COL7655 (3-0-2)
Foundations of Visual Computing

Overlaps: 780, 781, 783

Subodh Kumar
subodh@cse.iitd.ac.in

Course Administration

- Instructor: Subodh Kumar, subodh@cse.iitd.ac.in, (Bharti 429)
- Website: www.cse.iitd.ac.in/~subodh/courses/COL7655
- Assignments to be submitted on Moodle (moodlenew.iitd)
- Course discussion on Moodle, MS Teams (2601-COL7655: **5bjt9ca**)

Evaluation

- Assignments: 28% (will include viva)
  - Individual, must pass similarity check (even if you use LLM)
- Exams 60% (Minor: 30%, Major: 30%)
- Quizzes (including class participation): 12%
- 4 in Quizzes, 8 in Assignments, and 18 in Exams required to pass
  - Or, 35 total for D
- 22 total for E
  - 45 total marks are required for NP

Assignments

1. Image enhancement and filtering (7 marks, Due 17 Aug)
2. 3D Modeling and rendering (7 marks, Due 7 Sep)
3. Volume rendering (7 marks, Due 11 Oct)
4. Object tracking for AR (7 marks, Due 17 Nov)

Other Policies

Attendance:

100% required. No re-test, or NP/I grade if attendance < 75%

+2 for class participation and best 3 quizzes taken if attendance >= 90%

Plagiarism:

Similar codes will result in *every matching student* receiving negative of the assignment marks

Second similarity offense gets an F

Cheating in written exam gets an F + Disco

Late assignment submission:

5 free late days for the semester, beyond that -1 mark per extra late day (or part)

Textbooks

No required text, but the following are highly recommended for reference:

- Aditi Majumder, M. Gopi, *Introduction to Visual Computing Core Concepts in Computer Vision, Graphics, and Image Processing.*
- Peter Shirley, Steve Marschner, *Fundamentals of Computer Graphics*
- Rafael Gonzalez, Richard Woods, *Digital Image Processing*
- David Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*

What is “Visual” Computing

COL7655: Generate, transform, and understand Images

- Related to sight 
  - What can be seen (scene) or imagined (idea)
  - … and communicated
- Images
  - Can describe light distribution, diagrams and illustrations, icons, even text
  - Encapsulate shapes and color
  - Convey and inform

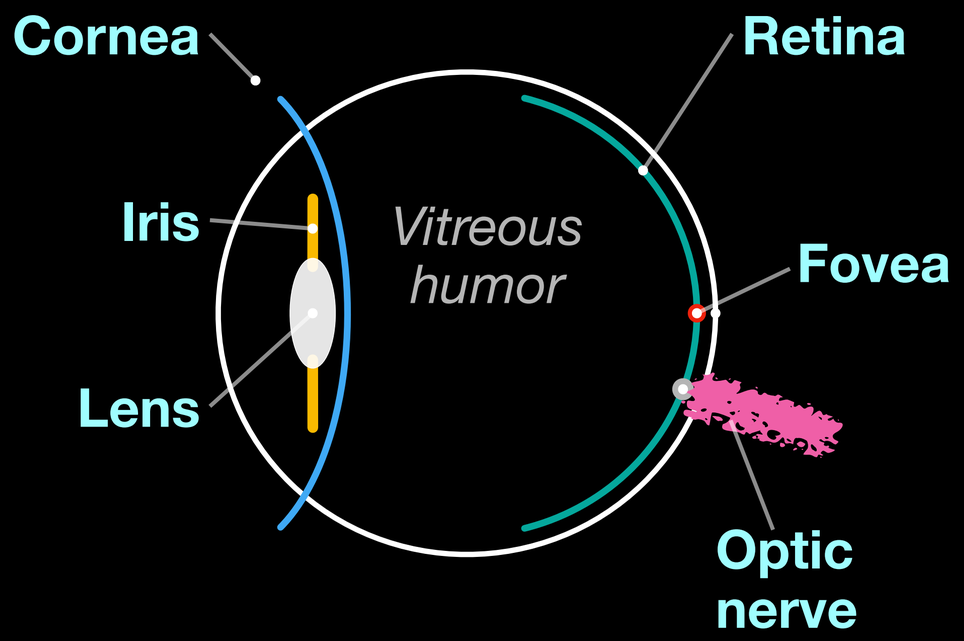
“*Mind’s eye*”

Imagination cannot exceed the original stock of ideas, furnished by the internal and external senses [Hume, 1777]

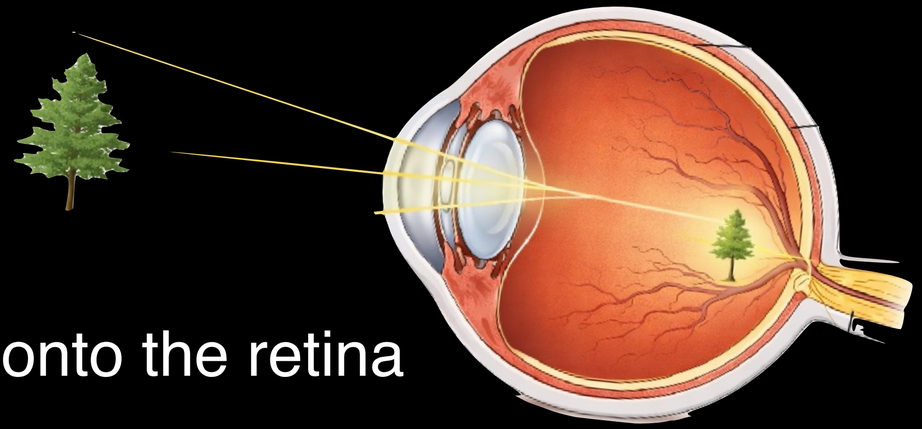
Course objective:
Convey information through images
Understand information in images

Anatomy of the Eye

- Cornea and lens: focus incoming light onto the retina
- Iris: adjusts pupil size to control how much light enters
- Retina: converts light into electrical signals using photoreceptors
- Fovea: the small pit that provides sharpest central vision
- Optic nerve: carries visual signals from the retina to the brain



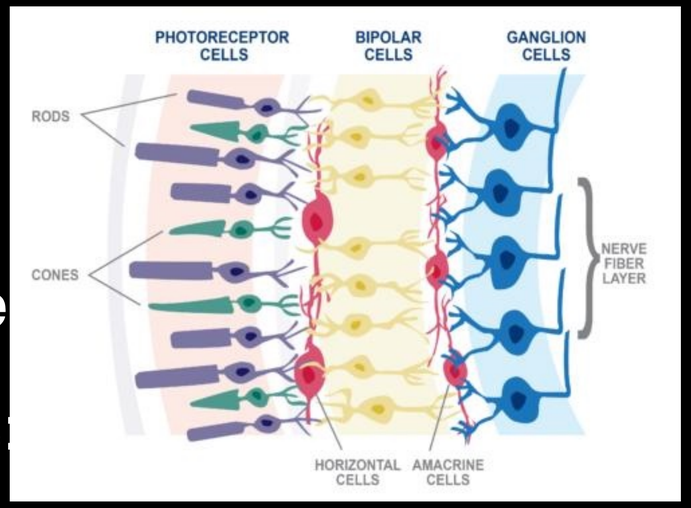
Seeing an Image



- Light entering iris is focused by the adjustable lens onto the retina
- Photoreceptors on retina convert light into electrical signals:
  - ~120 million rod cells: highly sensitive; night and peripheral vision; no color
  - ~6 million cone cells: less sensitive; daylight, color vision, and high acuity
    - Three types: More sensitive to Short, Medium, Long wavelengths, respectively
- Helper cells: Horizontal, Amacrine, Bipolar, Ganglion
  - Integrate from neighbors and over time, convert to neural currents, control pathways

Seeing an Image

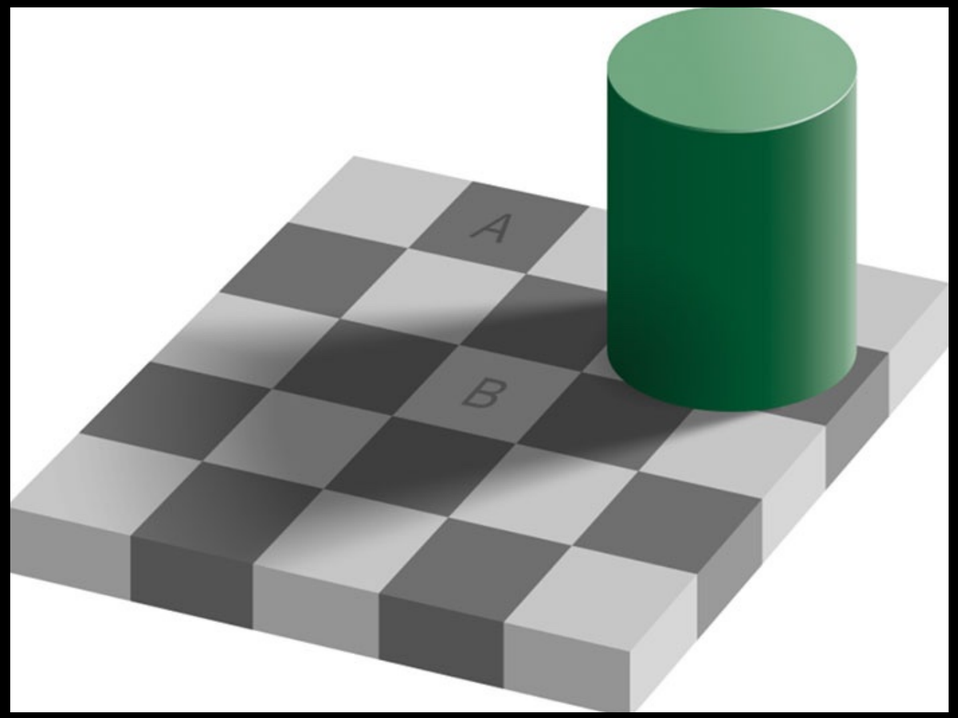
[Lambdavision]



