



Computer Vision & Cognitive Systems

SCQ5109806 - LM CS,DS,CYB,PD

Visual Intelligence and Machine Perception

Prof. Lamberto Ballan

Cognitive Computing

- Applied Machine Learning (e.g. Deep Learning)
- Natural Language Processing
- Speech Recognition
- **Computer Vision**

*learn from &
make sense of*
• Big Data

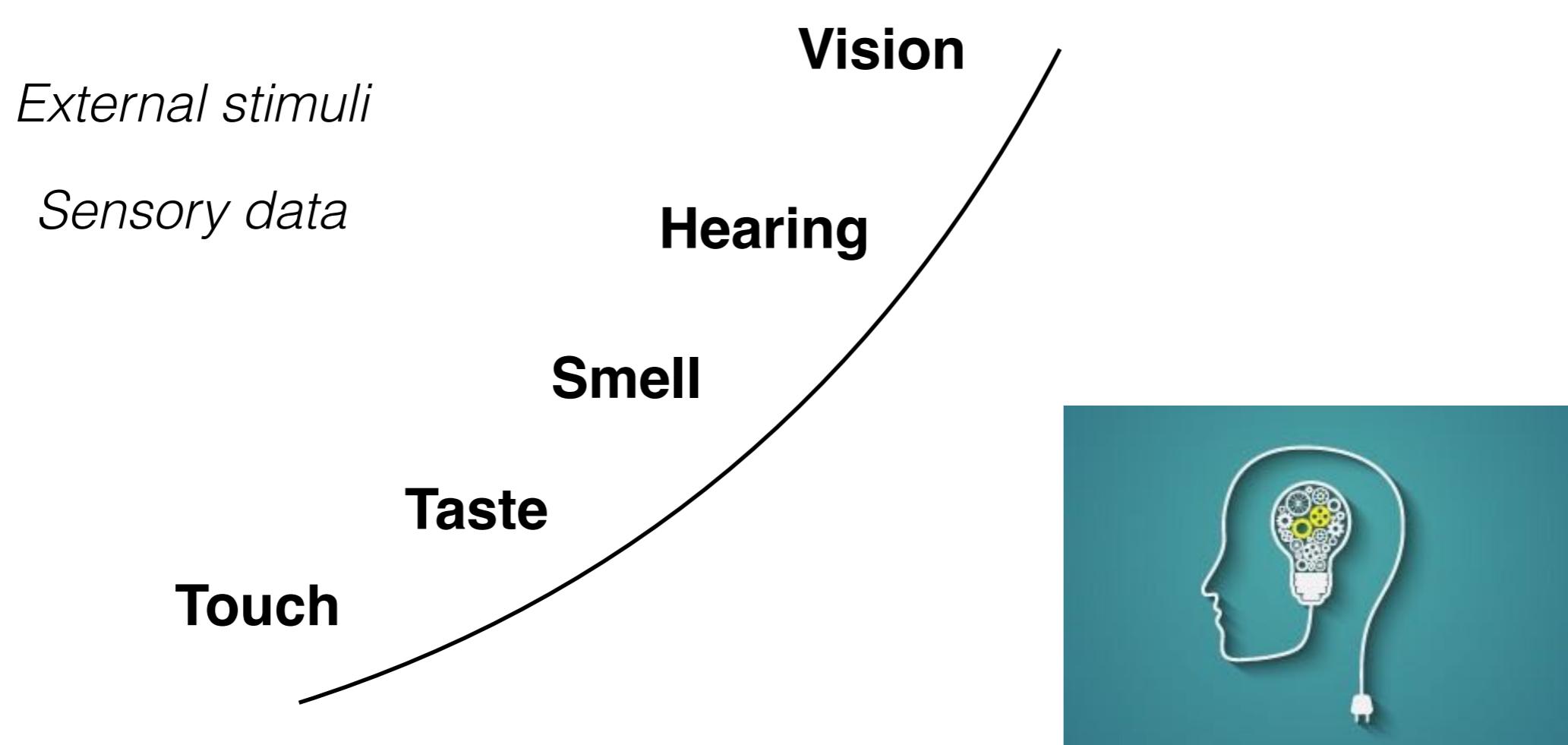


(A)Intelligent
Applications

Data-driven
decisions

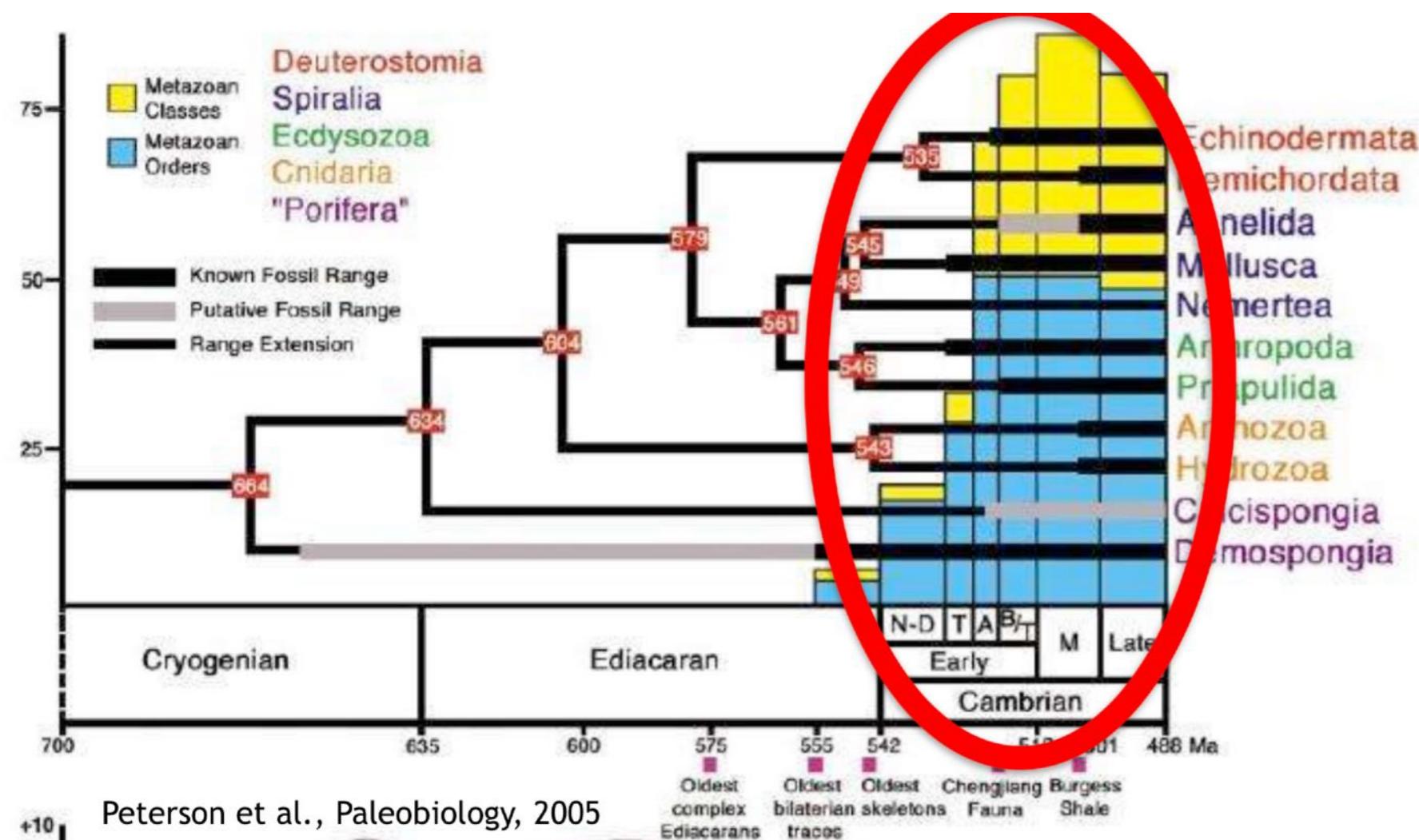
(Machine) Perception

- Perception is the ability to capture, process, and make sense of the information that our senses receive



Why is vision so important?

Evolution's Big Bang: the
“Cambrian Explosion”



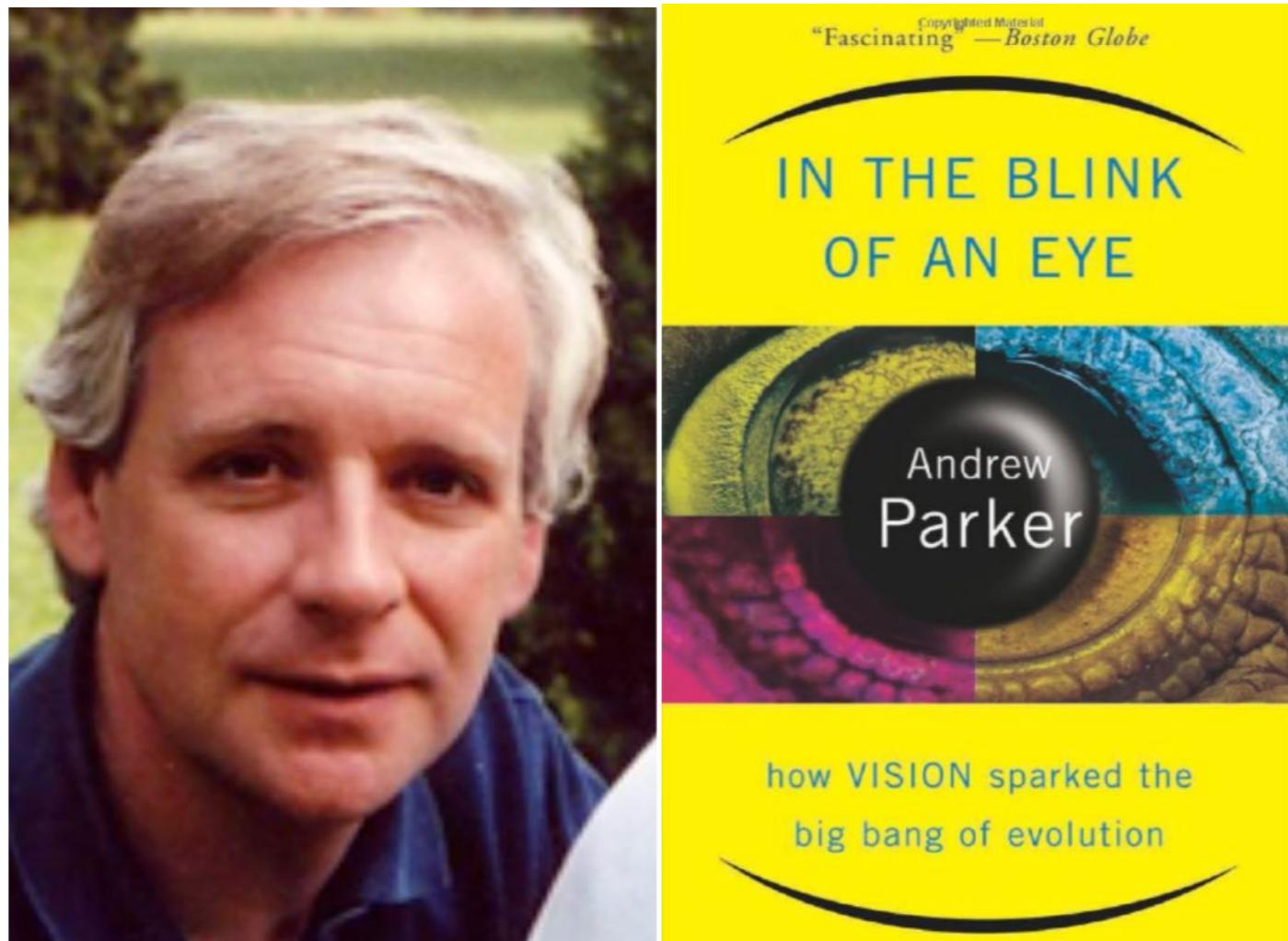
approx. 543M years B.C.



Why is vision so important?

- “The Cambrian Explosion is triggered by the sudden evolution of vision”, which set off an evolutionary arms race where animals either evolved or died.

Andrew Parker,
zoologist



Why is vision so important?

Visual Intelligence is
the cornerstone of (Artificial) Intelligence

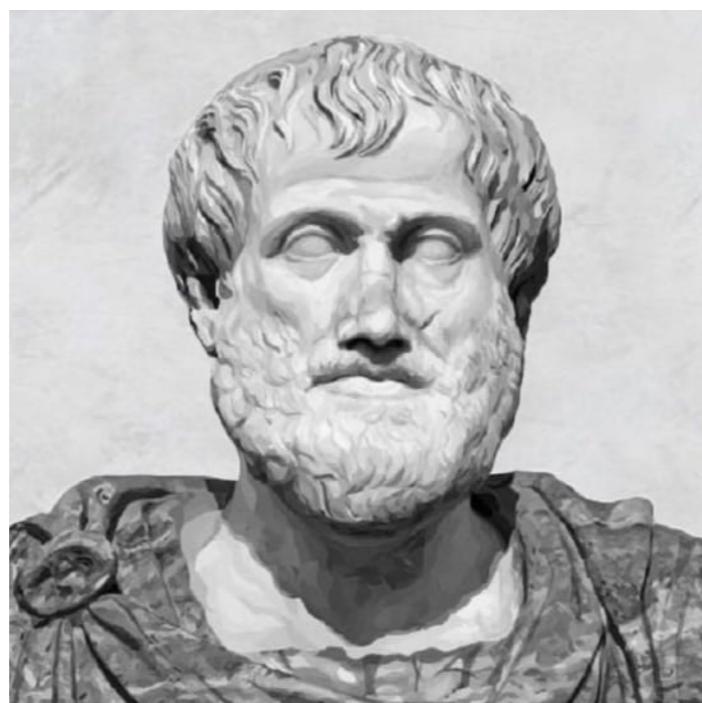
What is (computer) vision?

(in other words)

“What does it mean, to see?”

**“To know what is
where by looking.”**

Aristotle,
~300 BC



VISION

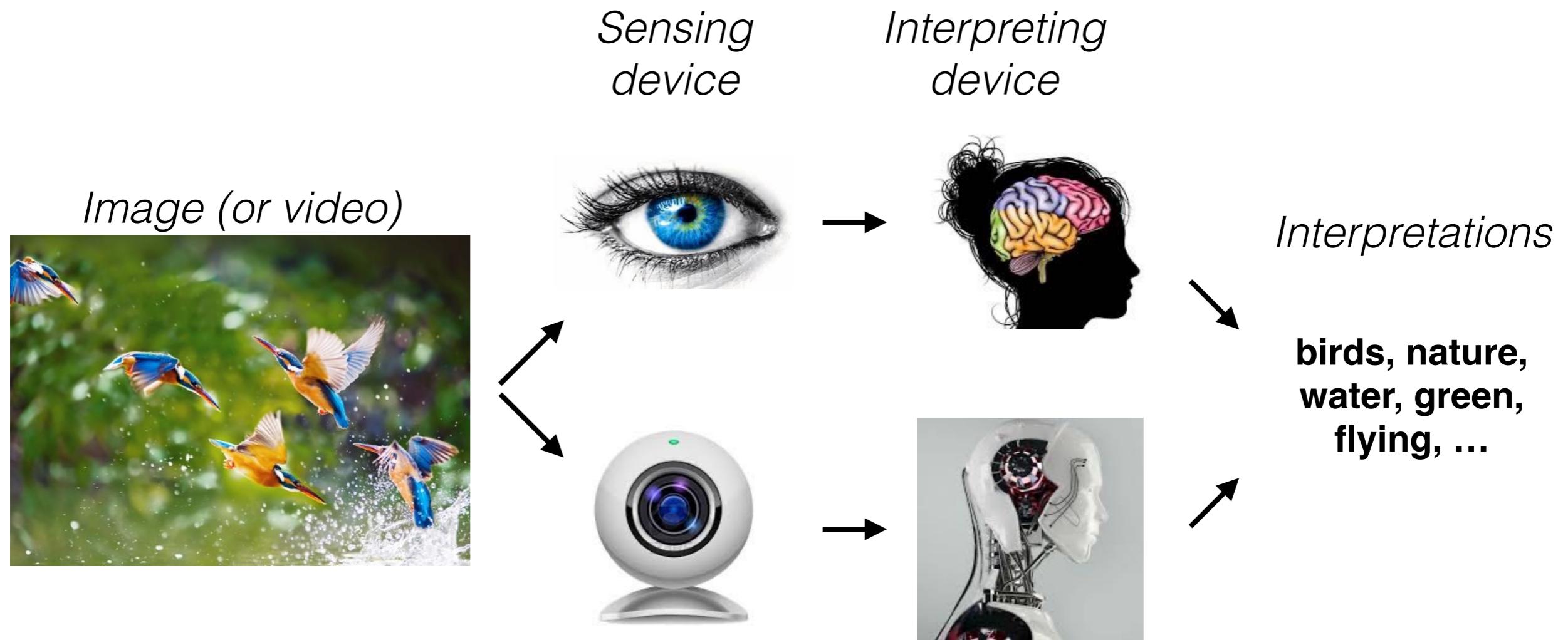


David Marr

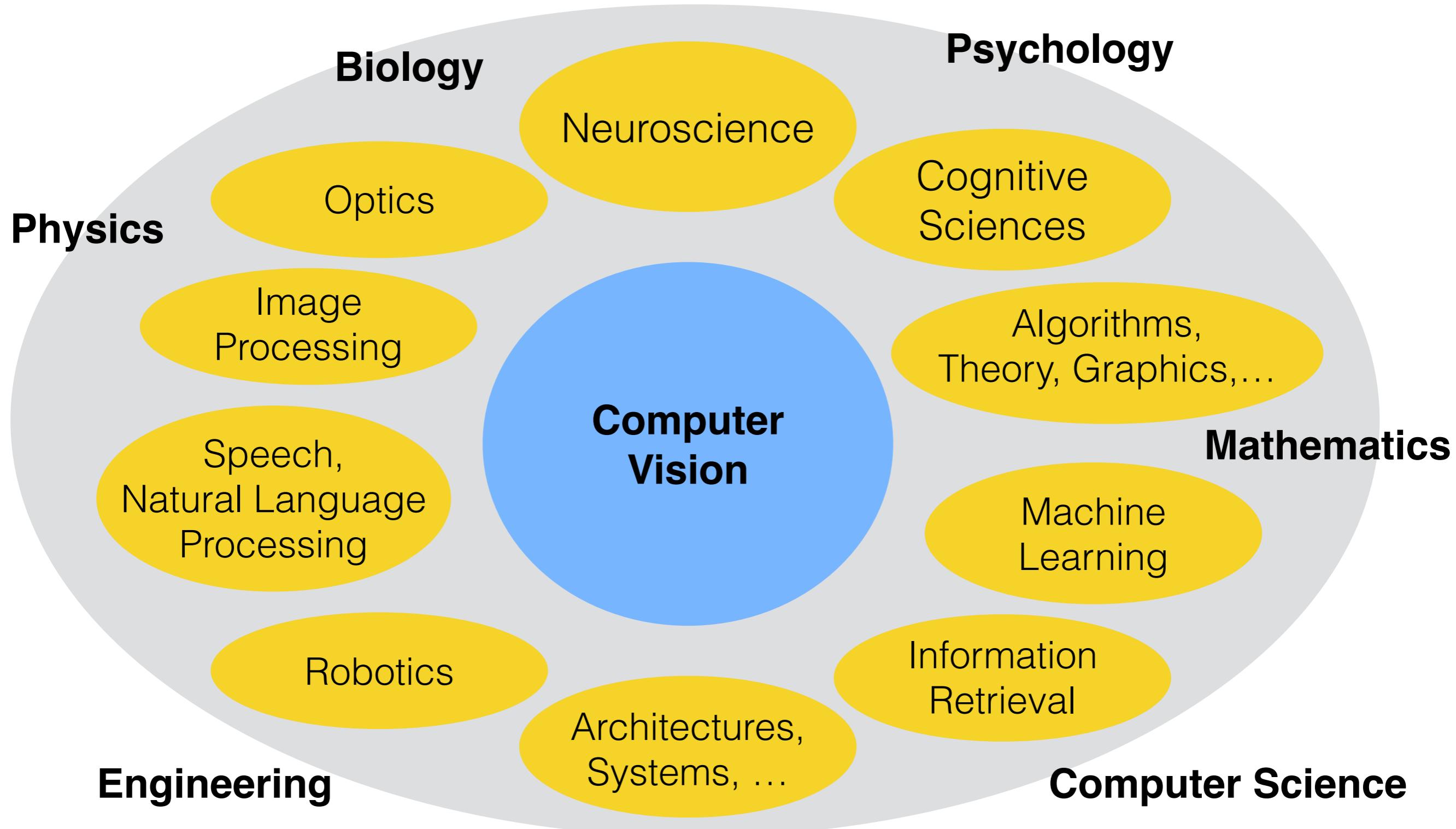
FOREWORD BY
Shimon Ullman
AFTERWORD BY
Tomaso Poggio

What is (computer) vision?

