the quest for robotic vision

Peter Corke

Queensland University of Technology

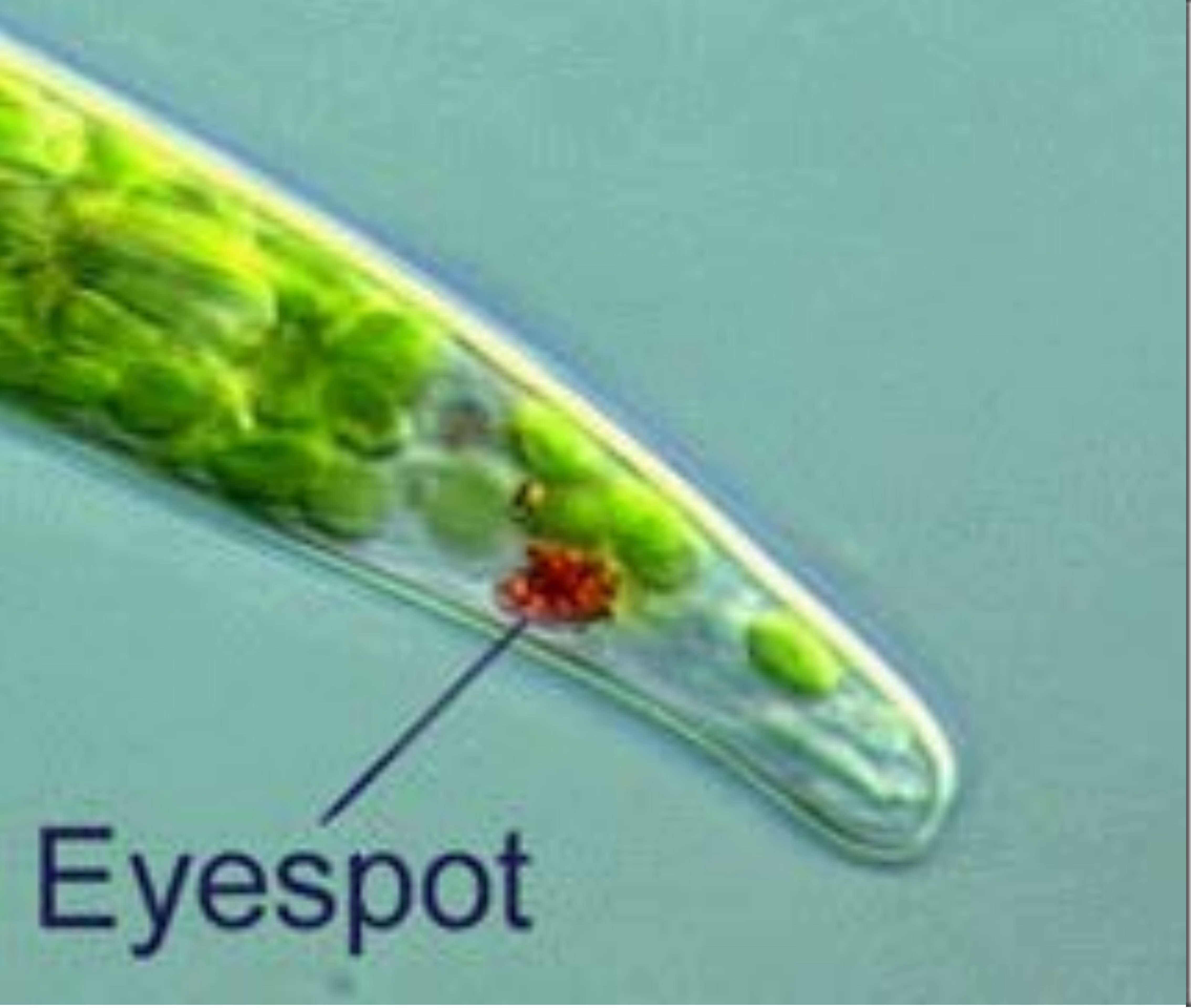
ARC Centre of Excellence for ROBOTIC VISION

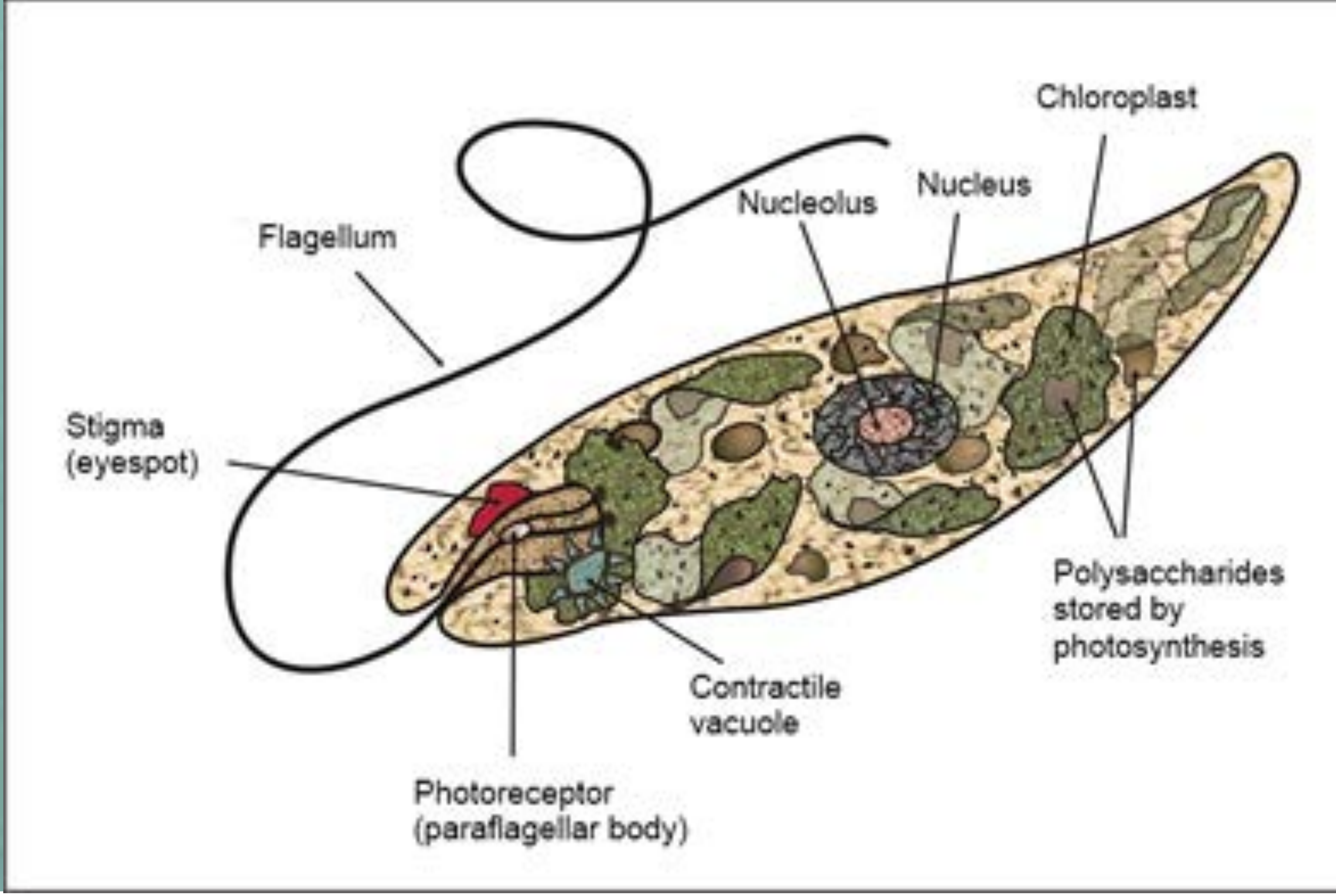
in the beginning


t = -13.8B years


t = -4B years

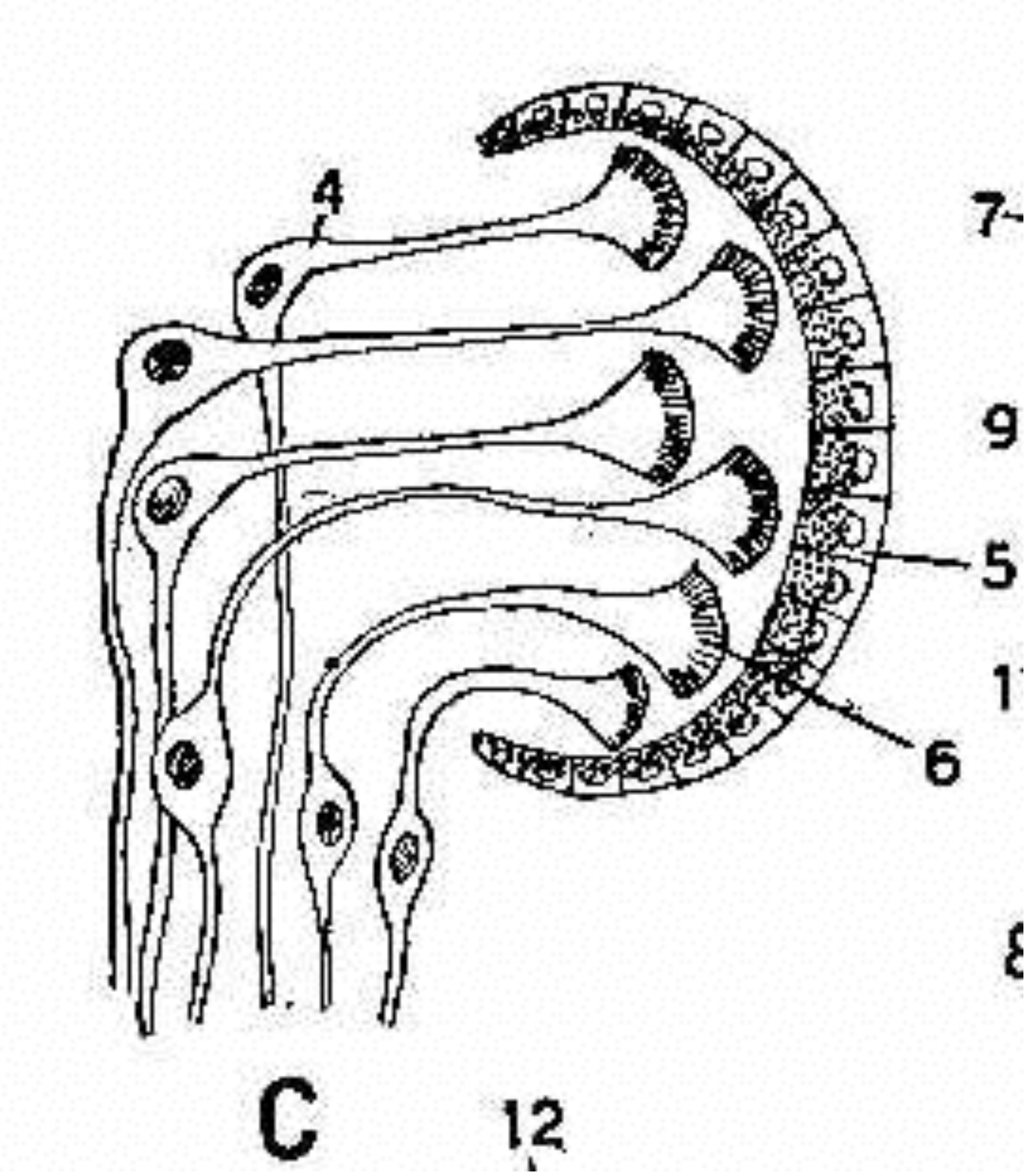

t = -3.5 years





evolution invents a light sensor

t = -600M years



Planarium flat worm

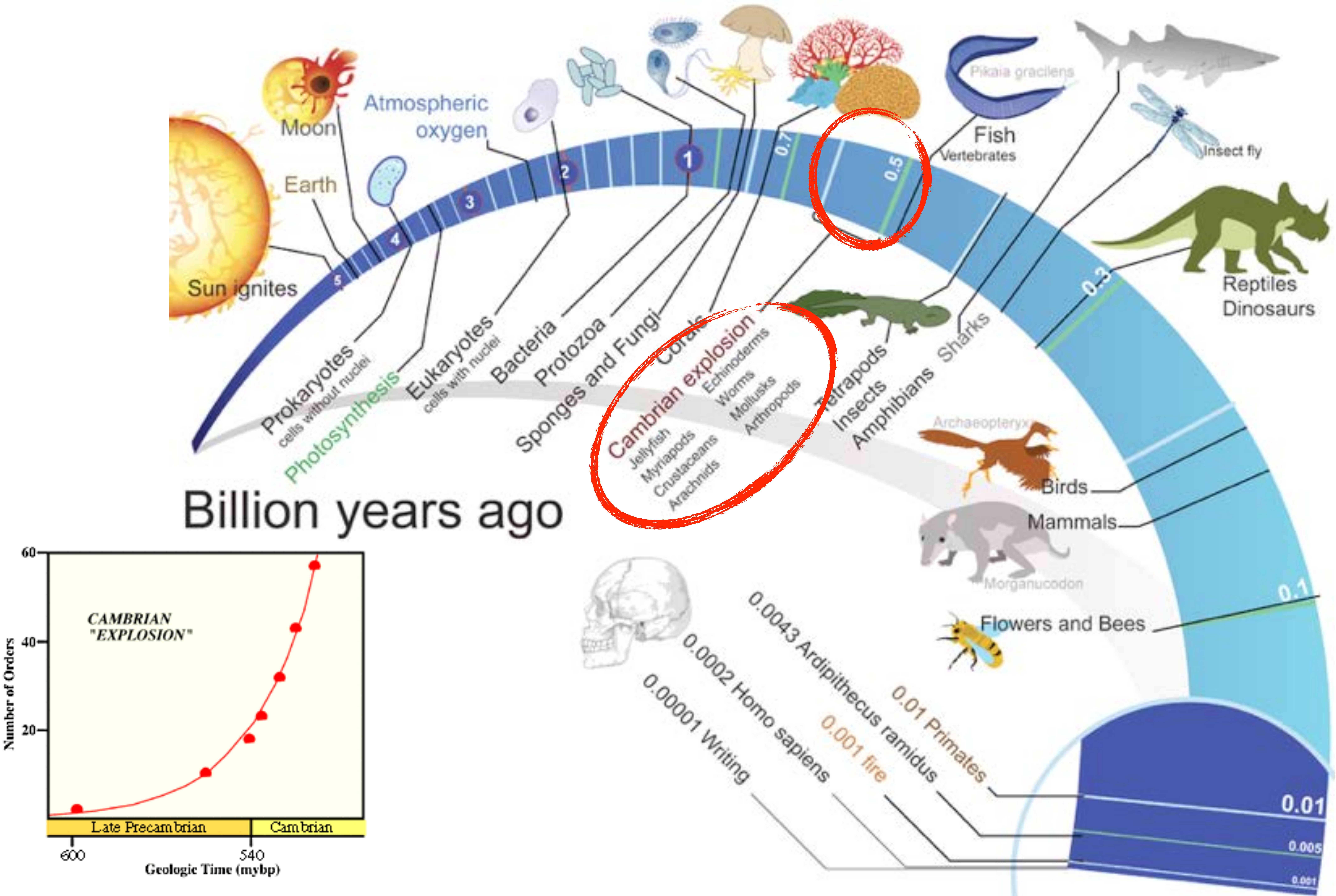
several eyespots make an eye

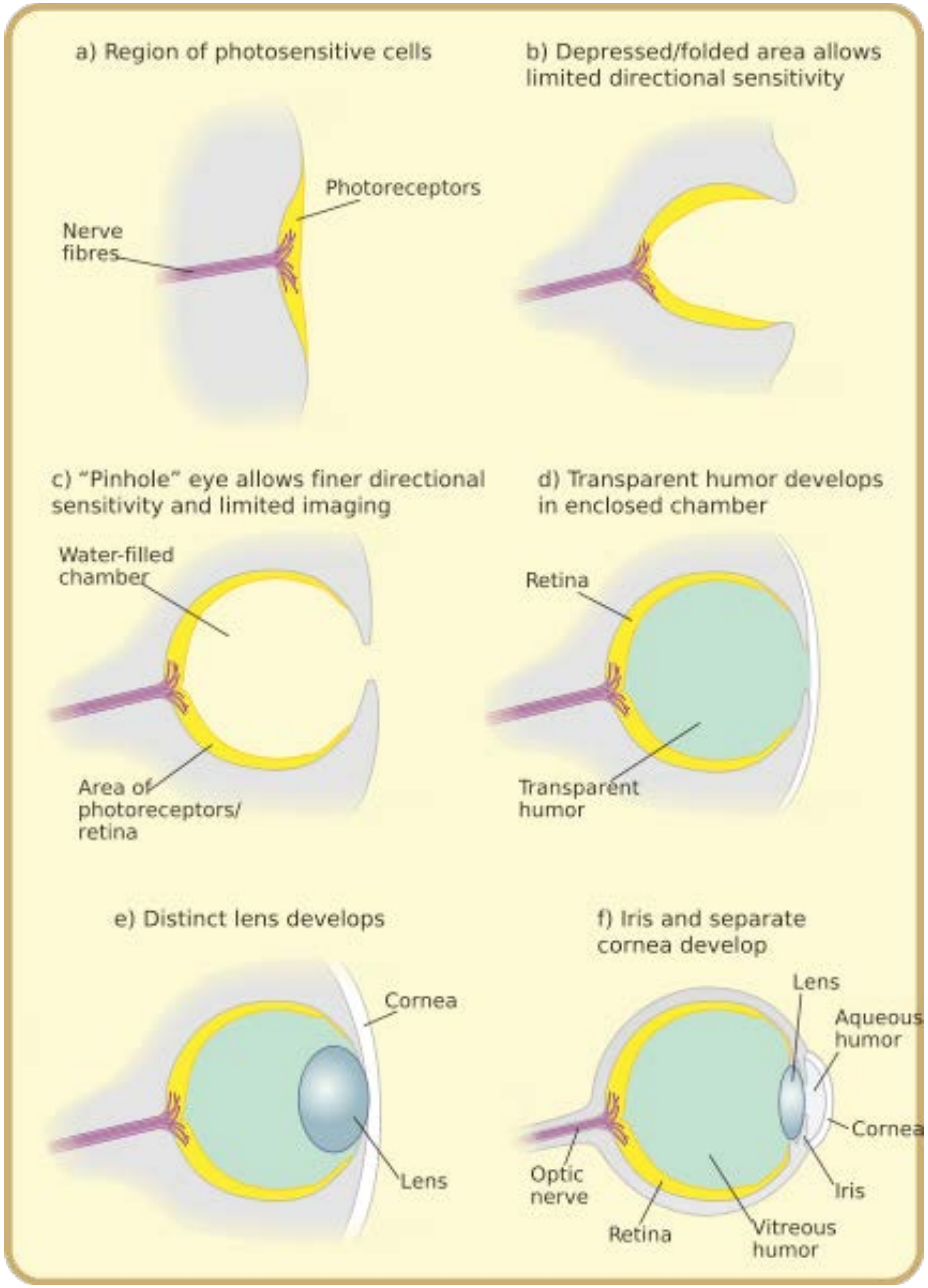
t = -550M years

trilobyte: successful marine animal for 270M years

t = -521M years

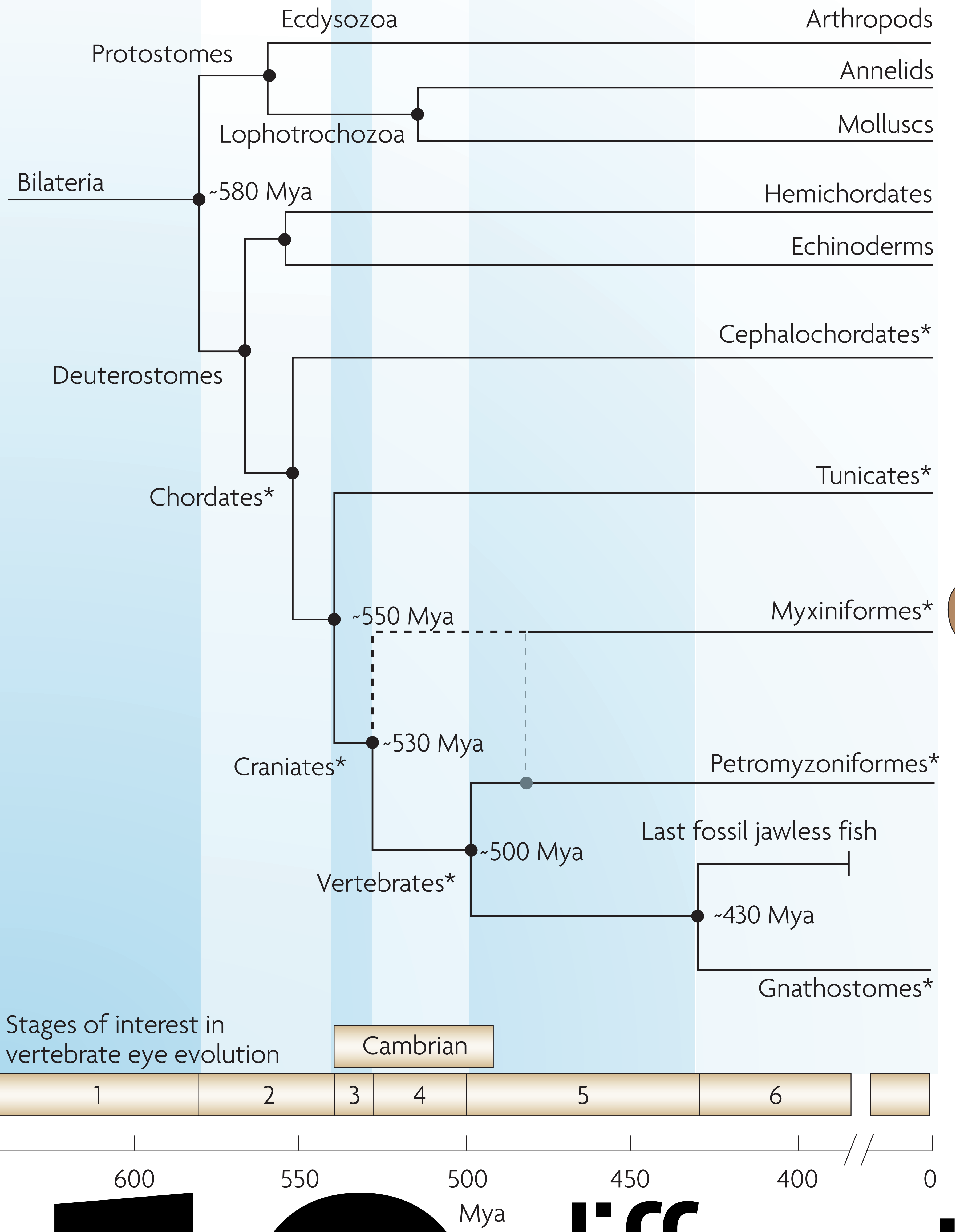

lamprey eye is very
similar to our own eye
t = -500M years

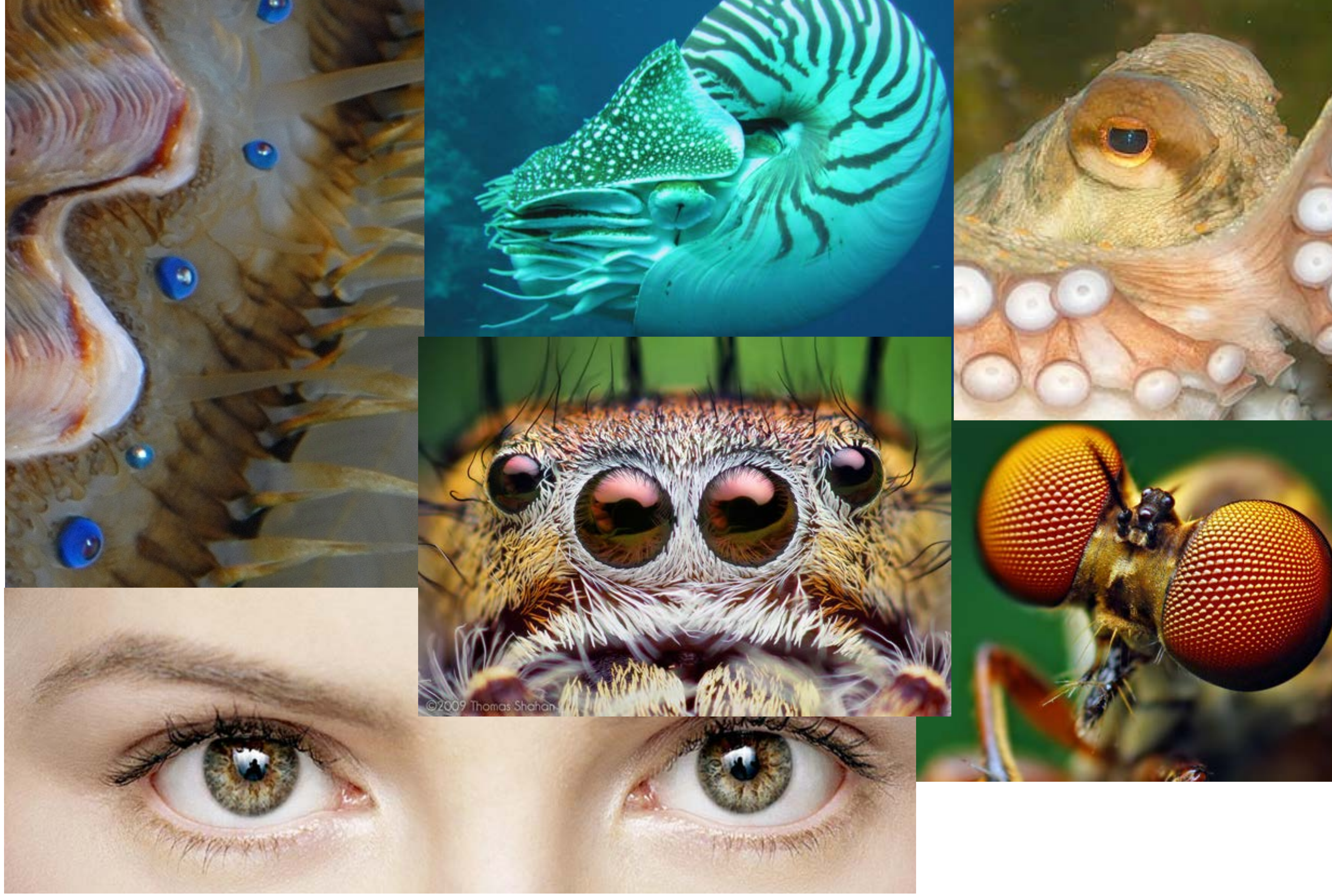

Billion years ago
Sun ignites
Earth
Moon
Atmospheric oxygen
Prokaryotes
cells without nuclei
Photosynthesis
Eukaryotes
cells with nuclei
Bacteria
Protozoa
Sponges and Fungi
Corals
Cambrian explosion
Jellyfish
Myriapods
Crustaceans
Arachnids
Echinoderms
Worms
Mollusks
Arthropods
Tetrapods
Insects
Amphibians
Sharks
Fish
Vertebrates
Pikaia gracilens
Insect fly
Reptiles
Dinosaurs
Archaeopteryx
Birds
Mammals
Morganucodon
Flowers and Bees
0.0043 Ardipithecus ramidus
0.0002 Homo sapiens
0.00001 Writing
0.01 Primates
0.001 fire
5
4
3
2
1
0.7
0.5
0.3
0.1
0.01
0.005
0.001
CAMBRIAN "EXPLOSION"
Number of Orders
60
40
20
Late Precambrian
Cambrian
600
540
Geologic Time (mybp)



To suppose that the eye, with all its inimitable contrivances… could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree… Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor…

- Charles Darwin (1809–1882)