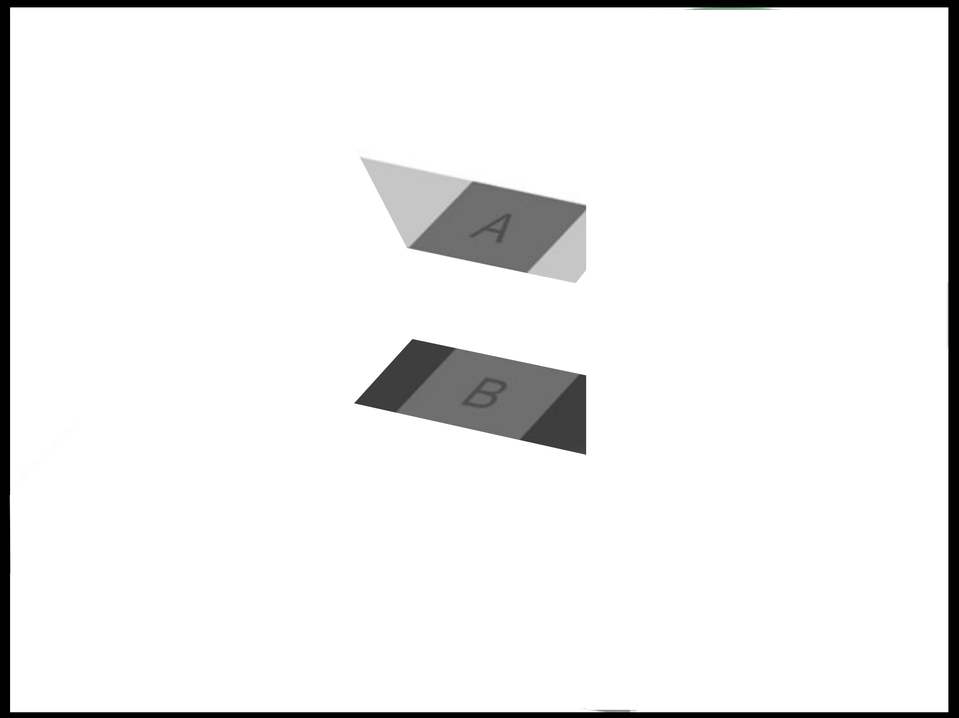
- Light entering iris is focused by the adjustable lens onto the
- Photoreceptors on retina convert light into electrical signals
  - ~120 million rod cells: highly sensitive; night and peripheral vision; no color
  - ~6 million cone cells: less sensitive; daylight, color vision, and high acuity
    - Three types: More sensitive to Short, Medium, Long wavelengths, respectively
- Helper cells: Horizontal, Amacrine, Bipolar, Ganglion
  - Integrate from neighbors and over time, convert to neural currents, control pathways

Visual Perception

- Brain fills gaps in visual signals based on expectation to construct information

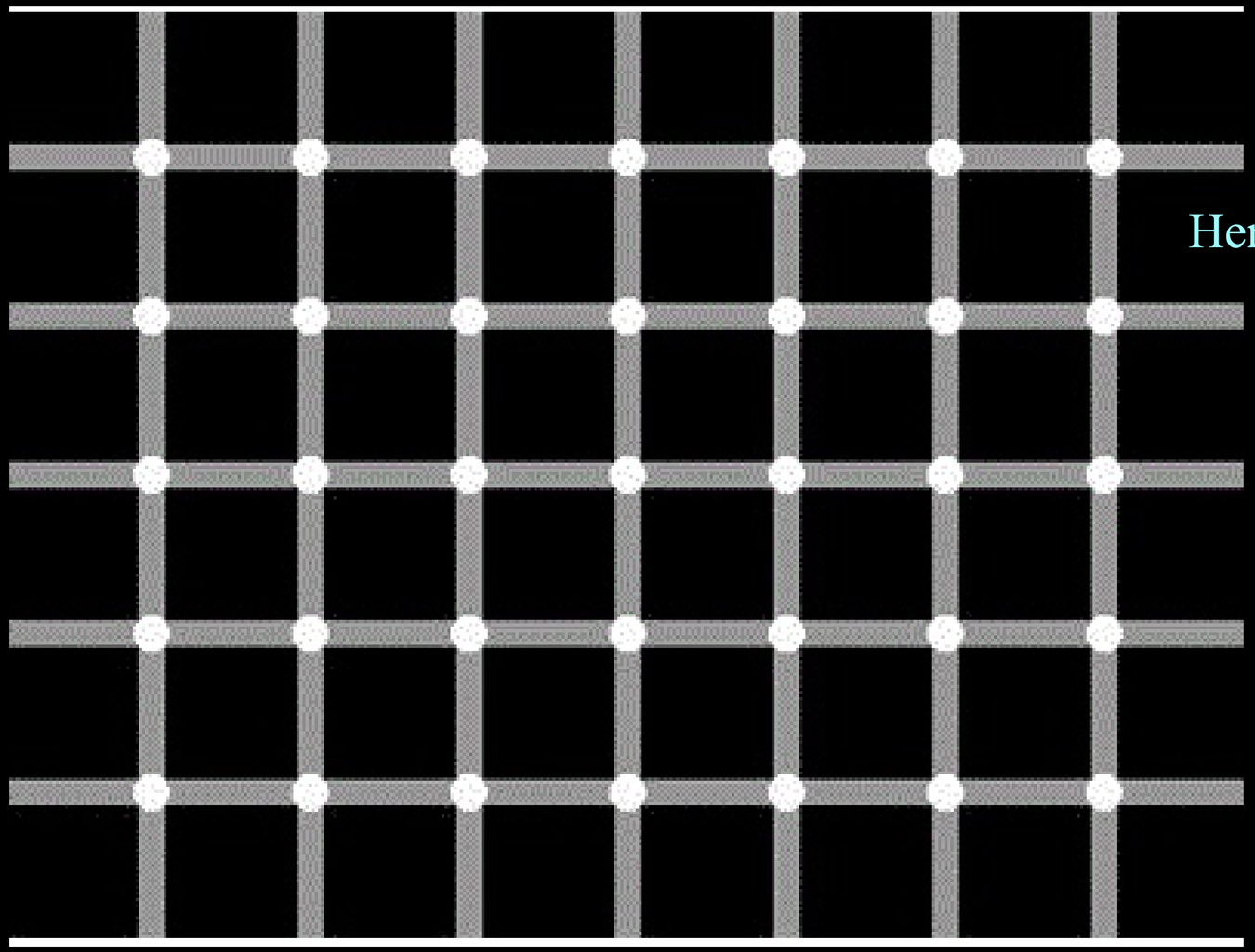


Adelson's Checker-board



Visual Perception

- Brain fills gaps in visual signals based on expectation to construct information



Hermann Grid

How to Convey Visual Signal

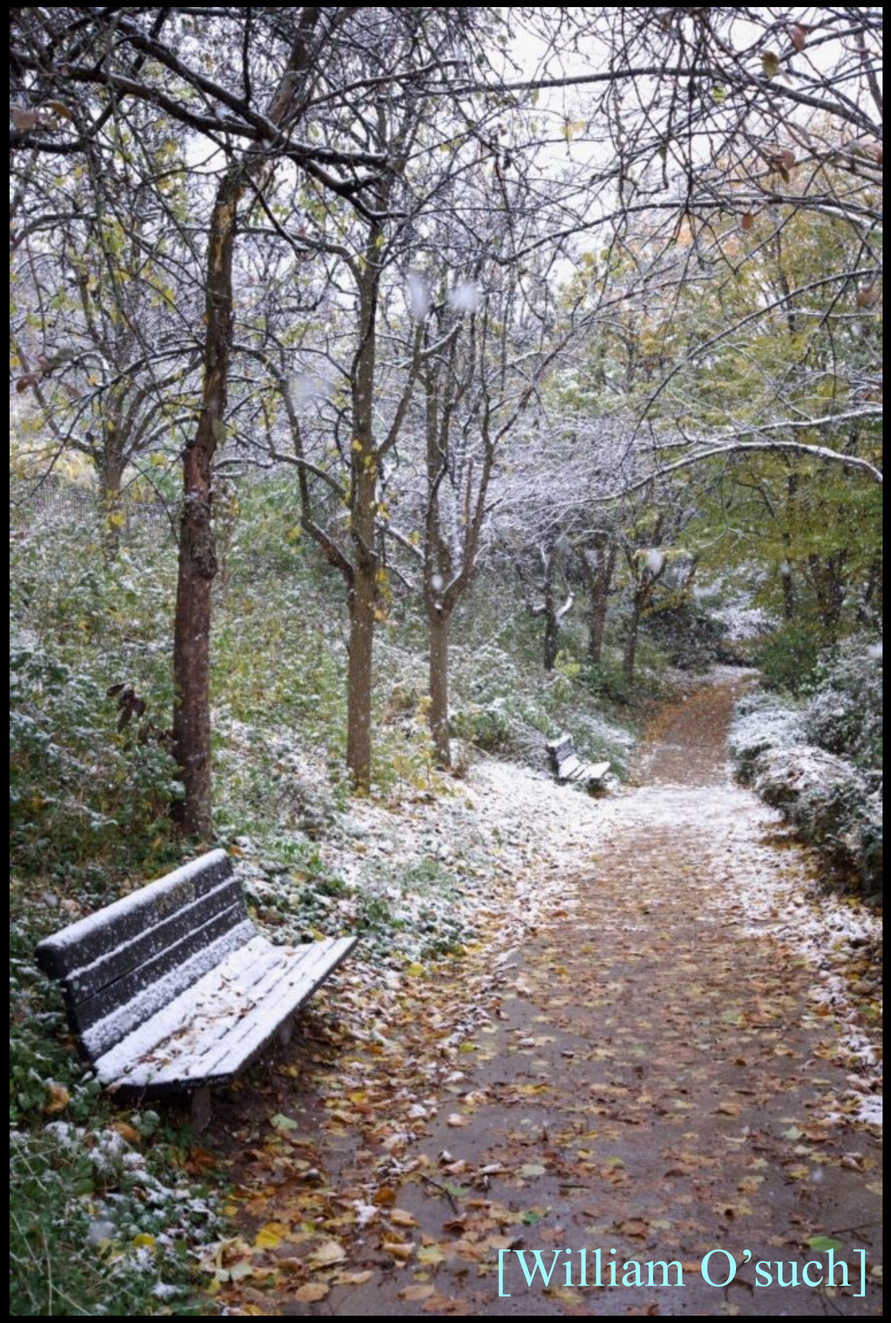
Paint a picture

- Activation of millions of photo-receptor cells
- Photons (or waves) representing the sum of all light reaching eye
- A high level description
  - Sources of light, transport of light, reflection of light
- Textual representation
  - Snow-clad mountains in the distance behind a lake reflecting the sky with 10 swans and grass-bank
- Icons