- Visual recognition is the *north star* of computer vision

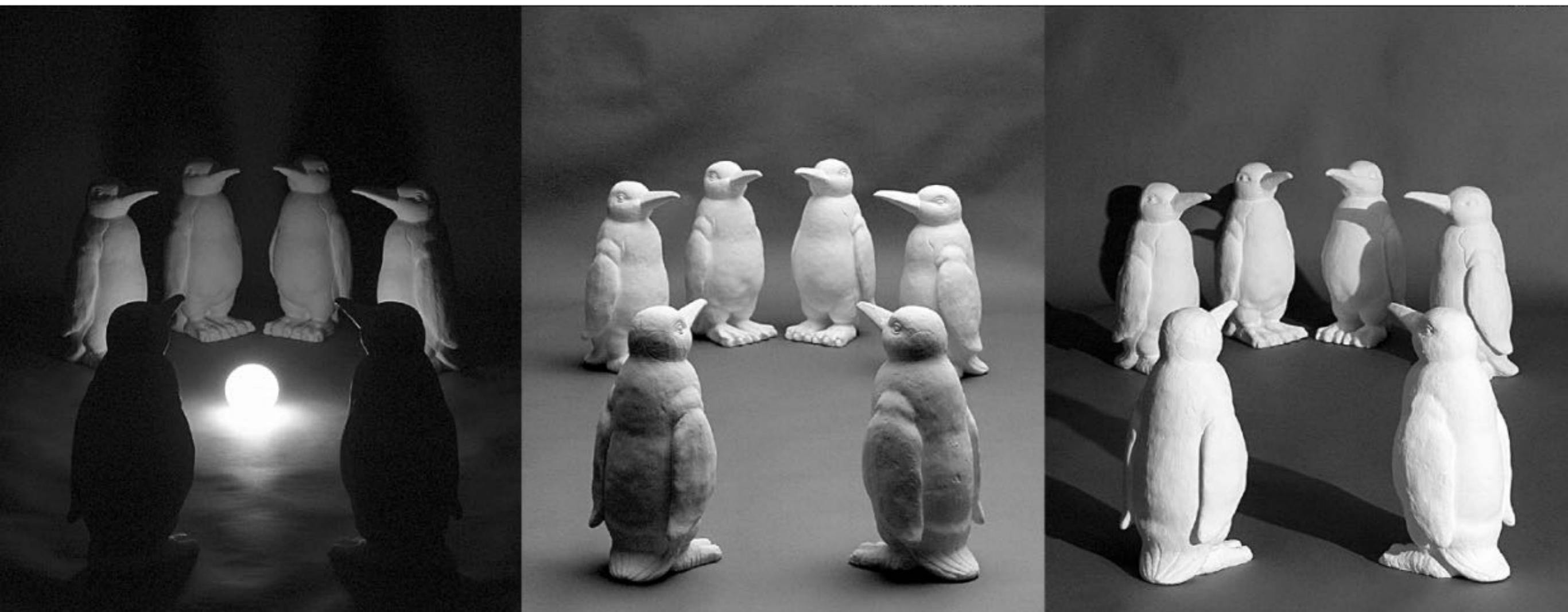


What is computer vision related to?



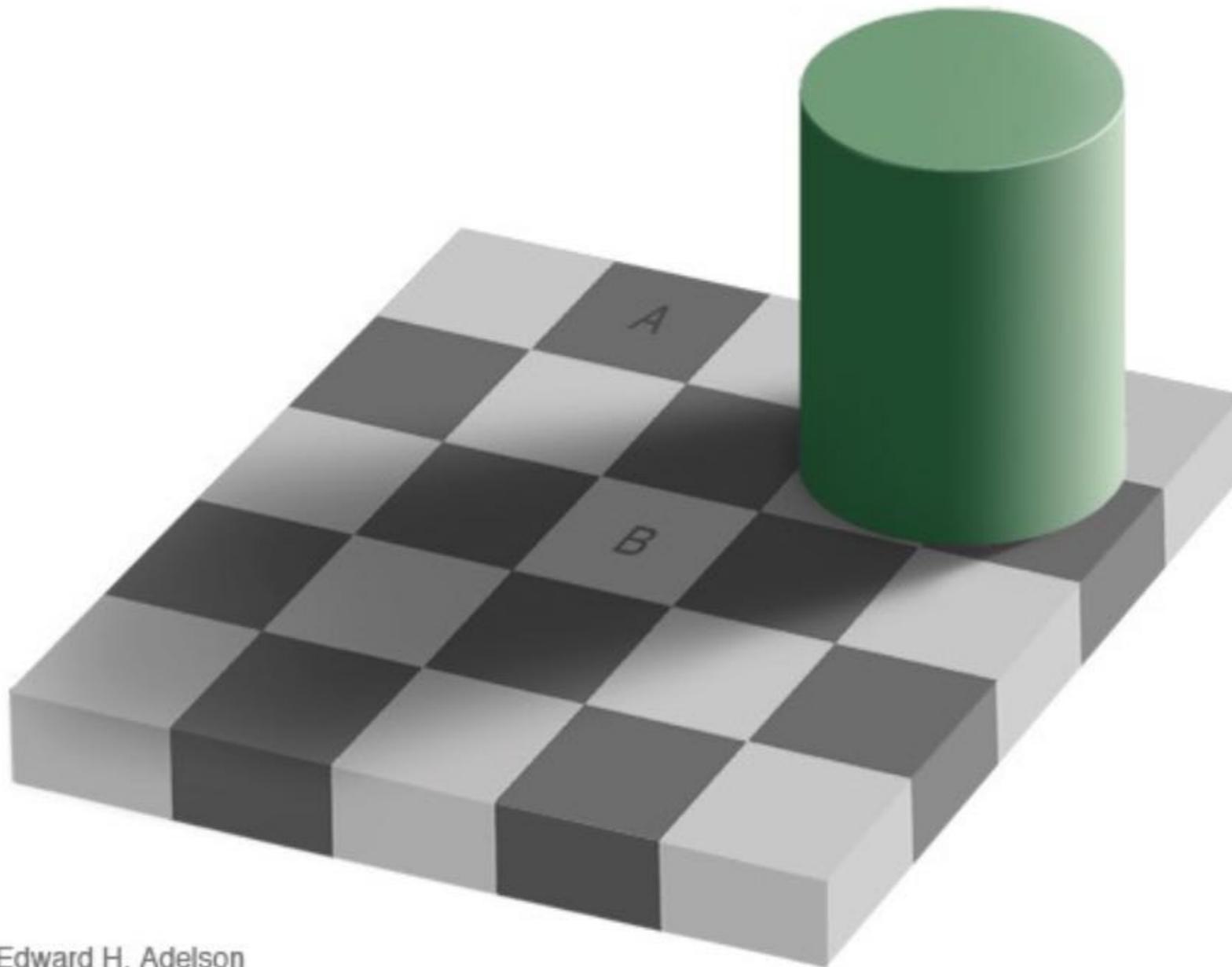
Visual perception: seeing is challenging