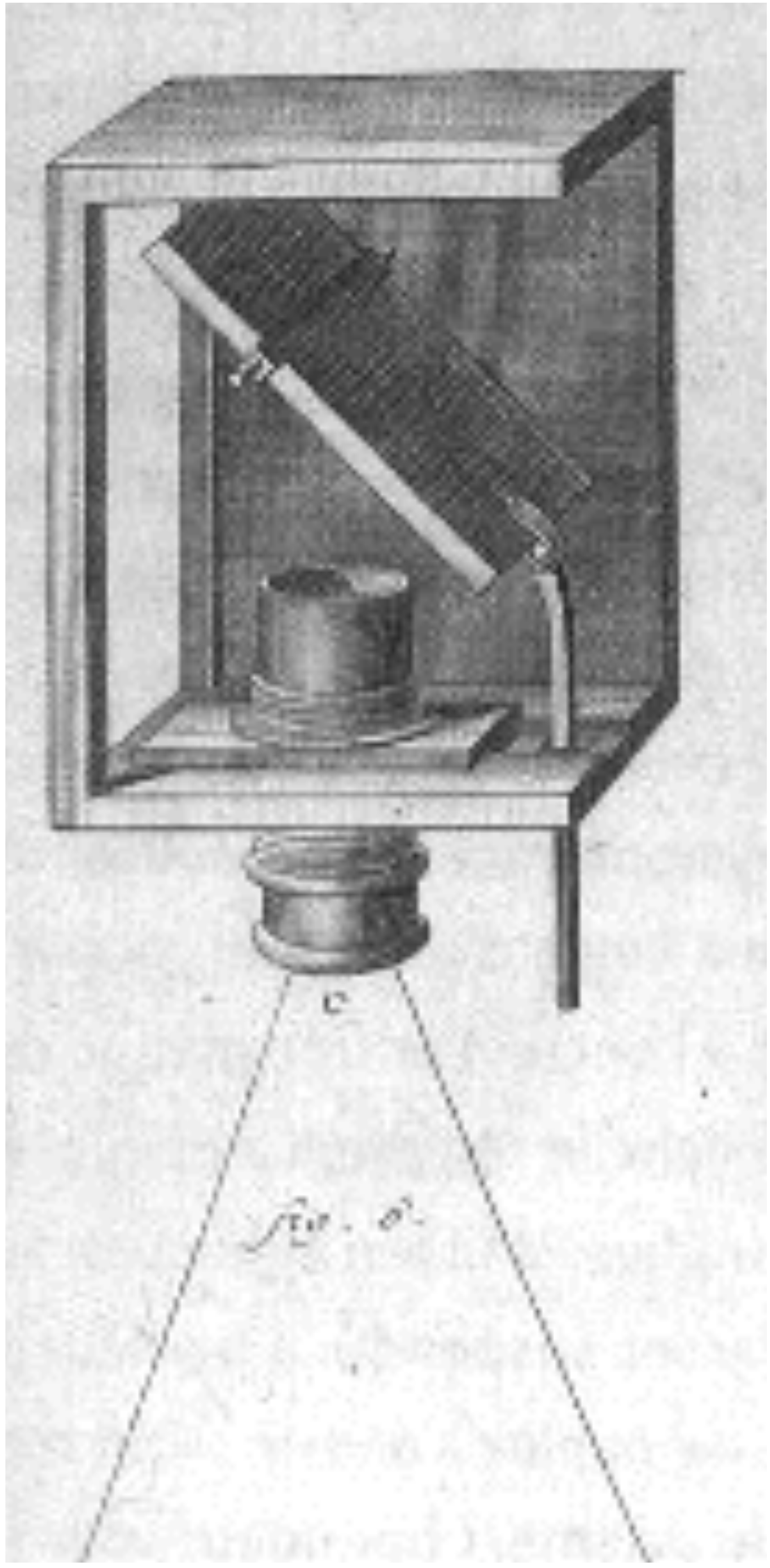
10 different eye designs

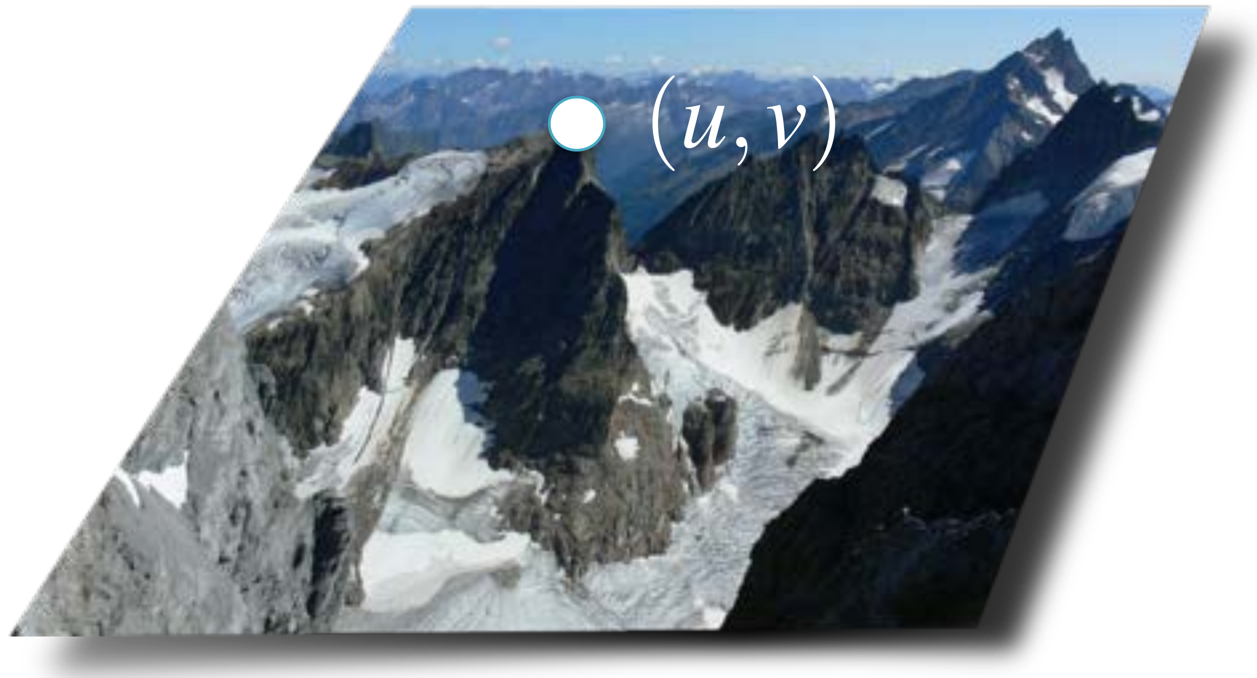
96% of animal species have eyes

retinal projection

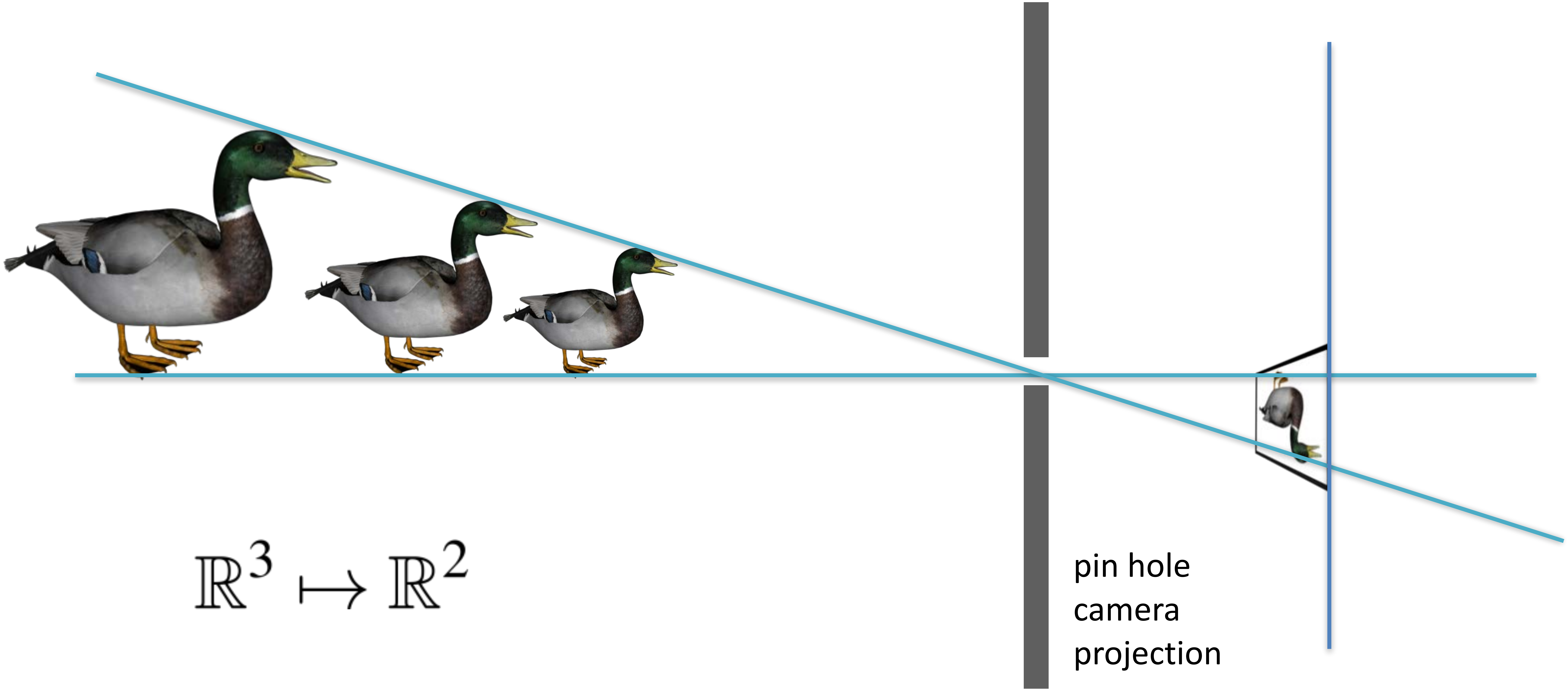
$$(X,Y,Z) \mapsto (\frac{fX}{Z}, \frac{fY}{Z})$$

$$\mathbb{R}^3 \mapsto \mathbb{R}^2$$



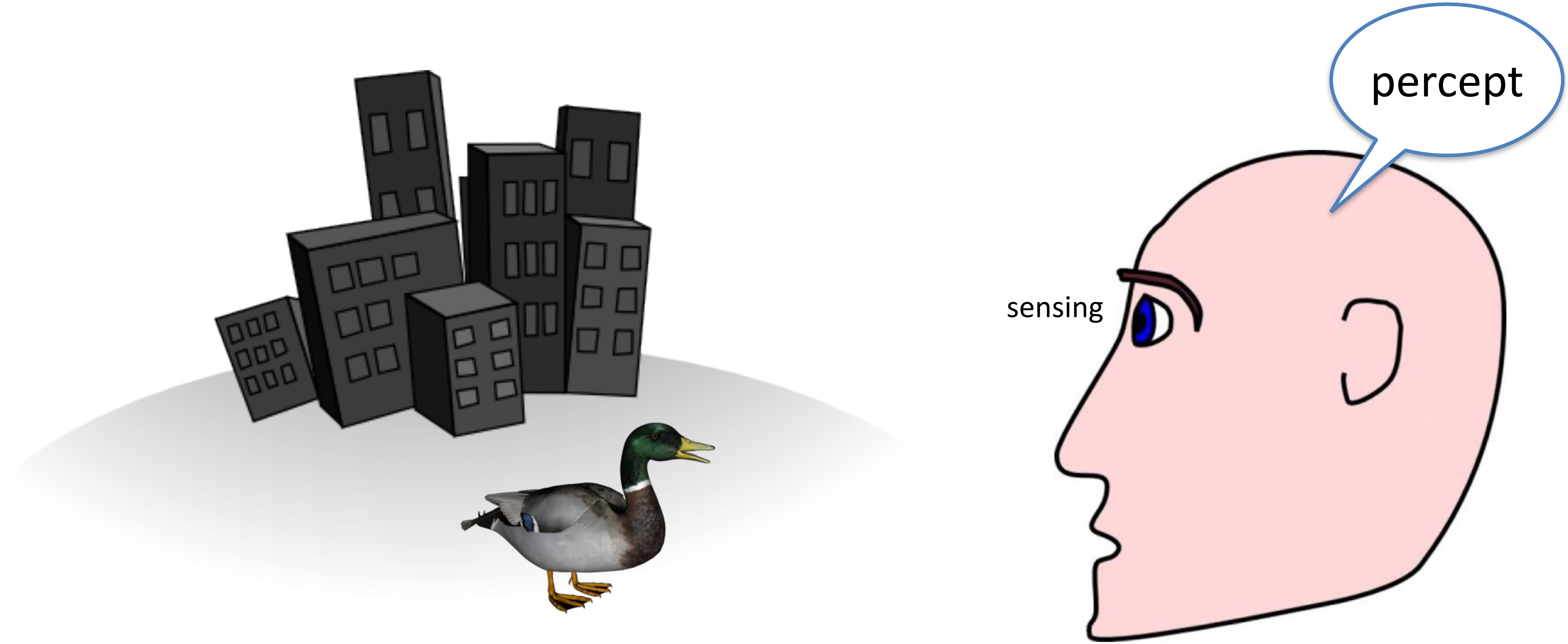


there is no unique inverse



$$\mathbb{R}^3 \mapsto \mathbb{R}^2$$

theories of vision


percept
sensing

Plato
428-348 BCE

Aristotle
384-322 BCE

Euclid
350-250 BCE?

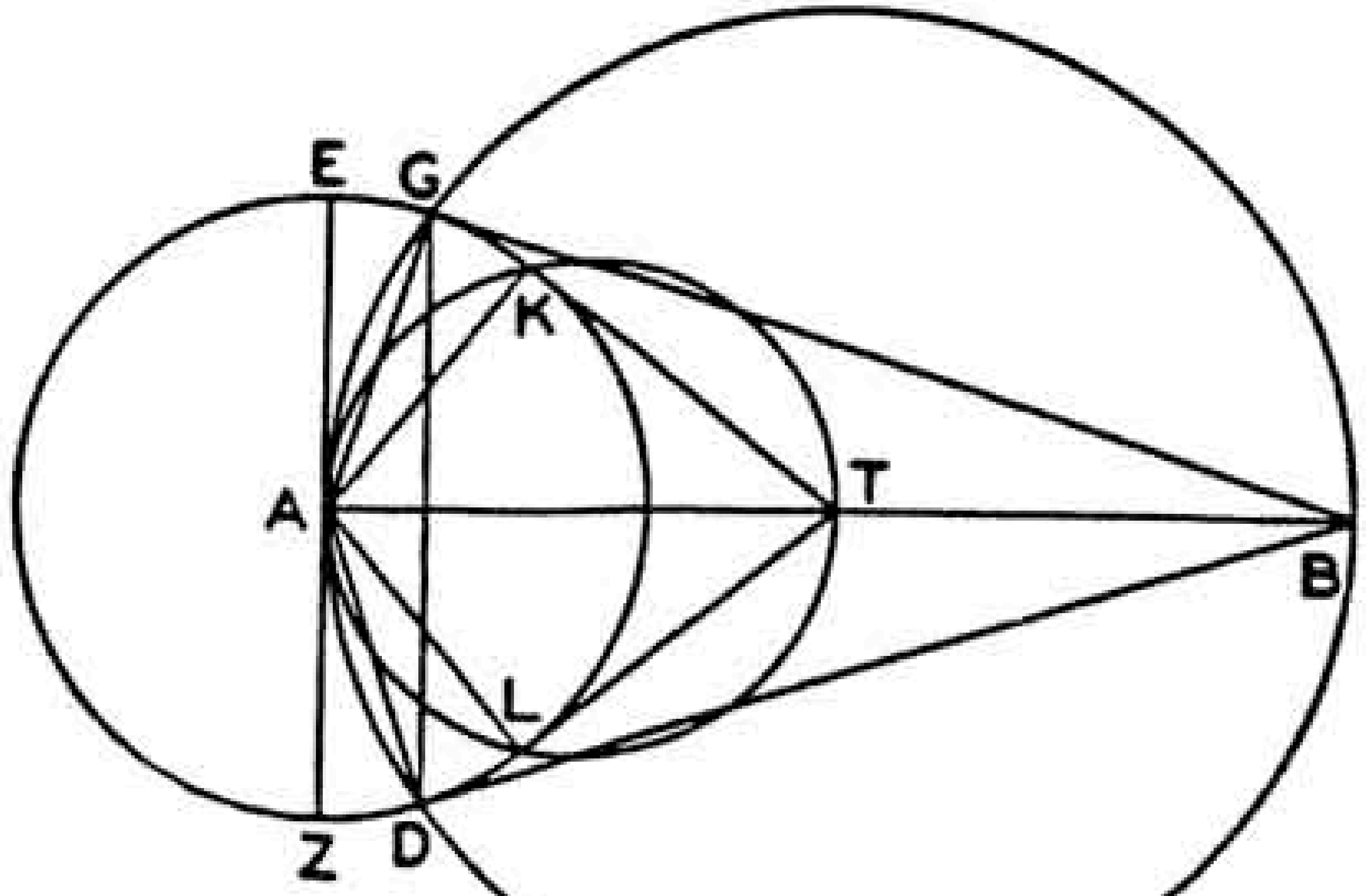
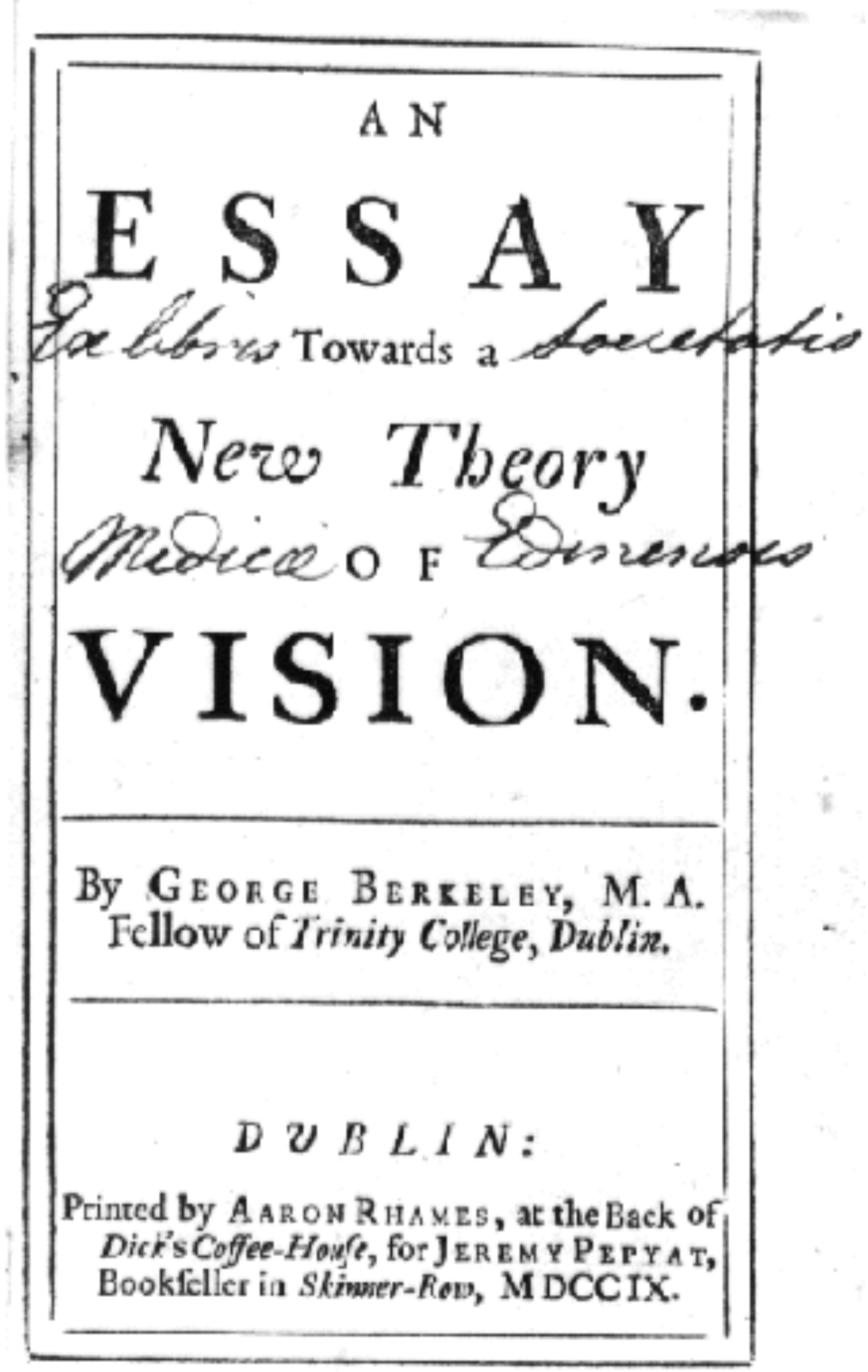


FIG. 24.

When the eye approaches the sphere, the part seen will be less, but will seem to be more. (Fig. 24.)

Let there be a sphere, of which the center is *A*, and let the eye be *B*, from which let the straight line *AB* be drawn. And around *AB* let the circle *GBD* be inscribed, and from the point *A* let the straight line *EZ* be drawn, perpendicular to the straight line *AB* in either direction, and let the plane be produced along *EZ* and *AB*. So it will make a circular section. Let it be *GEZD*, and let *GA*, *AD*, *DB*, *BG*, and *GD* be drawn. So, according to the theorem given before, the angles at the points *G* and *D* are right angles. Thus, *BG* and *BD*, whatever rays there are, touch the sphere. And the part of the sphere, *GD*, is seen by the eye, *B*. Now let the eye be moved nearer to the sphere, and let it be at *T*, from which let the straight line *TA* be drawn, and let the circle *ALK* be inscribed, and let the straight lines *TK*, *KA*,


AN

ESSAY

Ex libris Towards a Societatis

New Theory

Medicae OF Edinensis

VISION.

By GEORGE BERKELEY, M. A.
Fellow of *Trinity College, Dublin.*

DUBLIN:

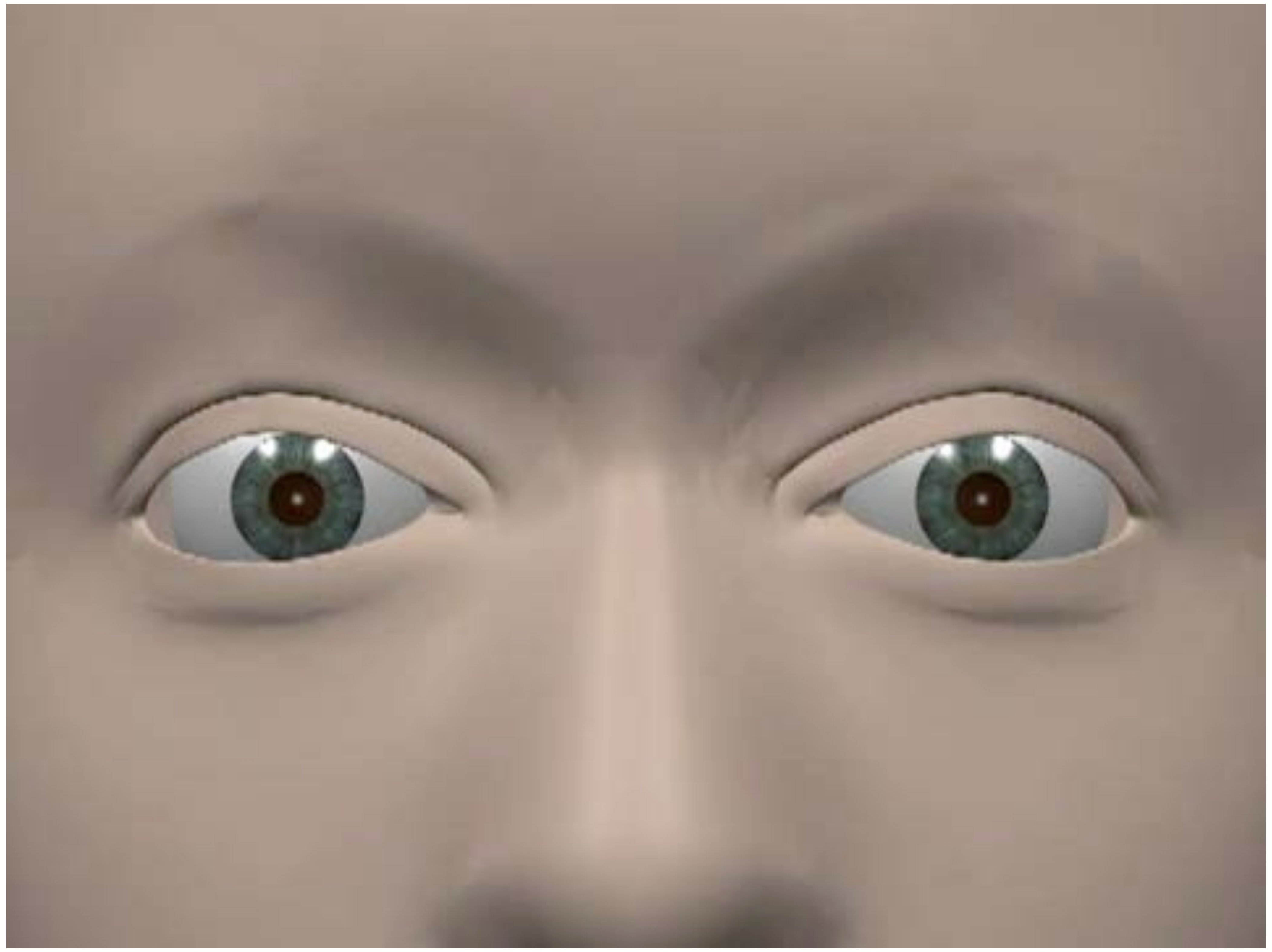
Printed by AARON RHAMES, at the Back of *Dick's Coffee-Houſe*, for JEREMY PEPYAT, Bookſeller in *Skinner-Row*, MDCCIX.

Bishop Berkeley
1709

- Distance is determined indirectly by visual cues.
- We **learn** to separate size from distance.

