliably compare the brightness of probe-patches even for separations as large as 10 degrees.

Interim summary:

Sticks appear to provide a stable encoding of images.

Non-local comparisons are perceptually plausible.

But, are sticks actually useful for vision tasks, such as recognition?

Interim summary:

Sticks appear to provide a stable encoding of images.

Non-local comparisons are perceptually plausible.

But, are sticks actually useful for vision tasks, such as recognition?

STICKS for recognition

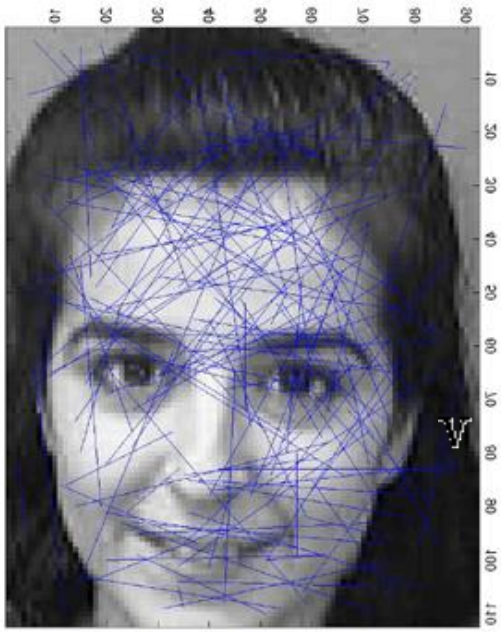
Computational explorations

We test whether the inclusion of non-local measurements confers any advantages to a recognition system.

Our sample domain: Faces

The basic STICKS algorithm

Step 1: ‘Drop’ a set of sticks randomly on an image
Step 2: Construct a feature vector from stick responses
Step 3: To compare two images, determine the distance (say, L2) between their stick feature vectors



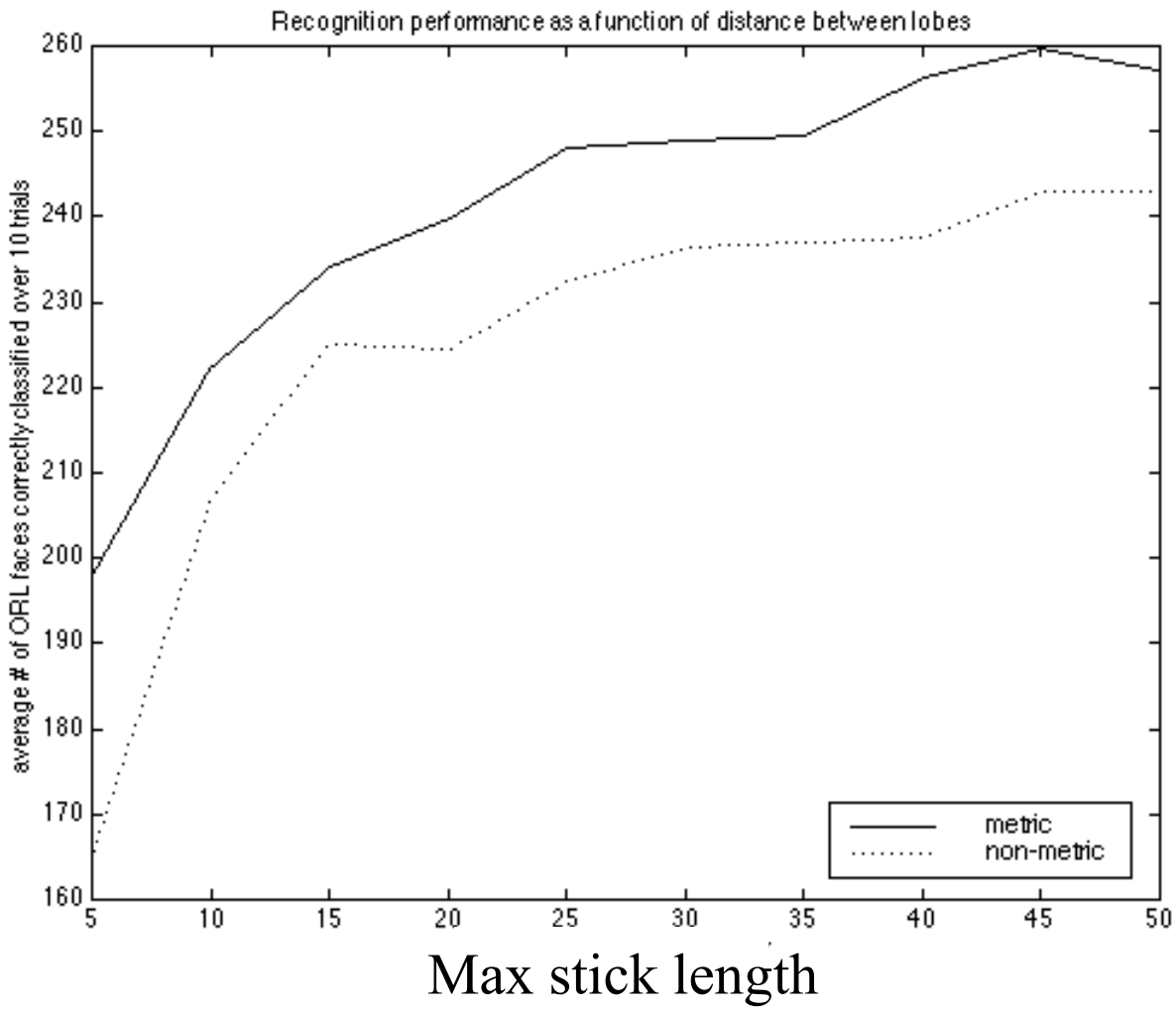
$$y = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