- Challenge: illumination



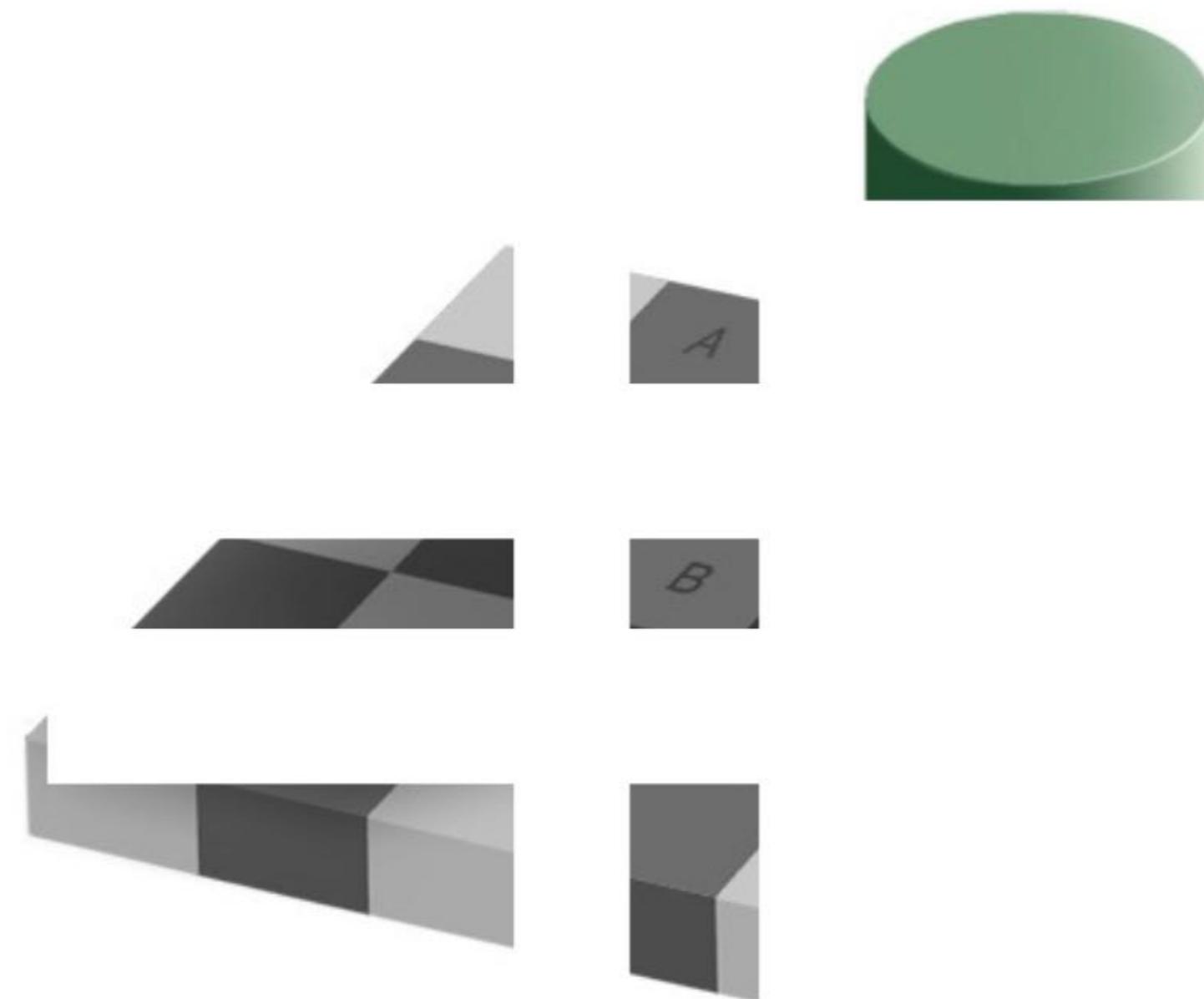
Seeing is challenging

- Challenge: light and shadow



Seeing is challenging

- Challenge: light and shadow



Edward H. Adelson

Seeing is challenging

- Challenge: illumination

WIRED

The Science of Why No One Agrees on the Color of This Dress

SL

ADAM ROGERS SCIENCE 02.26.15 10:28 PM

THE SCIENCE OF WHY NO ONE AGREES ON THE COLOR OF THIS DRESS



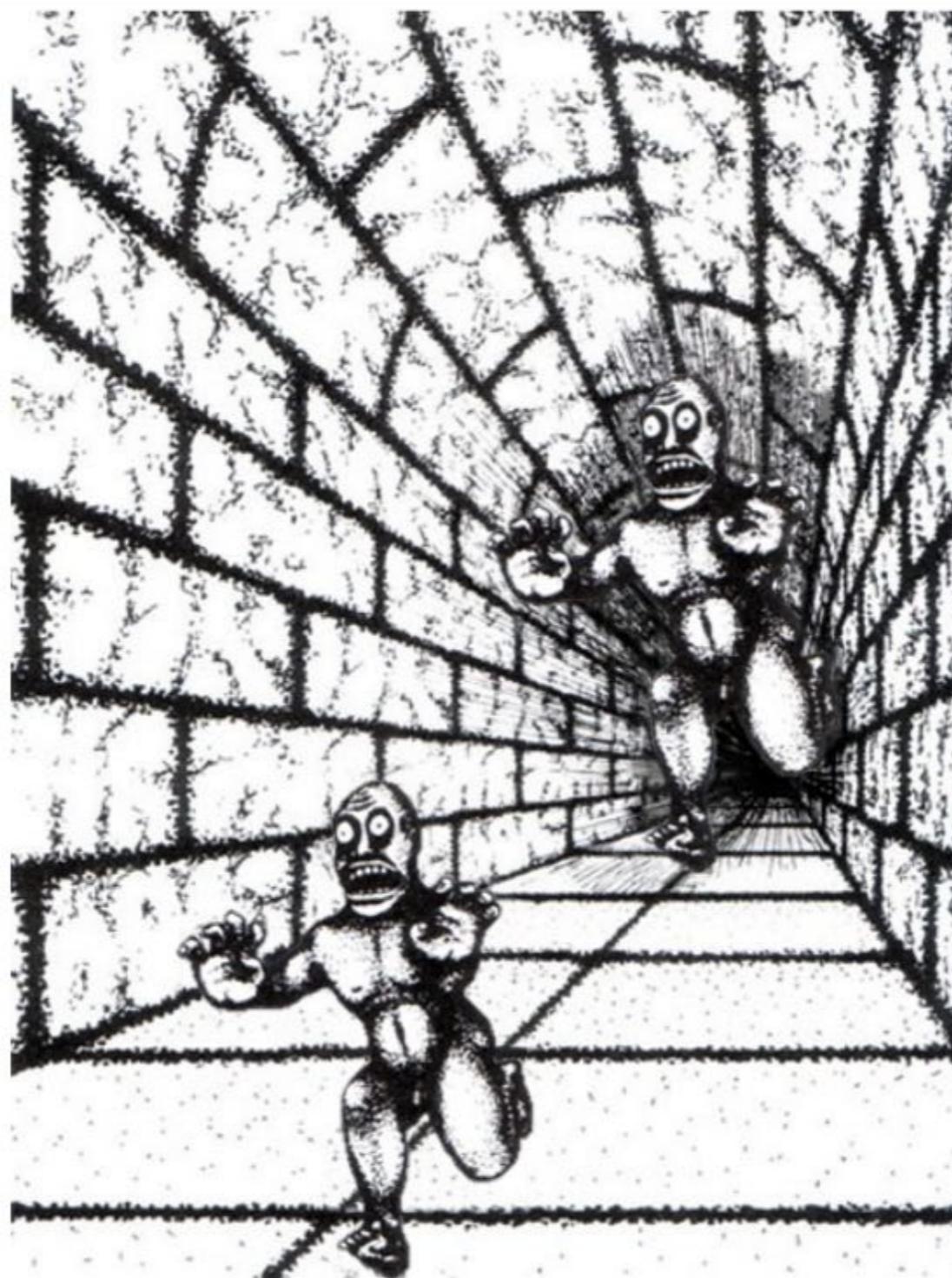
Seeing is challenging

- Challenge: scale



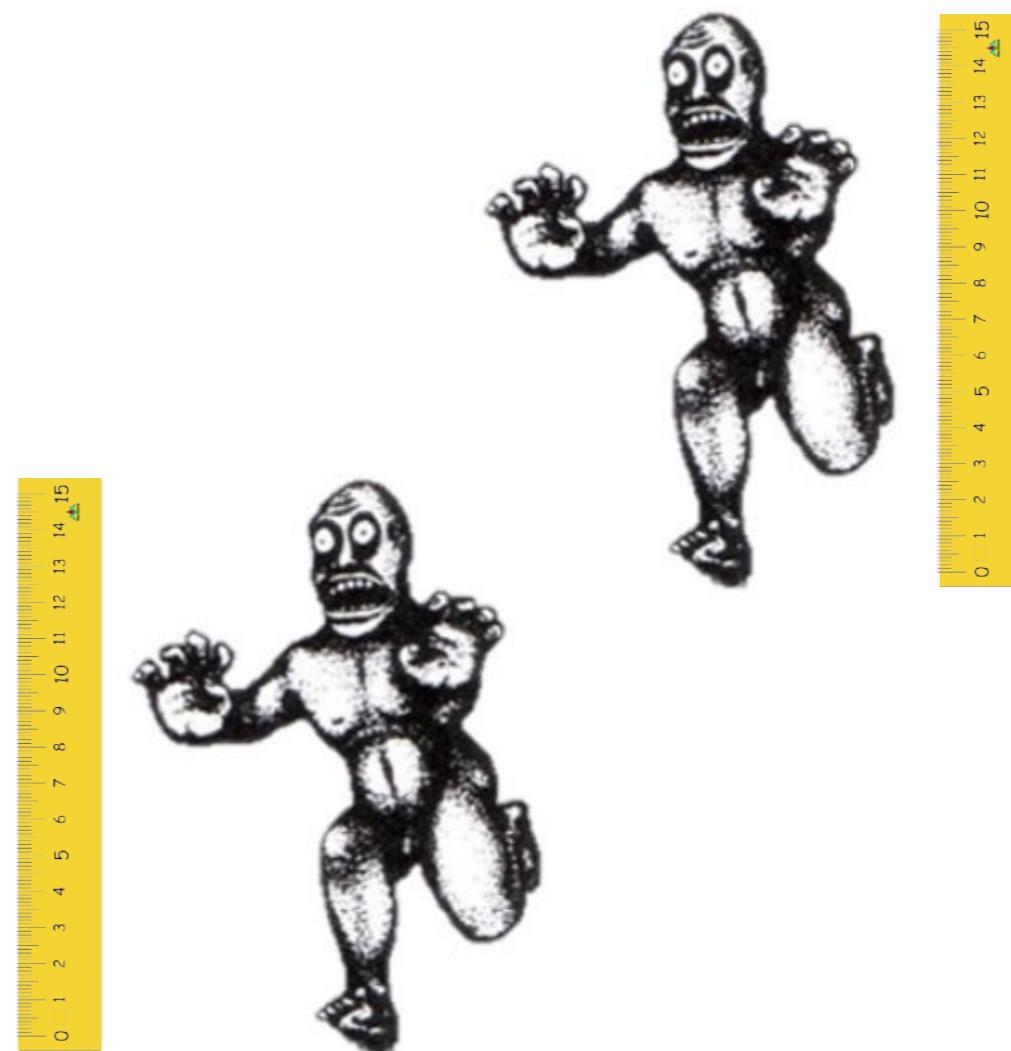
Seeing is challenging

- Challenge: scale and perspective view



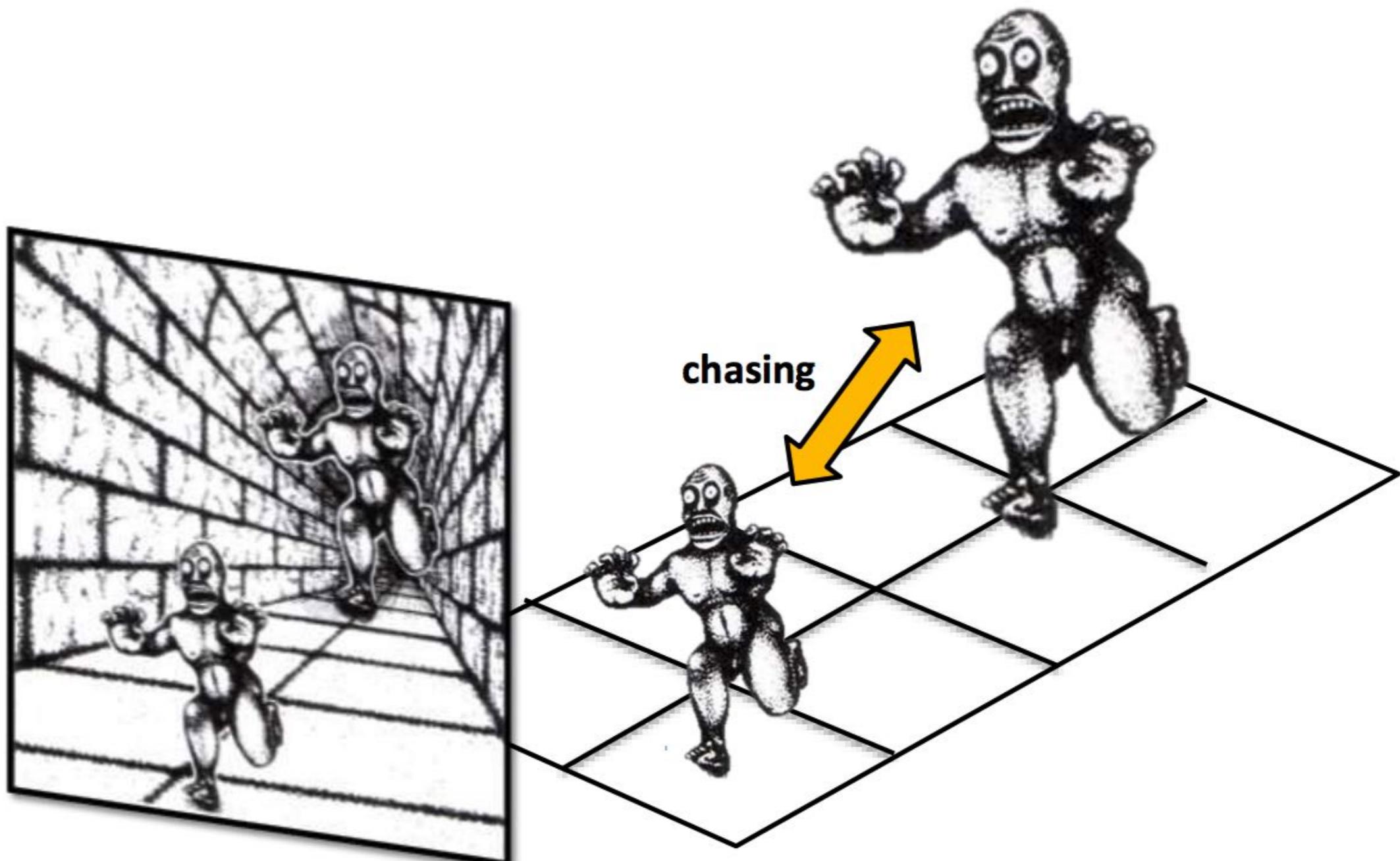
Seeing is challenging

- Challenge: scale and perspective view



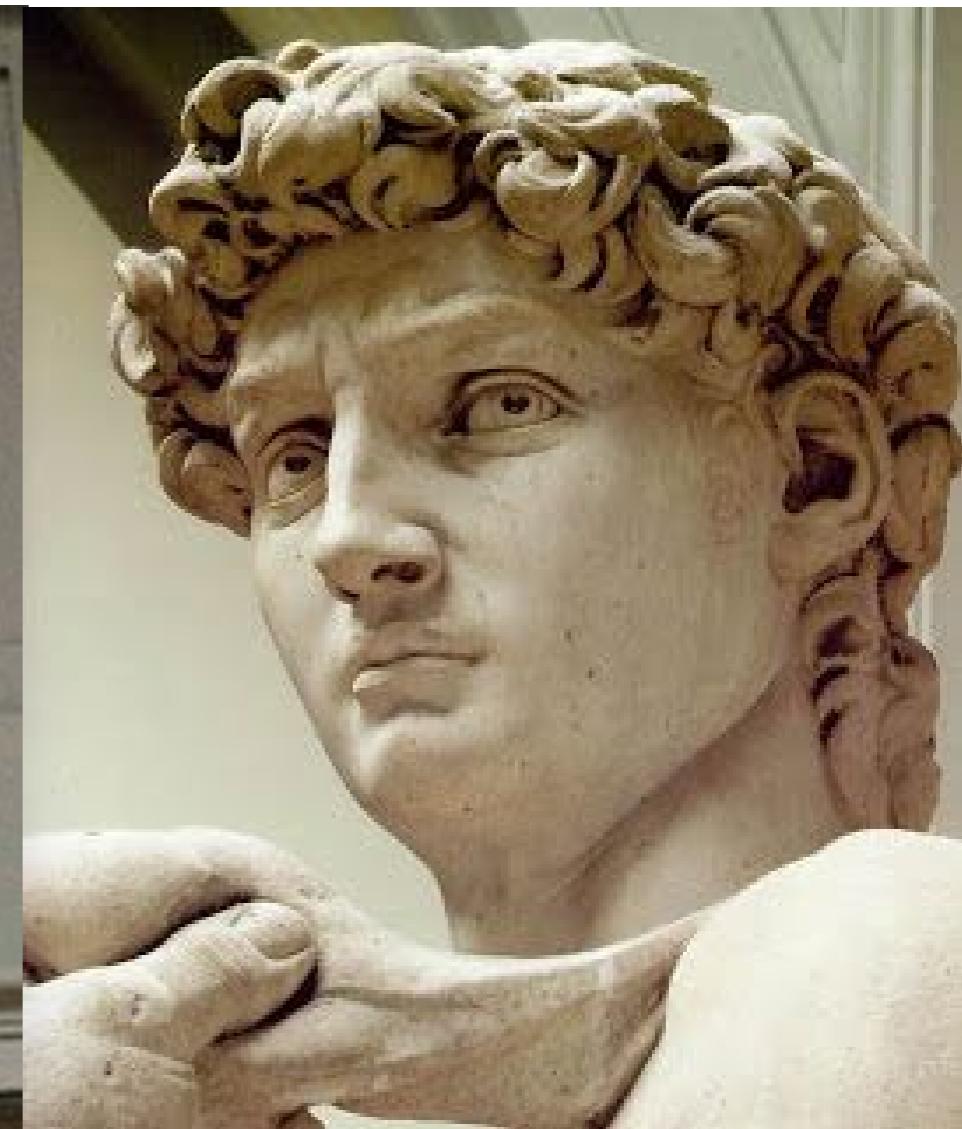
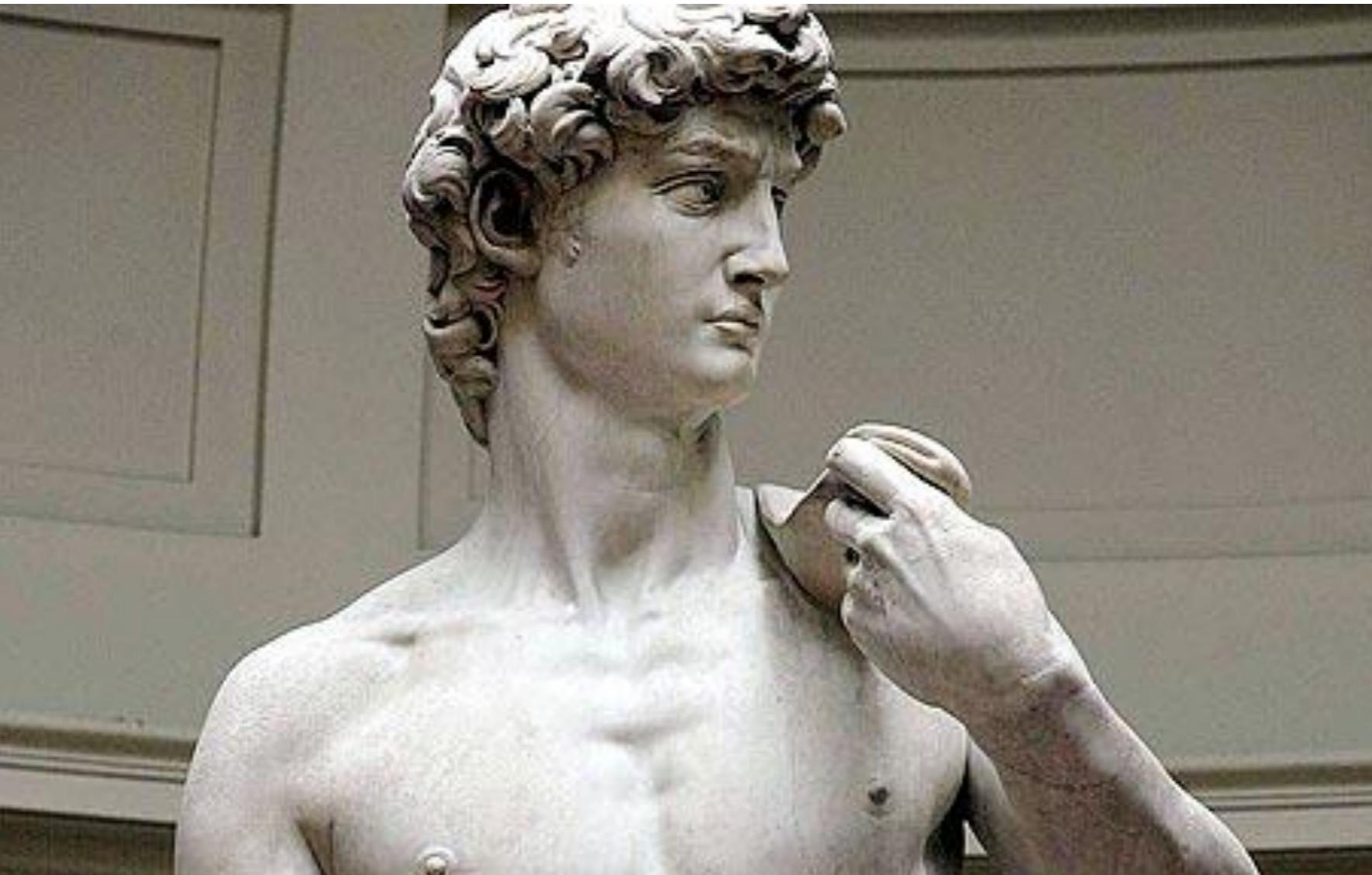
Seeing is challenging

- Challenge: scale and perspective view



Seeing is challenging

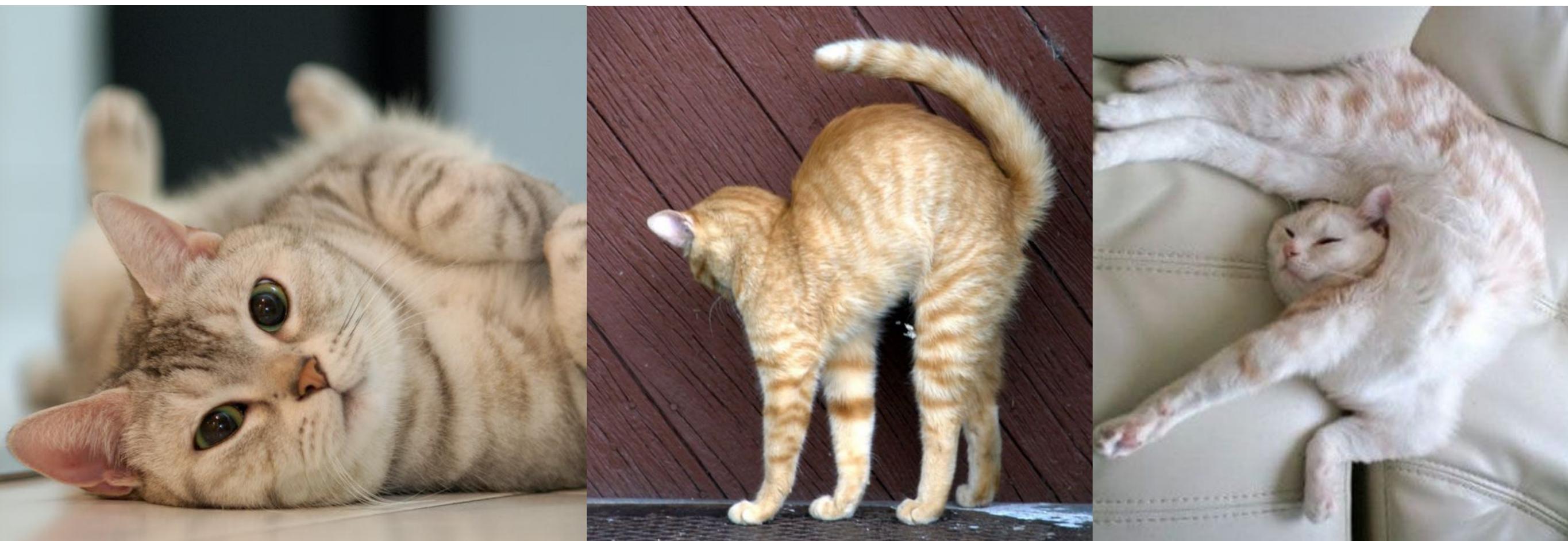
- Challenge: viewpoint



Michelangelo's David

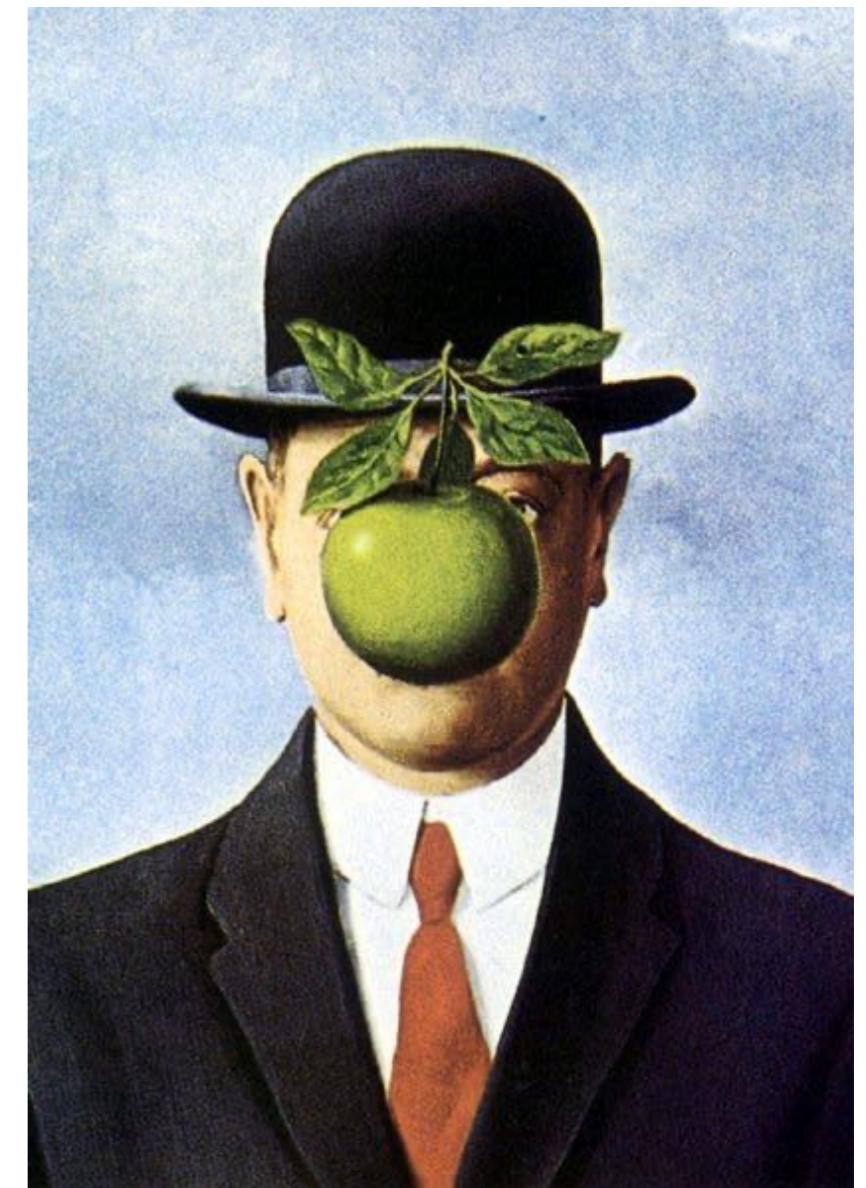
Seeing is challenging

- Challenge: deformation



Seeing is challenging

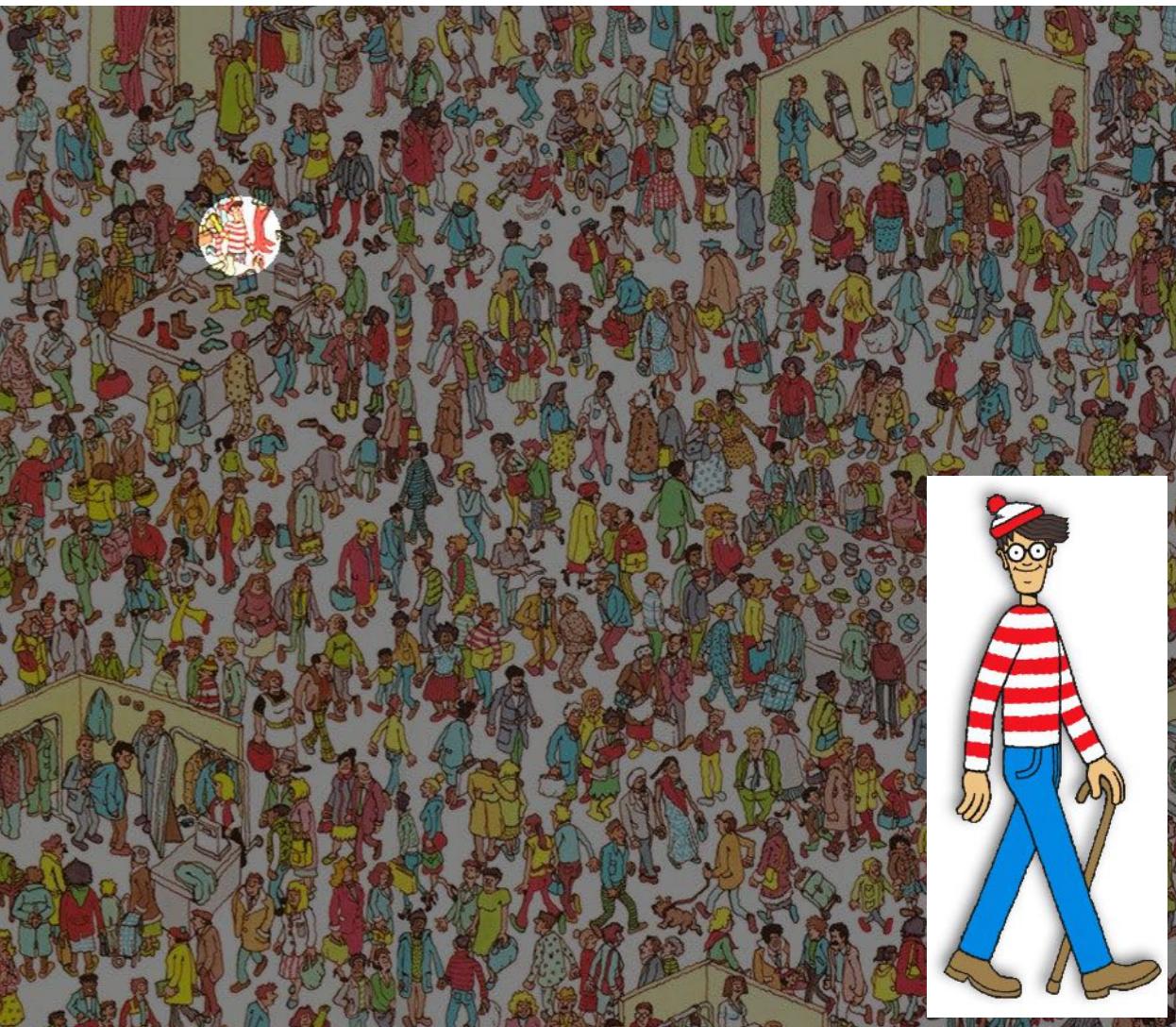
- Challenge: occlusion



Magritte's "The Son of Man"