Binocular stereo

Accommodation



Convergence

Occlusion

Height in visual field

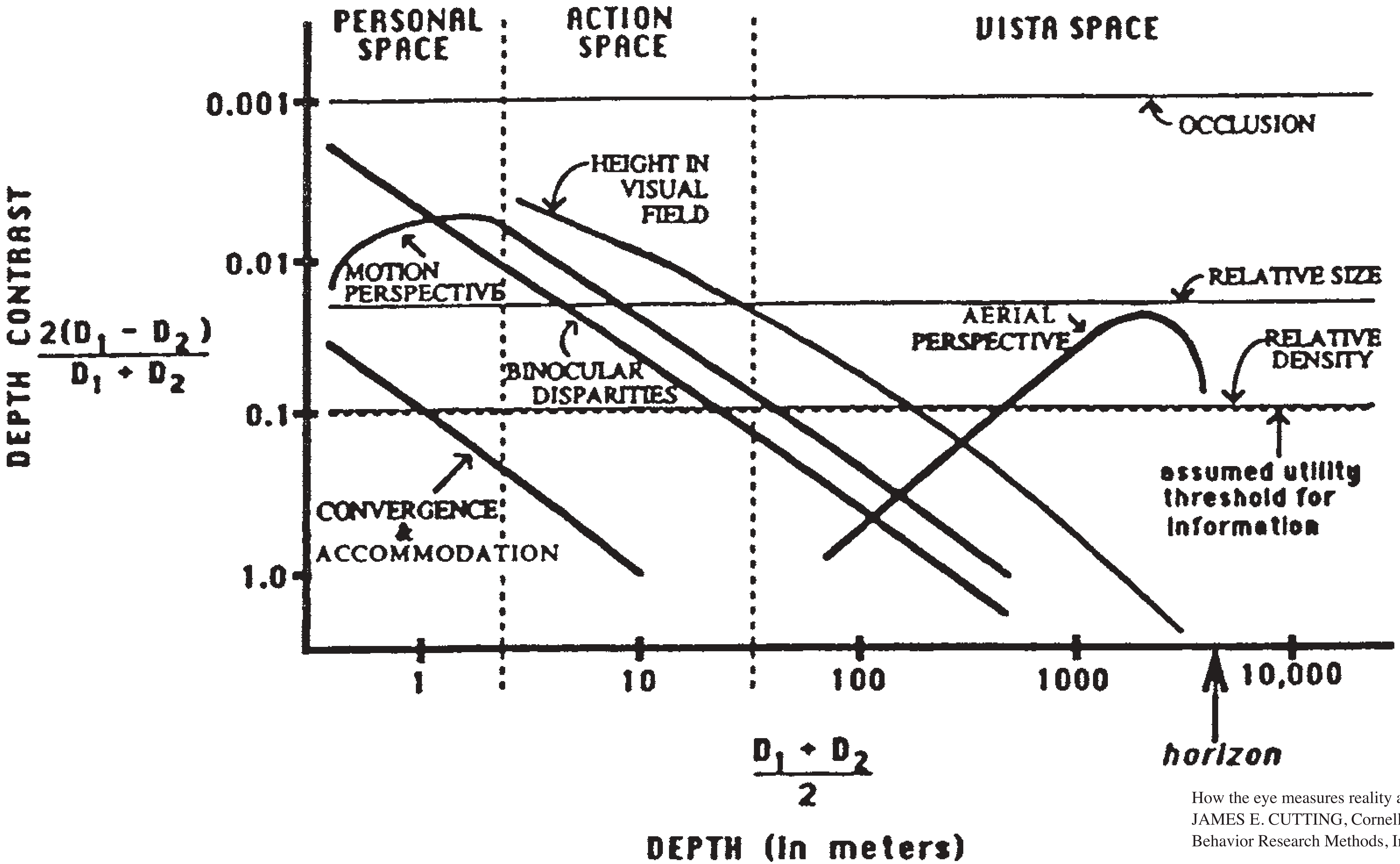
Relative size

distance: bush < cat < pillar

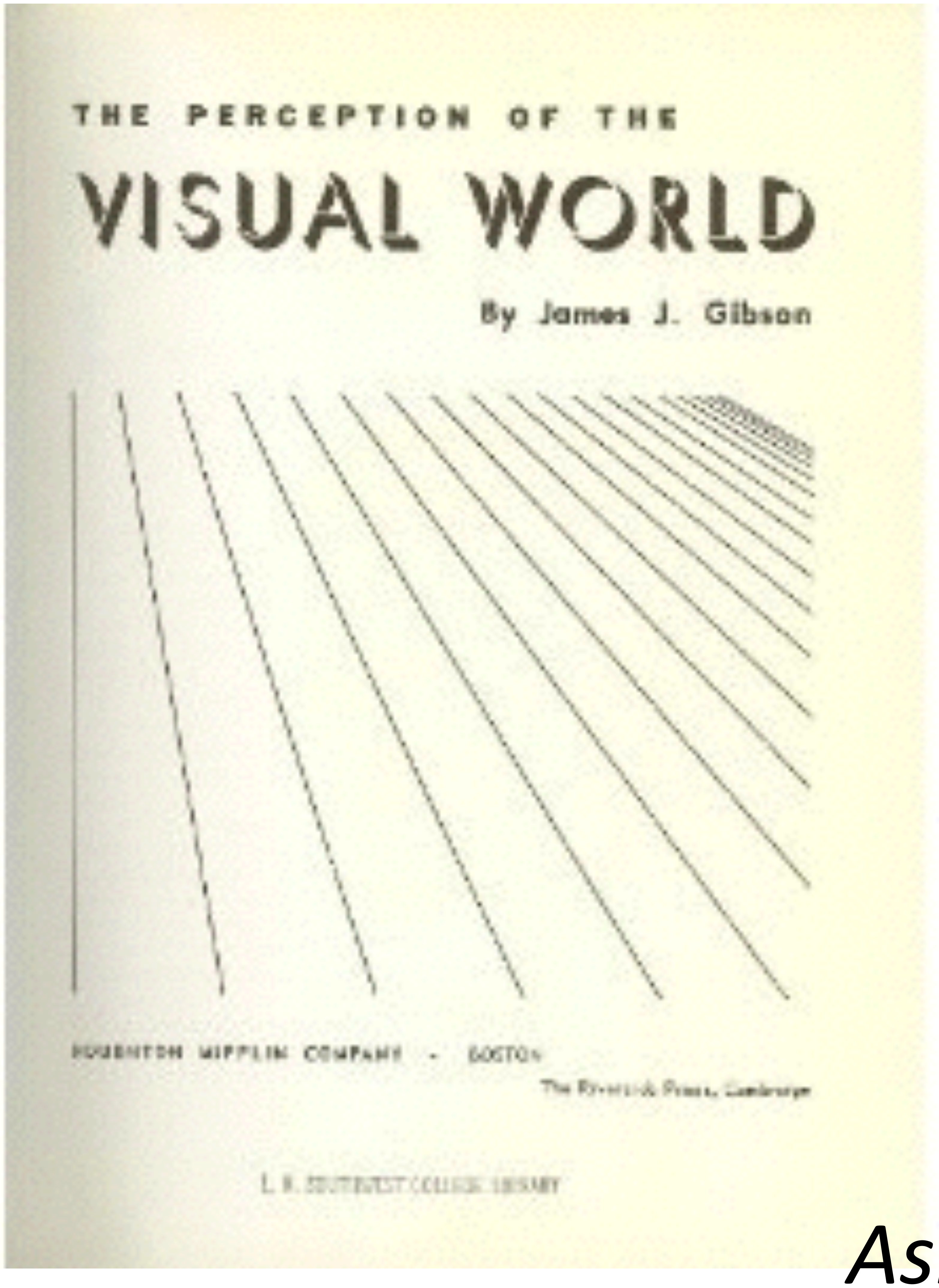
Texture density

Aerial perspective

Determining distance



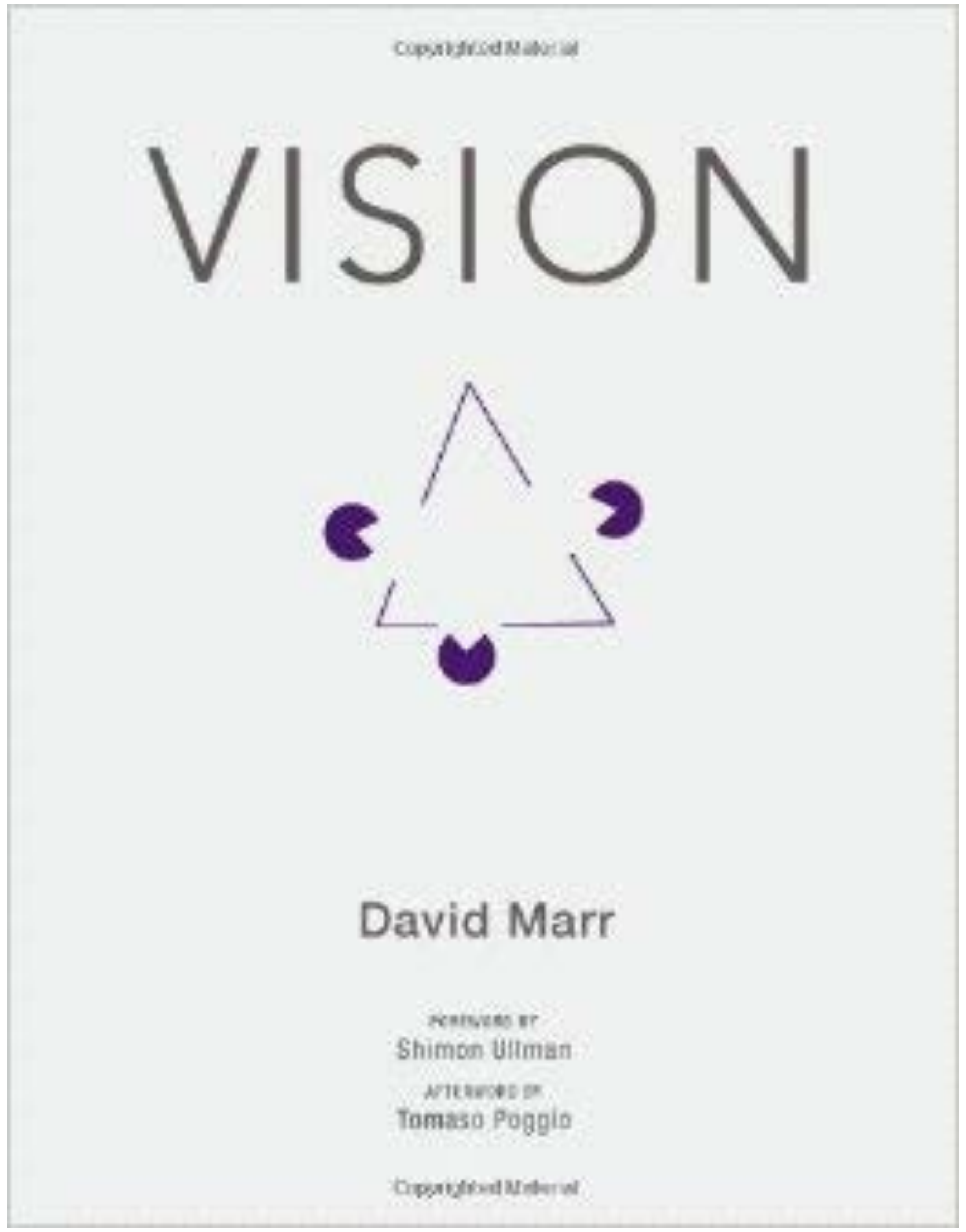
How the eye measures reality and virtual reality.
JAMES E. CUTTING, Cornell University, Ithaca, New York.
Behavior Research Methods, Instruments, & Computers 1997, 29 (1),


THE PERCEPTION OF THE
VISUAL WORLD
By James J. Gibson
HOUGHTON MIFFLIN COMPANY · BOSTON
The Riverside Press, Cambridge

1950

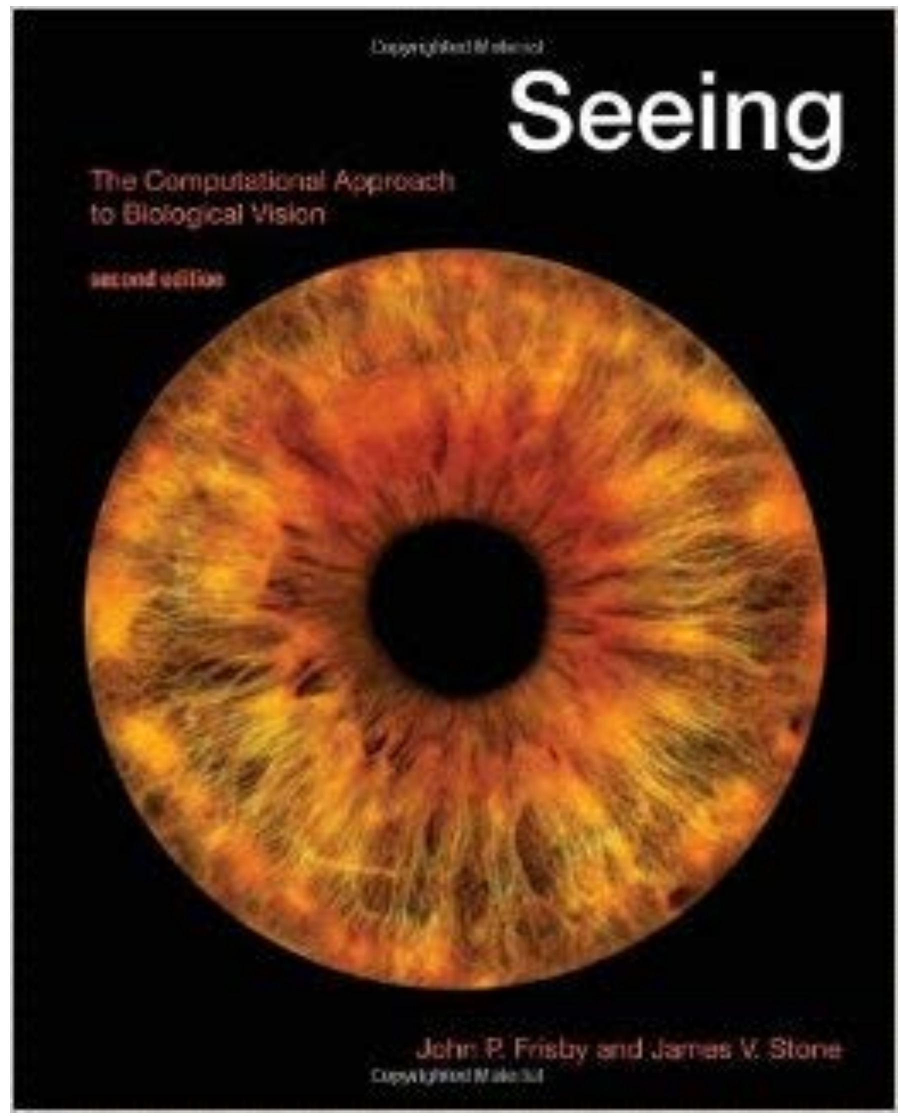
ecological vision

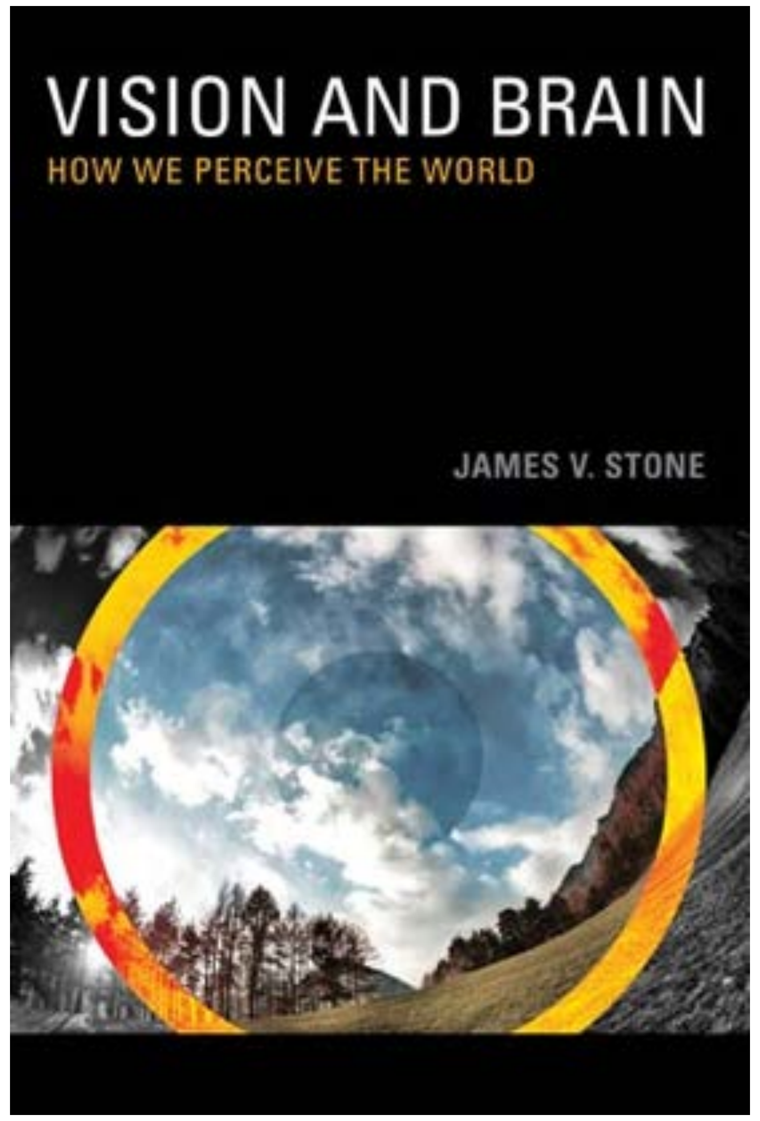
Ask not what's inside your head, but what your head's inside of.

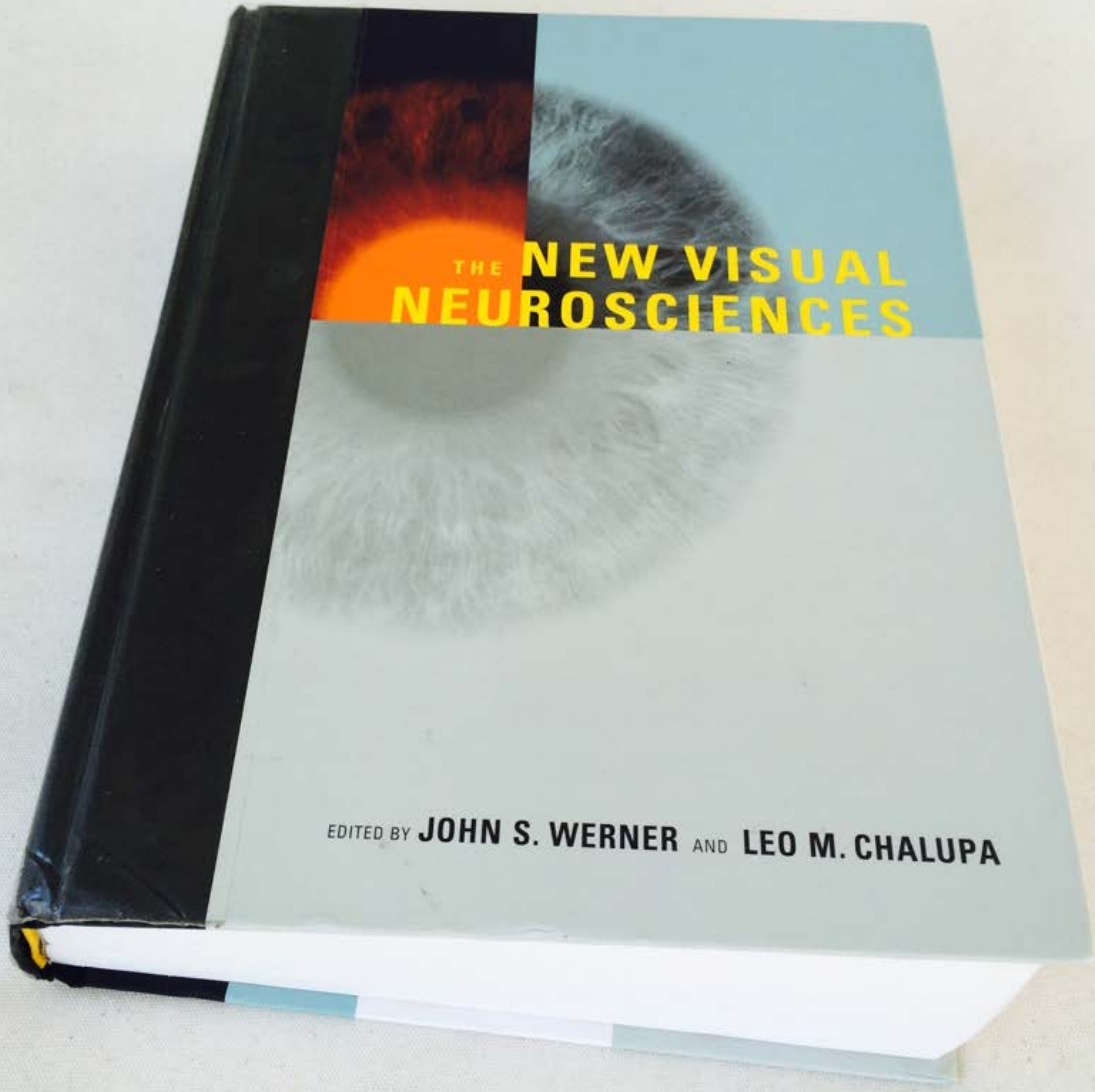


1982

- Computational theories
- Algorithms
- Circuitry





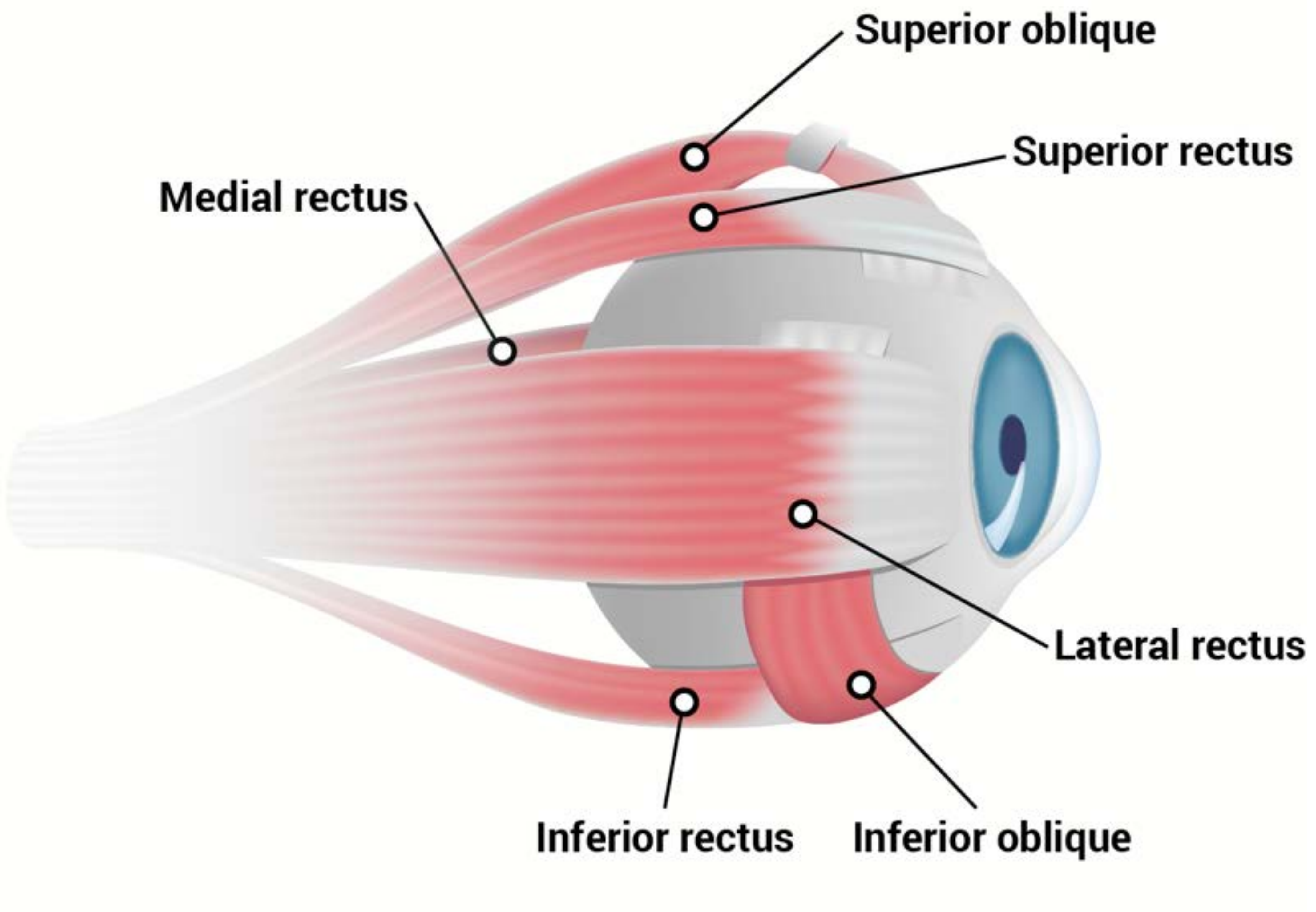


Human Brain Project

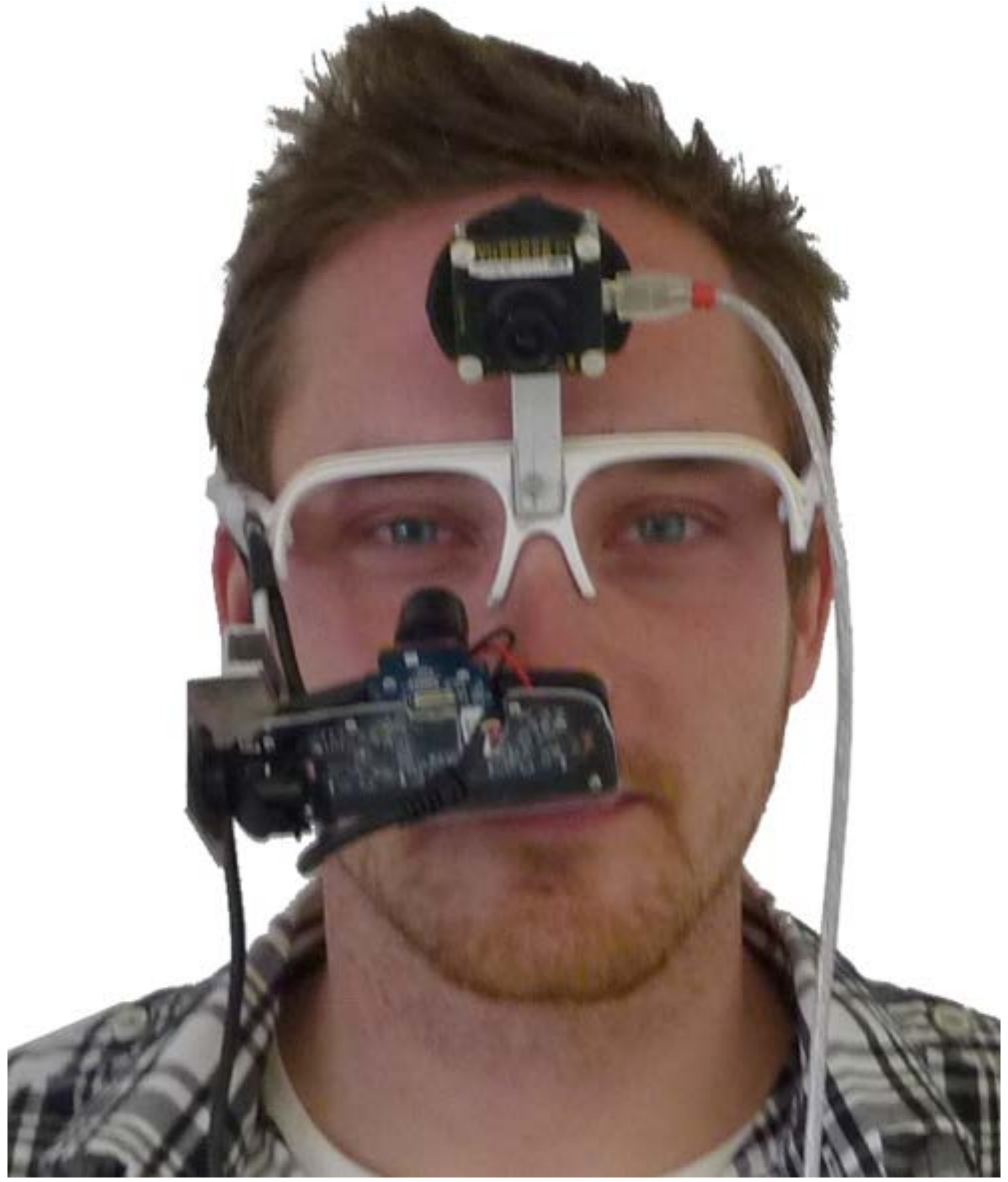
The BRAIN initiative

Visual neuroscience growing at a massive rate

Seeing is an active process



- Fastest moving muscles in the human body
  - ↪ rotational speed: 600deg/s
  - ↪ rotational acceleration: 35,000 deg/s^2




Ilya Efimovich Repin - The Complete Works

What are the material circumstances of the family?

Figure 109 “Seven Records of eye movements by the same subject” 1967
Yarbus, A. L.

Artwork: *Unexpected visitors*
Ilya Repin | 1884-1888

What age are the figures in the painting?

Figure 109 "Seven Records of eye movements by the same subject" 1967
Yarbus, A. L.

Artwork: *Unexpected visitors*
Ilya Repin | 1884-1888

What type of clothes are the family wearing?

Figure 109 “Seven Records of eye movements by the same subject” 1967
Yarbus, A. L.