Query image

...

Database images
(multiple images for several individuals)

Recognition Results



Inclusion of non-local measurements improves recognition performance

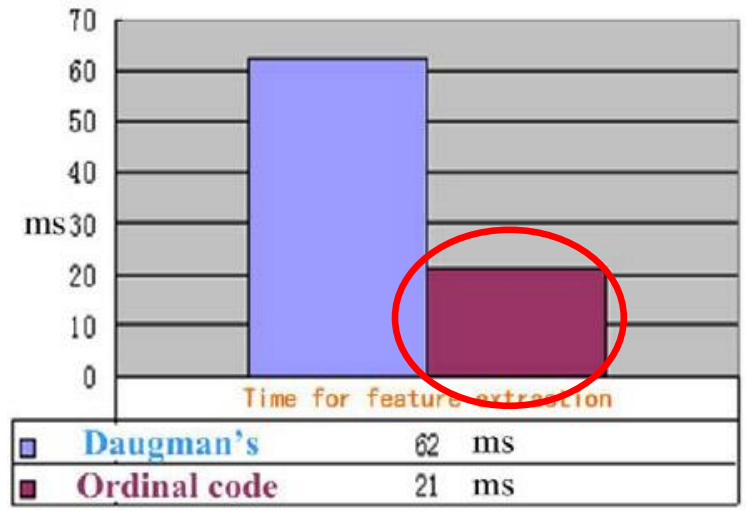
Independent tests of effectiveness of sticks

Results from other labs that have built computer vision systems based on the dissociated-dipoles proposal show that this is the most effective (more accurate and faster) representational vocabulary available today for diverse recognition tasks.

Results on UST Hand Image Database

Algorithm	EER	d'
Palm code ($\theta = 0^\circ$) [5]	1.68%	3.44
Palm code ($\theta = 45^\circ$) [5]	1.77%	3.39
Palm code ($\theta = 90^\circ$) [5]	2.95%	2.95
Palm code ($\theta = 135^\circ$) [5]	2.14%	3.25
Fusion code [14]	0.75%	3.40
Competitive code [15]	0.38%	3.51
Ordinal code	0.22%	4.77