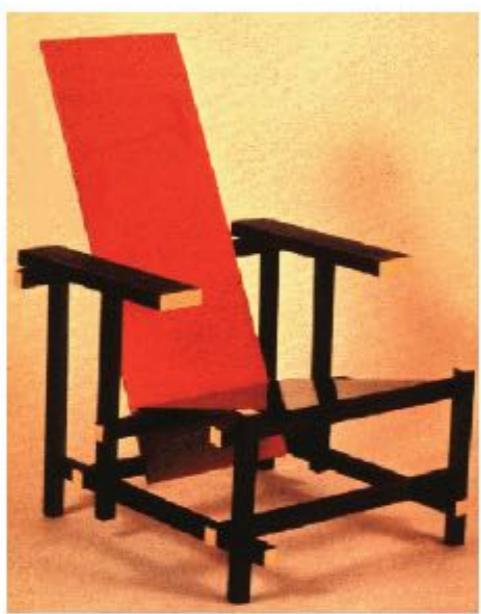
Seeing is challenging

- Challenge: clutter



Seeing is challenging

- Challenge (categorization): intra-class variation



Seeing is challenging

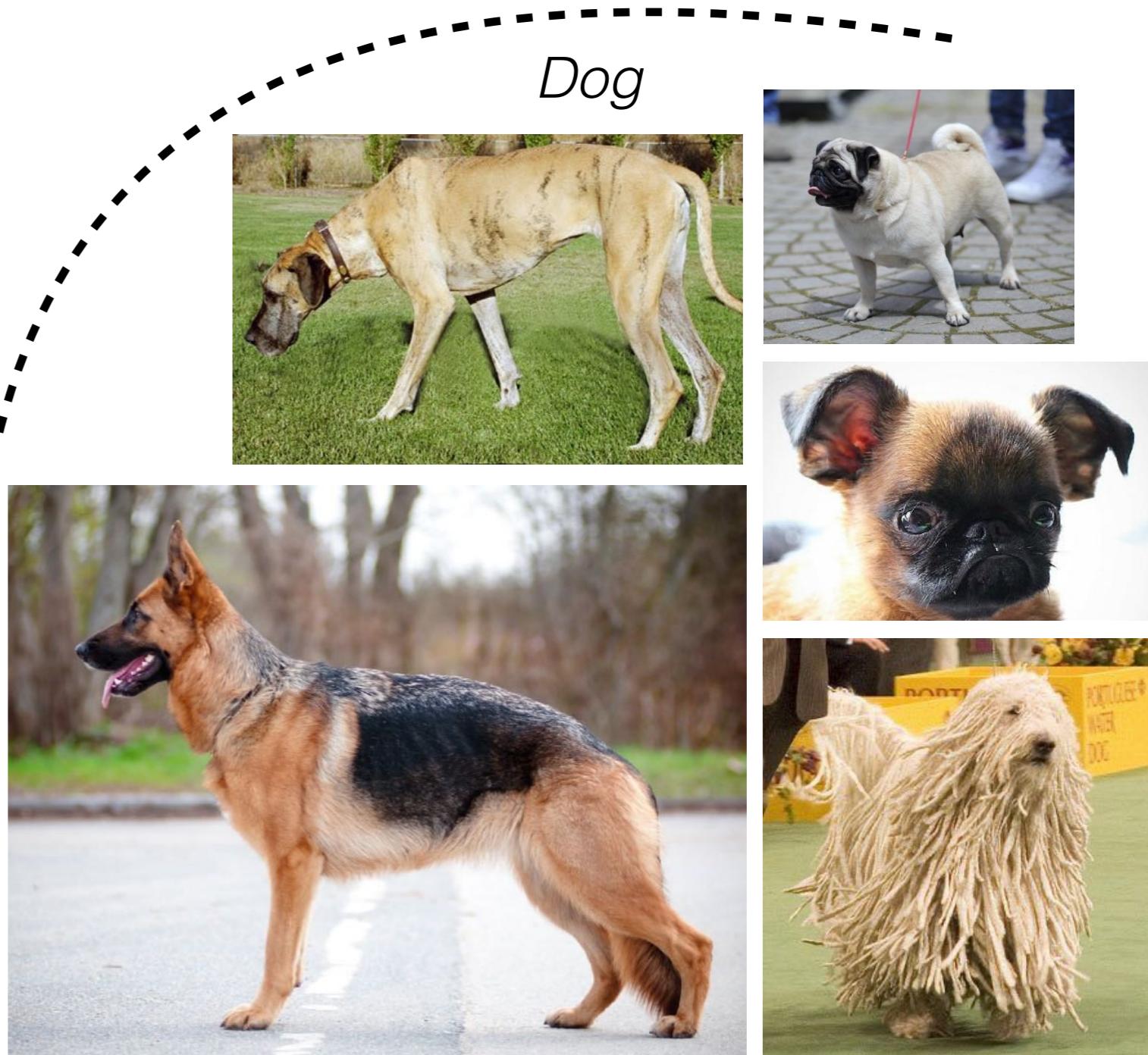
- Challenge (categorization): inter-class similarity

Wolf

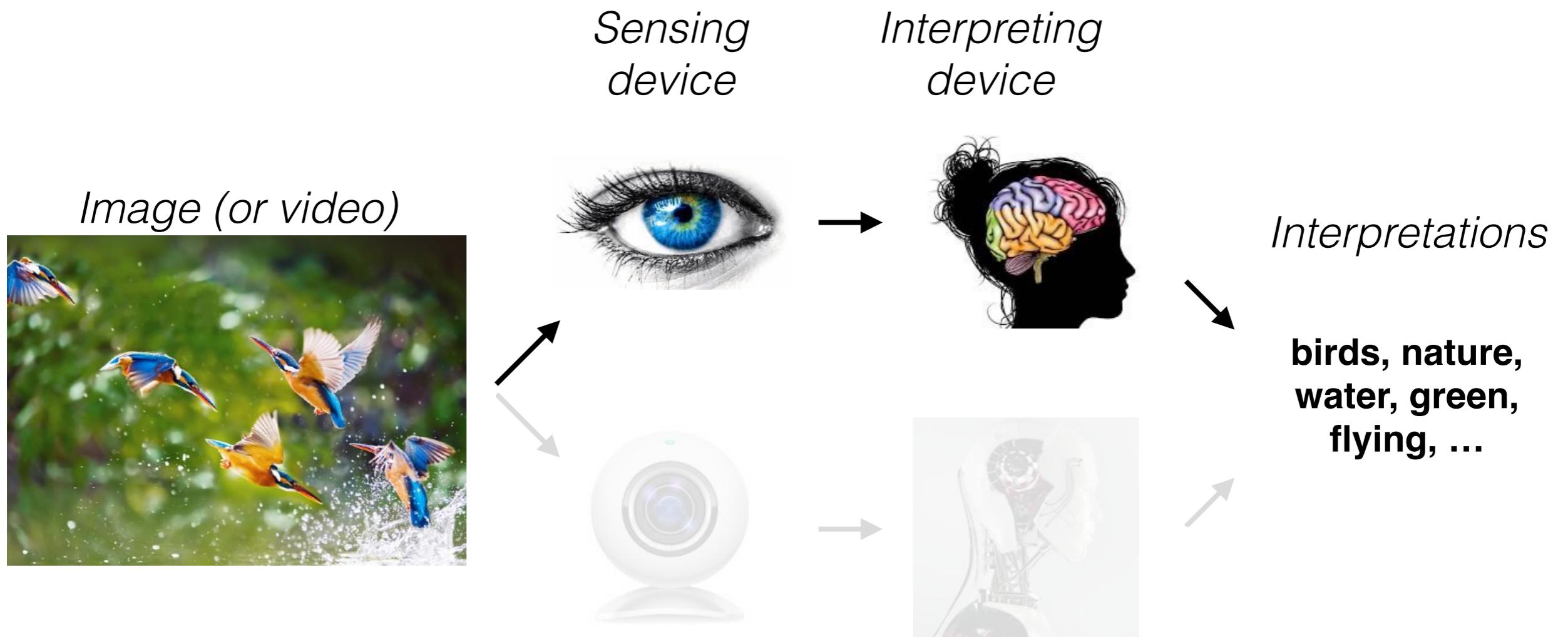


Copyright Jon Atkinson 2010

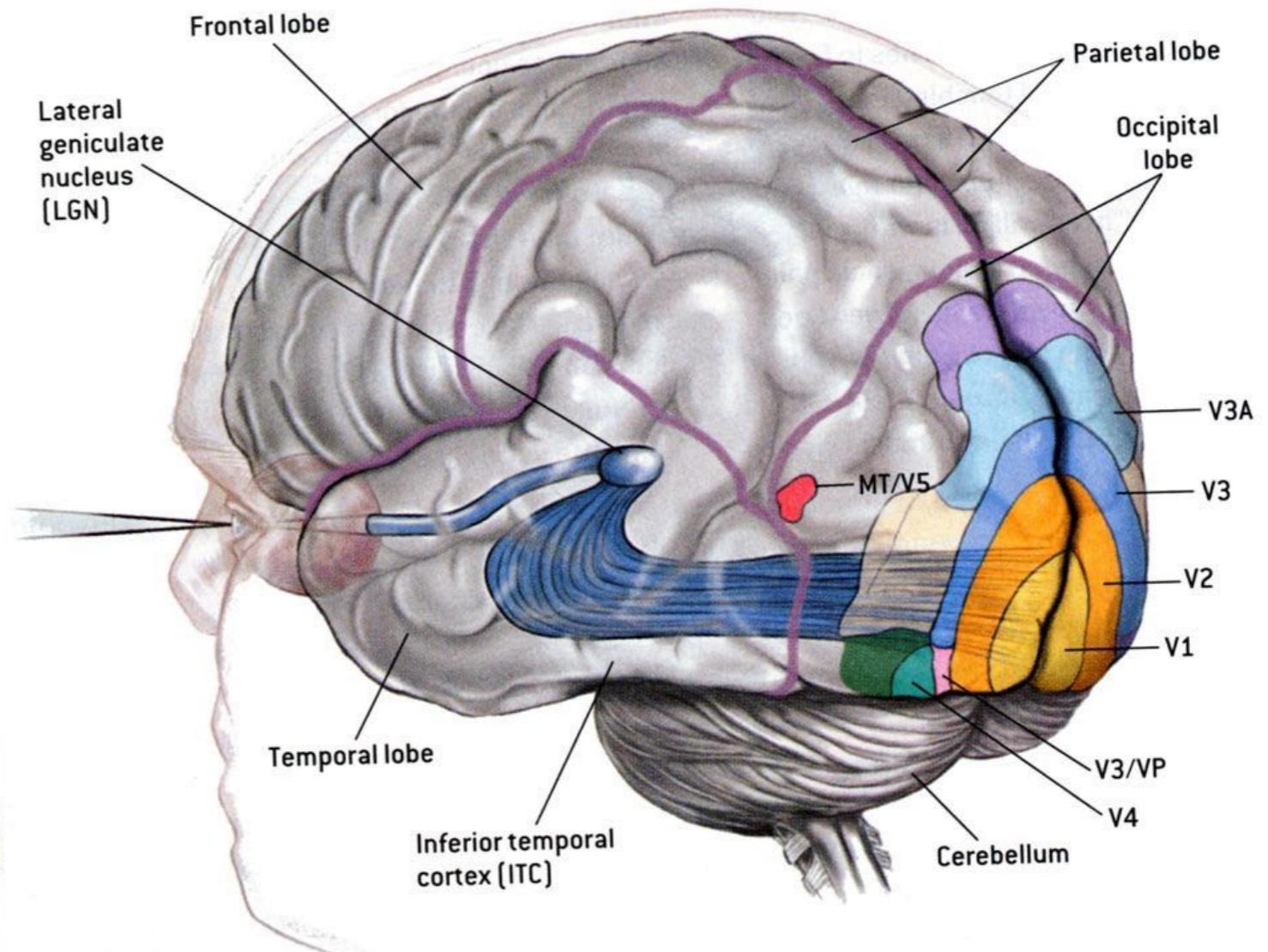
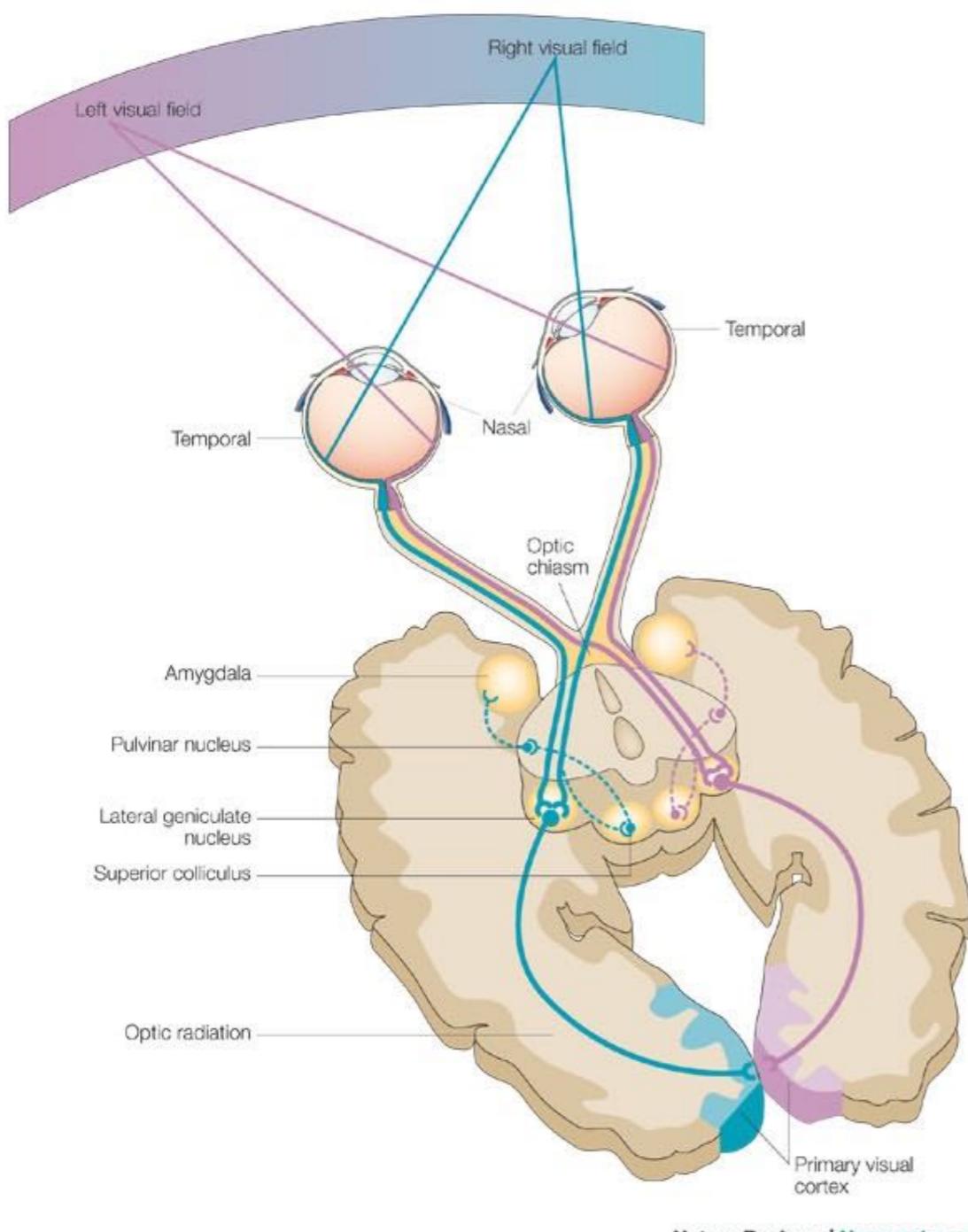
Dog



What is (computer) vision?



Human (mammalian) vision



Functional specialization:

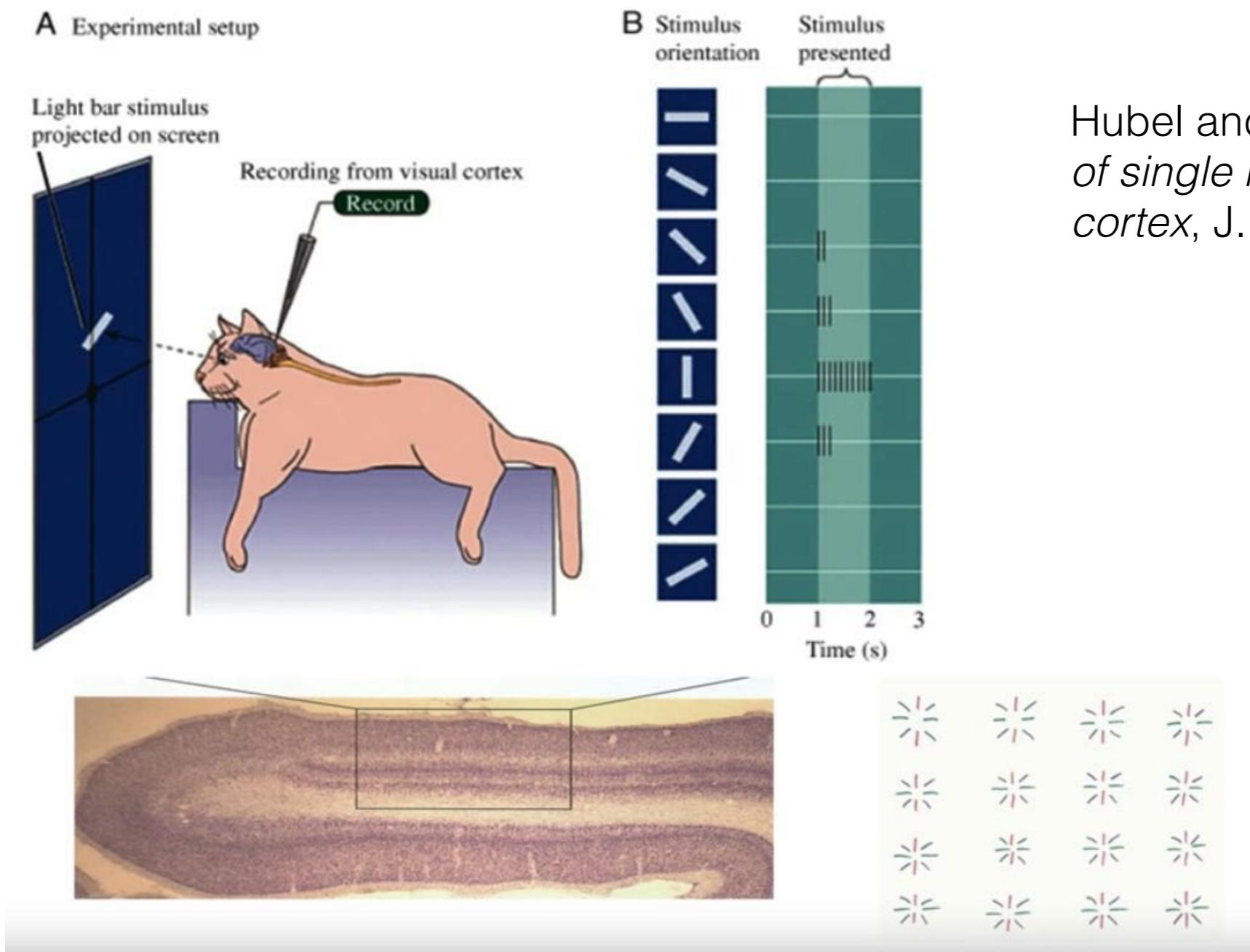
V1: primary visual cortex

V4: color, V3/VP: recognition

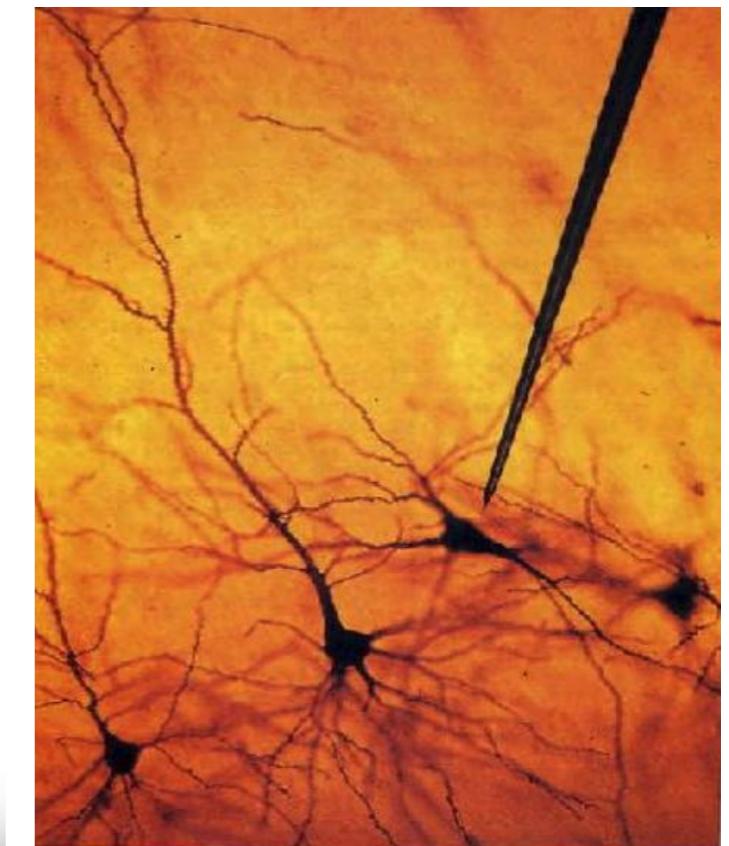
MT/V5: motion

Human (mammalian) vision

- 1981: Hubel & Wiesel won the Nobel prize in medicine



Hubel and Wiesel, *Receptive fields of single neurones in the cat's striate cortex*, J. Physiol 1959



