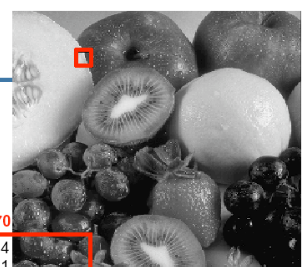
Introduction to visual computation and the primate visual system

- Problems in vision
- Basic facts about the visual system
- Mathematical models for early vision
- Marr's computational philosophy and proposal
- 2.5D sketch example stereo computation

15-883 Computational models of neural systems. Visual system lecture 1. Tai Sing Lee.

Grayscale Image



y = \ x =	58	59	60	61	62	63	64	65	66	67	68	69	70		
41	210	209	204	202	197	247	143	71	64	80	84	54	54		
42	206	196	203	197	195	210	207	56	63	58	53	53	61		
43	201	207	192	201	198	213	156	69	65	57	55	52	53	60	50
44	216	206	211	193	202	207	208	57	69	60	55	77	49	62	61
45	221	206	211	194	196	197	220	56	63	60	55	46	97	58	106
46	209	214	224	199	194	193	204	173	64	60	59	51	62	56	48
47	204	212	213	208	191	190	191	214	60	62	66	76	51	49	55
48	214	215	215	207	208	180	172	188	69	72	55	49	56	52	56
49	209	205	214	205	204	196	187	196	86	62	66	87	57	60	48
50	208	209	205	203	202	186	174	185	149	71	63	55	55	45	56
51	207	210	211	199	217	194	183	177	209	90	62	64	52	93	52
52	208	205	209	209	197	194	183	187	187	239	58	68	61	51	56
53	204	206	203	209	195	203	188	185	183	221	75	61	58	60	60
54	200	203	199	236	188	197	183	190	183	196	122	63	58	64	66
55	205	210	202	203	199	197	196	181	173	186	105	62	57	64	63

What make vision difficult?

1. Projection of 3D scene into 2D array of numbers - *recovering the lost dimension*
2. Variability of object manifestations -- *invariance*
3. Multiple causes for generating images -- *disambiguation*
4. Occlusion and clutters - *figure-ground, attention.*

What does it mean to understand something *computationally?*

David Marr (1945-1980)

1. *Computational theory*
2. *Algorithms*
3. *Implementations.*

Marr (1981) Vision.

Computational theory

- What is the goal of the computation?
- Why is it appropriate?
- What is the logic of the strategy by which it can be carried out?
  1. *Computational constraints*
  2. *Prior knowledge*

Representation and algorithms

- How can the computational theory be implemented?
- What is the representation for the input and output?
- What is the algorithm for the transformation?

Representation and algorithms

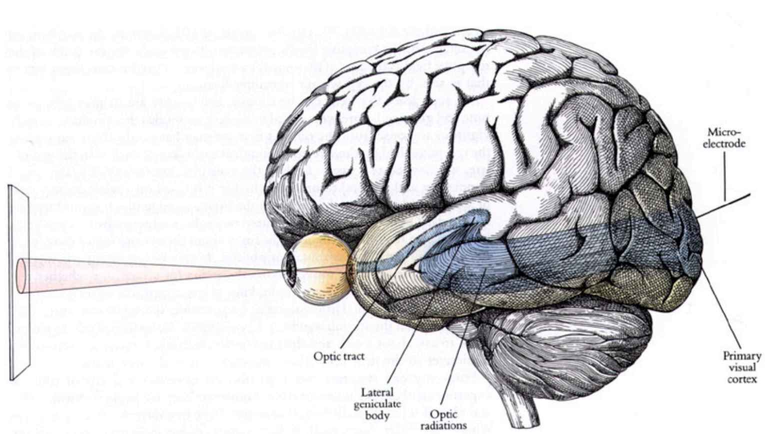
- How can the computational theory be implemented?
- What is the representation for the input and output?
- What is the algorithm for the transformation?

Processes and representations

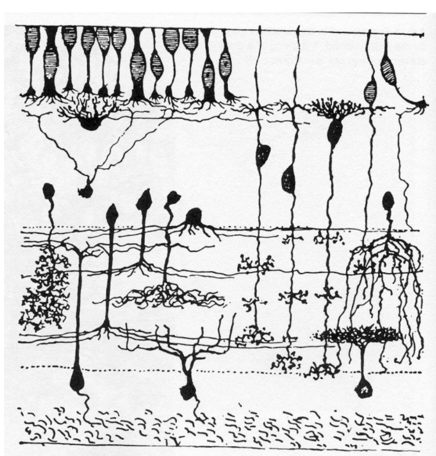
Hardware implementation

- How can the representation and algorithm be realized physically?