Image-based Reprsentation

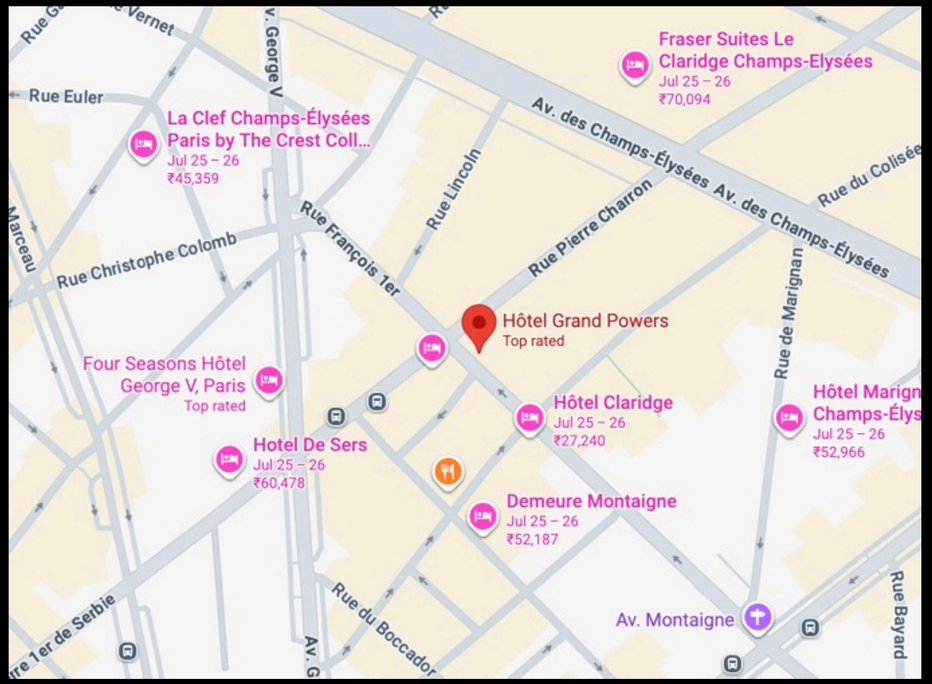


[Paris Snow Scene by Christopher Wood]



[William O’such]

Image-based Representation



[Shutterstock]



Image-based

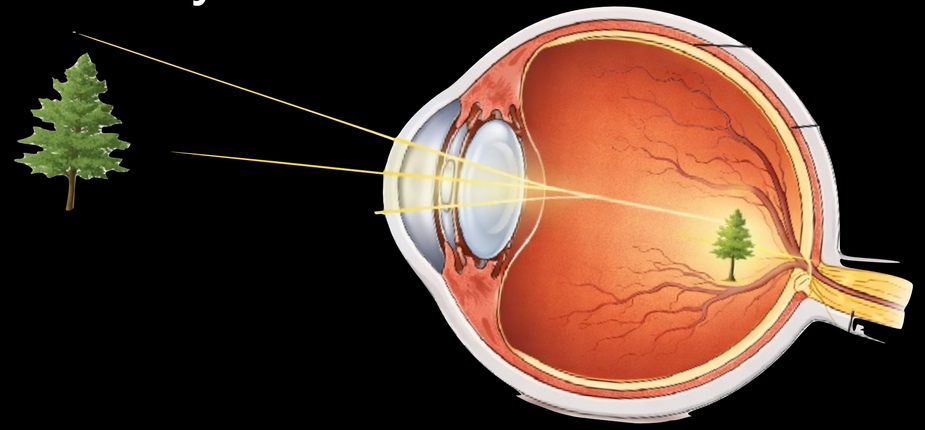
Downward, aerial perspective of a Parisian street and square as viewed from the window of a flat, framed by leafless, dark-grey trees against a snowy, white background.



Image Formed by Light

from scene or image?

- Light reflected by an object passes through eye lens and reaches a photoreceptor in the retina
- What is the difference if the “same” light is generated by an artificial source
  - Computer screen, photograph, painting …
  - Does point of origin matter?
    - Relates to depth perception
- How to generate that many (more than 10^{20}/s, e.g.) “same” photons?



Location of the Source of Light

- Humans have stereoscopic vision
- Depth cues
  - Parallax (Relative shift with motion, depending on the distance)
  - Convergence of eye to focus on nearby objects
  - Accommodation of lens to get object to focus on retina

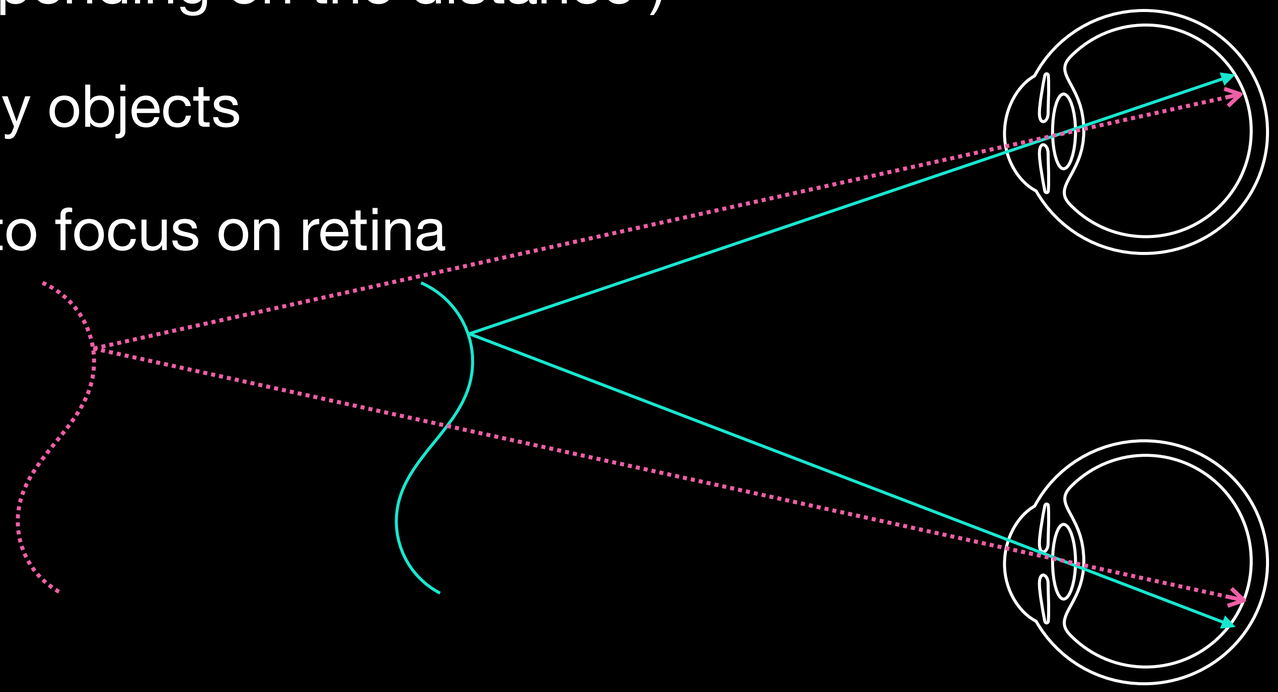
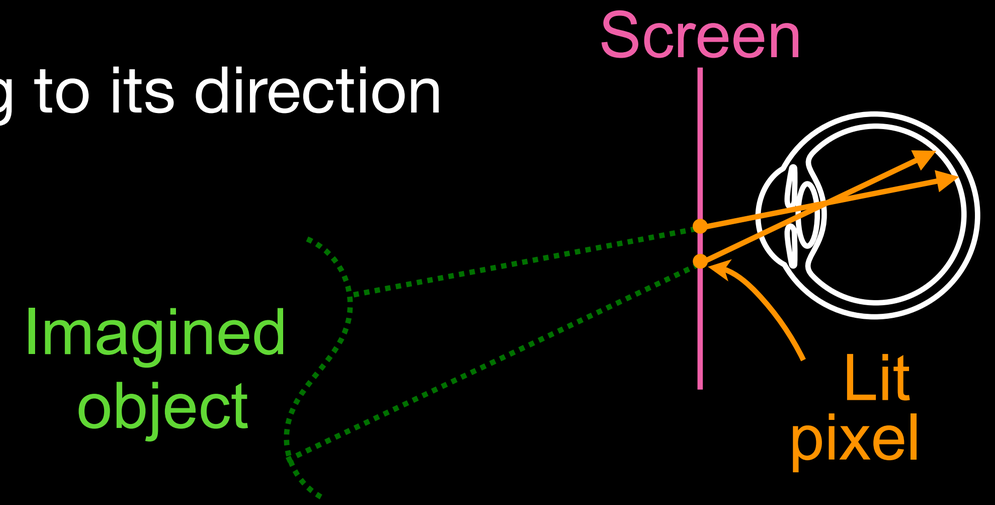