Human (mammalian) vision

- 1981: Hubel & Wiesel won the Nobel prize in medicine

Hubel and Wiesel, *Receptive fields of single neurones in the cat's striate cortex*, J. Physiol 1959

Simple Cortical Cell

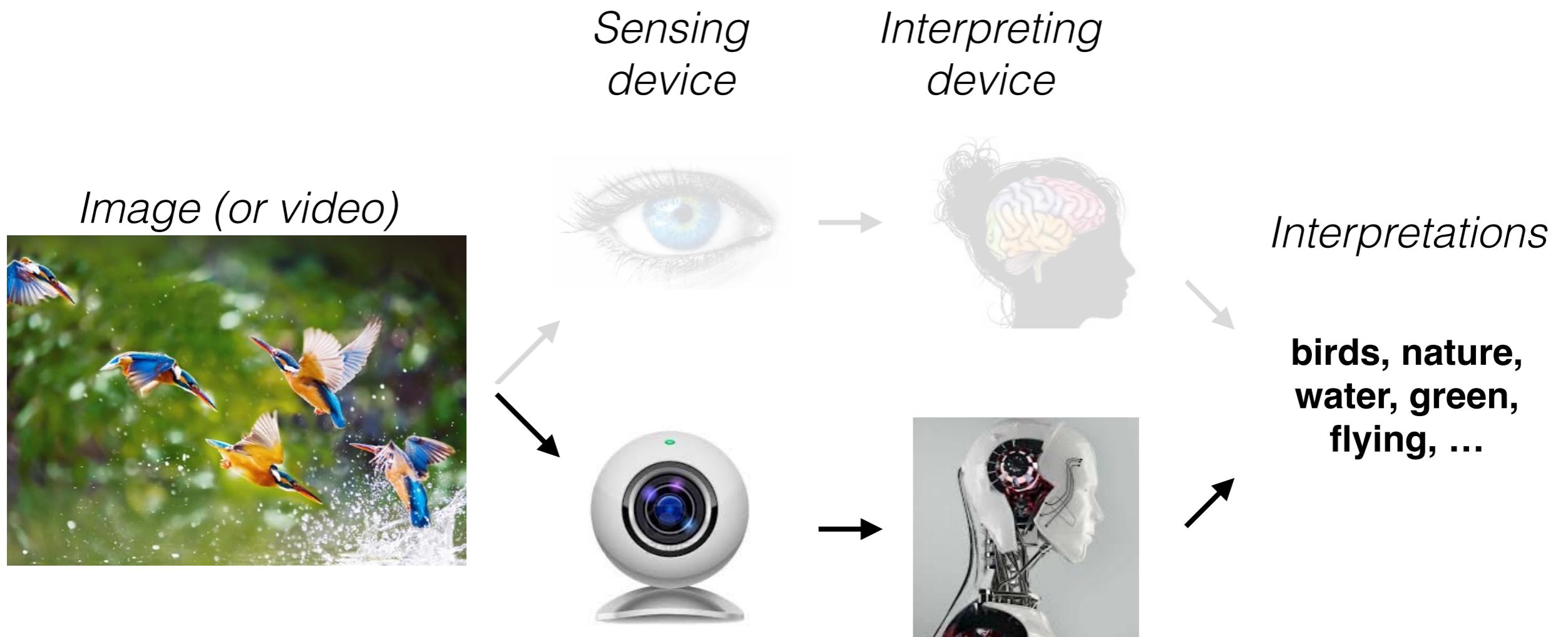
Human (mammalian) vision

- 1981: Hubel & Wiesel won the Nobel prize in medicine

Hubel and Wiesel, *Receptive fields of single neurones in the cat's striate cortex*, J. Physiol 1959

Simple Cortical Cell

What is computer vision?



The goal of computer vision

- To bridge the gap between pixels and meaning



What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2

What a computer sees

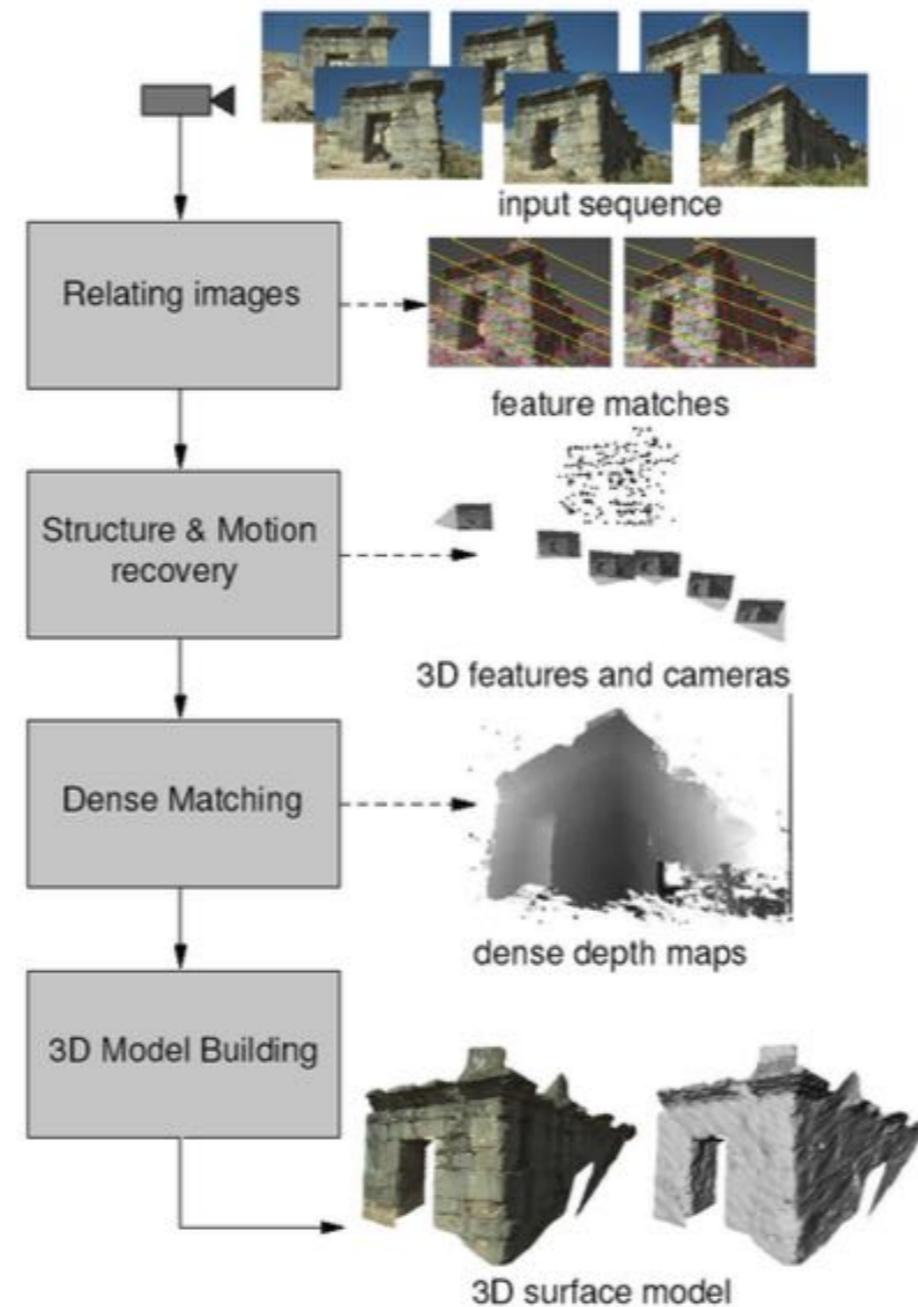
The goal of computer vision

- What kind of information can we extract from an image?
 - Metric 3D information
 - Semantic information



This is the focus of our course

Vision as measurement device

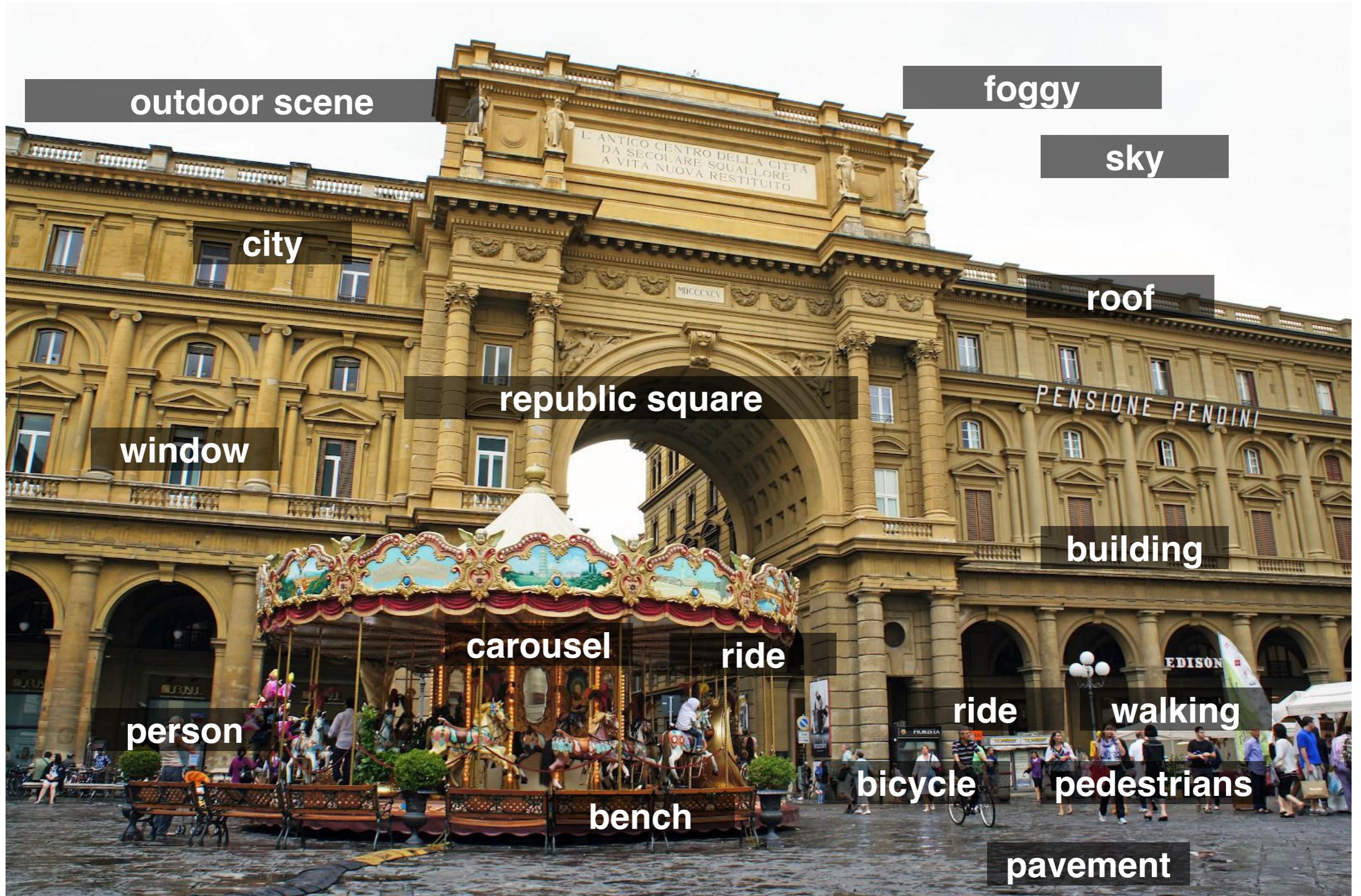


Pollefeys et al.



Goesele et al.

Vision as a source of semantic information



Origins of computer vision

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

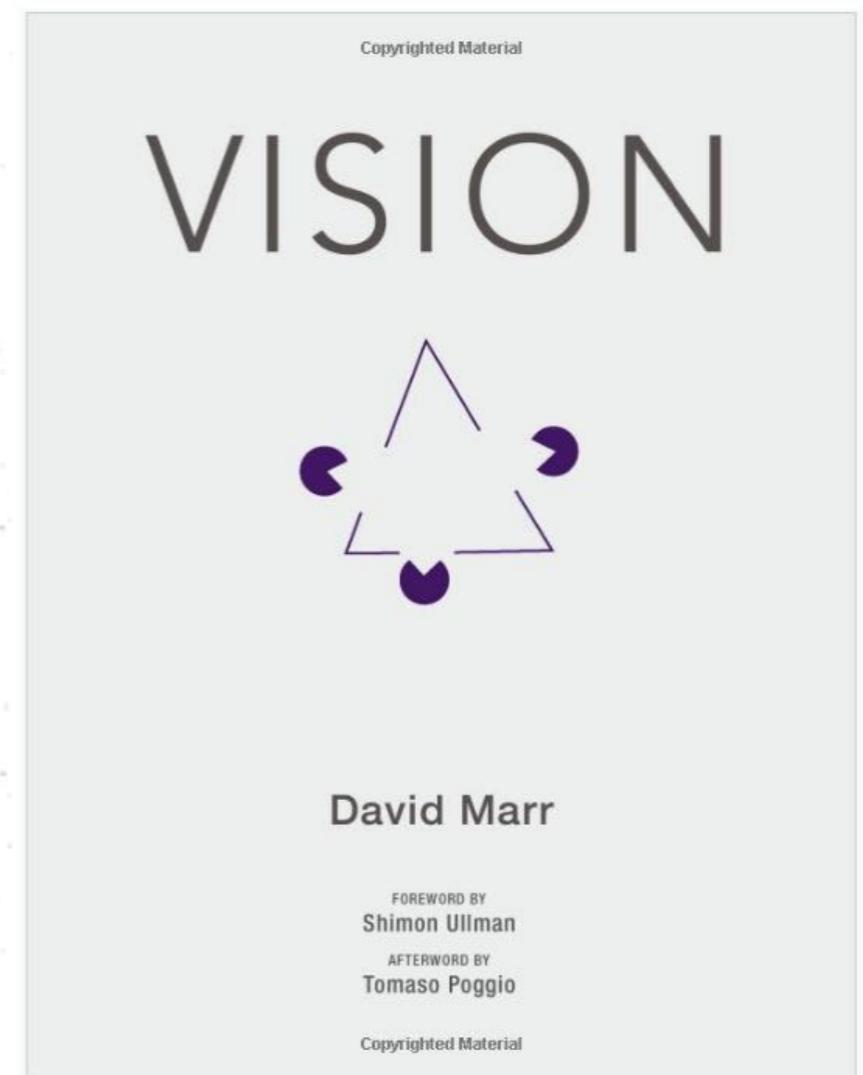
July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

CV “started” as an undergraduate summer project at MIT



David Marr’s influential work in the 1970s

A brief history of computer vision

1966: Marvin Minsky assigns computer vision as an undergrad summer project

1970s: interpretation of synthetic worlds and carefully selected images

1980s: shift towards geometry and increased mathematical rigor

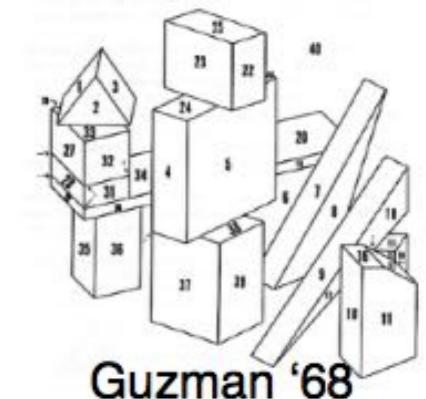
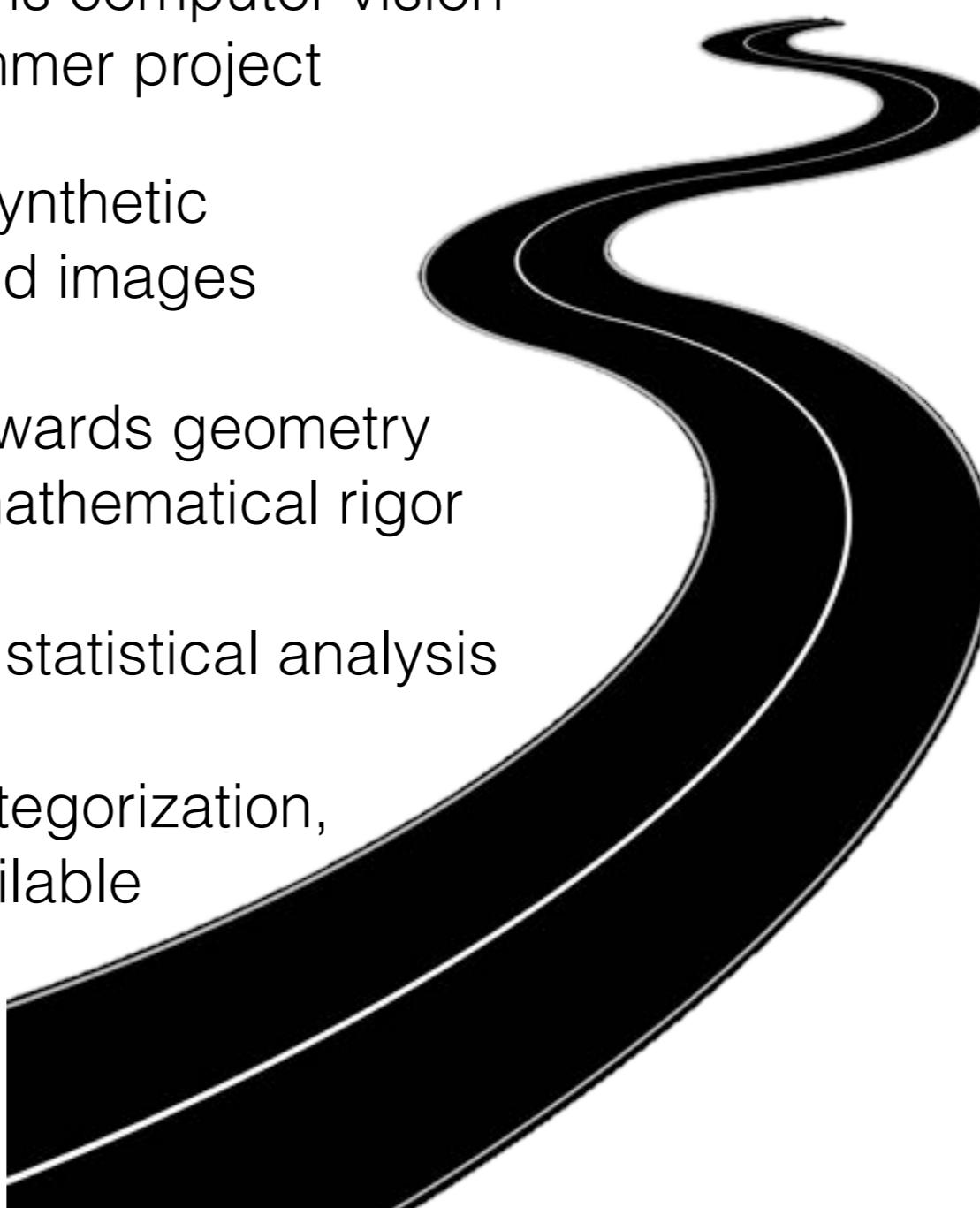
1990s: face recognition, statistical analysis

2000s: object recognition, categorization, annotated datasets available

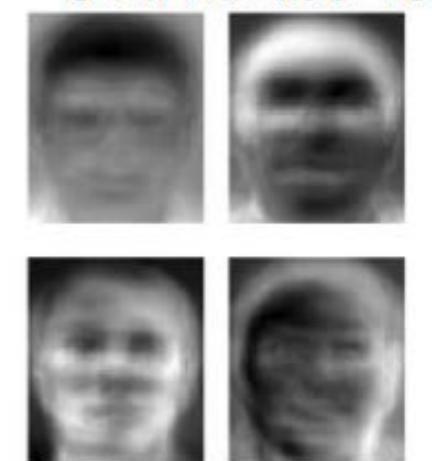
2010s: large-scale visual recognition, visual intelligence

2020s: ???

(big/foundation models, ...)