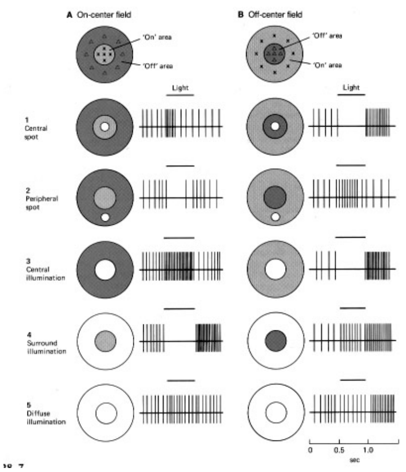
What is known about the visual system at the time?

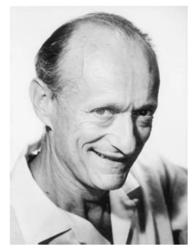


Cajal's microscopic study of the retina

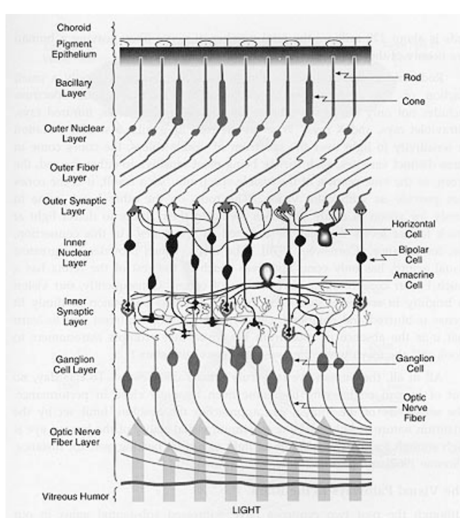


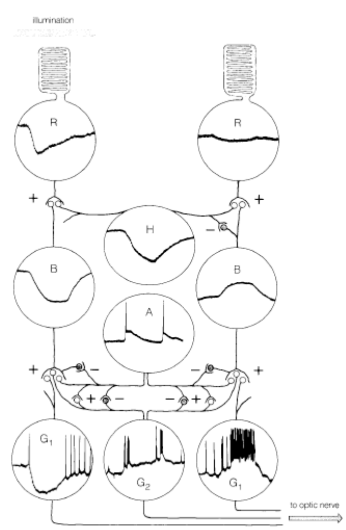
On-off center surround receptive fields of intact retina, cells responded primarily to contrast and to moving stimuli rather than diffused light.





Steven Kuffler (1953)

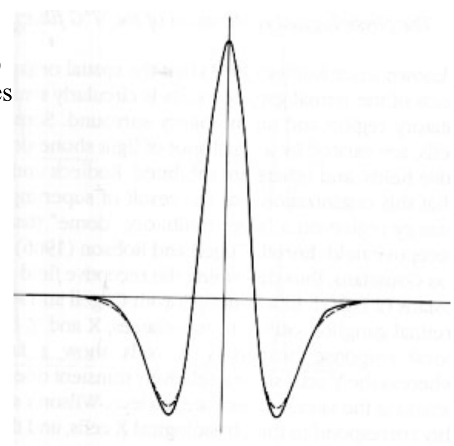




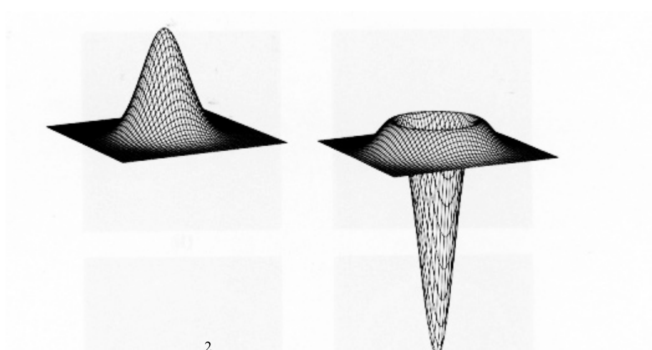
John Dowling

Laplacian of Gaussian operator

- DOG (difference of Gaussians) of ratio 1:1.6 best approximates a Laplacian of Gaussian filter (Marr and Hildreth,1980)