Image on a Screen

Picture masquerading as the source of light

- Light generator (pixel) on a grid
  - On a “Flat” surface (usually)
  - e.g., 2560x1664 pixels, ~224 pixels per inch
- “20/20” angular resolution is ~60 points per degree (~300 points at 1’ distance)
- Each pixel must produce light corresponding to its direction



A Simple Abstraction

Raw bits; Appearance; Content/Information

- Image is the *integration* of *light* coming from “all” directions for a certain duration
  - Parameterized directions, e.g., $I(\theta,\phi,t)$
- For digital processing, I may be sampled along (θ,ϕ) or (θ,ϕ,t)

✔ Our focus

  - Usually a uniform grid (x, y) instead of (θ,ϕ)
- Or, vectorized (as lines, curves or other 2D shapes)
- I values are themselves quantized
  - ***e.g.,*** triplet of numbers, called Red, Green, and Blue, respectively

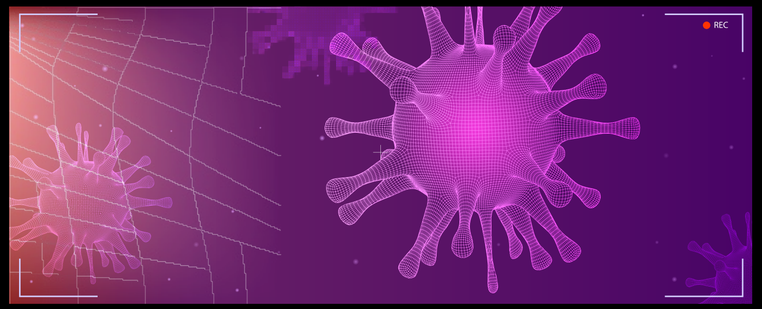
In many cases, the original continuous domain I can be reconstructed from the samples (aka PIXELS)

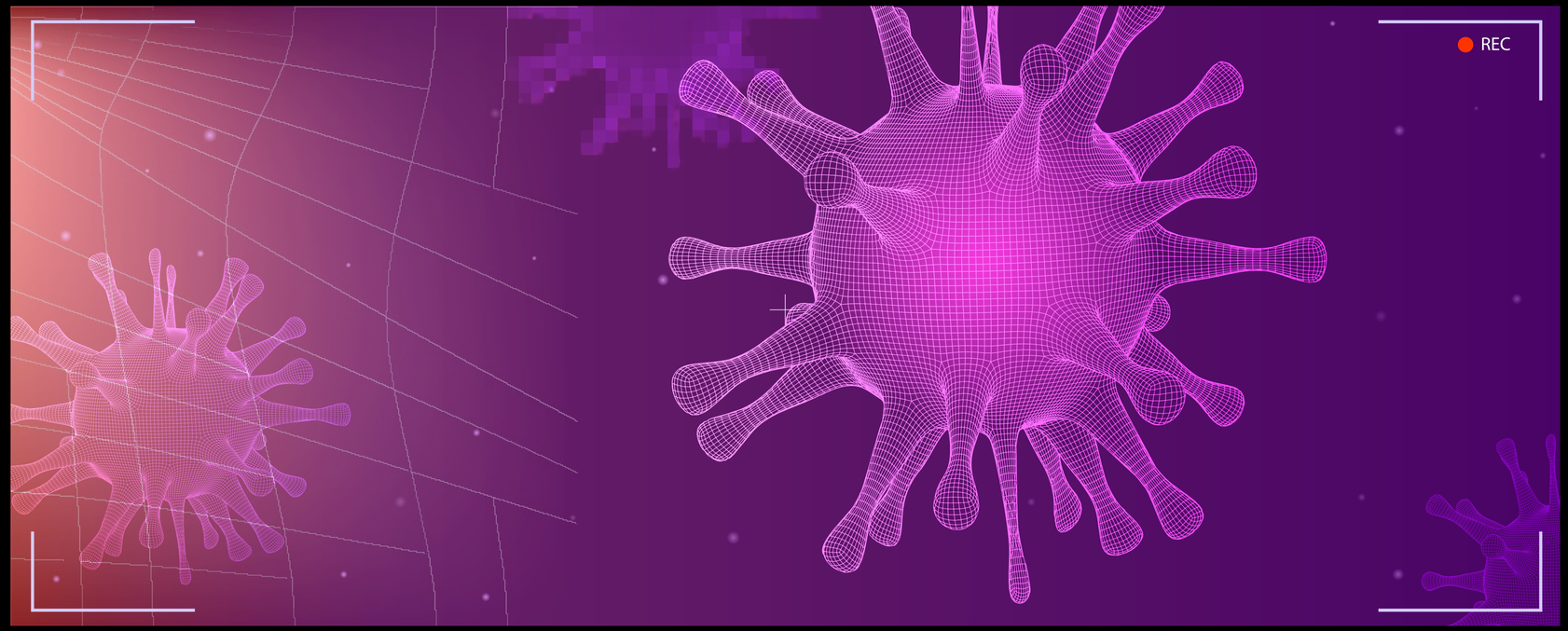
Digital Raster Image

- A 2D array of “color” values
  - Color may be a single number, integer or real representing brightness
  - Or, three numbers together representing chromacity and brightness
    - Usually a three dimensional space spanning human visual appearance
    - But multi-spectral images are used in specialized situations
- The most common value representation is RGB8
  - The three axis are “redness,” “greenness,” and “blueness”
  - An 8-bit value on each axis

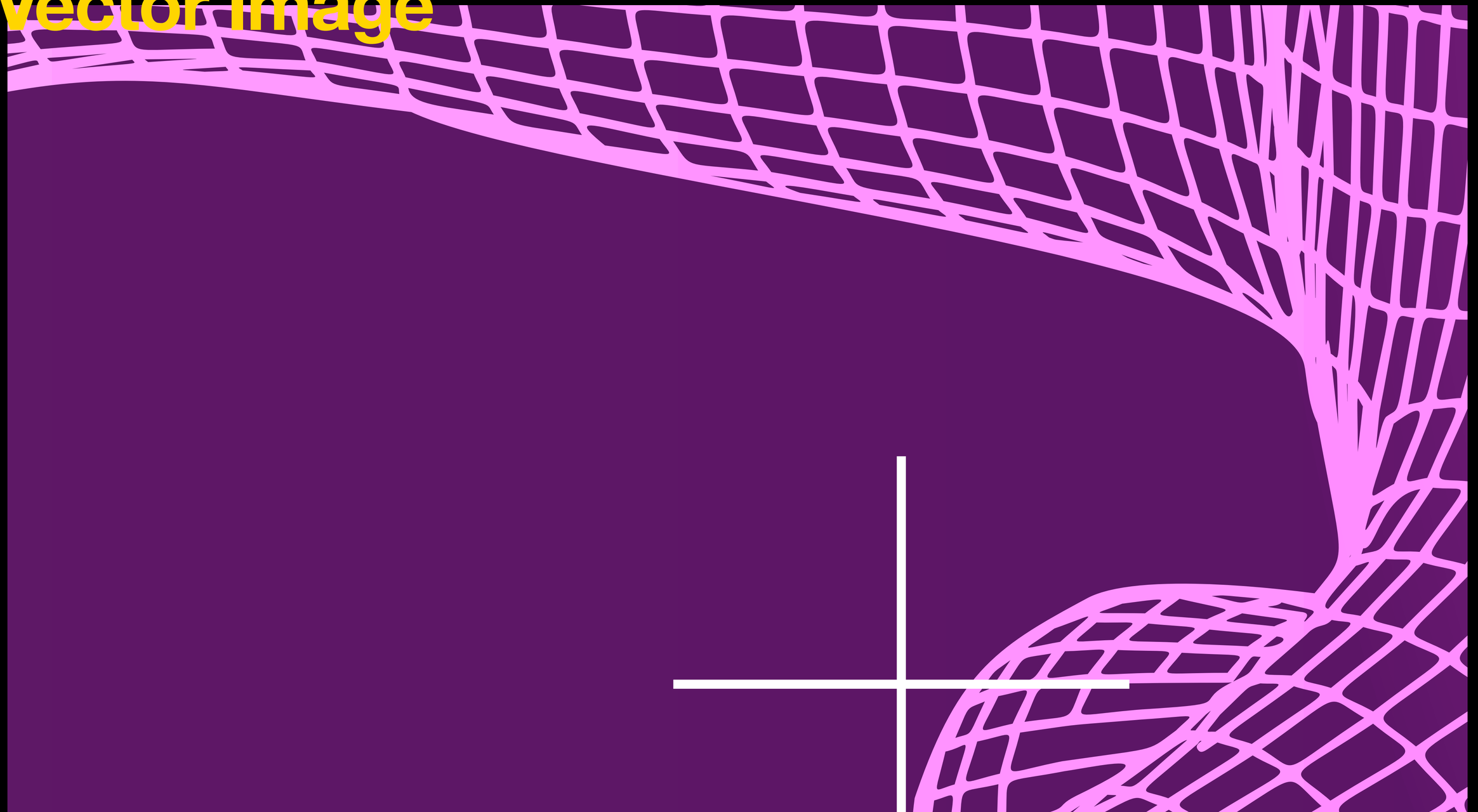
Many other spaces:
sRGB, Lab, YUV, HSV ..