Sun et al., CVPR 2005



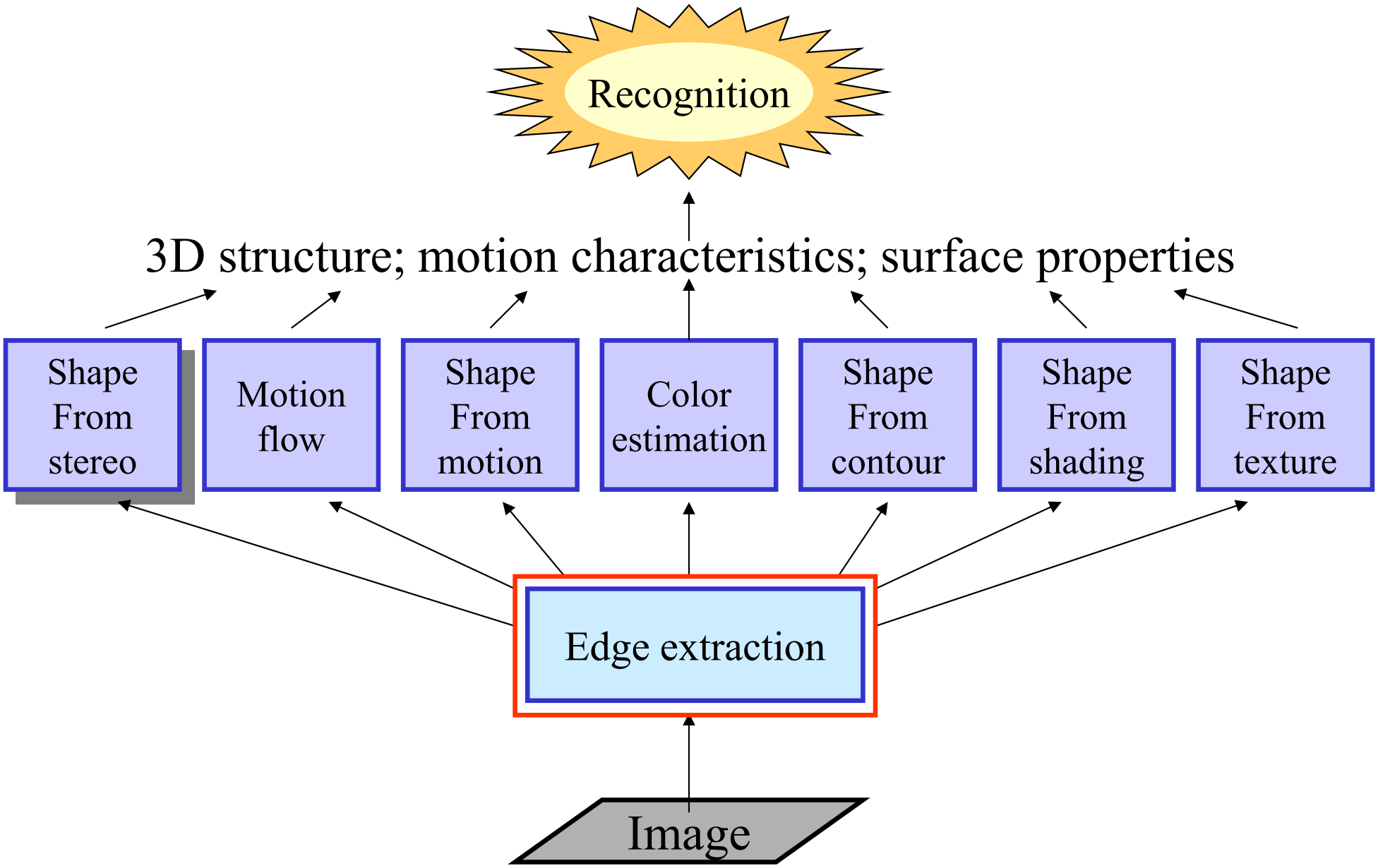
Tieniu Tan, Chinese Academy of Sciences, Biometrics Consortium Conf., 2006

	Feature Set	
Data Set	Haar-like	Weighted dipoles
Face Det.	55.22% ± 4.23(3.0)	87.74% ± 2.85(1.0)
Traffic Signs Det.	63.30% ± 4.41(2.0)	87.92% ± 3.61(1.0)
Pedestrians Det.	54.00% ± 6.68(3.0)	88.40% ± 2.40(1.0)
Cars Det.	69.65% ± 7.54(3.0)	95.21% ± 3.28(1.0)
Text Det.	45.84% ± 5.75(3.0)	80.35% ± 5.08(1.0)
Average Rank	2.80	1.00

Baro and Vitria, LNCS 2008

*"In a comparison with extensively used SIFT descriptors, Balas & Sinha's dipole descriptors are more efficient while containing 6 times less information. This allows the complete retrieval to be both more efficient and faster. This strategy ranked first in **ImagEval** benchmark, which is the only competition including a transformed image recognition task."*

A. Joly, INRIA, CIVR 2007

Fly in the ointment: Little physiological evidence of dissociated dipoles

Processing Framework Proposed by Marr