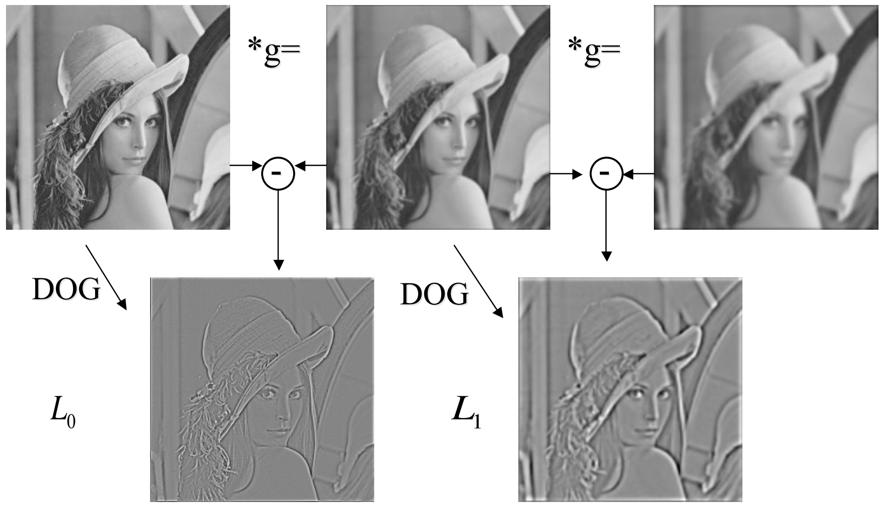
Laplacian of Gaussian



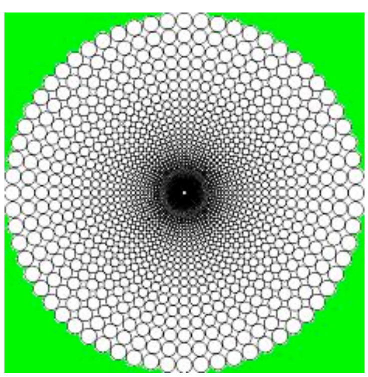
$$G_\sigma(r) = \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \qquad \nabla^2 G(r) = \frac{-1}{\pi\sigma^4}\left(1 - \frac{r^2}{2\sigma^2}\right) e^{\frac{-r^2}{2\sigma^2}}$$

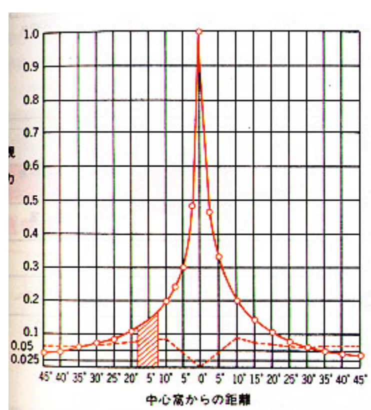
where r is the radial distance from the origin.

Difference of Gaussian smoothed images



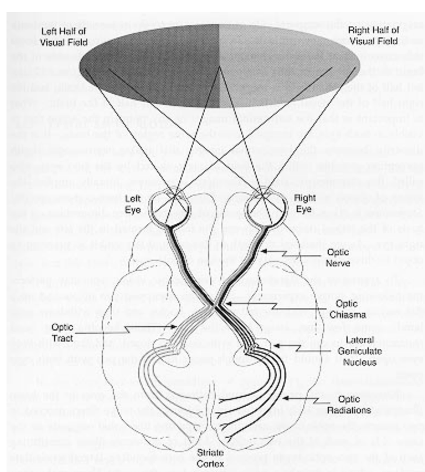
Retinal receptive fields and resolution



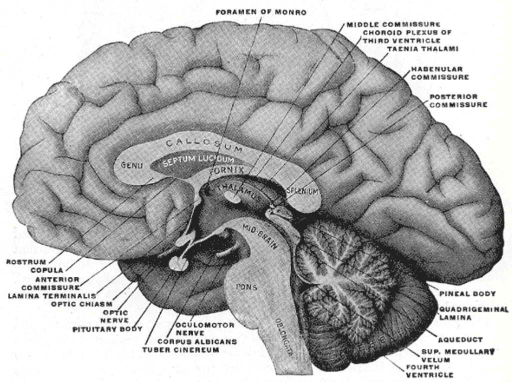


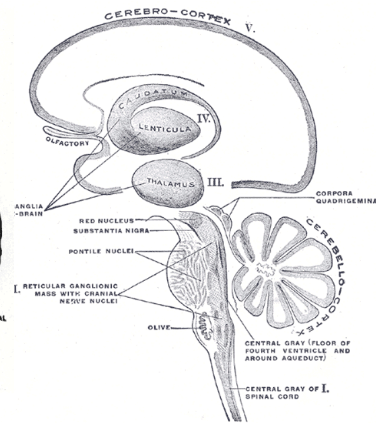
Organization of visual pathways from retina to cortex

- Optic Nerve - digital signal
- Optic Chiasma
- Optic tracts
- Lateral geniculate nucleus
- Optic radiation
- Primary visual cortex (Striate cortex, V1, area 17)
- Extrastriate cortex



Thalamus





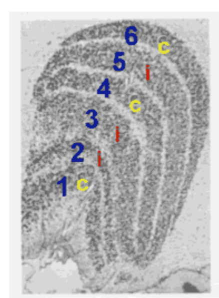
LGN anatomy

6 layers sandwiched together:

Layers 1 and 2: magnocellular (M) layers, large cells, fast processing and conducting, motion, gross features, monochromatic, transient respons

Layer 3,4,5,6: parvocellular (P) layers, small cell bodies, thin fibre, high-resolution, fine details, sustained responses, color coded.