Vector Image





Vector Image



Raster Image



Raster Image Example: Medical Imaging

- Imag
- Segr

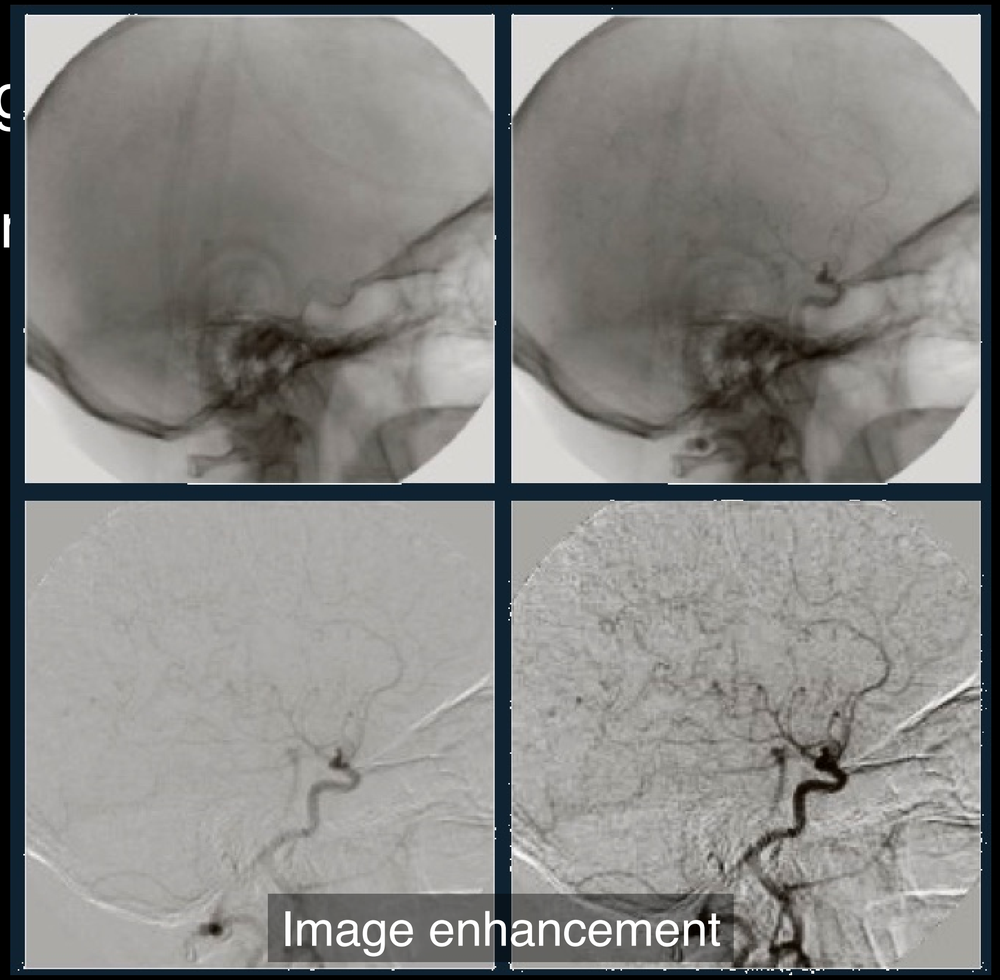
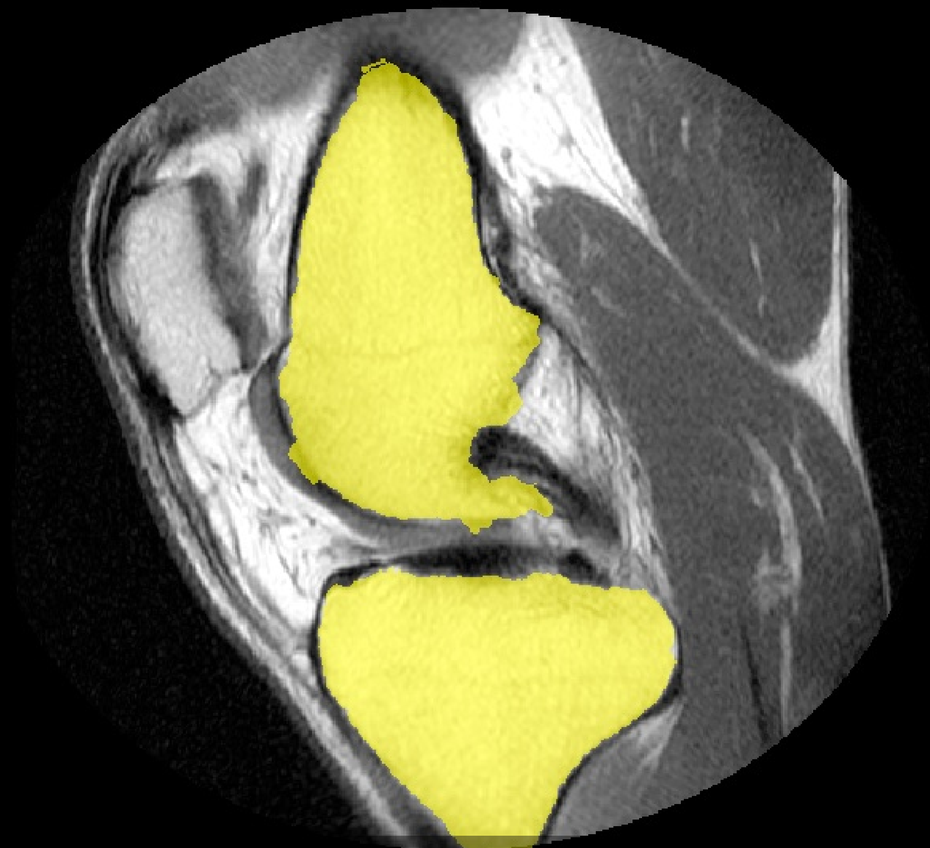