Artwork: *Unexpected visitors*
Ilya Repin | 1884-1888

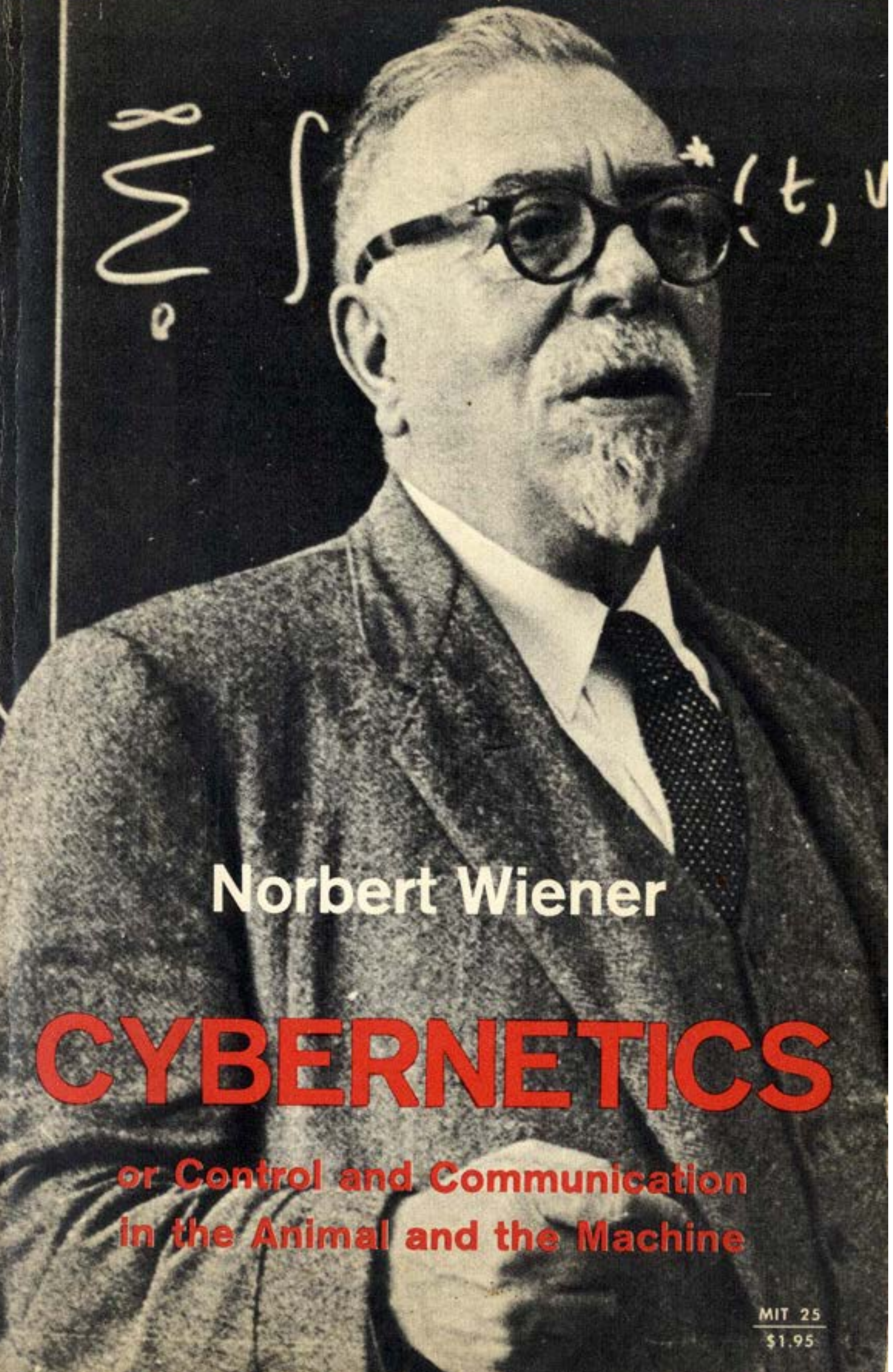
Vision is the process of **discovering** from images what is present in the world and where it is.

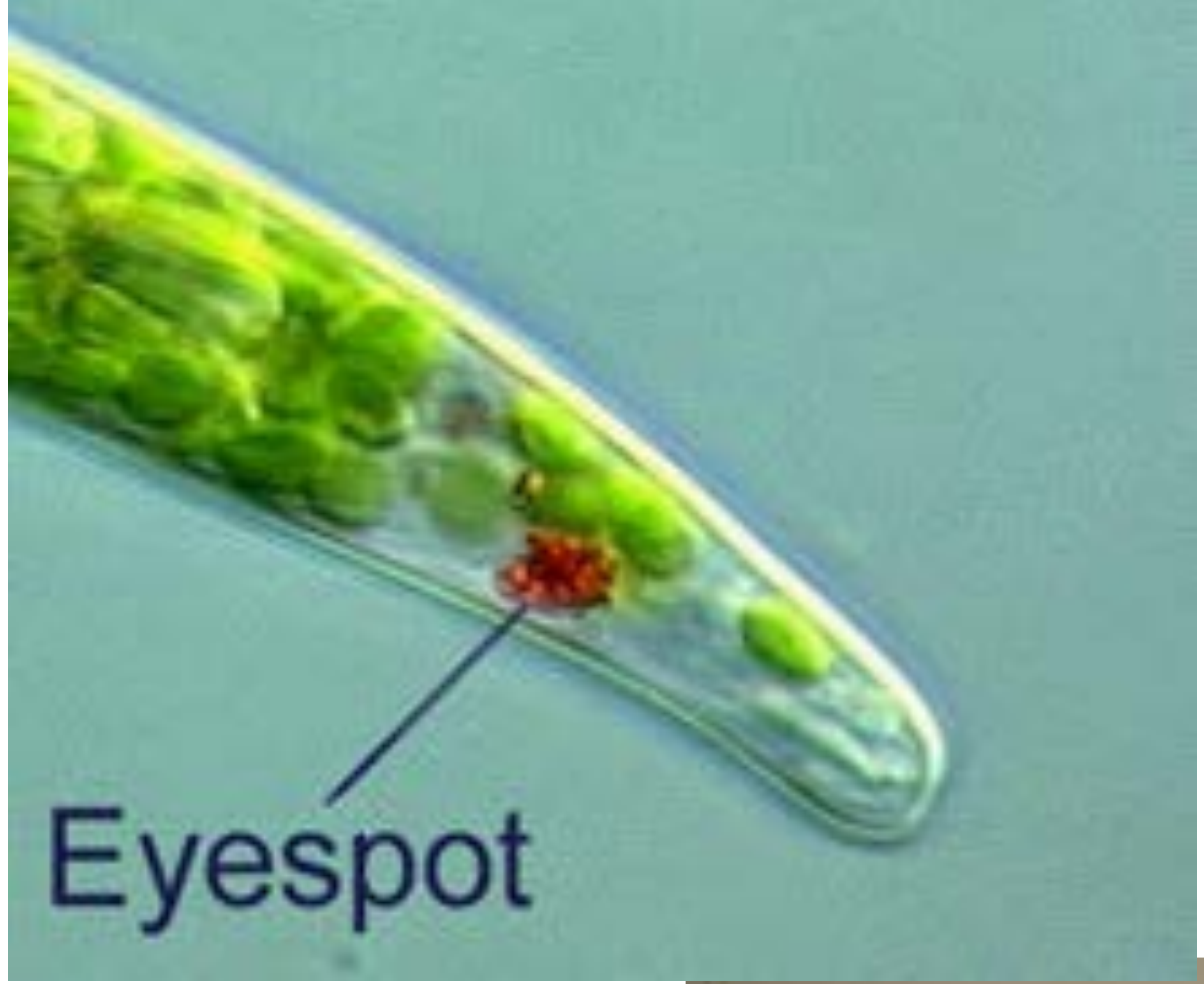
David Marr (1982)

Active vision/perception

- Active Vision, Yiannis Alomoinos etal., IJCV, 1988
- Active Perception, Ruzena Bajcsy, Proc IEEE, 76(8) August 1988

robotics and **vision**





William Grey Walter’s cybernetic tortoise

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

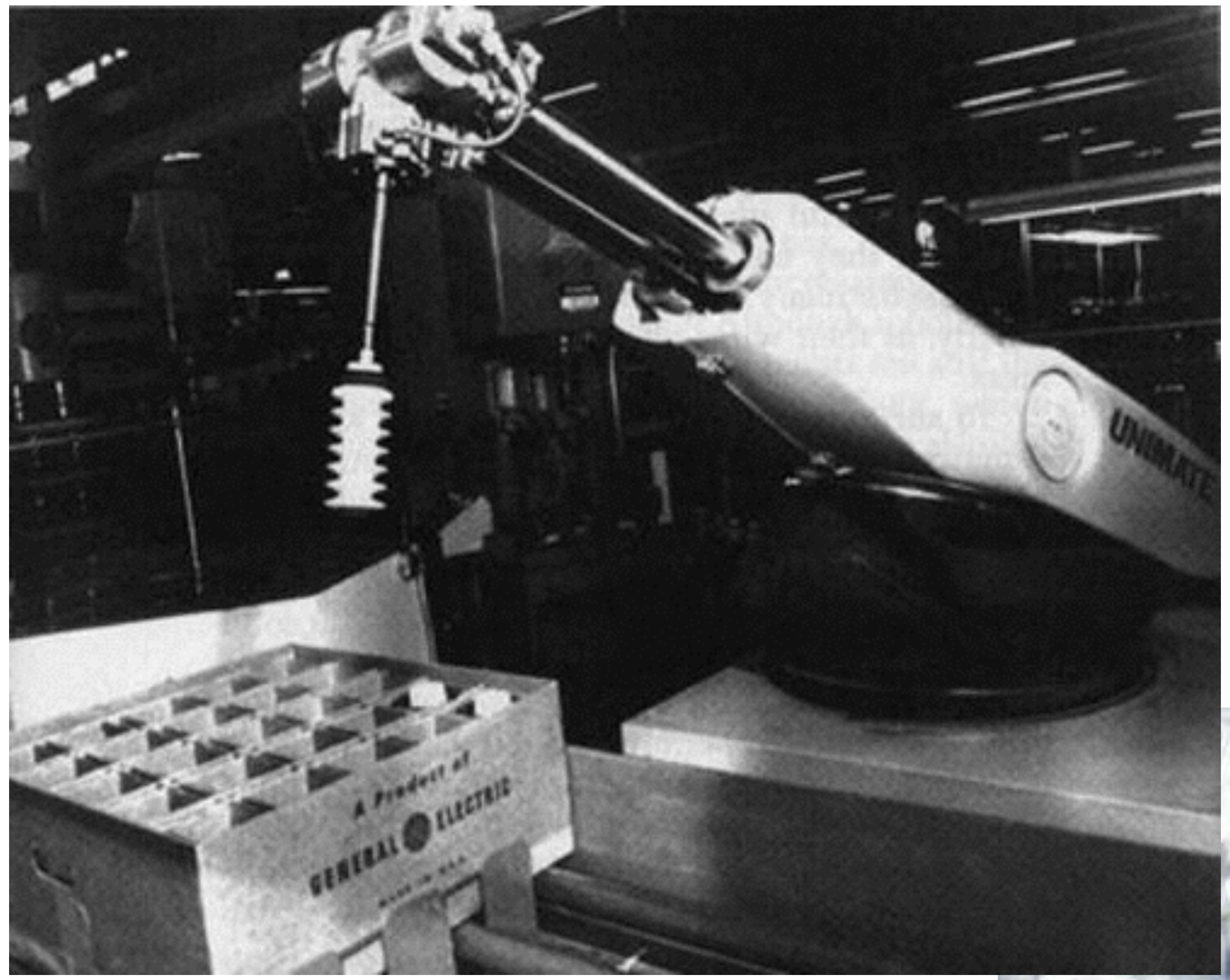
August 31, 1955

*John McCarthy, Marvin L. Minsky,
Nathaniel Rochester,
and Claude E. Shannon*

The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it…

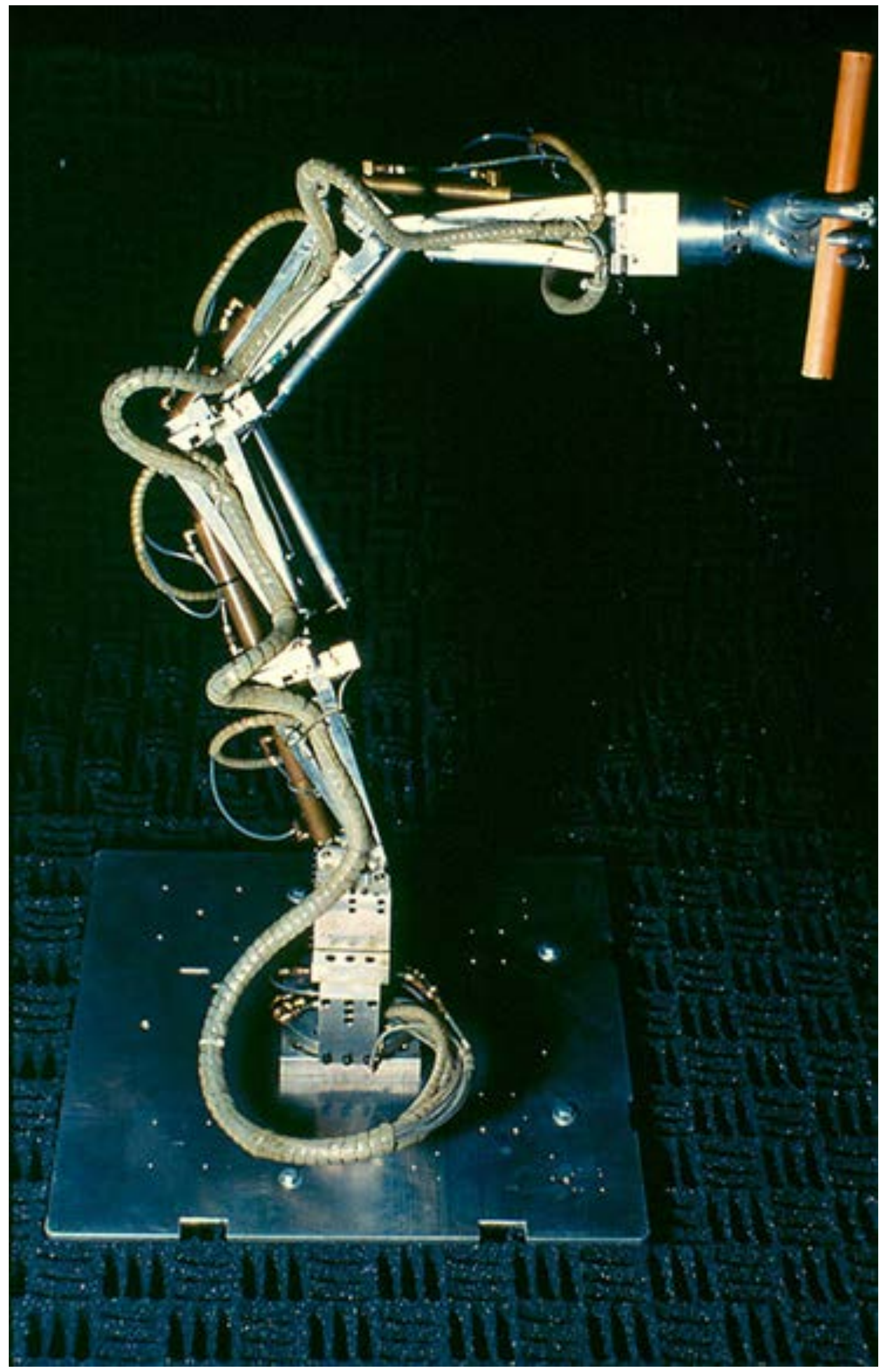
- *computers,*
- *natural language processing,*
- *neural networks,*
- *theory of computation,*
- *abstraction and*
- *creativity*

AI Magazine Volume 27 Number 4 (2006) (© AAAI)

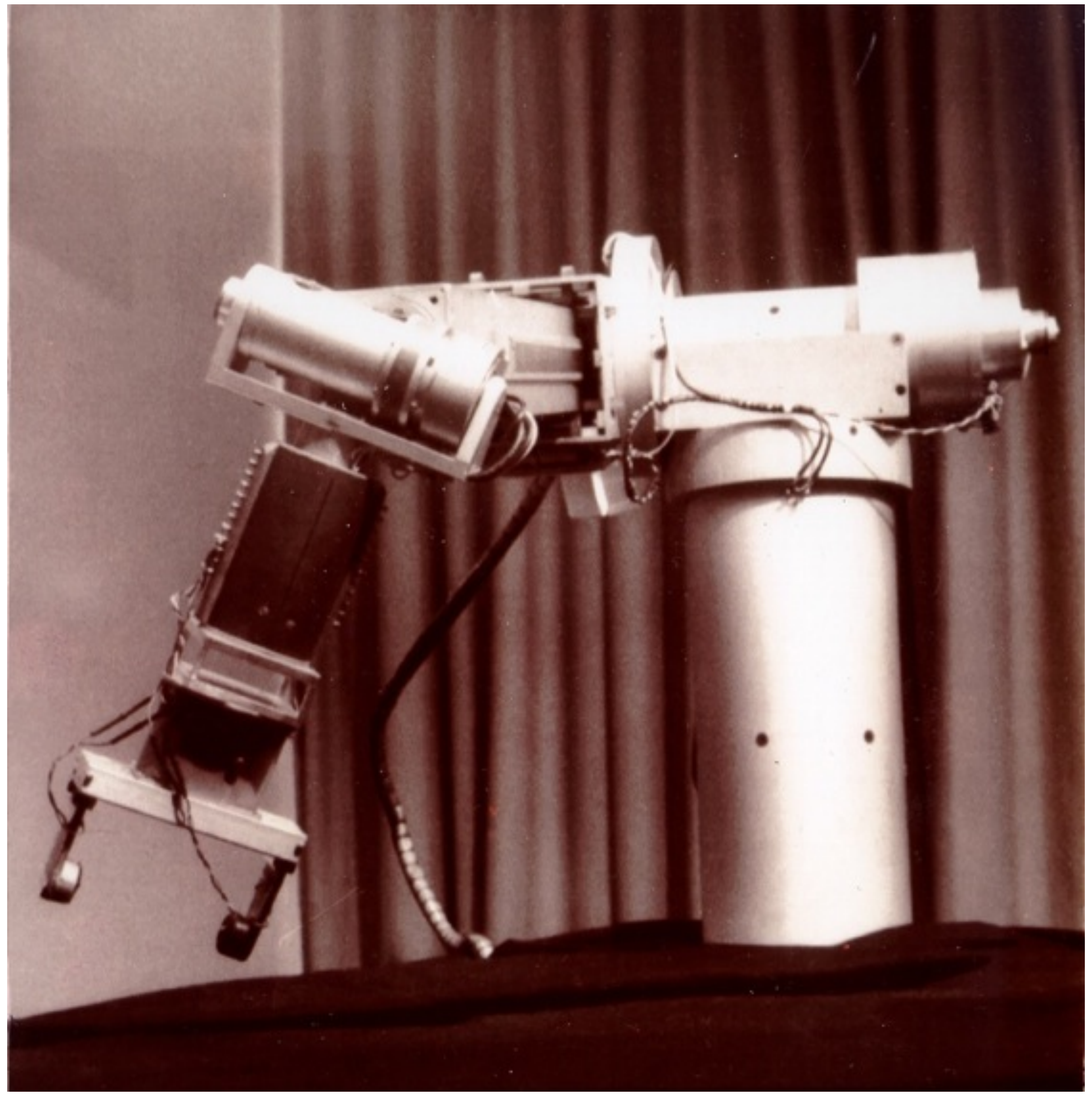


Unimate
Devol
Unimation Inc. 1961

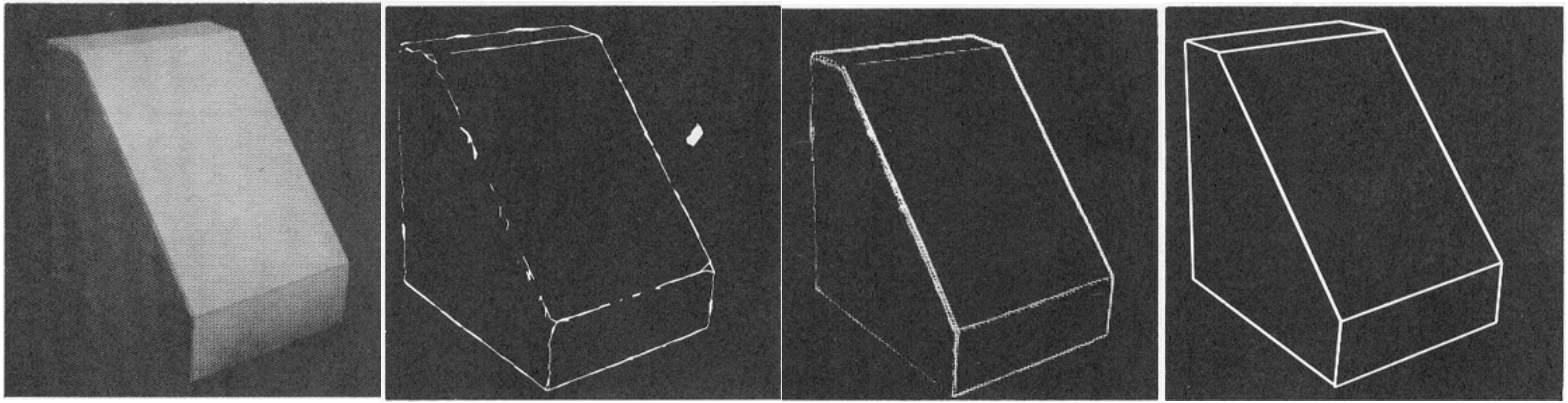
ORM
Scheinman & Leiffer
Stanford 1965



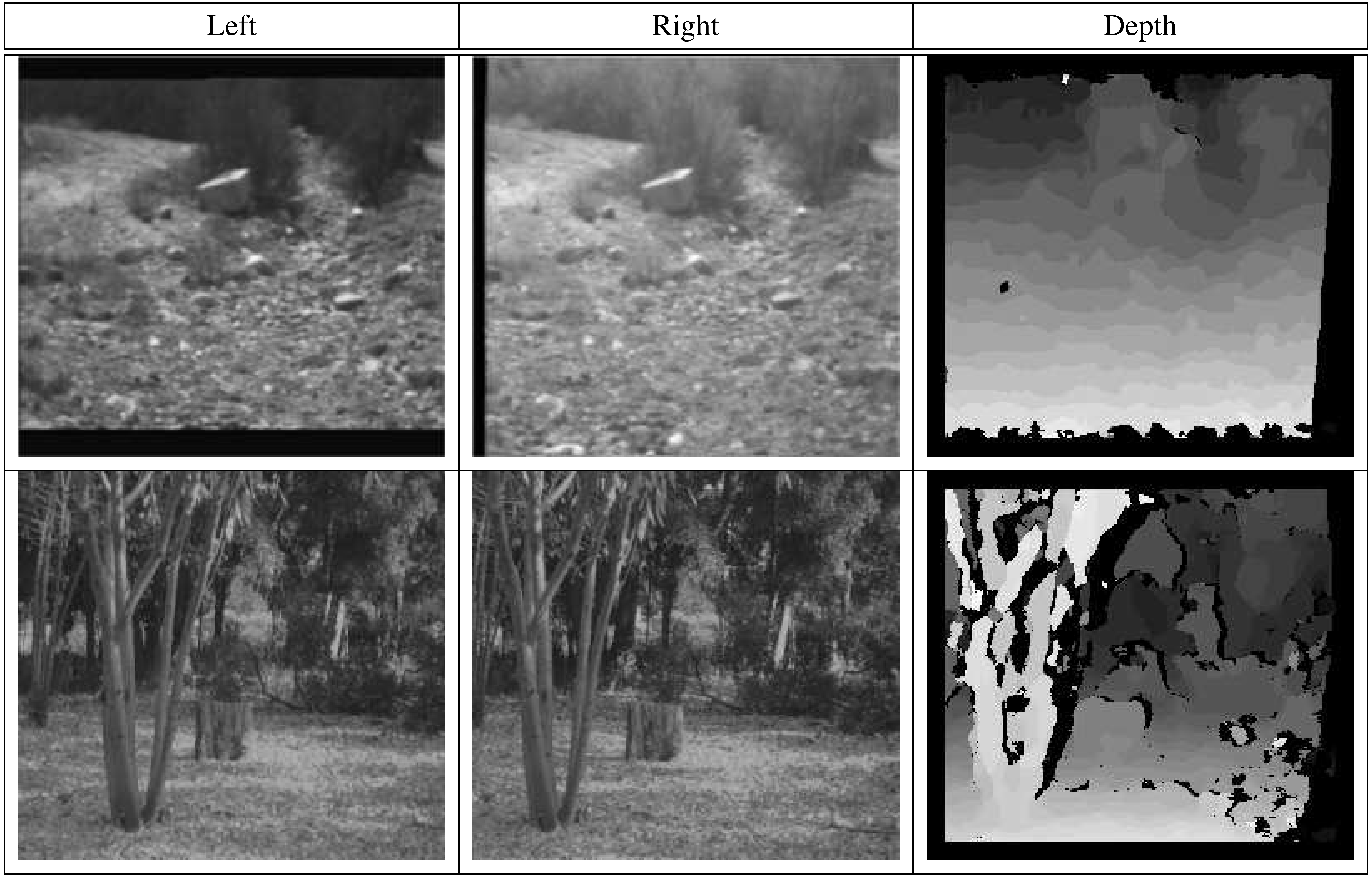
Tentacle arm
Minsk
MIT 1968



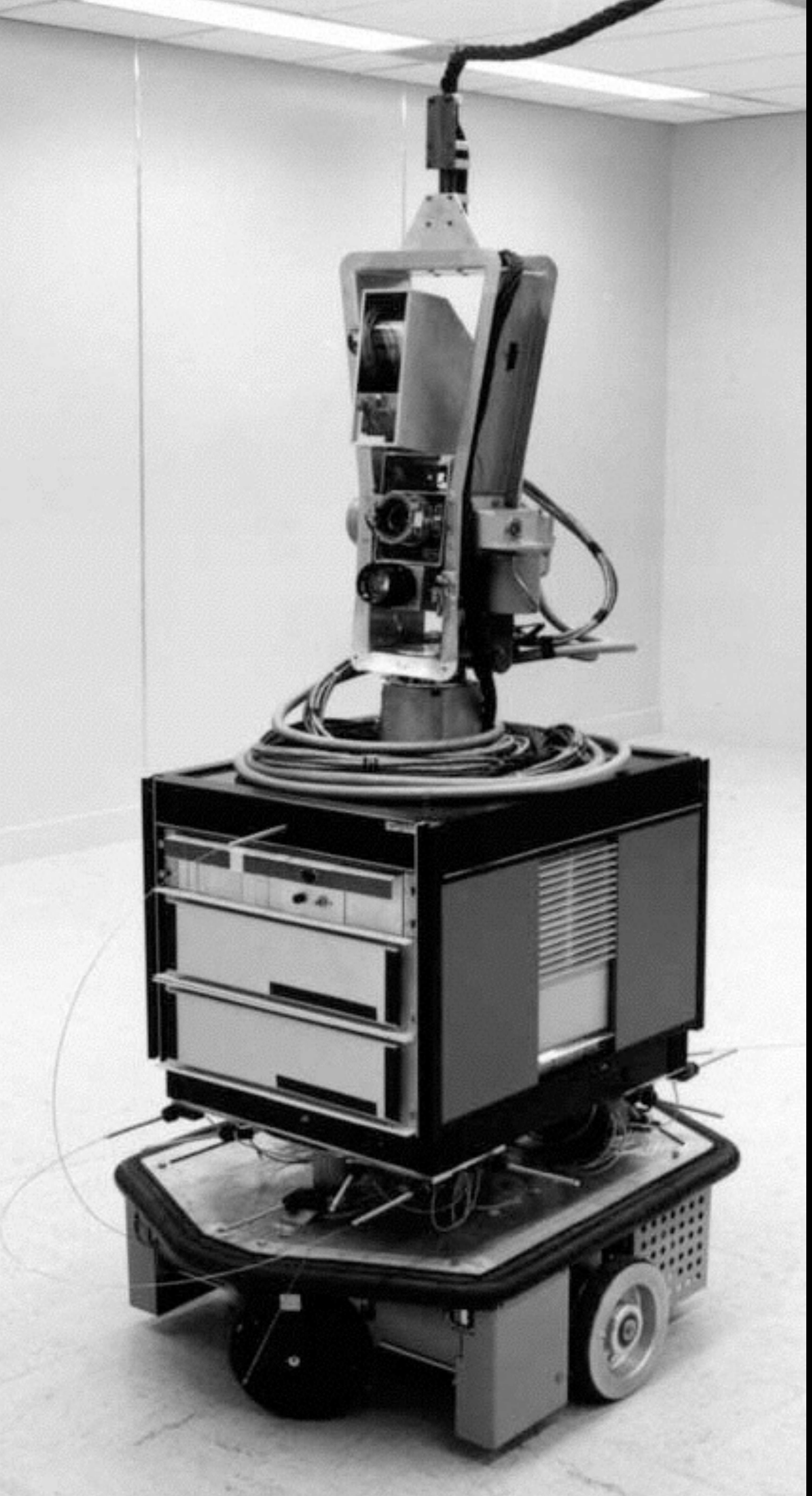
Stanford arm
Scheinman
Stanford 1968



Machine perception of three-dimensional solids
Larry Roberts
MIT 1964



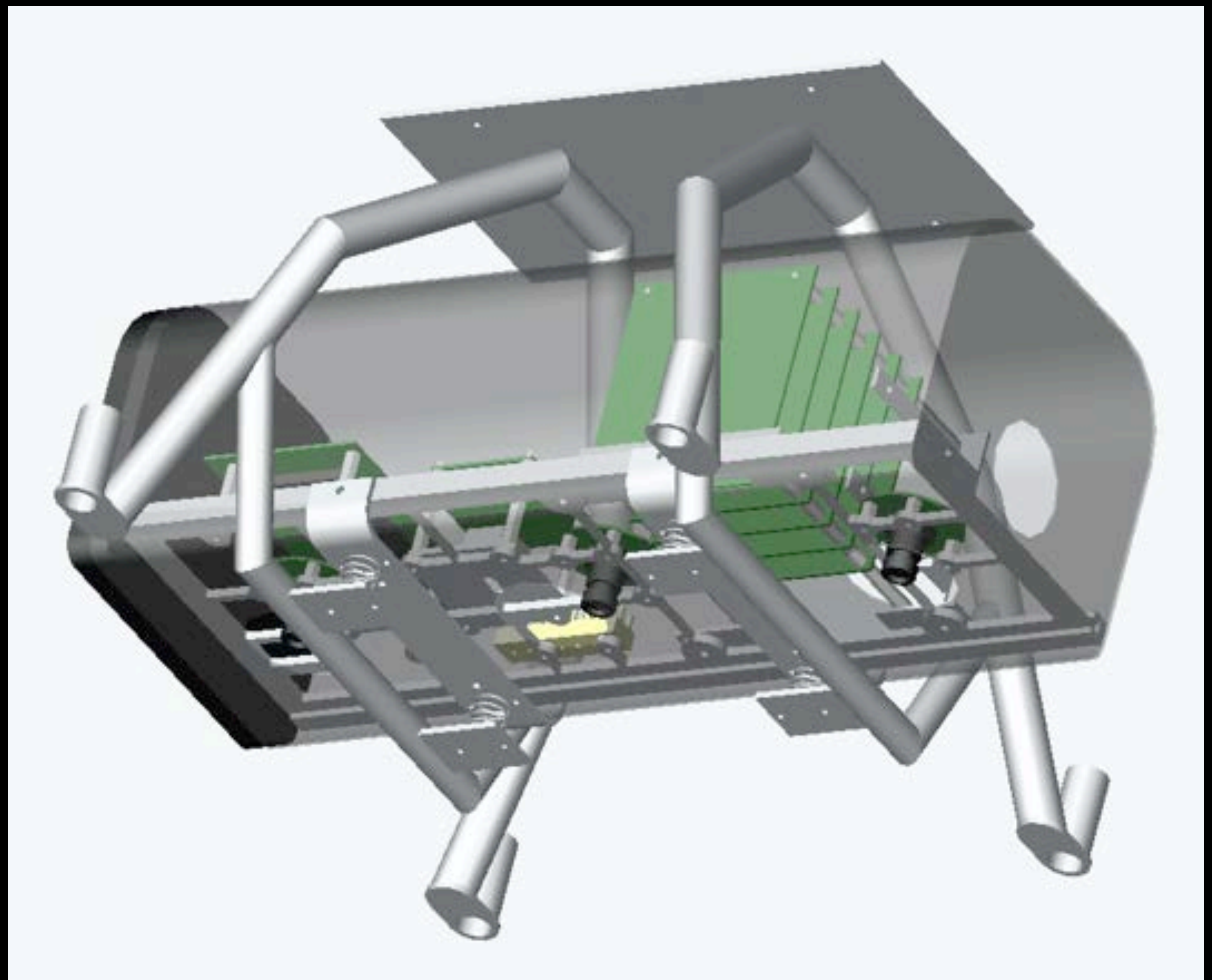
JISCT data set and early stereo results
1993