Between layers: unmyelinated neural dendrites and axons, also contains interlaminar or koniocellular (K) layer. Functionally distinct third channels.



layers 1 and 2
- *magnocellular*

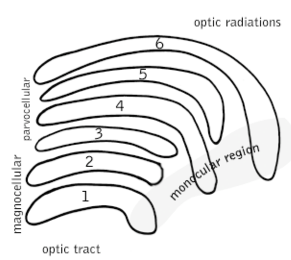
layers 3-6
- *parvocellular*

layers 2, 3 and 5
- *ipsilateral* eye

layers 1, 4 and 6
- *contralateral* eye

1 mm

lateral geniculate nucleus (LGN)

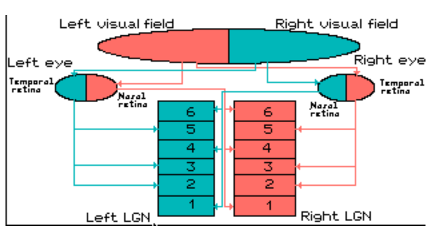


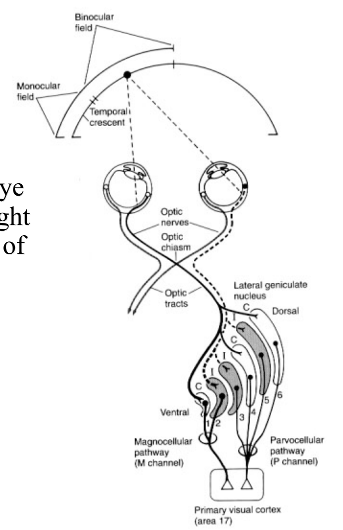
Functional difference between magnocellular and parvocellular LGN neurons

	Parvo	Magno
Color sensitivity	High (cones)	Low (cones+rods)
Contrast sensitivity	Low	High
Spatial resolution	High	Low
Temporal resolution	Slow	Fast
Receptive field size	Small	Large

LGN monocular retinotopic maps from both eyes

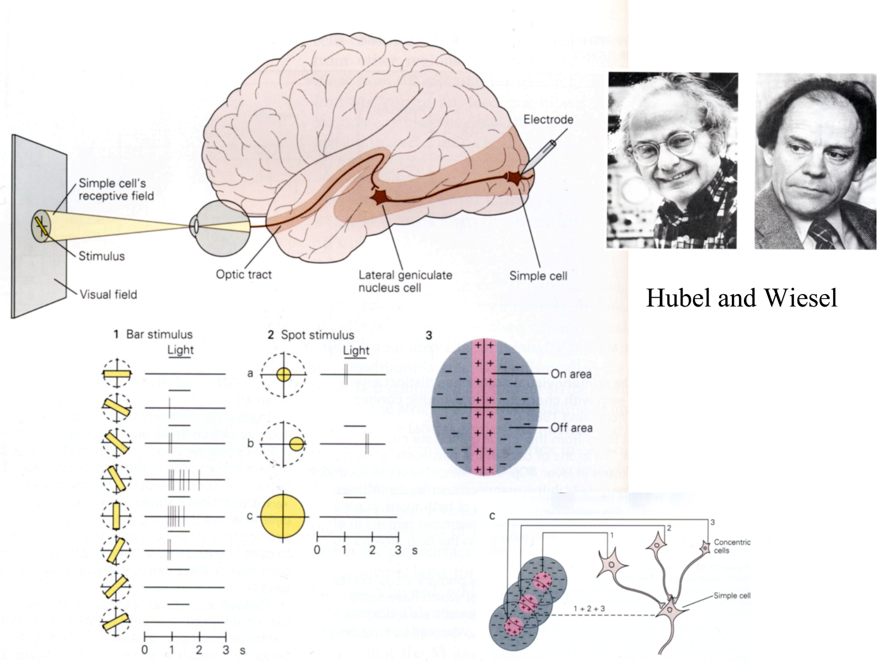
Input from the right hemi-retina of each eye project orderly to different layers of the right LGN to create 6 complete representations of the left visual hemi-field





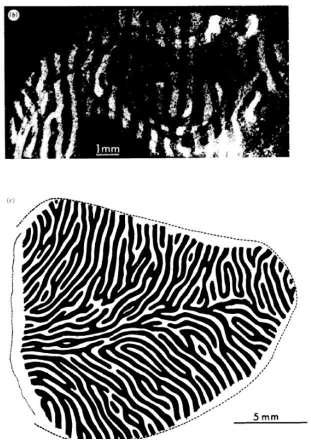
What are the differences between retinal and LGN neurons?