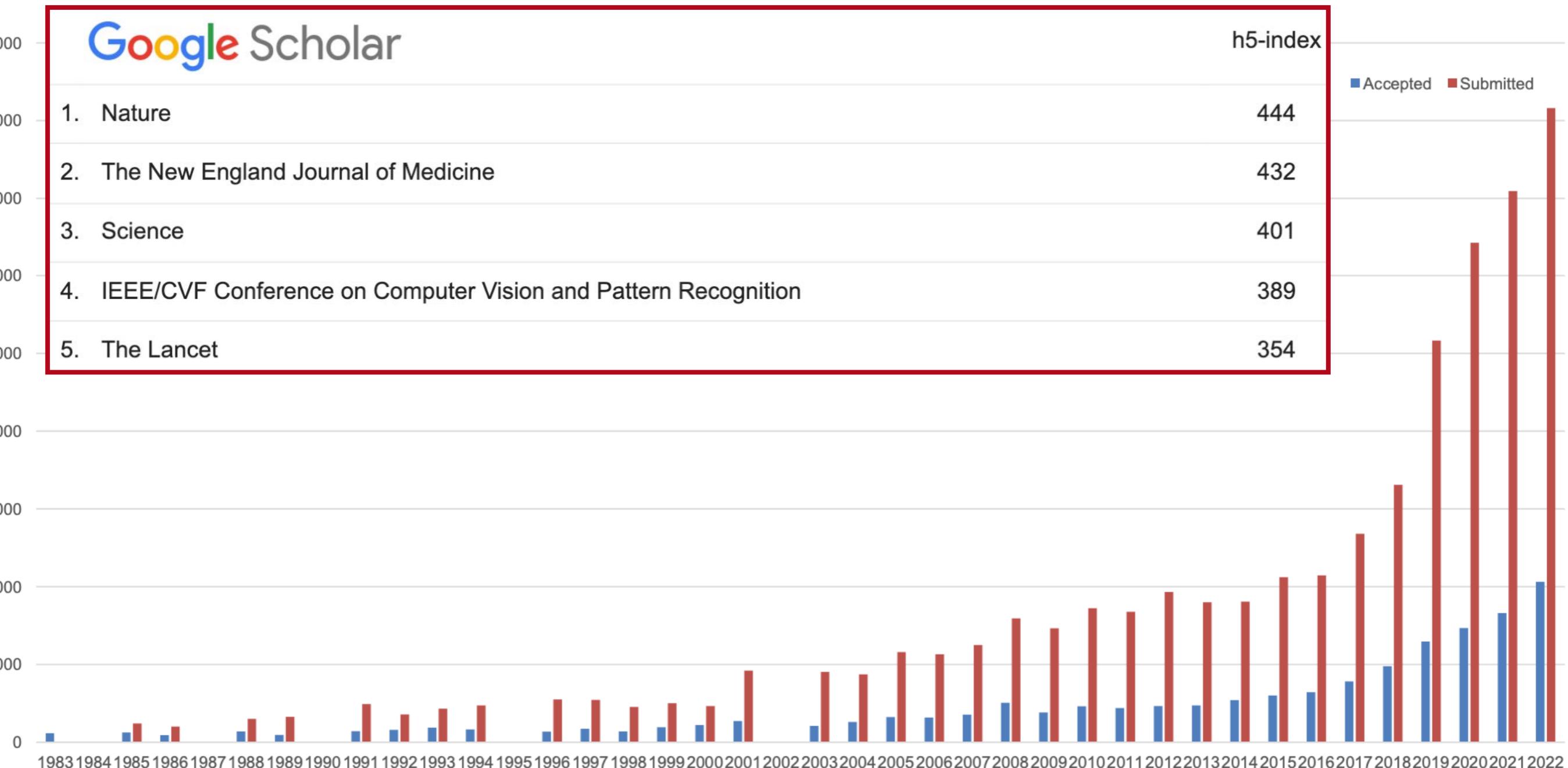
Ohta Kanade '78



Turk and Pentland '91

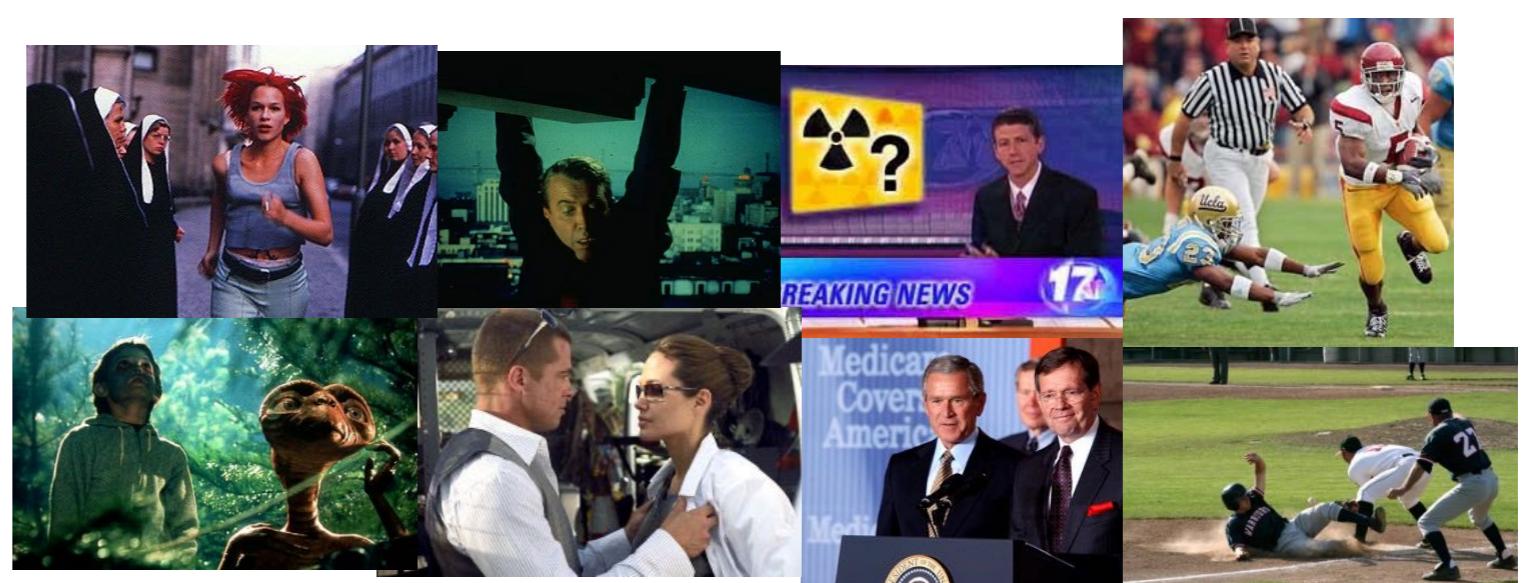
A brief history of computer vision

- Academic perspective (CVPR year-over-year-growth):



Why study computer vision?

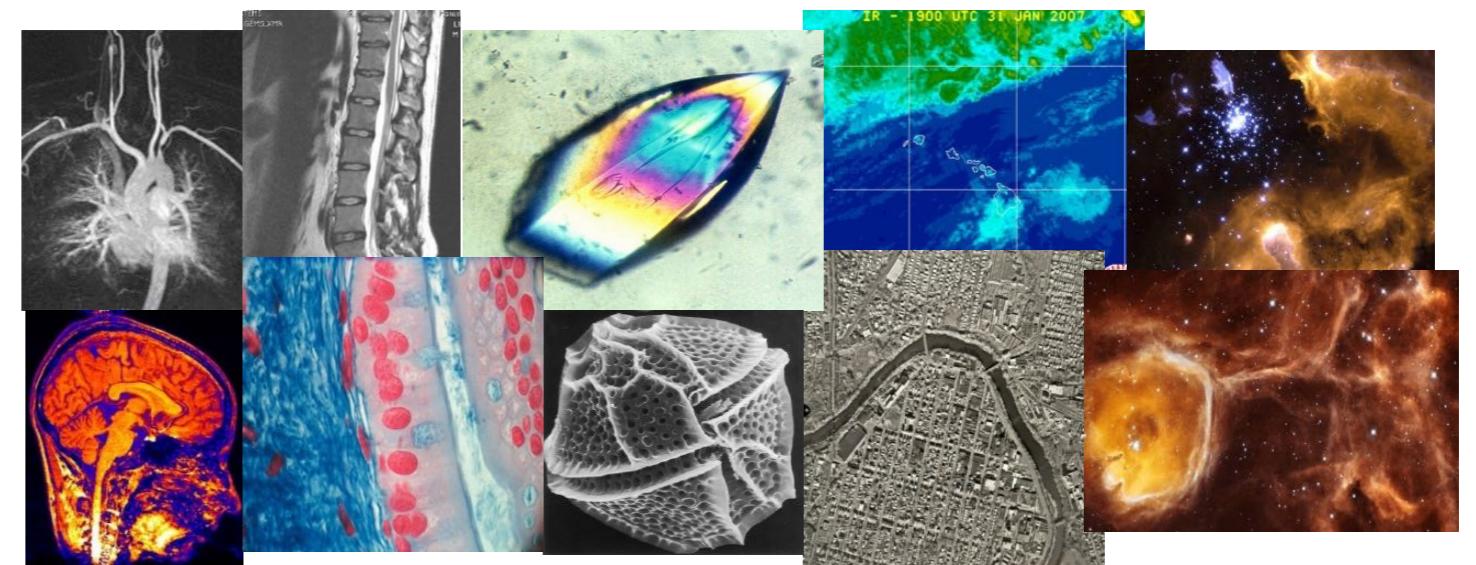
- Big visual data: e.g. almost 90% of web data is visual



Google Photos

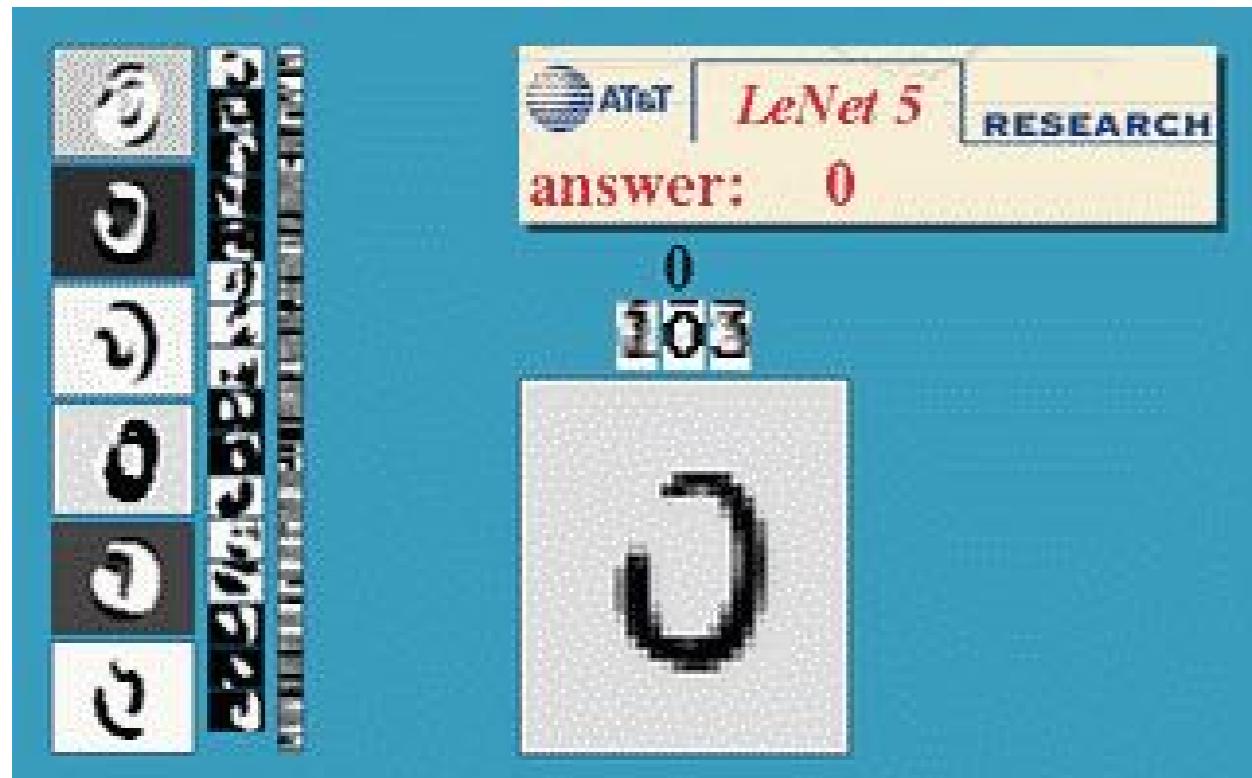


Surveillance and security



Medical and scientific images

Applications: optical character recognition

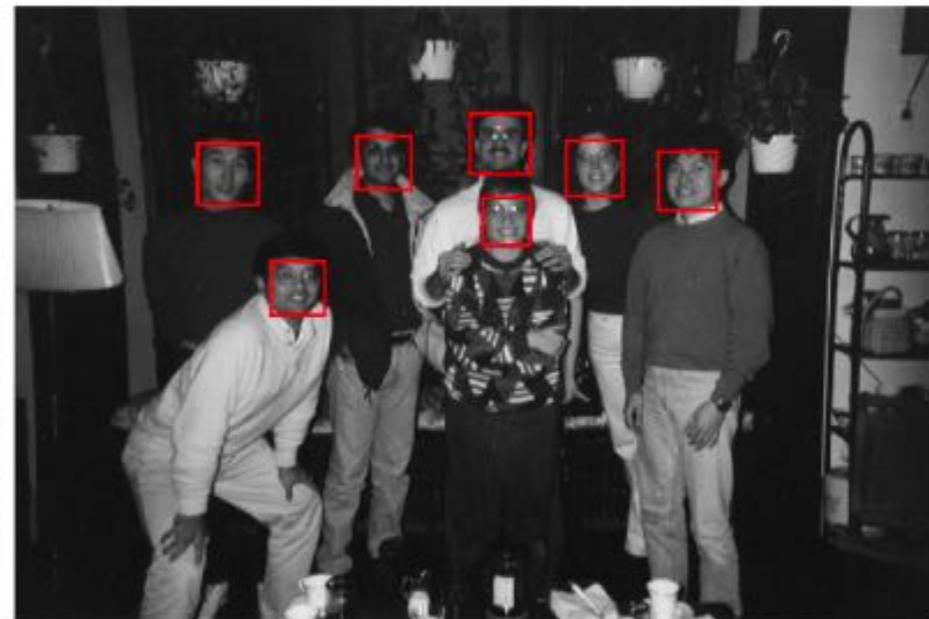
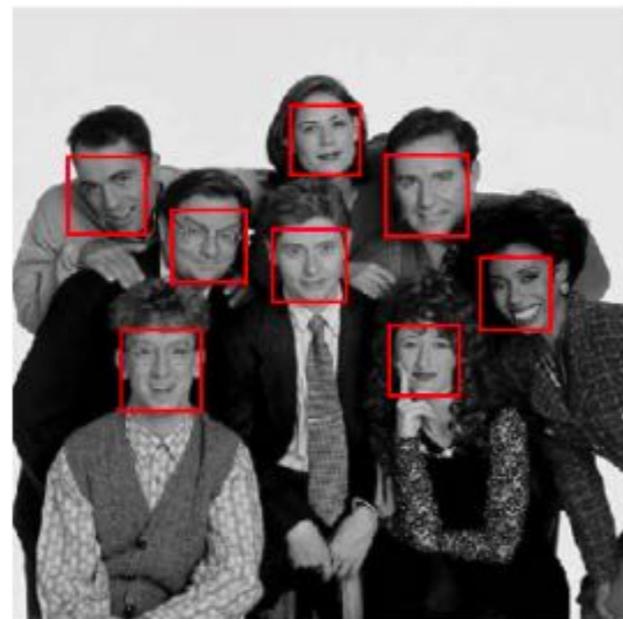
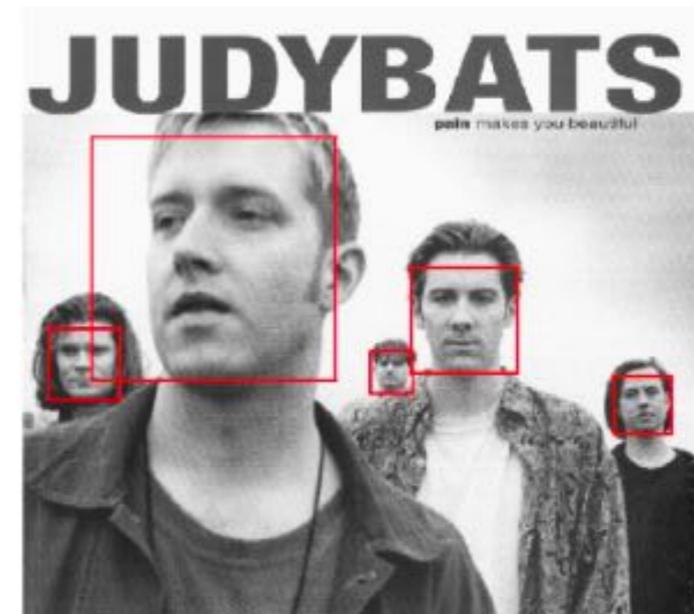
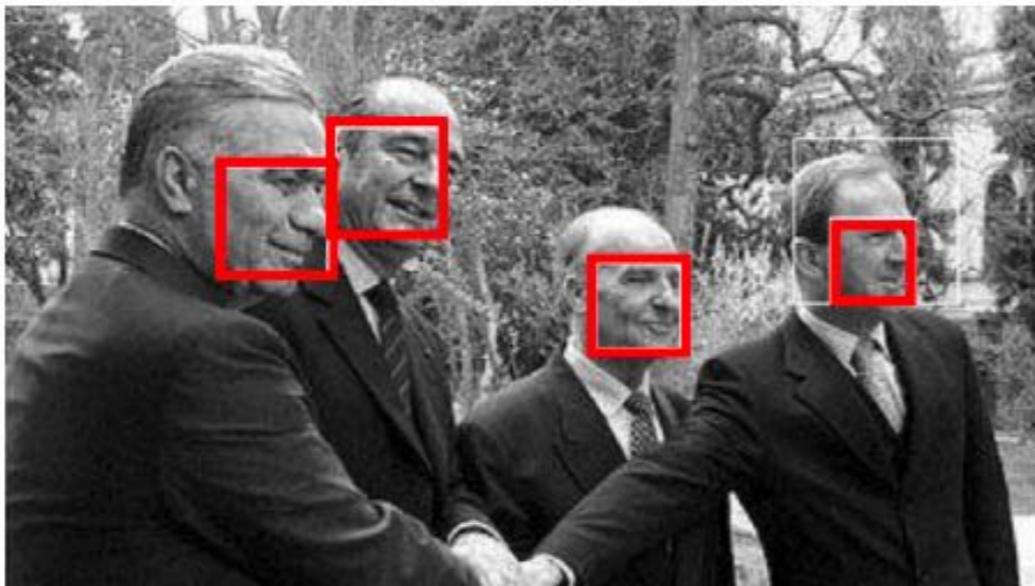


Digit recognition, AT&T labs



License plate readers

Applications: face detection



FinePix S6000fd, by Fujifilm, 2006

Viola & Jones, 2001

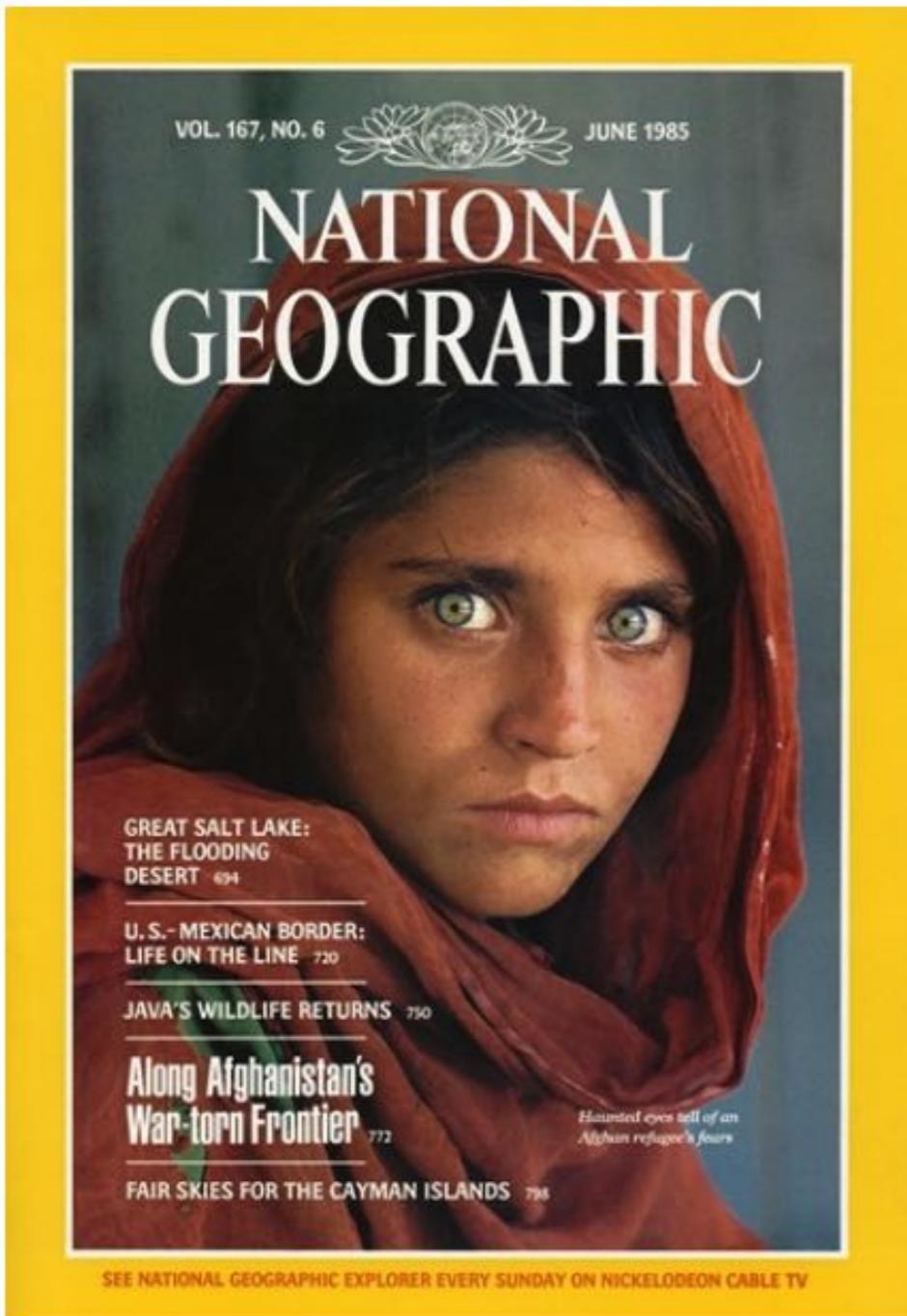
Applications: face recognition

Face ID

Your face is your secure password.