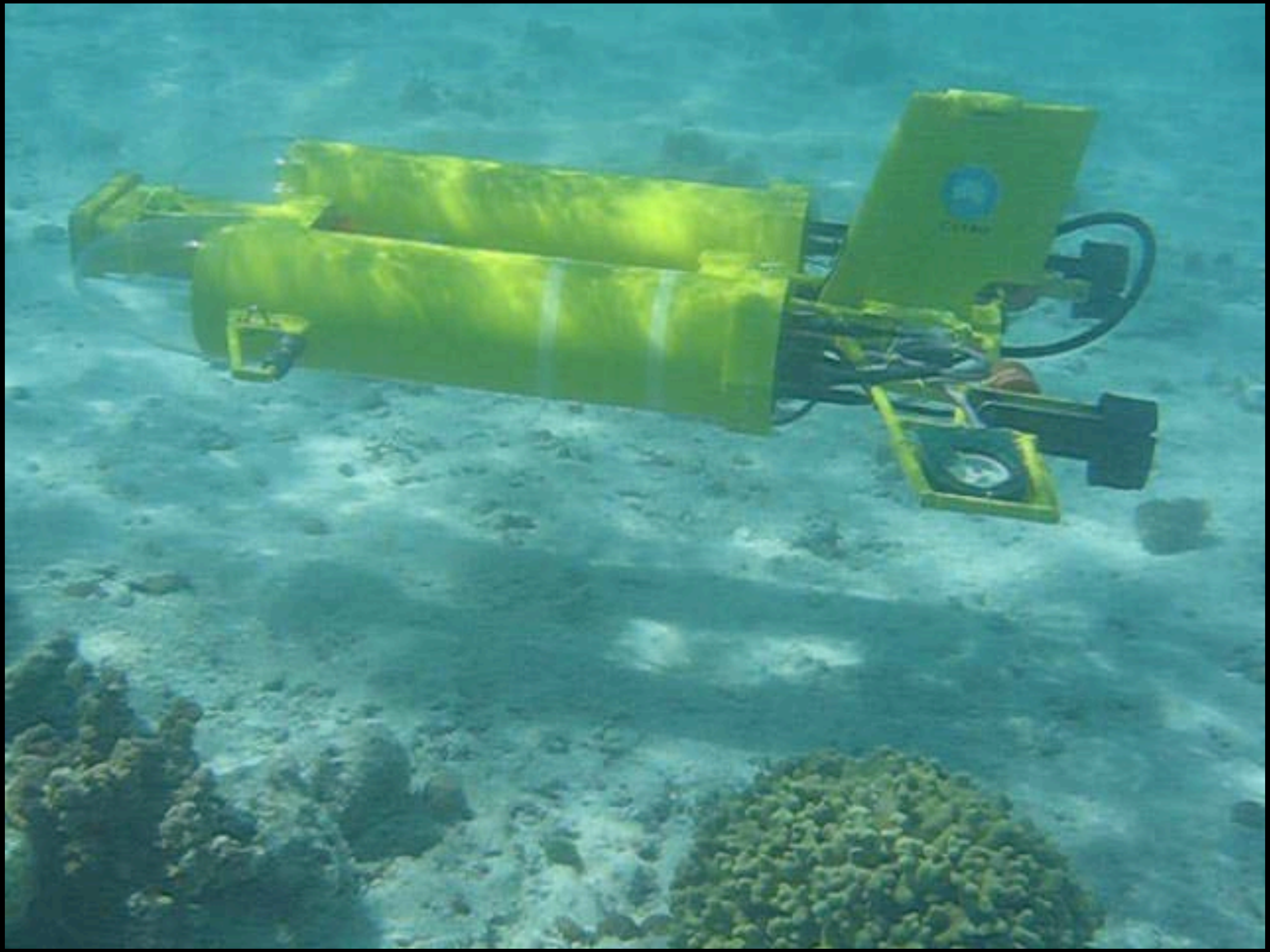
Shakey
SRI
1966-72

Stanford Cart
Moravec
1971-80

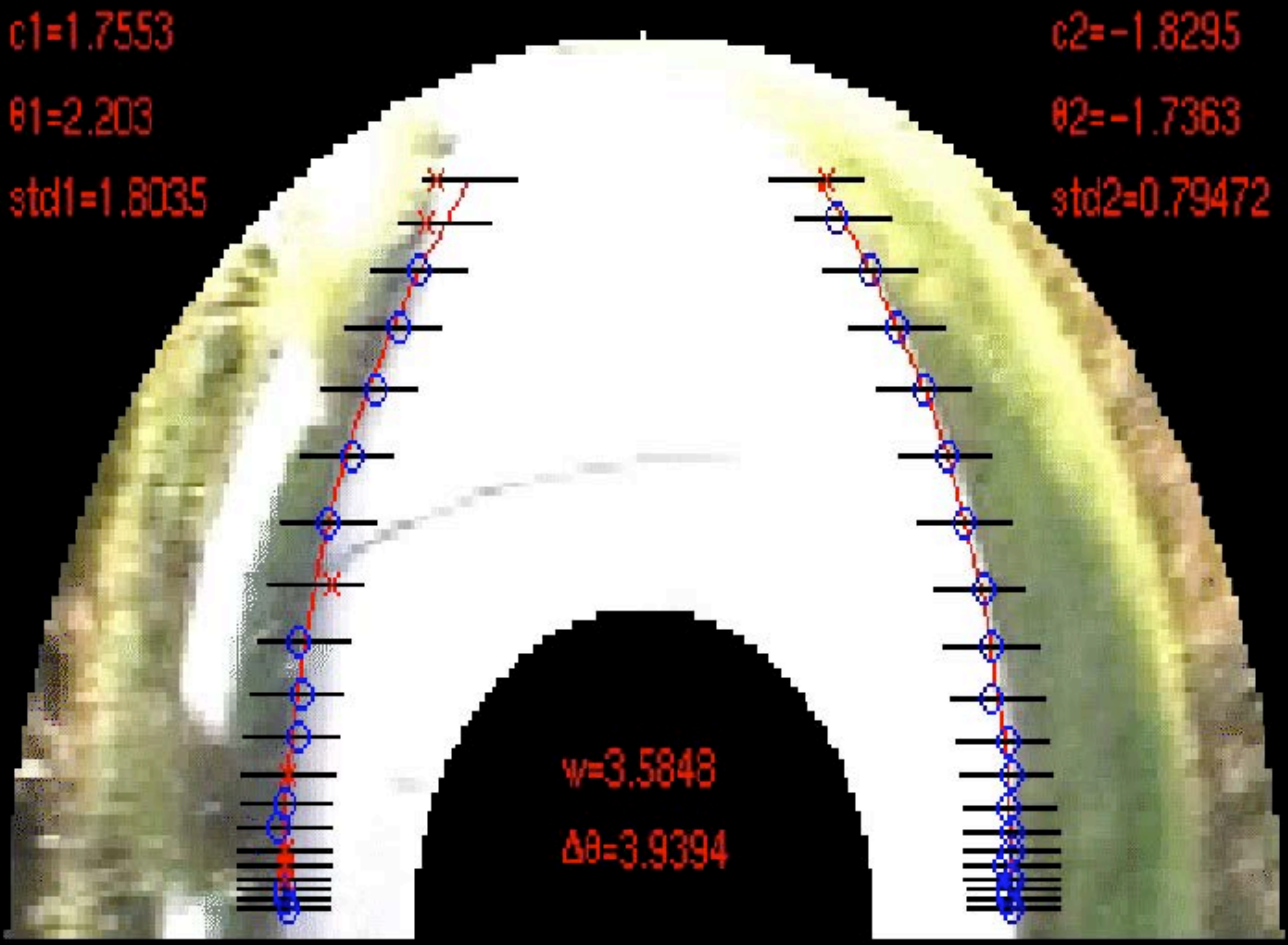
Corke
GRASP lab + CSIRO
1988-92



Buskey, Roberts, Corke
CSIRO
2004



Corke and Dunbabin
CSIRO
2002



Corke
CSIRO
2002

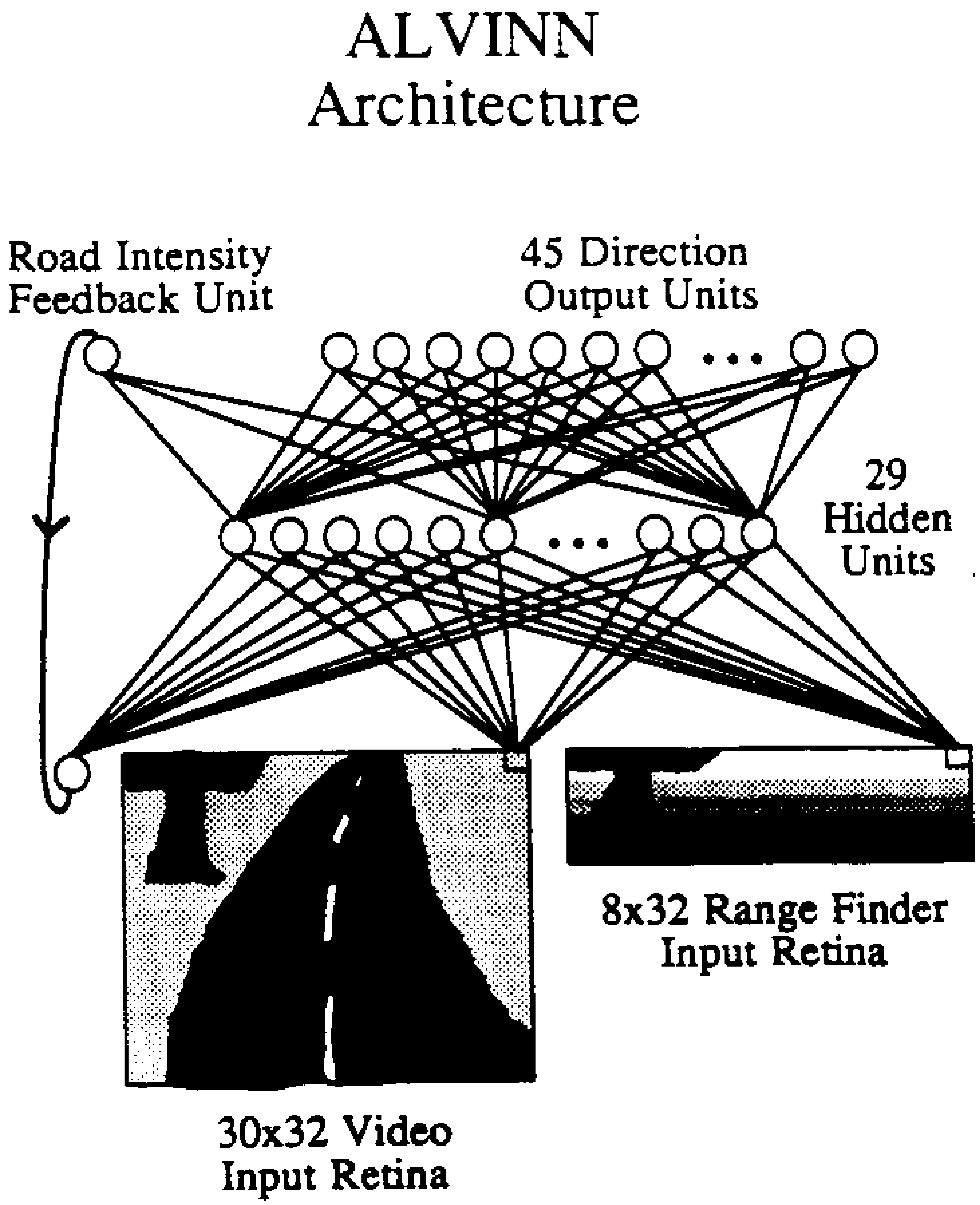
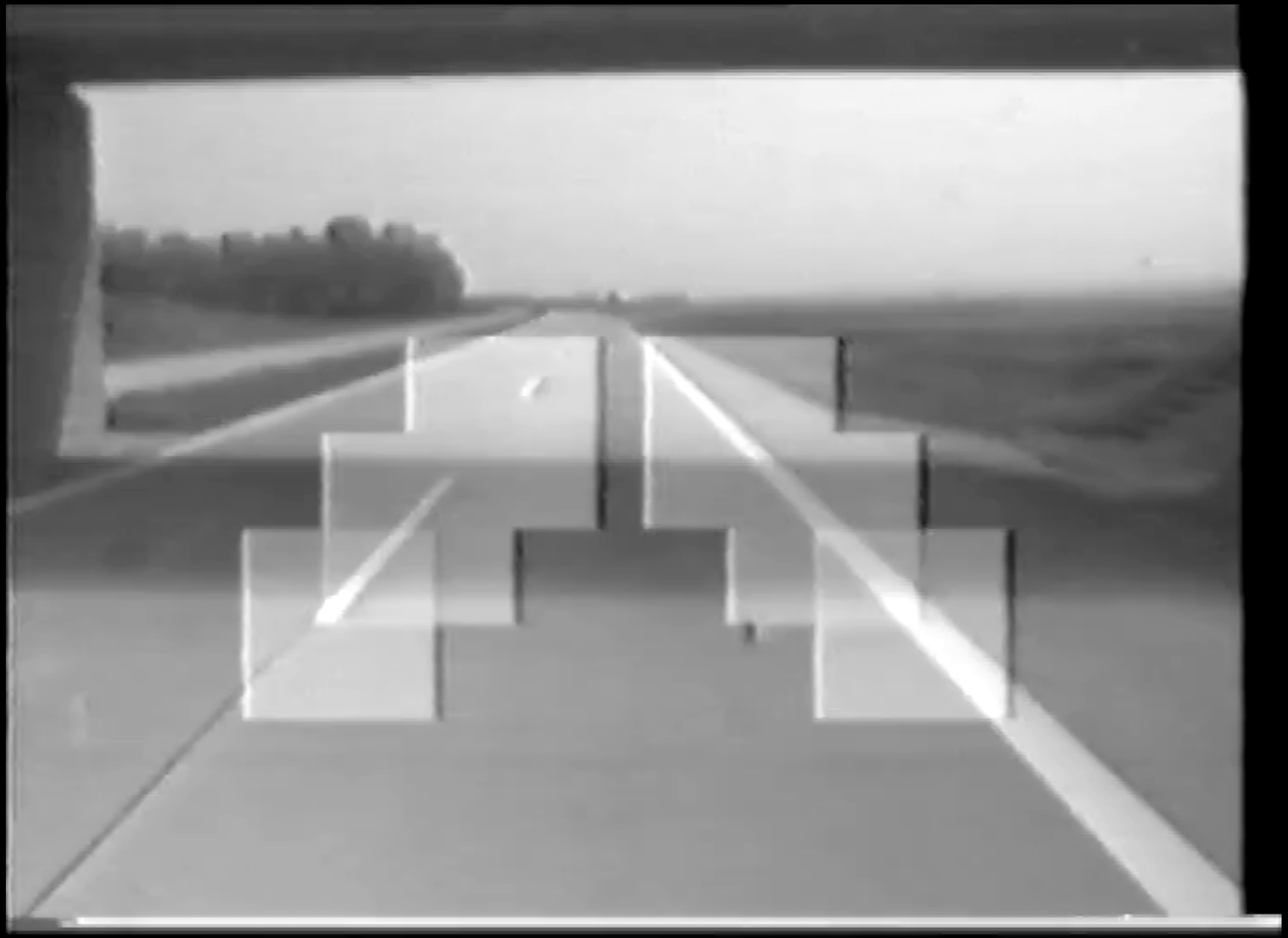


Figure 1: ALVINN Architecture

1989

3 x Sun computers
Warp systolic array computer (100 Mflop)

Ernst Dickmanns
Bundeswehr Universität München
1980-2004



movie courtesy of Joe Wuesche
Bundeswehr Universität München

PROgraMme for a European Traffic of Highest Efficiency and Unprecedented Safety

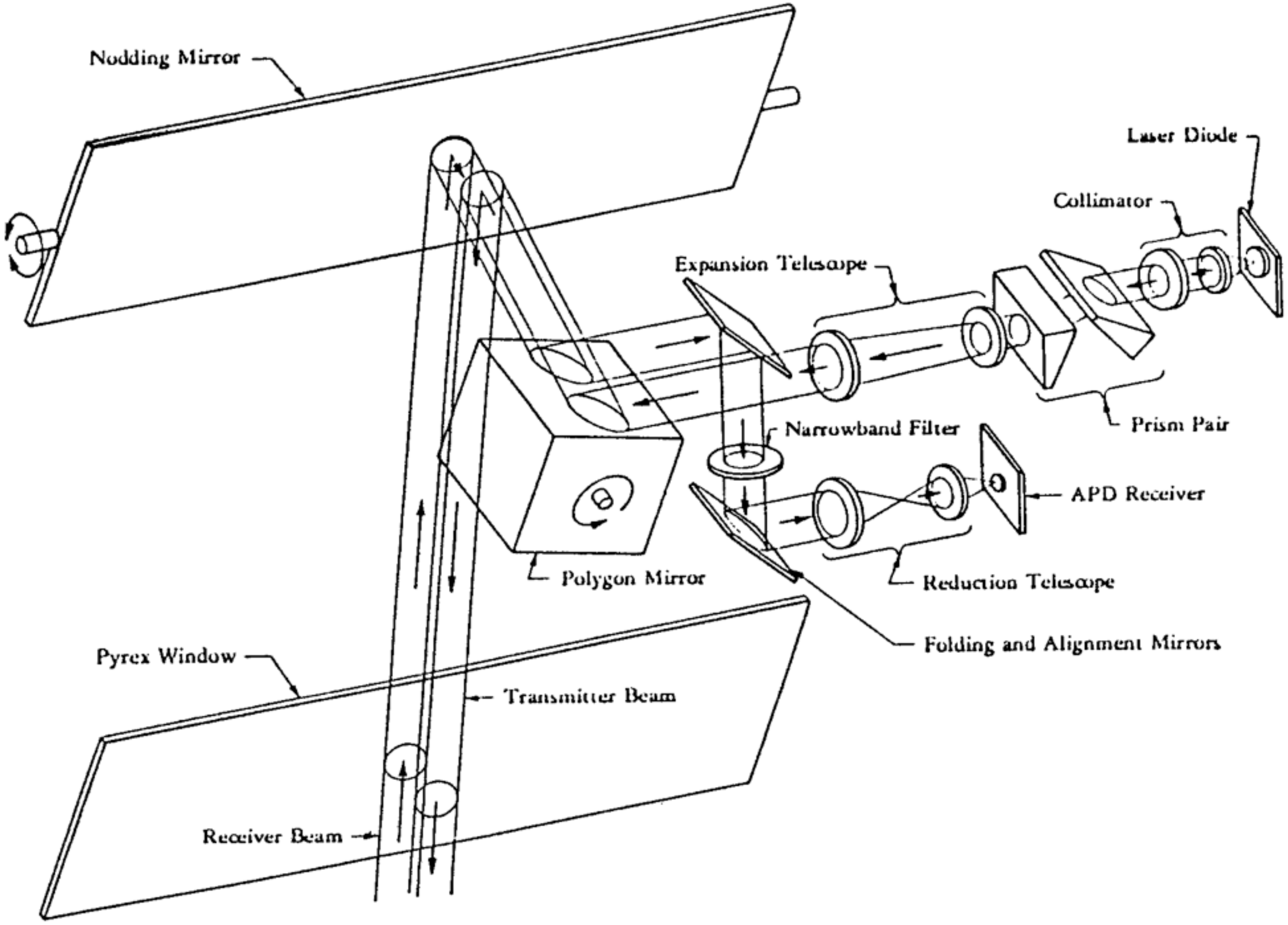
1987-95

Munich to Copenhagen return
Upto 175km/h
Overtaking
Distance between human interventions:
- Avg 9km
- Max 158km

the split

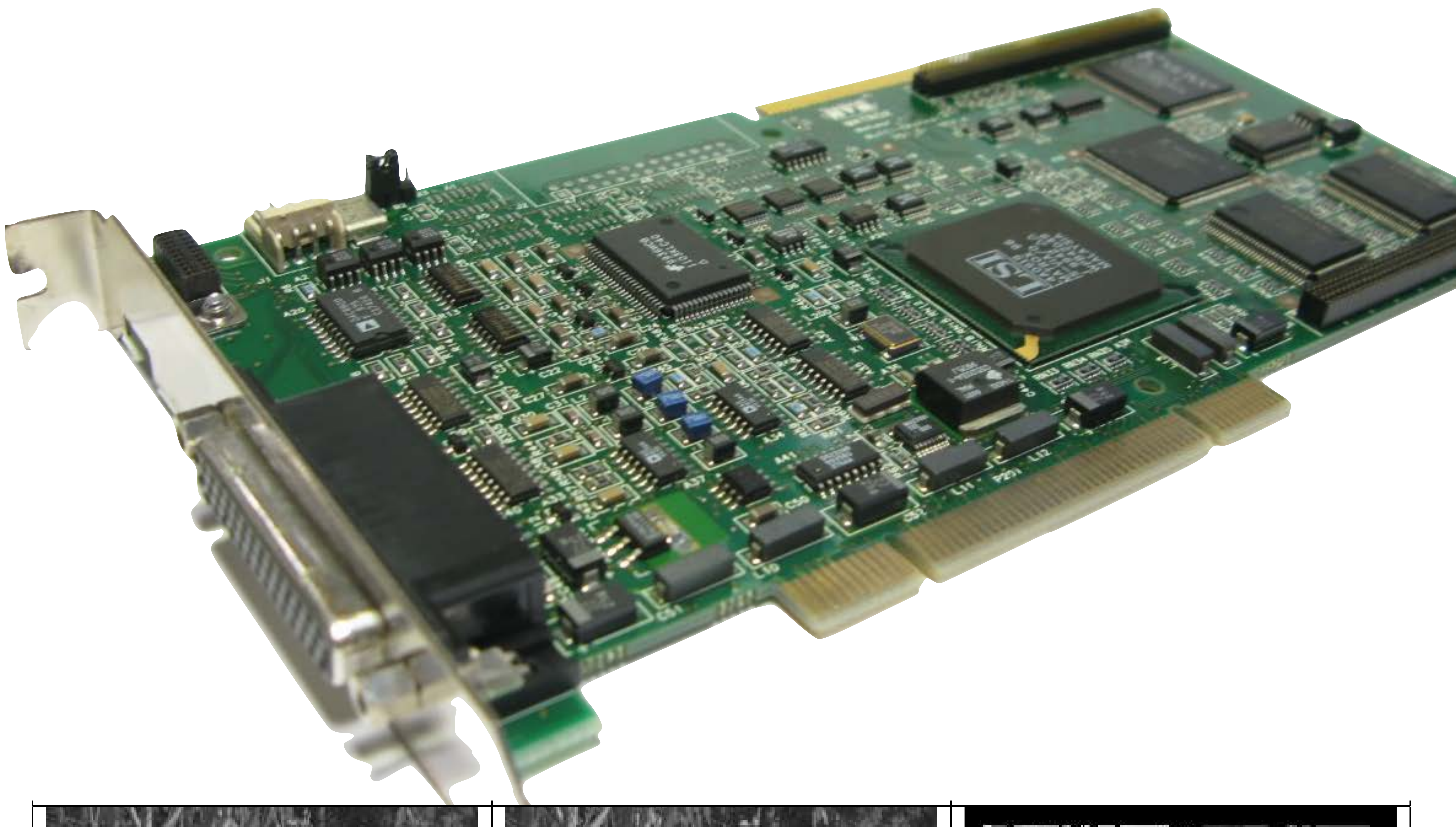
early 1990s

cost < $10,000



CMU Navlab 1 + ERIM laser scanner
1980s


PULNiX
PULNiX
TM-6CN



- Improved sensors appeared over time
- Increase in resolution, rotation rate, reflectance data quality, number of beams
- But maximum range and minimum cost little changed

Jericho, Oxford, August 2014. Mobile Robotics Group, University of Oxford

RGB-D
cameras

but it's only geometry

we're missing color, texture, object recogntion, face recognition, human actions, human intent etc.