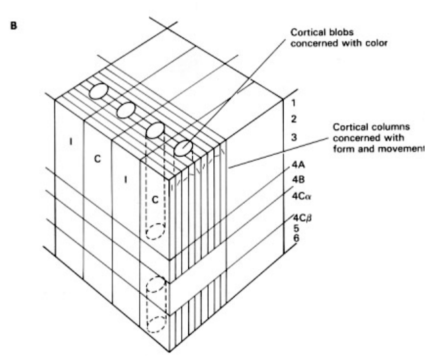
1. Broad attributes resemble retinal ganglian cells
2. Contrast gain control strengthened.
3. RF with a center and a larger surround.
4. Biphasic temporal kernel in both center and surround.
5. LGN receives feedback, but not retina.



Hubel and Wiesel

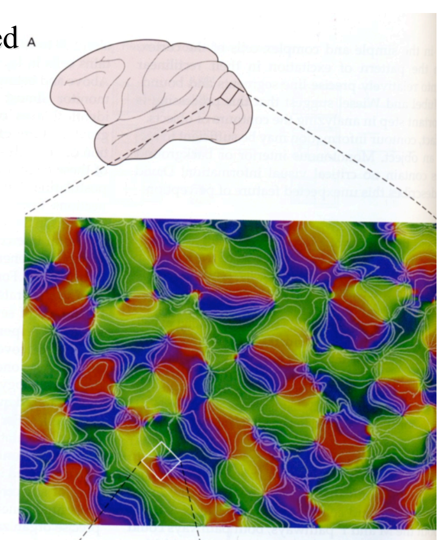
Ocular dominance columns and hypercolumns

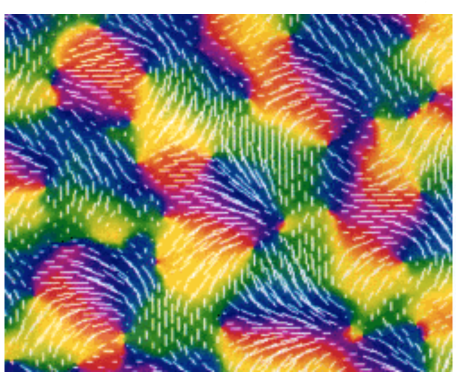




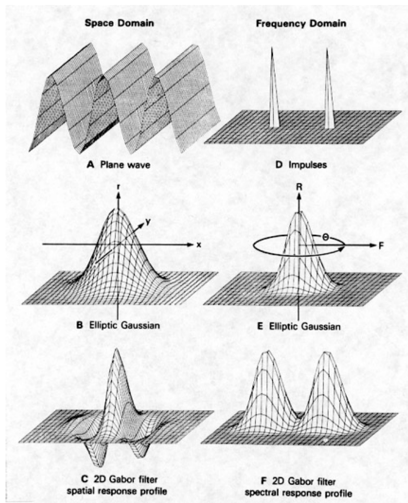
Cells tuned to a variety of visual cues: color, orientation, disparity, motion direction.

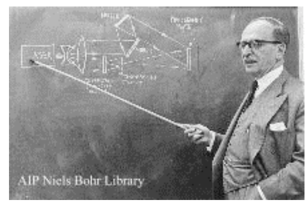
The actual topological map revealed by optical imaging.



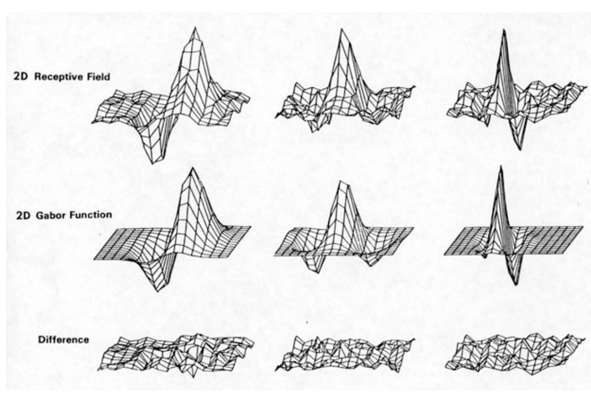


Gabor filters are spatial frequency analyzers

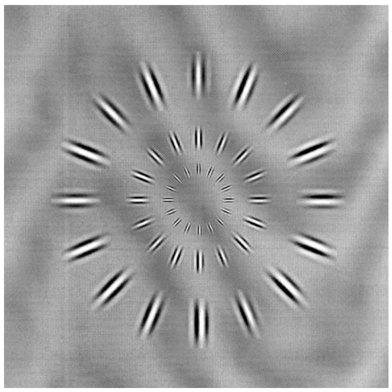


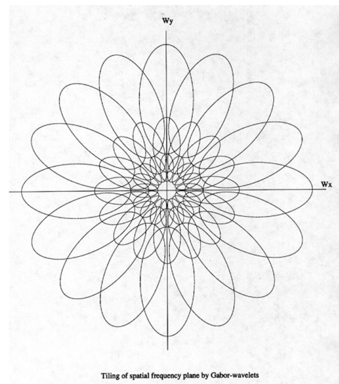


Daugman (1985) and others proposed simple cells can be modeled by Gabor filters. Jones and Palmer (1988) confirmed Gabor fit.