Image enhancement

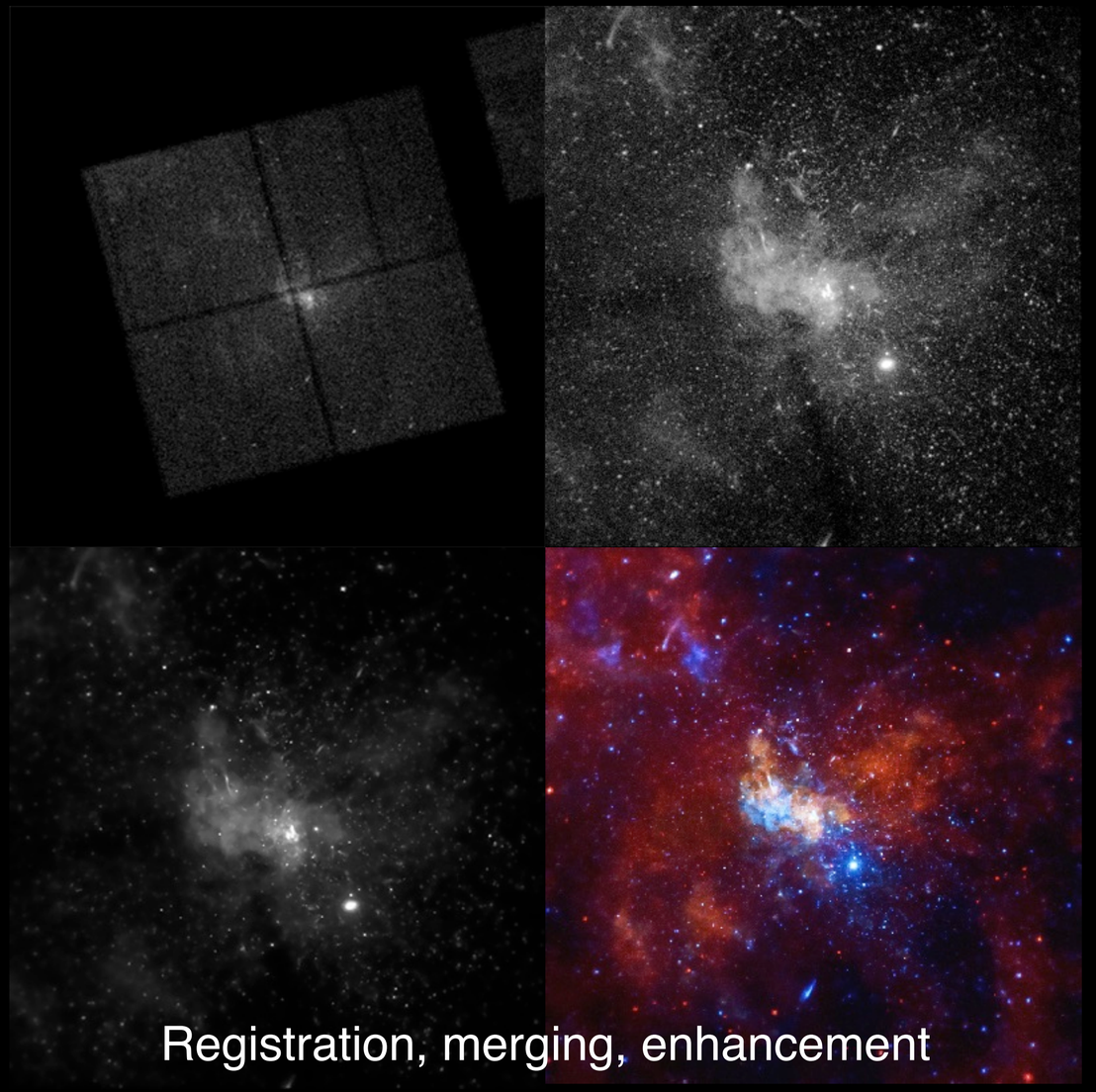


Segmentation

Astronomy



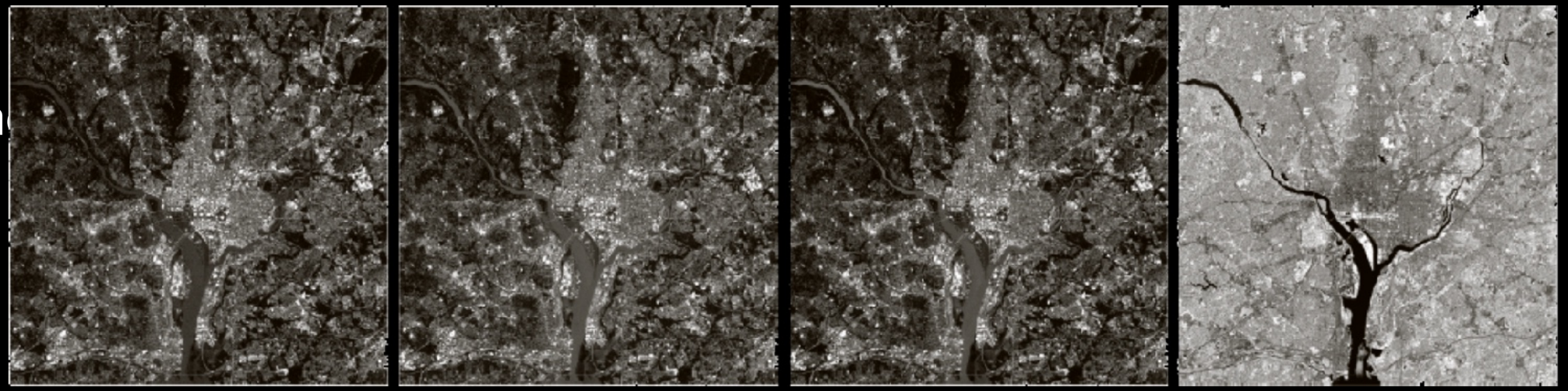
Deconvolution



Registration, merging, enhancement

Remote Sensing

- Rem





Land use classification

Object Detection and Recognition

- Fingerprints
- QR codes
- License plates
- Faces







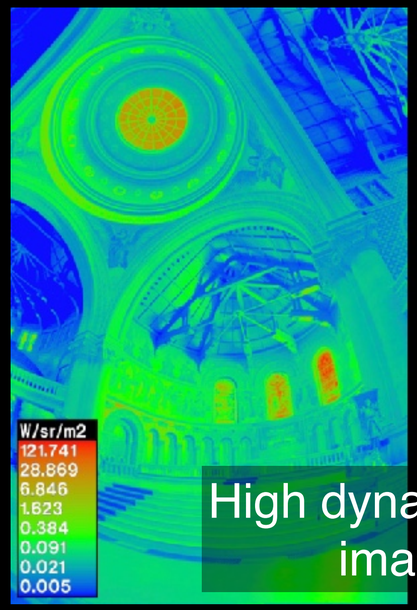


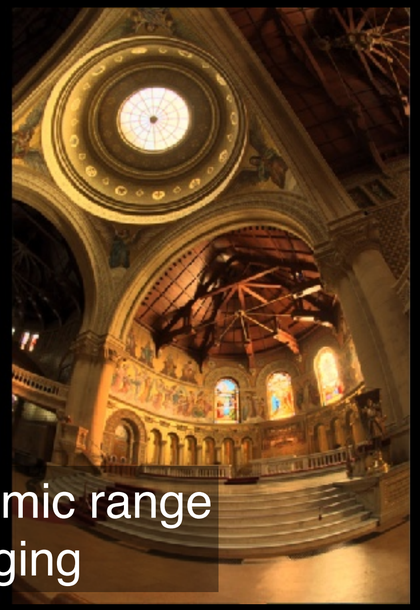
Computational Photography











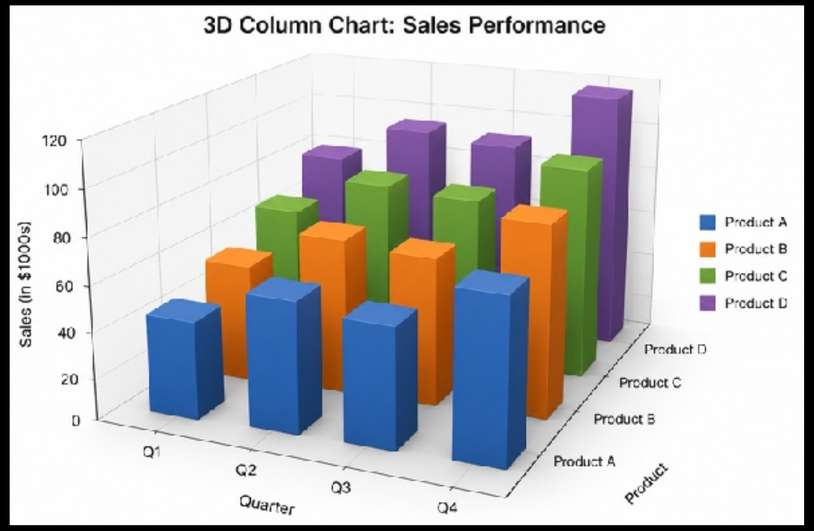
High dynamic range imaging

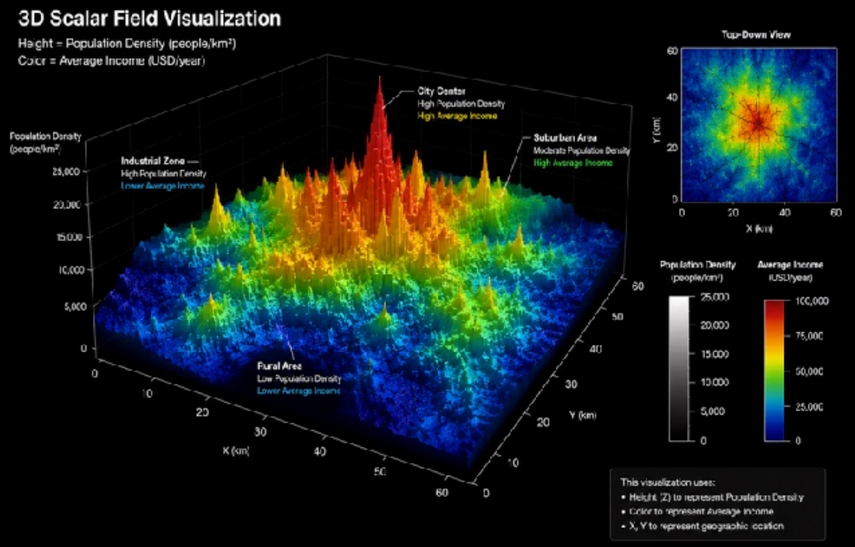
Data Visualization



Student	Math	Science	English	Total
Alice	85	90	88	263
Bob	78	82	75	235
Charlie	92	95	93	280
David	66	70	68	204
Eve	88	85	90	263