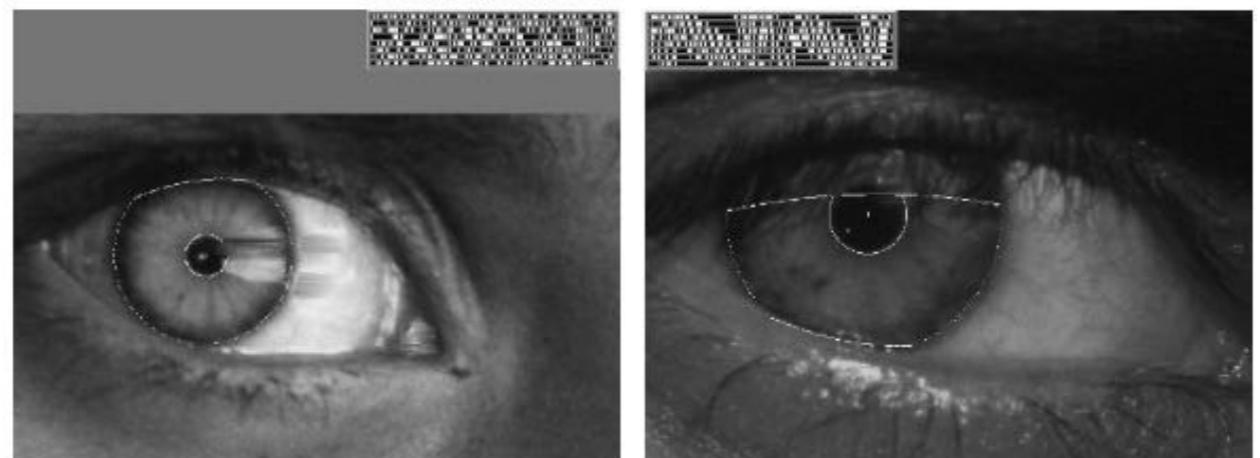
Applications: biometrics



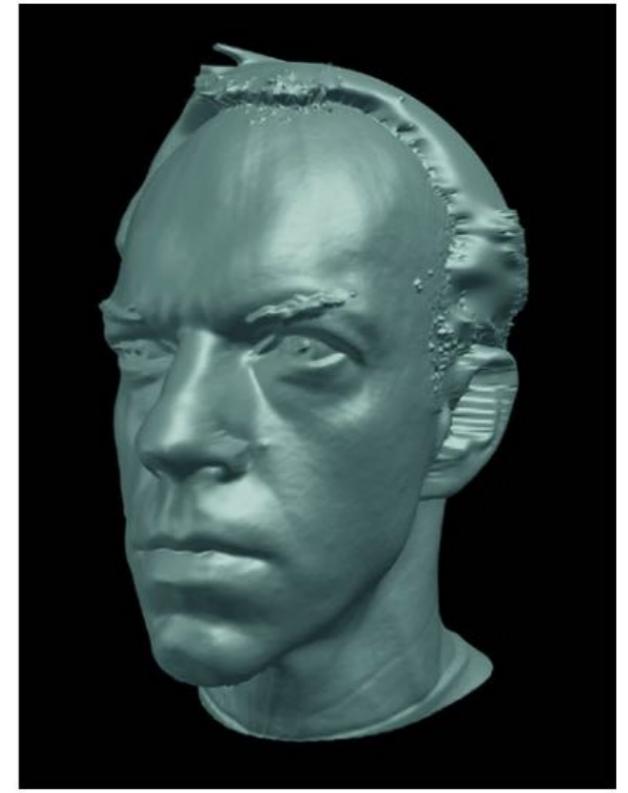
Who is she?



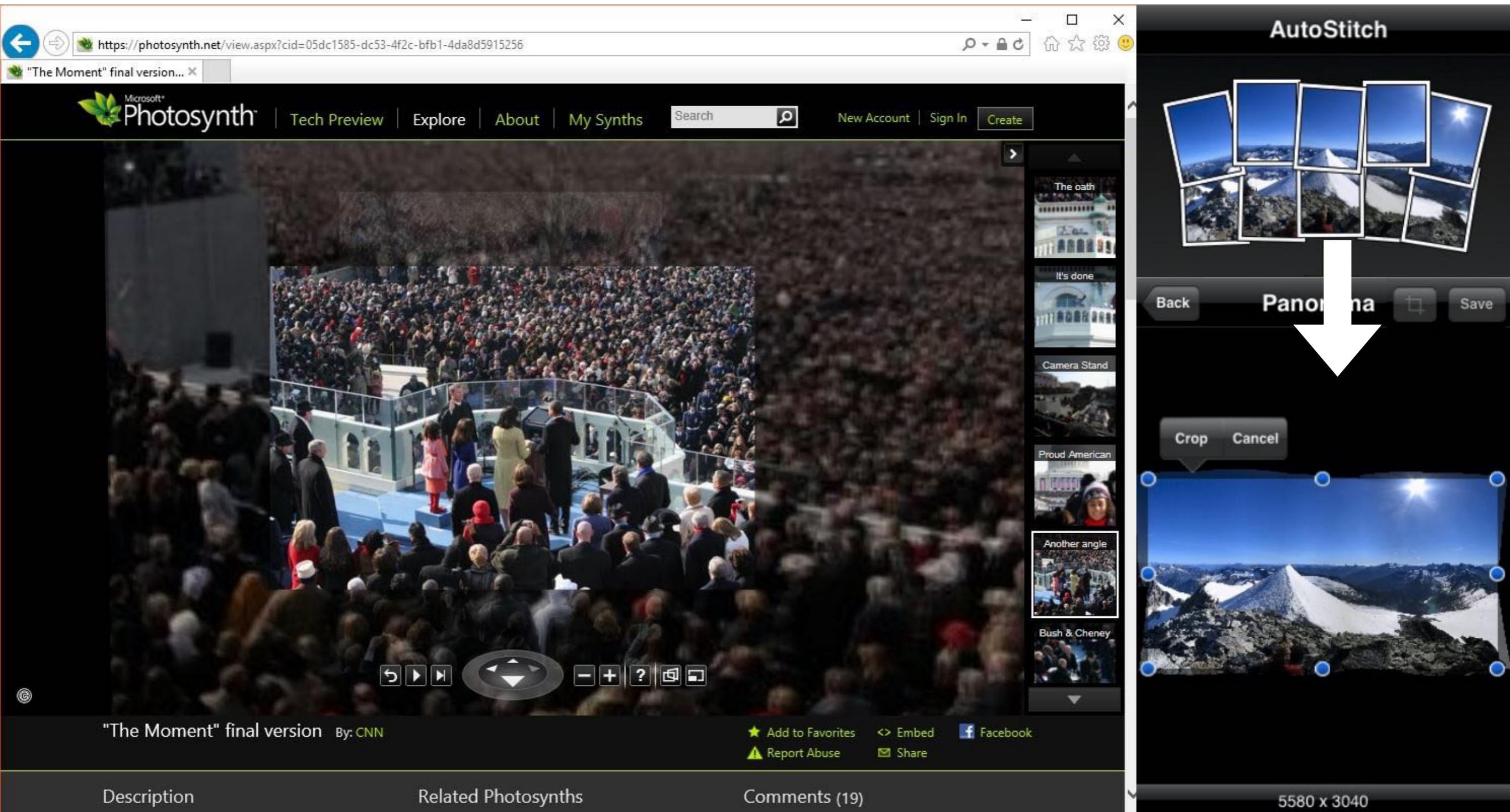
"How the Afghan Girl was Identified by Her Iris Patterns" [[read the story](#)]



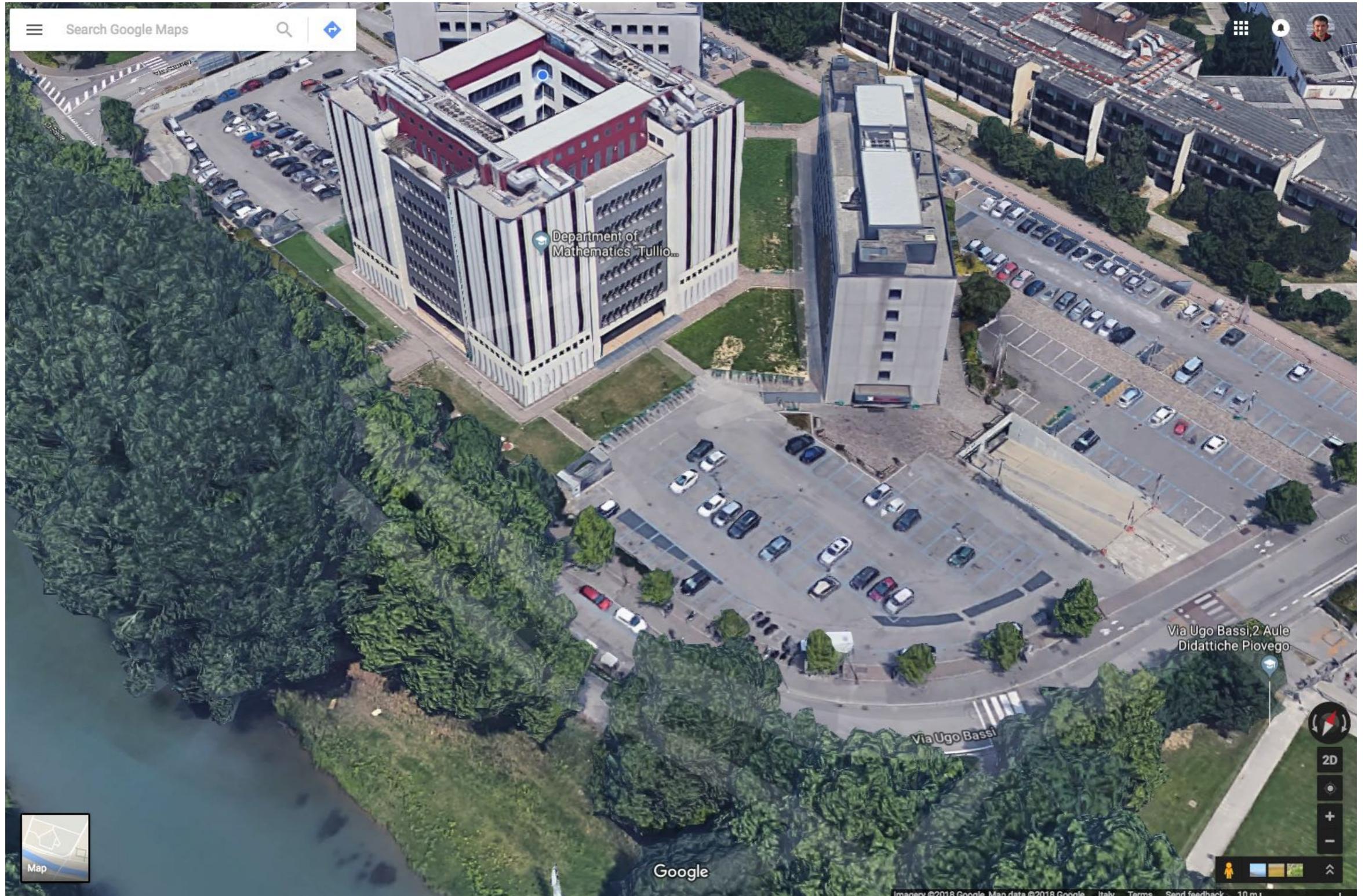
Applications: motion capture



Applications: photo stitching



Applications: 3D urban modeling



Applications: robotics



[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision system used for several tasks:

- Panorama stitching
- 3D terrain modeling
- Obstacle detection and position tracking

For more info read “computer vision on Mars”, Matthies et al.

Applications: interaction

Motion sensing for videogames
(Microsoft's Kinect)