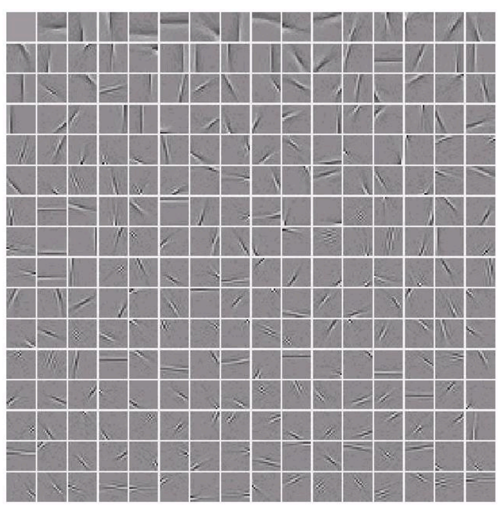
V1 neurons modeled as Gabor wavelets, wavelets can efficiently encode images



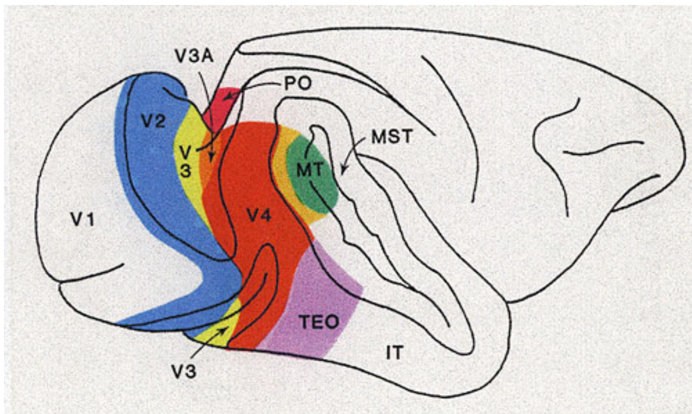


Lee (1996) Image representation using 2D Gabor wavelets. PAMI. 18(10): 959-971.

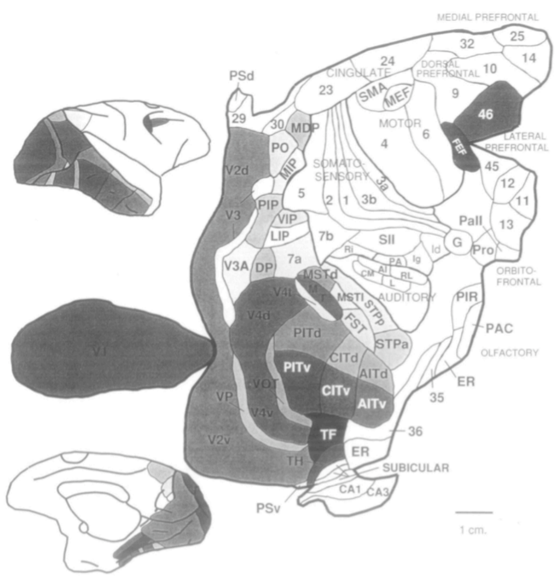
Gabor wavelet like structures can be learned as sparse efficient codes from natural image patches -- Olshausen and Field (1996),