Widespread with 1D and 2D data



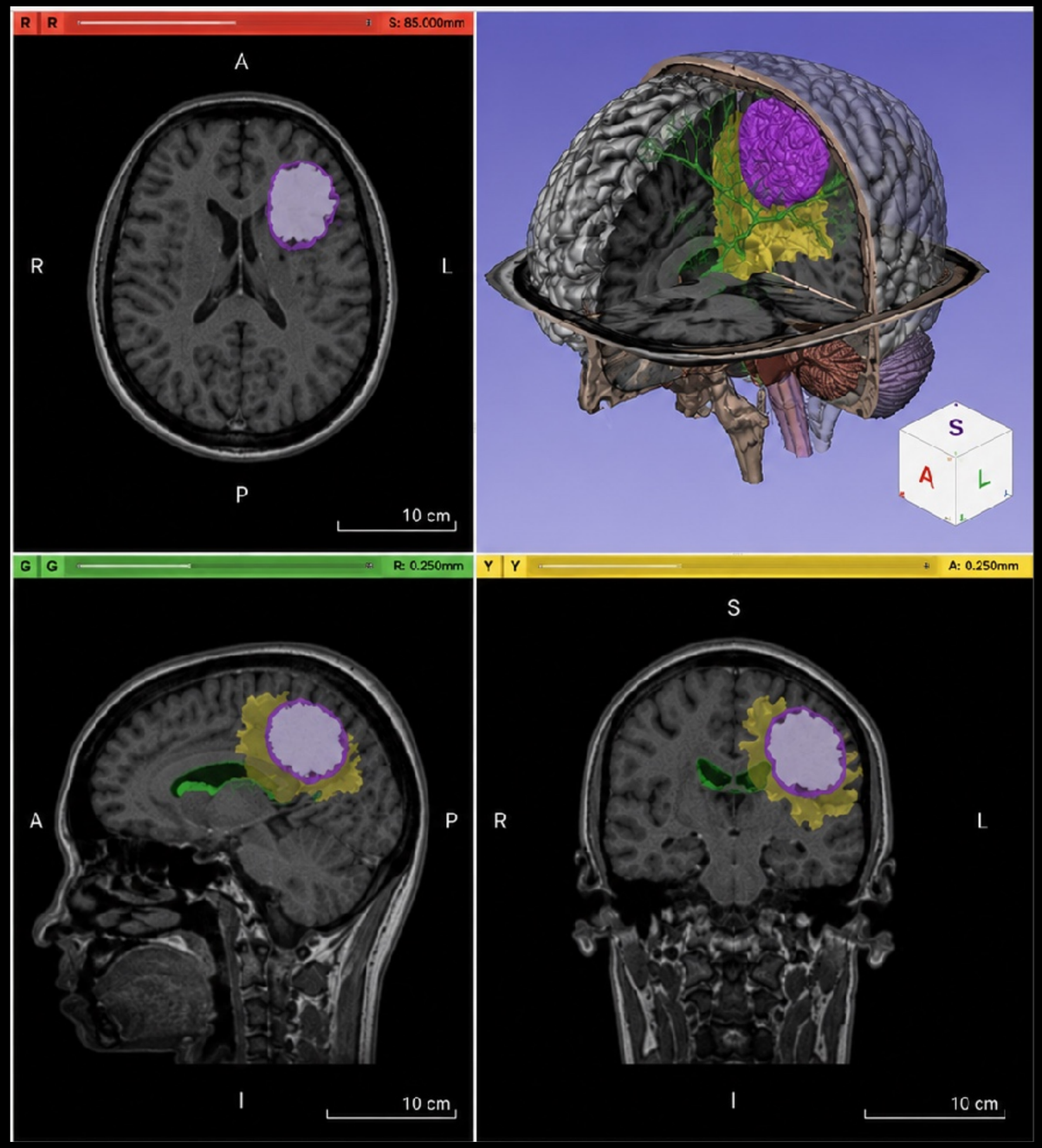


2.5/3.5D quite feasible

Data Visualization

Medical Volume

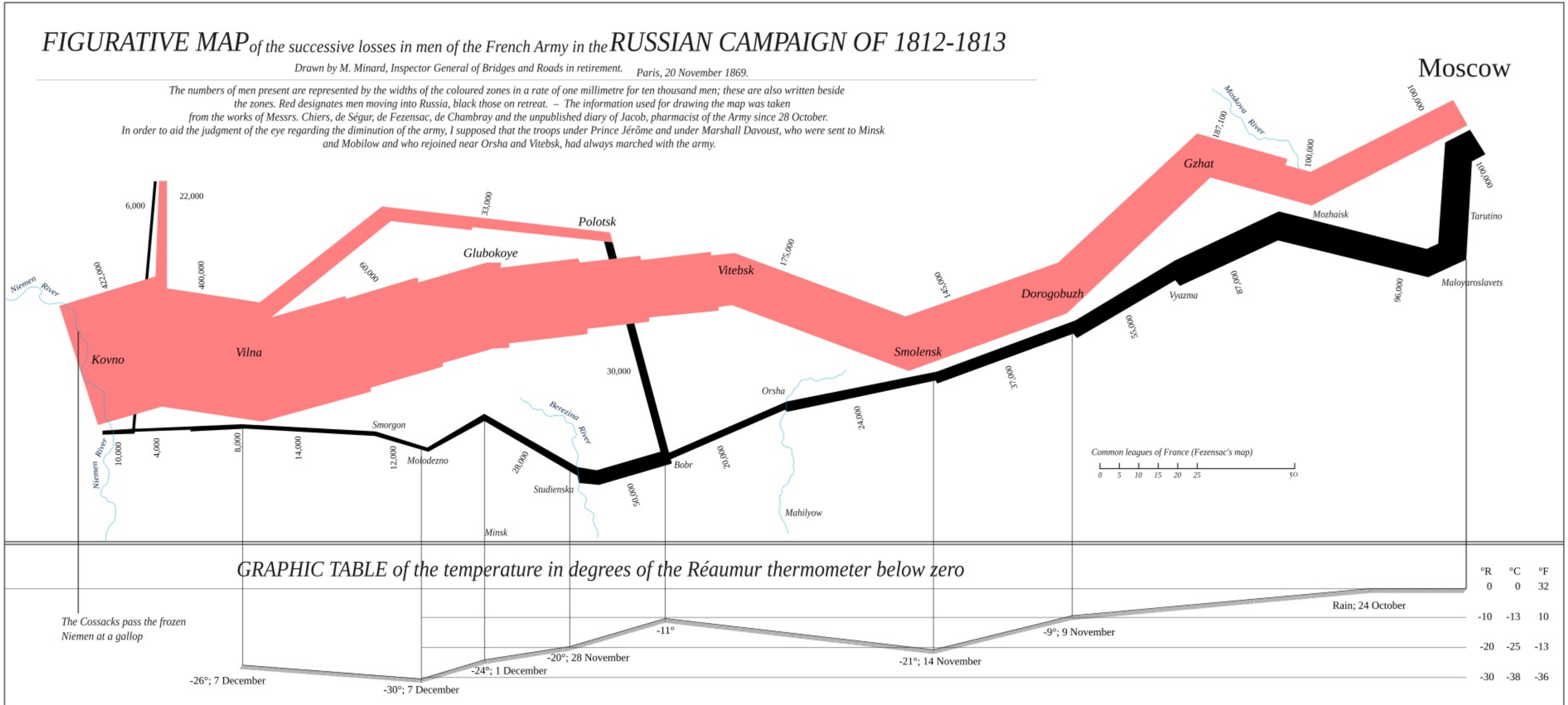
True 3D is somewhat less intuitive



Data Visualization

Many dimensions possible

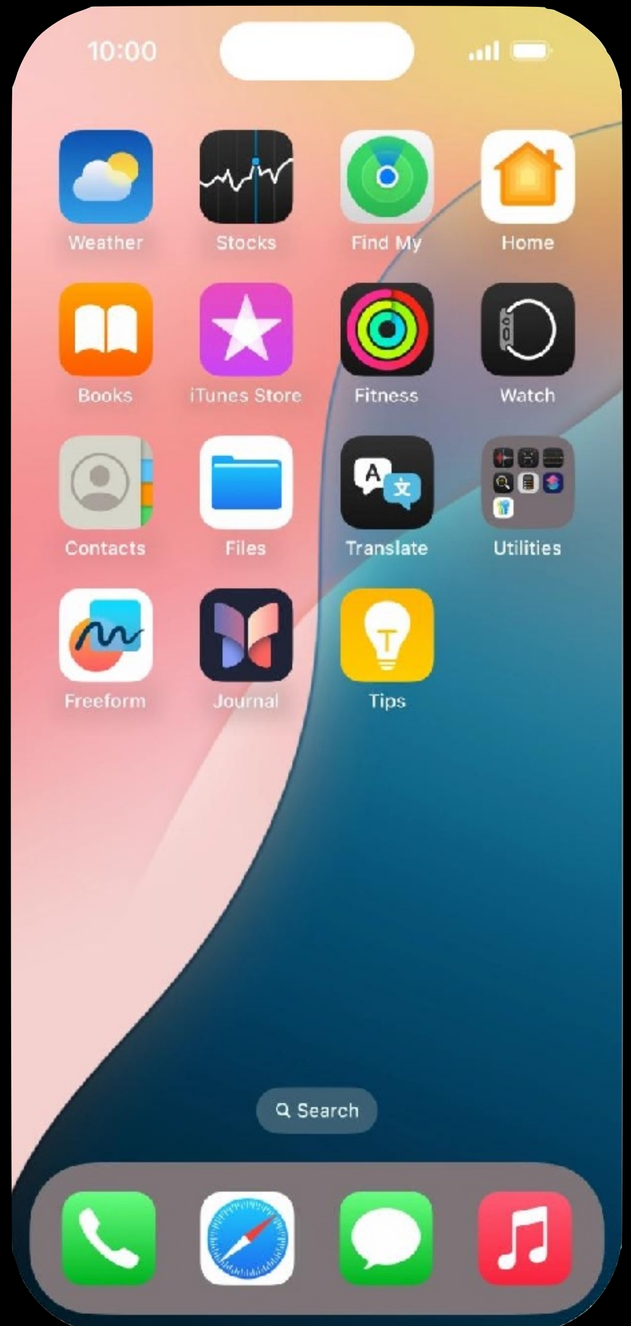
[Charles Minard]



Visual Interaction with Data

Touch Screen

Use of Icons (consistent and representative)
Clear visual hierarchy
Touch-screen with large touch-zones
Immediate feedback of user actions



Visual Interaction with Data

Virtual environment

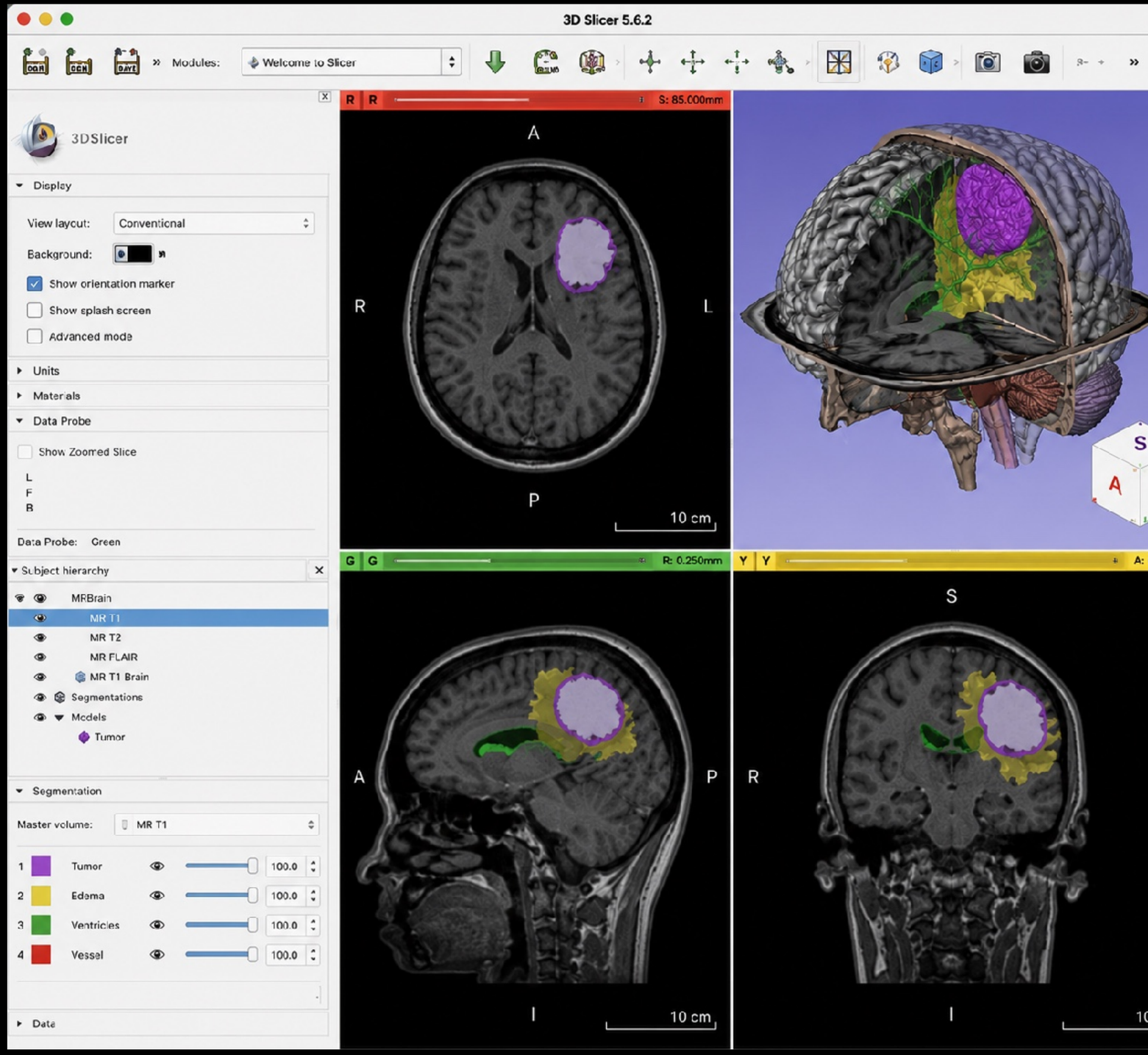
Large floating panels
Depth cues
Hand tracking
Gaze interaction
Minimal text
Natural spatial arrangement
Immediate feedback of user actions