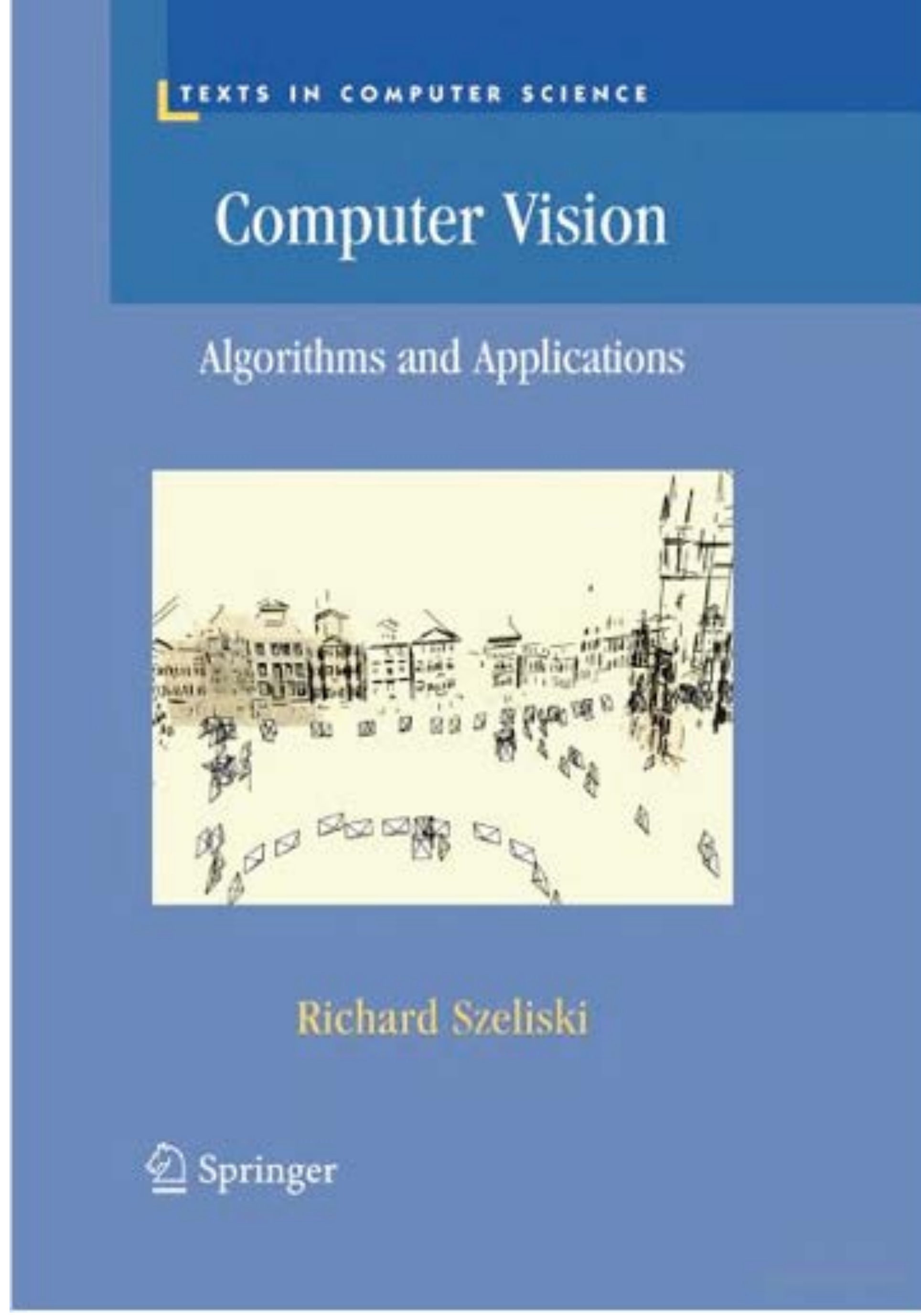
what have the vision people been doing?

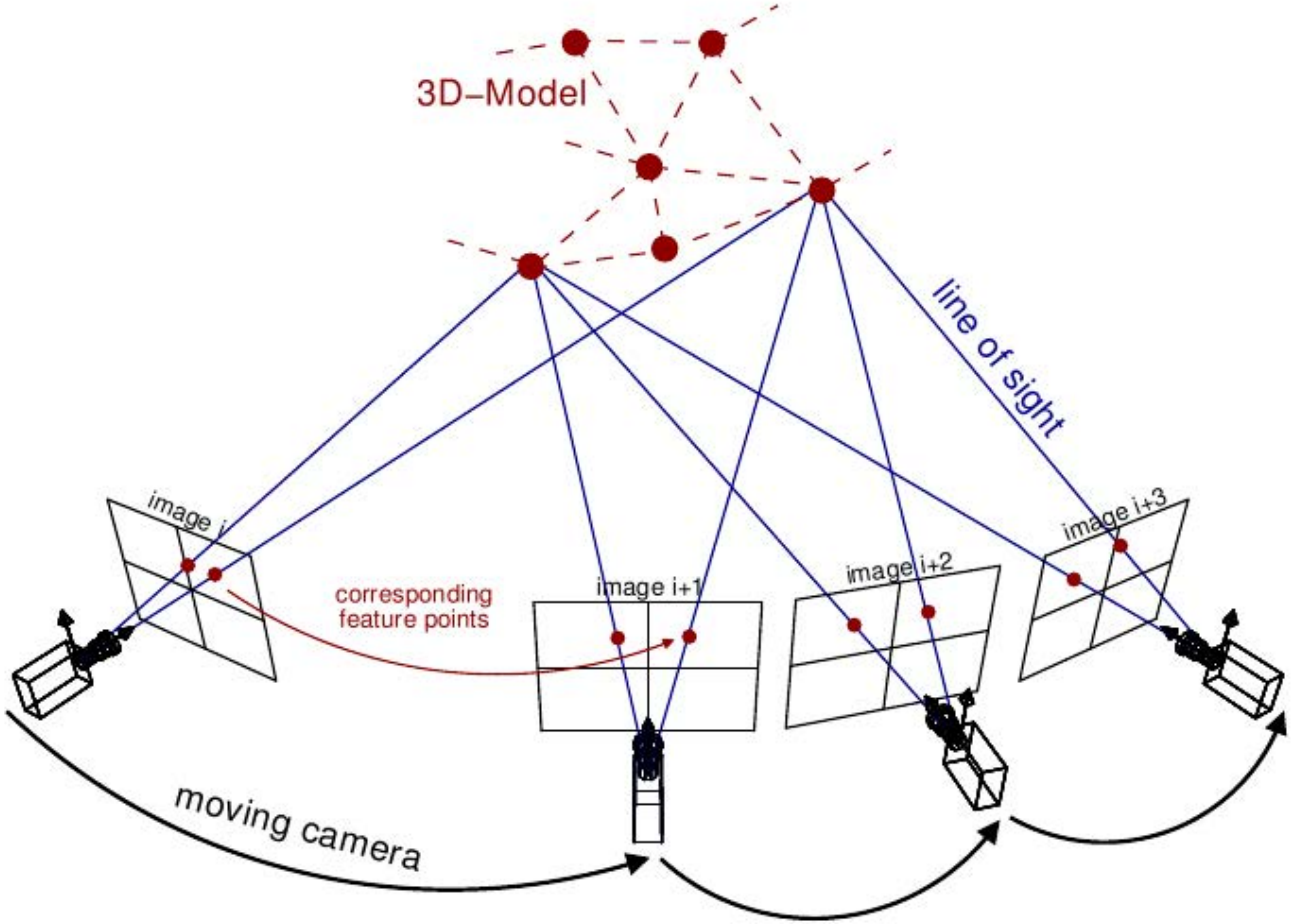
What makes Paris looks like Paris?

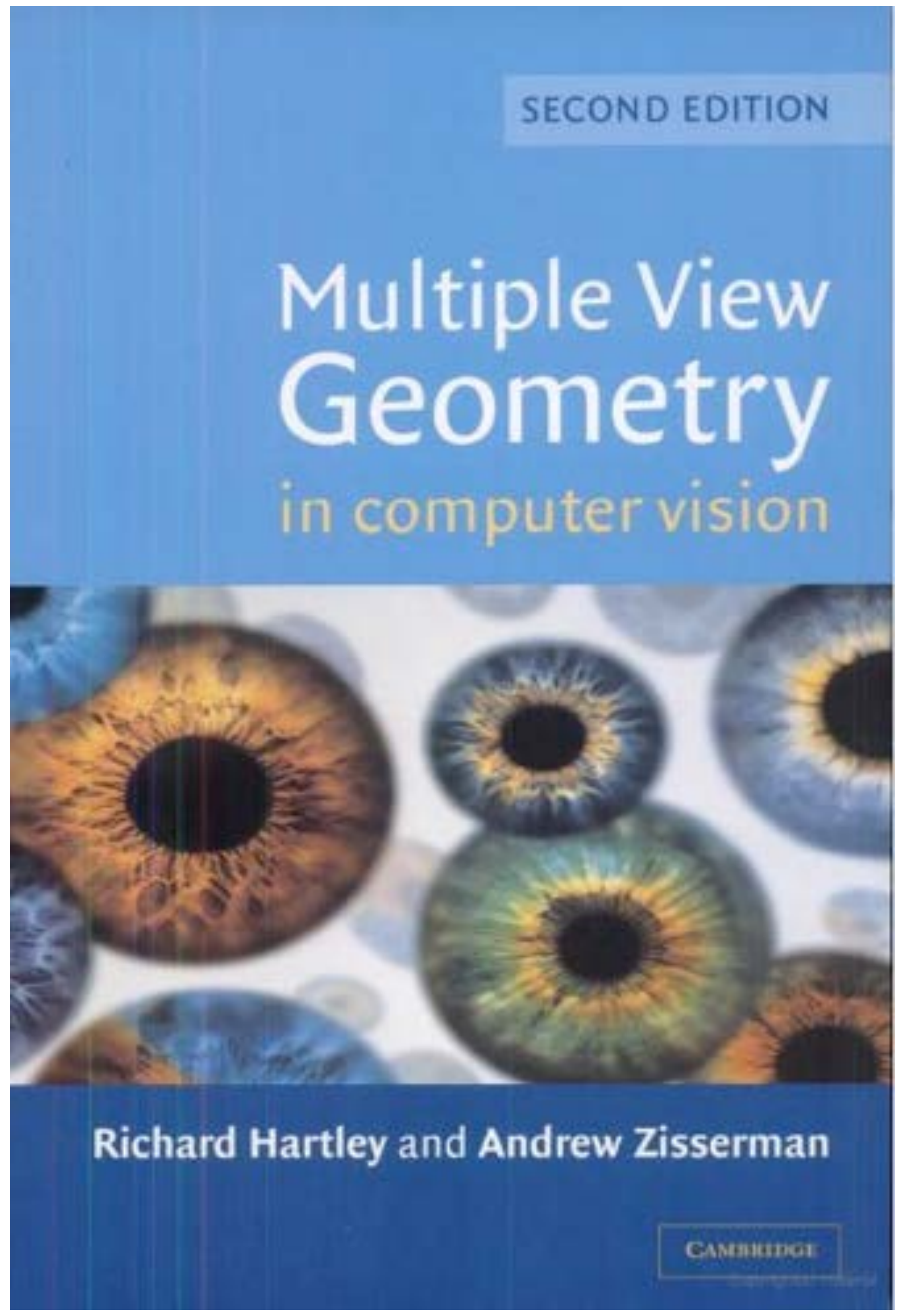
Built Rome in a day

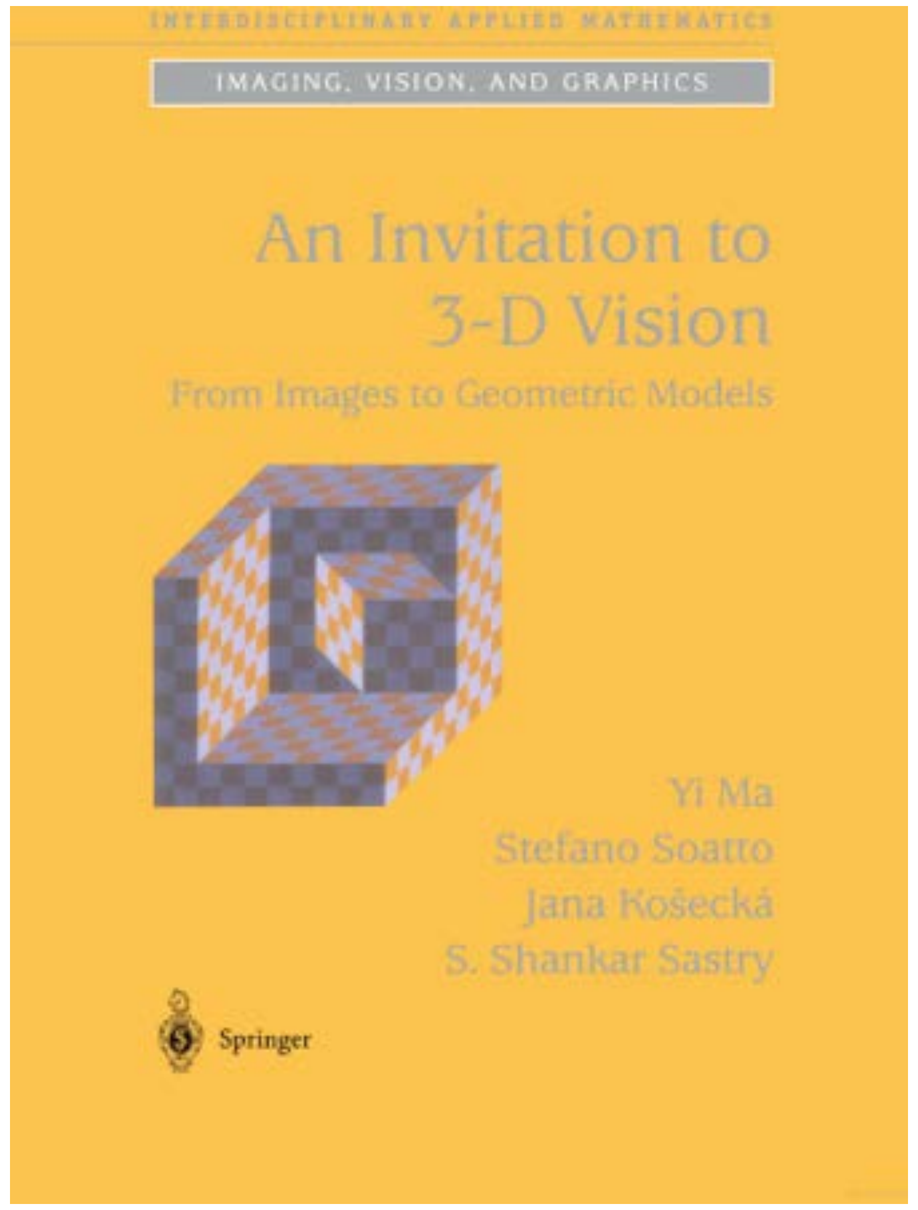
- Multi-view geometry
  - Structure from motion
  - Stereo
  - Visual odometry
  - Visual SLAM
- Stitching
- Super resolution
- Pose estimation
- Segmentation
- Image retrieval

- Object recognition
- Face recognition
- Action recognition
- Text recognition
- Pedestrian detection
- Calibration
- Feature detectors & descriptors: SIFT, SURF, FAST, BRIEF, BRISK, MSER, FREAK, HOG, CenSureE

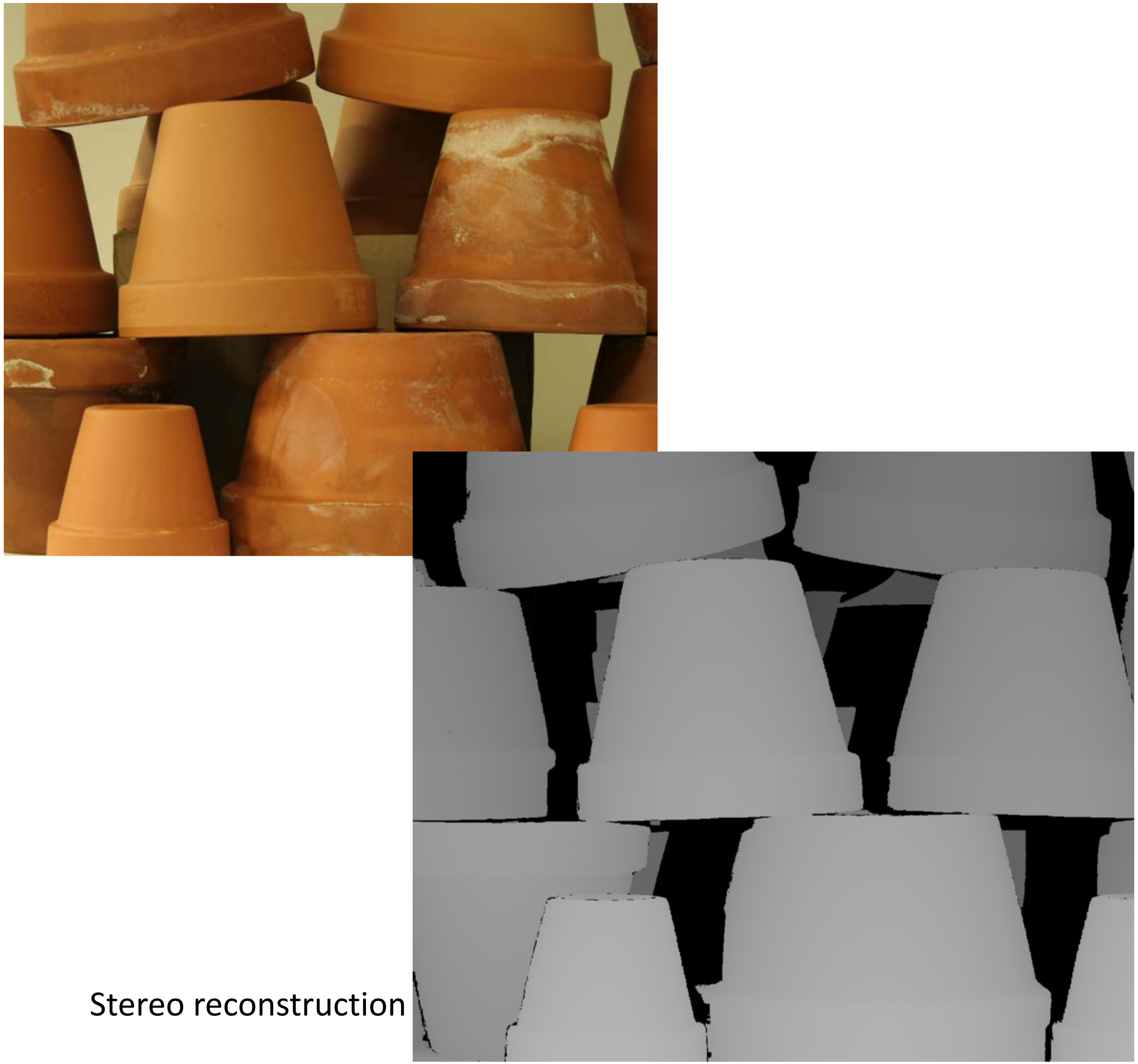



3D–Model
line of sight
image i
image i+1
image i+2
image i+3
corresponding
feature points
moving camera


SECOND EDITION
Multiple View
Geometry
in computer vision
Richard Hartley and Andrew Zisserman
CAMBRIDGE


INTERDISCIPLINARY APPLIED MATHEMATICS
IMAGING, VISION, AND GRAPHICS
An Invitation to
3-D Vision
From Images to Geometric Models
Yi Ma
Stefano Soatto
Jana Košecká
S. Shankar Sastry
Springer

Middlebury stereo dataset



Stereo reconstruction

Monocular camera reconstruction of Scott Reef
He, McKinnon, Upcroft
QUT

image sequence courtesy of U. Sydney

Alessandro Bissacco, Mark Cum- mins, Yuval Netzer, and Hartmut Neven.

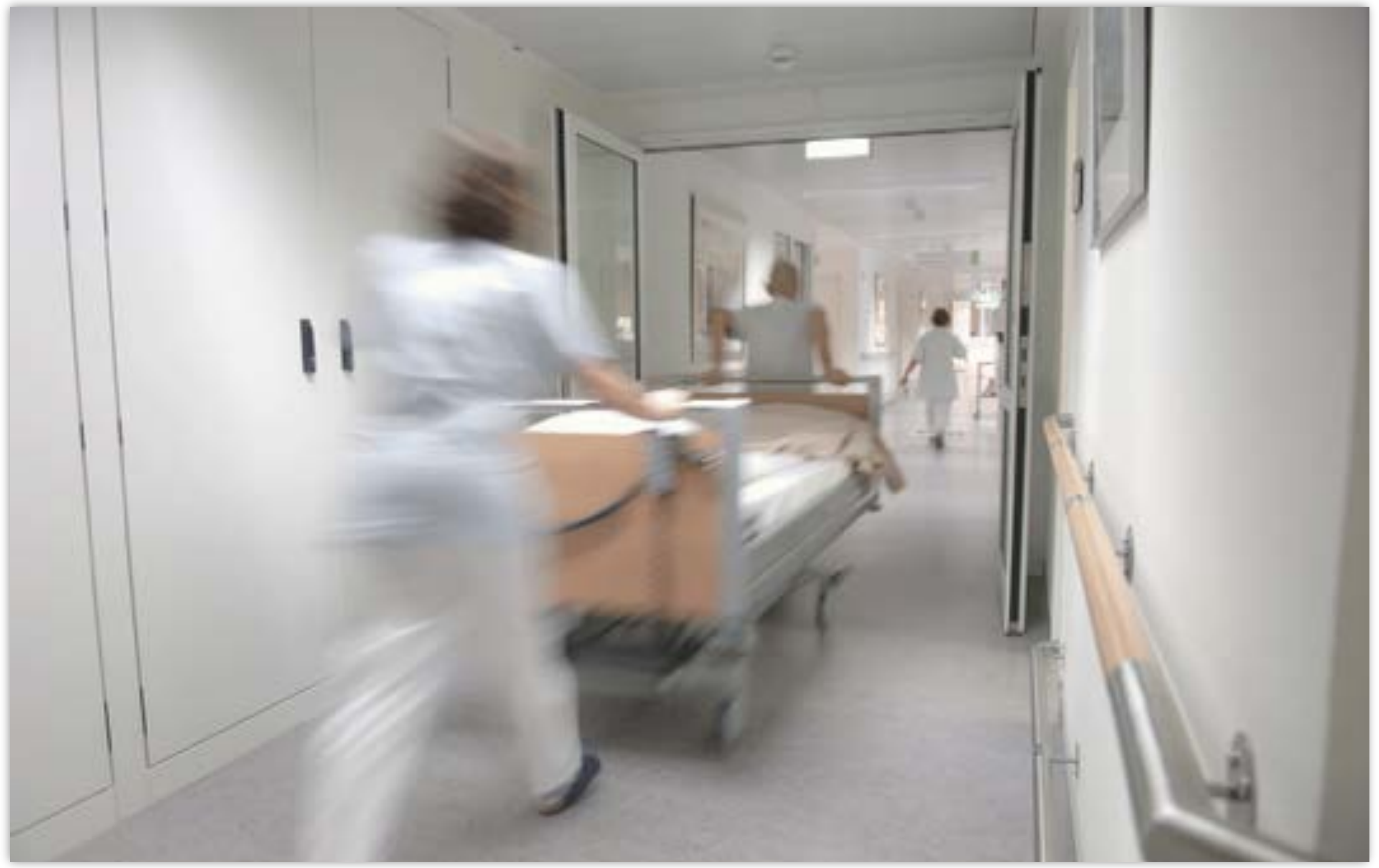
Photoocr: Reading text in uncontrolled conditions. In Computer Vision (ICCV), 2013 IEEE International Conference on, pages 785–792. IEEE, 2013.

lots of awesome (and useful) stuff roboticists need to be aware of

but the camera is generally passive

robots → computer vision

the future



where are all the robots?

Seeing

seeing creates memories
memory helps seeing

context for seeing
seeing for context

recognising objects & stuff

recognising places

see to move
move to see

detecting motion

paying attention

recognizing humans, their activities and intent

So why is it hard?

- Vision is a great sensor but
  - The rich visual information in encoded
    - 3D world is projected to 2D
  - To recover the “lost” information we need
    - assumptions (context, world knowledge)
    - to interact with the environment (move)
  - Many distractors
    - shadows, lighting change, seasons etc.
    - information is ambiguous

but we know it is possible

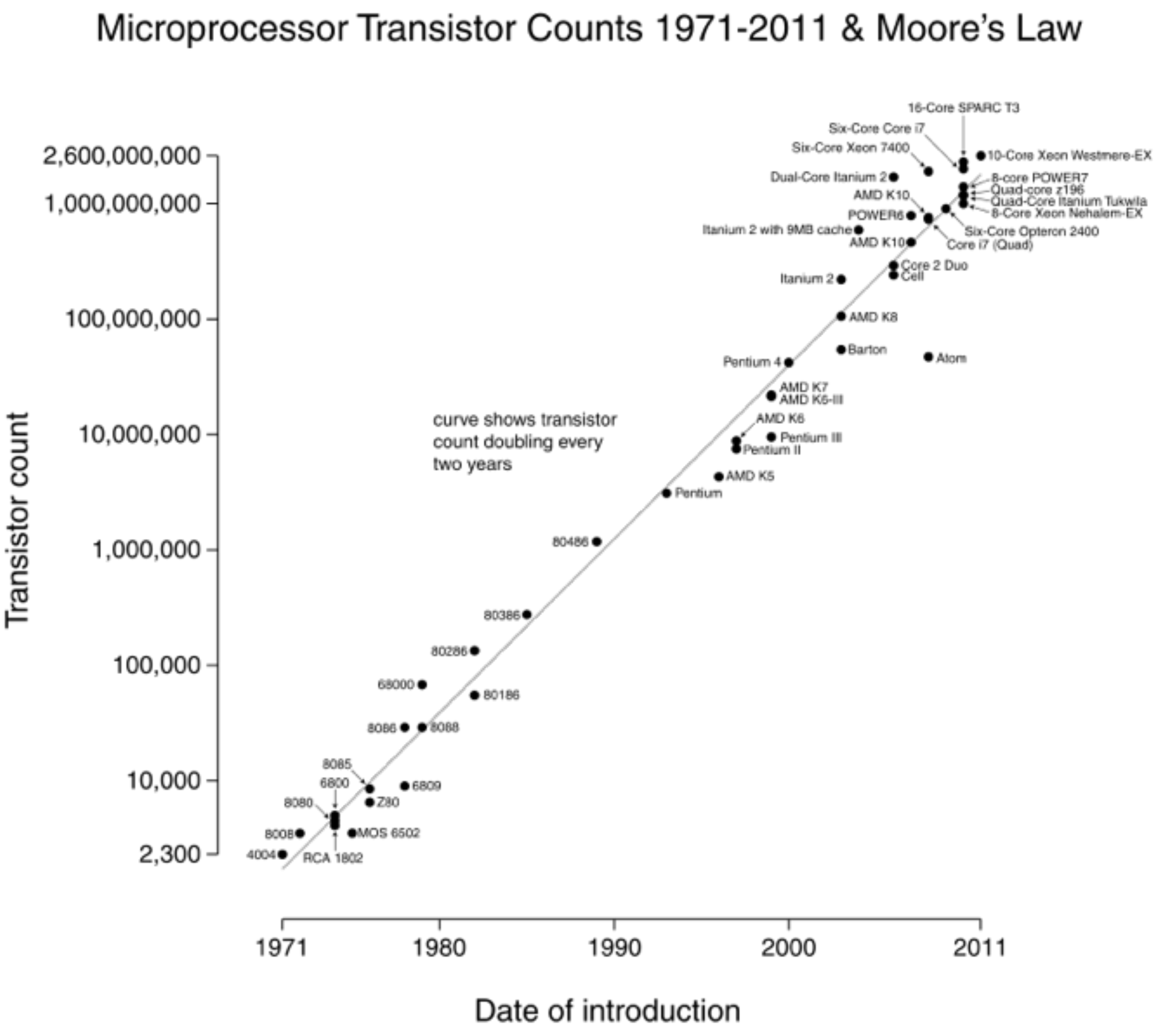
- 120Mpixel
- 20bit dynamic range
- 3 colors