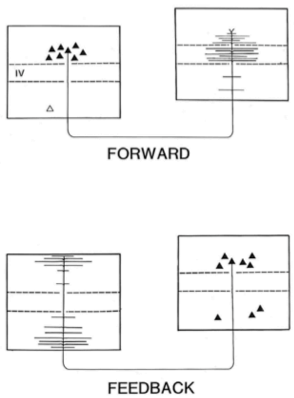
Visual areas in the visual system



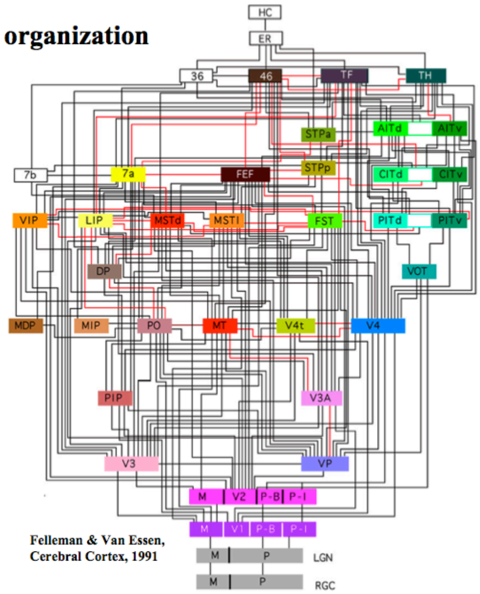
Cortical areas flat map



Distributed hierarchical organization

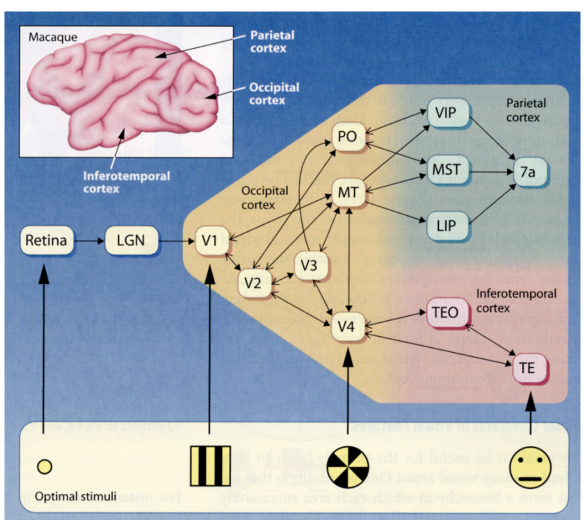


Maunsell & Van Essen, J. Neurophysiol., 1983

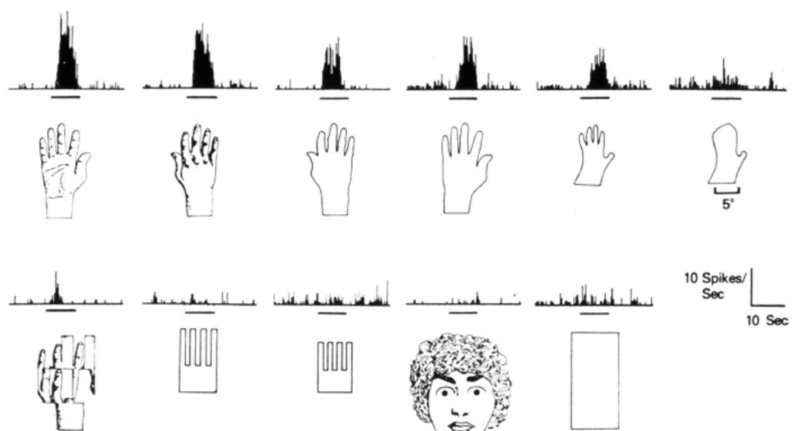


Felleman & Van Essen, Cerebral Cortex, 1991

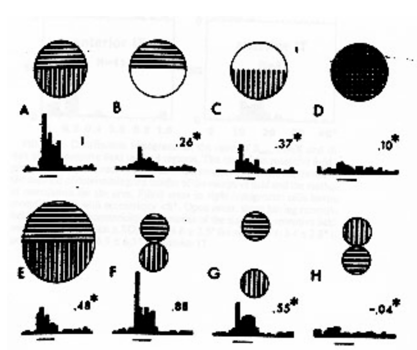
Ventral and dorsal streams

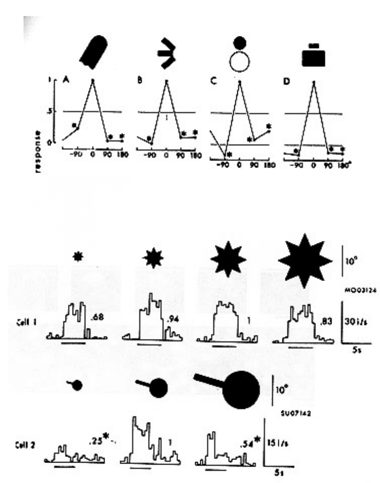


Object detector neurons in IT



Combination Coding and Invariance





Marr's proposal on visual processing

Digitized Image

Filtering, Edge detection, Chunking

Primal Sketch

Depth, surfaces, occlusion, figure-ground

2 1/2 D Sketch

3D structural model and parts

3D Model

Comparison with memory prototypes

Object Recognition /Scene Description

Marr's proposal on visual processing

Digitized Image

Filtering, Edge detection, Chunking

V1,V2 — Primal Sketch

Depth, surfaces, occlusion, figure-ground

V2,V4 — 2 1/2 D Sketch

3D structural model and parts

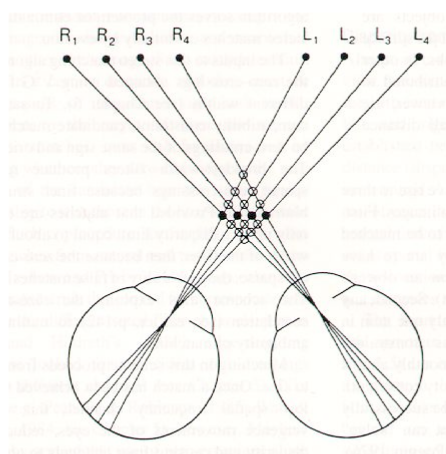
IT — 3D Model

Comparison with memory prototypes

Object Recognition /Scene Description

Julesz random dot stereogram

Stereo Correspondence is Hard

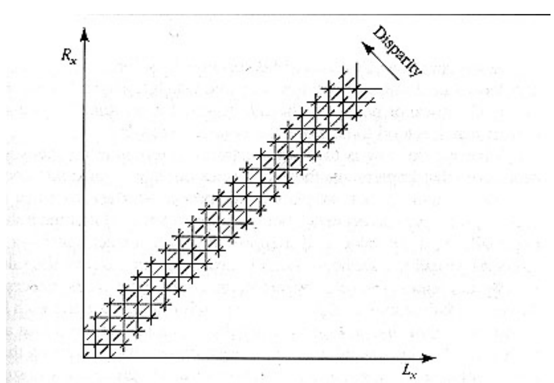


Computing 2.5D sketch -- e.g. stereopsis

Computational constraints

1. *Compatibility*: Black dots can match only black dots.
2. *Uniqueness*: Almost always, a black dot from one image can match no more than one black dot from the other image.
3. *Continuity*: The disparity of the matches varies smoothly almost everywhere over the image.

Marr and Poggio (Marr 1976).



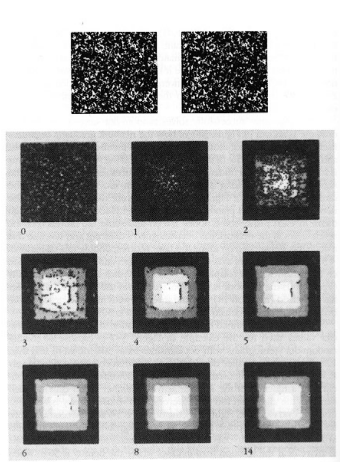
(a)

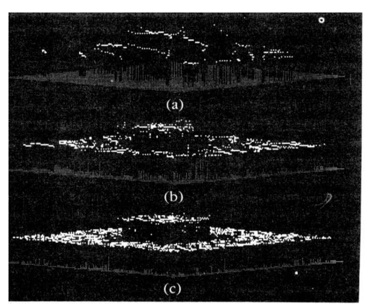
- Left and right eyes
- Continuous lines = line of sights