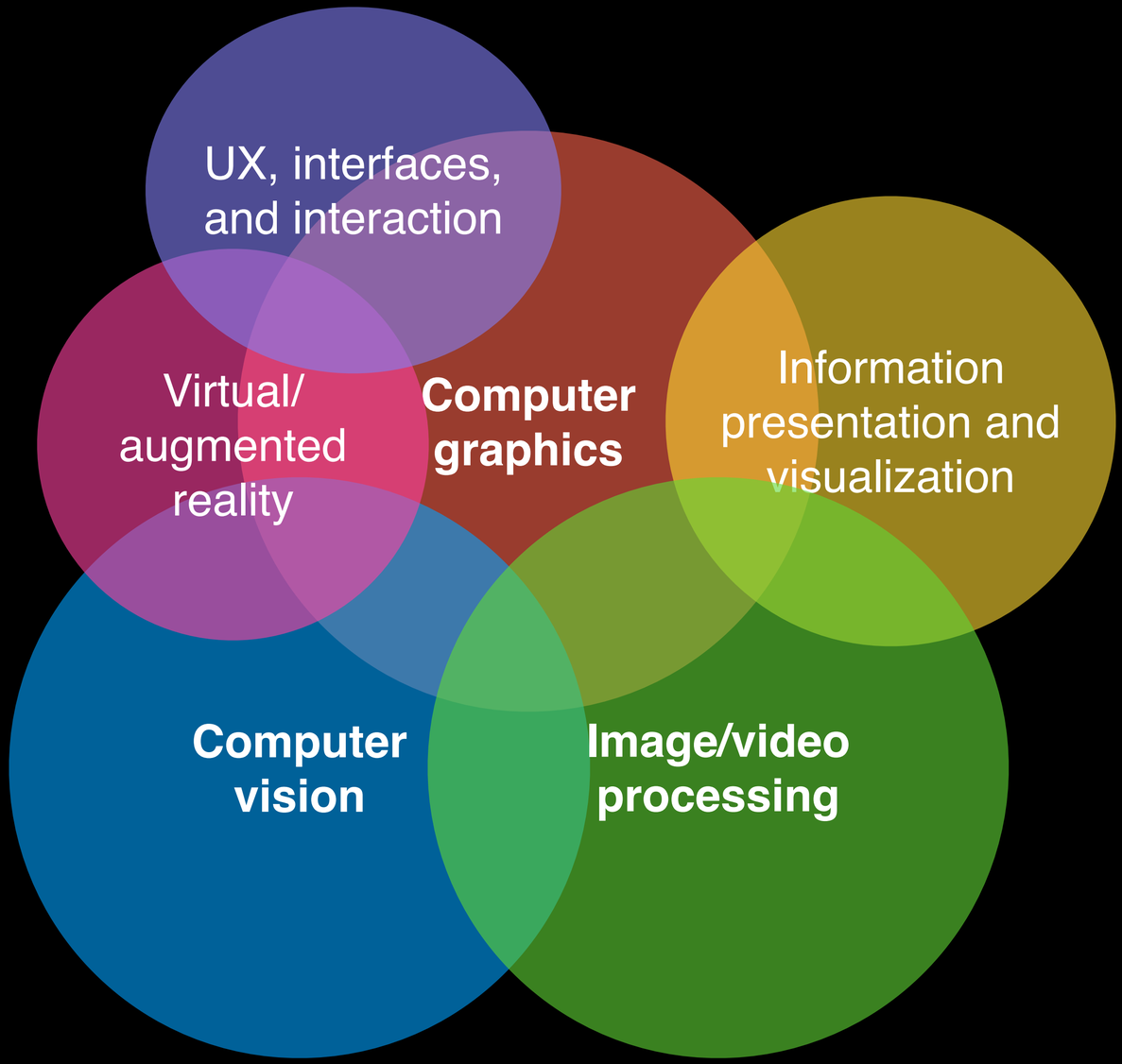
Visual Interaction

Mouse interfaces

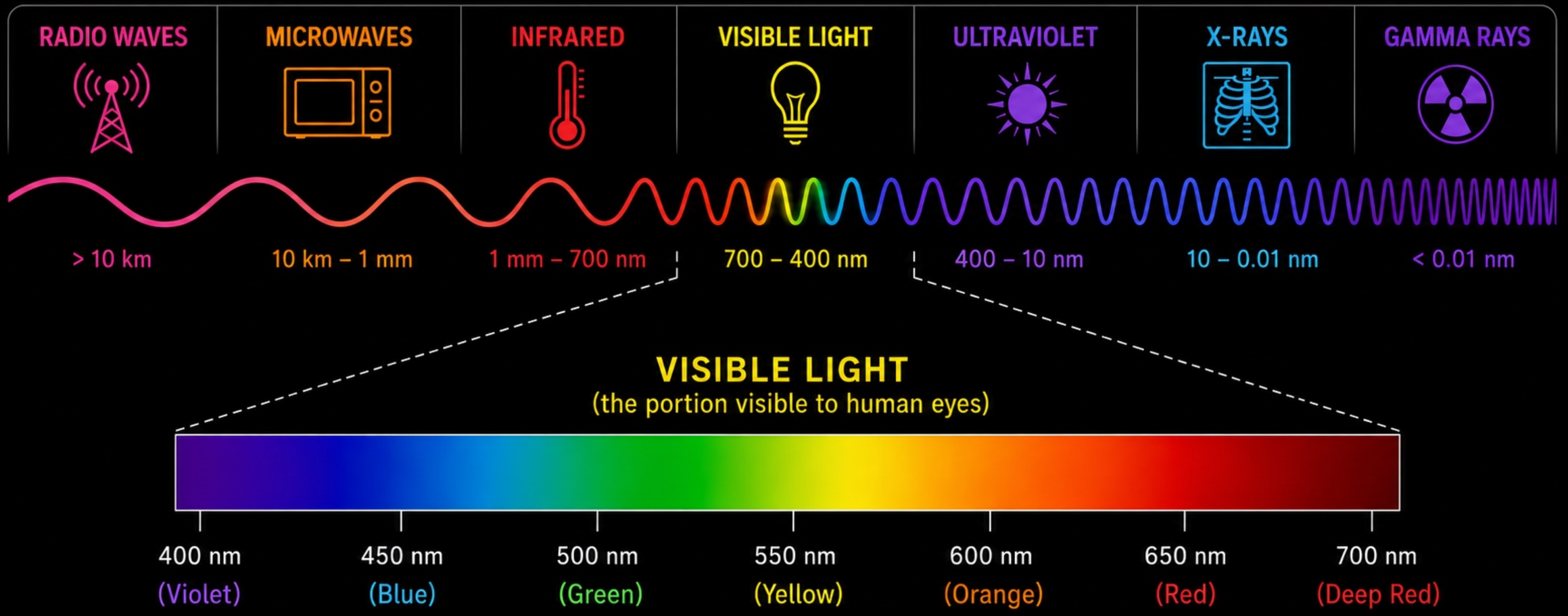
Minimal number of interactions
Recognizable labels and icons
… and navigation
Organized and grouped
Uncluttered
Clear fonts, visible color
Consistent style and appearance
Immediate feedback



Visual computing

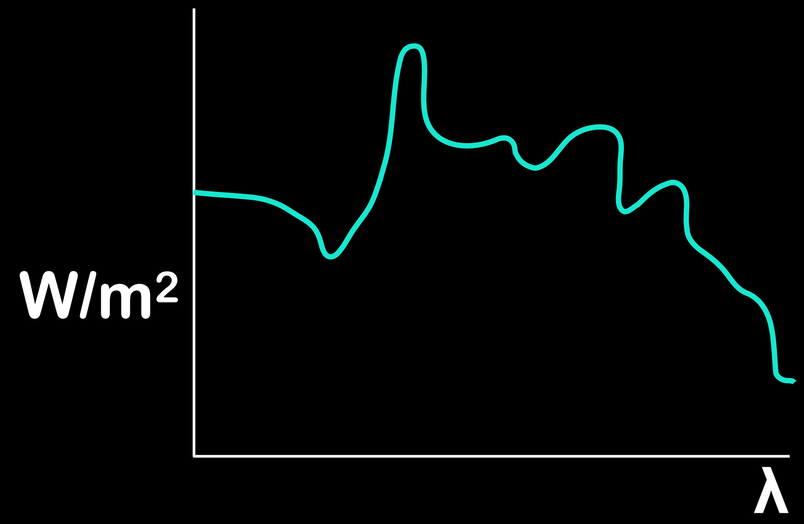


Light is EM Radiation



Sensing Light

- Eye senses light distribution
  - But light != R-G-B (or any other such model)
- Electromagnetic energy (waves or photons)
  - $h\nu == hc/\lambda$ (Joules)
  - c can vary with medium, λ can shift
- Eyes only four types of sensor cells (rods, SML cones)
- Equal *W/m²* with different λ do no appear equally bright



Luminous Efficiency Function