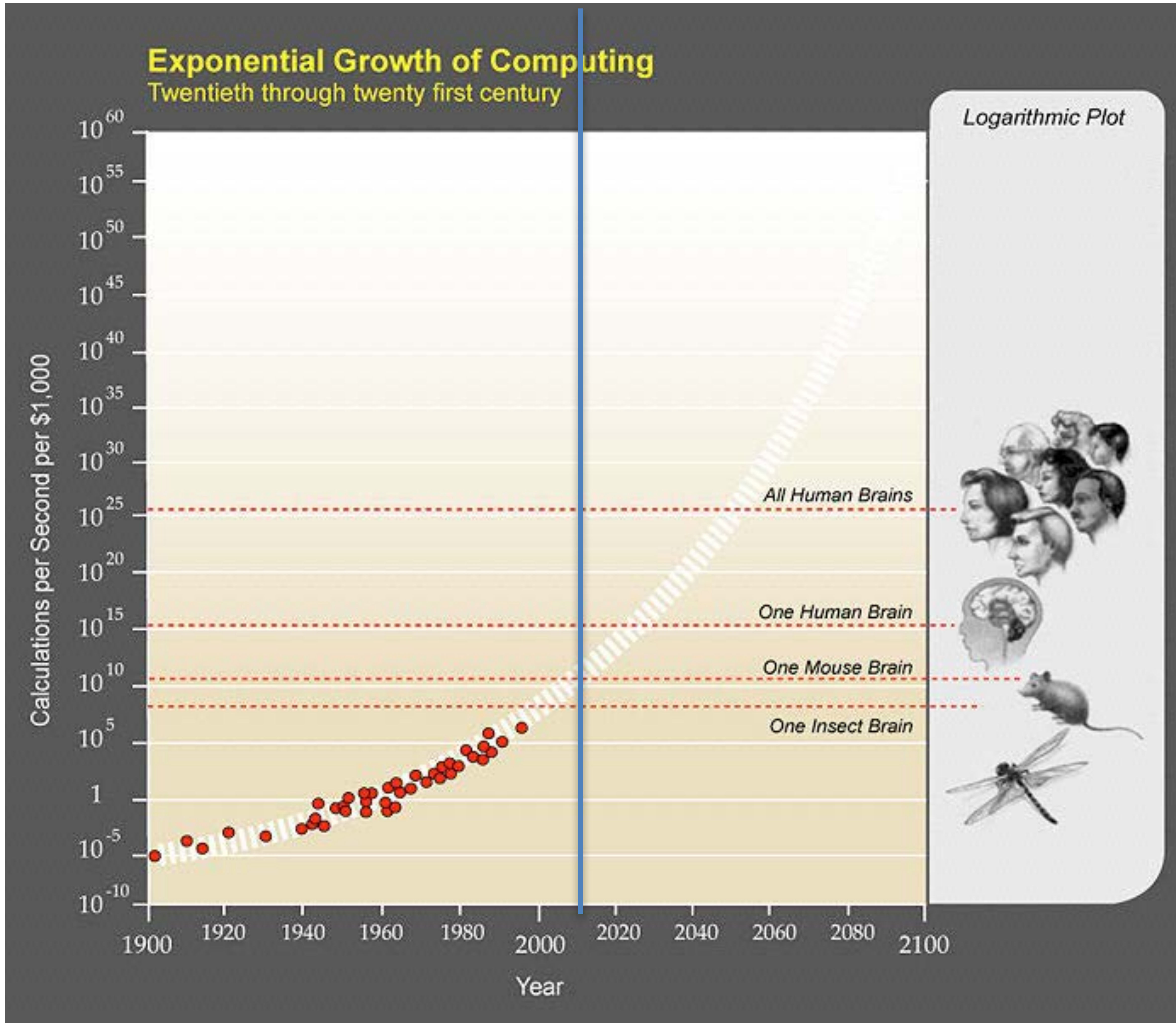
$\times 2$

- Vision engine
  - $3\text{x}10^{10}$ neurons
  - 500g
  - 6W

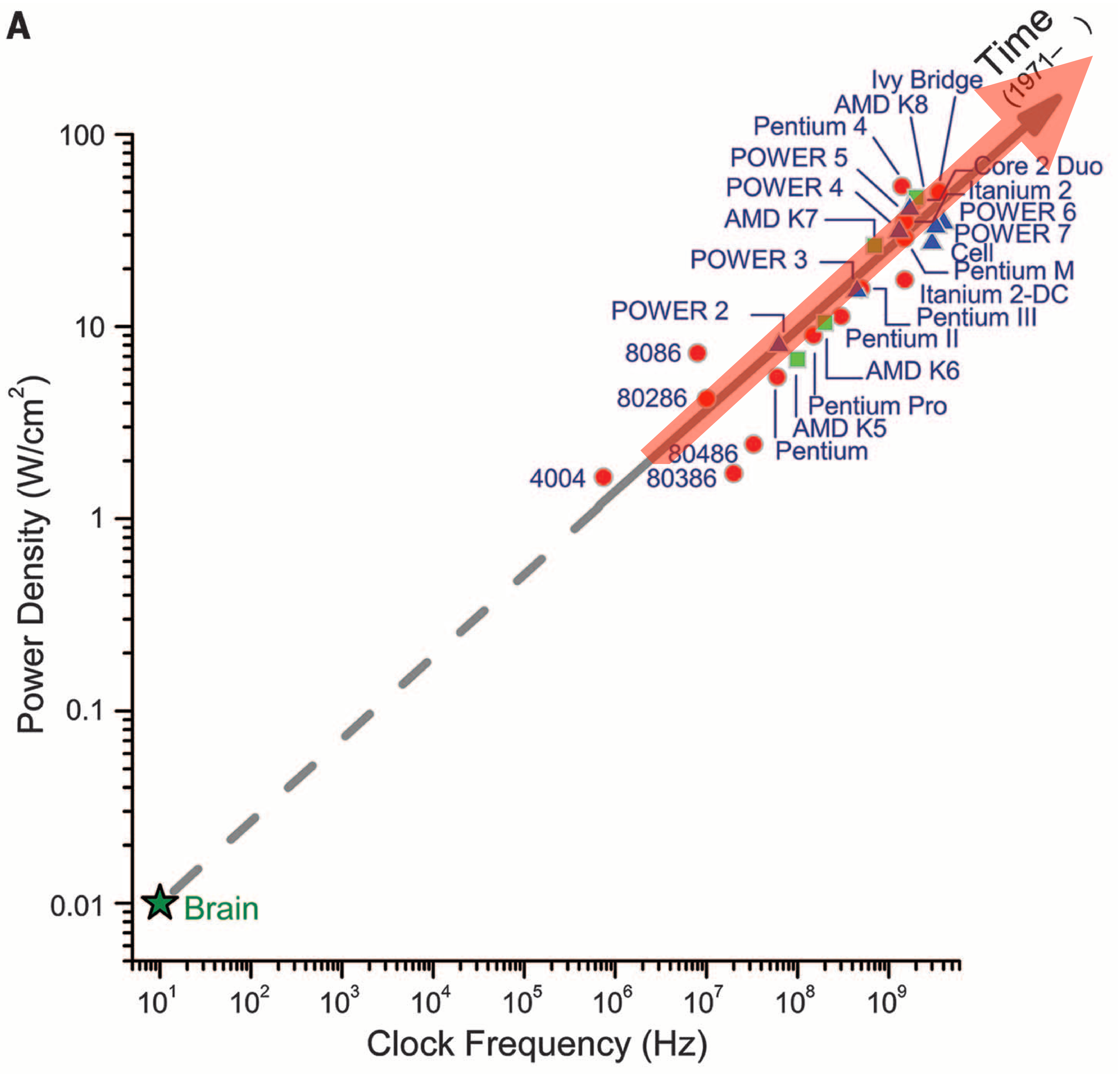
- 3 gyroscopes
- 2 accelerometers

$\times 2$

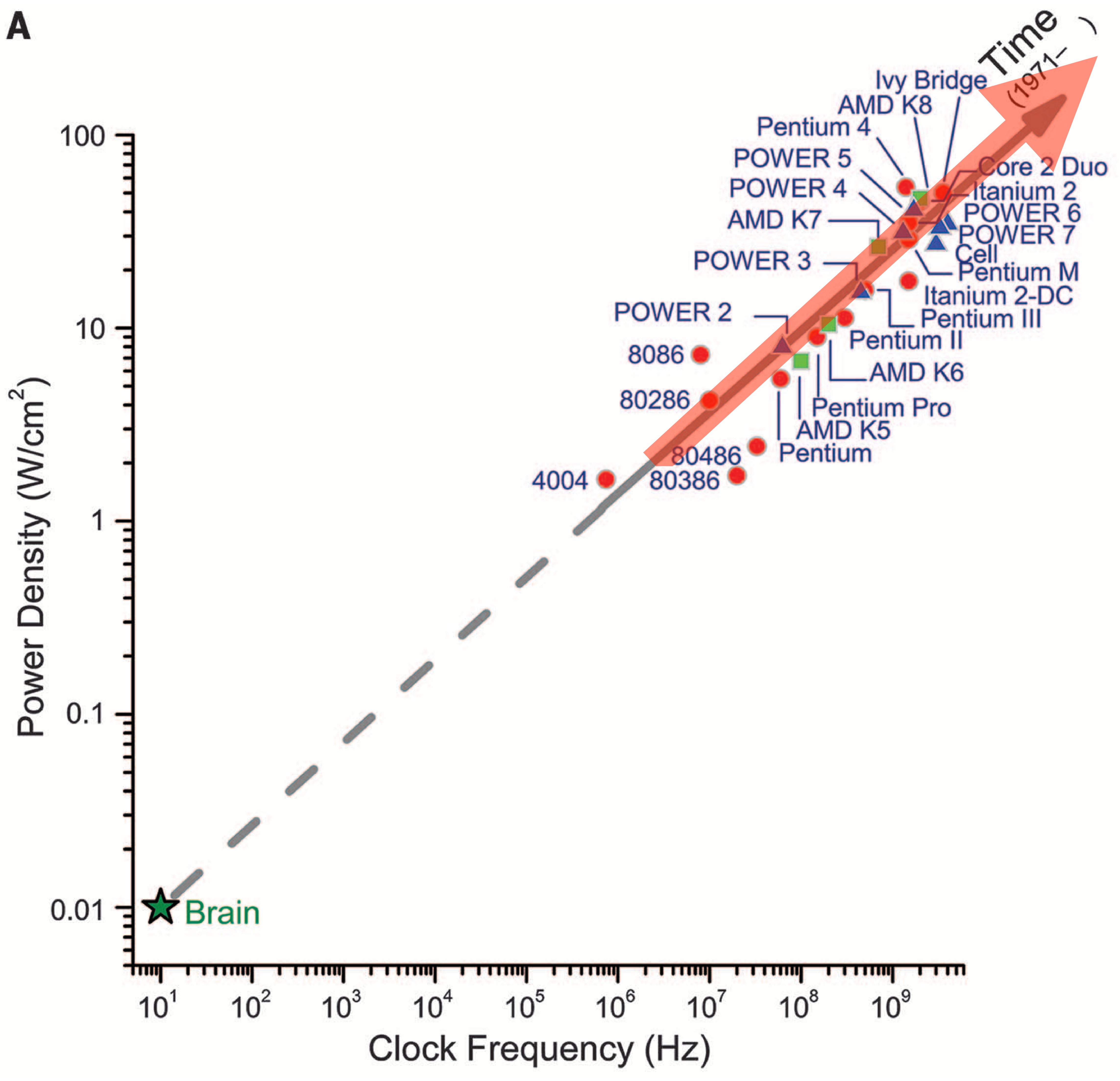




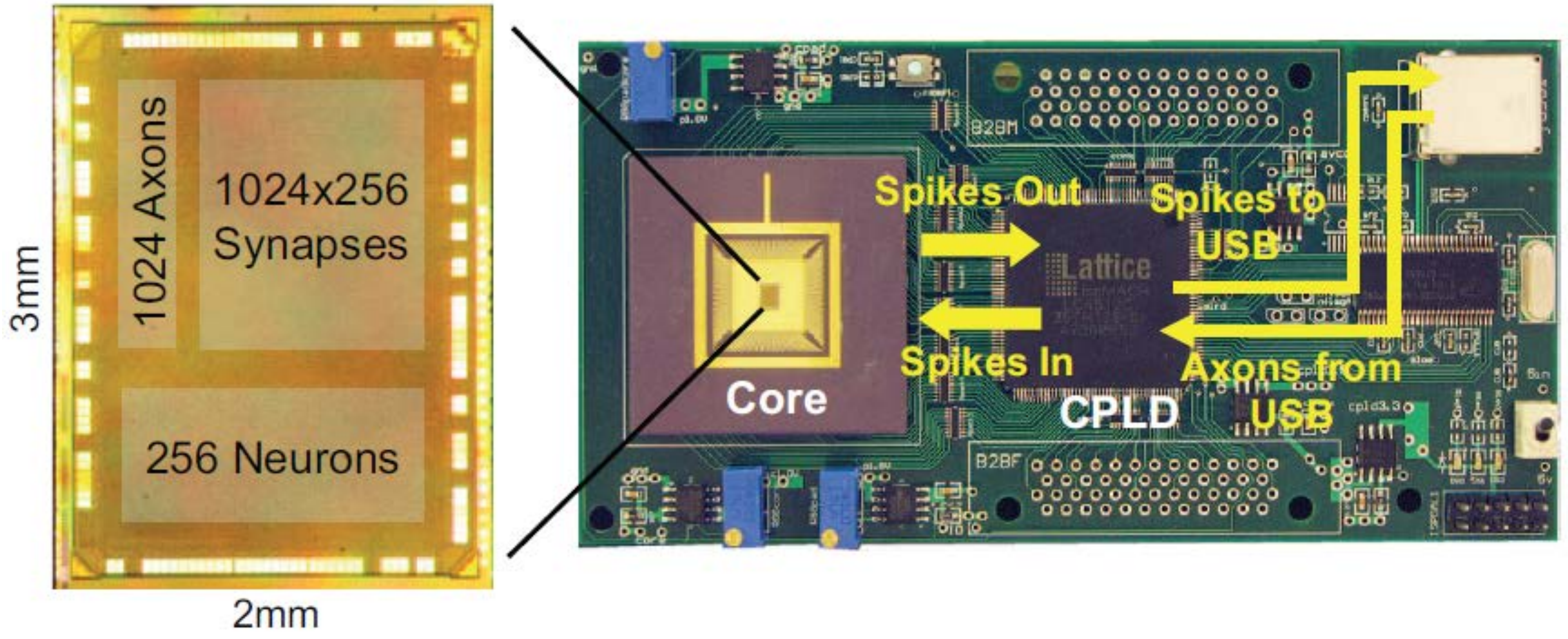
Ray Kurzweil


A
Time
(1971–)
Ivy Bridge
AMD K8
Pentium 4
POWER 5
POWER 4
AMD K7
POWER 3
POWER 2
Core 2 Duo
Itanium 2
POWER 6
POWER 7
Cell
Pentium M
Itanium 2-DC
Pentium III
Pentium II
AMD K6
Pentium Pro
AMD K5
Pentium
8086
80286
80486
4004
80386
Brain
Power Density (W/cm^2)
100
10
1
0.1
0.01
10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9
Clock Frequency (Hz)

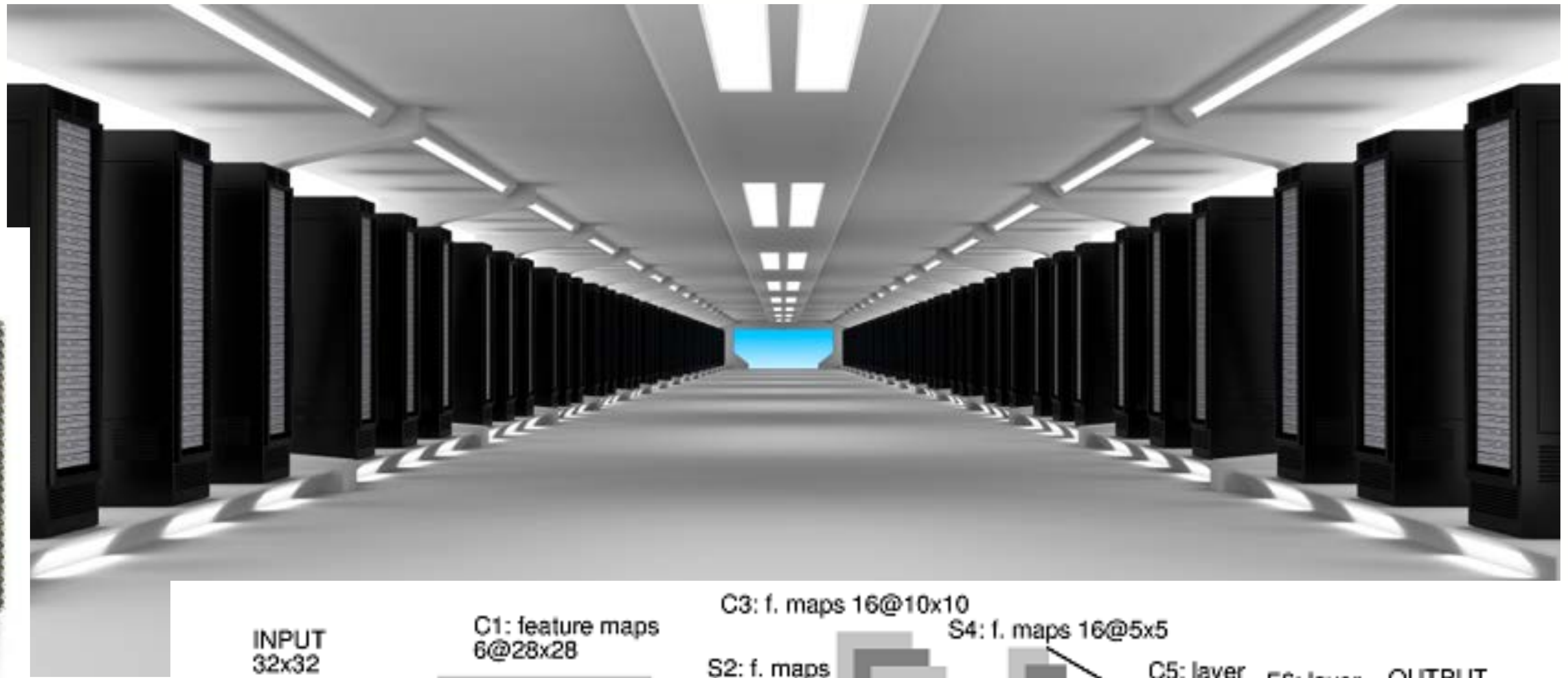

IBM
SEQUOIA


A
Time
(1971–)
Ivy Bridge
AMD K8
Pentium 4
POWER 5
POWER 4
AMD K7
POWER 3
POWER 2
8086
80286
80486
4004
80386
Core 2 Duo
Itanium 2
POWER 6
POWER 7
Cell
Pentium M
Itanium 2-DC
Pentium III
Pentium II
AMD K6
Pentium Pro
AMD K5
Pentium
Brain
Power Density (W/cm²)
100
10
1
0.1
0.01
10¹
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹
Clock Frequency (Hz)

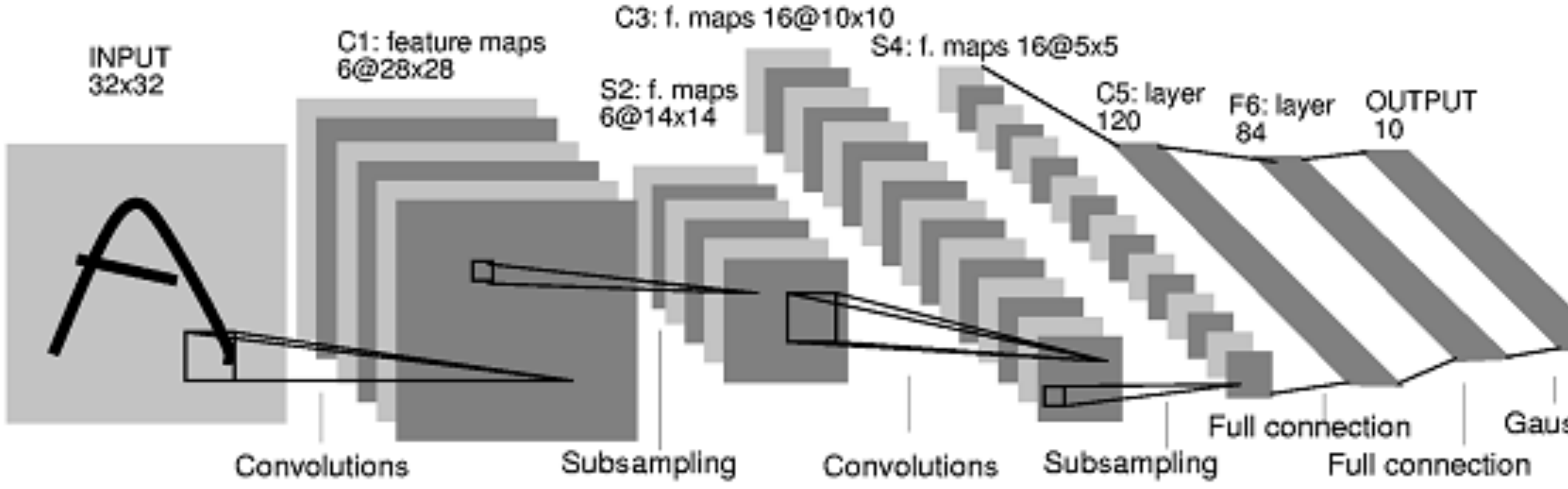

IBM
Neurosynaptic System

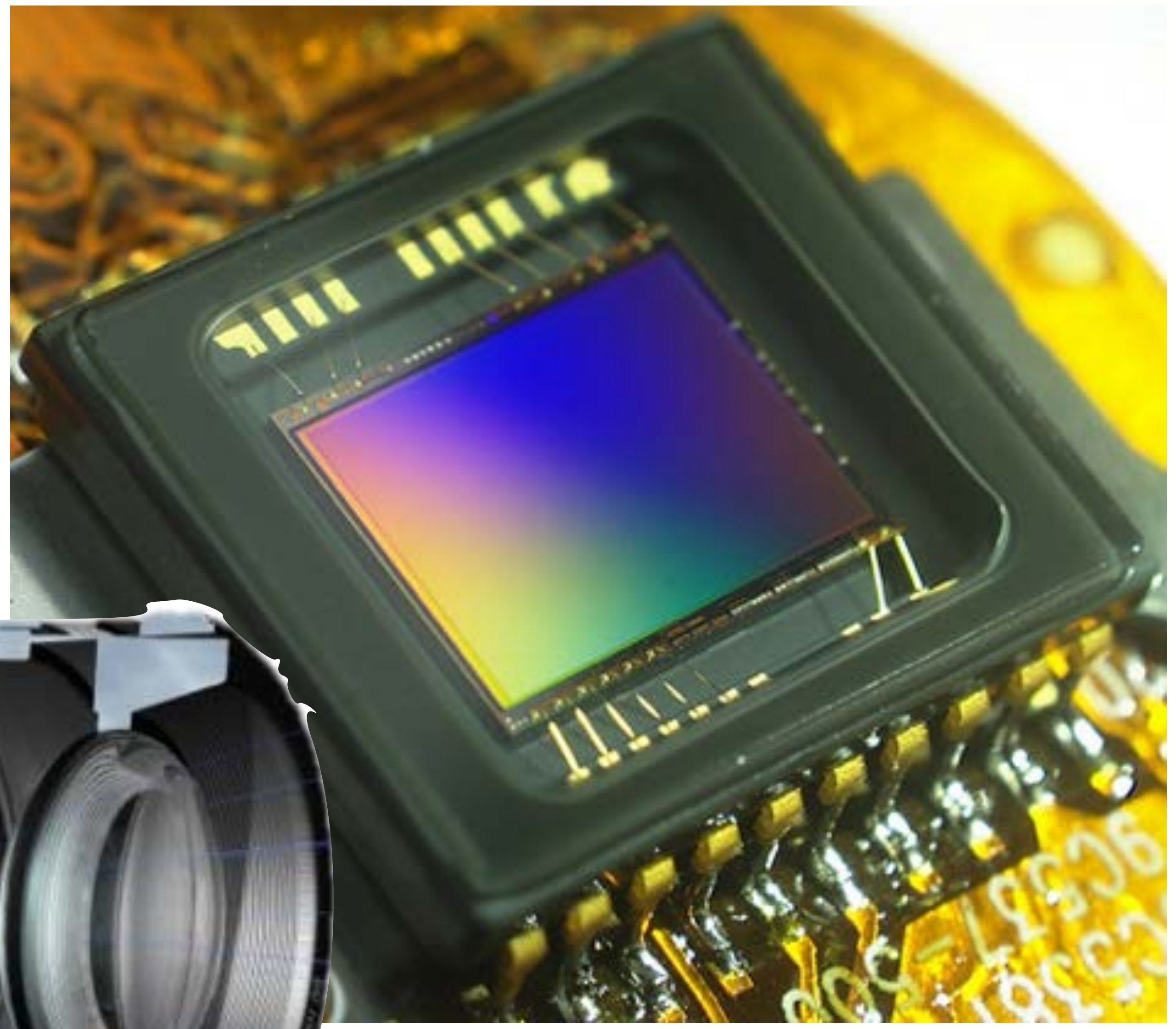

1024 Axons
1024x256
Synapses
256 Neurons
3mm
2mm
Spikes Out
Spikes to
USB
Spikes In
Axons from
USB
Core
CPLD
Lattice

What's different now?



- Computation (ops/$/W)
- Algorithms
  – all that computer vision stuff, particularly semantics
  – machine learning, CNNs etc.
- (Visual) neuroscience
- Sensors
  – light field
  – low light
  – high dynamic range





An ambition for a sensor

- Cost < $1,000
- Works in a usefully wide range of lighting conditions
- Provides geometric and semantic description of all objects in the scene
  - particularly people: activity, intent, etc.
- Low power
- Learns, exploits attention & context

ARC Centre of Excellence for
ROBOTIC VISION

$25M	16 new postdocs
7 years	50 PhD students
13 CIs	8 projects

Robust vision

- better sensors, comp. photography
- robust algorithms for poor images
- contextual priming

Semantic vision

- understanding from images
- lifelong learning

Vision & action

- seeing to move
- moving to see

Algorithms & architectures

- real-time, energy efficient
- new architectures
- local + cloud computing

QUT Queensland University of Technology

MONASH University

ANU THE AUSTRALIAN NATIONAL UNIVERSITY

THE UNIVERSITY of ADELAIDE

UNIVERSITY OF OXFORD

ETH Eidgenössische Technische Hochschule Zürich

INRIA

Imperial College London

Georgia Institute of Technology

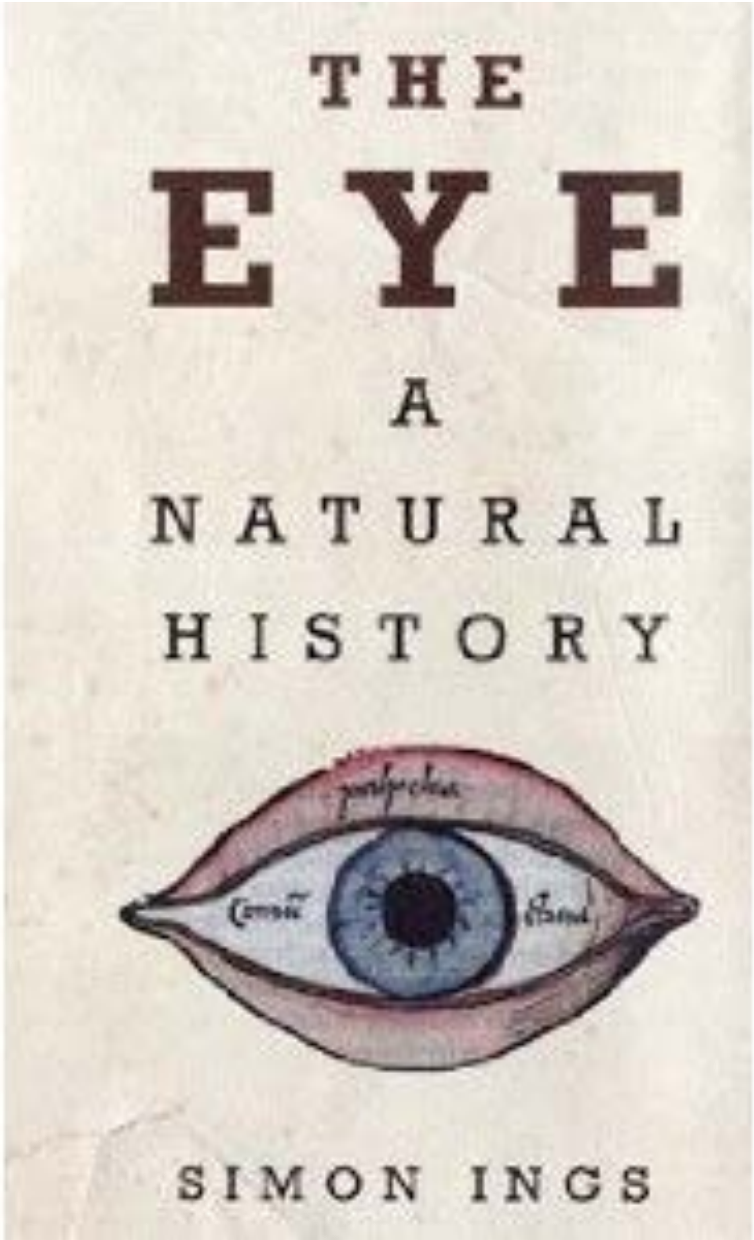
Vision is a **540M** year old technology