- Intersection = possible disparity values
- Dotted diagonal lines = lines of constant disparity (planar surface).
- How to implement the rules?

Iterative (Relaxation) Algorithm

$$C^{t+1}_{x,y;d} = \sigma\{ \sum_{x',y',d' \in S(x,y,d)} C^{t}_{x',y',d'} - \varepsilon \sum_{x',y',d' \in O(x,y,d)} C^{t}_{x',y',d'} + C^{0}_{x,y,d} \}$$

where $C^{t}_{x,y,d}$ denotes the state of the cell corresponding to the position (x, y) , disparity d and time t. It is binary. $S(x, y, d)$ is the local excitatory neighborhood, and $O(x, y, d)$ is the inhibitory neighborhood. ε is the inhibitory constant, and σ is the threshold function. C^0 is all the possible matches, including false targets, within the prescribed disparity range, added at each iteration to speed up convergence, can simply use to initialize.



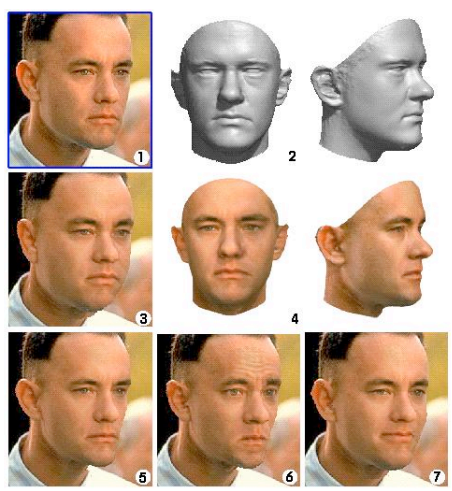


See also Samonds, Potetz and Lee (2007) NIPS for neural evidence of the computational constraints at work during stereo computation.

Computing 2.5D sketch -- e.g. shape from shading

Potetz (2007)

3D model



Blanz and Vetter (1999)

Summary

- Why vision is difficult?
- What Marr and we know about the biological visual system?
- Contrast, edges and Laplacian and Gabor filters.
- Pandemonium model and Fukushima's neocognitron
- Marr's computational philosophy and proposal
- Some outstanding realizations of Marr's vision.

- Next lecture: how the hierarchical visual system might compute?

Readings

- Van Essen, D. Anderson, C, Felleman, DJ (1992) Information processing in the primate visual system: an integrated systems perspective. *Science*, vol. 225, no. 5043, pp. 419-423.
- Marr, D. (1982) *Vision*, chapter 1. San Francisco: W. H. Freeman.
- Marr, D., and Poggio, T. (1976) Cooperative computation of stereo disparity. *Science*, vol. 194, no.462, pp. 283-287.