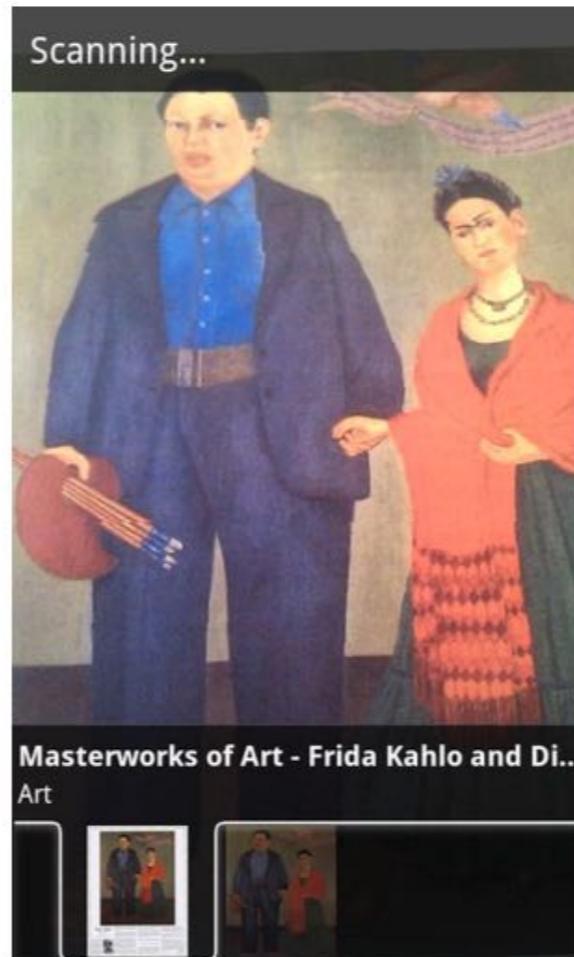
Assistive technologies



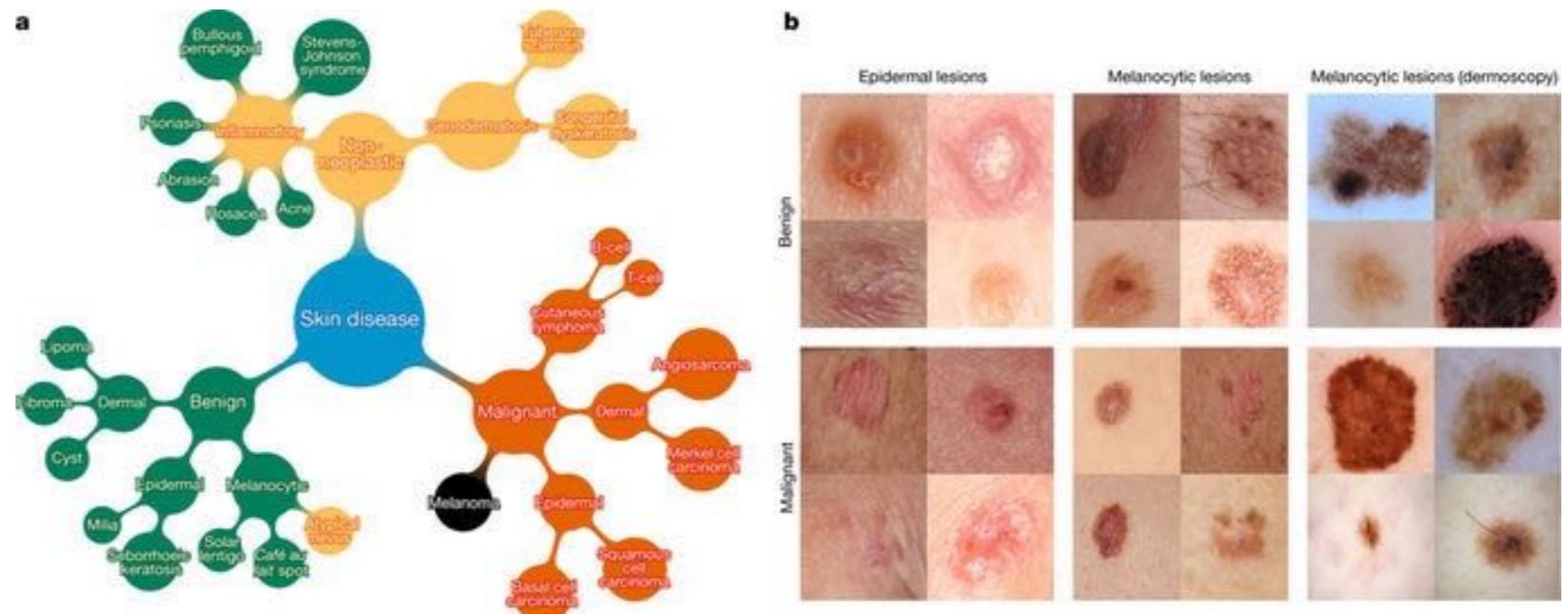
Applications: visual search



Google Goggles



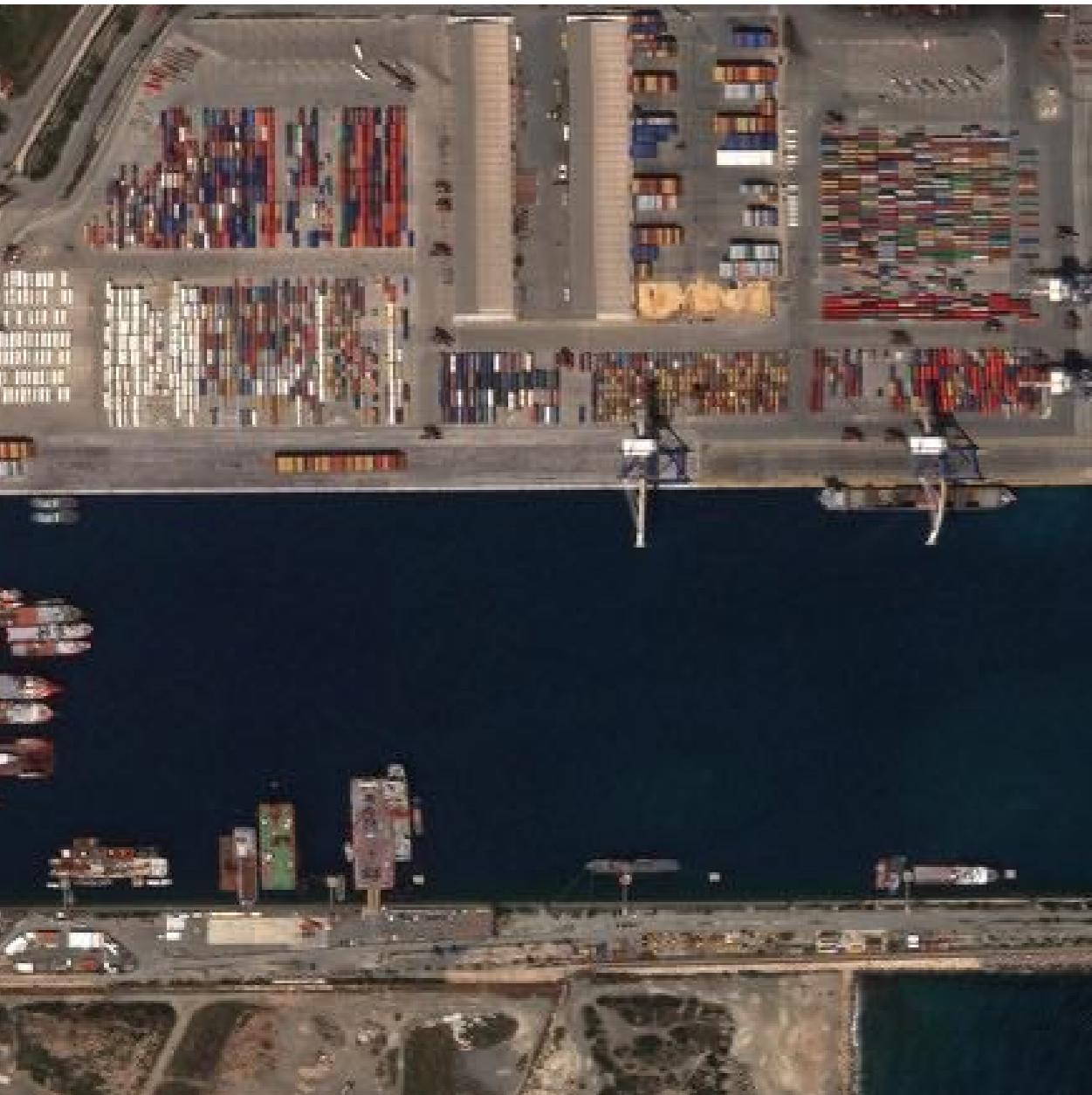
Applications: healthcare



Dermatologist-level Classification of Skin Cancer with Deep Neural Networks

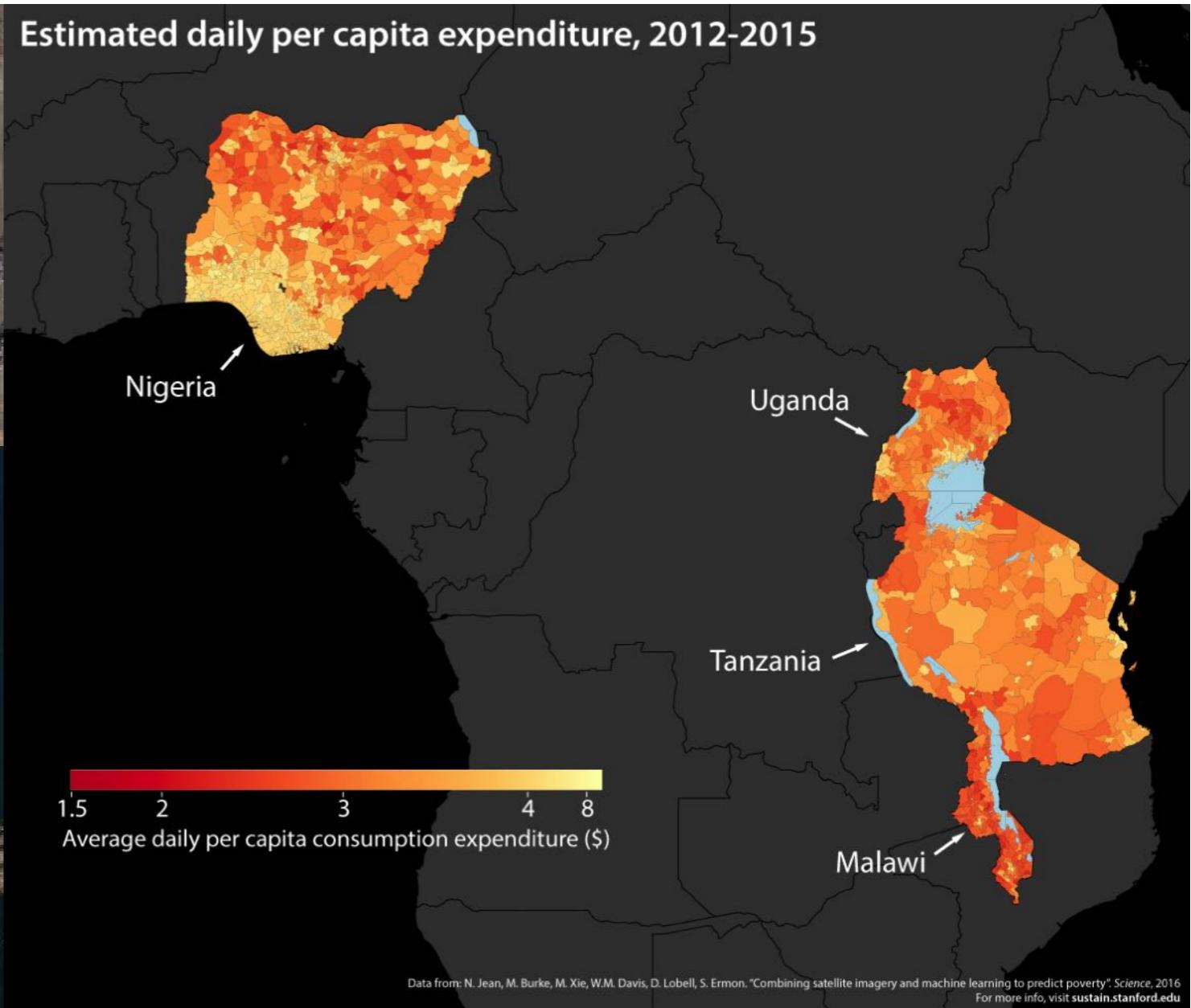
Applications: analysing satellite images

Security and defense



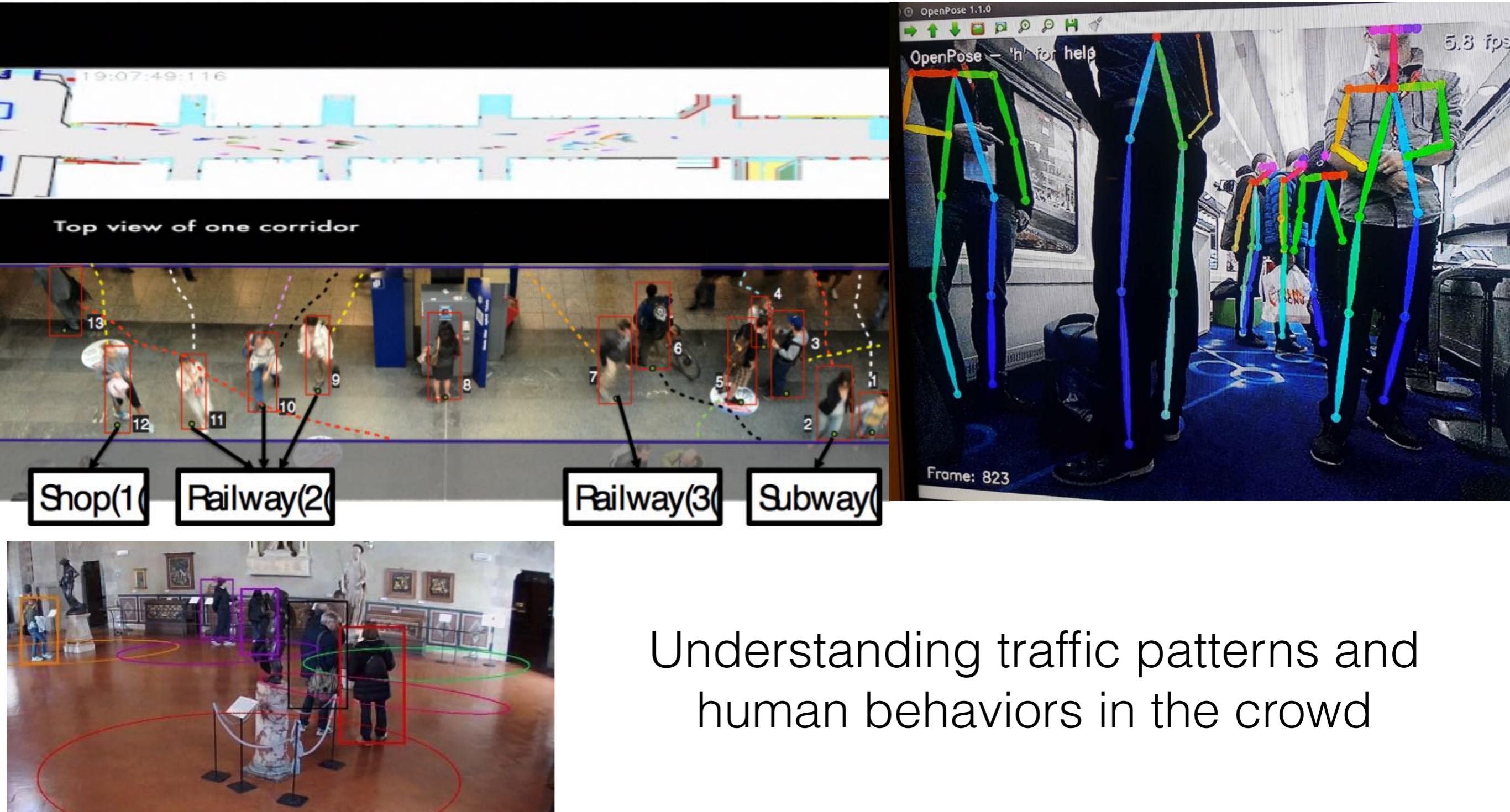
Pentagon: \$100,000 in prizes to develop algorithms that can interpret high-resolution satellite images

Environmental challenges



Combining Satellite Imagery and Machine Learning to Predict Poverty, *Science* 2016 (S. Ermon)

Applications: intelligent mobility



Applications: city-scale surveillance



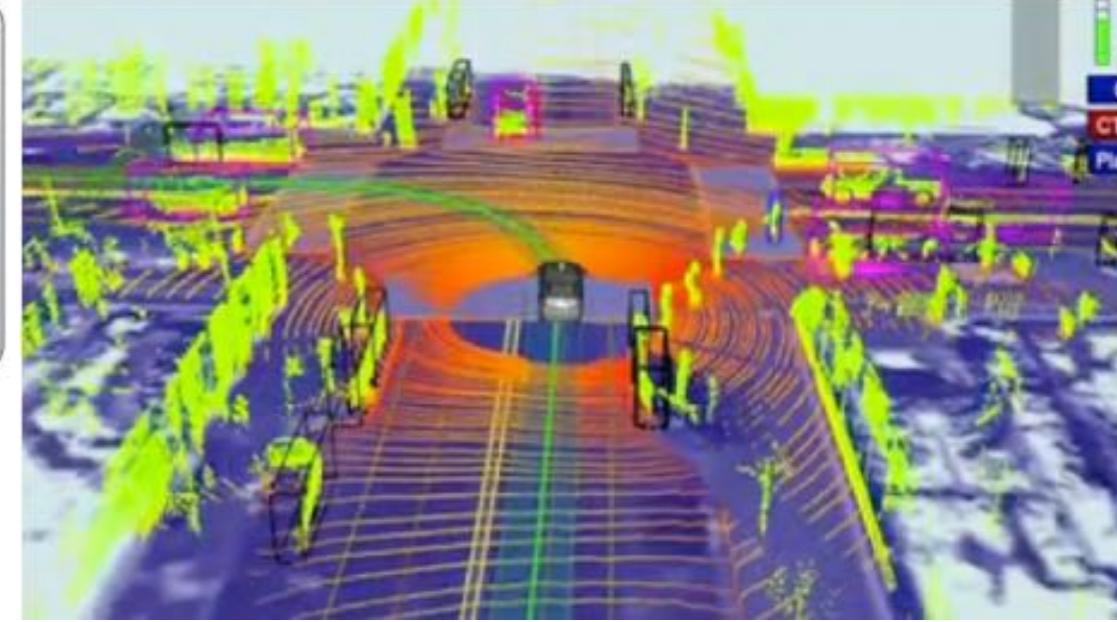
Applications: retail & supermarket 2.0



Applications: self-driving cars



A screenshot of a website page featuring three main sections: 'EyeQ Vision on a Chip' (with an image of a chip), 'Vision Applications' (with an image of a pedestrian crossing), and 'AWS Advance Warning System' (with an image of a screen displaying a car icon). Each section has a 'read more' link.



- Mobileye: vision system in high-end BMW, GM, Volvo Models
- All major companies and several startups

Cognitive services: vision

- Image tagging (classification)
- Face detection
- Landmark/logo detection
- Image properties
- OCR
- Image search / retrieval
- Inappropriate content detection
- Activity detection
- People tracking

The screenshot shows the Google Cloud Vision API landing page. At the top, there's a navigation bar with links like Why Google, Products (which is selected), Solutions, Launcher, Pricing, Security, Customers, Documentation, Support, Partners, TRY IT FREE, and CONTACT SALES. Below the navigation, there's a section titled "Powerful Image Analysis" with a description of the Google Cloud Vision API's capabilities. To the right of this text is a 3D diagram illustrating object detection and text recognition. Below this, there's a "Try the API" section with a dashed box for dragging an image file or browsing from a computer. To the right of this box is another dashed box containing a small image of a mountain. Further down, there's a section titled "Insight From Your Images" with a list of detected objects: CARS, FLOWERS, RABBITS, MOUNTAINS, and BIRDS. Next to this is a small icon of a person with a list of detected categories. To the right of this is a section titled "Detect Inappropriate Content" with a description of the SafeSearch feature. Below this is a section titled "Power of the Web" with a description of how the Vision API uses Google Image Search to find topical entities. At the bottom, there's a section titled "Extract Text" with a description of Optical Character Recognition (OCR) and its capabilities.

What are we going to learn?

- Focus on high-level visual recognition tasks
- Course (1st part) outline:
 - ▶ Foundations: images, convolution and filtering
 - ▶ Foundations: feature detectors and descriptors
 - ▶ Intro to visual recognition, classification, detection

A brief history of computer vision

1966: Marvin Minsky assigns computer vision as an undergrad summer project

1970s: interpretation of synthetic worlds and carefully selected images

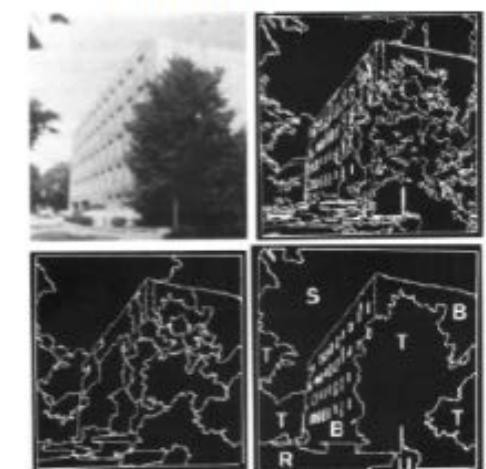
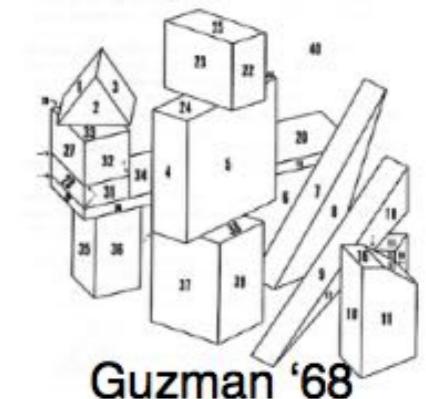
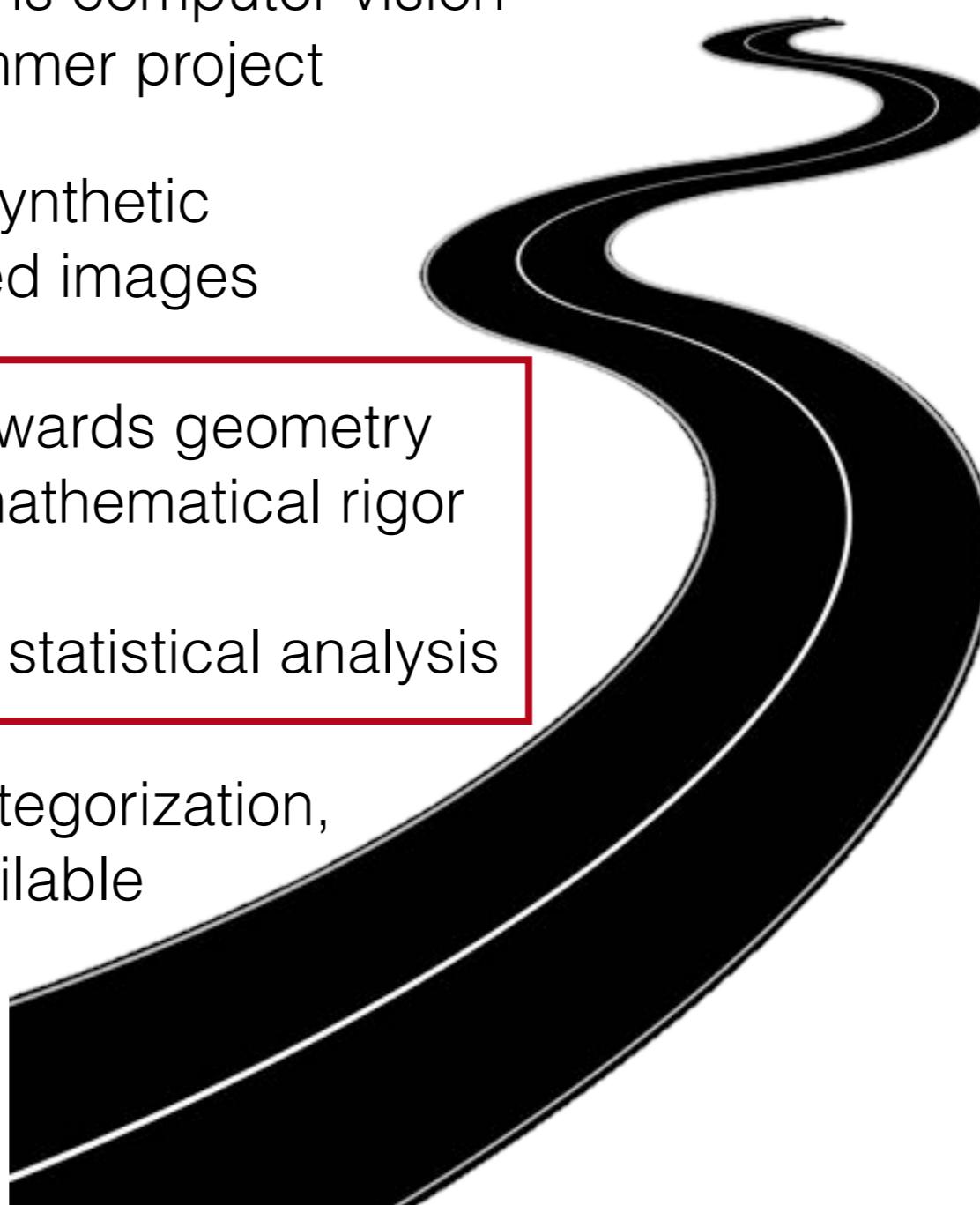
1980s: shift towards geometry and increased mathematical rigor

1990s: face recognition, statistical analysis

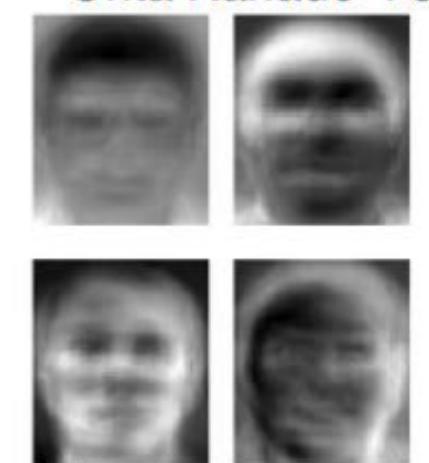
2000s: object recognition, categorization, annotated datasets available

2010s: large-scale visual recognition, visual intelligence

2020s: ???



Ohta Kanade '78



Turk and Pentland '91

Coming up

- **Next lecture(s):**

- ▶ Foundations: images, pixels, convolution, filtering