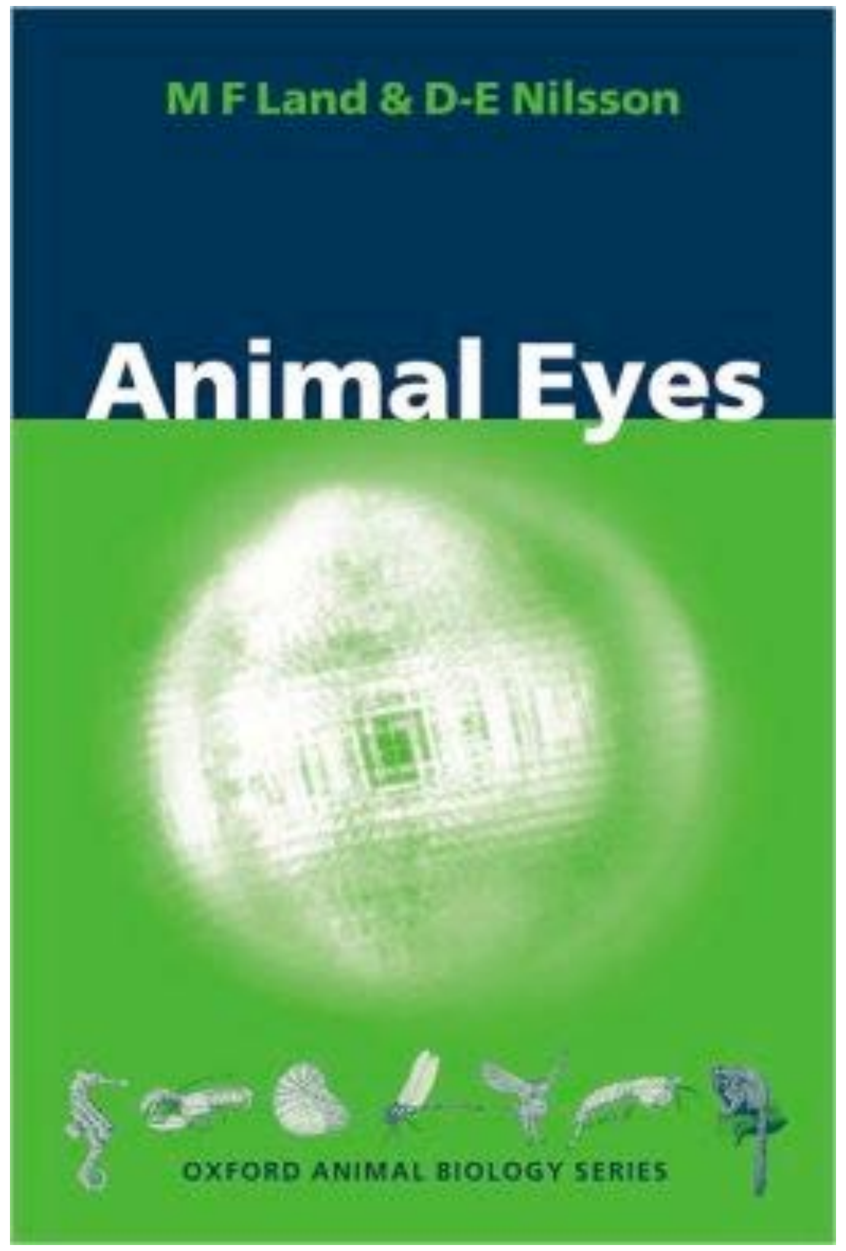
Radically transformed life on planet Earth

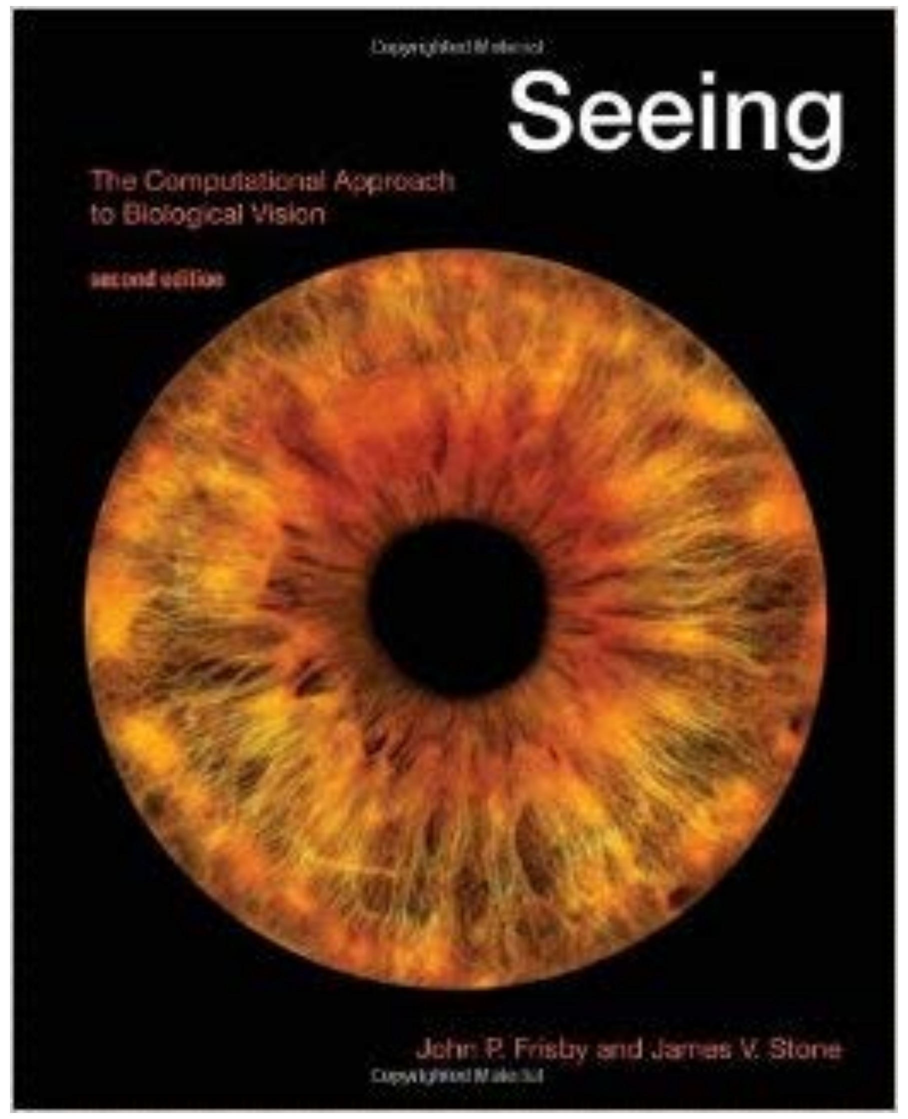
It is sufficient to enable almost all useful tasks

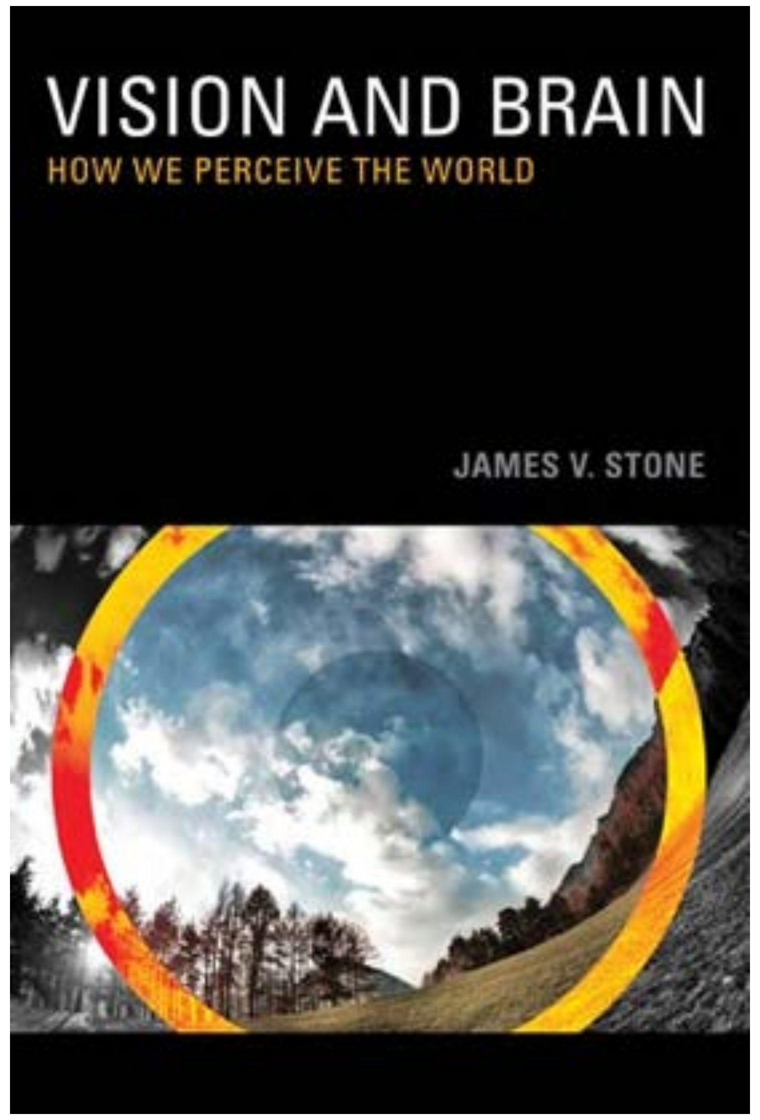
It's time to bring vision and robotics back together!

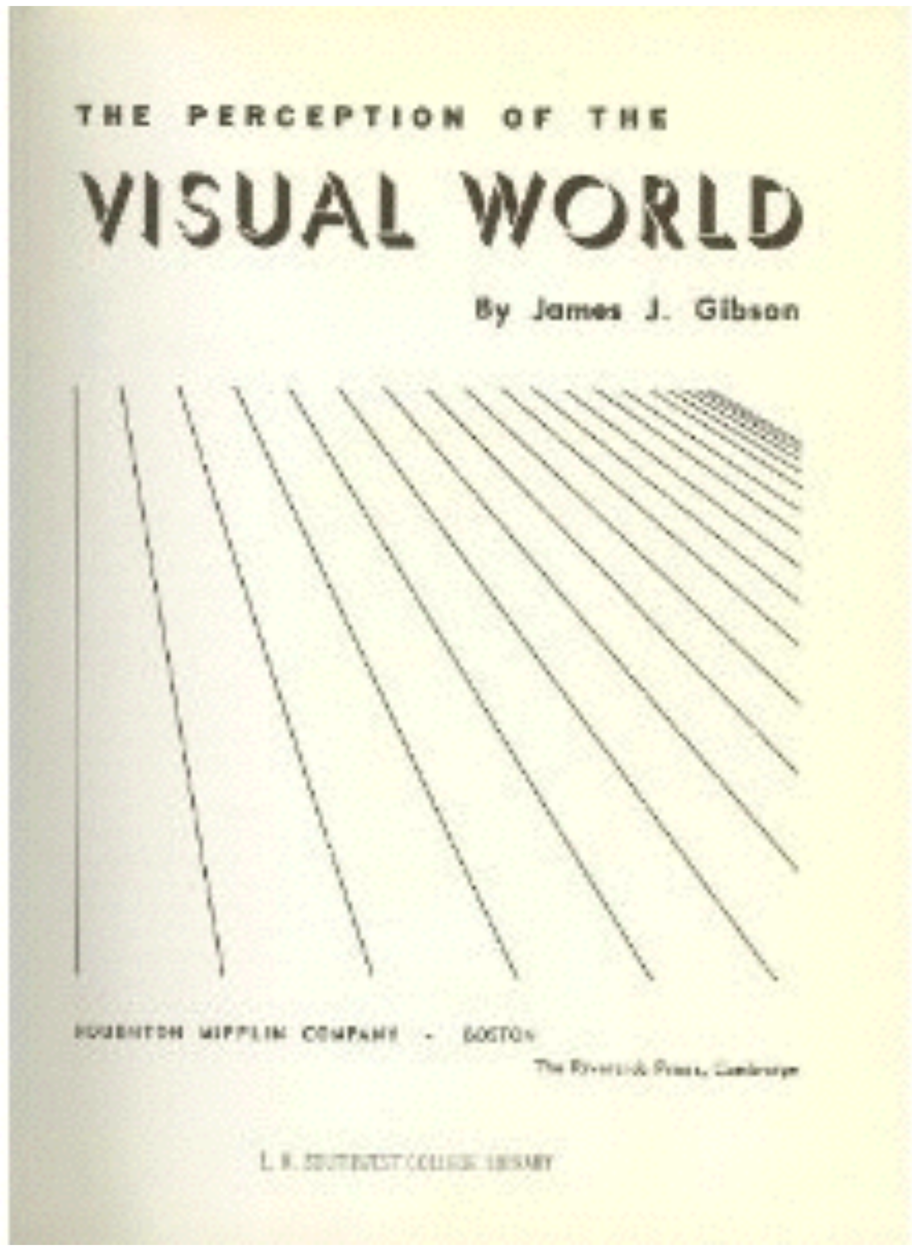
Further reading

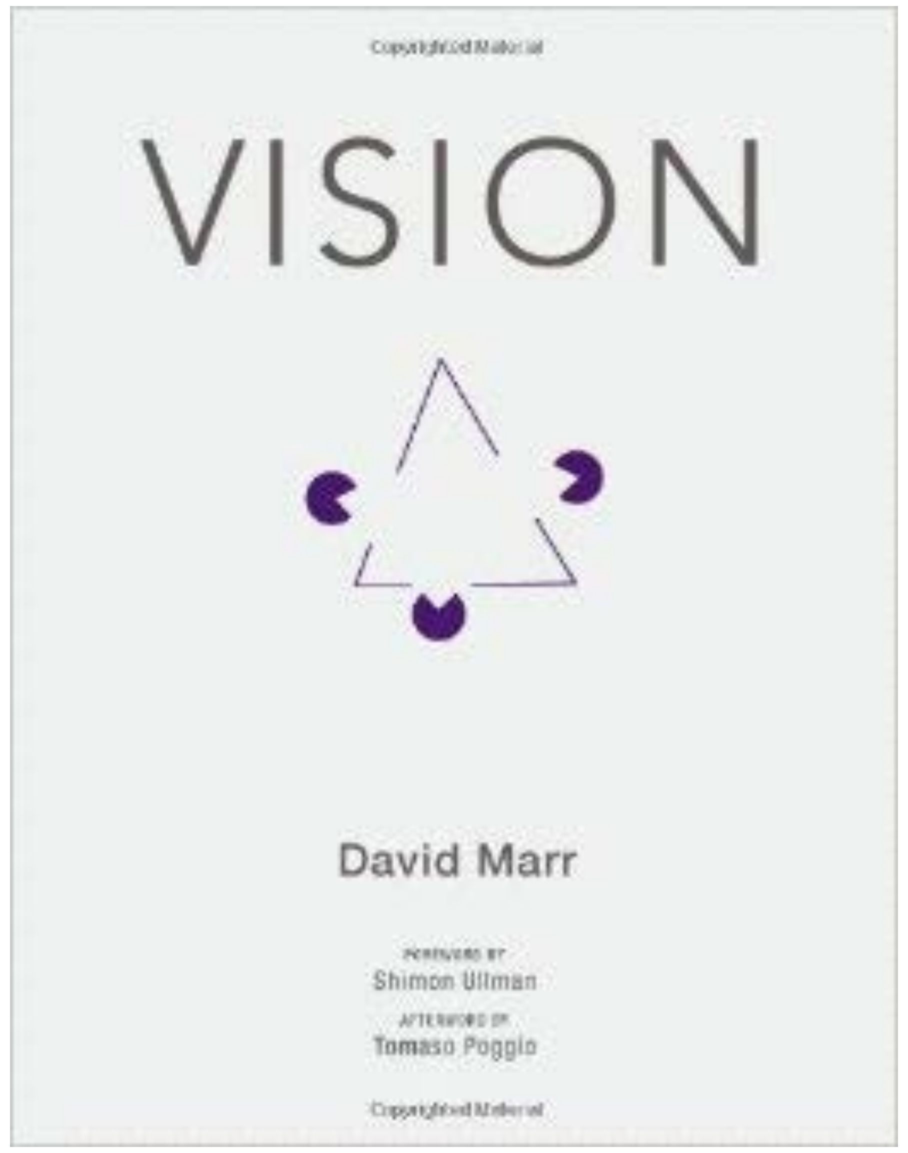


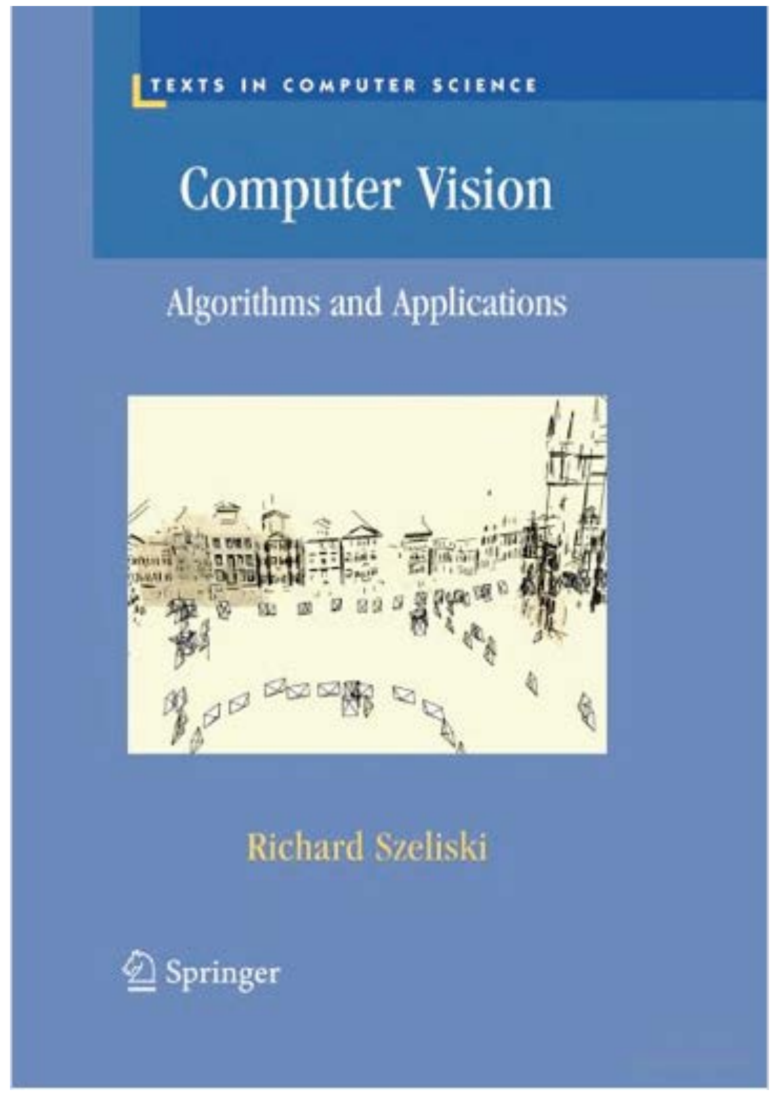


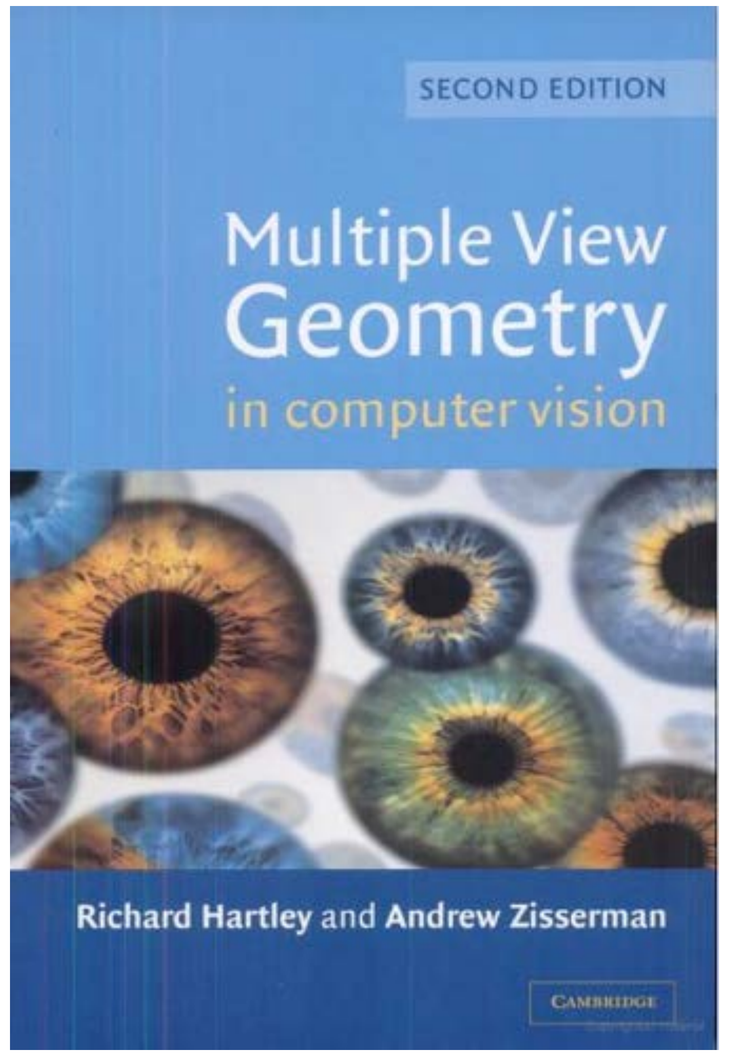


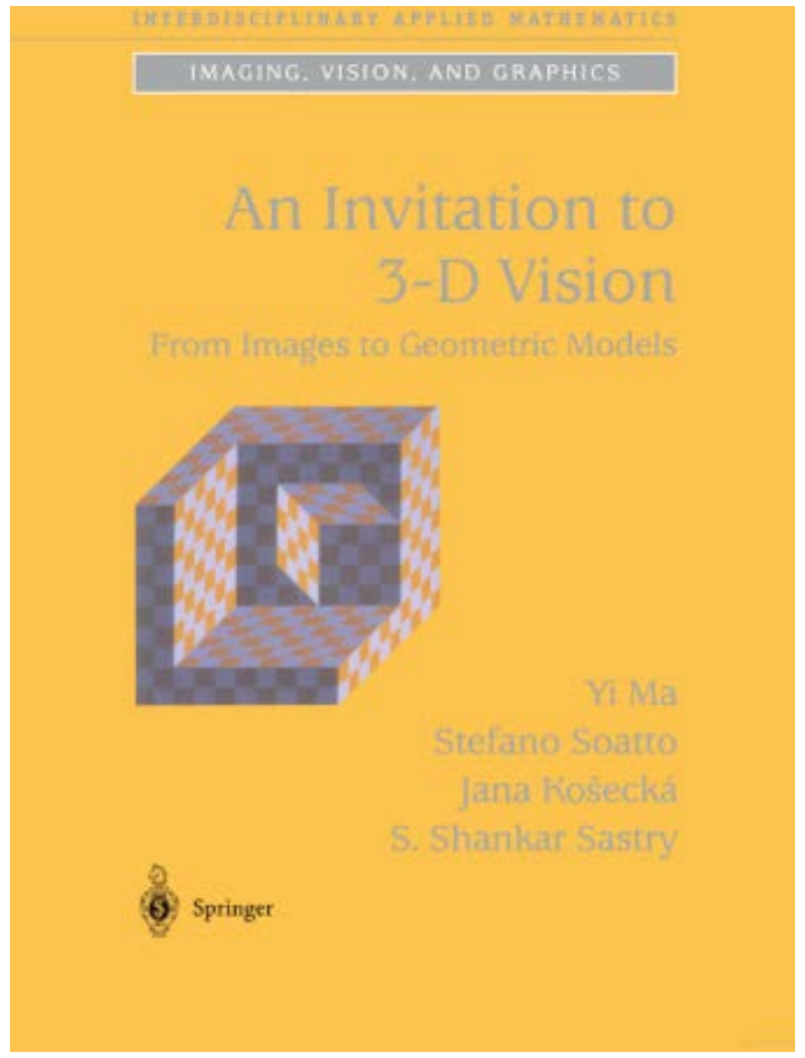












Vision is a **540M** year old technology

Radically transformed life on planet Earth

It is sufficient to enable almost all useful tasks

It's time to bring vision and robotics back together!

ARC Centre of Excellence for ROBOTIC VISION

www.roboticvision.org

Cave Paintings ~40,000 years ago

Ideal City (1470)
Piero della Francesca (1415–1492)

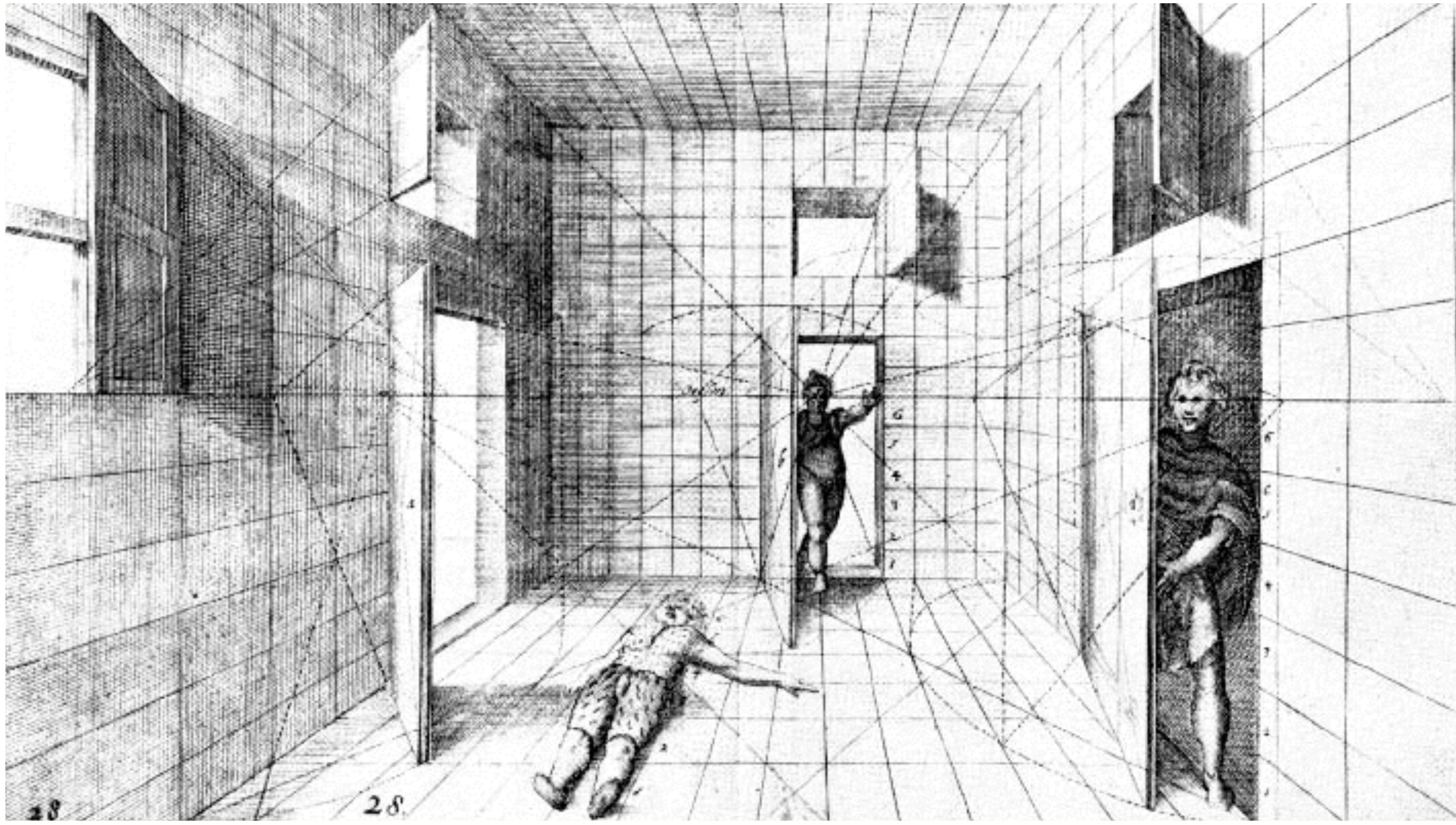


Figure 28 (Jan Vredeman de Vries, 1604).
Used with permission from Perspective, Dover Publications,1964.

People are actually avoiding walking in the "hole" 2007
Joe Beever | CC A2.0

Stunning 3D chalk drawing from Zebit stops Liverpool shoppers in their tracks on Bold Street. *2012*
Bill Hunt Original art: **Zebit** | CC A2.0

Edgar Meuller http://www.metanamorph.com